

**PEMANFAATAN DATA MINING DALAM PERANCANGAN
STRATEGI PEMASARAN: STUDI KASUS GROBBAK KOREA
NOLOGATEN**

TUGAS AKHIR

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



Nama : M. Aunur Rahim Ghaffar Tuasikal
No. Mahasiswa : 21522292

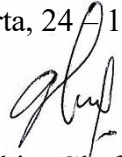
**PROGRAM STUDI TEKNIK INDUSTRI PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2025**

PERNYATAAN KEASLIAN

Saya mengakui bahwa tugas akhir yang berjudul “Pemanfaatan Data Mining Dalam Perancangan Strategi Pemasaran: Studi Kasus Grobbak Korea Nologaten” 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.



rta, 24/10 - 2025


Muhim Ghaffar Tuasikal)
21522292

SURAT BUKTI PENELITIAN



**FAKULTAS
TEKNOLOGI INDUSTRI**

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

SURAT KETERANGAN PENELITIAN

Nomor: 15/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 Aunur Rahim Ghaffar Tuasikal

No. Mhs : 21522292

Dosen Pembimbing : Annisa Uswatun Khasanah, S.T., M.Sc

Telah selesai melaksanakan penelitian yang berjudul "Pemanfaatan Data Mining Dalam Perancangan Strategi Pemasaran: Studi Kasus Grobbak Korea Nologaten" di Laboratorium Data Mining, Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia tercatat mulai tanggal 7 Mei sampai dengan tanggal 6 Juni 2025.

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

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Yogyakarta 10 September 2025

Kepala Laboratorium
Data Mining

Dr. Dwi Adi Purnama, S.T.

**LEMBAR PENGESAHAN PEMBIMBING
PEMANFAATAN DATA MINING DALAM PERANCANGAN
STRATEGI PEMASARAN: STUDI KASUS GROBBAK KOREA
NOLOGATEN**

TUGAS AKHIR

Disusun Oleh :

Nama : M. Aunur Rahim Ghaffar Tuasikal

No. Mahasiswa : 21522292



Yogyakarta, 09 September 2025

Dosen Pembimbing



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

**LEMBAR PENGESAHAN DOSEN PENGUJI
PEMANFAATAN DATA MINING DALAM PERANCANGAN
STRATEGI PEMASARAN: STUDI KASUS GROBBAK KOREA
NOLOGATEN**

TUGAS AKHIR

Disusun Oleh :

**Nama : M. Aunur Rahim Ghaffar Tuasikal
No. Mahasiswa : 21522292**

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

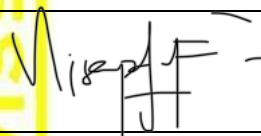
Yogyakarta, 01 - Oktober – 2025

Tim Penguji

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

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

Chancard Basumerda, S.T., M.Sc. 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

Atas seizin Allah SWT atas nikmat yang telah beliau berikan, tugas akhir ini saya persembahkan kepada:

Bapak Rusydi Tuasikal dan Ibu Dini Wulandari, kedua orang tua saya yang menjadi *support system*, mengasuh dan mendidik saya sedari kecil. Lalu kepada kakak, adik, dan teman-teman penulis yang menjadi pendukung atas tersusunnya laporan tugas akhir ini.

MOTTO

“Hope is a Good Thing, Maybe the Best of Thing, and No Good Thing Ever Dies.”

(Andy Dufresne)

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabaraakaatuh

Alhamdulillahirabbil'alamiin, segala puji dan syukur penulis sampaikan kepada Allah SWT atas berkah rahmat serta nikmat-Nya sehingga penelitian dan penyusunan laporan Tugas Akhir dengan judul “Peningkatan Produktivitas *Fruit Dehydrator* Pada Pengolahan Buah Kering Dengan Sistem *Vacuum* Menggunakan Metode *Craig Harris*” sebagai salah satu persyaratan untuk memperoleh gelar sarjana. Tidak lupa shalawat serta salam tercurah kepada junjungan kita Nabi Muhammad SAW yang telah membimbing manusia dari zaman kegelapan hingga ke zaman yang terang dan penuh dengan ilmu.

Laporan Tugas Akhir ini penulis susun berdasarkan hasil penelitian yang dilaksanakan selama kurang lebih enam bulan. Seluruh hasil yang disajikan dalam laporan ini merupakan data asli yang diperoleh dari proses uji coba secara langsung di lapangan, sehingga dapat dipertanggungjawabkan secara ilmiah dan relevan dengan kondisi sebenarnya. Selama kegiatan Penelitian dan pengerjaan Laporan Tugas Akhir, penulis sadar bahwa semua tidak akan berjalan lancar tanpa bimbingan dan dorongan dari semua pihak.

Pada kesempatan kali ini, penulis ingin mengucapkan teirma kasih dan penghargaan yang setinggi-tingginya kepada semua pihak yang telah berjasa memberikan bimbingan dan motivasi dalam menyelesaikan laporan ini. Maka dari itu, penulis ingin 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 Dr. Drs. Imam Djati Widodo, M.Eng.Sc., selaku Ketua Jurusan Teknik Industri FTI UII.
3. Bapak Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM, selaku Ketua Program Studi Teknik Industri program sarjana FTI UII.
4. Ibu Annisa Uswatun Khasanah, S.T., M.Sc., selaku dosen pembimbing yang telah memberikan bimbingan dan arahan pada Tugas Akhir penulis.
5. Kedua orang tua penulis yang selalu mendoakan, memberi motivasi, dan dukungan moral maupun materil kepada peneliti.
6. Rafi Al Harits, Grove Street, Affan Malik, Fadli Alfarizi, dan Bunga Aulia Maharani yang selalu memberikan dukungan kepada penulis.

Yogyakarta, 01 Oktober 2025



M. Aunur Rahim Ghaffar Tuasikal

ABSTRAK

Grobbak Korea merupakan UMKM makanan yang menghadapi tantangan stagnasi penjualan meskipun tren Korean Wave (Hallyu) mendorong peningkatan konsumsi produk makanan Korea di Indonesia. Persaingan yang semakin ketat akibat masuknya produk impor dan bertambahnya kompetitor lokal menuntut strategi pemasaran berbasis data agar mampu bertahan di pasar. Cabang Grobbak Korea Nologaten, Yogyakarta, yang sempat mencatat penjualan tertinggi pada September–Oktober 2024 dengan nilai mendekati Rp17.000.000, kini mengalami penurunan signifikan hingga kurang dari Rp8.000.000 pada April 2025. Penelitian ini bertujuan untuk mengidentifikasi pola preferensi pembelian konsumen dan menganalisis sentimen ulasan pelanggan, sehingga dapat dirumuskan strategi peningkatan penjualan yang relevan. Metode penelitian menggunakan Association Rule Mining (ARM) dengan algoritma apriori dan Over-Time Rule Consistency Ratio (OCVR) pada data transaksi, serta analisis sentimen berbasis Naive Bayes Classifier pada data ulasan konsumen. Hasil ARM menunjukkan adanya keterkaitan kuat antara produk Corndog Mozzarella dan Topokki Extra Mozzarella dengan nilai lift > 1 , menandakan kecenderungan konsumen membeli kedua produk tersebut secara bersamaan. Analisis OCVR menunjukkan mayoritas aturan asosiasi konsisten ($< 30\%$), meskipun terdapat aturan kurang stabil dengan nilai 37%. Sementara itu, hasil analisis sentimen memperlihatkan dominasi ulasan positif, namun sentimen negatif mengkritisi aspek kualitas topokki, kelembutan roti corndog, standar rasa sosis, kualitas mozzarella, serta inkonsistensi pelayanan. Berdasarkan temuan tersebut, dirumuskan enam strategi utama: bundling dan cross-selling, promosi produk jarang terbeli, evaluasi formulasi topokki, standarisasi rasa, peningkatan kualitas bahan baku, serta perbaikan konsistensi pelayanan. Strategi ini diharapkan mampu meningkatkan kepuasan konsumen dan mendongkrak performa penjualan Grobbak Korea.

Kata Kunci: *Association Rule Mining, OCVR, Sentiment Analysis, 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	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Penelitian	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Kajian Literatur	5
2.2 Landasan Teori	16
2.2.1 Data Mining.....	16
2.2.2 Association Rule	18
2.2.3 Algoritma Apriori.....	19
2.2.4 Knowledge Discovery in Data Mining (KDD)	20
2.2.5 Overall Variability of Association Rule (OCVR)	22
2.2.6 Analisis Sentimen.....	22
2.2.7 Klasifikasi Sentimen	23
2.2.8 Algoritma Naïve Bayes Classifier.....	24
BAB III METODE PENELITIAN.....	25
3.1 Pendekatan Penelitian	25
3.2 Objek dan Subjek Penelitian	25
3.2.1 Objek Penelitian	25
3.2.2 Subjek Penelitian.....	25
3.3 Data yang Digunakan	25
3.3.1 Data Transaksi Penjualan:	25
3.3.2 Data Review Konsumen:.....	26
3.4 Alat dan Perangkat Lunak	26
3.5 Bagan Alir Penelitian	26
BAB IV PENGUMPULAN DAN PENGOLAHAN DATA	30
4.1 Pengumpulan Data	30
4.1.1 Data Transaksi Konsumen	30
4.1.2 Data Review Pelanggan	30
4.2 Pengolahan Data.....	31
4.2.1 Data Transaksi.....	31

4.2.2 Data Review	42
BAB V PEMBAHASAN	56
5.1 Analisis Aturan Asosiasi (<i>Rule</i>) Data Transaksi 1 tahun	56
5.2 Analisis Aturan Asosiasi (<i>Rule</i>) Data Transaksi Pada setiap Periode (12 Bulan) Menggunakan OCVR	58
5.3 Analisis Sentimen Data Review	60
5.4 Strategi Marketing Berdasarkan Hasil Analisis Preferensi Pembelian dan Sentimen Konsumen.....	61
BAB VI PENUTUP	64
6.1 Kesimpulan.....	64
6.2 Saran.....	65
DAFTAR PUSTAKA	66
LAMPIRAN	69

DAFTAR TABEL

Tabel 2. 1 Kajian Literatur	14
Tabel 4. 2 Penentuan Parameter	32
Tabel 4. 3 <i>Rule Association Rule</i> 1 Tahun	32
Tabel 4. 4 Aturan yang Muncul Pada Setiap Bulan (Januari)	34
Tabel 4. 5 Aturan yang Muncul Pada Setiap Bulan (Februari)	35
Tabel 4. 6 Aturan yang Muncul Pada Setiap Bulan (Maret)	36
Tabel 4. 7 Aturan yang Muncul Pada Setiap Bulan (April)	36
Tabel 4. 8 Aturan yang Muncul Pada Setiap Bulan (Mei)	37
Tabel 4. 9 Aturan yang Muncul Pada Setiap Bulan (Juni)	37
Tabel 4. 10 Aturan yang Muncul Pada Setiap Bulan (Juli)	38
Tabel 4. 11 Aturan yang Muncul Pada Setiap Bulan (Agustus)	39
Tabel 4. 12 Aturan yang Muncul Pada Setiap Bulan (September)	39
Tabel 4. 13 Aturan yang Muncul Pada Setiap Bulan (Oktober)	40
Tabel 4. 14 Aturan yang Muncul Pada Setiap Bulan (November)	40
Tabel 4. 15 Aturan yang Muncul Pada Setiap Bulan (Desember)	41
Tabel 4. 16 Perhitungan OCVR	42
Tabel 4. 17 Tahapan <i>Case Folding</i>	42
Tabel 4. 18 Tahapan <i>Remove Punctuation</i>	43
Tabel 4. 19 Tahapan <i>Emoji Removal</i>	44
Tabel 4. 20 Tahapan <i>Tokenization</i>	45
Tabel 4. 21 Tahapan <i>Stopword Removal</i>	47
Tabel 4. 22 Tahapan <i>Normalization</i>	48
Tabel 4. 23 Tahapan <i>Stemming</i>	49
Tabel 4. 24 Pelabelan <i>Lexicon</i>	50
Tabel 4. 25 Klasifikasi	51
Tabel 5. 1 <i>Rule Association Rule</i> 1 tahun	56
Tabel 5. 2 Jumlah <i>Rule</i> Tiap Periode	58
Tabel 5. 3 OCVR	59

DAFTAR GAMBAR

Gambar 1. 1 Data Penjualan.....	2
Gambar 2. 1 KDD	21
Gambar 4. 1 Data Transaksi Konsumen.....	30
Gambar 4. 2 Data Transaksi Setelah Preprocessing.....	31
Gambar 4. 3 Distribusi Sentimen	50
Gambar 4. 4 Confusion Matrix	52
Gambar 4. 5 Wordcloud Sentimen Negatif.....	54
Gambar 5. 1 Frequent Itemset.....	57
Gambar 5. 2 Kata Paling Sering Muncul pada Sentimen Negatif.....	60

BAB I

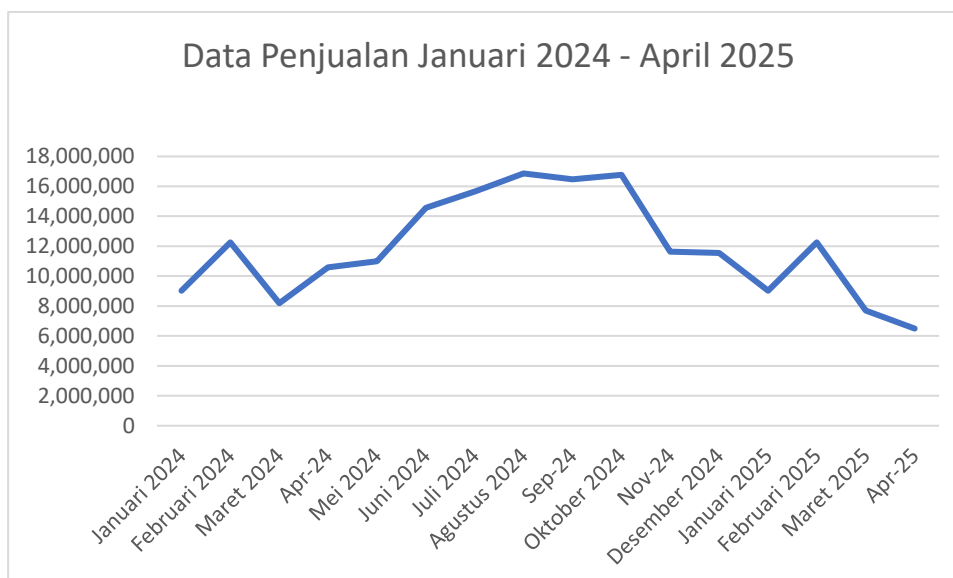
PENDAHULUAN

1.1 Latar Belakang

Grobbak Korea merupakan UMKM makanan yang telah tersebar di beberapa kota di Indonesia dan sedang menghadapi tantangan stagnasi penjualan di tengah maraknya tren makanan Korea di Indonesia. *Fenomena Korean Wave (Hallyu)* telah mendorong pertumbuhan signifikan pada industri makanan Korea, yang terlihat dari meningkatnya jumlah restoran dan warung makan Korea di berbagai kota besar di Indonesia (Yoga, 2023). Data dari Korea Agro-Fisheries & Food Trade Corporation mencatat bahwa ekspor makanan dan minuman Korea ke Indonesia terus meningkat setiap tahun, dengan nilai ekspor makanan halal Korea mencapai USD236 juta atau sekitar Rp3,7 triliun hingga September 2024, naik 2,7% dibandingkan tahun sebelumnya. Produk seperti mie instan Korea bahkan mencatat kenaikan ekspor sebesar 25,8% pada tahun 2024 (Nanang, 2024).

Peningkatan popularitas ini menjadikan Indonesia sebagai pasar utama bagi produk makanan Korea, dengan permintaan yang tinggi terhadap berbagai jenis kuliner seperti ramyun, tteokbokki, hingga minuman dan buah-buahan premium asal Korea. Namun, di balik pertumbuhan tersebut, persaingan di industri makanan juga semakin ketat. Gempuran produk impor dan perubahan selera konsumen akibat tren budaya Korea menuntut pelaku usaha untuk terus berinovasi dan mencari strategi pemasaran yang lebih efektif (Khoirul, 2021).

Dalam beberapa waktu terakhir, banyak cabang Grobbak Korea yang mengalami penurunan performa dan akhirnya tutup. Salah satu cabang yang masih bertahan hingga saat ini berada di wilayah Nologaten, Yogyakarta. Cabang ini mulai beroperasi pada tahun 2023 dan sempat menjadi satu-satunya gerai makanan Korea di daerah tersebut, yang membuat penjualan cukup tinggi pada saat awal beroperasi. Namun dengan munculnya berbagai kompetitor yang menawarkan produk serupa di daerah tersebut, penjualan Grobbak Korea Nologaten mengalami penurunan. Berikut merupakan data penjualan Januari 2024 – April 2025 Grobbak Korea Nologaten:



Gambar 1. 1 Data Penjualan

Berdasarkan grafik diatas, terlihat bahwa tren penjualan sempat meningkat secara signifikan hingga mencapai puncaknya pada bulan September dan Oktober 2024, dengan nilai penjualan mendekati Rp17.000.000. Setelah itu, penjualan mengalami penurunan tajam dan cenderung tidak stabil, bahkan mencapai titik terendah pada bulan April 2025 dengan nilai penjualan kurang dari Rp8.000.000. Fenomena ini menunjukkan bahwa meskipun usaha ini sempat menunjukkan performa positif, kondisi pasar yang semakin kompetitif dan kurangnya strategi adaptif berdampak langsung terhadap performa penjualan. Pemilik Grobbak Korea Nologaten ingin mencari cara untuk dapat menaikkan penjualan, Oleh karena itu diperlukan suatu analisis yang mendalam untuk meningkatkan daya saing dan volume penjualan.

Dalam konteks ini, Pemanfaatan data transaksi dan ulasan pelanggan menjadi sangat penting untuk merumuskan strategi pemasaran berbasis data. Penelitian terdahulu menunjukkan bahwa metode association rule mining efektif dalam menemukan pola pembelian konsumen dan dapat digunakan untuk merancang promosi produk yang lebih tepat sasaran (Elvin, Sulaiha, Laily, & Lailatul, 2022). *Association rule mining*, khususnya dengan algoritma apriori, mampu mengidentifikasi kombinasi produk yang sering dibeli bersamaan, sehingga dapat membantu UMKM dalam merancang strategi cross-selling dan penempatan produk yang optimal (Harlinda & Ramdan, 2025).

Selain itu, analisis sentimen terhadap ulasan pelanggan juga memberikan wawasan berharga mengenai persepsi dan kepuasan konsumen terhadap produk dan layanan yang ditawarkan. Dengan mengklasifikasikan opini konsumen menjadi sentimen positif, negatif, atau

netral, UMKM dapat memahami kebutuhan dan harapan pelanggan secara lebih mendalam, serta meningkatkan kualitas produk dan layanan yang diberikan. Penelitian terkait analisis sentimen pada UMKM telah membuktikan bahwa pendekatan ini dapat membantu mengidentifikasi kekuatan dan kelemahan produk mereka, sehingga dapat melakukan perbaikan yang berdampak positif pada penjualan (Prasetyo & Wibowo, 2020).

Berdasarkan hal tersebut, penelitian ini penting dilakukan untuk membantu Grobbak Korea dalam menggali pola preferensi pembelian dan sentimen pelanggan, sehingga dapat merumuskan strategi peningkatan penjualan yang relevan dan berbasis data.

1.2 Rumusan Masalah

Berdasarkan permasalahan yang dijelaskan pada bagian latar belakang, berikut ini rumusan masalah yang didapatkan:

1. Bagaimana pola preferensi pembelian konsumen dapat diidentifikasi menggunakan metode *Association Rule*?
2. Bagaimana cara mengidentifikasi dan memahami ulasan produk menggunakan analisis sentimen?
3. Strategi apa yang dapat diterapkan untuk meningkatkan penjualan berdasarkan hasil analisis preferensi pembelian dan sentimen konsumen?

1.3 Tujuan Penelitian

Berikut ini tujuan penelitian yang diusulkan berdasarkan rumusan masalah di atas:

1. Mengidentifikasi pola preferensi pembelian konsumen berdasarkan data transaksi dengan menggunakan metode *Association Rule Mining* untuk mengetahui keterkaitan antar produk yang sering dibeli secara bersamaan.
2. Menganalisis dan memahami ulasan konsumen terhadap produk melalui pendekatan *sentiment analysis* guna mengklasifikasikan opini konsumen ke dalam sentimen positif, negatif, atau netral.
3. Merumuskan strategi peningkatan penjualan yang relevan berdasarkan hasil analisis pola pembelian dan sentimen konsumen, sehingga dapat memberikan rekomendasi yang lebih tepat sasaran bagi pengambilan keputusan bisnis.

1.4 Manfaat Penelitian

Berikut ini merupakan manfaat dari penelitian ini:

1. Bagi Penulis Mengetahui arti penting dalam penerapan Data Mining, sehingga apabila penulis telah bekerja di bidang yang sama dapat menjadi nilai tambah yang dapat membantu dalam pemecahan masalah dan pemberian rekomendasi sesuai dengan keilmuan teknik industri.
2. Bagi Perusahaan Memberikan manfaat kepada perusahaan sebagai masukan untuk melakukan strategi pemasaran produk yang lebih efektif sehingga dapat meningkatkan profit penjualan dari perusahaan.
3. Bagi Pihak Lain Sebagai bahan acuan dan referensi bagi peneliti untuk melakukan penelitian dalam bidang yang sama dan mengembangkan metode yang lebih baik.

1.5 Batasan Penelitian

Penelitian ini memiliki Batasan seperti berikut:

1. Data transaksi yang digunakan terbatas pada periode Januari hingga Desember 2024. Transaksi di luar periode tersebut tidak dianalisis.
2. Ulasan pelanggan yang dianalisis diambil dalam kurun waktu yang sama, dan tidak mencakup ulasan sebelum atau setelahnya.
3. Teknik data mining yang digunakan untuk pola pembelian terbatas pada *Association Rule Mining* dengan menggunakan *algoritma apriori*.
4. Analisis sentimen dilakukan dengan pendekatan Naive Bayes, tanpa membandingkan kinerja dengan algoritma lain.
5. Penelitian ini hanya sampai pada tahapan perancangan strategi pemasaran, dan tidak mencakup implementasi langsung atau pengukuran hasil nyata setelah strategi diterapkan oleh Grobbak Korea Nologaten.

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Literatur

Berikut merupakan kajian literatur yang dipakai peneliti sebagai acuan dalam membuat penelitian: Penelitian pertama dilakukan oleh Krishna Kumar et al., (2024). Dokumen ini membahas tentang segmentasi pelanggan menggunakan teknik clustering, terutama K-Means, untuk menganalisis pola pembelian dan perilaku pelanggan. Penelitian ini menyoroti bagaimana analisis data pelanggan dari berbagai sumber seperti POS, CRM, dan platform online dapat membantu mengidentifikasi kelompok pelanggan yang berbeda. Hasil dari segmentasi ini memungkinkan perusahaan untuk melakukan pemasaran yang lebih personal, meningkatkan pengalaman pelanggan, dan mendukung pengambilan keputusan strategis. Proses yang dilakukan meliputi pengumpulan data, pra-pemrosesan, pemilihan fitur yang relevan, serta penentuan jumlah kluster optimal menggunakan metode seperti *Elbow* dan *Silhouette*, kemudian hasilnya divisualisasikan untuk analisis lebih lanjut. Meskipun pendekatan ini efektif, terdapat tantangan seperti skalabilitas data dan kebutuhan pemrosesan data secara real-time. Ke depan, diharapkan pengembangan dapat mencakup analisis data secara real-time, penggunaan model machine learning yang lebih canggih, serta integrasi dengan sistem CRM dan analisis geospasial. Tujuan utama dari penelitian ini adalah untuk meningkatkan pemasaran yang ditargetkan, pengelolaan inventaris, dan pengalaman pelanggan, sehingga dapat meningkatkan loyalitas dan pertumbuhan bisnis secara berkelanjutan.

Penelitian selanjutnya dilakukan oleh Pleerux & Nardkulpat (2023). Tujuan dari penelitian ini adalah untuk menganalisis dan membandingkan sentimen ulasan pelanggan restoran sebelum dan selama pandemi COVID-19, serta mengidentifikasi kata-kata emosional dan isi opini yang terkandung dalam ulasan tersebut. Penelitian ini bertujuan memberikan wawasan tentang persepsi pelanggan terkait layanan restoran, termasuk aspek makanan, kebersihan, dan layanan, sehingga dapat membantu pemilik restoran dalam meningkatkan kualitas layanan di masa krisis dan masa depan. Metode yang digunakan meliputi analisis sentimen berbasis NLP dan VADER untuk menilai ulasan online dari *TripAdvisor*, serta analisis frekuensi kata dan uji statistik *Wilcoxon signed-rank* untuk membandingkan pola sentimen sebelum dan selama

pandemi. Manfaat dari penelitian ini adalah memberikan wawasan yang dapat digunakan untuk strategi pemasaran dan peningkatan layanan restoran, serta membantu pemilik memahami aspek yang perlu diperbaiki, terutama terkait kebersihan dan *higiene*. Hasil utama menunjukkan penurunan volume ulasan, peningkatan sentimen negatif, dan fokus keluhan beralih ke kebersihan dan *higiene* selama pandemi, sementara kata-kata positif seperti "*food*", "*good*", dan "*great*" tetap dominan. Kekurangan dari penelitian ini meliputi keterbatasan sumber data yang hanya dari satu platform, potensi bias ulasan online, penggunaan satu model analisis sentimen, kurangnya analisis kualitatif mendalam, serta fokus geografis yang terbatas di Pattaya, sehingga hasilnya tidak dapat digeneralisasi secara luas.

Penelitian selanjutnya dilakukan oleh Khasanah & Baihaqie (2024). Jurnal ini menganalisis perilaku pembelian konsumen di toko ABC Yogyakarta menggunakan metode *clustering* (K-Means) dan *association rule mining* (FP-Growth) berdasarkan data transaksi tiga bulan. Tujuannya adalah mengidentifikasi segmen pelanggan dan pola pembelian produk tertentu, seperti telur dan mie instan. Ditemukan empat segmen pelanggan, dengan fokus pada "*Diverse Afternoon Buyers*" yang aktif di sore hari dan memiliki transaksi tinggi. Sebanyak 104 aturan asosiasi berhasil diidentifikasi, dan direkomendasikan strategi pemasaran seperti bundling, diskon, poin loyalitas, dan penempatan produk strategis. Namun, keterbatasannya meliputi durasi data yang singkat, fokus berlebih pada satu segmen, metode analisis yang terbatas, belum diuji nyatanya strategi yang disarankan, serta tidak adanya analisis faktor demografis.

Penelitian ini dilakukan oleh Pandya & Mehta (2020). Jurnal ini membahas berbagai aspek penting dalam analisis sentimen dan pemrosesan bahasa alami, termasuk teknik seperti *POS tagging*, *stemming*, penghilangan *stop words*, serta penanganan konjungsi dan negasi. Data yang digunakan berasal dari berbagai sumber seperti media sosial (Twitter, Facebook, YouTube), ulasan produk, artikel berita, dan *platform e-commerce*. Metode yang dibahas meliputi pendekatan *lexicon-based* dan *machine learning* seperti *Naïve Bayes*, *Support Vector Machines*, dan *neural networks* dengan tingkat analisis pada level dokumen, kalimat, maupun aspek tertentu. Selain itu, jurnal ini juga menyoroti aplikasi analisis sentimen dalam pengambilan keputusan bisnis, analisis tren pasar, dan riset sosial, termasuk penggunaan dalam berbagai bahasa seperti Arab dan Cina. Namun, terdapat beberapa tantangan yang dihadapi, seperti penanganan data tidak terstruktur dan multibahasa, kesulitan membedakan konten subjektif dan objektif, serta masalah polarity shift yang kompleks. Hasil penelitian menunjukkan bahwa model *machine learning* seperti SVM dan *Naïve Bayes* mampu mencapai

akurasi antara 57% hingga lebih dari 90%, dengan pentingnya *preprocessing* dan pemilihan sumber daya yang tepat. Jurnal juga menyarankan pengembangan model *hybrid* untuk meningkatkan performa analisis sentimen. Meski demikian, jurnal ini memiliki beberapa kekurangan. Penanganan data tidak terstruktur dari media sosial belum sepenuhnya optimal, terutama dalam konteks dan ambiguitas. Masalah *polarity shift* dan negasi masih menjadi tantangan besar tanpa solusi yang benar-benar efektif. Evaluasi metode yang ada juga terbatas pada beberapa metrik sehingga kurang memberikan gambaran menyeluruh tentang kelebihan dan kekurangan tiap metode. Selain itu, analisis sentimen secara *real-time* yang penting untuk aplikasi media sosial kurang dibahas, dan penggunaan dataset yang relatif kecil atau terbatas domainnya membuat hasil kurang dapat digeneralisasi.

Penelitian selanjutnya dilakukan oleh Hartmann et al. (2023). Penelitian ini membahas perbandingan antara tiga metode utama dalam analisis sentimen, yaitu *lexicon*, *machine learning* tradisional, dan *transfer learning*. *Transfer learning*, terutama model-model seperti *BERT*, *RoBERTa*, dan *SiBERT*, secara konsisten menunjukkan performa yang lebih tinggi dengan akurasi lebih dari 80% pada dataset besar dan teks panjang, serta mampu memanfaatkan tugas terkait untuk mengurangi kebutuhan data pelatihan. Metode *lexicon* dikenal karena kesederhanaan dan interpretabilitasnya, cocok untuk skenario yang membutuhkan kecepatan dan kejelasan, meskipun akurasinya cenderung lebih rendah sekitar 69%. Sementara itu, *machine learning* tradisional membutuhkan data berlabel dan fitur yang ekstensif, dan performanya masih kalah dibandingkan transfer learning, meskipun mampu memberikan interpretasi yang lebih baik terkait ekspresi bahasa yang digunakan. Kekurangan utama dari *transfer learning* adalah kurangnya transparansi dan interpretabilitas, serta ketergantungan pada *domain* tertentu yang memerlukan *fine-tuning* tambahan. Selain itu, performa model sangat dipengaruhi oleh karakteristik dataset seperti panjang teks dan jumlah kelas sentimen, dan tantangan di dunia nyata seperti data yang tidak bersih dan variabilitas konteks tetap menjadi hambatan. Beberapa studi juga tidak secara eksplisit membahas aspek etis dan potensi bias dari penggunaan model ini. Secara keseluruhan, tren saat ini menunjukkan bahwa transfer learning merupakan pendekatan paling menjanjikan, tetapi pemilihan metode harus disesuaikan dengan kebutuhan aplikasi, terutama terkait *trade-off* antara akurasi dan *interpretabilitas*.

Penelitian selanjutnya dilakukan oleh Wang (2025). Jurnal ini membahas penggunaan model aturan asosiasi dan *decision tree* dalam menganalisis data perilaku konsumen untuk

mengidentifikasi kombinasi produk yang memiliki korelasi tinggi serta pola perilaku pembelian. *Threshold support* sebesar 0,01, *confidence* 0,7, dan *lift* 1,2 diterapkan untuk menyeimbangkan efisiensi komputasi dan relevansi bisnis, meskipun hal ini dapat menyebabkan beberapa aturan yang kurang mencolok terabaikan. Data yang digunakan berasal dari 1.000 konsumen di Shenyang, dengan pembagian 70% untuk pelatihan dan 30% untuk pengujian, serta melalui validasi silang dan evaluasi menggunakan metrik seperti akurasi, *F1-score*, dan *AUC-ROC*. Model dioptimalkan melalui proses pruning dan *tuning hiperparameter* agar lebih *robust*. Analisis ini memungkinkan segmentasi pasar berdasarkan frekuensi pembelian, jumlah belanja, preferensi produk, dan saluran belanja, sehingga mendukung strategi pemasaran yang lebih personal seperti promosi tertarget, penetapan harga dinamis, dan optimalisasi saluran penjualan untuk meningkatkan keterlibatan pelanggan dan pertumbuhan penjualan. Selain itu, pendekatan ini menyoroti pentingnya pemasaran personalisasi, harga dinamis, dan program keanggotaan yang didukung *big data analytics* dalam meningkatkan penjualan, loyalitas pelanggan, dan pangsa pasar. Model decision tree dan aturan asosiasi memberikan wawasan yang dapat diinterpretasikan untuk memahami perilaku konsumen secara lebih transparan, sehingga memungkinkan pengembangan strategi yang lebih tepat sasaran. Meskipun efektif, model ini memiliki keterbatasan seperti risiko *overfitting* dan tantangan dalam pasar yang dinamis atau berbeda secara budaya, sehingga perlu pengembangan algoritma yang lebih canggih dan pembelajaran adaptif secara *real-time* agar akurasi dan relevansi dapat terus ditingkatkan. Penelitian ini juga menekankan pentingnya strategi retensi pelanggan jangka panjang, seperti personalisasi dan pengalaman pelanggan yang optimal, untuk memastikan pertumbuhan dan daya saing yang berkelanjutan. Secara keseluruhan, pendekatan berbasis data ini menunjukkan bahwa strategi pemasaran yang didukung analitik *big data* sangat penting untuk mengoptimalkan efektivitas pemasaran dan mempertahankan keunggulan kompetitif di pasar yang semakin kompleks dan cepat berubah.

Penelitian selanjutnya dilakukan oleh Gurudath (2020). Penelitian ini membahas secara mendalam tentang teknik data mining, khususnya asosiasi aturan dan algoritma seperti *Apriori* dan *FP-Growth* yang digunakan untuk mengidentifikasi hubungan antar produk dalam konteks ritel. Algoritma ini membantu dalam mengoptimalkan penempatan produk, meningkatkan penjualan, serta mengembangkan sistem rekomendasi dengan menganalisis pola pembelian pelanggan, serta menggunakan metrik seperti *support*, *confidence*, dan *lift* untuk menilai relevansi aturan yang dihasilkan. Prosesnya meliputi pengumpulan dan preprocessing data

transaksi, di mana data diubah menjadi format yang sesuai untuk analisis, kemudian algoritma tersebut digunakan untuk mengekstrak *itemset* yang sering muncul bersama dan menghasilkan aturan asosiasi. Dalam pengembangan sistem rekomendasi berbasis *web*, *framework* seperti *Flask* digunakan untuk visualisasi dan penyajian rekomendasi produk kepada pengguna. Hasil analisis menunjukkan pola pembelian pelanggan dan asosiasi produk yang sering muncul, yang dapat dimanfaatkan untuk meningkatkan pengalaman pelanggan dan penjualan. Perbandingan antara algoritma *Apriori* dan *FP-Growth* menunjukkan bahwa *FP-Growth* lebih efisien dan *scalable* untuk *dataset* besar. Selain itu, dokumen ini juga menyoroti tantangan seperti pertumbuhan eksponensial kombinasi *itemset* yang membuat proses pencarian *itemset* sering menjadi sangat komputasi intensif. Penelitian ini juga menekankan pentingnya pengaturan *threshold support* dan *confidence* yang tepat agar aturan yang dihasilkan relevan dan berkualitas. Secara umum, teknik data mining ini mendukung pengambilan keputusan berbasis data di bidang retail, membantu retailer memahami pola konsumsi pelanggan, mengelola inventaris secara lebih efektif, serta meningkatkan personalisasi rekomendasi produk. Rekomendasi pengembangan ke depan meliputi integrasi algoritma yang lebih canggih, sistem hybrid, serta penerapan pada *dataset* yang lebih besar dan beragam domain, guna memperluas manfaat dari analisis asosiasi ini.

Penelitian selanjutnya dilakukan oleh Xu et al. (2022). Jurnal ini merupakan tinjauan sistematis mengenai penelitian analisis sentimen berbasis media sosial yang membahas berbagai metode, tantangan, dan aplikasi dalam menganalisis opini serta emosi pengguna dari platform seperti Twitter, Facebook, dan Instagram. Pendekatan yang digunakan meliputi metode berbasis leksikon, *machine learning*, *hybrid*, serta kombinasi dengan teknik lain seperti analisis jejaring sosial dan regresi. Metode *machine learning* mendominasi penelitian dengan penggunaan algoritma seperti *Random Forest*, *SVM*, *Naive Bayes*, serta model *deep learning* seperti *CNN*, *LSTM*, dan *BERT* yang menunjukkan peningkatan performa signifikan. *Dataset* yang umum digunakan berasal dari berbagai sumber seperti Reddit, Twitter, IMDb, Yelp, Facebook, TripAdvisor, dan Amazon Reviews, dengan mayoritas penelitian menggunakan bahasa Inggris, meskipun ada peningkatan studi pada bahasa non-Inggris yang memerlukan *lexicon* khusus. Evaluasi model biasanya menggunakan metrik seperti akurasi, *F-measure*, *precision*, *recall*, *RMSE*, dan metrik khusus untuk klasifikasi *multi-label* seperti *Hamming Loss* dan *Jaccard Score*. Namun, jurnal ini juga menyoroti pentingnya mempertimbangkan efisiensi model, skalabilitas, dan waktu pelatihan untuk aplikasi dunia nyata. Tantangan utama yang

dihadapi meliputi data sosial media yang berisik dan tidak terstruktur, keterbatasan sumber daya bahasa, kesulitan menangkap nuansa seperti sarkasme dan konteks, serta kurangnya standarisasi dalam dataset dan metrik evaluasi. Selain itu, interpretabilitas model, terutama model *deep learning* yang kompleks, masih menjadi kendala dalam penerapan praktis. Penelitian masa depan disarankan untuk fokus pada pengembangan dataset terpadu, desain *classifier multi-domain* yang lebih fleksibel, pemilihan metrik evaluasi yang tepat, serta penilaian holistik terhadap performa model dalam konteks praktis. Secara keseluruhan, kombinasi metode analisis sentimen dengan teknik lain dapat meningkatkan wawasan, namun masih menghadapi tantangan terkait kualitas data, generalisasi model, dan konsistensi evaluasi.

Penelitian selanjutnya dilakukan oleh Kewsuwun & Kajornkasirat (2022). Jurnal ini membahas analisis sentimen komentar Facebook terkait startup agritech di Thailand menggunakan algoritma *Naive Bayes classifier*. Penelitian ini bertujuan untuk mengklasifikasikan komentar menjadi sentimen positif dan negatif guna memahami persepsi publik serta sikap investor terhadap inovasi agritech di Thailand. Data dikumpulkan dari 12 posting Facebook antara April 2016 hingga Maret 2021 dengan total 1.001 komentar yang kemudian dibersihkan untuk menghilangkan kata-kata slang atau tidak terstruktur agar mengurangi kesalahan analisis. Algoritma *Naive Bayes* yang digunakan terbukti efektif dalam mengklasifikasikan komentar dengan hasil menunjukkan bahwa sekitar 63,7% komentar bernada positif dan 36,3% bernada negatif. Hasil ini mencerminkan minat dan penerimaan yang kuat terhadap *startup agritech*, sekaligus adanya skeptisisme atau kekhawatiran dari sebagian pengguna. Penelitian ini juga menyoroti tantangan dalam analisis sentimen bahasa Thailand yang kompleks dan penggunaan data hanya dari satu platform media sosial, sehingga hasilnya mungkin kurang representatif secara luas. Selain itu, penelitian ini belum menggabungkan metode lain yang dapat meningkatkan akurasi serta belum melakukan analisis kualitatif yang lebih mendalam. Temuan dari studi ini dapat digunakan untuk mendukung strategi pemasaran, pengambilan kebijakan, serta pengembangan model analisis sentimen di konteks lain seperti perilaku konsumen dan pemasaran

Penelitian selanjutnya dilakukan oleh Horng et al. (2022). Penelitian ini menyajikan tinjauan sistematis terhadap penelitian manajemen pemasaran hotel dari tahun 2010 hingga 2020 dengan menganalisis 417 artikel dari 27 jurnal hospitality dan pariwisata menggunakan teknik *text mining*, analisis klaster, dan algoritma association rule. Studi ini mengelompokkan

literatur ke dalam empat klaster utama, yaitu kerangka dan strategi pemasaran, strategi pemasaran khusus seperti pemasaran media sosial, kualitas layanan dan perilaku pelanggan, serta ulasan konsumen online dan inovasi. Temuan menunjukkan bahwa branding, pengalaman pelanggan, pemasaran digital, dan pengukuran kinerja seperti kualitas layanan, loyalitas, dan kepuasan pelanggan menjadi fokus utama dalam penelitian pemasaran hotel. Selain itu, penggunaan *big data*, media sosial, dan pendekatan pemasaran berkelanjutan semakin menonjol sebagai tren terkini. Implikasi praktis dari penelitian ini menekankan pentingnya pemanfaatan media sosial dan ulasan *online* secara efektif serta personalisasi pengalaman pelanggan untuk meningkatkan nilai dan loyalitas. Namun, jurnal ini juga memiliki beberapa keterbatasan, seperti cakupan data yang terbatas pada artikel terindeks tertentu dan fokus yang sempit pada sektor hotel tanpa memperluas ke pariwisata secara umum. Selain itu, penelitian ini lebih bersifat deskriptif dan kurang mengembangkan teori baru, serta kurang menyoroti aspek organisasi internal hotel dan dinamika pasar yang cepat. Oleh karena itu, penelitian selanjutnya disarankan untuk memperdalam aspek teoritis, memperluas cakupan studi, dan mengikuti perkembangan teknologi serta tren pemasaran yang terus berubah.

Penelitian selanjutnya dilakukan oleh Giovani et al. (2020). Penelitian ini membahas analisis sentimen terhadap aplikasi Ruang Guru menggunakan data Twitter. Isi utama dari penelitian ini meliputi pengumpulan 513 tweet berbahasa Indonesia yang berkaitan dengan Ruang Guru, kemudian dilakukan proses preprocessing seperti mengubah huruf menjadi kecil, menghapus tautan, mention, kata pendek, dan stopwords. Selanjutnya, data diekstraksi menggunakan algoritma *Naive Bayes (NB)*, *Support Vector Machine (SVM)*, dan *K-Nearest Neighbor (K-NN)*, serta dilakukan fitur *selection* dengan *Particle Swarm Optimization (PSO)* untuk meningkatkan performa model. Hasil penelitian menunjukkan bahwa SVM dengan PSO memberikan performa terbaik dengan akurasi 78,55% dan AUC 0,853, lebih tinggi dibandingkan tanpa fitur PSO. Penelitian ini juga membandingkan efektivitas algoritma dan menunjukkan bahwa penggunaan fitur selection PSO dapat meningkatkan akurasi dan keandalan model dalam mengklasifikasi sentimen positif dan negatif dari komentar pengguna terhadap aplikasi Ruang Guru. Namun, terdapat beberapa kekurangan dalam penelitian ini. Jumlah data yang digunakan terbatas, hanya 513 tweet, sehingga mungkin belum cukup representatif untuk generalisasi yang lebih luas. Selain itu, fokus pada data berbahasa Indonesia membatasi penerapan hasil pada konteks multibahasa. Analisis juga kurang mendalam terhadap faktor-faktor yang mempengaruhi sentimen pengguna, seperti demografi atau waktu

pengumpulan data. Model yang digunakan terbatas pada algoritma klasik tanpa mencoba metode yang lebih canggih seperti deep learning. Selain itu, evaluasi hanya dilakukan melalui akurasi dan AUC tanpa mempertimbangkan metrik lain seperti *precision*, *recall*, dan *F1-score*, yang penting dalam analisis sentimen. Terakhir, hasil penelitian belum diuji pada data lain atau dalam konteks nyata untuk memastikan keandalannya.

Penelitian selanjutnya dilakukan oleh Safira et al. (2023). Penelitian ini membahas analisis sentimen masyarakat terhadap penggunaan fitur *PayLater* dalam belanja *online* di Indonesia. Penelitian ini menggunakan data sebanyak 405 twit yang dikumpulkan melalui metode *crawling* dengan pustaka *Python* bernama *snsrape*. Proses analisis dimulai dengan tahap *cleansing* data, di mana karakter tidak penting seperti *mention* dan URL dihilangkan dari *tweet*. Selanjutnya, dilakukan proses transformasi untuk memperbaiki kata tidak formal atau singkatan menjadi bentuk formal menggunakan database tertentu. Setelah itu, data diberi label sentimen (positif, negatif, netral) sebagai data uji untuk algoritma *Naive Bayes*. Implementasi algoritma *Naive Bayes* dan *TextBlob* dilakukan untuk klasifikasi sentimen, dengan hasil menunjukkan bahwa *Naive Bayes* mencapai akurasi tinggi sebesar 96%, sedangkan *TextBlob* lebih rendah, yaitu 61%. Distribusi sentimen menunjukkan bahwa mayoritas opini masyarakat bersentimen negatif terhadap *PayLater*. Evaluasi performa model dilakukan menggunakan confusion matrix, yang memperlihatkan keunggulan *Naive Bayes*. Kesimpulan dari penelitian ini adalah masyarakat cenderung memiliki persepsi negatif terhadap penggunaan *PayLater*. Secara umum, studi ini menunjukkan bahwa teknik data *analytics* dan *machine learning* dapat digunakan secara efektif untuk memahami persepsi masyarakat melalui data digital dan media sosial, seperti yang juga didukung oleh berbagai penelitian lain yang fokus pada analisis opini terkait topik-topik seperti vaksinasi, pembelajaran daring, layanan *e-commerce*, dan pengalaman pengguna layanan keuangan digital.

Penelitian selanjutnya dilakukan oleh Isra Almahsa et al. (2023). Penelitian ini membahas penerapan algoritma *FP-Growth* dalam berbagai konteks, seperti analisis pola pembelian pelanggan di CV Sukses Bersama dan Ghania Mart, serta aplikasi lain seperti analisis preferensi mahasiswa dan deteksi pencurian daya listrik. Temuan utama menunjukkan adanya hubungan kuat antar produk, seperti '*HRM*', '*MCC*', dan '*EXCELENTE*', dengan tingkat confidence yang tinggi, yang dapat digunakan untuk strategi pemasaran dan pengelolaan stok. Selain itu, penelitian ini juga menyoroti pentingnya transformasi data dan penyusunan tata letak toko

untuk meningkatkan pengalaman berbelanja pelanggan. Namun, jurnal ini memiliki beberapa kekurangan, seperti kurangnya penjelasan rinci tentang metodologi pengumpulan data, proses *preprocessing*, dan parameter yang digunakan dalam algoritma *FP-Growth*. Tidak disebutkan pula jumlah data yang dianalisis, sehingga sulit menilai generalisasi hasilnya. Selain itu, tidak ada analisis kritis mengenai keterbatasan metode, seperti potensi *overfitting* atau data yang tidak representatif, serta tidak adanya perbandingan performa *FP-Growth* dengan algoritma lain seperti Apriori. Implementasi dan evaluasi penggunaan *tools* atau *library* dalam analisis juga tidak dijelaskan secara mendetail, sehingga menimbulkan keterbatasan dalam menilai keberhasilan penelitian ini. Secara keseluruhan, jurnal ini memberikan gambaran umum yang baik tentang penerapan *FP-Growth*, tetapi membutuhkan penjelasan lebih mendalam dan evaluasi kritis untuk memperkuat temuan dan aplikasinya.

Penelitian selanjutnya dilakukan oleh Miftahul Jannah et al. (2023). Penelitian ini menerapkan metode *association rule mining* menggunakan algoritma Apriori untuk menganalisis data transaksi di Cafe X, sebuah kedai kopi modern di Jakarta, dengan tujuan mengembangkan strategi bundling produk guna meningkatkan penjualan. Data yang digunakan sebanyak 3.100 catatan transaksi dari bulan Maret 2021, dengan fokus pada transaksi yang berisi lebih dari satu item sebanyak 983 data. Proses analisis meliputi pembersihan data, integrasi, seleksi, reduksi, dan *preprocessing* untuk mengidentifikasi itemset yang sering muncul. Hasilnya, ditemukan delapan aturan asosiasi produk yang memenuhi ambang *support* dan *confidence* tertentu, yang menunjukkan peluang untuk melakukan bundling produk seperti *MD001*, *TH002*, *LB001*, dan *TU001*. Temuan ini kemudian digunakan untuk merancang strategi CRM dan promosi bundling produk yang diharapkan dapat meningkatkan penjualan dan keterlibatan pelanggan. Validasi melalui survei pelanggan menunjukkan bahwa empat bundling produk tersebut cukup menjanjikan untuk meningkatkan penjualan. Meskipun demikian, penelitian ini memiliki beberapa kekurangan, seperti keterbatasan data yang hanya berasal dari satu lokasi dan periode tertentu, jumlah aturan yang terbatas, serta validasi yang bersifat subjektif dan belum diuji secara langsung di lapangan. Selain itu, analisis tidak menyertakan aspek ekonomi dari strategi yang diusulkan dan belum dijelaskan secara rinci bagaimana implementasi operasionalnya. Penggunaan algoritma Apriori juga tidak dibandingkan dengan metode lain yang mungkin lebih efisien. Oleh karena itu, penelitian ini menyarankan perlunya pengembangan lebih lanjut dengan data yang lebih luas dan penerapannya nyata di lapangan untuk memperoleh hasil yang lebih komprehensif dan aplikatif.

Penelitian selanjutnya dilakukan oleh Khirdany et al. (2022). Penelitian ini membahas penerapan metode association rule, khususnya algoritma *Apriori*, dalam analisis data transaksi penjualan di sebuah kafe. Proses dimulai dengan transformasi data transaksi menjadi format *dummy* variabel (*binary*), di mana nilai 1 menunjukkan barang yang terjual dan 0 untuk yang tidak terjual. Selanjutnya, data tersebut dianalisis menggunakan algoritma *association rules* untuk menemukan pola pembelian yang sering terjadi, berdasarkan parameter *support* dan *confidence*, dengan bantuan perangkat lunak seperti *Rapid Miner*. Hasil analisis menunjukkan adanya kombinasi produk yang sering dibeli bersama, seperti onion ring dengan jus orange maupun latte americano, yang kemudian digunakan untuk merancang strategi promosi seperti diskon *buy 1 get 1 free*. Aturan-aturan yang ditemukan dapat membantu pemilik kafe dalam meningkatkan penjualan dan pengalaman pelanggan dengan menata produk secara lebih menarik dan sesuai kebiasaan pembeli. Penelitian ini menegaskan bahwa metode *association rule* efektif dalam mengidentifikasi pola pembelian dan mendukung pengambilan keputusan promosi produk. Pengembangan selanjutnya diharapkan dapat meningkatkan performa analisis dan strategi pemasaran berbasis data transaksi. Selain itu, penulis juga membahas perbandingan efisiensi antara algoritma Apriori dengan algoritma lain dalam penentuan *frequent itemset*, serta menekankan pentingnya proses *preprocessing* dan transformasi data dalam memperoleh *insight* yang akurat dari data penjualan.

Tabel 2. 1 Kajian Literatur

Penulis	Objek (Industri)		Metode			
	Umum	Restoran	SVM	NB	AR	KNN
(Krishna Kumar et al., 2024)	√					
(Pleerux & Nardkulpat, 2023)		√		√		
(Khasanah & Baihaqie, 2024)	√				√	
(Pandya & Mehta, 2020)	√		√	√		

Penulis	Objek (Industri)			Metode		
	Umum	Restoran	SVM	NB	AR	KNN
(Hartmann et al., 2023)	√		√			
(Wang, 2025)	√				√	
(Gurudath, 2020)	√				√	
(Xu et al., 2022)	√		√	√		
(Kewsuwun & Kajornkasirat, 2022)		√		√		
(Horng et al., 2022)		√			√	
(Giovani et al., 2020)	√		√	√		√
(Safira et al., 2023)	√			√		
(Isra Almahsa et al., 2023)	√				√	
(Miftahul Jannah et al., 2023)		√			√	
(Khirdany et al., 2022)		√			√	
Usulan		√		√	√	

Berdasarkan hasil kajian literatur yang tersaji pada Tabel 2.1, dapat diketahui bahwa mayoritas penelitian terdahulu berfokus pada penerapan satu metode analisis saja, baik *Association Rule Mining* maupun analisis sentimen berbasis algoritma klasifikasi seperti *Naïve Bayes* atau *Support Vector Machine* (SVM). Beberapa penelitian seperti yang dilakukan oleh Krishna Kumar et al. (2024), Pandya & Mehta (2020), serta Hartmann et al. (2023) menunjukkan

penerapan metode klasifikasi (SVM dan NB) dalam konteks analisis umum, sedangkan penelitian oleh Khasanah & Baihaqie (2024), Wang (2025), dan Gurudath (2020) lebih menekankan pada penerapan *Association Rule* untuk menemukan pola keterkaitan antarproduk.

Sementara itu, penelitian yang berfokus pada sektor restoran masih relatif terbatas. Studi oleh Pleerux & Nardkulpat (2023), Kewsuwun & Kajornkasirat (2022), Horng et al. (2022), serta Miftahul Jannah et al. (2023) menyoroti potensi penerapan analisis data dalam konteks industri kuliner, namun sebagian besar hanya mengimplementasikan satu pendekatan, seperti *Naïve Bayes* untuk analisis sentimen atau *Association Rule* untuk identifikasi pola transaksi.

Berdasarkan kajian tersebut, penelitian ini memiliki kebaruan (novelty) dengan menggabungkan dua pendekatan analitik, yaitu *Association Rule Mining* menggunakan algoritma *Apriori* untuk menemukan pola keterkaitan antarproduk dari data transaksi dan *Naïve Bayes* untuk analisis sentimen pada ulasan konsumen. Integrasi kedua metode ini diharapkan mampu memberikan gambaran yang lebih komprehensif mengenai pola preferensi dan persepsi konsumen, khususnya pada sektor restoran cepat saji, seperti studi kasus Grobbak Korea Nologaten. Selain itu, penelitian ini juga menambahkan tahap pengukuran *Overall Variability of Association Rule (OCVR)* untuk menilai stabilitas pola asosiasi sepanjang periode pengamatan, yang belum banyak dilakukan pada penelitian terdahulu.

2.2 Landasan Teori

Berikut merupakan landasan teori pada penelitian ini:

2.2.1 Data Mining

Data mining atau penambangan data merupakan bagian dari proses *Knowledge Discovery in Databases (KDD)* yang bertujuan untuk mengekstrak pola-pola atau informasi tersembunyi dari kumpulan data yang besar. Proses ini tidak hanya sekadar pengumpulan data, tetapi lebih kepada penemuan pola, relasi, dan tren yang bermakna dari data, yang dapat digunakan untuk pengambilan keputusan strategis. Han (2011) mendefinisikan *data mining* sebagai “proses untuk menemukan pola-pola yang menarik dan pengetahuan dari jumlah data yang sangat besar”.

Menurut Fayyad et al. (1996), proses *Knowledge Discovery in Databases (KDD)* terdiri atas beberapa tahapan penting, dimulai dari seleksi data untuk menentukan data yang relevan, pembersihan data guna menghilangkan *noise* maupun inkonsistensi, transformasi data agar

sesuai dengan kebutuhan analisis, hingga tahap inti yaitu *data mining* yang berfokus pada penerapan algoritma untuk mengekstrak pola. Tahap terakhir dalam proses ini adalah evaluasi pola dan interpretasi hasil, di mana pola yang ditemukan dinilai dan dipahami agar dapat memberikan manfaat nyata bagi pengguna akhir.

Dalam perkembangannya, *data mining* banyak diaplikasikan di berbagai bidang seperti perbankan, pemasaran, kedokteran, manufaktur, hingga pendidikan. Dalam konteks bisnis, teknik ini memiliki peran penting untuk membantu perusahaan memahami perilaku pelanggan, melakukan segmentasi pasar, mendeteksi penipuan, serta menyusun strategi penjualan dan pemasaran yang lebih efektif (Larose, 2015).

Lebih lanjut, Tan (2019) membagi teknik *data mining* menjadi dua kategori utama, yaitu *descriptive mining* dan *predictive mining*. *Descriptive mining* bertujuan menggambarkan pola atau hubungan yang terdapat dalam data, contohnya *association rules* dan *clustering*. Sementara itu, *predictive mining* berfungsi untuk memprediksi nilai atau kelas tertentu berdasarkan data historis, misalnya melalui metode *classification* dan *regression*.

Beberapa teknik populer yang sering digunakan dalam *data mining* antara lain *classification* untuk mengelompokkan data ke dalam kategori yang telah ditentukan sebelumnya, *clustering* untuk mengelompokkan data berdasarkan kemiripan tanpa label awal, *association rule mining* untuk menemukan keterkaitan antar item dalam satu dataset, *regression* untuk memprediksi nilai numerik suatu variabel berdasarkan variabel lain, serta *anomaly detection* yang digunakan dalam mendeteksi data menyimpang, seperti pada kasus deteksi penipuan.

Keberhasilan penerapan *data mining* sangat dipengaruhi oleh kualitas data yang dianalisis. Data dengan *noise*, *missing values*, atau ketidakkonsistenan akan menghambat proses ekstraksi informasi yang valid. Oleh karena itu, tahap *preprocessing* menjadi aspek yang sangat penting untuk memastikan integritas dan kualitas data. Selain itu, tantangan lain yang sering muncul adalah pemilihan algoritma yang tepat sesuai tujuan analisis, kemampuan metode untuk menangani data berukuran besar, serta kebutuhan pemahaman konteks domain oleh pengguna akhir. Keselarasan antara tujuan bisnis dan pemilihan teknik *data mining* harus diperhatikan agar hasil analisis benar-benar mampu memberikan nilai tambah dalam proses pengambilan keputusan.

2.2.2 Association Rule

Association Rule adalah metode data mining yang digunakan untuk menemukan pola hubungan atau keterkaitan antar item dalam data transaksi besar. Tujuannya adalah mengidentifikasi kombinasi item yang sering muncul bersama, sehingga dapat dimanfaatkan untuk strategi bisnis seperti promosi dan pengelolaan stok. Misalnya, algoritma Apriori, yang merupakan teknik populer dalam Association Rule, digunakan untuk menemukan kombinasi item dengan nilai support dan confidence tertentu sehingga menghasilkan aturan asosiasi yang kuat. Parameter support menunjukkan frekuensi kemunculan itemset dalam data, confidence merepresentasikan kekuatan aturan, dan lift mengukur kekuatan asosiasi antara item X dan Y melebihi probabilitas kemunculannya secara independent (Fitria, 2025). Untuk menentukan seberapa kuat dan relevan suatu aturan asosiasi, digunakan tiga ukuran utama, yaitu *support*, *confidence*, dan *lift*.

1. Support

Support mengukur proporsi transaksi dalam dataset yang mengandung X dan Y secara bersamaan.

$$Support(X \rightarrow Y) = P(X \cup Y) \quad (2.1)$$

2. Confidence

Confidence mengukur seberapa besar kemungkinan item Y muncul dalam transaksi yang sudah mengandung item X .

$$Confidence(X \rightarrow Y) = P(Y | X) = \frac{P(X \cup Y)}{P(X)} \quad (2.2)$$

3. Lift

Lift mengukur tingkat keterkaitan antara X dan Y dibandingkan dengan jika keduanya tidak saling tergantung. Nilai lift > 1 menunjukkan adanya korelasi positif.

$$Lift(X \rightarrow Y) = \frac{P(Y | X)}{P(Y)} = \frac{Confidence(X \rightarrow Y)}{Support(Y)} \quad (2.3)$$

Ketiga metrik ini penting untuk menyaring dan mengevaluasi aturan asosiasi yang dianggap signifikan dan relevan dalam konteks bisnis atau operasional tertentu. Beberapa algoritma yang banyak digunakan dalam pembangkitan aturan asosiasi di antaranya adalah *Apriori Algorithm*

dan *FP-Growth Algorithm*. Algoritma *Apriori* yang diperkenalkan oleh Agrawal et al. (1994) merupakan algoritma klasik yang menggunakan pendekatan *bottom-up* serta mengandalkan properti *downward-closure*. Prinsip ini menyatakan bahwa jika suatu *itemset* tidak sering muncul, maka seluruh *superset*-nya juga dapat dipastikan tidak sering muncul.

Sementara itu, algoritma *FP-Growth* atau *Frequent Pattern Growth* yang diperkenalkan oleh Han et al. (2000) hadir sebagai pengembangan dari *Apriori* dengan tujuan meningkatkan efisiensi. *FP-Growth* menghilangkan proses pembangkitan kandidat *itemset* dan menggantinya dengan pembangunan struktur pohon yang disebut *FP-tree*. Melalui struktur ini, eksplorasi pola yang sering muncul dapat dilakukan dengan lebih cepat dan efisien dibandingkan metode *Apriori*.

2.2.3 Algoritma *Apriori*

Algoritma *Apriori* merupakan salah satu algoritma paling fundamental dan banyak digunakan dalam proses *association rule mining*, yaitu untuk mengekstraksi pola-pola *itemset* yang sering muncul (*frequent itemsets*) dari *database* transaksi yang besar. Tujuan utama algoritma ini adalah untuk menemukan hubungan asosiatif antar item dalam dataset transaksi sehingga dapat diturunkan aturan-aturan yang bermanfaat dalam pengambilan Keputusan (Witten, Frank, & Hall, 2011)

Apriori bekerja berdasarkan prinsip *downward-closure* atau disebut juga *Apriori property*, yang menyatakan bahwa: “Jika suatu *itemset* bersifat *frequent*, maka semua subset-nya juga bersifat *frequent*. Sebaliknya, jika suatu *itemset* tidak memenuhi *minimum support*, maka semua *superset*-nya tidak perlu dipertimbangkan” (Sharma & Bhardwaj, 2012). Algoritma ini menggunakan pendekatan iteratif yang dikenal sebagai metode *level-wise search*, dimulai dari *itemset* berukuran satu dan secara bertahap memperluasnya ke *itemset* yang lebih besar hingga tidak ada lagi *itemset* yang memenuhi ambang *minimum support*.

Menurut Srikant dan Agrawal (1996) serta dipertegas oleh Witten et al. (2011), proses kerja algoritma *Apriori* dapat dirangkum ke dalam tiga tahap utama. Tahap pertama adalah *frequent itemset generation*, yaitu menentukan seluruh kombinasi *itemset* yang memiliki nilai *support* lebih besar atau sama dengan *minimum support*. Proses ini diawali dari *itemset* tunggal yang kemudian diperluas menjadi pasangan item dan seterusnya. Tahap kedua adalah *pruning*, yaitu menghapus kandidat *itemset* yang subset-nya tidak termasuk dalam *frequent itemset* sebelumnya sehingga dapat mengurangi jumlah perhitungan yang tidak diperlukan. Tahap

ketiga adalah *rule generation*, di mana dari *frequent itemset* yang terbentuk kemudian dihasilkan aturan asosiasi yang memenuhi nilai *minimum confidence*.

Sebagai ilustrasi, apabila dalam suatu transaksi ritel ditemukan aturan asosiasi {roti} → {mentega}, maka hal tersebut menunjukkan bahwa pelanggan yang membeli roti cenderung juga membeli mentega. Informasi seperti ini dapat dimanfaatkan dalam strategi penjualan silang (*cross-selling*) maupun dalam pengaturan tata letak produk agar lebih efektif (Tan et al., 2006). Algoritma apriori memiliki kelebihan dan kekurangan, kelebihanannya antara lain adalah konsepnya sederhana dan mudah diimplementasikan dan cocok untuk dataset yang ukurannya tidak terlalu besar. Sedangkan kekurangannya antara lain adalah waktu komputasi tinggi karena menghasilkan banyak kandidat itemset terutama ketika jumlah item besar dan membutuhkan multiple scan terhadap database yang menyebabkan inefisiensi pada skala besar (Soni, 2014).

2.2.4 Knowledge Discovery in Data Mining (KDD)

Knowledge Discovery in Databases (KDD) merupakan suatu proses sistematis untuk mengekstraksi *knowledge* atau pengetahuan yang valid, baru, potensial berguna, dan dapat dimengerti dari sekumpulan besar data. Istilah ini pertama kali diperkenalkan oleh (Fayyad, Piatetsky-Shapiro, & Smyth, 1996), yang menekankan bahwa *data mining* hanyalah salah satu tahap dalam proses KDD, bukan keseluruhan proses itu sendiri.

Menurut (Turban, Sharda, & Delen, 2011), KDD adalah proses yang melibatkan berbagai teknik statistik, pembelajaran mesin, dan *database management system* untuk menemukan pola dan tren yang tersembunyi dalam data yang bersifat besar, kompleks, dan heterogen.

Berdasarkan pemodelan umum dari (Fayyad, Piatetsky-Shapiro, & Smyth, 1996), proses KDD terdiri dari beberapa tahapan utama:

1. *Selection*

Tahap ini berfokus pada pemilihan data yang relevan dari database atau sumber data lainnya yang akan digunakan dalam proses KDD.

2. *Preprocessing (Cleaning)*

Data yang dipilih dibersihkan dari duplikasi, noise, dan data yang hilang agar kualitas data terjaga.

3. *Transformation*

Data dibentuk ulang ke dalam format yang sesuai untuk proses data mining, misalnya melalui normalisasi, pengkodean, atau reduksi dimensi.

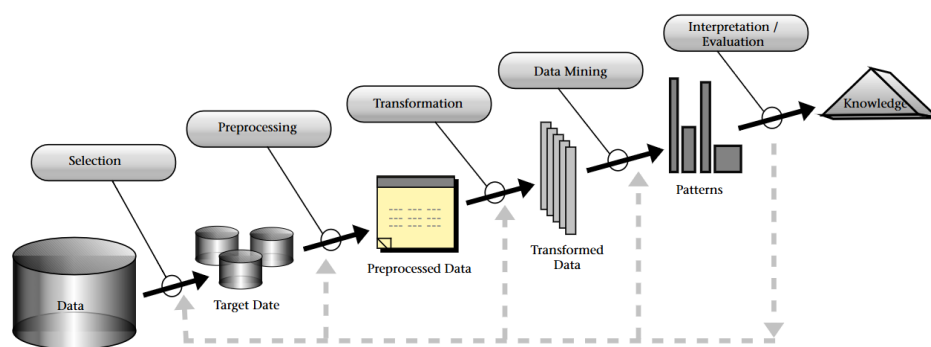
4. *Data Mining*

Pada tahap ini dilakukan penerapan teknik komputasi untuk menemukan pola atau model yang menarik. Metode yang digunakan bisa berupa *classification*, *clustering*, *association rules*, *regression*, dan lainnya.

5. Interpretation/Evaluation

Pola atau model yang ditemukan dievaluasi untuk menilai validitas, relevansi, dan kegunaannya. Hasil ini kemudian diterjemahkan menjadi pengetahuan eksplisit yang dapat digunakan untuk mendukung pengambilan keputusan.

Berikut merupakan gambar dari proses KDD:



Gambar 2. 1 KDD

Fayyad et al. (1996) menjelaskan bahwa pengetahuan yang diperoleh melalui proses *Knowledge Discovery in Databases* (KDD) harus memenuhi beberapa kriteria tertentu. Pertama, pengetahuan yang dihasilkan harus valid, yaitu pola yang ditemukan secara statistik signifikan serta benar dalam konteks data. Kedua, pengetahuan tersebut harus bersifat novel, yakni memberikan informasi baru yang sebelumnya belum diketahui. Ketiga, pengetahuan yang diperoleh harus memiliki potensi kegunaan, yaitu dapat mendukung proses pengambilan keputusan maupun tindakan lanjutan. Terakhir, hasil yang diperoleh harus dapat dimengerti, sehingga dapat dijelaskan dan dipahami oleh pengguna.

Lebih lanjut, Witten et al. (2011) menekankan adanya perbedaan lingkup antara istilah KDD dan *data mining* yang sering kali tertukar penggunaannya. *Data mining* merupakan tahap inti dari KDD yang berfokus pada proses penggalian pola dari data, sedangkan KDD mencakup ruang lingkup yang lebih luas, mulai dari pengumpulan, pengolahan, analisis, hingga interpretasi hasil. Dengan demikian, *data mining* dapat dipandang sebagai salah satu komponen penting dalam siklus KDD, namun bukan keseluruhan prosesnya.

2.2.5 Overall Variability of Association Rule (OCVR)

Overall Variability of Association Rule (OCVR) adalah suatu pendekatan evaluatif dalam analisis aturan asosiasi yang digunakan untuk mengukur tingkat kestabilan dari aturan yang dihasilkan dari proses *data mining* berbasis *market basket analysis*. Pendekatan ini muncul sebagai respons terhadap permasalahan dalam penerapan aturan asosiasi, yaitu bagaimana konsistensi suatu aturan berubah seiring waktu atau dalam dataset yang berbeda beda.

Stabilitas aturan sangat penting dalam konteks strategi pemasaran, penataan produk, dan sistem rekomendasi. Aturan yang tidak konsisten berpotensi menghasilkan keputusan yang tidak akurat. Oleh karena itu, metrik OCVR dikembangkan sebagai indikator untuk menyaring aturan yang memiliki variabilitas rendah, yang berarti lebih dapat diandalkan. OCVR hanya memperhitungkan dua metrik utama dari aturan asosiasi, yaitu confidence dan lift. Perhitungan dilakukan melalui koefisien variasi (Coefficient of Variation/CV) dari masing-masing metrik (Alfiqra & Khasanah, 2020).

Rumus untuk OCVR adalah sebagai berikut:

$$OVCR = \frac{CVC + CVL}{2} \quad (2.4)$$

Dengan

$$CV = \frac{\sigma}{\mu} \times 100\% \quad (2.4)$$

Keterangan:

- σ = standar deviasi nilai confidence/lift dalam beberapa periode waktu
- μ = rata-rata nilai confidence/lift dalam periode yang sama

2.2.6 Analisis Sentimen

Analisis sentimen atau *sentiment analysis* merupakan salah satu cabang dari *text mining* dan *natural language processing* (NLP) yang berfokus pada identifikasi, ekstraksi, dan pengklasifikasian opini atau emosi yang terkandung dalam teks. Tujuan utama dari analisis sentimen adalah untuk mengetahui kecenderungan sikap pengguna terhadap suatu entitas, seperti produk, layanan, individu, atau peristiwa (Liu, 2012).

Analisis sentimen biasanya diklasifikasikan dalam tiga kategori utama:

1. Sentimen positif
2. Sentimen negatif
3. Sentimen netral

Dalam praktiknya, teknik ini banyak digunakan dalam pengolahan data opini dari media sosial, ulasan produk, atau data survei terbuka untuk mendukung pengambilan keputusan dalam pemasaran, politik, atau layanan publik.

Menurut (Cambria, Schuller, Xia, & Havasi, 2013), proses analisis sentimen umumnya mencakup beberapa tahapan sebagai berikut:

1. Preprocessing:
Meliputi penghapusan karakter khusus, stopword removal, stemming, tokenisasi, dan normalisasi teks.
2. Feature Extraction:
Mengubah teks menjadi representasi numerik, misalnya menggunakan metode *bag-of-words*, *TF-IDF*, atau *word embedding*.
3. Sentiment Classification:
Melibatkan proses klasifikasi menggunakan pendekatan berbasis:
 - *Lexicon-based*: menggunakan kamus kata dengan label sentimen.
 - *Machine learning*: menggunakan algoritma klasifikasi seperti Naive Bayes, SVM, Decision Tree, atau Neural Network.
 - *Hybrid*: kombinasi antara pendekatan leksikal dan pembelajaran mesin.

Analisis sentiment memiliki kekurangan antara lain adalah ambiguitas Bahasa alami, penggunaan sarkasme, ironi, atau metafora, keterbatasan kamus sentiment pada domain tertentu, dan perubahan makna kata tergantung konteks kalimat.

2.2.7 Klasifikasi Sentimen

Klasifikasi sentimen merupakan bagian integral dari analisis sentimen yang bertujuan untuk mengelompokkan opini dalam teks ke dalam kategori polaritas tertentu, seperti positif, negatif, atau netral. Proses ini memungkinkan komputer untuk memahami dan menafsirkan emosi atau pendapat manusia yang tersirat dalam bahasa alami (Pang & Lee, 2008).

Secara umum, klasifikasi sentimen menggunakan pendekatan supervised learning, di mana algoritma belajar dari data yang telah diberi label sebelumnya untuk mengklasifikasikan data baru berdasarkan pola yang telah dikenali (Sebastiani, 2002).

Tujuan utama dari klasifikasi sentimen adalah untuk mengidentifikasi kecenderungan opini dari pengguna terhadap suatu produk, jasa, atau fenomena tertentu, kemudian untuk mendukung pengambilan keputusan bisnis atau kebijakan public, dan untuk menyaring informasi berdasarkan emosi pengguna, seperti dalam sistem rekomendasi atau monitoring media sosial.

2.2.8 *Algoritma Naïve Bayes Classifier*

Algoritma *Naïve Bayes Classifier* merupakan salah satu metode klasifikasi berbasis probabilistik yang didasarkan pada penerapan *Teorema Bayes* dengan asumsi bahwa antar fitur bersifat independen secara kondisional. Dalam konteks klasifikasi teks, algoritma ini sangat populer karena kesederhanaannya, efisiensinya dalam komputasi, dan kemampuannya menghasilkan akurasi yang kompetitif, terutama ketika jumlah data pelatihan terbatas (Manning, Raghavan, & Schütze, 2008). Prinsip utama dari *Naïve Bayes* adalah menghitung probabilitas posterior dari suatu kelas berdasarkan fitur-fitur yang ada dalam data.

Meskipun algoritma ini mengasumsikan independensi antar fitur yang dalam praktiknya jarang sepenuhnya terpenuhi namun pendekatan ini terbukti cukup kuat dalam berbagai aplikasi, terutama pada klasifikasi dokumen, *spam detection*, dan analisis sentimen (McCallum & Nigam, 1998). Keunggulan utama dari algoritma ini adalah efisiensi dalam proses pelatihan dan klasifikasi, bahkan pada dataset besar, serta kemampuannya untuk menangani data diskrit maupun kontinu. Untuk data teks, biasanya fitur berupa frekuensi kata atau representasi *TF-IDF* digunakan dalam proses perhitungan *likelihood*.

Dalam pengembangannya, terdapat beberapa varian dari *Naïve Bayes*, seperti *Multinomial Naïve Bayes* yang cocok untuk data teks, serta *Gaussian Naïve Bayes* yang digunakan untuk data numerik yang terdistribusi normal. Kelemahan utama dari algoritma ini adalah sensitivitas terhadap korelasi antar fitur dan performa yang menurun apabila asumsi independensi tidak terpenuhi secara signifikan (Rish, 2001). Namun demikian, *Naïve Bayes* tetap menjadi pilihan utama dalam berbagai studi klasifikasi karena kesederhanaan dan efektivitasnya.

BAB III

METODE PENELITIAN

3.1 Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksploratif. Pendekatan ini dipilih karena bertujuan untuk menggali pola-pola tersembunyi dalam data transaksi dan data review pelanggan Grobbak Korea cabang Nologaten. Teknik data mining digunakan untuk mengeksplorasi pola pembelian (menggunakan Market Basket Analysis) dan klasifikasi sentimen pelanggan (menggunakan algoritma Naive Bayes).

3.2 Objek dan Subjek Penelitian

3.2.1 *Objek Penelitian*

Objek dalam penelitian ini adalah konsumen *Grobbak Korea Nologaten* yang telah melakukan transaksi pembelian dan memberikan ulasan terhadap produk.

3.2.2 *Subjek Penelitian*

Subjek dalam penelitian ini adalah data transaksi dan data ulasan pelanggan di restoran *Grobbak Korea* cabang Nologaten. Data transaksi diperoleh dari perusahaan dan mencakup seluruh pembelian produk dari Januari hingga Desember 2024. Data ini dianalisis menggunakan metode *Association Rule Market Basket Analysis (AR-MBA)* dan *Over-Time Rule Consistency Ratio (OCVR)*. Data ulasan pelanggan diperoleh melalui web scraping dari platform GoFood dan digunakan untuk analisis sentimen menggunakan Naive Bayes dan pencarian asosiasi kata.

3.3 Data yang Digunakan

Jenis data yang digunakan dalam penelitian ini adalah data sekunder, yaitu:

3.3.1 *Data Transaksi Penjualan:*

Catatan pembelian pelanggan yang terdiri dari nama produk, tanggal transaksi, dan informasi pembelian lainnya dari Januari sampai Desember 2024.

3.3.2 Data Review Konsumen:

Kumpulan komentar dan ulasan dari pelanggan pada platform GoFood yang diperoleh melalui proses web scraping.

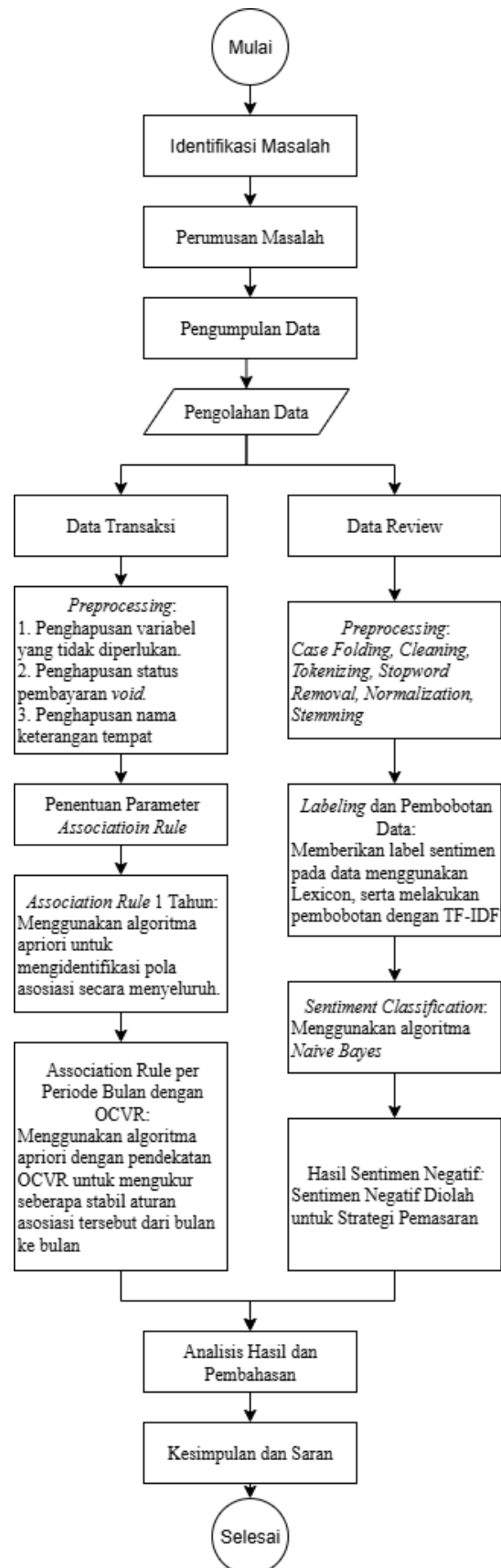
3.4 Alat dan Perangkat Lunak

Penelitian ini menggunakan perangkat lunak dan pustaka berikut:

- *Python*: untuk preprocessing, data mining, dan analisis sentimen.
- *Pandas, Numpy*: untuk manipulasi data.
- *MLxtend & Scikit-Learn*: untuk Market Basket Analysis dan Naive Bayes.
- *NLTK, Sastrawi, emoji, re*: untuk preprocessing teks Bahasa Indonesia.
- *Google Colab*: sebagai platform pemrograman.

3.5 Bagan Alir Penelitian

Berikut adalah bagan alir dari proses penelitian:



Gambar 3. 1 Bagan Alir Penelitian

1. Identifikasi Masalah

Menentukan permasalahan utama yang dihadapi oleh objek penelitian, dalam hal ini Grobbak Korea Nologaten. Permasalahan yang diidentifikasi adalah stagnasi penjualan yang berpotensi disebabkan oleh pola pembelian dan persepsi pelanggan.

2. Rumusan Masalah

Merumuskan masalah secara terstruktur berdasarkan hasil identifikasi. Ini menjadi dasar penentuan tujuan dan ruang lingkup penelitian, seperti apakah terdapat pola pembelian yang signifikan dan bagaimana persepsi konsumen terhadap produk.

3. Pengumpulan Data

Pengumpulan data didapatkan berdasarkan sumber data yang dimiliki.

Mengumpulkan dua jenis data:

- Data transaksi dari perusahaan selama 1 tahun.
- Data review pelanggan hasil *scraping* dari GoFood.

4. Pengolahan Data

Tahap pengolahan data dalam penelitian ini dibagi menjadi dua jalur utama sesuai dengan jenis data yang digunakan, yaitu data transaksi dan data review. Pada jalur data transaksi, pengolahan dimulai dengan tahap *preprocessing* berupa pembersihan data. Proses ini dilakukan dengan menghapus variabel yang tidak relevan, menghilangkan status pembayaran yang berstatus *void* atau gagal, serta menghapus keterangan lokasi yang tercantum pada nama produk agar data lebih seragam. Setelah data dibersihkan, dilakukan analisis Association Rule dengan menggunakan algoritma *Apriori*. Analisis ini dilakukan dalam dua skala, yaitu skala tahunan dan skala bulanan. Pada skala tahunan, *Association Rule* digunakan untuk mengidentifikasi pola asosiasi pembelian produk secara menyeluruh sepanjang periode penelitian, sehingga dapat ditemukan keterkaitan produk yang paling dominan. Sementara itu, pada skala bulanan, analisis dilakukan dengan menambahkan pendekatan *Overall Variability of Confidence Ratio (OCVR)* untuk mengukur stabilitas aturan asosiasi dari bulan ke bulan. Nilai OCVR yang rendah menunjukkan bahwa suatu aturan relatif konsisten sepanjang periode, sedangkan nilai yang tinggi mengindikasikan adanya variasi atau ketidakstabilan pola asosiasi. Dengan demikian, kombinasi analisis tahunan dan bulanan ini memungkinkan diperolehnya gambaran pola pembelian yang tidak hanya dominan secara agregat, tetapi juga konsisten dalam jangka waktu yang lebih detail.

Pada jalur data review, tahap pertama yang dilakukan adalah *preprocessing* teks untuk memastikan data ulasan dapat dianalisis secara tepat. Proses ini meliputi *case folding* (mengubah semua huruf menjadi kecil), *data cleaning* (menghapus tanda baca, karakter ASCII, dan emoji), *tokenization* (memecah kalimat menjadi kata per kata), *stopword removal* (menghapus kata-kata umum yang tidak bermakna dengan bantuan NLTK), *normalization* (menstandarkan kata tidak baku atau singkatan), serta *stemming* (mengubah kata ke bentuk dasar). Setelah teks dibersihkan, data diberi label sentimen menggunakan pendekatan berbasis *lexicon* untuk membaginya ke dalam tiga kategori, yaitu positif, netral, dan negatif. Selanjutnya, dilakukan pembobotan kata menggunakan metode *TF-IDF* (*Term Frequency–Inverse Document Frequency*) agar setiap kata memiliki nilai bobot sesuai tingkat kepentingannya dalam dokumen.

Tahap berikutnya adalah klasifikasi sentimen menggunakan algoritma *Naïve Bayes*, yang dipilih karena kesederhanaan, kecepatan, serta kinerjanya yang baik dalam klasifikasi teks. Proses klasifikasi ini menghasilkan distribusi sentimen konsumen yang memberikan gambaran persepsi umum terhadap produk. Dari hasil tersebut, penelitian kemudian memfokuskan analisis lebih lanjut pada ulasan dengan sentimen negatif. Analisis ini bertujuan untuk mengidentifikasi aspek-aspek yang paling banyak dikeluhkan konsumen, sehingga dapat diperoleh wawasan (*knowledge*) yang relevan untuk perbaikan produk maupun pelayanan.

5. . Analisis Hasil dan Pembahasan

Membahas temuan penelitian berdasarkan hasil pengolahan dan interpretasi data. Termasuk membandingkan hasil dengan studi sebelumnya (literatur) dan membahas implikasinya bagi Grobbak Korea.

6. Kesimpulan dan Saran

Menarik kesimpulan dari keseluruhan proses dan hasil penelitian. Memberikan saran yang aplikatif untuk Grobbak Korea terkait strategi penjualan, promosi, atau peningkatan layanan serta untuk penelitian selanjutnya yang sejenis.

BAB IV

PENGUMPULAN DAN PENGOLAHAN DATA

4.1 Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan untuk memperoleh informasi yang relevan dan mendalam mengenai preferensi pembelian konsumen dan persepsi pelanggan terhadap produk *Grobbak Korea* di Nologaten. Data yang digunakan terdiri dari dua jenis sumber utama, yaitu data transaksi dan data ulasan pelanggan.

4.1.1 Data Transaksi Konsumen

Data transaksi merupakan data primer yang diperoleh langsung dari pihak *Grobbak Korea*. Data ini mencatat seluruh transaksi penjualan selama periode Januari hingga Desember 2024. Data transaksi ini digunakan untuk menemukan pola pembelian konsumen menggunakan metode Association Rule Mining. Berikut merupakan contoh data transaksi bulan Januari 2024

Tanggal & Waktu	ID Struk	Status	Nama Produk	Kategori	Jumlah Produk	Total	Metode Pembayaran
31-01-2025 20:39:49	BMMRSG6Y8ZQD7	Success	Corndog Matcha - Kios	Sweet Corndog	1	28000	Tunai
31-01-2025 20:39:49	BMMRSG6Y8ZQD7	Success	Corndog Strawberry - Kios	Sweet Corndog	1		
31-01-2025 20:35:43	ZVVPTR84GYM84	Success	Corndog Sosis Mozzarella - Shopee	Corndog	1	71400	Kartu
31-01-2025 20:35:43	ZVVPTR84GYM84	Success	+ Oles Langsung		1		
31-01-2025 20:35:43	ZVVPTR84GYM84	Success	Corndog Tiramisu - Shopee	Sweet Corndog	1		
31-01-2025 20:35:43	ZVVPTR84GYM84	Success	Bukokki - Shopee	Tteobokki	1		
31-01-2025 20:35:43	ZVVPTR84GYM84	Success	+ Level 0		1		
31-01-2025 20:34:25	RQQKF7ZNR8N9	Success	Odeng Fish Soup - Grab	Snack	1	40800	Kartu
31-01-2025 20:34:25	RQQKF7ZNR8N9	Success	+ Level 0		1		
31-01-2025 20:34:25	RQQKF7ZNR8N9	Success	Bukokki - Grab	Tteobokki	1		
31-01-2025 20:34:25	RQQKF7ZNR8N9	Success	+ Level 1		1		

Gambar 4. 1 Data Transaksi Konsumen

4.1.2 Data Review Pelanggan

Data review diperoleh melalui *web scraping* dari platform pemesanan online *GoFood*, dengan mengambil ulasan konsumen terhadap *Grobbak Korea*. Data ini bersifat sekunder dan berisi opini pelanggan berupa kalimat bebas.

4.2 Pengolahan Data

Tahap pengolahan data pada penelitian ini dibagi menjadi 2 bagian, yaitu pengolahan data transaksi yang menggunakan *Association Rule*, dan pengolahan data review yang menggunakan analisis sentimen.

4.2.1 Data Transaksi

Data transaksi diolah menggunakan metode *Association Rule*, berikut merupakan tahapan-tahapan dari pengolahan data transaksi:

4.2.1.1 Preprocessing Data

Tahap awal pengolahan data transaksi pada penelitian ini dilakukan melalui proses *preprocessing*. Tujuan dari tahap ini adalah memastikan data yang digunakan berada dalam kondisi bersih, relevan, serta sesuai dengan kebutuhan metode ARMB (Association Rule Mining-Based Analysis). Beberapa langkah yang dilakukan meliputi penghapusan variabel yang tidak diperlukan untuk menyederhanakan data dan meningkatkan efisiensi analisis, penghapusan status pembayaran *void* karena dianggap tidak valid, dan penghapusan nama keterangan tempat pembelian (Grab/Gofood/Shopeefood/Kios) karena Jika variabel ini tetap digunakan, maka hasil aturan asosiasi cenderung bias karena pola yang terbentuk lebih menekankan pada platform pembelian, bukan pada keterkaitan antarproduk. Dengan melalui tahapan tersebut, data transaksi menjadi lebih terstruktur dan siap digunakan untuk analisis lebih lanjut.

ID Struk	Nama Produk	Kategori	Jumlah Produk
BMMRSG6Y8ZQD7	Corndog Matcha - Kios	Sweet Corndog	1
BMMRSG6Y8ZQD7	Corndog Strawberry - Kios	Sweet Corndog	1
ZVVPTR84GYM84	Corndog Sosis Mozzarella - Shopee	Corndog	1
ZVVPTR84GYM84	Corndog Tiramisu - Shopee	Sweet Corndog	1
ZVVPTR84GYM84	Bukokki - Shopee	Tteobokki	1
RQQKF7ZNR8N9	Odeng Fish Soup - Grab	Snack	1
RQQKF7ZNR8N9	Bukokki - Grab	Tteobokki	1

Gambar 4. 2 Data Transaksi Setelah *Preprocessing*

4.2.1.2 Penentuan Parameter *Association Rule*

Penentuan parameter dalam penelitian ini dilakukan dengan metode sederhana, yaitu melalui proses *trial and error*. Langkah tersebut didasarkan pada hasil aturan asosiasi (*rule*) yang terbentuk. Pada *association rule*, terdapat dua parameter utama yaitu *support* dan *confidence*.

Dalam penerapan Algoritma Apriori, perlu ditetapkan terlebih dahulu nilai minimum *support* dan *confidence*. Berikut merupakan penentuan parameter pada penelitian ini.

Tabel 4. 1 Penentuan Parameter

Percobaan	min. support	min. confidence	Hasil Rule
1	0,1	0,1	Tidak ada rule terbentuk
2	0,01	0,2	Tidak ada rule terbentuk Ada rule terbentuk tetapi hanya sedikit
3	0,01	0,1	(4)
4	0,01	0,01	Ada Rule Terbentuk (6)

Berdasarkan Tabel 4.1 pilihan minimal *support* dan *confidence* yang sesuai adalah percobaan ke-4. Sehingga nilai minimal *support* dan *confidence* tersebut yang dipilih untuk melakukan pengolahan data menggunakan Algoritma Apriori.

4.2.1.3 Association Rule 1 Tahun

Dalam penelitian ini, analisis *Association Rule Mining* dilakukan dalam dua tahapan, yaitu berdasarkan data transaksi selama satu tahun penuh dan berdasarkan data transaksi per periode bulanan. Analisis selama satu tahun penuh bertujuan untuk mengidentifikasi pola asosiasi secara menyeluruh atau global. Dengan cakupan waktu yang panjang, pola yang muncul dapat mencerminkan keterkaitan antar produk yang paling dominan dan stabil sepanjang tahun. Hasil ini penting sebagai dasar untuk merumuskan strategi pemasaran jangka panjang, seperti penentuan paket *best seller* atau produk andalan yang terbukti konsisten diminati oleh konsumen. Berikut merupakan *rule* yang muncul pada *Association Rule* 1 tahun:

Tabel 4. 2 Rule Association Rule 1 Tahun

No	antecedents	consequents	support	confidence	lift	count
1	{Bukokki}	{Corndog Mozzarella}	0.014	0.249	0.725	56
2	{Corndog Mozzarella}	{Bukokki}	0.014	0.040	0.725	56
3	{Corndog Mozzarella}	{Corndog Choco}	0.023	0.066	0.439	91
4	{Corndog Choco}	{Corndog Mozzarella}	0.023	0.151	0.439	91
5	{Corndog Sosis Mozzarella}	{Corndog Choco}	0.023	0.080	0.534	94
6	{Corndog Choco}	{Corndog Sosis Mozzarella}	0.023	0.156	0.534	94

No	antecedents	consequents	support	confidence	lift	count
7	{Corndog Sosis Mozzarella}	{Corndog Matcha}	0.012	0.040	0.760	47
8	{Corndog Matcha}	{Corndog Sosis Mozzarella}	0.012	0.222	0.760	47
9	{Corndog Sosis}	{Corndog Mozzarella}	0.012	0.181	0.528	47
10	{Corndog Mozzarella}	{Corndog Sosis}	0.012	0.034	0.528	47
11	{Corndog Sosis Mozzarella}	{Corndog Mozzarella}	0.053	0.183	0.533	215
12	{Corndog Mozzarella}	{Corndog Sosis Mozzarella}	0.053	0.155	0.533	215
13	{Odeng Fish Soup}	{Corndog Mozzarella}	0.019	0.294	0.856	75
14	{Corndog Mozzarella}	{Odeng Fish Soup}	0.019	0.054	0.856	75
15	{Odeng Saus Korea}	{Corndog Mozzarella}	0.023	0.303	0.883	91
16	{Corndog Mozzarella}	{Odeng Saus Korea}	0.023	0.066	0.883	91
17	{Corndog Mozzarella}	{Rabokki}	0.012	0.035	0.739	48
18	{Rabokki}	{Corndog Mozzarella}	0.012	0.254	0.739	48
19	{Topokki Extra Mozzarella}	{Corndog Mozzarella}	0.030	0.374	1.089	122
20	{Corndog Mozzarella}	{Topokki Extra Mozzarella}	0.030	0.088	1.089	122
21	{Topokki Extra Odeng}	{Corndog Mozzarella}	0.011	0.239	0.695	43
22	{Corndog Mozzarella}	{Topokki Extra Odeng}	0.011	0.031	0.695	43
23	{Topokki Original}	{Corndog Mozzarella}	0.032	0.334	0.973	127
24	{Corndog Mozzarella}	{Topokki Original}	0.032	0.092	0.973	127
25	{Corndog Sosis Mozzarella}	{Hot Spicy Bibimmyeon}	0.012	0.042	0.865	49
26	{Hot Spicy Bibimmyeon}	{Corndog Sosis Mozzarella}	0.012	0.253	0.865	49
27	{Corndog Sosis Mozzarella}	{Odeng Fish Soup}	0.016	0.054	0.860	64
28	{Odeng Fish Soup}	{Corndog Sosis Mozzarella}	0.016	0.251	0.860	64
29	{Corndog Sosis Mozzarella}	{Odeng Saus Korea}	0.021	0.071	0.948	83
30	{Odeng Saus Korea}	{Corndog Sosis Mozzarella}	0.021	0.277	0.948	83
31	{Corndog Sosis Mozzarella}	{Rabokki}	0.012	0.040	0.852	47
32	{Rabokki}	{Corndog Sosis Mozzarella}	0.012	0.249	0.852	47
33	{Corndog Sosis Mozzarella}	{Topokki Extra Mozzarella}	0.017	0.059	0.725	69
34	{Topokki Extra Mozzarella}	{Corndog Sosis Mozzarella}	0.017	0.212	0.725	69
35	{Corndog Sosis Mozzarella}	{Topokki Extra Odeng}	0.010	0.035	0.780	41
36	{Topokki Extra Odeng}	{Corndog Sosis Mozzarella}	0.010	0.228	0.780	41
37	{Corndog Sosis Mozzarella}	{Topokki Original}	0.025	0.085	0.902	100

No	antecedents	consequents	support	confidence	lift	count
38	{Topokki Original}	{Corndog Sosis Mozzarella}	0.025	0.263	0.902	100

Berdasarkan table diatas, hasil pengolahan data menggunakan algoritma apriori pada data transaksi 1 tahun, yaitu dari Januari 2024 – Desember 2024 menghasilkan sebanyak 38 *rule*. Namun, hanya terdapat 2 *rule* yang memiliki nilai *lift* > 1 yaitu *rule* no 19 dan 20. Table diatas menunjukkan *antecedents*, *consequents*, *support*, *confidence*, *lift*, dan *count*. Berikut ini merupakan cara pembacaan aturan asosiasi:

Aturan Asosiasi {Topokki Extra Mozzarella} → {Corndog Mozzarella}. Jika pelanggan membeli {Topokki Extra Mozzarella} maka kemungkinan pelanggan juga membeli {Corndog Mozzarella} adalah sebesar 37,4% (*confidence*). Nilai *support* sebesar 0,030 mengindikasikan bahwa kombinasi kedua produk ini muncul dalam 3% dari seluruh transaksi yaitu terdapat 122 transaksi (*count*) dari total data transaksi yang sebanyak 6798 transaksi. Lalu Aturan ini juga sudah valid dikarenakan nilai *lift* > 1, yaitu sebesar 1.089, ini menandakan bahwa peluang kedua produk ini dibeli bersama 1.089 kali lebih besar dibandingkan jika pembelian keduanya terjadi secara kebetulan.

4.2.1.4 AR Setiap Periode (12 Bulan) Menggunakan OCVR

Analisis per periode bulanan dilakukan untuk menangkap dinamika dan variasi pola pembelian konsumen dari waktu ke waktu. Pendekatan ini tidak hanya memperlihatkan asosiasi produk pada tiap bulan, tetapi juga memungkinkan untuk menilai tingkat konsistensi suatu aturan asosiasi dengan menggunakan metrik *Overall Variability of Association Rule (OCVR)*. Aturan dengan nilai OCVR rendah menunjukkan kestabilan hubungan antar produk. Berikut merupakan *rule* yang muncul tiap periode bulan yang nantinya akan diolah untuk menemukan nilai OCVR:

Tabel 4. 3 Aturan yang Muncul Pada Setiap Bulan (Januari)

No	antecedents	consequents	support	confidence	lift	count
1	frozenset({'Corndog Choco'})	frozenset({'Corndog Sosis Mozzarella'})	0.030	0.125	0.470	7
2	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Choco'})	0.030	0.111	0.470	7
3	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.038	0.143	0.513	9

No	antecedents	consequents	support	confidence	lift	count
4	frozenset({'Corndog Mozzarella'})	frozenset({'Corndog Sosis Mozzarella'})	0.038	0.136	0.513	9
5	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Topokki Extra Mozzarella'})	0.013	0.048	0.451	3
6	frozenset({'Topokki Extra Mozzarella'})	frozenset({'Corndog Sosis Mozzarella'})	0.013	0.120	0.451	3
7	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Topokki Original'})	0.034	0.127	1.254	8
8	frozenset({'Topokki Original'})	frozenset({'Corndog Sosis Mozzarella'})	0.034	0.333	1.254	8

Tabel 4. 4 Aturan yang Muncul Pada Setiap Bulan (Februari)

No	antecedents	consequents	support	confidence	lift	count
1	frozenset({'Corndog Choco'})	frozenset({'Corndog Sosis Mozzarella'})	0.037	0.239	0.886	11
2	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Choco'})	0.037	0.136	0.886	11
3	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.047	0.173	0.529	14
4	frozenset({'Corndog Mozzarella'})	frozenset({'Corndog Sosis Mozzarella'})	0.047	0.143	0.529	14
5	frozenset({'Topokki Extra Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.033	0.417	1.276	10
6	frozenset({'Corndog Mozzarella'})	frozenset({'Topokki Extra Mozzarella'})	0.033	0.102	1.276	10
7	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Topokki Original'})	0.017	0.062	0.514	5
8	frozenset({'Topokki Original'})	frozenset({'Corndog Sosis Mozzarella'})	0.017	0.139	0.514	5

Tabel 4. 5 Aturan yang Muncul Pada Setiap Bulan (Maret)

No	antecedents	consequents	support	confidence	lift	count
5	frozenset({'Corndog Choco'})	frozenset({'Corndog Sosis Mozzarella'})	0.024	0.208	0.549	5
6	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Choco'})	0.024	0.063	0.549	5
13	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.081	0.213	0.623	17
14	frozenset({'Corndog Mozzarella'})	frozenset({'Corndog Sosis Mozzarella'})	0.081	0.236	0.623	17
19	frozenset({'Topokki Extra Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.033	0.438	1.282	7
20	frozenset({'Corndog Mozzarella'})	frozenset({'Topokki Extra Mozzarella'})	0.033	0.097	1.282	7
37	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Topokki Original'})	0.014	0.038	0.791	3
38	frozenset({'Topokki Original'})	frozenset({'Corndog Sosis Mozzarella'})	0.014	0.300	0.791	3

Tabel 4. 6 Aturan yang Muncul Pada Setiap Bulan (April)

No	antecedents	consequents	support	confidence	lift	count
1	frozenset({'Corndog Choco'})	frozenset({'Corndog Sosis Mozzarella'})	0.014	0.160	0.544	4
2	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Choco'})	0.014	0.049	0.544	4
3	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.043	0.146	0.434	12
4	frozenset({'Corndog Mozzarella'})	frozenset({'Corndog Sosis Mozzarella'})	0.043	0.128	0.434	12
5	frozenset({'Topokki Extra Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.018	0.185	0.550	5

No	antecedents	consequents	support	confidence	lift	count
6	frozenset({'Corndog Mozzarella'})	frozenset({'Topokki Extra Mozzarella'})	0.018	0.053	0.550	5
7	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Topokki Original'})	0.039	0.134	1.248	11
8	frozenset({'Topokki Original'})	frozenset({'Corndog Sosis Mozzarella'})	0.039	0.367	1.248	11

Tabel 4. 7 Aturan yang Muncul Pada Setiap Bulan (Mei)

No	antecedents	consequents	support	confidence	lift	count
3	frozenset({'Corndog Choco'})	frozenset({'Corndog Sosis Mozzarella'})	0.027	0.200	0.681	8
4	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Choco'})	0.027	0.093	0.681	8
13	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.027	0.093	0.262	8
14	frozenset({'Corndog Mozzarella'})	frozenset({'Corndog Sosis Mozzarella'})	0.027	0.077	0.262	8
27	frozenset({'Topokki Extra Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.020	0.429	1.207	6
28	frozenset({'Corndog Mozzarella'})	frozenset({'Topokki Extra Mozzarella'})	0.020	0.058	1.207	6
49	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Topokki Original'})	0.027	0.093	0.973	8
50	frozenset({'Topokki Original'})	frozenset({'Corndog Sosis Mozzarella'})	0.027	0.286	0.973	8

Tabel 4. 8 Aturan yang Muncul Pada Setiap Bulan (Juni)

No	antecedents	consequents	support	confidence	lift	count
1	frozenset({'Corndog Choco'})	frozenset({'Corndog Sosis Mozzarella'})	0.015	0.140	0.472	6

No	antecedents	consequents	support	confidence	lift	count
2	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Choco'})	0.015	0.052	0.472	6
3	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.046	0.155	0.431	18
4	frozenset({'Corndog Mozzarella'})	frozenset({'Corndog Sosis Mozzarella'})	0.046	0.128	0.431	18
5	frozenset({'Topokki Extra Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.043	0.459	1.277	17
6	frozenset({'Corndog Mozzarella'})	frozenset({'Topokki Extra Mozzarella'})	0.043	0.121	1.277	17
7	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Topokki Original'})	0.020	0.069	0.795	8
8	frozenset({'Topokki Original'})	frozenset({'Corndog Sosis Mozzarella'})	0.020	0.235	0.795	8

Tabel 4. 9 Aturan yang Muncul Pada Setiap Bulan (Juli)

No	antecedents	consequents	support	confidence	lift	count
1	frozenset({'Corndog Choco'})	frozenset({'Corndog Sosis Mozzarella'})	0.021	0.123	0.440	8
2	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Choco'})	0.021	0.075	0.440	8
3	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.058	0.208	0.583	22
4	frozenset({'Corndog Mozzarella'})	frozenset({'Corndog Sosis Mozzarella'})	0.058	0.163	0.583	22
5	frozenset({'Topokki Extra Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.026	0.370	1.040	10
6	frozenset({'Corndog Mozzarella'})	frozenset({'Topokki Extra Mozzarella'})	0.026	0.074	1.040	10
7	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Topokki Original'})	0.026	0.094	0.894	10

8	frozenset({'Topokki Original'})	frozenset({'Corndog Sosis Mozzarella'})	0.026	0.250	0.894	10
---	------------------------------------	--	-------	-------	-------	----

Tabel 4. 10 Aturan yang Muncul Pada Setiap Bulan (Agustus)

No	antecedents	consequents	support	confidence	lift	count
1	frozenset({'Corndog Choco'})	frozenset({'Corndog Sosis Mozzarella'})	0.012	0.069	0.242	5
2	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Choco'})	0.012	0.041	0.242	5
3	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.059	0.207	0.561	25
4	frozenset({'Corndog Mozzarella'})	frozenset({'Corndog Sosis Mozzarella'})	0.059	0.161	0.561	25
5	frozenset({'Topokki Extra Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.026	0.379	1.030	11
6	frozenset({'Corndog Mozzarella'})	frozenset({'Topokki Extra Mozzarella'})	0.026	0.071	1.030	11
7	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Topokki Original'})	0.029	0.099	0.819	12
8	frozenset({'Topokki Original'})	frozenset({'Corndog Sosis Mozzarella'})	0.029	0.235	0.819	12

Tabel 4. 11 Aturan yang Muncul Pada Setiap Bulan (September)

No	antecedents	consequents	support	confidence	lift	count
1	frozenset({'Corndog Choco'})	frozenset({'Corndog Sosis Mozzarella'})	0.014	0.162	0.524	6
2	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Choco'})	0.014	0.044	0.524	6
3	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.055	0.178	0.513	24
4	frozenset({'Corndog Mozzarella'})	frozenset({'Corndog Sosis Mozzarella'})	0.055	0.159	0.513	24

No	antecedents	consequents	support	confidence	lift	count
5	frozenset({'Topokki Extra Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.039	0.386	1.116	17
6	frozenset({'Corndog Mozzarella'})	frozenset({'Topokki Extra Mozzarella'})	0.039	0.113	1.116	17
7	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Topokki Original'})	0.023	0.074	0.923	10
8	frozenset({'Topokki Original'})	frozenset({'Corndog Sosis Mozzarella'})	0.023	0.286	0.923	10

Tabel 4. 12 Aturan yang Muncul Pada Setiap Bulan (Oktober)

No	antecedents	consequents	support	confidence	lift	count
5	frozenset({'Corndog Choco'})	frozenset({'Corndog Sosis Mozzarella'})	0.027	0.174	0.584	12
6	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Choco'})	0.027	0.092	0.584	12
11	frozenset({'Corndog Mozzarella'})	frozenset({'Corndog Sosis Mozzarella'})	0.059	0.182	0.611	26
12	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.059	0.198	0.611	26
19	frozenset({'Topokki Extra Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.027	0.429	1.319	12
20	frozenset({'Corndog Mozzarella'})	frozenset({'Topokki Extra Mozzarella'})	0.027	0.084	1.319	12
29	frozenset({'Topokki Original'})	frozenset({'Corndog Sosis Mozzarella'})	0.027	0.279	0.937	12
30	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Topokki Original'})	0.027	0.092	0.937	12

Tabel 4. 13 Aturan yang Muncul Pada Setiap Bulan (November)

No	antecedents	consequents	support	confidence	lift	count
1	frozenset({'Corndog Choco'})	frozenset({'Corndog Sosis Mozzarella'})	0.028	0.145	0.510	9
2	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Choco'})	0.028	0.097	0.510	9

No	antecedents	consequents	support	confidence	lift	count
3	frozenset({'Corndog Mozzarella'})	frozenset({'Corndog Sosis Mozzarella'})	0.058	0.181	0.636	19
4	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.058	0.204	0.636	19
5	frozenset({'Topokki Extra Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.018	0.240	0.747	6
6	frozenset({'Corndog Mozzarella'})	frozenset({'Topokki Extra Mozzarella'})	0.018	0.057	0.747	6
7	frozenset({'Topokki Original'})	frozenset({'Corndog Sosis Mozzarella'})	0.024	0.348	1.223	8
8	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Topokki Original'})	0.024	0.086	1.223	8

Tabel 4. 14 Aturan yang Muncul Pada Setiap Bulan (Desember)

No	antecedents	consequents	support	confidence	lift	count
5	frozenset({'Corndog Choco'})	frozenset({'Corndog Sosis Mozzarella'})	0.042	0.203	0.780	13
6	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Choco'})	0.042	0.160	0.780	13
11	frozenset({'Corndog Mozzarella'})	frozenset({'Corndog Sosis Mozzarella'})	0.068	0.176	0.678	21
12	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.068	0.259	0.678	21
21	frozenset({'Topokki Extra Mozzarella'})	frozenset({'Corndog Mozzarella'})	0.048	0.500	1.307	15
22	frozenset({'Corndog Mozzarella'})	frozenset({'Topokki Extra Mozzarella'})	0.048	0.126	1.307	15
29	frozenset({'Topokki Original'})	frozenset({'Corndog Sosis Mozzarella'})	0.016	0.192	0.738	5
30	frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Topokki Original'})	0.016	0.062	0.738	5

Data yang digunakan untuk perhitungan selanjutnya adalah *confidence* dan *lift* yang nantinya akan digunakan untuk perhitungan *Index Variability Lift* (CVL) dan *Index Variability Confidence* (CVC) dari masing-masing aturan asosiasi (rule). Kemudian nilai *Index Variability Lift* (CVL)

dan *Index Variability Confidence* (CVC) digunakan untuk perhitungan OCVR. Berikut merupakan hasil perhitungannya:

Tabel 4. 15 Perhitungan OCVR

antecedents	consequents	CVC	CVL	OCVR
frozenset({'Corndog Choco'})	frozenset({'Corndog Sosis Mozzarella'})	29%	30%	16%
frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Choco'})	45%	30%	8%
frozenset({'Corndog Mozzarella'})	frozenset({'Corndog Sosis Mozzarella'})	25%	21%	16%
frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Corndog Mozzarella'})	24%	21%	18%
frozenset({'Topokki Extra Mozzarella'})	frozenset({'Corndog Mozzarella'})	26%	23%	37%
frozenset({'Corndog Mozzarella'})	frozenset({'Topokki Extra Mozzarella'})	29%	23%	9%
frozenset({'Topokki Original'})	frozenset({'Corndog Sosis Mozzarella'})	24%	24%	27%
frozenset({'Corndog Sosis Mozzarella'})	frozenset({'Topokki Original'})	32%	24%	9%

Tabel diatas menunjukkan nilai OCVR berdasarkan aturan asoisasi yang muncul pada tiap bulan. Nilai OCVR menunjukkan variabilitas pada masing-masing rule. Semakin besar nilai OCVR maka semakin tinggi kerentanan perubahannya. Sebaliknya, semakin kecil nilai OCVR maka semakin stabil atau semakin sering aturan asosiasi muncul pada tiap bulan Dalam 1 tahun.

4.2.2 Data Review

Data review diolah menggunakan analisis sentimen untuk mengetahui ulasan negatif konsumen, berikut merupakan tahapan-tahapan dari pengolahan data review:

4.2.2.1 Preprocessing Data

Proses *preprocessing data* dilakukan untuk membersihkan data dari informasi-informasi yang tidak diperlukan. Proses ini dilakukan dengan menggunakan *google collab* dan Bahasa pemrograman *python*. Berikut merupakan tahap-tahap pada proses *preprocessing data*:

a. Case Folding

Tahap *case folding* dilakukan untuk mengubah semua huruf kapital dalam teks menjadi huruf kecil. Tahap *case folding* ditunjukkan pada table dibawah ini:

Tabel 4. 16 Tahapan Case Folding

Sebelum Case Folding	Setelah Case Folding
enak bngt gada tandingan	enak bngt gada tandingan

corndognya enak , tp buat topokki nya kurenggg bgt, saus encer, terlalu kenyal, manis dan asin ngga balance, sayang bgt buat harga segituu isinya jg dikit bgt

porsinya banyak. corndog nya enak. tapi rabokkinya ga enak, ga jelas bgt rasanya, berasa manis doang, kuahnya terlalu encer. buat sekelas topokki gerobakan mending topokki yg di depan superindo jakal atau topokki oppa oppa aja, itu jauhhhh lebih enak dgn harga yg 11 12

Corndog nya enak. Topokki nya enak kenyal pas, tapi kuahnya kurang kental.

dengan harga semurah itu ga berharap pake bahan kualitas premium sih tapi at least tekstur topokkinya bisa lah dibuat lebih chewy dan ga terlalu lembek. rasa saus topokkinya juga ga ada sama sekali terlalu banyak air. corndognya b aja masih kebantu sama kejunya yang masih bisa ditarik. (+) packagingnya oke dan aman, harga murah

saya pesan nya tiramisu kenapa malah vanilla yg dikasih, saya gamau makan lah kalo gini mah. 😞😞

corndognya enak , tp buat topokki nya kurenggg bgt, saus encer, terlalu kenyal, manis dan asin ngga balance, sayang bgt buat harga segituu isinya jg dikit bgt

porsinya banyak. corndog nya enak. tapi rabokkinya ga enak, ga jelas bgt rasanya, berasa manis doang, kuahnya terlalu encer. buat sekelas topokki gerobakan mending topokki yg di depan superindo jakal atau topokki oppa oppa aja, itu jauhhhh lebih enak dgn harga yg 11 12

corndog nya enak. topokki nya enak kenyal pas, tapi kuahnya kurang kental.

dengan harga semurah itu ga berharap pake bahan kualitas premium sih tapi at least tekstur topokkinya bisa lah dibuat lebih chewy dan ga terlalu lembek. rasa saus topokkinya juga ga ada sama sekali terlalu banyak air. corndognya b aja masih kebantu sama kejunya yang masih bisa ditarik. (+) packagingnya oke dan aman, harga murah

saya pesan nya tiramisu kenapa malah vanilla yg dikasih, saya gamau makan lah kalo gini mah. 😞😞

b. Data Cleaning

Tahap *Data Cleaning* digunakan untuk menghilangkan tanda baca yang tidak diperlukan di dalam teks agar data menjadi lebih bersih dan mudah dianalisis. Proses ini menggunakan *library string* dengan atribut *string.punctuation* untuk mengenali seluruh tanda baca yang akan dihapus. Tahap ini ditunjukkan pada tabel di bawah ini:

Tabel 4. 17 Tahapan *Remove Punctuation*

Sebelum <i>Data Cleaning</i>	Setelah <i>Data Cleaning</i>
enak bngt gada tandingan	enak bngt gada tandingan
corndognya enak , tp buat topokki nya kurenggg bgt, saus encer, terlalu kenyal, manis dan asin ngga balance, sayang bgt buat harga segituu isinya jg dikit bgt	corndognya enak tp buat topokki nya kurenggg bgt saus encer terlalu kenyal manis dan asin ngga balance sayang bgt buat harga segituu isinya jg dikit bgt

Sebelum <i>Data Cleaning</i>	Setelah <i>Data Cleaning</i>
porsinya banyak. corndog nya enak. tapi rabokkinya ga enak, ga jelas bgt rasanya, berasa manis doang, kuahnya terlalu encer. buat sekelas topokki gerobakan mending topokki yg di depan superindo jakal atau topokki oppa oppa aja, itu jauh lebih enak dgn harga yg 11 12	porsinya banyak corndog nya enak tapi rabokkinya ga enak ga jelas bgt rasanya berasa manis doang kuahnya terlalu encer buat sekelas topokki gerobakan mending topokki yg di depan superindo jakal atau topokki oppa oppa aja itu jauh lebih enak dgn harga yg
corndog nya enak. topokki nya enak kenyal pas, tapi kuahnya kurang kental.	corndog nya enak topokki nya enak kenyal pas tapi kuahnya kurang kental
dengan harga semurah itu ga berharap pake bahan kualitas premium sih tapi at least tekstur topokkinya bisa lah dibuat lebih chewy dan ga terlalu lembek. rasa saus topokkinya juga ga ada sama sekali terlalu banyak air. corndognya b aja masih kebantu sama kejunya yang masih bisa ditarik. (+) packagingnya oke dan aman, harga murah	dengan harga semurah itu ga berharap pake bahan kualitas premium sih tapi at least tekstur topokkinya bisa lah dibuat lebih chewy dan ga terlalu lembek rasa saus topokkinya juga ga ada sama sekali terlalu banyak air corndognya b aja masih kebantu sama kejunya yang masih bisa ditarik packagingnya oke dan aman harga murah
saya pesan nya tiramisu kenapa malah vanilla yg dikasih, saya gamau makan lah kalo gini mah. 😞😞	saya pesan nya tiramisu kenapa malah vanilla yg dikasih saya gamau makan lah kalo gini mah 😞😞

- *Emoji Removal*

Tahap *Emoji Removal* digunakan untuk menghapus emoji yang terdapat di dalam teks agar hasil data menjadi lebih bersih. Proses ini menggunakan *library re (regular expression)* untuk mengenali pola karakter emoji berdasarkan kode *Unicode*, kemudian menghapusnya dari teks. Tahap *Emoji Removal* ditunjukkan pada tabel di bawah ini:

Tabel 4. 18 Tahapan *Emoji Removal*

Sebelum <i>Emoji Removal</i>	Setelah <i>Emoji Removal</i>
enak bngt gada tandingan	enak bngt gada tandingan
corndognya enak tp buat topokki nya kurenggg bgt saus encer terlalu kenyal manis dan asin ngga balance sayang bgt buat harga segituu isinya jg dikit bgt	corndognya enak tp buat topokki nya kurenggg bgt saus encer terlalu kenyal manis dan asin ngga balance sayang bgt buat harga segituu isinya jg dikit bgt

Sebelum <i>Emoji Removal</i>	Setelah <i>Emoji Removal</i>
porsinya banyak corndog nya enak tapi rabokkinya ga enak ga jelas bgt rasanya berasa manis doang kuahnya terlalu encer buat sekelas topokki gerobakan mending topokki yg di depan superindo jakal atau topokki oppa oppa aja itu jauh lebih enak dgn harga yg	porsinya banyak corndog nya enak tapi rabokkinya ga enak ga jelas bgt rasanya berasa manis doang kuahnya terlalu encer buat sekelas topokki gerobakan mending topokki yg di depan superindo jakal atau topokki oppa oppa aja itu jauh lebih enak dgn harga yg
corndog nya enak topokki nya enak kenyal pas tapi kuahnya kurang kental	corndog nya enak topokki nya enak kenyal pas tapi kuahnya kurang kental
dengan harga semurah itu ga berharap pake bahan kualitas premium sih tapi at least tekstur topokkinya bisa lah dibuat lebih chewy dan ga terlalu lembek rasa saus topokkinya juga ga ada sama sekali terlalu banyak air corndognya b aja masih kebantu sama kejunya yang masih bisa ditarik packagingnya oke dan aman harga murah	dengan harga semurah itu ga berharap pake bahan kualitas premium sih tapi at least tekstur topokkinya bisa lah dibuat lebih chewy dan ga terlalu lembek rasa saus topokkinya juga ga ada sama sekali terlalu banyak air corndognya b aja masih kebantu sama kejunya yang masih bisa ditarik packagingnya oke dan aman harga murah
saya pesan nya tiramisu kenapa malah vanilla yg dikasih saya gamau makan lah kalo gini mah 😞😞	saya pesan nya tiramisu kenapa malah vanilla yg dikasih saya gamau makan lah kalo gini mah

c. *Tokenization*

Tahap *Tokenization* digunakan untuk memisahkan kalimat menjadi potongan-potongan kata (*token*) agar teks lebih mudah dianalisis. Proses ini menggunakan *library nltk (Natural Language Toolkit)* yang menyediakan fungsi *tokenizer* untuk memecah teks secara otomatis berdasarkan struktur bahasa. Tahap *Tokenization* ditunjukkan pada tabel di bawah ini:

Tabel 4. 19 Tahapan *Tokenization*

Sebelum <i>Tokenization</i>	Setelah <i>Tokenization</i>
enak bngt gada tandingan	['enak', 'bngt', 'gada', 'tandingan']
corndognya enak tp buat topokki nya kurenggg bgt saus encer terlalu kenyal manis dan asin ngga balance sayang bgt buat harga segituu isinya jg dikit bgt	['corndognya', 'enak', 'tp', 'buat', 'topokki', 'nya', 'kurenggg', 'bgt', 'saus', 'encer', 'terlalu', 'kenyal', 'manis', 'dan', 'asin', 'ngga', 'balance', 'sayang', 'bgt', 'buat', 'harga', 'segituu', 'isinya', 'jg', 'dikit', 'bgt']

Sebelum <i>Tokenization</i>	Setelah <i>Tokenization</i>
porsinya banyak corndog nya enak tapi rabokkinya ga enak ga jelas bgt rasanya berasa manis doang kuahnya terlalu encer buat sekelas topokki gerobakan mending topokki yg di depan superindo jakal atau topokki oppa oppa aja itu jauhhhh lebih enak	['porsinya', 'banyak', 'corndog', 'nya', 'enak', 'tapi', 'rabokkinya', 'ga', 'enak', 'ga', 'jelas', 'bgt', 'rasanya', 'berasa', 'manis', 'doang', 'kuahnya', 'terlalu', 'encer', 'buat', 'sekelas', 'topokki', 'gerobakan', 'mending', 'topokki', 'yg', 'di', 'depan', 'superindo', 'jakal', 'atau', 'topokki', 'oppa', 'oppa', 'aja', 'itu', 'jauh', 'lebih', 'enak']
corndog nya enak topokki nya enak kenyal pas tapi kuahnya kurang kental	['corndog', 'nya', 'enak', 'topokki', 'nya', 'enak', 'kenyal', 'pas', 'tapi', 'kuahnya', 'kurang', 'kental']
dengan harga semurah itu ga berharap pake bahan kualitas premium sih tapi at least tekstur topokkinya bisa lah dibuat lebih chewy dan ga terlalu lembek rasa saus topokkinya juga ga ada sama sekali terlalu banyak air corndognya b aja masih kebantu sama kejunya yang masih bisa ditarik packagingnya oke dan aman harga murah	['dengan', 'harga', 'semurah', 'itu', 'ga', 'berharap', 'pake', 'bahan', 'kualitas', 'premium', 'sih', 'tapi', 'at', 'least', 'tekstur', 'topokkinya', 'bisa', 'lah', 'dibuat', 'lebih', 'chewy', 'dan', 'ga', 'terlalu', 'lembek', 'rasa', 'saus', 'topokkinya', 'juga', 'ga', 'ada', 'sama', 'sekali', 'terlalu', 'banyak', 'air', 'corndognya', 'b', 'aja', 'masih', 'kebantu', 'sama', 'kejunya', 'yang', 'masih', 'bisa', 'ditarik', 'packagingnya', 'oke', 'dan', 'aman', 'harga', 'murah']
saya pesan nya tiramisu kenapa malah vanilla yg dikasih saya gamau makan lah kalo gini mah	['saya', 'pesan', 'nya', 'tiramisu', 'kenapa', 'malah', 'vanilla', 'yg', 'dikasih', 'saya', 'gamau', 'makan', 'lah', 'kalo', 'gini', 'mah']

d. *Stopword Removal*

Tahap *Stopword Removal* digunakan untuk menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis, seperti “dan”, “yang”, atau “di”. Proses ini menggunakan *library nltk* dengan memanfaatkan daftar *stopwords* bahasa Indonesia yang telah disediakan di dalamnya. Selain itu, dilakukan juga penambahan beberapa kata umum lain secara manual agar proses pembersihan teks menjadi lebih optimal. Tahap *Stopword Removal* ditunjukkan pada tabel di bawah ini:

Tabel 4. 20 Tahapan *Stopword Removal*

Sebelum <i>Stopword Removal</i>	Setelah <i>Stopword Removal</i>
['enak', 'bngt', 'gada', 'tandingan']	['enak', 'bngt', 'gada', 'tandingan']
['corndognya', 'enak', 'tp', 'buat', 'topokki', 'nya', 'kurenggg', 'bgt', 'saus', 'encer', 'terlalu', 'kenyal', 'manis', 'dan', 'asin', 'ngga', 'balance', 'sayang', 'bgt', 'buat', 'harga', 'segituu', 'isinya', 'jg', 'dikit', 'bgt']	['corndognya', 'enak', 'topokki', 'kurenggg', 'bgt', 'saus', 'encer', 'kenyal', 'manis', 'asin', 'ngga', 'balance', 'sayang', 'bgt', 'harga', 'segituu', 'isinya', 'jg', 'dikit', 'bgt']
['porsinya', 'banyak', 'corndog', 'nya', 'enak', 'tapi', 'rabokkinya', 'ga', 'enak', 'ga', 'jelas', 'bgt', 'rasanya', 'berasa', 'manis', 'doang', 'kuahnya', 'terlalu', 'encer', 'buat', 'sekelas', 'topokki', 'gerobakan', 'mending', 'topokki', 'yg', 'di', 'depan', 'superindo', 'jakal', 'atau', 'topokki', 'oppa', 'oppa', 'aja', 'itu', 'jauh', 'lebih', 'enak']	['porsinya', 'corndog', 'enak', 'rabokkinya', 'enak', 'bgt', 'berasa', 'manis', 'doang', 'kuahnya', 'encer', 'sekelas', 'topokki', 'gerobakan', 'mending', 'topokki', 'superindo', 'jakal', 'topokki', 'oppa', 'oppa', 'jauh', 'enak', 'harga']
['corndog', 'nya', 'enak', 'topokki', 'nya', 'enak', 'kenyal', 'pas', 'tapi', 'kuahnya', 'kurang', 'kental']	['corndog', 'enak', 'topokki', 'enak', 'kenyal', 'pas', 'kuahnya', 'kental']
['dengan', 'harga', 'semurah', 'itu', 'ga', 'berharap', 'pake', 'bahan', 'kualitas', 'premium', 'sih', 'tapi', 'at', 'least', 'tekstur', 'topokkinya', 'bisa', 'lah', 'dibuat', 'lebih', 'chewy', 'dan', 'ga', 'terlalu', 'lembek', 'rasa', 'saus', 'topokkinya', 'juga', 'ga', 'ada', 'sama', 'sekali', 'terlalu', 'banyak', 'air', 'corndognya', 'b', 'aja', 'masih', 'kebantu', 'sama', 'kejunya', 'yang', 'masih', 'bisa', 'ditarik', 'packagingnya', 'oke', 'dan', 'aman', 'harga', 'murah']	['harga', 'semurah', 'berharap', 'pake', 'bahan', 'kualitas', 'premium', 'at', 'least', 'tekstur', 'topokkinya', 'chewy', 'lembek', 'saus', 'topokkinya', 'air', 'corndognya', 'b', 'kebantu', 'kejunya', 'ditarik', 'packagingnya', 'oke', 'aman', 'harga', 'murah']
['saya', 'pesan', 'nya', 'tiramisu', 'kenapa', 'malah', 'vanilla', 'yg', 'dikasih', 'saya', 'gamau', 'makan', 'lah', 'kalo', 'gini', 'mah']	['pesan', 'tiramisu', 'vanilla', 'dikasih', 'gamau', 'makan', 'gini', 'mah']

e. *Normalization*

Tahap *Normalization* digunakan untuk menormalkan kata tidak baku atau singkatan menjadi bentuk kata yang sesuai dengan kaidah bahasa Indonesia agar teks lebih konsisten. Proses ini dilakukan menggunakan kamus normalisasi (*dictionary*) yang berisi pasangan kata tidak baku dan padanannya dalam bentuk baku. Melalui kamus tersebut, setiap kata dalam teks diperiksa dan diganti dengan bentuk bakunya secara otomatis. Tahap *Normalization* ditunjukkan pada tabel di bawah ini:

Tabel 4. 21 Tahapan *Normalization*

Sebelum Normalisasi	Setelah Normalisasi
['enak', 'bngt', 'gada', 'tandingan']	['enak', 'banget', 'gada', 'tandingan']
['corndognya', 'enak', 'topokki', 'kurenggg', 'bgt', 'saus', 'encer', 'kenyal', 'manis', 'asin', 'ngga', 'balance', 'sayang', 'bgt', 'harga', 'segituu', 'isinya', 'jg', 'dikit', 'bgt']	['corndognya', 'enak', 'topokki', 'kurenggg', 'banget', 'saus', 'encer', 'kenyal', 'manis', 'asin', 'ngga', 'balance', 'sayang', 'banget', 'harga', 'segituu', 'isinya', 'juga', 'dikit', 'banget']
['porsinya', 'corndog', 'enak', 'rabokkinya', 'enak', 'bgt', 'berasa', 'manis', 'doang', 'kuahnya', 'encer', 'sekelas', 'topokki', 'gerobakan', 'mending', 'topokki', 'superindo', 'jakal', 'topokki', 'oppa', 'oppa', 'jauh', 'enak', 'harga']	['porsinya', 'corndog', 'enak', 'rabokkinya', 'enak', 'banget', 'berasa', 'manis', 'doang', 'kuahnya', 'encer', 'sekelas', 'topokki', 'gerobakan', 'mending', 'topokki', 'superindo', 'jakal', 'topokki', 'oppa', 'oppa', 'jauh', 'enak', 'harga']
['corndog', 'enak', 'topokki', 'enak', 'kenyal', 'pas', 'kuahnya', 'kental']	['corndog', 'enak', 'topokki', 'enak', 'kenyal', 'pas', 'kuahnya', 'kental']
['harga', 'semurah', 'berharap', 'pake', 'bahan', 'kualitas', 'premium', 'at', 'least', 'tekstur', 'topokkinya', 'chewy', 'lembek', 'saus', 'topokkinya', 'air', 'corndognya', 'b', 'kebantu', 'kejunya', 'ditarik', 'packagingnya', 'oke', 'aman', 'harga', 'murah']	['harga', 'semurah', 'berharap', 'pake', 'bahan', 'kualitas', 'premium', 'at', 'least', 'tekstur', 'topokkinya', 'chewy', 'lembek', 'saus', 'topokkinya', 'air', 'corndognya', 'b', 'kebantu', 'kejunya', 'ditarik', 'packagingnya', 'oke', 'aman', 'harga', 'murah']
['pesan', 'tiramisu', 'vanilla', 'dikasih', 'gamau', 'makan', 'gini', 'mah']	['pesan', 'tiramisu', 'vanilla', 'dikasih', 'gamau', 'makan', 'begini', 'mah']

f. *Stemming*

Tahap *Stemming* digunakan untuk mengubah kata berimbuhan menjadi bentuk dasar agar makna kata lebih konsisten dan mudah dianalisis. Proses ini menggunakan *library Sastrawi* yang berfungsi melakukan stemming pada teks berbahasa Indonesia. Selain itu, digunakan juga

library re (regular expression) untuk menghapus akhiran kepemilikan seperti “-nya”, “-ku”, atau “-mu” sebelum proses stemming dilakukan. Tahap *Stemming* ditunjukkan pada tabel di bawah ini:

Tabel 4. 22 Tahapan *Stemming*

Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
['enak', 'banget', 'gada', 'tandingan']	['enak', 'banget', 'gada', 'tanding']
['corndognya', 'enak', 'topokki', 'kurenggg', 'banget', 'saus', 'encer', 'kenyal', 'manis', 'asin', 'ngga', 'balance', 'sayang', 'banget', 'harga', 'segituu', 'isinya', 'juga', 'dikit', 'banget']	['corndog', 'enak', 'topokki', 'kurenggg', 'banget', 'saus', 'encer', 'kenyal', 'manis', 'asin', 'ngga', 'balance', 'sayang', 'banget', 'harga', 'segituu', 'isi', 'juga', 'dikit', 'banget']
['porsinya', 'corndog', 'enak', 'rabokkinya', 'enak', 'banget', 'berasa', 'manis', 'doang', 'kuahnya', 'encer', 'sekelas', 'topokki', 'gerobakan', 'mending', 'topokki', 'superindo', 'jakal', 'topokki', 'oppa', 'oppa', 'jauh', 'enak', 'harga']	['porsi', 'corndog', 'enak', 'rabokki', 'enak', 'banget', 'asa', 'manis', 'doang', 'kuah', 'encer', 'kelas', 'topokki', 'goba', 'mending', 'topokki', 'superindo', 'jakal', 'topokki', 'oppa', 'oppa', 'jauh', 'enak', 'harga']
['corndog', 'enak', 'topokki', 'enak', 'kenyal', 'pas', 'kuahnya', 'kental']	['corndog', 'enak', 'topokki', 'enak', 'kenyal', 'pas', 'kuah', 'kental']
['harga', 'semurah', 'berharap', 'pake', 'bahan', 'kualitas', 'premium', 'at', 'least', 'tekstur', 'topokkinya', 'chewy', 'lembek', 'saus', 'topokkinya', 'air', 'corndognya', 'b', 'kebantu', 'kejunya', 'ditarik', 'packagingnya', 'oke', 'aman', 'harga', 'murah']	['harga', 'murah', 'harap', 'pake', 'bahan', 'kualitas', 'premium', 'at', 'least', 'tekstur', 'topokki', 'chewy', 'lembek', 'saus', 'topokki', 'air', 'corndog', 'b', 'bantu', 'keju', 'tarik', 'packaging', 'oke', 'aman', 'harga', 'murah']
['pesan', 'tiramisu', 'vanilla', 'dikasih', 'gamau', 'makan', 'begini', 'mah']	['pesan', 'tiramisu', 'vanilla', 'kasih', 'gamau', 'makan', 'begini', 'mah']

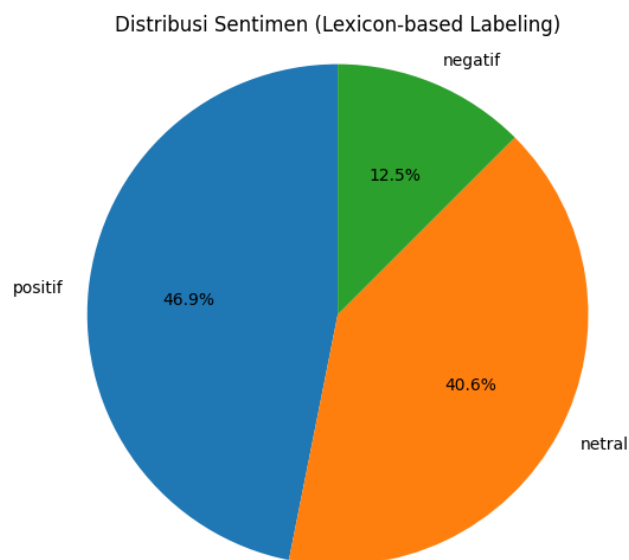
4.2.2.2 Pelabelan dan Pembobotan Data

Setelah dilakukan tahapan *pre-processing*, maka data ulasan yang sudah dibersihkan akan melalui tahapan pelabelan sentimen. Dalam penelitian ini, pelabelan dilakukan dengan menggunakan kamus *lexicon*, yang membagi sentimen ulasan kedalam 3 kategori yaitu positif, netral, serta negatif. Dalam penelitian ini pelabelan data dilakukan dengan bantuan pemrograman *python*.

Tabel 4. 23 Pelabelan *Lexicon*

Review	Skor Lexicon	Sentiment
“enak niii worth dengan harga yg pas bakal re order nih”	3	Positif
“kurang meleleh kejunya”	0	Netral
“saos topokki terlalu encer topokki nya masih agak keras odeng nya bagian yg tidak terendam kuah masih keras”	-3	Negatif

Kemudian didapat hasil sentimen setelah dilakukan pelabelan data, dari 64 data sebanyak 30 data berulasan positif, 24 data berulasan netral, dan 10 data berulasan negatif. Adapun visualisasi persentase hasil pelabelan sentimen ditunjukkan pada grafik pie chart yang berada di gambar:



Gambar 4. 3 Distribusi Sentimen

Tahap selanjutnya setelah data diberi label dengan kamus lexicon adalah melakukan pembobotan kata menggunakan fitur TF-IDF. Fitur ini berperan dalam memberikan nilai bobot pada kata berdasarkan frekuensi kemunculannya dalam dokumen, baik pada tingkat dokumen individu maupun keseluruhan kumpulan dokumen. Dalam penelitian ini, proses perhitungan bobot TF-IDF diimplementasikan melalui pemrograman *Python*.

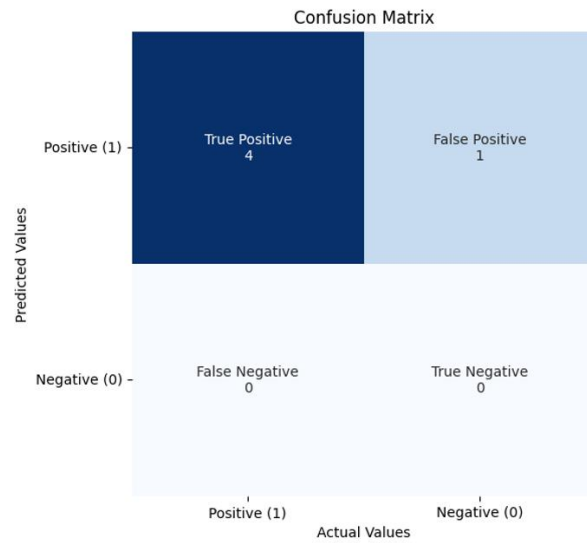
4.2.2.3 Klasifikasi

Setelah dilakukan pelabelan dan pembobotan data, tahap selanjutnya dalam penelitian ini adalah melakukan pembagian data (*splitting data*) menjadi *data training* dan *data testing*. Tujuan dari *splitting data* ini adalah agar model *machine learning* dapat dilatih menggunakan *data training* dan kemudian diuji menggunakan *data testing* untuk mengevaluasi kemampuan model dalam menggeneralisasi data baru secara objektif (Wijiyanto et al., 2024). Pada penelitian ini, pembagian data dilakukan dengan rasio 90% untuk *training* dan 10% untuk *testing*. Pemilihan rasio 90/10 ini didasarkan pada hasil beberapa studi yang menunjukkan bahwa rasio ini sering dipakai untuk memberikan performa awal yang cukup baik dan efisien terutama pada dataset dengan jumlah data yang sedang, sekaligus menjaga agar data testing tetap representatif (Wijiyanto et al., 2024).

Selanjutnya, proses klasifikasi dilakukan menggunakan algoritma *Naïve Bayes*. Algoritma ini dipilih karena memiliki keunggulan dalam kesederhanaan, kecepatan komputasi, serta kemampuan menghasilkan akurasi yang baik pada berbagai jenis data yang telah melalui proses pembobotan dan pelabelan. *Naïve Bayes* bekerja berdasarkan prinsip probabilitas sehingga efektif untuk klasifikasi data dengan banyak fitur dan distribusi yang seragam (Adinugroho et al., 2020; Nurhadi et al., 2022). Selain itu, dalam penelitian-penelitian sebelumnya, algoritma *Naïve Bayes* menunjukkan performa yang lebih baik dibandingkan beberapa metode lain seperti *Neural Network* pada konteks data serupa (Wijiyanto et al., 2024). Berikut merupakan hasil klasifikasi pada penelitian ini:

Tabel 4. 24 Klasifikasi	
<i>Output</i>	<i>Naïve Bayes Classifier</i>
<i>Accuracy</i>	71.43%
<i>Precision</i>	64%
<i>Recall</i>	71%
<i>F1 Score</i>	67%

Dari tabel diatas, diperoleh nilai akurasi sebesar 71.43%, *precision* sebesar 64%, *recall* sebesar 71%, dan *f1 score* sebesar 67% . Selanjutnya, nilai kinerja pada suatu hasil model prediksi klasifikasi diperoleh dengan menggunakan *confusion matrix*. Berikut merupakan hasil *confusion matrix*:



Gambar 4. 4 *Confusion Matrix*

Berdasarkan hasil *confusion matrix*, diketahui bahwa model klasifikasi menghasilkan nilai *True Positive (TP)* sebesar 4, yang berarti terdapat empat data ulasan dengan label asli positif yang berhasil diprediksi positif oleh model. Sementara itu, terdapat *False Positive (FP)* sebanyak 1, yang menunjukkan adanya satu data ulasan dengan label asli negatif tetapi diprediksi positif oleh model. Pada sisi lain, nilai *False Negative (FN)* sebesar 0 menandakan tidak ada data ulasan positif yang salah diprediksi menjadi negatif. Demikian pula, nilai *True Negative (TN)* sebesar 0 menunjukkan bahwa tidak ada data ulasan negatif yang berhasil dikenali dengan benar oleh model. Kondisi ini mengindikasikan bahwa model memiliki kecenderungan bias terhadap kelas positif, sehingga performa klasifikasi lebih baik dalam mengenali ulasan positif dibandingkan dengan ulasan negatif. Hal tersebut diduga dipengaruhi oleh distribusi data yang tidak seimbang, di mana jumlah data positif lebih dominan dibandingkan data negatif, sehingga model kurang optimal dalam mempelajari karakteristik kelas negatif.

4.2.2.4 Hasil Sentimen

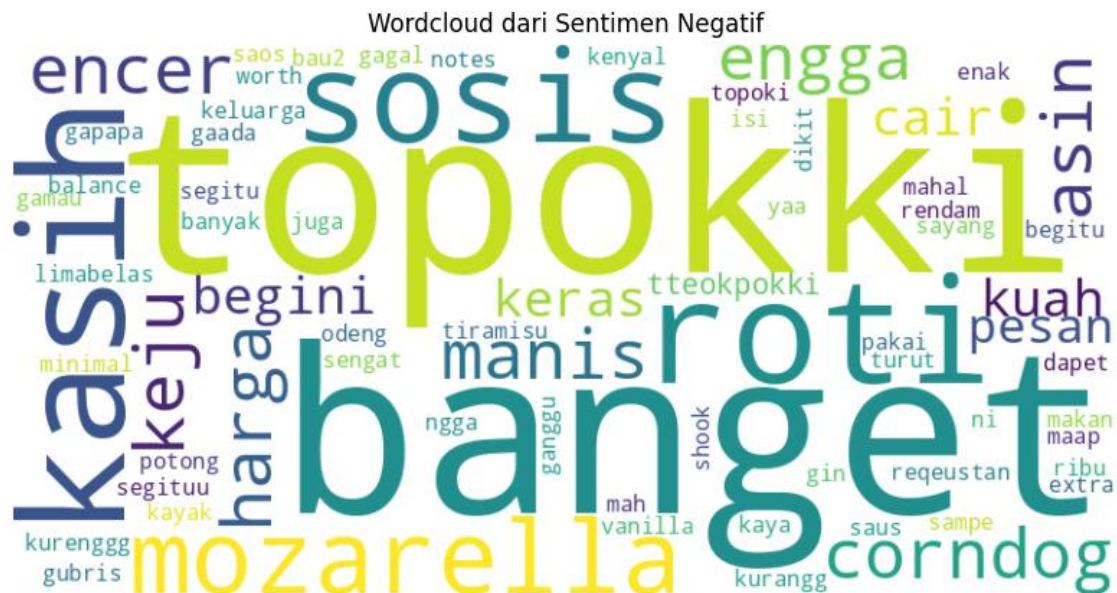
Hasil analisis sentimen divisualisasikan menggunakan *wordcloud* untuk menggambarkan kata-kata yang paling sering muncul pada setiap kategori sentimen, yaitu positif, netral, dan negatif. Visualisasi ini membantu dalam mengidentifikasi fokus opini konsumen terhadap produk dan layanan Grobbak Korea Nologaten.

Pada *wordcloud* sentimen positif, kata “enak”, “corndog”, “harga”, “mantap”, dan “topokki” mendominasi. Hal ini menunjukkan bahwa konsumen memberikan apresiasi terhadap cita rasa produk, terutama pada menu corndog dan topokki, serta menilai harga produk relatif sesuai dengan kualitas yang diterima.



Pada *wordcloud* sentimen netral, kata “pesan”, “saus”, “kuah”, “coba”, dan “corndog” paling sering muncul. Pola ini mengindikasikan bahwa konsumen banyak memberikan komentar

bersifat deskriptif mengenai pengalaman membeli, mencoba produk, dan kualitas penyajian tanpa menunjukkan emosi positif maupun negatif yang kuat.



Gambar 4. 7 Wordcloud Sentimen Negatif

Analisis sentimen dilakukan terhadap data ulasan konsumen dengan tujuan untuk mengetahui persepsi pelanggan terhadap produk yang ditawarkan. Proses analisis menghasilkan daftar kata-kata yang paling sering muncul dalam ulasan dengan label sentimen negatif. Hasilnya ditampilkan pada Gambar 4.5 yang menggambarkan 10 kata paling dominan dalam ulasan negatif. Berdasarkan visualisasi tersebut, kata “topokki” dan “banget” muncul paling sering dengan frekuensi empat kali. Selanjutnya, kata “roti”, “kasih”, “sis”, dan “mozzarella” memiliki frekuensi masing-masing tiga kali. Sementara itu, kata “encer”, “manis”, dan “asin” muncul sebanyak dua kali. Secara umum, hasil ini memperlihatkan bahwa ulasan negatif konsumen cenderung berfokus pada dua aspek utama, yaitu produk utama yang ditawarkan (contohnya topokki, corndog, sosis, mozzarella, roti) serta karakteristik rasa dan tekstur (contohnya asin, manis, encer). Temuan ini memberikan gambaran awal mengenai bagian mana dari produk yang menjadi perhatian khusus dari konsumen. Lalu kata “corndog” muncul sebanyak dua kali, namun setelah diteliti kembali, kata “corndog” pada sentiment negatif ini memiliki makna positif, sebagai contoh “*corndognya enak tp buat topokki nya kurenggg bgt saos encer terlalu kenyal manis dan asin ngga balance sayang bgt buat harga segituu isinya jg dikit bgt*”, review tersebut secara keseluruhan bersentimen negatif, namun kata “corndog” yang

muncul memiliki makna positif, sehingga kata “corndog” bisa dikatakan hal yang positif, hal tersebut sesuai dengan hasil *wordcloud* sentimen positif, sehingga tidak menimbulkan kontradiksi dalam penelitian.

BAB V

PEMBAHASAN

5.1 Analisis Aturan Asosiasi (*Rule*) Data Transaksi 1 tahun

Penelitian ini menggunakan data transaksi 1 tahun yaitu dari Januari 2024 – Desember 2024. Sebanyak 6798 data telah diolah menggunakan metode *Association Rule* yang merupakan salah satu metode di Dalam pengaplikasian *Market Basket Analysis*. Algoritma yang digunakan Dalam penggunaan Metode *Association Rule* ini adalah Algoritma Apriori dengan bantuan *Python*.

Berdasarkan perhitungan *association rule* 1 tahun didapatkan *rule* sebanyak 38 dari 6798 transaksi, namun hanya terdapat 2 *rule* yang memiliki nilai *lift* > 1 yaitu dapat dilihat pada tabel dibawah:

Tabel 5. 1 *Rule Association Rule* 1 tahun

No	antecedents	consequents	support	confidence	lift	count
19	{Topokki Extra Mozzarella}	{Corndog Mozzarella}	0.030	0.374	1.089	122
20	{Corndog Mozzarella}	{Topokki Extra Mozzarella}	0.030	0.088	1.089	122

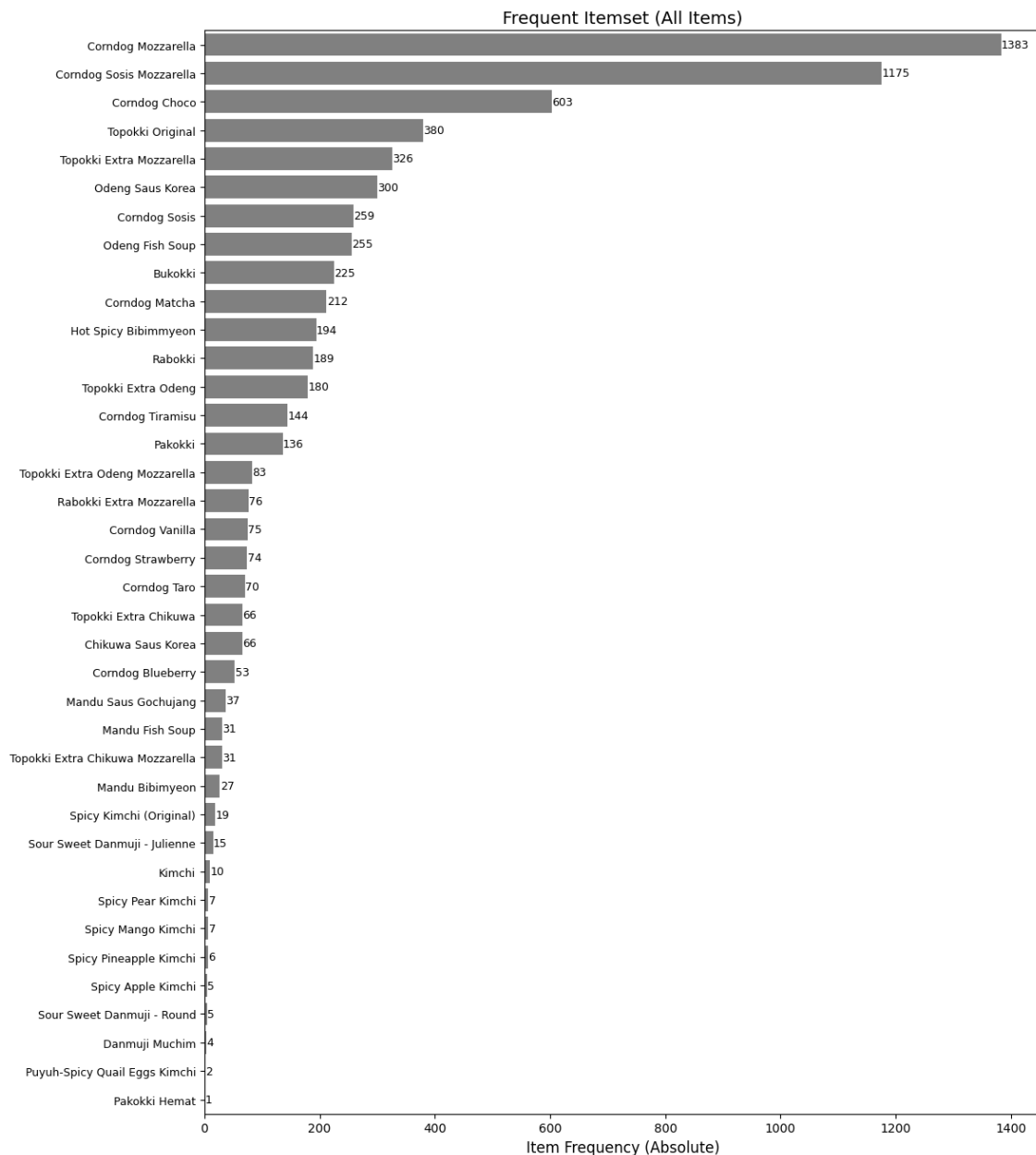
Aturan pertama menunjukkan bahwa dari seluruh transaksi, 3% di antaranya mengandung kombinasi *Topokki Extra Mozzarella* dan *Corndog Mozzarella*, dengan nilai *confidence* sebesar 37,4%. Hal ini berarti bahwa dari seluruh transaksi yang melibatkan *Topokki Extra Mozzarella*, sekitar 37,4% juga mencakup pembelian *Corndog Mozzarella*. Nilai *lift* sebesar 1,089 menunjukkan bahwa keterkaitan antara kedua produk tersebut lebih tinggi dibandingkan dengan kemunculan acak.

Sementara itu, aturan kedua menunjukkan bahwa dari seluruh transaksi yang mencakup *Corndog Mozzarella*, sebanyak 8,8% juga mencakup *Topokki Extra Mozzarella*, dengan nilai *lift* yang sama, yaitu 1,089. Meskipun nilai *confidence*-nya lebih rendah, namun *support* dan *jumlah kemunculan* dari kedua aturan ini sama, yaitu sebesar 0,030 atau setara dengan 122 transaksi.

Temuan ini menunjukkan bahwa terdapat hubungan asosiasi yang saling menguatkan antara dua produk tersebut. Kedua produk memiliki kecenderungan untuk muncul bersama dalam satu transaksi lebih sering dibandingkan dengan kombinasi produk lainnya. Meskipun

jumlah aturan asosiasi yang memiliki lift > 1 tergolong sedikit, informasi ini tetap memberikan gambaran bahwa terdapat produk-produk yang memiliki korelasi pembelian yang cukup signifikan di antara konsumen Grobbak Korea Nologaten sehingga *knowledge* ini nantinya dapat digunakan sebagai analisis strategi marketing.

Hasil pengolahan data juga menunjukkan produk-produk yang sering terbeli (*frequent itemset*). Berikut merupakan hasil *frequent itemset* pada data transaksi 1 tahun:



Gambar 5. 1 *Frequent Itemset*

Berdasarkan Gambar diatas, dapat dilihat hasil perhitungan *frequent itemset*, Produk yang paling sering dibeli adalah Corndog Mozzarella yaitu sebanyak 1383 transaksi, kemudian

Corndog Sosis Mozarella yaitu sebanyak 1175 transaksi, dan Corndog Choco yaitu sebanyak 603 transaksi. Sedangkan produk yang paling jarang terbeli yaitu Pakokki Hemat yaitu sebanyak 1 transaksi, Puyuh-Spicy Quail Eggs Kimchi yaitu sebanyak 2 transaksi, Danmujin Muchim yaitu sebanyak 4 transaksi, dan produk kimchi lainnya. Produk yang paling sering dan paling jarang terbeli dapat digunakan untuk strategi pemasaran nantinya.

5.2 Analisis Aturan Asosiasi (*Rule*) Data Transaksi Pada setiap Periode (12 Bulan) Menggunakan OCVR

Selain dilakukan *association rule* selama 1 tahun, dilakukan juga *association rule* untuk menemukan aturan asosiasi dilakukan setiap bulan selama satu tahun. Dengan kata lain, data transaksi diolah sebanyak 12 kali, sesuai dengan jumlah bulan Dalam setahun. Berikut ini merupakan jumlah transaksi, jumlah item, serta jumlah aturan asosiasi (*rule*) yang berhasil dihasilkan.

Tabel 5. 2 Jumlah *Rule* Tiap Periode

Bulan	Jumlah Transaksi	Jumlah Item	Jumlah Rule
Januari	237	398	44
Februari	300	514	80
Maret	211	355	44
April	279	477	60
Mei	293	480	58
Juni	392	644	56
Juli	379	658	46
Agustus	421	730	46
September	436	748	46
Oktober	440	731	32
November	327	543	36
Desember	311	520	34

Berdasarkan tabel diatas, dapat dilihat jumlah transaksi, jumlah item, dan jumlah *rule* yang terbentuk pada tiap bulan. Hasil tersebut didapatkan dengan penggunaan parameter *minimal support* 0,01 dan *minimal confidence* 0,01. Transaksi terbanyak terdapat pada bulan September

yaitu berjumlah 436 transaksi, sedangkan *rule* terbanyak terdapat pada bulan Februari yaitu sebanyak 80 *rule*.

Pengolahan data dilakukan setiap periode (Bulan) dalam satu tahun bertujuan untuk memperoleh nilai *Overall Change in Variability of Rules* (OCVR). Nilai OCVR ini merepresentasikan seberapa besar perubahan atau variasi pada *association rule* di setiap periode mingguan selama satu bulan. Melalui analisis ini, aturan-aturan asosiasi yang terbentuk dapat dikelompokkan berdasarkan tingkat variabilitasnya masing-masing.

Berdasarkan hasil pengolahan data, ditemukan sebanyak 8 *association rule* yang konsisten muncul setiap periode (Bulan) selama satu tahun. *Rule* tersebut kemudian digunakan dalam proses analisis untuk menghitung nilai *Overall Change in Variability of Rules* (OCVR) masing-masing. Perhitungan OCVR menggunakan dua parameter utama, yaitu nilai *confidence* dan *lift*. Dari kedua parameter ini, dibentuk dua indikator, yaitu *Confidence Variability Index* (CVC) dan *Lift Variability Index* (CVL), yang berperan dalam menentukan nilai akhir OCVR. Selanjutnya, hasil OCVR dari masing-masing *rule* diurutkan dari nilai yang paling rendah hingga yang paling tinggi. Berikut adalah urutan *association rule* berdasarkan nilai OCVR tersebut:

Tabel 5. 3 OCVR

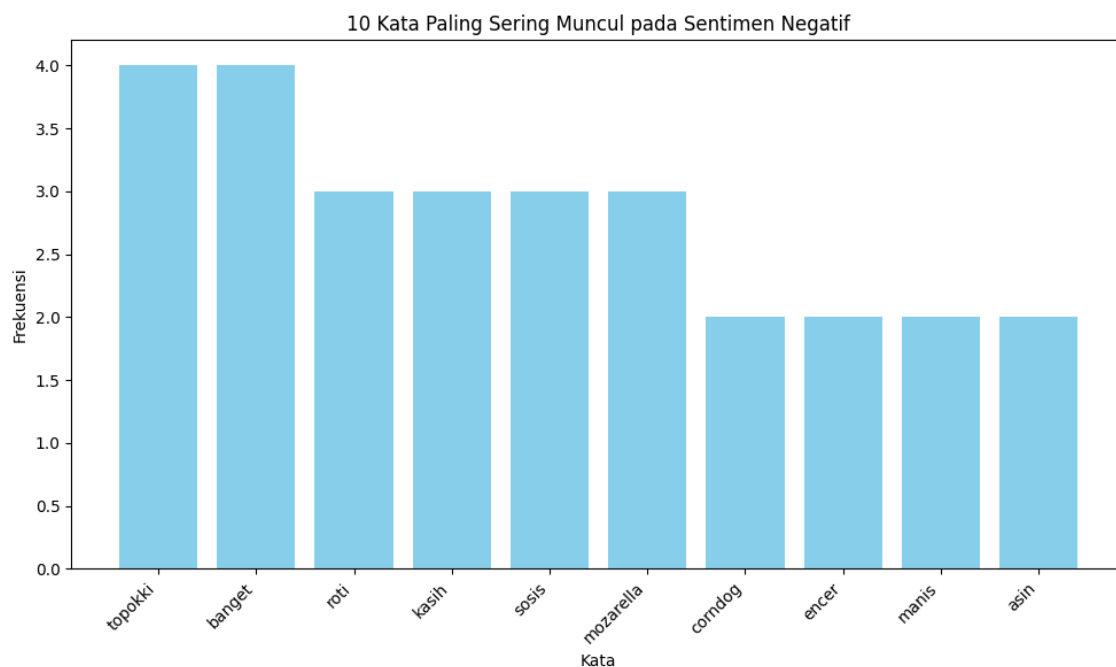
No	antecedents	consequents	OCVR
1	{Corndog Sosis Mozzarella}	{Corndog Choco}	8%
2	{Corndog Sosis Mozzarella}	{Topokki Original}	9%
3	{Corndog Mozzarella}	{Topokki Extra Mozzarella}	9%
4	{Corndog Mozzarella}	{Corndog Sosis Mozzarella}	16%
5	{Corndog Choco}	{Corndog Sosis Mozzarella}	16%
6	{Corndog Sosis Mozzarella}	{Corndog Mozzarella}	18%
7	{Topokki Original}	{Corndog Sosis Mozzarella}	27%
8	{Topokki Extra Mozzarella}	{Corndog Mozzarella}	37%

Tabel diatas menyajikan urutan *association rule* berdasarkan nilai *Overall Variability of Association Rule* (OCVR), dimulai dari nilai yang paling kecil hingga yang paling besar. Mengacu pada penelitian Papavasileiou & Tsadiras (2011), nilai OCVR yang berada dalam rentang 1% hingga 30% menunjukkan tingkat variabilitas yang rendah, sehingga *rule* tersebut cenderung stabil dan tidak mudah terpengaruh oleh perubahan antar periode. Sebaliknya, apabila nilai OCVR melebihi 30%, maka *rule* tersebut dianggap sangat fluktuatif dan tidak

dapat digunakan secara konsisten di setiap periode, melainkan hanya muncul pada waktu-waktu tertentu saja.

Dari tabel tersebut dapat diamati bahwa mayoritas rule memiliki nilai OCVR di bawah 30%, yang menunjukkan bahwa rule tersebut relatif konsisten dan dapat digunakan dalam strategi asosiasi produk secara berkelanjutan. Namun, rule pada urutan kedelapan menunjukkan nilai OCVR sebesar 37%, yang berada di atas ambang batas stabilitas. Hal ini mengindikasikan bahwa rule tersebut kurang dapat diandalkan untuk digunakan secara konsisten, karena hanya relevan pada kondisi atau periode tertentu. *Knowledge* ini digunakan untuk strategi marketing pada Grobbak Korea Nologaten untuk meningkatkan penjualan.

5.3 Analisis Sentimen Data Review



Gambar 5. 2 Kata Paling Sering Muncul pada Sentimen Negatif

Hasil analisis memberikan insight yang lebih mendalam ketika ditafsirkan. Kata “topokki” menjadi yang paling sering muncul dalam ulasan negatif. Hal ini mengindikasikan bahwa produk topokki masih menjadi titik lemah dari segi kepuasan konsumen. Keluhan yang sering muncul antara lain terkait rasa saus yang dianggap kurang sesuai selera lokal serta tekstur saus yang “encer”. Kondisi ini menunjukkan perlunya evaluasi formulasi produk topokki agar lebih sesuai dengan ekspektasi konsumen, misalnya dengan memperkaya cita rasa atau menyesuaikan tingkat kekentalan saus. Kemunculan kata “banget” memperkuat intensitas ketidakpuasan konsumen. Kata ini biasanya muncul dalam bentuk frasa seperti “asin banget”,

“keras banget”, atau “kurang banget”, yang menunjukkan keluhan dengan tekanan emosional. Artinya, meskipun jumlah ulasan negatif tidak selalu dominan, ekspresinya kuat dan dapat memengaruhi citra produk jika tidak segera ditangani. Kata “roti” dan “sosis” menegaskan adanya masalah kualitas bahan utama, khususnya pada produk *corndog*. Keluhan yang berkaitan dengan roti biasanya mencakup tekstur yang kurang lembut, sementara sosis dikaitkan dengan rasa yang dianggap standar atau tidak berbeda dari produk sejenis di pasaran. Begitu pula dengan kata “mozzarella”, yang sering dikaitkan dengan kualitas lelehan keju serta porsi yang dianggap kurang memuaskan. Hal ini dapat mengarah pada persepsi bahwa produk kurang memberikan nilai lebih dibanding ekspektasi konsumen. Kata “kasih” banyak muncul dalam konteks keluhan mengenai porsi atau jumlah tambahan, seperti “nggak dikasih saus”, “nggak dikasih banyak topping”, atau “nggak dikasih sesuai pesanan”. Hal ini mengindikasikan bahwa aspek pelayanan, terutama konsistensi dalam pemberian topping atau saus, juga menjadi perhatian penting. Sementara itu, kata “encer”, “manis”, dan “asin” berhubungan dengan persepsi konsumen terhadap cita rasa dan keseimbangan bumbu. Saus yang terlalu encer, rasa yang terlalu asin, atau manis yang berlebihan menunjukkan bahwa standar rasa masih perlu distandarisasi agar lebih konsisten. Konsumen cenderung menginginkan produk dengan cita rasa yang stabil, sesuai ekspektasi, dan tidak berbeda jauh dari satu pesanan ke pesanan lain. *Knowledge* yang didapat ini nantinya dapat dipakai sebagai acuan dalam perancangan strategi marketing untuk meningkatkan penjualan/pendapatan *Grobbak Korea Nologaten*.

5.4 Strategi Marketing Berdasarkan Hasil Analisis Preferensi Pembelian dan Sentimen Konsumen

Berdasarkan hasil analisis data yang telah dilakukan, dapat dirumuskan strategi pemasaran produk baru yang sejalan dengan temuan penelitian guna meningkatkan penjualan. Dari hasil tersebut, terdapat beberapa alternatif strategi pemasaran yang dapat diterapkan baik itu yang didapat dari data transaksi maupun dari data review.

1. Pemanfaatan data transaksi

- Strategi Bundling dan Cross-Selling Berbasis Pola Asosiasi Konsumen

Hasil analisis ARMBA selama satu tahun menunjukkan bahwa keterkaitan paling kuat dan paling sering muncul terdapat pada produk *Corndog Mozzarella* dan *Topokki Extra Mozzarella*. Hubungan ini ditunjukkan oleh nilai lift yang lebih besar dari 1, sehingga dapat diinterpretasikan bahwa konsumen yang membeli salah satu dari produk tersebut memiliki kecenderungan tinggi untuk membeli produk pasangannya. Selanjutnya, analisis ARMBA per

bulan dengan pengukuran OCVR memperlihatkan bahwa asosiasi antara *Corndog Mozzarella* dan *Topokki Extra Mozzarella* tidak hanya muncul secara konsisten pada setiap periode, tetapi juga memiliki tingkat variabilitas yang rendah. Artinya, pola keterkaitan kedua produk ini stabil dari waktu ke waktu. Konsistensi temuan ini memperkuat dasar bagi penyusunan strategi pemasaran. Oleh karena itu, penerapan strategi bundling berupa paket *best seller*, misalnya “*Combo Mozzarella Delight*” yang berisi *Corndog Mozzarella* dan *Topokki Extra Mozzarella*, menjadi sangat relevan. Strategi ini tidak hanya didukung oleh frekuensi asosiasi yang tinggi sepanjang tahun, tetapi juga diperkuat oleh kestabilan hubungan kedua produk di setiap periode, sehingga dapat diimplementasikan sebagai program promosi berkelanjutan untuk meningkatkan nilai transaksi rata-rata pelanggan sekaligus memperkuat citra menu andalan. Selain melalui bundling, hasil asosiasi ini juga dapat dimanfaatkan dalam sistem penjualan digital, misalnya dengan menampilkan rekomendasi “*Orang yang membeli produk ini juga membeli ...*” di platform e-commerce atau aplikasi pemesanan. Pendekatan ini dapat meningkatkan peluang pembelian silang (*cross-selling*) secara otomatis dengan memanfaatkan pola perilaku konsumen yang konsisten.

- Strategi Promosi Pengenalan dan Revitalisasi Produk

Berdasarkan hasil analisis ARMB, produk varian Kimchi jarang muncul dalam aturan asosiasi, yang menunjukkan bahwa menu ini relatif jarang dipilih sebagai side dish. Hal ini menandakan perlunya strategi pengenalan agar konsumen lebih aware terhadap keberadaan menu tersebut. Pendekatan yang dapat diterapkan adalah *sampling* atau *free add-on*, misalnya memberikan Kimchi sebagai pelengkap gratis untuk pembelian menu utama tertentu. Strategi ini efektif untuk meningkatkan eksposur produk tanpa harus menuntut konsumen membeli secara terpisah. Selain itu, hasil analisis frekuensi item tunggal menunjukkan bahwa beberapa varian menu utama, khususnya Corndog Strawberry, Corndog Taro, dan Corndog Blueberry, memiliki tingkat pembelian yang rendah dibandingkan varian populer lainnya. Untuk itu, diperlukan strategi *revitalisasi* agar produk-produk tersebut tidak tertinggal. Salah satu pendekatan adalah dengan mengombinasikannya bersama produk best seller dalam bentuk paket promo, seperti “*Mix & Match Combo*” yang menggabungkan satu corndog varian baru dengan corndog populer. Dengan cara ini, konsumen terdorong mencoba varian yang jarang diminati tanpa merasa berisiko, sekaligus membuka peluang bagi produk tersebut untuk menemukan segmen pasar yang tepat.

2. Pemanfaatan data review

- Strategi Evaluasi Kualitas Produk Topokki serta Standarisasi Rasa

Berdasarkan hasil analisis ulasan konsumen, ditemukan beberapa keluhan yang menyoroti aspek kualitas produk, khususnya pada menu topokki. Beberapa konsumen menyebutkan bahwa saus topokki memiliki tekstur yang terlalu encer dan tingkat kekenyalan topokki yang kurang optimal. Selain itu, terdapat pula ulasan yang menyinggung ketidakseimbangan cita rasa, di mana rasa saus dinilai terlalu manis atau terlalu asin. Oleh karena itu, strategi yang diusulkan adalah melakukan evaluasi ulang terhadap formulasi saus serta proses pemasakan topokki, disertai penerapan standarisasi takaran bumbu untuk menjaga konsistensi rasa pada setiap penyajian.

- Strategi Konsistensi dan Ketepatan Pesanan

Hasil analisis sentimen juga menunjukkan adanya keluhan konsumen terkait ketidaksesuaian pesanan, seperti menu atau topping yang tidak sesuai dengan pesanan awal (misalnya, konsumen memesan varian tiramisu namun menerima rasa vanilla, atau mendapatkan tteokbokki tanpa tambahan mozzarella). Untuk mengatasi hal tersebut, strategi yang diambil adalah memperkuat prosedur pemeriksaan pesanan melalui penerapan *standard operating procedure* (SOP) pemeriksaan ganda sebelum pesanan diberikan kepada pelanggan, serta memastikan setiap pesanan diverifikasi dengan label atau tanda pengecekan agar kesalahan dapat diminimalisir.

BAB VI

PENUTUP

6.1 Kesimpulan

1. Berdasarkan analisis data transaksi menggunakan metode Association Rule Mining dengan algoritma apriori (baik ARMBA satu tahun penuh maupun per periode), ditemukan bahwa produk *Corndog Mozzarella* dan *Topokki Extra Mozzarella* memiliki keterkaitan yang saling menguatkan dengan nilai lift > 1 . Pola ini menunjukkan kecenderungan konsumen membeli kedua produk tersebut secara bersamaan lebih sering dibandingkan kombinasi lainnya. Selain itu, hasil analisis OCVR memperlihatkan bahwa sebagian besar aturan asosiasi relatif konsisten (di bawah 30%), meskipun terdapat satu aturan dengan OCVR 37% yang kurang stabil dan hanya relevan pada periode tertentu. Temuan ini menegaskan adanya pola pembelian yang dapat dimanfaatkan sebagai dasar strategi pemasaran.
2. Hasil analisis sentimen ulasan konsumen menunjukkan mayoritas ulasan bernuansa positif, namun ulasan negatif menyoroti aspek penting yang perlu diperhatikan, yaitu: kualitas *topokki* (saus dianggap terlalu encer dan kurang sesuai selera), tekstur roti *corndog* yang kurang lembut, rasa sosis yang standar, serta kualitas *mozzarella* yang belum memenuhi ekspektasi. Dari sisi pelayanan, muncul keluhan mengenai porsi atau topping yang tidak konsisten, sementara standar rasa (asin, manis, dan encer) juga dianggap belum stabil. Hal ini mengindikasikan perlunya evaluasi cita rasa, kualitas bahan baku, serta konsistensi pelayanan untuk meningkatkan kepuasan konsumen.
3. Berdasarkan hasil analisis pola pembelian dan sentimen, dirumuskan 6 strategi pemasaran utama, yaitu:
 - Strategi Bundling dan Cross-Selling pada produk dengan asosiasi kuat (misalnya paket *Combo Mozzarella Delight* yang menggabungkan *Corndog Mozzarella* dan *Topokki Extra Mozzarella*).
 - Strategi Promosi Pengenalan dan Revitalisasi Produk, khususnya pada menu yang jarang terbeli (seperti varian Kimchi dan *corndog* rasa manis), melalui sampling, free add-on, atau paket kombinasi dengan menu best seller.

- Strategi Evaluasi Kualitas Produk Topokki serta Standarisasi Rasa.
- Strategi Konsistensi dan Ketepatan Pesanan.

6.2 Saran

Berikut merupakan saran dari penelitian ini:

1. Saran untuk Peneliti Selanjutnya

Penelitian selanjutnya disarankan untuk memperluas ruang lingkup analisis dengan menambahkan variabel atau dimensi lain yang belum dibahas dalam penelitian ini, seperti faktor waktu transaksi, segmentasi konsumen, atau pengaruh promosi terhadap pola pembelian. Selain itu, penerapan metode *Association Rule* dapat dikembangkan dengan algoritma lain seperti *FP-Growth* atau *Eclat* untuk membandingkan tingkat efisiensi dan akurasi hasil aturan asosiasi. Lalu penelitian selanjutnya juga disarankan untuk melakukan scrapping data review dari berbagai platform seperti *Grab* dan *ShopeeFood* yang tidak penulis *scrape*, sehingga dapat memperbesar ukuran *dataset*.

2. Saran untuk Grobbak Korea Nologaten

Diharapkan bagi pihak Grobbak Korea Nologaten dapat meninjau hasil penelitian ini sebagai dasar dalam merancang strategi peningkatan penjualan. Temuan dari analisis data transaksi menunjukkan adanya kombinasi produk yang berpotensi untuk dikembangkan melalui strategi bundling atau promosi bersama, sedangkan hasil analisis ulasan pelanggan menyoroti aspek kualitas produk dan konsistensi pelayanan yang perlu diperbaiki. Usulan tersebut diharapkan dapat menjadi masukan untuk meningkatkan daya saing serta mendorong pertumbuhan penjualan Grobbak Korea Nologaten.

DAFTAR PUSTAKA

- Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules. *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*, 487–499.
- Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 22(2), 207–216.
- Alfiqra, & Khasanah, A. U. (2020). Implementation of Market Basket Analysis Based on Overall Variability of Association Rule (OCVR) on Product Marketing Strategy. *IOP Conference Series: Materials Science and Engineering*, 722(1), 012068.
- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems*, 28(2), 15–21.
- Costa, T., Silva, F., & Ferreira, L. P. (2017). Improve the extrusion process in tire production using Six Sigma methodology. *Procedia Manufacturing*, 1104-1111.
- Elvin, N. K., Sulaiha, N. I., Laily, N. A., & Lailatul, M. (2022). Aplikasi metode association rule pada promosi produk. *Journal of Management and Digital Business*.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *From Data Mining to Knowledge Discovery in Databases*. AI Magazine.
- Fransiscus, H., Cynthia, P. J., & Isabella, S. A. (2014). Implementasi Metode Six Sigma DMAIC untuk Mengurangi Paint Bucket Cacat di PT X. *Jurnal Rekayasa Sistem Industri*, 3(2), 53-63.
- Ghiffari Ibrahim, A. H. (2013, Juli). Analisis Six Sigma Untuk Mengurangi Jumlah Cacat di Stasiun Kerja Sablon (Studi Kasus: CV. Miracle). *Jurnal Online Institut Teknologi Nasional*, 1.
- Han, J. P. (2011). *Data Mining: Concepts and Techniques (3rd ed.)*. San Francisco: Morgan Kaufmann.
- Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques (3rd ed.)*. San Francisco: Morgan Kaufmann.
- Han, J., Pei, J., Yin, Y., & Mao, R. (2000). Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach. *Data Mining and Knowledge Discovery*, 8(1), 53–87.
- Harlinda, L., & Ramdan, S. (2025). Data Mining Approach to Improve Minimarket Sales using Association Rule Method. *JURNAL INFORMATIKA*, 10-14 .
- Ibrahim, G., Harsono, A., & Bakar, A. (2013, Juli). Analisis Six Sigma Untuk Mengurangi Jumlah Cacat di Stasiun Kerja Sablon (Studi Kasus: CV. Miracle). *Jurnal Online Institut Teknologi Nasional*, 1.
- Indrawati Sri, M. R. (2015). Manufacturing Continuous Improvement Using Lean Six Sigma: An Iron Ores Industry Case Application. *Procedia Manufacturing*, 528-534.
- Indrawati, S., & Ridwansyah, M. (2015). Manufacturing Continuous Improvement Using Lean Six Sigma: An Iron Ores Industry Case Application. *Procedia Manufacturing*, 528-534.
- Khoirul, A. (2021, November 30). *Mayora Sebut Tren Korea Bikin Ketat Bisnis Makanan Ringan RI*. Retrieved from [cnbcindonesia.com: https://www.cnbcindonesia.com/news/20211130191255-4-295636/mayora-sebut-tren-korea-bikin-ketat-bisnis-makanan-ringan-ri](https://www.cnbcindonesia.com/news/20211130191255-4-295636/mayora-sebut-tren-korea-bikin-ketat-bisnis-makanan-ringan-ri)
- Kusuma, & Fendy, M. (2014). Analisis Kualitas Produk Pakan Ternak Dengan Metode Six Sigma Di PT. Charoen Pokphand Indonesia (Tbk). *JTM*, 54-62.
- Larose, D. T. (2015). *Data Mining and Predictive Analytics (2nd ed.)*. Hoboken, NJ: John Wiley & Sons.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
- McCallum, A., & Nigam, K. (1998). A comparison of event models for Naive Bayes text classification. *In AAAI-98 Workshop on Learning for Text Categorization*, 41–48.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113.
- Nanang, W. (2024, November 21). *Nilai Ekspor Makanan Halal Korea ke Indonesia Tembus Rp3,7 Triliun*. Retrieved from Sindonews.com: <https://ekbis.sindonews.com/read/1492037/34/nilai-ekspor-makanan-halal-korea-ke-indonesia-tembus-rp37-triliun-1732191089>
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1–135.
- Papavasileiou, V., & Athanasios, T. (2011). Time Variations of Association Rules in Market Basket Analysis. *IFIP Advances in Information and Communication Technology (IFIP AICT)*, 191–200.
- Prasetyo, E., & Wibowo, A. (2020). Analisis Sentimen Ulasan Pelanggan untuk Meningkatkan Kualitas Produk UMKM Menggunakan Metode Naive Bayes. *Jurnal Teknologi Informasi dan Komunikasi*, 7(2), 123-130.
- Putri, & Fatma, C. (2010). Upaya Menurunkan Jumlah Cacat Produk Shuttlecock Dengan Metode Six Sigma. *Widya Teknika*, 18(2), 14-23.
- Putri, C. F. (2010). Upaya Menurunkan Jumlah Cacat Produk Shuttlecock Dengan Metode Six Sigma. *Widya Teknika*, 18(2), 14-23.
- Rish, I. (2001). An empirical study of the naive Bayes classifier. *In IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, Vol. 3, No. 22, pp. 41–46.
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys (CSUR)*, 34(1), 1–47.
- Sharma, N., & Bhardwaj, A. (2012). A Review of Data Mining Techniques for Finding Interesting Patterns. *International Journal of Computer Applications*, 60(19), 35–40.
- Soni, D. (2014). Comparative Analysis of Apriori Algorithm and FP-Growth Algorithm. *International Journal of Engineering and Computer Science*, 3(11), 9282–9285.
- Sucipto, Sulistyowati, D. P., & Anggarini, S. (2017). Pengendalian Kualitas Pengalengan Jamur dengan Metode Six Sigma di PT Y.Pasuruan, Jawa Timur. *Jurnal Teknologi dan Manajemen Agroindustri*, 1-7.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2), 267–307.
- Tan, H. T. (2012). Metode DMAIC Sebagai Solusi Pengendalian Kualitas Produksi Sepatu Tambang: Studi Kasus PT Mangul Jaya-Bekasi. *ComTech*, 3, 509-523.
- Tan, P. N. (2019). *Introduction to Data Mining (2nd ed.)*. Boston: Pearson.
- Tan, P. N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*. Boston: Pearson Education.
- Turban, E., Sharda, R., & Delen, D. (2011). *Decision Support and Business Intelligence Systems (9th ed.)*. Boston: Pearson Education.
- Vitho, I., Ginting, E., & Anizar. (2013). Aplikasi Six Sigma Untuk Menganalisis Faktor-faktor Penyebab Kecacatan Produk Crumb Rubber Sir 20 Pada Pt. XYZ. *e-Jurnal Teknik Industri FT USU Vol 3, No. 4*, 23-28.
- Wisnubroto, P., & Rukmana, A. (2015). Pengendalian Kualitas Produk dengan Pendekatan Six Sigma dan Analisis Kaizen serta New Seven Tools Sebagai Usaha Pengurangan Kecacatan Produk. *Jurnal Teknologi*, 65-74.

- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques (3rd ed.)*. Burlington, MA: Morgan Kaufmann.
- Yoga, S. (2023, Agustus 6). *Indonesia Jadi Pasar Penting Bisnis Makanan Korea*. Retrieved from Kompas.com: <https://money.kompas.com/read/2023/08/06/082245726/indonesia-jadi-pasar-penting-bisnis-makanan-korea>
- Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, 13(3), 55–75.
- Yuliana, Nasution, Y. N., & Wasono. (2017). Penggunaan Metode Kaizen Pada Tahap Improve Dalam Six Sigma (Studi Kasus: Perusahaan Air Minum Dalam Kemasan (AMDK) Merk RAMA Produksi PT Ranam Mahakam Indonesia). *Jurnal Eksponensial*.

LAMPIRAN

1. Olah Data Transaksi

```
import pandas as pd

import re

from mlxtend.frequent_patterns import apriori, association_rules

import matplotlib.pyplot as plt

import networkx as nx

import seaborn as sns

import glob


# Load data transaksi

file_path = "/content/data_1_tahun.csv"

df = pd.read_csv(file_path)


# PREPROCESSING: Hapus keterangan tempat dari Nama Produk

def clean_product_name(product_name):

    return re.sub(r" - (Kios|Shopee|Grab|GoFood|GrabFood)$", "", product_name)


df["Nama Produk"] = df["Nama Produk"].apply(clean_product_name)


# Transformasi data ke format transaksi

basket = df.groupby(['ID Struk', 'Nama Produk'])['Nama Produk'].count().unstack().fillna(0)


# Binarisasi data (ubah angka menjadi 1 atau 0)

basket = basket.applymap(lambda x: 1 if x > 0 else 0)


# Mencari frequent itemsets menggunakan algoritma Apriori

frequent_itemsets = apriori(basket, min_support=0.01, use_colnames=True)
```

```

# Hitung jumlah transaksi tiap item untuk visualisasi
item_counts = basket.sum().sort_values(ascending=False)

# Mencari association rules berdasarkan confidence
rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.01)

# Menambahkan count ke dalam hasil (jumlah transaksi yang memenuhi aturan)
rules['count'] = rules['support'] * len(basket)

# Menampilkan hasil association rules
pd.set_option('display.max_rows', None) # Tampilkan semua baris
pd.set_option('display.max_columns', None) # Tampilkan semua kolom
pd.set_option('display.expand_frame_repr', False) # Hindari pemotongan tampilan dataframe

print("Hasil Association Rules:")
print(rules)

# Simpan hasil ke CSV
output_path = "/content/data_tahun.csv"
rules.to_csv(output_path, index=False)

print(f"Hasil analisis telah disimpan di {output_path}")

# ===== VISUALISASI ===== #

# 1. Bar Chart untuk Frequent Itemsets

plt.figure(figsize=(12, 6))
top_items = item_counts.head(20) # Ambil 20 item teratas

```

```
sns.barplot(x=top_items.values, y=top_items.index, color='gray')
```

```
plt.xlabel("Item Frequency (Absolute)", fontsize=12)
```

```
plt.ylabel("")
```

```
plt.title("Frequent Itemset (Top 20)", fontsize=14)
```

```
plt.xticks(fontsize=10)
```

```
plt.yticks(fontsize=10)
```

```
plt.show()
```

```
# 2. Network Graph untuk Association Rules
```

```
plt.figure(figsize=(10, 8))
```

```
G = nx.DiGraph()
```

```
# Tambahkan edge berdasarkan aturan asosiasi
```

```
for i, row in rules.iterrows():
```

```
    antecedents = ', '.join(row['antecedents'])
```

```
    consequents = ', '.join(row['consequents'])
```

```
    G.add_edge(antecedents, consequents, weight=row['confidence'])
```

```
# Posisi node
```

```
pos = nx.spring_layout(G, seed=42)
```

```
# Gambar nodes dan edges
```

```
nx.draw(G, pos, with_labels=True, node_size=3000, node_color="lightblue", font_size=10,
font_weight="bold", edge_color="gray")
```

```
labels = {edge: f'{round(G.edges[edge]["weight"], 2)}' for edge in G.edges}
```

```
nx.draw_networkx_edge_labels(G, pos, edge_labels=labels, font_size=8)
```

```
plt.title("Network Graph of Association Rules")
```

```

plt.show()

# Path ke semua file association rules bulanan
file_paths = sorted(glob.glob("/content/Association_Rules_*.csv")) # Sesuaikan lokasi

# Dictionary untuk menyimpan DataFrame tiap bulan
rules_dict = {}

# Dictionary untuk menyimpan rule string tiap bulan
rules_per_month = {}

# Membaca semua file dan memprosesnya
for file in file_paths:
    df = pd.read_csv(file)

    # Pastikan kolom yang digunakan tersedia
    df = df[['antecedents', 'consequents', 'support', 'confidence', 'lift', 'count']]

    # Buat kolom rule
    df['rule'] = "{" + df['antecedents'].astype(str) + "}" + "=> {" + df['consequents'].astype(str) + "}"

    # Ambil nama bulan dari nama file
    bulan = file.split("_")[-1].replace(".csv", "").replace("Association_Rules_", "").strip()

    # Simpan daftar aturan dalam bentuk set
    rules_per_month[bulan] = set(df['rule'])

# Set rule sebagai index dan ubah nama kolom agar unik per bulan
df = df.set_index('rule')

```

```

df = df.rename(columns={
    'support': f'support_{bulan}',
    'confidence': f'confidence_{bulan}',
    'lift': flift_{bulan}',
    'count': f'count_{bulan}'
})

# Simpan dataframe ke dictionary
rules_dict[bulan] = df

# Ambil irisan aturan yang muncul di SEMUA bulan
common_rules = set.intersection(*rules_per_month.values())

# Gabungkan semua dataframe berdasarkan index
all_rules = pd.concat(rules_dict.values(), axis=1, join='outer')

# Filter hanya aturan yang muncul di semua bulan
rules_stable = all_rules.loc[all_rules.index.isin(common_rules)].reset_index()

# Jika tidak ada aturan yang stabil
if rules_stable.empty:
    print("\n Tidak ada aturan asosiasi yang muncul di semua 12 bulan.\n")
else:
    # Hitung CVC, CVL, dan OVCR
    confidence_cols = [col for col in rules_stable.columns if "confidence_" in col]
    lift_cols = [col for col in rules_stable.columns if "lift_" in col]

    rules_stable['CVC'] = rules_stable[confidence_cols].std(axis=1) /
rules_stable[confidence_cols].mean(axis=1)

```

```

rules_stable['CVL'] = rules_stable[lift_cols].std(axis=1) /
rules_stable[lift_cols].mean(axis=1)

rules_stable['OVCR'] = rules_stable[confidence_cols].mean(axis=1)

# Tampilkan hasil
print("\n Aturan asosiasi yang muncul setiap bulan beserta nilai CVC, CVL, dan OVCR:\n")
print(rules_stable[['rule', 'CVC', 'CVL', 'OVCR']].to_string(index=False))

```

2. Data Review

```

import pandas as pd
import string
import re

from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from imblearn.over_sampling import SMOTE
from sklearn.naive_bayes import MultinomialNB

from sklearn.metrics import accuracy_score, classification_report, confusion_matrix,
ConfusionMatrixDisplay

import matplotlib.pyplot as plt
from collections import Counter
import numpy as np
import seaborn as sns
from wordcloud import WordCloud

data=pd.read_csv('/content/scrapping_data_grobbak_korea_nologaten.csv')

# Melakukan casefolding pada kolom 'Keterangan', casefolding dilakukan untuk Mengubah
semua huruf dalam teks menjadi huruf kecil
data['Keterangan'] = data['Keterangan'].str.lower()

```

```

# Menghapus tanda baca
def remove_punctuation(text):
    text = "".join([char for char in text if char not in string.punctuation])
    return text

data['Keterangan'] = data['Keterangan'].apply(remove_punctuation)

# menghapus emoji, atau melakukan proses emoji removal

# Menghapus emoji
def remove_emoji(text):
    emoji_pattern = re.compile("[
        u"\U0001F600-\U0001F64F" # emoticons
        u"\U0001F300-\U0001F5FF" # symbols & pictographs
        u"\U0001F680-\U0001F6FF" # transport & map symbols
        u"\U0001F1E0-\U0001F1FF" # flags (iOS)
        u"\U00002702-\U000027B0"
        u"\U000024C2-\U0001F251"
    ]+", flags=re.UNICODE)
    return emoji_pattern.sub(r'', text)

data['Keterangan'] = data['Keterangan'].apply(remove_emoji)

# melakukan proses tokenisasi

# Tokenisasi
from nltk.tokenize import word_tokenize
import nltk

```

```

nltk.download('punkt')
nltk.download('punkt_tab') # Download the punkt_tab data

def word_tokenize_wrapper(text):
    return word_tokenize(text)

data['Keterangan_tokens'] = data['Keterangan'].apply(word_tokenize_wrapper)
# melakukan stop word removal atau Menghapus Kata Umum yang Tidak Bermakna

# Download required NLTK data
nltk.download('stopwords')
from nltk.corpus import stopwords

# Stop word removal
list_stopwords = stopwords.words('indonesian')

# Add additional stopwords if needed
list_stopwords.extend(["yg", "dg", "rt", "dgn", "ny", "d", 'klo',
                      'kalo', 'amp', 'biar', 'bikin', 'bilang',
                      'gak', 'ga', 'krn', 'nya', 'nih', 'sih',
                      'si', 'tau', 'tdk', 'tuh', 'utk', 'ya',
                      'jd', 'jgn', 'sdh', 'aja', 'n', 't', 'nyg',
                      'hehe', 'pen', 'u', 'nan', 'loh', 'rt', 'wkwk',
                      'wkwk', 'haha', 'gakk', 'gk', 'lg', 'blm', 'tp',
                      'loh', 'doi', 'banget', 'bang', 'iya', 'g'])

# Convert list of stopwords to set for faster lookups
list_stopwords = set(list_stopwords)

```

```

def remove_stopwords(text):
    tokens = text
    text = [word for word in tokens if word not in list_stopwords]
    return text

data['Keterangan_tokens'] = data['Keterangan_tokens'].apply(remove_stopwords)

# Kamus normalisasi slang ke kata baku
kamus_normalisasi = {
    'gini': 'begini',
    'gitu': 'begitu',
    'pesen': 'pesan',
    'makasih': 'terima kasih',
    'makasi': 'terima kasih',
    'makasii': 'terima kasih',
    'mks': 'makasih',
    'bgt': 'banget',
    'bangettt': 'banget',
    'bener': 'benar',
    'bgt': 'banget',
    'ga': 'tidak',
    'gak': 'tidak',
    'nggak': 'tidak',
    'gatau': 'tidak tahu',
    'gue': 'saya',
    'aku': 'saya',
    'lu': 'kamu',
    'loe': 'kamu',
    'dpt': 'dapat',

```

```

'tp': 'tapi',
'udh': 'sudah',
'udah': 'sudah',
'blm': 'belum',
'tdk': 'tidak',
'trs': 'terus',
'klo': 'kalau',
'klau': 'kalau',
'yg': 'yang',
'dr': 'dari',
'lg': 'lagi',
'jg': 'juga',
'aja': 'saja',
'ajaah': 'saja',
'bikin': 'membuat',
'dgn': 'dengan',
'bsk': 'besok',
'kmrn': 'kemarin',
'sblm': 'sebelum',
'bngt': 'banget',
'sy': 'saya'
}

# Fungsi normalisasi berdasarkan kamus
def normalisasi_kata(tokens):
    return [kamus_normalisasi.get(kata, kata) for kata in tokens]

# Terapkan normalisasi ke kolom token
data['Keterangan_tokens'] = data['Keterangan_tokens'].apply(normalisasi_kata)

```

prompt: melakukan stemming atau mengubah kata ke bentuk asal

!pip install Sastrawi

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import re
```

Inisialisasi stemmer Sastrawi

```
factory = StemmerFactory()
stemmer = factory.create_stemmer()
```

Fungsi untuk menghapus akhiran kepemilikan seperti -nya, -ku, -mu

```
def remove_possessive_suffix(word):
    return re.sub(r'(nya|ku|mu)$', '', word)
```

Fungsi lemmatization akhir

```
def lemmatize_tokens(tokens):
    lemmatized = []
    for word in tokens:
        # Step 1: hapus akhiran seperti 'nya'
        word = remove_possessive_suffix(word)
        # Step 2: lakukan stemming/lemmatization
        lemma = stemmer.stem(word)
        lemmatized.append(lemma)
    return lemmatized
```

Terapkan ke DataFrame

```
data['Keterangan_tokens'] = data['Keterangan_tokens'].apply(lemmatize_tokens)
```

```

# 1. Baca dataset

# =====

# data = pd.read_csv("scrapping_data_grobbak_korea_nologaten.csv")

# =====

# 2. Buat Lexicon Dictionary

# =====

lexicon = {
    # Positif
    "enak": 1, "mantap": 1, "lumer": 1, "oke": 1, "pas": 1, "worth": 1,
    "rekomendasi": 1, "recommended": 1, "mantul": 1, "sukses": 1, "bagus": 1,
    "lezat": 1, "nikmat": 1, "murah": 1, "puas": 1, "beneran": 1, "banhet": 1,
    "kriuk": 1, "top": 1, "mantappp": 1, "juara": 1, "maknyus": 1, "good": 1,
    "segar": 1, "fresh": 1, "best": 1, "ramah": 1, "cepat": 1, "bersih": 1,
    "juicy": 1, "gurih": 1, "favorit": 1, "istimewa": 1, "nikmattt": 1,
    "juaraaa": 1, "worthit": 1, "hemat": 1, "kenyang": 1, "murmer": 1,
    "crispy": 1, "recommended banget": 1, "autopilot": 1,
    "praktis": 1, "halal": 1, "wangi": 1, "sehat": 1, "murce": 1,
    "ramah banget": 1, "helpful": 1, "keren": 1, "terbaik": 1, "top markotop": 1,
    "super": 1, "juara banget": 1, "maknyuss": 1, "ajibb": 1, "the best": 1,
    "worth it": 1, "harga pas": 1, "recommended seller": 1, "bagus banget": 1,
    "asli": 1, "ori": 1, "mantappu jiwa": 1, "premium": 1, "nyaman": 1,
    "wajib coba": 1, "senang": 1, "amazing": 1, "luar biasa": 1,
    "banyak bonus": 1, "fast respon": 1, "delivery cepat": 1, "packing rapi": 1,

    # Negatif
    "kurang": -1, "tidak": -1, "ga": -1, "g": -1, "kecewa": -1, "gagal": -1,
    "keras": -1, "asin": -1, "pahit": -1, "hambar": -1, "encer": -1, "bau": -1,

```

```

"mentah": -1, "gosong": -1, "overcooked": -1, "basi": -1, "lama": -1,
"lambat": -1, "jelek": -1, "mahal": -1, "kureng": -1, "kurenggg": -1,
"tidak enak": -1, "mengecewakan": -1, "payah": -1, "jorok": -1, "kotor": -1,
"anyep": -1, "amis": -1, "berantakan": -1, "menyedihkan": -1, "bermasalah": -1,
"bikin sakit": -1, "menjijikkan": -1, "busuk": -1, "tidak sesuai": -1,
"palsu": -1, "zonk": -1, "tidak fresh": -1, "hancur": -1, "kurang matang": -1,
"sakit perut": -1, "bau tengik": -1, "bocor": -1, "busaan": -1,
"cair": -1, "penyok": -1, "luntur": -1, "tidak ori": -1, "rusak": -1,
"expired": -1, "palsuuu": -1, "asin banget": -1, "kemanisan": -1,
"kebanyakan msg": -1, "ngaret": -1, "lelet": -1, "ribet": -1,
"sok": -1, "jutek": -1, "tidak ramah": -1, "bohong": -1, "tipu": -1,
"ulasan palsu": -1, "tidak sesuai foto": -1, "mengecewa": -1, "garing": -1,
"benyek": -1, "nggak worth": -1, "overprice": -1, "zonkk": -1, "gubris": -1, "kurangg": -1,
"ganggu": -1, "keras": -1, "gamau": -1
}

```

```
# =====
```

```
# 3. Fungsi pelabelan dengan lexicon
```

```
# =====
```

```
def lexicon_sentiment(tokens):
```

```
    if not tokens:
```

```
        return "netral"
```

```
    score = 0
```

```
    for word in tokens:
```

```
        score += lexicon.get(word, 0)
```

```
    if score > 0:
```

```
        return "positif"
```

```
    elif score < 0:
```

```
        return "negatif"
```

```
    else:
```

```

        return "netral"

# =====

# 4. Tambahkan label sentimen ke dataset

# =====

data["Sentimen"] = data["Keterangan_tokens"].apply(lexicon_sentiment)

# =====

# 5. TF-IDF + Split Data

# =====

X = data["Keterangan_tokens"].fillna("").astype(str)
y = data["Sentimen"]

vectorizer = TfidfVectorizer()
X_tfidf = vectorizer.fit_transform(X)

X_train, X_test, y_train, y_test = train_test_split(
    X_tfidf, y, test_size=0.1, random_state=42
)

# =====

# 6. Oversampling dengan SMOTE

# =====

min_samples = min(Counter(y_train).values())
k_neighbors = max(1, min_samples - 1)

smote = SMOTE(random_state=42, k_neighbors=k_neighbors)
X_train_resampled, y_train_resampled = smote.fit_resample(X_train, y_train)

```

```

# =====
# 7. Train Model Naive Bayes
# =====

nb_model = MultinomialNB()
nb_model.fit(X_train_resampled, y_train_resampled)

# =====

# 8. Prediksi & Evaluasi
# =====

y_pred = nb_model.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
classification_rep = classification_report(
    y_test, y_pred, target_names=["negatif", "netral", "positif"]
)

print(f'Akurasi Model: {accuracy*100:.2f} %')
print("\nLaporan Klasifikasi:\n", classification_rep)

# =====

# 9. Confusion Matrix (Custom Style seperti gambar)
# =====

cm = confusion_matrix(y_test, y_pred, labels=["positif", "negatif"])

cm_values = np.array([[cm[0,0], cm[1,0]], # TP, FP
                      [cm[0,1], cm[1,1]]]) # FN, TN

labels_matrix = np.array([["True Positive", "False Positive"],
                          ["False Negative", "True Negative"]])

```

```

plt.figure(figsize=(6,6))

sns.heatmap(cm_values, annot=labels_matrix.astype(str) + "\n" + cm_values.astype(str),
            fmt="", cmap="Blues", cbar=False, square=True)

plt.xlabel("Actual Values")
plt.ylabel("Predicted Values")
plt.title("Confusion Matrix")
plt.xticks([0.5,1.5], ["Positive (1)", "Negative (0)"])
plt.yticks([0.5,1.5], ["Positive (1)", "Negative (0)"], rotation=0)
plt.show()

# =====
# 10. Visualisasi Pie Chart
# =====

sentiment_counts = data["Sentimen"].value_counts()
labels = sentiment_counts.index.tolist()
sizes = sentiment_counts.values.tolist()

plt.figure(figsize=(6, 6))
plt.pie(
    sizes,
    labels=labels,
    autopct=lambda p: '{:.0f}'.format(p * sum(sizes) / 100),
    startangle=90
)
plt.title("Distribusi Sentimen (Lexicon-based Labeling)")
plt.axis('equal')
plt.show()

```

```

#wordcloud
def show_results():

    # Gabungkan semua token dalam kolom 'Keterangan_tokens' menjadi satu string
    all_words = ''.join([' '.join(tokens) for tokens in data['Keterangan_tokens']])

    # Buat WordCloud
    wordcloud = WordCloud(width=800, height=400,
background_color='white').generate(all_words)

    # Tampilkan WordCloud
    plt.figure(figsize=(10, 5))
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis("off")
    plt.title("Wordcloud dari Keterangan")
    plt.show()

show_results()

#wordcloud negatif

from collections import Counter

def show_results():

    #wordcloud dari sentimen yang negatif
    negative_data = data[data["Sentimen"] == "negatif"]
    all_negative_words = ' '.join([' '.join(tokens) for tokens in
negative_data['Keterangan_tokens']])

```

```

wordcloud = WordCloud(width=800, height=400,
background_color='white').generate(all_negative_words)

plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.title("Wordcloud dari Sentimen Negatif")
plt.show()

# Count word frequencies in negative sentiment
words = all_negative_words.split()
word_counts = Counter(words)
# Get the 10 most common words
most_common_words = word_counts.most_common(10)
print("10 Kata yang paling sering muncul pada Wordcloud Negatif:")
for word, count in most_common_words:
    print(f"{word}: {count}")
show_result

```

