ANALISIS SENTIMEN ULASAN PELANGGAN BERBASIS *NAÏVE BAYES* DAN *ASSOCIATION RULE* DALAM PERBAIKAN KUALITAS LAYANAN UMKM GROBBAK KOREA

TUGAS AKHIR

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

Nama : Muhammad Faisal Siraj Ramadhan
No. Mahasiswa : 21522295

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

2025

PERNYATAAN KEASLIAN

Saya mengakui bahwa tugas akhir yang berjudul "Analisis Sentimen Ulasan Pelanggan Berbasis Naïve Bayes Dan Association Rule Dalam Perbaikan Kualitas Layanan Umkm Grobbak Korea" adalah hasil karya saya sendiri kecuali kutipan dan ringkasan yang seluruhnya sudah saya jelaskan sumbernya. Jika dikemudian hari ternyata terbukti pengakuan saya ini tidak benar dan melanggar peraturan yang sah maka saya bersedia ijazah yang telah saya terima ditarik kembali oleh Universitas Islam Indonesia.

Yogyakarta, 10 – 10 - 2025



(Muhammad Faisal Siraj Ramadhan)
21522295

SURAT BUKTI PENELITIAN

UNIVERSITAS ISLAM INDONESIA

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. ft@uii.ac.id
W. ft.uii.ac.id

SURAT KETERANGAN PENELITIAN

Nomor: 16/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 Faisal Siraj Ramadhan
No. Mhs	:	21522295
Dosen Pembimbing	:	Annisa Uswatun Khasanah, S.T., M.Sc

Telah selesai melaksanakan penelitian yang berjudul "Analisis Sentimen Ulasan Pelanggan Berbasis Naïvebayes Dan Association Rule Dalam Perbaikan Kualitas Layanan Umkm Grobbak Korea" di Laboratorium Data Mining, Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia tercatat mulai tanggal 10 Mei sampai dengan tanggal 20 Juni 2025.

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

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Yogyakarta 12 September 2025

Kepala Laboratorium
Data Mining



Dr. Dwi Adi Purnama, S.T.

LEMBAR PENGESAHAN PEMBIMBING

ANALISIS SENTIMEN ULASAN PELANGGAN BERBASIS *NAÏVE BAYES* DAN *ASSOCIATION RULE* DALAM PERBAIKAN KUALITAS LAYANAN UMKM GROBBAK KOREA

TUGAS AKHIR

Disusun Oleh :

Nama : Muhammad Faisal Siraj Ramadhan
No. Mahasiswa : 21522295

Yogyakarta, 09 September 2025

Dosen Pembimbing



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

LEMBAR PENGESAHAN DOSEN PENGUJI

ANALISIS SENTIMEN ULASAN PELANGGAN BERBASIS *NAÏVE BAYES* DAN *ASSOCIATION RULE* DALAM PERBAIKAN KUALITAS LAYANAN UMKM GROBBAK KOREA

TUGAS AKHIR

Disusun Oleh :
Nama : Muhammad Faisal Siraj Ramadhan
No. Mahasiswa : 21522295

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

Yogyakarta, *10 - 10 – 2025*

Tim Penguji

Annisa Uswatun Khasanah, S.T., M.Sc.	
Ketua	
Wahyudhi Sutrisno, S.T., M.M.	101025
Anggota I	
Elanjati Worldailmi, 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

HALAMAN PERSEMBAHAN

Bismillahirrahmanirrahim

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

Bapak Akhmad Azmi dan Ibu Eny Hasanah, kedua orang tua saya yang menjadi *support system*, mengasuh dan mendidik saya sedari kecil. Lalu kepada adik, dan teman-teman penulis yang menjadi pendukung atas tersusunnya laporan tugas akhir ini.

MOTTO

“Tidak ada suatu jiwapun (diri) melainkan ada penjaganya.”
(Q.S At-Tariq : 4)

"Dan apabila hamba-hamba-Ku bertanya kepadamu tentang Aku, maka (jawablah), sesungguhnya Aku dekat. Aku mengabulkan doa orang yang berdoa apabila ia berdoa kepada-Ku. Maka hendaklah mereka itu memenuhi (perintah)-Ku dan beriman kepada-Ku, agar mereka selalu berada dalam kebenaran."
(Q.S Al-Baqarah : 186)

“The harder you practice, the luckier you get.”
(Gary Player)
Apa yang sering kita sebut sebagai “keberuntungan” sebenarnya adalah buah dari persiapan yang konsisten dan usaha yang dilakukan dengan sungguh-sungguh. Sehingga apa itu keberuntungan? Ia lahir dari kerja keras, konsistensi, dan keberanian untuk terus mencoba setiap hari. Berlatihlah dengan cara yang benar, maka hidup akan menghadiahimu kesempatan yang orang lain sebut “keberuntungan”.

KATA PENGANTAR

Alhamdulillahi Rabbil`alamin puji dan syukur senantiasa penulis ucapkan kepada Allah S.W.T karena atas limpahan rahmat dan hidayah-Nya penulis dapat menyelesaikan Laporan tugas akhir dengan baik dan sesuai waktu yang telah ditentukan. Shalawat serta salam senantiasa penulis sampaikan kepada junjungan kita yakni Nabi Muhammad S.A.W yang telah membawa risalah kepada umat manusia dari zaman jahiliyyah menuju zaman modern seperti saat ini.

Penulis pun menyadari bahwa laporan ini tidak akan tersusun dengan baik tanpa adanya bantuan dan bimbingan dari pihak-pihak terkait. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada semua pihak yang telah membantu dalam keberlangsungan tugas akhir penulis. Ucapan terima kasih penulis sampaikan kepada:

1. Bapak Prof. Dr. Ir. Hari Purnomo, M.T., IPU., ASEAN. Eng., selaku Dekan dari 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 tugas akhir, yang telah mendukung dan banyak memberikan arahan dan masukan kepada penulis dalam melaksanakan kegiatan tugas akhir dan juga dalam menyelesaikan laporan tugas akhir.
5. Kedua orang tua tercinta, yang telah membantu mendoakan dan memberikan dukungan dalam pelaksanaan tugas akhir maupun penyusunan laporan tugas akhir.
6. Teman-teman dan sahabat yang penulis cintai, yang telah memberi energi semangat agar penulis dapat menyelesaikan tugas akhir dengan baik.
7. Keluarga, saudara, teman-teman dan semua pihak yang tidak bisa penulis sebutkan satu-persatu yang telah memberikan dukungan moral.

Penulis menyadari bahwa masih terdapat kekurangan dalam penyusunan laporan ini karena terbatasnya pengetahuan dan kemampuan penulis. Oleh karena itu, apabila terdapat kekeliruan informasi dan penulisan pada laporan tugas akhir ini, penulis mengucapkan permohonan maaf. Kritik dan saran yang membangun untuk perbaikan laporan ini penulis harapkan demi peningkatan kualitas laporan ini. Besar harapan penulis semoga laporan tugas akhir ini dapat memberikan beragam informasi dan manfaat dalam menambah ilmu pengetahuan mengenai penggunaan *machine learning*, terkhusus dalam lingkup *sentiment analysis*. Akhirul kalam, Wassalamu'alaikum Warahmatullahi Wabarakatuh.

ABSTRAK

Perkembangan UMKM kuliner di Indonesia semakin kompetitif, terutama dalam bidang kuliner khas Korea yang mulai merebak seiring tren budaya Korea yang masuk ke Indonesia. Begitupun dengan Grobbak Korea yang merupakan *franchise* gerobak makanan khas Korea pertama, juga mengalami dampak dari persaingan yang ditunjukkan tren penurunan *revenue* yang didapat dalam kurun waktu 6 bulan yakni November 2024-April 2025. Ditinjau dari permasalahan tersebut, dalam ulasan pelanggan Grobbak Korea terdapat sekitar 30% *rating* ≤ 3, sehingga perlu ditinjau terkait keluhan pelanggan melalui analisis sentimen ulasan, yang kemungkinan didalamnya dapat ditemukan permasalahan terkait penurunan *revenue* yang dialami. Penelitian ini bertujuan untuk memahami sebaran umum sentimen pelanggan, mengetahui hasil uji model klasifikasi dan aturan asosiasi dari ulasan, serta mengidentifikasi aspek permasalahan untuk menawarkan solusi perbaikan kualitas layanan pada UMKM Grobbak Korea. Penelitian ini mengusulkan pendekatan analisis sentimen berbasis Naïve Bayes yang dipadukan dengan *association rule* untuk mengevaluasi kualitas layanan UMKM Grobbak Korea. Data berupa ulasan pelanggan diperoleh dari platform gojek, yang kemudian diproses melalui tahapan *pre-processing*, dan dilanjutkan dengan tahapan klasifikasi dan *association rule* secara paralel. Hasil penelitian menunjukkan bahwa metode Naïve Bayes, dengan *testing size* sebesar 20% dapat mengklasifikasikan ulasan pelanggan dengan tingkat akurasi sebesar 84.8%, *precision* sebesar 84.9%, dan *recall* sebesar 84.8%, *precision* sebesar 84.9%, dan *recall* sebesar 84.8%. Lalu pada *association rule* yang dicari pada ulasan berlabel negatif, ditemukan 180 aturan asosiatif yang selanjutnya diseleksi berdasarkan 10 kata yang paling sering muncul, menghasilkan *associative tree* yang digunakan untuuk mendapatkan informasi terkait kemungkinan permasalahan pada Grobbak Korea. Dari permasalahan yang ditemukan, digunakan penyelesaian akhir dengan pembuatan *standard operational procedure* (SOP) dengan ruang lingkup prosedur pembuatan produk, manajemen & pengadaan bahan baku, serta kebersihan dan sanitasi. Diharapkan dengan dibuatnya dokumen SOP ini dapat membantu Grobbak Korea dalam menetapkan standar pelayanan kearah yang lebih baik, terutama dalam pembuatan produk dari Grobbak Korea.

Kata kunci: *Analisis Sentimen, Naïve Bayes, Association rule, Text Mining*

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 5
1.3 Tujuan Penelitian 5
1.4 Manfaat Penelitian 6
1.5 Batasan Penelitian 6
BAB II TINJAUAN PUSTAKA 7
2.1 Kajian Literatur 7
2.2 Landasan Teori 13
2.2.1 Data Mining 13
2.2.2 Knowledge Discovery from Data (KDD) 13
2.2.3 Machine Learning 14
2.2.4 Text Mining 15
2.2.5 Analisis Sentimen 16
2.2.6 Text Pre-processing 16
2.2.7 TF-IDF 17
2.2.8 Klasifikasi 17
2.2.9 Naïve Bayes classifier 18
2.2.10 Support Vector Machine 19
2.2.11 Association rule 19
BAB III METODE PENELITIAN 21
3.1 Subjek Penelitian 21
3.2 Objek Penelitian 21
3.3 Jenis Data & Sumber Data 21
3.4 Pengumpulan Data 21
3.5 Alur Penelitian 22
BAB IV PENGUMPULAN DAN PENGOLAHAN DATA 25
4.1 Pengumpulan Data 25
4.2 Pengolahan Data 27
4.2.1 Text Pre-processing 27
4.2.2 Pelabelan & pembobotan data 31
4.2.3 Dataset splitting 32
4.2.4 Klasifikasi 34
4.2.5 Association rule 36

BAB V PEMBAHASAN 41
5.1 Gambaran umum pengolahan data 41
5.2 Hasil klasifikasi menggunakan *Naive bayes classifier* 41
5.3 Visualisasi kata dan *association rule* 42
5.4 Analisis strategi berdasarkan asosiasi kata dan observasi 44
BAB VI PENUTUP 47
6.1 Kesimpulan 47
6.2 Saran 48
DAFTAR PUSTAKA 49
LAMPIRAN 53
A-*Dataset* 53
B-*Script python preprocessing data* 53
C-*Labelling data* 55
D-*Script python* klasifikasi 56
E-*Script python association rule* 60
F-Dokumen *Standard Operation Procedure* 63

DAFTAR TABEL

Tabel 2. 1 Kajian Literatur 12
Tabel 4. 1 Contoh hasil *scraping* data 26
Tabel 4. 2 Tahapan *case folding* 27
Tabel 4. 3 Tahapan normalisasi 28
Tabel 4. 4 Tahapan *slang words replacement* 29
Tabel 4. 5 Tahapan *tokenize* 29
Tabel 4. 6 Tahapan *stopwords removal* 30
Tabel 4. 7 Tahapan *stemming* 30
Tabel 4. 8 Contoh hasil pelabelan dengan kamus *lexicon* 31
Tabel 4. 9 Variasi *dataset split* 1 33
Tabel 4. 10 Variasi *dataset split* 2 33
Tabel 4. 11 Variasi *dataset split* 3 33
Tabel 4. 12 Metriks hasil uji skenario 1 34
Tabel 4. 13 *Confusion matrix* skenario 1 34
Tabel 4. 14 Metriks hasil uji skenario 2 35
Tabel 4. 15 *Confusion matrix* skenario 2 35
Tabel 4. 16 Metriks hasil uji skenario 3 35
Tabel 4. 17 *Confusion matrix* skenario 3 36
Tabel 4. 18 Parameter pencarian *association rule* 38
Tabel 5. 1 Metriks hasil uji klasifikasi 42
Tabel 5. 2 Kode permasalahan 44
Tabel 5. 3 Analisis pemetaan permasalahan & solusi 45

DAFTAR GAMBAR

Gambar 1. 1 *Franchise* Grobbak Korea 2
Gambar 1. 2 Distribusi ulasan pelanggan Grobbak Korea 3
Gambar 1. 3 Revenue Grobbak Korea Pandeyan 3
Gambar 2. 1 Langkah proses *knowledge discovery* Sumber: (Han et al., 2024) 14
Gambar 2. 2 Kategori pembelajaran *machine learning* 15
Gambar 2. 3 Proses *Text mining* Sumber: (Gaikwad, 2014) 16
Gambar 3. 1 Diagram alur penelitian 22
Gambar 4. 1 Laman *Gofood* Grobbak Korea Pandeyan Sumber: https://*Gofood*.co.id/ .25
Gambar 4. 2 Proses *login* pada *website* Gojek 26
Gambar 4. 3 Distribusi sentimen ulasan pelanggan 31
Gambar 4. 4 Distribusi sentimen pasca perhitungan *polarity* 32
Gambar 4. 5 *Wordcloud* ulasan negatif pelanggan 37
Gambar 4. 6 Visualisasi data ulasan negatif pelanggan 37
Gambar 4. 7 Asosiasi kata sentimen negatif 38
Gambar 5. 1 Keterangan warna pada tabel rekomendasi 44

BAB I

PENDAHULUAN

1.1 Latar Belakang

Adanya perkembangan industri saat ini tidak lepas dari perkembangan teknologi informasi. Perkembangan teknologi informasi ini mengantarkan dunia kepada zaman serba digital, dimana digitalisasi teknologi mendatangkan kehidupan yang serba instan, serba cepat, serta pertukaran informasi dalam hitungan detik (Megawati & Lawi, 2021). Teknologi informasi sendiri memungkinkan pengolahan data yang cepat dan efisien, serta memungkinkan komunikasi dan serta pertukaran informasi dimana saja dan kapan saja (Saputra et al., 2023). Salah satu contoh bidang industri yang berkembang dengan pesat adalah industri kuliner.

Dalam beberapa tahun terakhir, industri *F&B* mengalami transformasi digital yang pesat. Munculnya platform *online* seperti *Gofood, GrabFood, dan ShopeeFood* telah mengubah cara konsumen membeli dan menikmati makanan. Sehingga dengan kemudahan yang diciptakan, *online ordering* membuka pasar yang lebih luas dan kompetitif. Persaingan tidak hanya terjadi antar bisnis *F&B* yang menawarkan produk serupa, tetapi juga dengan berbagai alternatif lain seperti makanan rumahan, *street food*, dan makanan berat. Hal ini mengharuskan bisnis *F&B* untuk terus berinovasi guna menarik minat konsumen, serta memastikan profitabilitas dan kelayakan produk yang diluncurkan ke pasar seperti menurut Torres et al. (2024), sehingga mampu memahami dan merespons perubahan di masa yang akan datang.

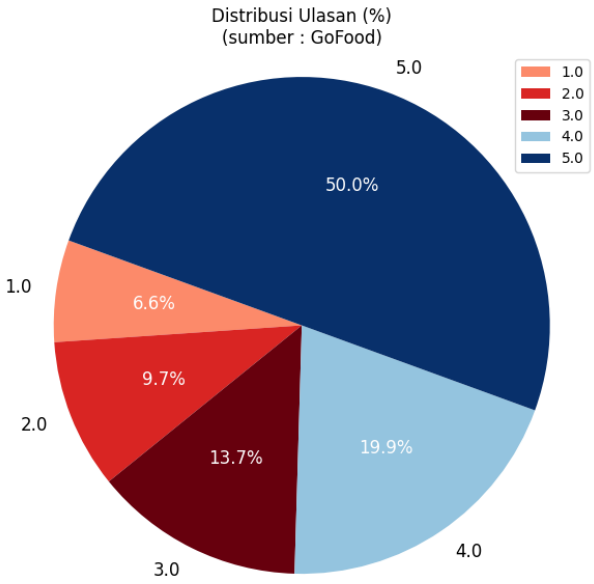
Dalam produk *street food* khususnya, bisnis kuliner khas Korea yang merebak di Indonesia merupakan respons terhadap akulturasi budaya Korea yang masuk lewat paparan industri hiburan Korea Selatan. Sebut saja seperti “drakor”, lalu budaya “*k-pop*”, yang secara tidak langsung memengaruhi permintaan konsumen terhadap *street food* khas Korea. Hal tersebut mencetuskan berdirinya usaha kuliner *street food* khas Korea Bernama “Grobbak Korea”, yang merupakan pelopor *franchise* kuliner *street food* khas Korea di Indonesia.



Gambar 1. 1 *Franchise* Grobbak Korea

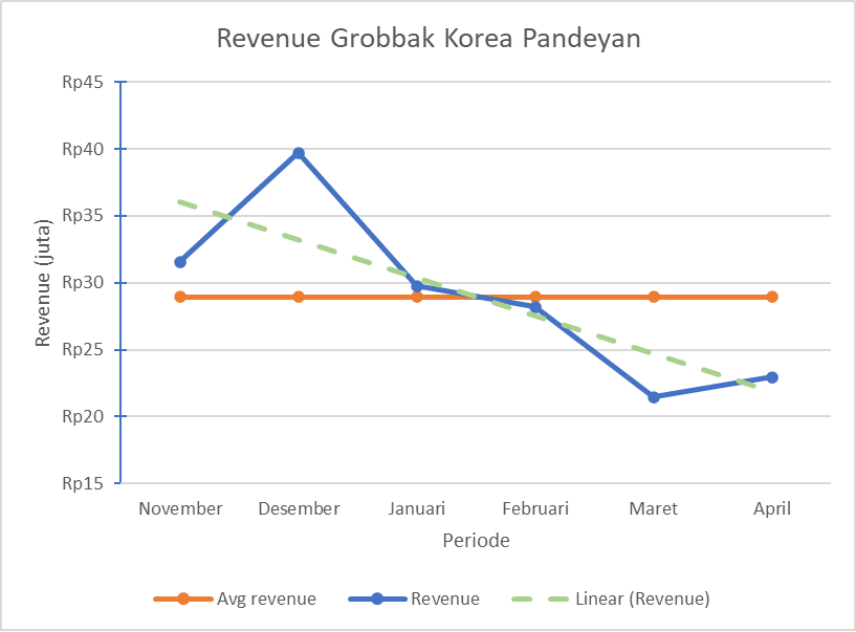
Sebagai pelopor *franchise street food* khas Korea, Grobbak Korea terus berupaya dalam melakukan *improvement* terhadap standar pelayanan yang ditawarkan kepada pelanggan. Hal ini merupakan respons dari Grobbak Korea terhadap perubahan pasar kuliner yang terus mengalami pergeseran tren yang cenderung abstrak.

Dalam menghadapi perubahan pergeseran tren tersebut, upaya *improvement* harus dilakukan selaras dengan permintaan pelanggan, agar meningkatkan kepuasan pelanggan (Oluwajinmi & Longe, 2024). Contohnya, pada *outlet* Grobbak Korea cabang Pandeyan, terdapat cukup banyak ulasan negatif dari pelanggan, sehingga perlu dilakukannya perbaikan pelayananan.



Gambar 1. 2 Distribusi ulasan pelanggan Grobbak Korea

Berdasarkan distribusi ulasan diatas, dimana 30% ulasan yang diungkapkan oleh pelanggan memiliki *rating* ≤ 3 yang tentunya terjadi bukan tanpa alasan, sehingga perlu dilakukannya perbaikan berdasarkan ulasan-ulasan tersebut.



Gambar 1. 3 Revenue Grobbak Korea Pandeyan

Lalu permasalahan terjadi dari *revenue* Grobbak Korea Pandeyan yang cenderung memiliki tren menurun, dan oleh karena itu harus dilakukan perbaikan yang terukur. Tanpa upaya *improvement* yang terukur dan mengindahkan permintaan pelanggan, bisnis *F&B* berisiko

kehilangan daya saing, dan bahkan gagal bertahan dalam jangka panjang. Seperti halnya Grobbak Korea yang merupakan pelopor bisnis *franchise* jajanan korea di Indonesia. Sebagai pelopor, Grobbak Korea tentunya harus memiliki rencana perbaikan kualitas pelayanan, agar dapat mempertahankan posisi bisnis mereka. Salah satu Upaya yang dapat dilakukan dalam Menyusun rencana perbaikan kualitas pelayanan dapat dilakukan berdasarkan ulasan pelanggan Grobbak Korea. Ulasan pelanggan ini dapat diolah dengan cara diekstrak sehingga didapatkan *knowledge* dari ulasan-ulasan tersebut. *Knowledge* disini berarti pola-pola dari kumpulan data besar yang valid, baru, dan berpotensi berguna, dan pada akhirnya dapat dipahami (Cios & Kurgan, 2005). *Knowledge* dapat diekstrak dari sebuah ulasan dengan bantuan sebuah konsep penambangan data, atau yang lebih dikenal dengan istilah *data mining.*

Data mining merupakan proses ekstraksi pengetahuan yang berguna dari kumpulan data yang besar dan kompleks. *Data Mining* bertujuan untuk mengidentifikasi pola-pola atau hubungan yang tersembunyi dalam data dan mengubah hal tersebut menjadi sebuah informasi (Tarigan et al., 2023). Dengan memanfaatkan berbagai teknik *data mining*, bisnis *F&B* dapat mengidentifikasi pola-pola tersembunyi, tren pasar, dan *feedback* pelanggan yang dapat digunakan untuk pengambilan keputusan yang lebih tepat dan efektif, dan tentunya berdasarkan data. Dengan menganalisis data yang berasal dari sistem POS, data penjualan *online*, ulasan pelanggan, dan media sosial, Grobbak Korea dapat memperoleh pemahaman komprehensif mengenai perilaku konsumen, kinerja operasional usaha, serta efektivitas strategi pemasaran. Penerapan *data mining* tidak hanya terbatas pada identifikasi masalah, tetapi juga pada pengembangan solusi inovatif untuk meningkatkan profitabilitas dan kepuasan pelanggan.

Dalam penelitian tugas akhir ini, akan menggunakan beberapa cabang keilmuan *data mining*, seperti *text mining* yang berkaitan dengan pengolahan analisis sentimen. Sentimen sendiri merupakan opini, sikap, dan perasaan seseorang tentang produk, layanan, organisasi, kepribadian, topik atau masalah, dan peristiwa. Sehingga dari sebuah sentimen yang terkandung dalam ulasan pelanggan tersebut, dapat diolah untuk mendapatkan *insight* dari ulasan pelanggan. Dimana dalam industri *F&B* ulasan pelanggan sangatlah penting, analisis sentimen dapat membantu dalam perbaikan menu, suasana, serta kualitas restoran secara keseluruhan(Rokhva et al., 2025). Pada penelitian ini akan digunakan metode klasifikasi *Naïve Bayes Classifier* dalam proses analisis sentimen, guna memprediksi apakah sentimen dari sebuah ulasan termasuk kedalam sentimen negatif ataupun positif. Selanjutnya, digunakan *association rule* untuk mengetahui pola asosiasi kata dari ulasan pelanggan Grobbak Korea,

agar nantinya didapatkan informasi asosiasi kata pada ulasan pelanggan Grobbak Korea. Untuk *association rule*, peneliti berencana menggunakan algoritma *Apriori* dalam proses penelitian ini.

Dari penelitian tugas akhir ini, diharapkan Grobbak Korea dapat memahami dan memeroleh masukan mengenai aspek-aspek produk dan layanan mereka, baik aspek positif maupun negatif. Lalu, diharapkan Grobbak Korea dapat menerapkan perbaikan baik layanan maupun produk, berdasarkan pola asosiasi kata yang sering muncul dalam ulasan pelanggan. Lebih jauh lagi, penelitian ini diharapkan dapat memberikan panduan praktis tentang bagaimana memanfaatkan data ulasan pelanggan untuk meningkatkan kepuasan pelanggan, dan meningkatkan profitabilitas bisnis secara keseluruhan. Dengan pemahaman yang lebih mendalam tentang *feedback* pelanggan, Grobbak Korea dapat beradaptasi dengan lebih efektif terhadap perubahan tren dan preferensi pasar, serta mempertahankan daya saing dalam dunia bisnis kuliner, khususnya kuliner Korea yang semakin kompetitif.

1.2 Rumusan Masalah

Berdasarkan latar belakang dari penelitian, berikut merupakan rumusan masalah pada penelitian ini:

1. Bagaimana sebaran umum mengenai sentimen pelanggan terhadap Grobbak Korea?
2. Bagaimana hasil uji model klasifikasi dari ulasan pelanggan dari Grobbak Korea?
3. Bagaimana hasil dari *associaton rule* berdasarkan ulasan pelanggan dari Grobbak Korea?
4. Aspek apa saja yang menjadi permasalahan, serta solusi apa yang dapat ditawarkan berdasarkan ulasan pelanggan terhadap upaya peningkatan kualitas pelayanan dari Grobbak Korea?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disusun, berikut ini merupakan tujuan dari penelitian ini:

1. Mengetahui sebaran umum mengenai sentimen ulasan pelanggan dari Grobbak Korea.
2. Mengetahui metriks hasil uji model klasifikasi ulasan dari pelanggan Grobbak Korea.
3. Mengetahui hasil *association rule* dari ulasan pelanggan Grobbak Korea.
4. Mengetahui aspek permasalahan, serta solusi yang ditawarkan dalam perbaikan berdasarkan ulasan pelanggan Grobbak Korea.

1.4 Manfaat Penelitian

Manfaat yang didapatkan berbagai pihak yang bersangkutan dari terlaksananya penelitian ini:

1. Bagi Grobbak Korea:
    a. Hasil analisis sentimen dan *association rule* dapat dijadikan bahan perbaikan kualitas dan pelayanan dari Grobbak Korea.
    b. Mengetahui Gambaran & sebaran sentimen pelanggan Grobbak Korea secara umum.
2. Bagi Penulis:

    Penulis dapat mengembangkan pemahaman dan menerapkan ilmu teknik industri yang diperoleh selama perkuliahan, terutama dalam disiplin ilmu *data mining*, dan khususnya terkait dengan analisis sentimen, klasifikasi, serta *association rule*.
3. Bagi pembaca:

    Penelitian ini dapat memberikan referensi terkait penggunaan analisis sentimen, klasifikasi dan *association rule*, dalam memahami sentimen ulasan, sehingga diharapkan penelitian ini dapat menjadi pembanding untuk penelitian-penelitian selanjutnya.

1.5 Batasan Penelitian

Berikut merupakan batasan dari penelitian yang akan dilakukan:

1. Data ulasan yang akan menjadi *input* dari penelitian ini diambil dari laman *Gofood* Grobbak Korea Pandeyan
2. Data ulasan yang diambil adalah ulasan yang telah diunggah maksimum 20 Mei 2025.
3. *Outlet* Grobbak Korea yang diteliti adalah *outlet* yang beralamatkan di Jalan Perintis Kemerdekaan No.35D, Pandeyan, Kec. Umbulharjo, Kota Yogyakarta, Daerah Istimewa Yogyakarta 55182.

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Literatur

Penelitian yang dilakukan oleh Singgalen (2022), bertujuan untuk menganalisis sentimen konsumen terhadap produk dan layanan dari empat restoran ternama di Labuan Bajo, yaitu *Happy Banana* Komodo, *MadeInItaly, Mediterraneo,* dan *La Cucina*. Ulasan pelanggan diambil dari platform TripAdvisor, dengan bantuan aplikasi *WebHarvy*. Selanjutnya data dilakukan *pre-processing* dan dilakukan *SMOTE* untuk mengatasi ketidakseimbangan kelas data, dan dilanjutkan pengolahan dengan metode klasifikasi yaitu *k-Nearest Neighbor (k-NN), Naïve Bayes (NB), Decision Tree (DT), dan Support Vector Machine (SVM).* Hasil penelitian menunjukkan bahwa Ulasan yang dianalisis mayoritas mengandung sentimen positif terhadap produk dan layanan. Lalu hasil uji algoritma k-NN dengan integrasi *SMOTE* memberikan performa terbaik dengan akurasi sebesar 99,27%, dibandingkan dengan *NB, DT, dan SVM* yang memiliki akurasi sekitar 93%. Penggunaan k-NN yang dikombinasikan dengan SMOTE terbukti efektif dalam klasifikasi sentimen, sehingga dapat digunakan dalam penyusunan strategi peningkatan kualitas.

Analisis sentimen pelanggan terhadap 6 restoran di Wakatobi dilakukan oleh Ode et al. (2024). Pertama, data ulasan dikumpulkan melalui kuesioner dengan pelanggan sebagai responden, dimana masing-masing restoran memperoleh 100 ulasan terkait 3 aspek penelitian, yaitu makanan, pelayanan, dan suasana. Data dianalisis secara otomatis dengan mengklasifikasikan ulasan menjadi sentimen positif, negatif, dan netral, analisis dilakukan menggunakan algoritma *Naïve Bayes (NB)*. Hasil analisis menunjukkan bahwa akurasi tertinggi diperoleh pada aspek pelayanan sebesar 88%, diikuti oleh suasana sebesar 86%, dan makanan sebesar 83%. Sentimen positif paling kuat muncul pada aspek pelayanan dengan *precision* sebesar 0.90, serta *F1-score* 0.93, sedangkan *precision* dan *recall* terendah terletak pada sentimen netral aspek makanan. Implementasi sistem ini mampu menjadi sistem rekomendasi restoran berdasarkan akumulasi sentimen positif, sehingga memudahkan wisatawan dalam memilih tempat makan yang sesuai dengan preferensi mereka.

Selanjutnya terdapat penelitian yang dilakukan oleh Ikhsan et al. (2025) yang bertujuan untuk menganalisis opini publik pada *X / Twitter* terhadap kenaikan Pajak Pertambahan Nilai (PPN) menggunakan dua algoritma klasifikasi: *Naïve Bayes* dan *Support Vector Machine (SVM).* Data yang didapat yaitu sebanyak 2.852 data, dilakukan *pre-processing* dan dilanjutkan dengan *TF-IDF* untuk vektorisasi kata. Kelas data yang awalnya tidak seimbang, diseimbangkan menggunakan teknik *SMOTE*. Kedua model kemudian dievaluasi dengan metrik akurasi, presisi, *recall,* dan *f1-score*. Hasilnya menunjukkan bahwa sebelum *balancing*, *SVM* menghasilkan akurasi 98% dan *Naïve Bayes* 97%, sementara setelah *balancing SVM* tetap stabil di 97%, namun *Naïve Bayes* menurun ke 90%. Hasil penelitian mengindikasikan bahwa algoritma *SVM* lebih *robust* terhadap ketidakseimbangan data dibandingkan *Naïve Bayes*, sehingga lebih direkomendasikan untuk analisis sentimen dalam konteks serupa.

Pada penelitian yang dilakukan oleh Amalia et al. (2021), dilakukan analisis sentimen terhadap ulasan pelanggan restoran Solaria. Dalam penelitian ini digunakan 2 metode klasifikasi yaitu *Support Vector Machine* (*SVM*) dan *K- Nearest Neighbor* (*K-NN*), dalam *workframe CRISP-DM.* Setelah data dilakukan *pre-processing*, selanjutnya dilakukan uji klasifikasi. Hasil uji klasifikasi menunjukkan bahwa *SVM* menghasilkan akurasi tertinggi sebesar *81,92%* dan *AUC 0,918*, sedangkan *k-NN* hanya mencapai akurasi 59,03% dan *AUC* 0,590. Dengan demikian, *SVM* terbukti lebih unggul dalam mengklasifikasikan sentimen positif dan negatif dalam ulasan pelanggan restoran di media sosial, khususnya pada penelitian ini.

Sedangkan pada penelitian yang dilakukan oleh Suryana et al. (2024) menerapkan algoritma *Naïve Bayes* dalam analisis sentimen terhadap 1000 ulasan pengguna aplikasi Grab yang dikumpulkan dari *Google Play Store* melalui *web scraping*. Selanjutnya data diproses melalui tahapan *text preprocessing* termasuk *case folding*, *tokenizing*, *filtering*, *stemming*, dan pembobotan *TF-IDF*. Hasil uji klasifikasi yang ditampilkan dalam *confusion matrix* menunjukkan akurasi sebesar 89%, dengan *precision* sebesar 88% dan *recall* sebesar 97%. Evaluasi ini menunjukkan bahwa *Naïve Bayes* merupakan pendekatan andal untuk menganalisis kepuasan pengguna terhadap layanan digital.

Penelitian dengan pendekatan kuantitatif berbasis *text mining, sentiment analysis,* dan *association rule* yang dilakukan oleh Saleh & AlShafeey (2025) mengkaji hubungan antara Industri 4.0 dan tiga pilar keberlanjutan (ekonomi, sosial, dan lingkungan). Hasil dari penelitian ini menunjukkan bahwa industri 4.0 memiliki persepsi positif pada tiap dimensi, dengan 94,5% positif pada dimensi ekonomi, 92,8% pada sosial, dan 88,5% pada lingkungan. *Association rule*

menunjukkan hubungan kuat antara Industri 4.0 dengan tiap dimensi, dimana dimensi ekonomi memiliki *support* tertinggi (0,444) dan *confidence* (0,855), diikuti oleh sosial (0,430 dan 0,831) dan lingkungan (0,380 dan 0,827). Hasil penelitian ini menunjukkan peran penting industri 4.0 dalam mendukung efisiensi energi, pengurangan limbah, pertumbuhan ekonomi sirkular, serta peningkatan keselamatan kerja dan kesejahteraan sosial, sekaligus menegaskan pendekatan integratif sebagai landasan pengembangan berkelanjutan berbasis teknologi.

Selanjutnya pada penelitian yang dilakukan oleh Anggina et al. (2022) yang menganalisis sentimen terhadap 741 ulasan pelanggan *Formaggio Coffee and Resto*, digunakan pendekatan *Multinomial Naïve Bayes* yang dikombinasikan dengan pendekatan *lexicon-based (InSet Lexicon)* dan *TF-IDF*. Pada awal tahapan, dilakukan *pre-proccessing data* (normalisasi, *stemming, filtering*), pelabelan otomatis berdasarkan polaritas kata, dan klasifikasi dengan validasi silang. Lalu untuk hasil uji secara keseluruhan menunjukkan performa yang cukup menjanjikan, dengan akurasi 95%, *precision* 85%, *recall* 68%, dan *f1-score* 72%. Pada hasil klasifikasi tiap aspek, aspek layanan memiliki performa klasifikasi terbaik, sedangkan aspek umum menunjukkan kelemahan dalam prediksi sentimen negatif. Penelitian ini menunjukkan bahwa integrasi metode leksikal dan *machine learning* dapat membantu memahami persepsi pelanggan secara efektif.

Pada penelitian yang dilakukan Laghbi & Al-Dhoayan (2024) digunakan pendekatan yang berbeda yaitu metode *ordinal logistic regression* dan *sentiment analysis VADER.* Penelitian ini menganalisis persepsi pelanggan terhadap 512 apotek di Riyadh, dimana data dikumpulkan dengan *scraping API* dan dibersihkan untuk validasi. Hasil dari penelitian ini menunjukkan bahwa skor apotek secara signifikan meningkatkan peluang skor ulasan yang lebih tinggi (OR = 3,734; $p < 0,001$), sementara jumlah *likes* dan total ulasan memiliki korelasi negatif. Sentimen positif mendominasi skor tinggi, dan terdapat hubungan signifikan antara sentimen dan nilai ulasan ($\chi^2 = 2991.315$; $p < 0,001$). Studi ini menegaskan bahwa ulasan pelanggan mencerminkan persepsi pelanggan, sehingga dapat dimanfaatkan untuk meningkatkan layanan farmasi.

Analisis sentimen berbasis aspek dan analisis emosi digunakan pada penelitian yang dilakukan oleh Sukwadi et al. (2025). Penelitian ini mengevaluasi persepsi pelanggan terhadap lima jaringan restoran cepat saji di Indonesia. Data yang diperoleh dari *review gmaps* dan kuesioner, dilakukan *pre-processing* serta klasifikasi aspek layanan (makanan, pelayanan, atmosfer, harga), dan dilanjut dengan pemrosesan *TF-IDF* dan uji klasifikasi dengan *Naïve*

Bayes Classifier. Hasil uji klasifikasi menunjukkan bahwa sebagian besar restoran menerima sentimen positif pada semua aspek, kecuali *Richeese* pada aspek pelayanan yang mendapat sentimen negatif dan emosi *"disappointed"*. Dengan hasil yang didapat, maka restoran yang masih memiliki sentimen negatif pada suatu aspek harus berbenah untuk memperbaiki permasalahan tersebut.

Selanjutnya pada penelitian yang dilakukan oleh Nofandi et al. (2023), dilakukan analisis sentimen untuk meningkatkan kualitas layanan warung wareg. Penelitian ini dimulai dengan dilakukannya *pre-processing* data mentah, yang selanjutnya dilakukan pelabelan(positif & negatif) secara manual, dan dilanjutkan dengan proses pembobotan *TF-IDF* dan klasifikasi dengan *SVM.* Hasil dari *confusion matrix* menunjukkan akurasi sebesar 94%, dengan *precision* dan *recall* tinggi di kedua kelas sentimen. Selain klasifikasi, dalam penelitian ini dilakukan *root cause analysis* untuk mengidentifikasi akar masalah dari ulasan negatif. Hasil rekomendasi yang diberikan pada penelitian ini diantaranya adalah penambahan pelayan saat *high season*, pelatihan sikap ramah, dan peningkatan kualitas makanan serta kebersihan lingkungan.

Penelitian yang dilakukan Gunawan et al. (2022) ini melakukan analisis emosi dan sentimen terhadap ulasan restoran di Kota Jakarta, menggunakan data ulasan yang diperoleh dari *TripAdvisor.* Setelah data didapat, proses *pre-processing* yang mencakup tokenisasi, *stopword removal, lowercasing,* dan *stemming* dilakukan. Selanjutnya dilakukan pelabelan tiap token dengan *total sentiment index (TSI).* Selanjutnya dilakukan pemodelan klasifikasi dengan algoritma *Naïve Bayes & Support Vector Machine*, dimana *SVM* menghasilkan akurasi terbaik sebesar 79%, lebih tinggi dari *NB* yang hanya mencapai 77%. Evaluasi menggunakan *confusion matrix* menunjukkan bahwa *SVM* lebih unggul dalam hal precision dan recall, hal ini menjadikan *SVM* metode yang lebih efektif untuk mengklasifikasi ulasan pelanggan berdasarkan polaritas sentimen yang telah dibuat.

Penelitian yang dilakukan Patil et al. (2024) menggabungkan algoritma *machine learning* dan *deep learning* untuk menganalisis klasifikasi sentimen ulasan sebuah restoran. Dataset uji yang dikumpulkan dari *Kaggle* dan *Zomato Hyderabad* diproses menggunakan teknik NLP (*tokenisasi, stopword removal, lemmatization*), dan setelah itu melalui tahapan klasifikasi dengan berbagai algoritma, dan pada penelitian ini digunakan *hyperparameter tuning* untuk optimasi model dengan *GridSearch*. Algoritma yang digunakan diantaranya adalah *Naïve Bayes, Logistic Regression, KNN, dan SVM,* serta terdapat juga uji dengan model *deep learning* yaitu *Convolutional Neural Networks (CNN)* dan *Bidirectional LSTM (Bi-LSTM).* Hasil

menunjukkan bahwa *Bi-LSTM* memberikan akurasi tertinggi sebesar 90%, diikuti *CNN* (89%), *Logistic Regression* (89.9%), dan *Multinomial Naïve Bayes* (89.6%). Pada penelitian ini, model deep learning khususnya *Bi-LSTM*, lebih unggul dalam memahami konteks dan struktur kalimat yang kompleks dalam ulasan pelanggan.

Li et al. (2021) dalam penelitiannya, mengembangkan model klasifikasi sentimen untuk ulasan restoran dengan menggabungkan *Word2Vec*, *Bi-GRU (Bidirectional Gated Recurrent Unit)*, dan *Attention sentiment analysis* yang nantinya akan dibandingkan dengan berbagai metode klasifikasi lainnya. Data dikumpulkan dari platform *Dianping* dan diproses melalui tahapan *pre-processing*, pelabelan POS, dan *word embedding* dengan *Word2Vec*. Hasil uji dan evaluasi menunjukkan bahwa model *Bi-GRU & ATT* mencapai F1-score lebih dari 89%, melampaui metode klasifikasi / *machine learning* lainnya. Studi ini menegaskan bahwa *deep learning* berbasis konteks dalam hal ini *Bi-GRU & ATT* secara signifikan meningkatkan performa klasifikasi dalam ulasan restoran, jika dibandingkan dengan metode *machine learning* lainnya.

Dalam penelitian yang dilakukan oleh Peretz et al. (2024), mereka memperkenalkan metode *machine learning* yaitu *Naïve Bayes Enrichment Method (NBEM)*, dimana terdapat prosedur ensemble yang meningkatkan performa klasifikasi Naïve Bayes melalui pembobotan otomatis dan pengelompokan fitur berdasarkan tipe data. Proses dimulai dengan seleksi fitur otomatis menggunakan AutoTFS, dilanjutkan dengan pelatihan tiga varian Naïve Bayes (Bernoulli, Gaussian, dan Categorical) pada sub-dataset terpisah. Hasil prediksi dari masing-masing model kemudian digabungkan menggunakan fungsi klasifikasi berbobot. Metode ini diuji pada 20 dataset multivariat dan menghasilkan peningkatan recall hingga 76,3% dan F1-score hingga 43,9% dibandingkan model Naïve Bayes standar. Studi ini menunjukkan bahwa NBEM efektif mengatasi sensitivitas terhadap data bising.

Dalam penelitian studi komparatif yang dilakukan Selimovic et al. (2024), mereka melakukan evaluasi peran dari analisis sentimen dalam pengelolaan & strategi pemasaran menggunakan ulasan dari 20.000 hotel di web *TripAdvisor*. Metode yang diuji meliputi *Naïve Bayes, Support Vector Machine (SVM), Recurrent Neural Networks (RNN), dan Long Short-Term Memory (LSTM).* Dalam studi ini, algoritma *SVM* memeroleh performa terbaik, dengan tingkat presisi sebesar 83,86%, diikuti oleh *RNN* (81,46%)*, LSTM* (81,53%)*, NB* (73,81%). Hal ini menggambarkan pentingnya memilih metode yang tepat yang sesuai dengan kegiatan pengumpulan data. Secara keseluruhan, analisis sentimen adalah *tools* yang dapat selalu

berperan dalam pelengkap hubungan bisnis dengan pelanggan untuk pengambilan keputusan dalam bisnis.

Tabel 2. 1 Kajian Literatur

Penulis	Obyek (Industri)		Metode			
	Umum	Restoran	SVM	NB	AR	KNN
(Singgalen, 2022)		√	√	√		
(Ode et al., 2024)		√		√		
(Ikhsan et al., 2025)	√		√	√		
(Amalia et al., 2021)		√	√			√
(Suryana et al., 2024)	√			√		
(Saleh & AlShafeey, 2025)				√	√	
(Laghbi & Al-Dhoayan, 2024)	√					
(Anggina et al., 2022)		√		√		
(Sukwadi et al., 2025)		√		√		
(Nofandi et al., 2023)	√		√			
(Gunawan et al., 2022)		√	√	√		
(Patil et al., 2024)		√		√		√
(Li et al., 2021)		√	√			√
(Peretz et al., 2024)	√			√		
(Selimovic et al., 2024)			√	√		
Usulan		√		√	√	

Pembahasan mengenai kajian literatur diatas menggambarkan bagaimana penelitian terdahulu melakukan proses analisis sentimen terhadap ulasan pelanggan. Proses tersebut

secara umum melewati tahapan data *pre-processing* yang meliputi (*normalization, case folding, stemming,* dsb) yang berguna untuk membuat data lebih *compact*, dengan menyisakan kata-kata penting dan berbobot sebelum uji klasifikasi dilakukan. Adapun berbagai metode klasifikasi digunakan dalam penelitian-penelitian terdahulu, seperti *Naïve Bayes, Support Vector machine, K-Nearest Neighbor*, dan berbagai metode klasifikasi lainnya yang kurang populer untuk digunakan. Dalam kajian diatas, berbagai penelitian menunjukkan bahwa dalam setiap kasus yang berbeda, berbagai metode klasifikasi yang digunakan memiliki keunggulannya masing-masing. Sehingga dalam rencana penelitian analisis sentimen ulasan pelanggan Grobbak Korea yang akan peneliti lakukan, peneliti berencana menggunakan metode klasifikasi seperti *Naïve Bayes* yang cukup populer digunakan dalam uji klasifikasi dan terbukti memiliki performa yang baik berdasarkan penelitian-penelitian terdahulu. Lalu untuk analisis lanjutan, peneliti berencana untuk menggunakan metode *Association rule* untuk mengetahui asosiasi dari kata-kata yang ada dalam ulasan. Sehingga, dengan metode tersebut peneliti dapat menarik kesimpulan dari asosiasi kata-kata dalam ulasan negatif, dan mengubahnya menjadi usulan perbaikan yang dapat ditawarkan kepada grobbak Korea.

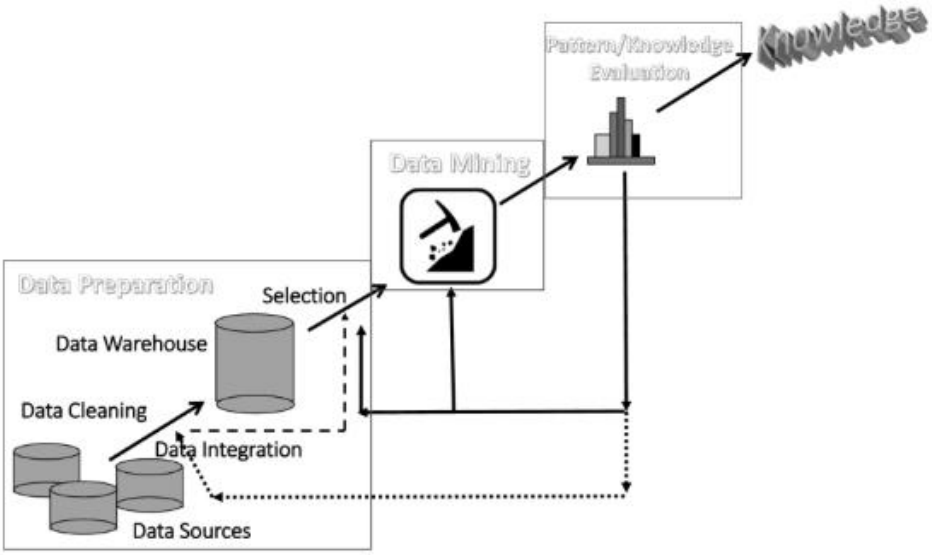
2.2 Landasan Teori

2.2.1 *Data Mining*

Data Mining yang secara harfiah berarti penambangan data, dapat diartikan sebagai penambangan pengetahuan dari sebuah kumpulan data(*knowledge discovery from data*). Pada dasarnya, *data mining* adalah proses menemukan pola, model, dan berbagai jenis pengetahuan menarik dalam kumpulan data yang besar (Han et al., 2024). *Data Mining* penting, dan diperlukan dikarenakan *data mining* berusaha untuk mengidentifikasi pola dan hubungan tersembunyi yang tidak selalu jelas, sehingga mudah diidentifikasi dibawah keadaan asumsi tertentu (Gorunescu, 2011).

2.2.2 *Knowledge Discovery from Data (KDD)*

KDD memiliki beberapa tahapan dan iterasi didalamnya, berikut merupakan tahapan KDD yang dilakukan dalam *knowledge discovery process* (Han et al., 2024):



Gambar 2. 1 Langkah proses *knowledge discovery*

Sumber: (Han et al., 2024)

1. *Data Preparation*
    a. *Data Cleaning* (menghilangkan data bising dan inkonsisten)
    b. *Data Integration* (menyatukan data dari berbagai sumber data)
    c. *Data Transformation* (data diubah kedalam bentuk yang sesuai untuk proses *data mining*)
    d. *Data Selection* (pengambilan data relevan dari *database*)
2. *Data Mining* (proses penting dimana berbagai metode diterapkan untuk mengekstraksi pola atau membangun model)
3. *Pattern / model evaluation* (mengidentifikasi pola atau model yang benar-benar menarik yang merepresentasikan pengetahuan berdasarkan ukuran keterarikan)
4. *Knowledge presentation* (teknik visualisasi dan representasi pengetahuan digunakan untuk menyampaikan *discovered knowledge* yang telah ditambang)

2.2.3 *Machine Learning*

Machine learning adalah cabang ilmu komputer yang secara luas bertujuan untuk memungkinkan komputer untuk "belajar" tanpa diprogram secara langsung (Bi et al., 2019). *Machine learning* merupakan model yang bertujuan untuk menemukan fungsi , ketergantungan, serta struktur yang tidak diketahui antara variabel *input* dan *output* (Kwekha-Rashid et al., 2023). Dalam penelitian yang dilakukan oleh Somvanshi et al. (2016), *machine learning* terbagi

kedalam beberapa kategori, yaitu *Supervised learning, unsupervised learning,* dan *Reinforcement learning.* Ketiga kategori *machine learning* dapat dijabarkan sebagai berikut:

1. *Supervised learning*

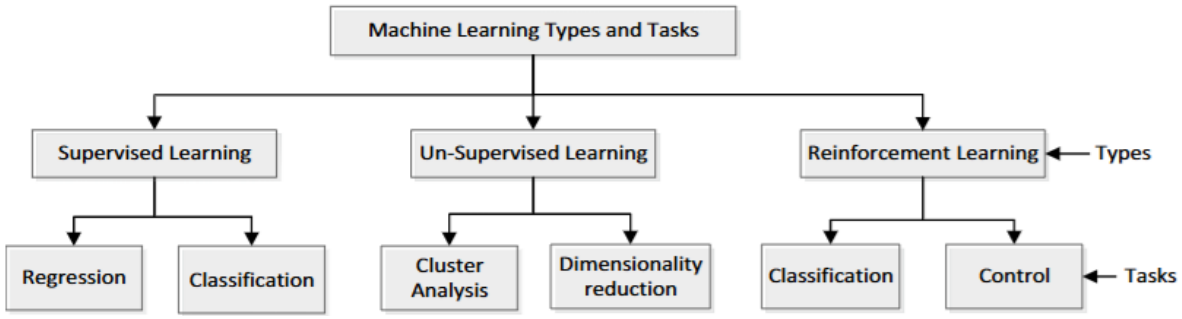
   Dalam jenis pembelajaran ini, sistem diberikan contoh *input* dan dipetakan dengan *output*, atau dengan kata lain *input & output* telah diketahui sebelumnya. Algoritma *supervised learning* memungkinkan untuk menentukan label kelas dengan tepat untuk sebuah *unseen instances.*

2. *Unsupervised learning*

   Dalam jenis pembelajaran *unsupervised learning*, system memiliki *input*, namun tidak memiliki *output* yang tersedia, atau dalam kata lain system tidak mengetahui seperti apakah *output* yang benar. Fungsi dari algoritma ini adalah mendefinisikan fungsi, untuk menggambarkan *hidden structure* dari data tidak berlabel.

3. *Reinforcement learning*

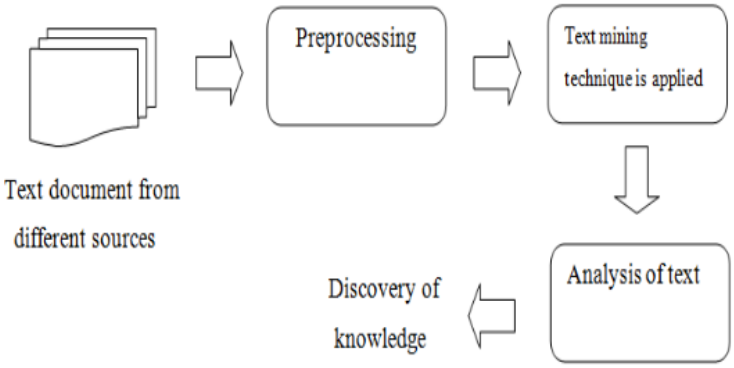
   Merupakan cabang *machine learning* yang didasarkan pada psikologi perilaku. *Reinforcement learning* berkaitan dengan bagaimana *software agent* harus bertindak dalam lingkungan untuk memaksimalkan gagasan tentang imbalan kumulatif.



Gambar 2. 2 Kategori pembelajaran *machine learning*
Sumber: (Kwekha-Rashid et al., 2023)

2.2.4 *Text Mining*

Menurut penelitian yang dilakukan oleh Aleqabie et al. (2024), *text mining* yang juga dikenal sebagai *natural language processing (NLP)*, adalah metode yang berguna untuk mengungkap nilai / *value* tersembunyi dalam suatu data tekstual. *Text mining* merupakan proses fleksibel yang dapat digunakan sebagai langkah *pre-processing* untuk proses *data mining* yang lebih lanjut (Chowdhury & Alzarrad, 2023). Dalam penelitian yang dilakukan oleh Gaikwad (2014), *text mining* terdiri dari beberapa langkah. Berikut merupakan langkah-langkah dari *text mining*:



Gambar 2. 3 Proses *Text mining*
Sumber: (Gaikwad, 2014)

2.2.5 Analisis Sentimen

Menurut Medhat et al. (2014), analisis sentimen dapat diartikan sebagai studi komputasional yang meneliti terkait opini, sikap, dan emosi yang berkaitan atau bertindak sebagai umpan balik / *feedback* terhadap suatu entitas. Entitas dapat mewakili individu, peristiwa, topik, dan lain sebagainya. Menurut Taboada (2016), sentimen dapat diartikan sebagai evaluasi baik positif maupun negatif terhadap suatu entitas, yang diungkapkan melalui bahasa. Dalam penelitian analisis sentimen, informasi dikumpulkan dan diolah dari kata-kata positif dan negatif yang ada dalam teks, konteks kata-kata tersebut, dan struktur linguistik suatu teks.

2.2.6 *Text Pre-processing*

Text pre-processing berfungsi untuk mengubah data tekstual yang tidak terstruktur seperti dokumen, kedalam data terstruktur dan disimpan kedalam *database* (Langgeni et al., 2010). Menurut Tanggraeni & Sitokdana (2022), untuk mengoptimalkan hasil analisis sentimen, terdapat beberapa tahapan *text pre-processing*, yaitu *case folding, punctuation removal, tokenizing, stemming,* dan *removal of stopwords*. Berikut merupakan penjelasan tahapan-tahapan *text pre-processing*:

1. *Case folding*

   Case folding merupakan proses pengubahan pada keseluruhan huruf kapital menjadi huruf kecil. Pada proses ini contohnya karakter 'A'-'Z' yang terdapat pada data diubah menjadi karakter 'a'-'z'.

2. *Punctuation removal / cleaning / normalization*

   Merupakan tahapan penghapusan tanda baca, karakter yang tidak diperlukan (seperti titik, koma, tanda tanya, tanda seru), huruf ganda, HTML, URL, *hashtag*, *mention*, *emoji*, dan karakter lain yang tidak relevan.

3. *Stopwords removal*

   Stopwords adalah kata-kata umum yang secara sering muncul dalam suatu bahasa, dan umumnya memberikan sedikit nilai informasi dalam analisis teks karena keberadaannya yang luas dan merata pada suatu dokumen. Dalam pemrosesan teks atau analisis *Natural Language Processing* (*NLP*), kata-kata ini dilakukan penghapusan karena makna yang tidak signifikan terhadap pemahaman isi teks. Proses ini membantu menyederhanakan teks dan memusatkan perhatian pada kata-kata yang lebih bermakna dan informatif.

4. *Tokenizing*

   Tokenizing adalah proses pemotongan teks menjadi bagian-bagian yang lebih kecil disebut token. Teks yang dipisah adalah kata-kata yang berdiri sendiri dengan batasan spasi dan tanda baca.

5. *Stemming*

   Tahap *stemming* adalah tahap untuk mengubah kata yang berimbuhan menjadi kata dasar melalui penghilangan imbuhan yang menjadi awalan, akhiran, maupun sisipan (mengubah kata menjadi bentuk kata asli).

2.2.7 *TF-IDF*

TF-IDF merupakan metrik yang cukup umum digunakan dalam kategorisasi teks. Menurut Kay et al. (2009), *TF-IDF* terdiri dari 2 jenis skor yaitu *term frequency & inverse document frequency. TF* dapat ditemukan dengan menghitung berapa kali sebuah istilah muncul dalam dokumen tertentu, sedangkan *IDF* ditemukan dengan membagi total jumlah dokumen dengan jumlah dokumen tempat kata tersebut muncul. Maka dari itu, *TF-IDF* dapat ditemukan dengan mengalikan nilai keduanya. Dengan metode ini, peneliti dapat mengetahui fitur-fitur yang penting akan memiliki nilai tinggi dan fitur yang tidak penting akan memiliki nilai rendah.

2.2.8 Klasifikasi

Menurut penelitian Han et al. (2011), klasifikasi merupakan bentuk analisis data yang mengekstrak model yang berguna untuk menggambarkan kelas data penting. Model pengklasifikasi (*classifier*), memprediksi label kelas kategoris (diskrit, tidak berurutan). Kategori ini dapat direpresentasikan oleh nilai-nilai diskrit, dimana urutan di antara nilai-nilai tidak memiliki arti. Klasifikasi merupakan proses 2 langkah, yang terdiri dari langkah pembelajaran (di mana model klasifikasi dibangun) dan langkah klasifikasi (di mana model digunakan untuk memprediksi label kelas untuk data yang diberikan).

Untuk hasil evaluasi dari metode klasifikasi, digunakan tabel *confusion matrix.* Uji ini menunjukkan berapa banyak prediksi yang terbukti benar (TP & TN) dan salah (FP & FN) pada setiap kelas yang ada. Berikut merupakan penjelasan komponen *confusion matrix* (Han et al., 2012):

- *True* Positive (TP): Merupakan tupel positif yang diberi label dengan benar oleh pengklasifikasi. TP adalah jumlah positif benar.
- *True Negatives* (TN): Merupakan tupel negatif yang diberi label dengan benar oleh pengklasifikasi. TN adalah jumlah negatif benar.
- *False Positive* (FP): Merupakan tupel negatif yang salah dilabeli sebagai positif. FP adalah jumlah positif salah.
- *False Negative* (FN): Merupakan tupel positif yang salah dilabeli sebagai negatif. FN merupakan jumlah negatif salah.

Lalu untuk metriks yang digunakan untuk mengevaluasi performa klasifikasi, digunakan metriks akurasi, presisi, dan *recall.* Berikut merupakan perumusan dari perhitungan metriks klasifikasi (Cynthia & Qoiriah, 2024):

- Menghitung akurasi

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.1)$$

- Menghitung presisi

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2.2)$$

- Menghitung *recall*

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (2.3)$$

2.2.9 *Naïve Bayes classifier*

Bayes classifier merupakan salah satu pengklasifikasi statistik, *classifier* ini dapat memprediksi probabilitas keanggotaan kelas suatu data tupel yang akan masuk ke dalam kelas tertentu, tentunya sesuai dengan perhitungan probabilitas. *Bayes classifier* didasari oleh teorema bayes yang ditemukan oleh Thomas Bayes pada abad ke-18 (Handayani et al., 2015). Berikut merupakan persamaan dari teorema bayes menurut Han et al. (2011) dan (Simanjuntak & Simatupang, 2022):

$$\mathrm{P(H|X)} = \frac{\mathrm{P(X|H)P(H)}}{\mathrm{P(X)}} \times 100\% \quad (2.4)$$

Keterangan:

X : Sampel data yang memiliki kelas(label) yang tidak diketahui

H : Hipotesis bahwa X adalah data kelas label

P(H) : Peluang hipotesis dari H

P(X) : Peluang dari data sampel yang diamati

P(X|H) : Peluang dari data sampel X bila diasumsikan hipotesis benar

2.2.10 *Support Vector Machine*

Support vector machine merupakan kategori *machine learning* yang termasuk dalam kategori *supervised learning.* Dengan kata lain, *Support Vector Machine* (SVM) adalah alat klasifikasi dan prediksi regresi yang menggunakan teori *machine learning* yang berguna dalam memaksimalkan akurasi prediktif sekaligus secara otomatis menghindari *over-fit* terhadap data (Jakkula, 2006).

2.2.11 *Association rule*

Association rule merupakan teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item atau suatu pekerjaan untuk menentukan atribut mana yang akan didapatkan bersamaan. *Association rule* adalah bentuk jika "kejadian sebelumnya" kemudian "konsekuensinya" *(IF antecedent, THEN consequent*) (Johan et al., 2019). Perhitungan *association rule* diikuti pula dengan perhitungan pada parameter *support*, *confidence*, dan *lift-ratio.* Berikut merupakan penjelasan ketiga parameter tersebut:

- *Support*

  Support merupakan presentasi gabungan item pada *database*, (jika didalamnya terdapat item A dan B) maka *support* dapat diartikan sebagai proporsi dari transaksi dalam database yang mengandung A dan B. Berikut merupakan rumus untuk mencari nilai *support* dari dua *item*:

$$Support\ (\mathrm{A}, \mathrm{B}) = \mathrm{P}(\mathrm{A} \cap \mathrm{B}) \tag{2.5}$$

$$Support\ (\mathrm{A}, \mathrm{B}) = \frac{\Sigma\ \text{Transaksi mengandung A dan B}}{\Sigma\ \text{Transaksi}} \tag{2.6}$$

- *Confidence*

  Pada *association rule*, *Confidence* adalah takaran ketepatan pada presentasi transaksi dalam database yang mengandung A dan B. Dengan penggunaan *confidence* dalam

association rule, kita dapat mengukur kuatnya hubungan antar *item*. Berikut adalah Rumus dalam mencari nilai confidence pada dua *item*:

$$Confidence = \mathrm{P}(\mathrm{B} \mid \mathrm{A}) \tag{2.7}$$

$$Confidence(\mathrm{A}, \mathrm{B}) = \frac{\Sigma \text{ Transaksi mengandung A dan B}}{\Sigma \text{ Transaksi mengandung A}} \tag{2.8}$$

- *Lift ratio*

  Lift ratio adalah suatu cara yang dilakukan untuk mengetahui seberapa besar ketahanan dari suatu aturan assosiasi yang sudah terbentuk sebelumnya. Nilai *lift ratio* memiliki kegunaan sebagai penentuan validitas suatu aturan asosiasi. Perhitungan *lift ratio* dirumuskan sebagai berikut:

$$Lift\ ratio = \frac{Support\ (A \cap B)}{Support\ (A) * Support\ (B)} \tag{2.9}$$

BAB III

METODE PENELITIAN

3.1 Subjek Penelitian

Subjek pada penelitian yang penulis lakukan pada laporan ini adalah salah satu *outlet* Grobbak Korea yang beralamatkan di Jalan Perintis Kemerdekaan No.35D, Pandeyan, Kec. Umbulharjo, Kota Yogyakarta, Daerah Istimewa Yogyakarta 55182. Dimana Grobbak Korea merupakan *franchise* yang bergerak dibidang kuliner.

3.2 Objek Penelitian

Pada penelitian ini, objek yang diteliti merupakan analisis sentimen yang terkait ulasan pelanggan Grobbak Korea *outlet* Pandeyan, sebagai upaya perbaikan kualitas pelayanan.

3.3 Jenis Data & Sumber Data

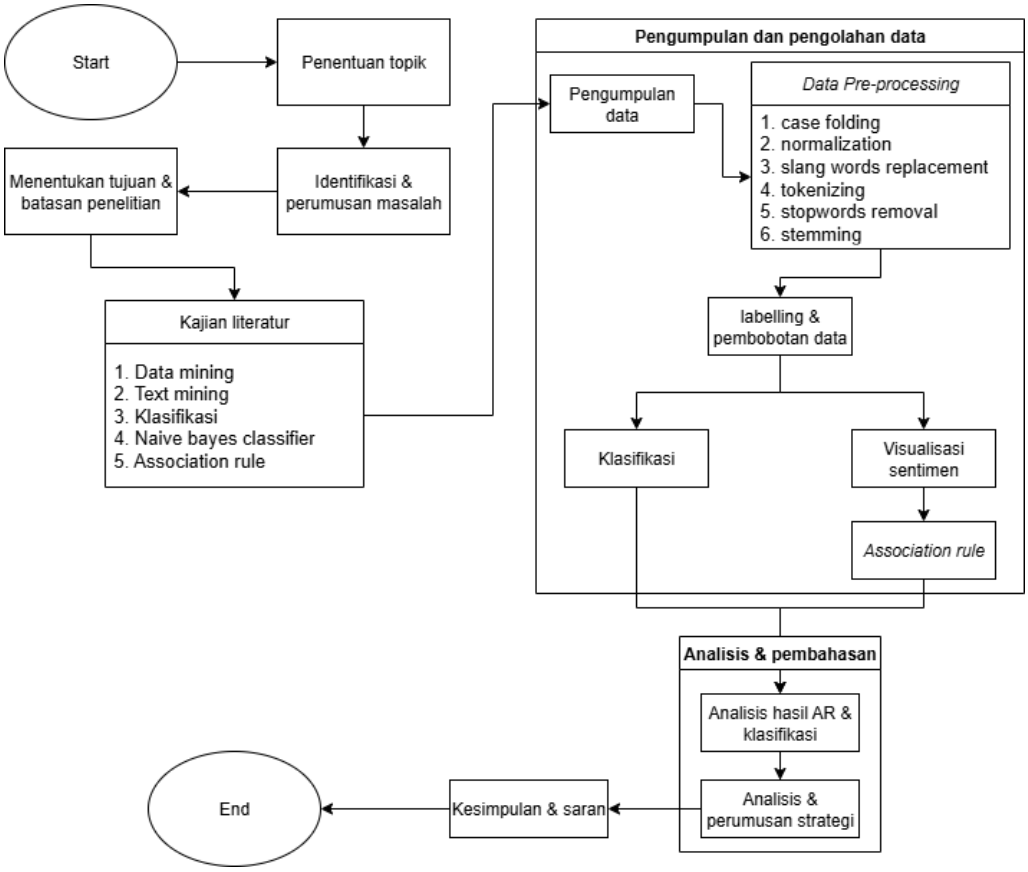
Dalam penelitian ini digunakan data sekunder, dimana data sekunder merupakan data yang dikumpulkan dari data yang telah ada sebelumnya. Untuk sumber data yang digunakan sendiri berasal dari data ulasan pelanggan yang terdapat pada *platform Gofood* dari Grobbak Korea *outlet* Pandeyan.

3.4 Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini dilakukan dengan cara *web scraping*. *Web scraping* merupakan *tools* yang memungkinkan ekstraksi sejumlah data yang secara konstan berkembang, dengan biaya yang relatif rendah (Khder, 2021). Dalam penelitian ini, digunakan bantuan pemrograman *python* sebagai alat dalam melakukan proses *web scraping.*

3.5 Alur Penelitian

Berikut merupakan alur dari penelitian yang penulis lakukan. Alur penelitian ini dapat digambarkan dengan diagram alir dibawah ini:



Gambar 3. 1 Diagram alur penelitian

Berikut merupakan penjelasan dari diagram alir dari penelitian yang penulis lakukan:

1. Identifikasi & perumusan masalah

   Identifikasi permasalahan disini merupakan tahap awal, dimana penulis mengidentifikasi terkait permasalahan yang terjadi pada subjek penelitian yaitu Grobbak Korea. Setelah permasalahan sudah teridentifikasi, selanjutnya dirumuskan kedalam pertanyaan-pertanyaan sehingga nantinya dapat ditentukan tujuan dari penelitian yang penulis lakukan.
2. Penentuan tujuan & Batasan penelitian

   Pada tahapan ini ditentukan tujuan penelitian yang akan menjawab rumusan masalah yang telah dibuat. Diharapkan dengan tujuan penelitian yang telah ditentukan dapat

menjawab permasalahan yang ada. Selanjutnya adalah penentuan batasan penelitian, dimana batasan penelitian akan menentukan lingkup subjek, objek, serta proses dari penelitian yang penulis lakukan.

3. Kajian literatur

   Peneliti melakukan kajian terhadap literatur-literatur terdahulu yang berkaitan dengan penelitian yang akan penulis lakukan. Kajian literatur yang penulis lakukan berguna untuk membandingkan berbagai metode penelitian terkait, keunggulan serta kelemahan dari sebuah metode, serta hasil dari penelitian terdahulu yang terkait dengan penelitian penulis.

4. Pengumpulan & pengolahan data

   Pada tahap ini, dilakukan pengumpulan data, dimana penulis menggunakan metode *web scraping* dengan bantuan permrograman *python*. Lalu untuk pengolahan data sendiri terbagi kedalam beberapa tahapan. Tahapan pertama adalah data *pre-processing*, dimana tahapan ini terdiri dari proses *case folding, normalization, stopwords removal, tokenizing,* dan *stemming.* Selanjutnya adalah tahapan *labelling* dan pembobotan data dimana *labelling* dilakukan dengna bantuan kamus *lexicon*. Lalu untuk pembobotan kata-kata, digunakan bantuan algoritma *TF-IDF* dimana kata-kata diubah menjadi bobot vektor.

   Pada tahapan selanjutnya dilakukan klasifikasi menggunakan *Naïve Bayes classifier* untuk mengetahui hasil klasifikasi label dari model yang telah dibuat. Bersamaan dengan tahapan ini, dilakukan visualisasi sentimen pelanggan dengan tampilan *wordcloud.* Selanjutnya dilakukan analisis *association rule* untuk mengetahui asosiasi kata yang terbentuk dari ulasan pelanggan Grobbak Korea. Selanjutnya hasil *association rule* ini akan dianalisis, sehingga ditemukan permasalahan yang mungkin terjadi pada Grobbak Korea *outlet* Pandeyan.

5. Analisis dan pembahasan

   Setelah dilakukan uji klasfikasi serta *association rule*, maka pada tahapan ini dilakukan analisis terhadap hasil uji yang telah dilakukan. Dari analisis hasil uji ini, akan diekstraksi menjadi sebuah solusi yang dapat ditawarkan yang merupakan hasil kajian dalam tahapan analisis & perumusan strategi. Serta diikuti pula dengan pembahasan terkait apa solusi yang dapat ditawarkan dalam upaya penyelesaian permasalahan yang terjadi pada *outlet* Grobbak Korea tersebut.

6. Kesimpulan dan saran

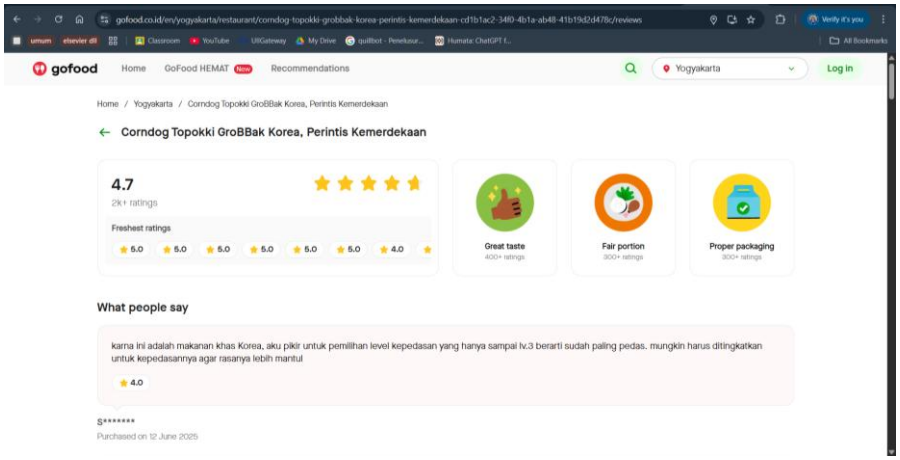
   Pada kesimpulan, penulis akan menjelaskan Kesimpulan dari penelitian yang penulis lakukan, dan tentunya dapat menjawab tujuan dari penelitian yang sudah ditentukan diawal. Lalu pada tahap terakhir penulis memberikan saran-saran bagi peneliti selanjutnya untuk dapat mengembangkan penelitian ini lebih lanjut, serta pemberian saran perbaikan bagi Grobbak Korea sebagai subjek dari penelitian ini.

BAB IV

PENGUMPULAN DAN PENGOLAHAN DATA

4.1 Pengumpulan Data

Pada tahapan ini, dilakukan pengambilan data yang akan penulis teliti. Data yang diambil pada penelitian ini merupakan data ulasan konsumen dari Grobbak korea *outlet* Pandeyan, yang diambil pada laman *Gofood* dari *outlet* Grobbak Korea Pandeyan.

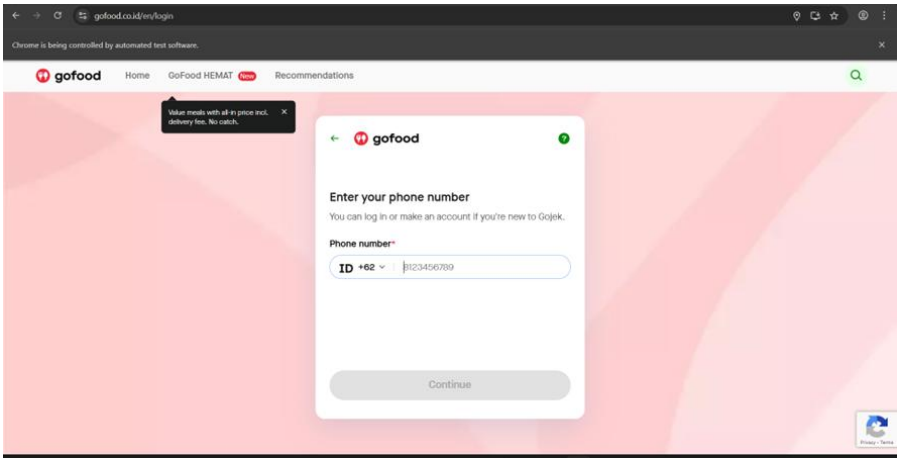


Gambar 4. 1 Laman *Gofood* Grobbak Korea Pandeyan

Sumber: https://*Gofood*.co.id/

Pengambilan data dilakukan dengan teknik *scraping* dengan bantuan pemrograman *python.* Dalam proses pengumpulan data yang penulis lakukan, terdiri dari beberapa tahapan yaitu:

1. Langkah pertama, domain dari laman *Gofood reviews* Grobbak Korea pandeyan dimasukkan kedalam kode yang telah dibuat. Setelah *web domain* dari laman *Gofood* Grobbak Korea Pandeyan sudah dimasukkan kedalam kode, maka kode siap dijalankan.
2. Pada langkah kedua, kode yang telah dibuat selanjutnya dijalankan. Lalu saat otomasi terbukanya laman *reviews* Grobbak Korea Pandeyan pada *web browser*, penulis akan melakukan proses *login* dengan akun Gojek.



Gambar 4. 2 Proses *login* pada *website* Gojek

Setelah proses *login* dilakukan, maka secara otomatis data ulasan akan mulai dilakukan *scraping* dari laman Grobbak Korea Pandeyan.

3. Selanjutnya, setelah proses scraping selesai dilakukan, data akan dijadikan dalam satu *file* csv yang nantinya akan diproses pada tahapan *text pre-processing.*

Tabel 4. 1 Contoh hasil *scraping* data

Item	Ulasan	*Rating*
Corndog Sosis Mozzarella - Corndog Mozarella	sosisnya kecil banget	4.0
Rabokki – Topokki extra odeng	Kuahnya tidak begitu enak	3.0
Rabokki - Corndog Mozarella - Corndog Mozarella	lebih teliti lg membaca pesanan ,terimakasih	4.0
Rabokki Extra Mozzarella - Corndog Matcha	tolong ya lain kali packing yang bener. sampe tumpah gini kuahnya.	2.0
Odeng Fish Soup - Corndog Sosis Mozzarella - Pakokki	Untuk mbk/mas yang jualan tolong di teliti lagi kalo ada orderan..	1.0

4.2 Pengolahan Data

Setelah data ulasan telah berhasil dilakukan *scraping* dari *website Gofood*, maka selanjutnya data ulasan akan melalui tahapan pengolahan data. Data ulasan akan diolah untuk nantinya didapatkan informasi yang dapat digunakan dalam perencanaan perbaikan kualitas pelayanan dari *outlet* Grobbak Korea Pandeyan. Berikut merupakan tahapan-tahapan dari pengolahan data dalam penelitian ini:

4.2.1 *Text Pre-processing*

Text pre-processing memiliki peran yang krusial dalam analisis data teks, khususnya pada kasus *text mining dan machine learning*. Dalam tahapan *pre-processing* dilakukan proses *cleaning, normalization, and preparation* dari data teks agar lebih terstruktur, sehingga mudah diolah pada tahapan analisis lanjutan. Dalam penelitian ini, dilakukan serangkaian tahapan *pre-processing* dengan bantuan pemrograman python agar data menjadi lebih siap digunakan untuk proses analisis sentimen serta ekstraksi informasi. Berikut merupakan tahapan *pre-processing* yang dilakukan:

1. *Case folding*

   Case folding adalah proses mengubah seluruh huruf pada data teks menjadi huruf kecil (*lowercase*). Tujuannya adalah untuk menyamakan representasi kata, sehingga kata yang sama namun berbeda penulisan kapitalnya (misal: “Enak” dan “enak”) dapat dianggap sama oleh sistem. Adapun contoh dari proses *case folding* dapat dilihat dibawah ini:

Tabel 4. 2 Tahapan *case folding*

Before	*After*
"MAKANANNYAAA ENAAAKKKKKK bangeeetttt, tp pesenan dtg lmaaaaaa, ga rekomen"	"makanannyaaa enaaakkkkkkk bangeeetttt, tp pesenan dtg lmaaaaaa, ga rekomen"
"Driver nya ramah, tp makanannya dingin bgt, pesanan telat dtg, kecewaaaaaa"	"driver nya ramah, tp makanannya dingin bgt, pesanan telat dtg, kecewaaaaaa"

2. Normalisasi

   Normalisasi dilakukan untuk menghilangkan karakter yang tidak diperlukan, seperti tanda baca, angka pada kata, serta pengulangan karakter yang tidak merepresentasikan kata. Selain itu, normalisasi juga menghapus karakter khusus, dan spasi berlebih sehingga teks menjadi lebih bersih dan konsisten. Adapun contoh dari proses normalisasi dapat dilihat dibawah ini:

Tabel 4. 3 Tahapan normalisasi

Before	*After*
"makanannyaaa enaaakkkkkkk bangeeetttt, tp pesenan dtg lmaaaaaa, ga rekomen"	"makanan enak banget tp pesanan dtg lama ga rekomen"
"driver nya ramah, tp makanannya dingin bgt, pesanan telat dtg, kecewaaaaaa"	"driver nya ramah tp makanan dingin bgt pesanan telat dtg kecewa"

3. *Slang words replacement*

   Pada tahap ini, kata-kata tidak baku atau slang (misal: “gk”, “bgt”, “mantep”) diganti dengan padanan kata baku sesuai dengan kamus yang telah disusun. Proses ini penting agar kata-kata yang memiliki makna sama dapat diolah secara konsisten pada tahap selanjutnya. Adapun contoh dari proses *slang words replacement* dapat dilihat dibawah ini:

Tabel 4. 4 Tahapan *slang words replacement*

Before	*After*
"makanan enak banget tp pesanan dtg lama ga rekomen"	"makanan enak banget tapi pesanan datang lama tidak rekomendasi"
"driver nya ramah tp makanan dingin bgt pesanan telat dtg kecewa"	"driver nya ramah tapi makanan dingin banget pesanan telat datang kecewa"

4. *Tokenize*

   Tokenisasi adalah proses memecah kalimat atau teks menjadi potongan-potongan kata (token). Atau dapat diartikan bahwa kata dijadikan sebuah *item*, sehingga sebuah kalimat dapat dikatakan sebagai *set item*. Setiap token akan diproses lebih lanjut pada tahap berikutnya. Adapun contoh dari proses *tokenizing* dapat dilihat dibawah ini:

Tabel 4. 5 Tahapan *tokenize*

Before	*After*
"makanan enak banget tapi pesanan datang lama tidak rekomendasi"	['makanan', 'enak', 'banget', 'tapi', 'pesanan', 'datang', 'lama', 'tidak', 'rekomendasi']
"driver nya ramah tapi makanan dingin banget pesanan telat datang kecewa"	['driver', 'nya', 'ramah', 'tapi', 'makanan', 'dingin', 'banget', 'pesanan', 'telat', 'datang', 'kecewa']

5. *Stopwords removal*

   Stopwords removal adalah proses penghapusan kata yang dianggap tidak memiliki makna penting dalam analisis, seperti “dan”, “di”, “ke”, dan sebagainya. Pada penelitian ini, kata “tidak” tetap dipertahankan karena memiliki peran penting dalam membentuk

dan mempertahankan makna pada ulasan. Adapun contoh dari proses *stoopwords removal* dapat dilihat dibawah ini:

Tabel 4. 6 Tahapan *stopwords removal*

Before	*After*
['makanan', 'enak', 'banget', 'tapi', 'pesanan', 'datang', 'lama', 'tidak', 'rekomendasi']	['makanan', 'enak', 'banget', 'pesanan', 'datang', 'lama', 'tidak', 'rekomendasi']
['driver', 'nya', 'ramah', 'tapi', 'makanan', 'dingin', 'banget', 'pesanan', 'telat', 'datang', 'kecewa']	['driver', 'ramah', 'makanan', 'dingin', 'banget', 'pesanan', 'telat', 'datang', 'kecewa']

6. *Stemming*

Stemming adalah proses mengubah kata berimbuhan menjadi bentuk dasarnya (*root word*). Proses ini dilakukan menggunakan *library Sastrawi*. Adapun contoh dari proses *stemming* dapat dilihat dibawah ini:

Tabel 4. 7 Tahapan *stemming*

Before	*After*
['makanan', 'enak', 'banget', 'pesanan', 'datang', 'lama', 'tidak', 'rekomendasi']	"makan enak banget pesan datang lama tidak rekomendasi"
['driver', 'ramah', 'makanan', 'dingin', 'banget', 'pesanan', 'telat', 'datang', 'kecewa']	"driver ramah makanan dingin banget pesan telat datang kecewa"

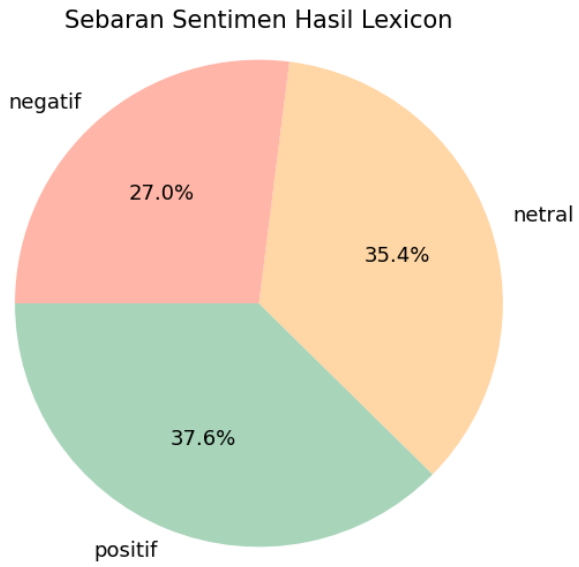
4.2.2 Pelabelan & pembobotan data

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. 8 Contoh hasil pelabelan dengan kamus *lexicon*

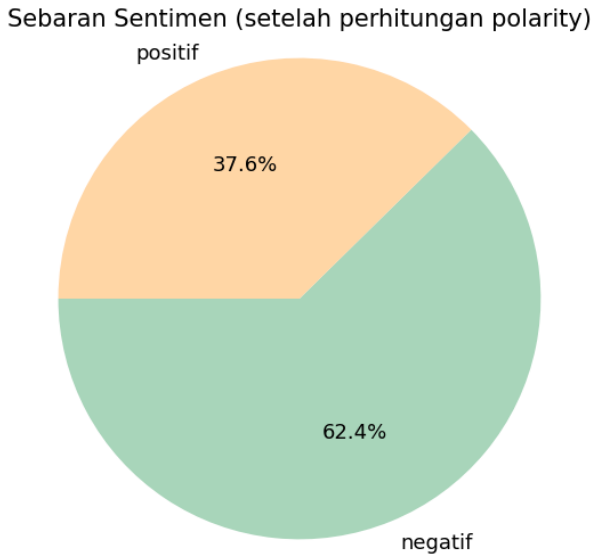
review	*Skor lexicon*	*sentiment*
"sosis kecil banget"	0	netral
"kaget pas datang tteok kecil banget chikuwa tipis banget potong saos super encer…."	-4	negatif
"enak banget corndog matcha the best"	1	positif
"corndog enak bibimyeon enak pedas chikuwa sama odeng banyak"	1	positif
"kira bakal kecil banget ternyata standar corndoglah"	0	netral

Untuk hasil keseluruhan dari pelabelan sentimen dari keseluruhan ulasan pelanggan yang telah didapatkan yaitu sebanyak 226 ulasan, memiliki sebaran sentimen sebanyak 85 ulasan bersentimen positif, 80 ulasan bersentimen netral, dan 61 ulasan bersentimen negatif.



Gambar 4. 3 Distribusi sentimen ulasan pelanggan

Dikarenakan dalam penelitian ini hanya digunakan label positif dan negatif dalam pembangunan model klasifikasi, maka sentimen dengan label netral dihitung kembali *polarity*-nya agar dapat diketahui ulasan yang masih berlabel netral tersebut sebenarnya lebih condong mengandung sentimen positif ataupun negatif. Namun jika ulasan benar-benar mengandung sentimen yang netral, maka ulasan tersebut tidak digunakan dalam pembangunan model klasifikasi dalam penelitian ini.



Gambar 4. 4 Distribusi sentimen pasca perhitungan *polarity*

Setelah perhitungan *polarity* dilakukan, sebaran ulasan pelanggan terdiri dari 37,6% atau 85 ulasan berlabel positif, dan 62,4% atau 141 ulasan berlabel negatif, sehingga jumlah *dataset* masih sama dengan sebelumnya yaitu 226 ulasan. Lalu setelah data melalui proses pelabelan dengan kamus *lexicon*, maka selanjutnya data akan dihitung bobotnya menggunakan fitur *TF-IDF*. Fitur *TF-IDF* dalam data ini berguna untuk menghitung bobot suatu kata berdasarkan frekuensi kemunculan kata dalam dokumen yang ada. Perhitungan bobot dalam penelitian ini juga dilakukan dengan bantuan pemrograman *python.*

4.2.3 *Dataset splitting*

Setelah melalui tahapan pelabelan dan pembobotan data, selanjutnya dilakukan tahapan klasifikasi. Pada penelitian ini, klasifikasi dilakukan dengan algoritma *Naïve Bayes.* Sebelum dilakukan proses klasifikasi, maka akan dilakukan *split dataset* menjadi *training dataset & testing dataset*. Pada dasarnya, *dataset* pelatihan digunakan untuk membangun model, sedangkan *dataset* pengujian digunakan untuk menilai kemampuan prediktif model (Nguyen et al., 2021). Menurut penelitian, model menjadi lebih stabil dan kinerja pelatihan meningkat

seiring dengan peningkatan kuantitas *dataset training*. Ukuran *training set* umumnya memiliki dampak yang signifikan terhadap kapabilitas prediksi dari sebuah model *machine learning* (Qi et al., 2018). Sehingga dalam penelitian ini, dilakukan uji klasifikasi dengan berbagai ukuran *data split.* Berikut merupakan pembagian *testing & training dataset* yang digunakan dalam penelitian ini:

1. Perbandingan *Data Training* sebesar 90% dan *Data Testing* sebesar 10%

Tabel 4. 9 Variasi *dataset split* 1

	Persentase	Jumlah
Training data	90	203
Testing data	10	23
Total	100	226

Berdasarkan tabel 4.9, digunakan perbandingan jumlah data sebesar 90% sebagai *data training*, dan 10% data digunakan sebagai *data testing*. Sehingga dari total 226 ulasan, 203 data digunakan sebagai *data training* dan 23 data digunakan sebagai *data testing* dengan perbandingan *data training* dan *data testing* adalah 9:1.

2. Perbandingan *Data Training* sebesar 80% dan *Data Testing* sebesar 20%

Tabel 4. 10 Variasi *dataset split* 2

	Persentase	Jumlah
Training data	80	181
Testing data	20	45
Total	100	226

Berdasarkan tabel 4.10, digunakan perbandingan jumlah data sebesar 80% sebagai *data training*, dan 20% data digunakan sebagai *data testing*. Sehingga dari total 226 ulasan, 181 data digunakan sebagai *data training* dan 45 data digunakan sebagai *data testing* dengan perbandingan *data training* dan *data testing* adalah 8:2.

3. Perbandingan *Data Training* sebesar 70% dan *Data Testing* sebesar 30%

Tabel 4. 11 Variasi *dataset split* 3

	Persentase	Jumlah

Training data	70	158
Testing data	30	68
Total	100	226

Berdasarkan tabel 4.11, digunakan perbandingan jumlah data sebesar 70% sebagai *data training*, dan 30% data digunakan sebagai *data testing*. Sehingga dari total 226 ulasan, 158 data digunakan sebagai *data training* dan 68 data digunakan sebagai *data testing* dengan perbandingan *data training* dan *data testing* adalah 7:3.

4.2.4 Klasifikasi

Setelah dilakukannya pembagian *dataset* latih dan uji, maka tahapan selanjutnya akan dilakukan proses *machine learning* yaitu klasifikasi menggunakan algoritma *Naïve Bayes*. Berikut merupakan skenario dan hasil pengujian klasifikasi pada penelitian ini:

1. Skenario uji dengan perbandingan *Data Training* 90% dan *Data Testing* 10%

Tabel 4. 12 Metriks hasil uji skenario 1

Output	Algoritma *Naïve Bayes Classifier*
Accuracy	0.783
Precision	0.814
Recall	0.783

Dari tabel diatas, diperoleh nilai akurasi sebesar 78.3%, *precision* sebesar 81.4%, dan *recall* sebesar 78.3%. Selanjuntya, nilai kinerja pada suatu hasil model prediksi klasifikasi diperoleh dengan menggunakan *confusion matrix*. Berikut merupakan hasil *confusion matrix* dari skenario uji 1:

Tabel 4. 13 *Confusion matrix* skenario 1

Prediction	*Negative*	*Positive*
Negative	*True Negative (TN)* 10	*False Negative (FN)* 4
Positive	*False Positive (FP)* 1	*True Positive (TP)* 8

2. Skenario uji dengan perbandingan *Data Training* 80% dan *Data Testing* 20%

Tabel 4. 14 Metriks hasil uji skenario 2

Output	Algoritma *Naïve Bayes Classifier*
Accuracy	0.848
Precision	0.849
Recall	0.848

Dari tabel diatas, diperoleh nilai akurasi pada model klasifikasinya sebesar 84.8%, nilai *precision* sebesar 84.9%, dan nilai *recall* sebesar 84.8%. Selanjuntya, nilai kinerja pada suatu hasil model prediksi klasifikasi diperoleh dengan menggunakan *confusion matrix*. Berikut merupakan hasil *confusion matrix* dari skenario uji 2:

Tabel 4. 15 *Confusion matrix* skenario 2

Prediction	*Negative*	*Positive*
Negative	*True Negative (TN)* 27	*False Negative (FN)* 2
Positive	*False Positive (FP)* 5	*True Positive (TP)* 12

3. Skenario uji dengan perbandingan *Data Training* 70% dan *Data Testing* 30%

Tabel 4. 16 Metriks hasil uji skenario 3

Output	Algoritma *Naïve Bayes Classifier*
Accuracy	0.794
Precision	0.800
Recall	0.794

Dari tabel diatas, diperoleh nilai akurasi pada model klasifikasinya sebesar 79.4%, nilai *precision* sebesar 80%, dan nilai *recall* sebesar 79.4%. Selanjuntya, nilai kinerja pada

suatu hasil model prediksi klasifikasi diperoleh dengan menggunakan *confusion matrix*. Berikut merupakan hasil *confusion matrix* dari skenario uji 3:

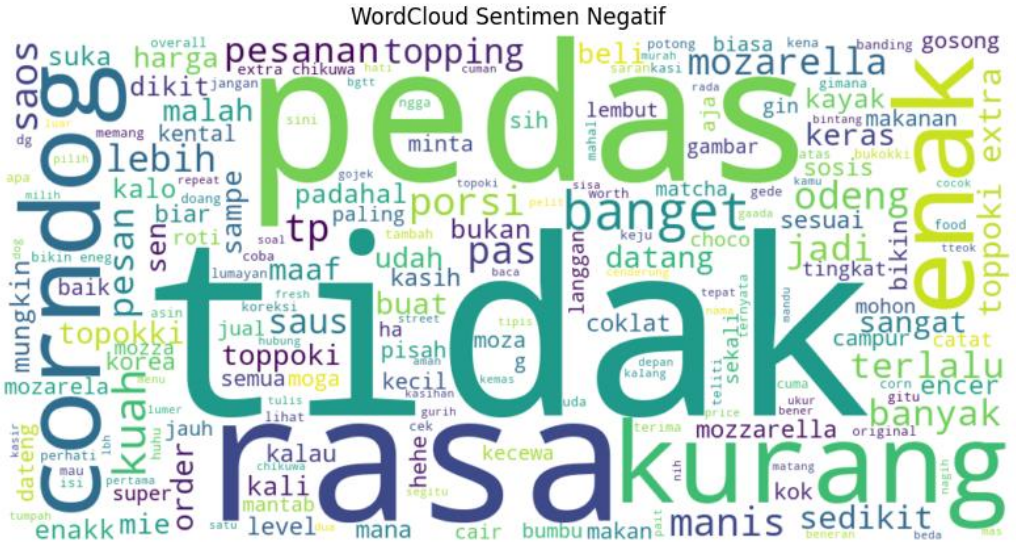
Tabel 4. 17 *Confusion matrix* skenario 3

Prediction	*Negative*	*Positive*
Negative	*True Negative (TN)* 39	*False Negative (FN)* 3
Positive	*False Positive (FP)* 11	*True Positive (TP)* 15

4.2.5 *Association rule*

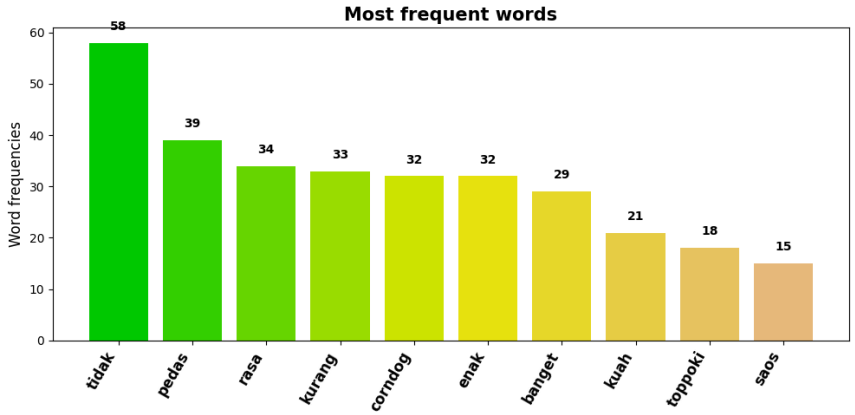
Pada tahapan ini, hasil klasifikasi ulasan yang memiliki label negatif akan menjadi *input* dari tahapan ini. Pada tahapan ini, ulasan berlabel negatif akan divisualisasikan menggunakan *wordcloud*, serta akan dicari asosiasi kata dari ulasan yang ada. Pencarian asosiasi kata ini berguna untuk mendapatkan informasi terkait hal apa yang sebenarnya pelanggan keluhkan terhadap Grobbak Korea Pandeyan, serta mengerucutkan informasi yang telah didapat dari tahapan-tahapan sebelumnya.

Dari ulasan pelanggan yang ada, terdapat beberapa kata yang sering muncul dalam ulasan yang diunggah. Dari kata yang sering muncul tersebut, dapat dilakukan ekstraksi informasi dengan visualisasi frekuensi kata dalam ulasan pelanggan Grobbak Korea Pandeyan. Berdasarkan hasil uji klasifikasi pada tahapan sebelumnya menunjukkan bahwa skenario uji ke-2, yang menggunakan *test split* sebesar 0.2 menjadi skenario yang paling optimal. Skenario tersebut dikatakan optimal dikarenakan memiliki metriks hasil uji dengan akurasi tertinggi dari ketiga skenario uji yang dilakukan. Sehingga, pada tahap ini ulasan negatif hasil klasifikasi skenario uji ke-2 akan digunakan sebagai *input* dari visualisasi *frequent words* dan *association rule*. Berikut merupakan visualisasi ”*frequent words*” dari ulasan pelanggan Grobbak Korea, yang ditampilkan dalam *wordcloud*:



Gambar 4. 5 *Wordcloud* ulasan negatif pelanggan

Wordcloud merupakan salah satu alat yang digunaka dalam visusalisasi kata. Semakin besar ukuran kata yang ditampilkan dalam *wordcloud* diatas, maka semakin tinggi frekuensi kemunculan kata tersebut dalam ulasan pelanggan. Dan sebaliknya jika semakin kecil sebuah kata ditampilkan, maka semakin jarang sebuah kata tersebut muncul didalam ulasan yang pelanggan tuliskan. Dari *wordcloud* tersebut, kata-kata yang paling sering muncul dapat divisualisasikan kembali dengan lebih jelas, sehingga dapat diketahui seberapa sering sebuah kata muncul dalam ulasan pelanggan. Berikut merupakan 10 *most frequent words* dalam *wordcloud* ulasan negatif dari pelanggan Grobbak korea:



Gambar 4. 6 Visualisasi data ulasan negatif pelanggan

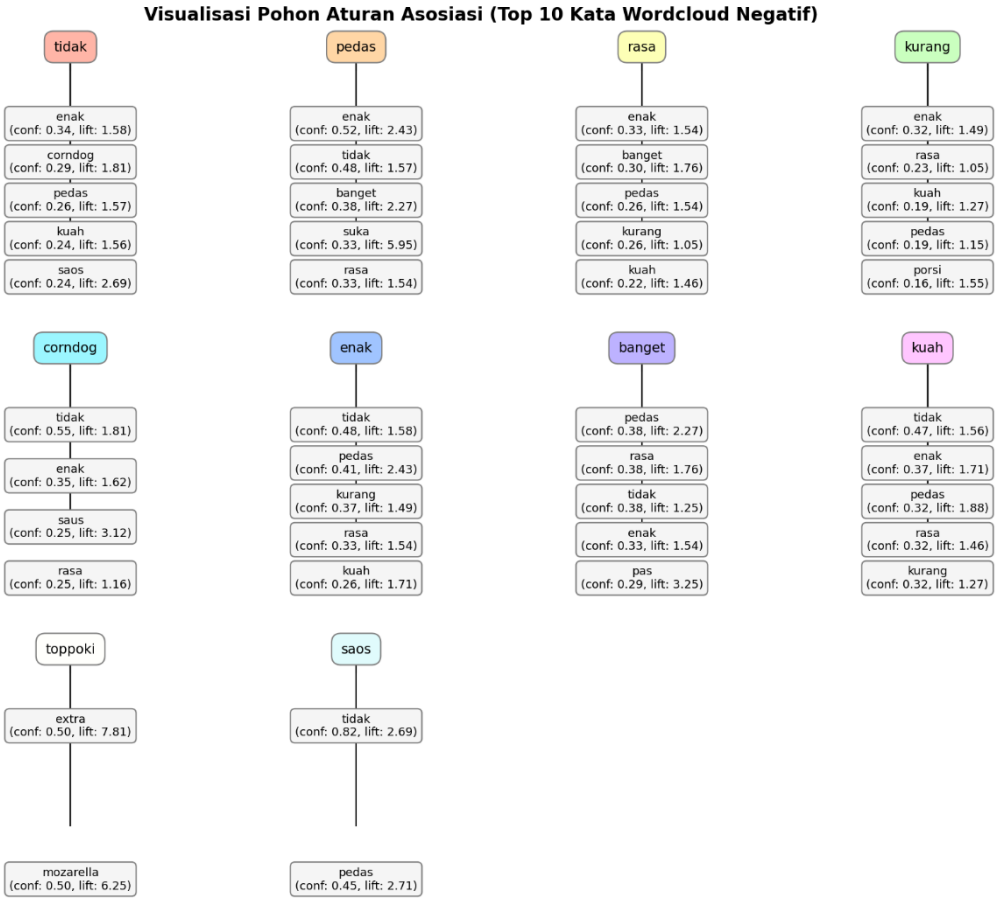
Berdasarkan gambar visualisasi diatas, dapat dilihat top 10 kata yang paling sering muncul dalam ulasan negatif pelanggan terhadap Grobbak Korea. Dari 141 ulasan yang memiliki label / kelas data negatif, dapat diketahui bahwa kata “tidak” muncul sebanyak 58 kali, diikuti dengan

kata “pedas” sebanyak 39 kali, lalu kata “rasa” sebanyak 34 kali, kata “kurang” sebanyak 33 kali, kata “corndog” sebanyak 32 kali, kata “enak” sebanyak 32 kali muncul, diikuti kata “banget” muncul 29 kali, kata “kuah” 21 kali muncul, kata “toppoki” 18 kali muncul, dan terakhir kata “sama” 15 kali muncul. Setelah diketahui kata-kata yang paling sering muncul dalam ulasan bersentimen negatif, maka selanjutnya dilakukan pencarian asosiasi kata dari dari top 10 kata yang sering muncul dari ulasan pelanggan tersebut.

Tabel 4. 18 Parameter pencarian *association rule*

Parameter	*Support*	*Confidence*
Besaran	0.04	0.2

Dalam pencarian aturan asosiatif, digunakan parameter *support* sebesar 0.04 dan *confidence* sebesar 0.2, seperti yang tertera pada tabel diatas. Setelah parameter digunakan dalam pencarian aturan asosiatif, terbentuk 180 *rule* dan diperoleh hasil visualisasi sebagai berikut:



Gambar 4. 7 Asosiasi kata sentimen negatif

Berdasarkan gambar 4.6, diperoleh asosiasi kata pada ulasan bersentimen negatif yang pelanggan berikan kepada Grobbak Korea. Kata-kata tersebut menggambarkan topik yang

paling sering diungkapkan dalam ulasan pelanggan, sehingga pada penelitian ini didapatkan 10 pohon asosiatif, sehingga dari 10 pohon asosiatif tersebut dapat disimpulkan relasi antar kata pada suatu pohon asosiatif, yang dapat menjadi informasi mengenai keluhan pelanggan Grobbak Korea. Berikut merupakan analisis kesimpulan relasi antar kata pada pohon asosiatif ulasan negatif.

Pada gambar, kata “tidak” berasosiasi dengan kata "enak", "corndog", "pedas", "kuah", dan "saos". Sehingga dapat diartikan, kemungkinan keluhan pelanggan adalah saus / kuah pada produk corndog tidak enak dikarenakan rasanya yang terlalu pedas atau sebaliknya yakni saus / kuah kurang pedas.

Lalu kata “pedas” berasosiasi dengan kata "enak", "tidak", "banget", "suka", dan "rasa". Hal ini menunjukkan kemungkinan keluhan pada sajian yang memiliki rasa pedas berlebih sehingga pelanggan tidak menyukai sajian.

Selanjutnya yakni kata “rasa” yang diasosiasikan dengan kata "enak", "banget", "pedas", "kurang", dan "kuah". Kemungkinan, pelanggan banyak menyoroti rasa makanan, terutama jika rasanya kurang enak, kurang pedas / terlalu pedas, atau kuahnya tidak sesuai.

Lalu pada kata “kurang”, berasosiasi dengan kata "enak", "rasa", "kuah", dan "porsi". Kemungkinan, pelanggan banyak mengeluhkan rasa kuah yang kurang enak, atau porsi hidangan yang dirasa kurang banyak dalam paket penjualan.

Selanjutnya pada kata “corndog” yang berasosiasi dengan kata "tidak", "enak", "saos", dan "rasa". Kemungkinan, konsumen mengeluhkan rasa saus pada produk corndog yang tidak enak. Ataupun produk corndog yang tidak enak.

Lalu pada kata “enak”, kata ini berasosiasi dengan kata "tidak", "pedas", "kurang", "rasa", dan "kuah". Kemungkinan, pelanggan mengeluhkan bahwa kuah yang diberikan pada paket penjualan dirasa tidak / kurang pedas. Atau, pelanggan juga mungkin mengeluhkan bahwa rasa kuah yang diberikan pada paket penjualan tidak / kurang enak.

Lalu pada kata “banget”, dimana kata tersebut berasosiasi dengan kata "pedas", "rasa", "tidak", "enak", dan "pas". Kemungkinan, pelanggan mengeluhkan bahwa rasa pedas pada suatu menu itu berlebihan sehingga tidak terasa enak, dimana hal ini digambarkan dengan kata “banget”.

Lalu pada kata “kuah” berasosiasi dengan kata "tidak", "enak", "pedas", "rasa", dan "kurang". Kemungkinan, pelanggan mengeluhkan bahwa kuah yang diberikan pada paket

penjualan dirasa tidak / kurang pedas. Atau, pelanggan juga mungkin mengeluhkan bahwa rasa kuah yang diberikan pada paket penjualan tidak / kurang enak.

Selanjutnya pada kata “topokki”, dimana kata tersebut berasosiasi dengan kata “extra” dan “mozarella”. Dimana pada aturan ini, kurang bisa diartikan apa hubungan asosiatif dari kata yang mengikuti kata “extra”.

Lalu yang terakhir yakni kata “saos”, dimana kata ini berasosiasi dengan kata "tidak" dan "pedas". Pada aturan ini, kurang bisa diartikan apa hubungan asosiatif dari kata yang mengikuti kata “saos”.

BAB V

PEMBAHASAN

5.1 Gambaran umum pengolahan data

Berdasarkan data ulasan terhadap Grobbak Korea yang diambil melalui *platform Gofood* dengan data terbaru per 20 Mei 2025, didapatkan 226 ulasan. Adapun setelah dilakukan pembobotan sentimen menggunakan kamus *lexicon*, didapatkan sebaran sentimen dengan 85 ulasan bersentimen positif, 80 ulasan bersentimen netral, dan 61 ulasan bersentimen negatif. Namun dalam penelitian ini, hanya akan digunakan kelas / label data positif dan negatif, sehingga kelas data netral diproses kembali melalui tahapan perhitungan *polarity.* Perhitungan *polarity* ini menghitung total kandungan makna suatu kata pada sebuah ulasan. Dimana, ulasan yang mendapatkan skor *polarity* positif (>0) akan masuk kedalam kelas positif, sebaliknya jika ulasan mendapatkan skor *polarity* negatif (<0) akan masuk kedalam kelas negatif.

Dengan perhitungan kembali kelas netral, maka penggunaan data ulasan secara keseluruhan dapat dimaksimalkan. Namun, jika suatu ulasan kelas netral tetap memiliki skor *polarity* 0 setelah dilakukannya perhitungan kembali, maka ulasan netral tersebut tidak digunakan dalam penelitian ini. Setelah perhitungan kembali kelas netral, maka didapatkan 85 ulasan berlabel positif, dan 141 ulasan berlabel negatif, sehingga tidak ada kelas data netral yang dibuang. Alasan dari perhitungan kembali kelas data netral ini adalah memaksimalkan *dataset* yang ada, mengingat jumlah data yang didapatkan sangat sedikit. Sehingga dengan dilakukannya perhitungan *polarity* tersebut, diharapkan metriks hasil klasifikasi memiliki performa yang lebih baik. Hal ini didukung dengan penelitian yang dilakukan oleh khan et al., (2025) yang meneliti 8 ukuran sampel yang bervariasi antara 50 dan 5000 dan hubungannya terhadap metriks model klasifikasi, penelitian menunjukkan bahwa jumlah data pelatihan merupakan salah satu faktor yang memiliki dampak signifikan dalam menentukan akurasi, presisi, *recall,* dan *f1 score* pada sebuah model klasifikasi.

5.2 Hasil klasifikasi menggunakan *Naive bayes classifier*

Proses uji klasifikasi diawali dengan pembagian *dataset* menjadi *dataset* uji dan *dataset* latih. Pembagian *dataset* dalam penelitian ini menghasilkan 3 skenario uji dengan persentase

perbandingan yang berbeda, dengan masing-masing skenario dilakukan uji sebanyak 1 kali. Berikut merupakan hasil percobaan uji ketiga skenario tersebut:

Tabel 5. 1 Metriks hasil uji klasifikasi

Output	Skenario uji		
	1(90:10)	2(80:20)	3(70:30)
Accuracy	0.783	0.848	0.794
Precision	0.814	0.849	0.800
Recall	0.783	0.848	0.794

Pada skenario uji 1 dengan perbandingan *dataset training & testing* 90:10 diketahui didapatkan nilai rata-rata akurasi sebesar 78.3%. Lalu pada skenario uji 2 dengan perbandingan *dataset training & testing* 80:20 diketahui didapatkan nilai rata-rata akurasi sebesar 84.8%. Lalu pada skenario uji 3 dengan perbandingan *dataset training & testing* 70:30 diketahui didapatkan nilai rata-rata akurasi sebesar 79.4%. Dari hasil yang didapat, skenario uji 2 dengan perbandingan *dataset training & testing* sebesar 80:20 mendapatkan nilai akurasi tertinggi, yakni sebesar 84.8%. Hasil ini selaras dengan pembangunan model dari penelitian Muslim et al., (2024), dimana pembangunan model klasifikasi dengan rasio *split* 80/20 memiliki metriks hasil klasifikasi terbaik.

Lalu untuk penggunaan algoritma *Naive bayes* pada penelitian ini, didasarkan pada penelitian dari khan et al., (2025), dimana algoritma *Naive bayes* menunjukkan performa metriks yang tergolong impresif, baik dalam ukuran data kecil maupun besar. Sehingga dari jumlah data ulasan yang didapatkan, yaitu sejumlah 226 data yang tergolong dalam ukuran data kecil, algoritma *Naive bayes* dapat bekerja dengan baik terutama pada metriks *recall* dan *f1 score.* Lalu didukung pula dengan penelitian yang dilakukan oleh Simarmata et al., (2024) yang menyatakan bahwa performa algoritma *Naive bayes* tidak secara signifikan dipengaruhi oleh ukuran sampel kecil maupun besar.

5.3 Visualisasi kata dan *association rule*

Setelah tahapan klasifikasi dilakukan, maka selanjutnya dilakukan proses visualisasi kata. Pada penelitian ini, visualisasi kata dilakukan pada ulasan bersentimen negatif agar nantinya dapat digunakan sebagai *input* dari tahapan *association rule* untuk pencarian asosiasi kata negatif dari ulasan pelanggan yang ada. Dari jumlah ulasan negatif sejumlah 141 ulasan, diperoleh beberapa

kata yang sering muncul diantaranya adalah kata kata “tidak” muncul sebanyak 58 kali, diikuti dengan kata “pedas” sebanyak 39 kali, lalu kata rasa sebanyak 34 kali, kata “kurang” sebanyak 33 kali, kata “corndog” sebanyak 32 kali, kata “enak” sebanyak 32 kali muncul, diikuti kata “banget” muncul 29 kali, kata “kuah” 21 kali muncul, kata “toppoki” 18 kali muncul, dan terakhir kata “sama” 15 kali muncul. Dari banyaknya kemunculan, dan urgensi terhadap identifikasi pola-pola yang terkait dengan kemunculan kata-kata tersebut, dengan hal ini menjadi dasar dari perlunya dilakukan pencarian asosiasi kata.

Association rule dalam penelitian ini berperan untuk menerjemahkan kesimpukan dari *frequent words* yang masih rancu. Hal ini ditinjau lebih detail dengan menemukan asosiasi dari kata-kata yang sering muncul bersamaan, sehingga analisis strategi yang akan dilakukan dapat lebih tepat sasaran terhadap substansi yang ada. Dari asosiasi kata yang ditemukan pada ulasan negatif yaitu kata “tidak”, “pedas”, “rasa”, “kurang”, “corndog”, “enak”, “banget”, “kuah”, “topokki”. Dari kata-kata tersebut, didapatkan informasi dengan kemungkinan keluhan sebagai berikut:

1. Pengguna menilai saus / kuah pada produk corndog tidak enak dikarenakan rasanya yang terlalu pedas atau sebaliknya yakni saus / kuah kurang pedas.
2. Keluhan pada sajian yang memiliki rasa pedas berlebih sehingga pelanggan tidak menyukai sajian.
3. Pelanggan juga banyak menyoroti rasa makanan, terutama jika rasanya kurang enak, kurang pedas / terlalu pedas, atau kuahnya tidak sesuai.
4. Pelanggan juga mengeluhkan rasa kuah yang kurang enak, atau porsi hidangan yang dirasa kurang banyak dalam paket penjualan.
5. Pelanggan mengeluhkan rasa saus pada produk corndog yang tidak enak. Ataupun produk corndog yang tidak enak.
6. Pelanggan mengeluhkan bahwa kuah yang diberikan pada paket penjualan dirasa tidak / kurang pedas. Atau, pelanggan juga mungkin mengeluhkan bahwa rasa kuah yang diberikan pada paket penjualan tidak / kurang enak.
7. Pelanggan mengeluhkan bahwa rasa pedas pada suatu menu itu berlebihan sehingga tidak terasa enak, dimana hal ini digambarkan dengan kata “banget”.
8. Pelanggan mengeluhkan bahwa kuah yang diberikan pada paket penjualan dirasa tidak / kurang pedas. Atau, pelanggan juga mungkin mengeluhkan bahwa rasa kuah yang diberikan pada paket penjualan tidak / kurang enak.

Sehingga dari kedelapan *rule* tersebut secara garis besar, pelanggan mengeluhkan kualitas produk terutama dalam segi rasa kuah atau saus, tingkat kepedasan, rasa *corndog*, dan porsi sajian. Keluhan pelanggan diatas merupakan hasil ekstraksi informasi dari aturan / *rule* yang terbentuk dari tahapan *association rule*. Dari hasil ekstraksi tersebut, sebagian besar keluhan pelanggan berkaitan dengan kuah, corndog, saus, serta tingkat kepedasan dari menu-menu yang ditawarkan oleh Grobbak Korea *outlet* Pandeyan. Selanjutnya, temuan keluhan ini akan dipetakan kedalam konsep *marketing mix* dan disusun solusi penyelesaiannya.

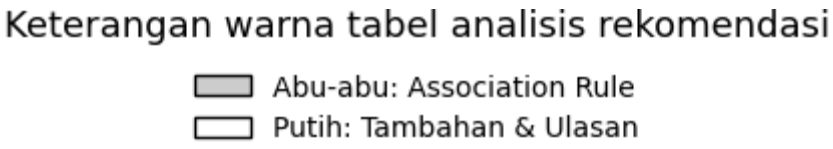
5.4 Analisis strategi berdasarkan asosiasi kata dan observasi

Dari permasalahan diatas, dapat ditinjau lebih lanjut dengan pengelompokan permasalahan kedalam 6p (*Product, Price, Place, Promotion, People, and Process*) yang terdapat dalam konsep *marketing mix.* Penggunaan kerangka 6p pada penelitian ini bertujuan untuk mempermudah pengelompokan permasalahan berdasarkan aspek bisnis yang ada. Untuk itu, 8 permasalahan yang telah didapatkan dari tahapan *association rule* yang menjadi *input* dalam penyelesaian permasalahan yang akan ditindak lanjut, dapat dikelompokkan sebagai berikut:

Tabel 5. 2 Kode permasalahan

Permasalahan	Keluhan	Kode permasalahan
Saus	Nomor 1 & 5	A
Kuah	Nomor 3,4,6, dan 8	B
Menu sajian	Nomor 3,4, dan 5	C
Tingkat kepedasan	Nomor 1,2,3,6,7,8	D

Dari tabel diatas, garis besar permasalahan telah dikelompokkan menjadi kode permasalahan yang nantinya akan digunakan dalam analisis solusi permasalahan. Sehingga dengan adanya kode permasalahan yang sudah dibuat, pembaca dapat memahami uraian analisis penyelesaian permasalahan dengan lebih baik. Berikut merupakan keterangan penjelas tabel analisis rekomendasi:



Gambar 5. 1 Keterangan warna pada tabel rekomendasi

Sesuai dengan keterangan pada gambar diatas, tabel dengan warna abu-abu berisi analisis rekomendasi dari permasalahan yang didapatkan dari hasil *association rule*. Lalu tabel

berwarna putih berisi analisis rekomendasi yang berasal dari ulasan lainnya, dan merupakan tambahan pelengkap. Berikut merupakan pemetaan permasalahan, serta solusi yang dapat ditawarkan dalam penyelesaian permasalahan tersebut:

Tabel 5. 3 Analisis pemetaan permasalahan & solusi

Aspek 6P	Permasalahan	Solusi	Rincian solusi
Product	A, B, C, D	Standarisasi bahan baku, takaran	• Gunakan sendok takar untuk setiap level kepedasan sajian, agar penyajian konsisten • Standarisasi bahan baku(merk, karakteristik) agar *output* produk dapat dijaga konsistensinya
People	Tidak ada supervisi *onsite*	Supervisi mandiri	• Gunakan *checklist* kontrol harian (bahan lengkap, peralatan siap, kebersihan terjaga) • Pemberian insentif tambahan jika *checklist* kontrol terpenuhi dengan sempurna selama suatu periode waktu, untuk menjaga motivasi karyawan
Process	C	Standarisasi waktu pengolahan, kelola bahan baku	• Gunakan sistem *FIFO* pada setiap bahan baku, untuk menjaga kesegaran bahan baku • Gunakan *timer* untuk kontrol waktu masak
Price	-	-	-

Aspek 6P	Permasalahan	Solusi	Rincian solusi
Place	-	Menjaga kebersihan area kerja secara rutin	• Bersihkan area *stall* & dapur setiap toko buka dan tutup • Foto kondisi toko setelah tutup untuk dokumentasi kebersihan toko
Promotion	-	-	-

Berdasarkan tabel solusi yang ditawarkan melalui kerangka *marketing mix 6P*, langkah strategis yang bersifat konseptual perlu diturunkan menjadi langkah operasional yang terukur. Dalam hal ini, penyusunan *Standard Operational Procedure* (*SOP*) menjadi kerangka untuk memastikan setiap solusi diterapkan dengan baik dan konsisten. SOP tidak hanya berfungsi sebagai panduan kerja, tetapi juga sebagai kontrol kualitas untuk menjaga standar produk, pelayanan, dan proses bisnis secara keseluruhan.

Adapun penyusunan dokumen SOP pada penelitian ini didasari dengan mempertimbangkan subjektivitas penerimaan pelanggan terhadap produk dari Grobbak Korea, sehingga hal tersebut harus distandarisasi guna mengakomodasi subjektivitas tersebut. Lalu untuk poin-poin dalam pengembangan dokumen *standard operational procedure* pada penelitian ini mempertimbangkan aspek pembuatan produk, manajemen & pengadaan bahan baku, kebersihan & sanitasi, serta *self supervising,* dengan *highlight* yakni aspek pembuatan produk. Hasil dari pembuatan dokumen SOP sendiri ditampilkan pada bagian lampiran laporan ini.

BAB VI

PENUTUP

6.1 Kesimpulan

1. Setelah dilakukannya pengumpulan sejumlah 226 ulasan pelanggan dari Grobbak Korea, serta dilakukannya proses pelabelan data, secara umum ulasan pelanggan memiliki sebaran sentimen dengan 85 atau 37.6% ulasan bersentimen positif, 80 atau 35.4% ulasan bersentimen netral, dan 61 atau 27.0% ulasan bersentimen negatif. Namun setelah dilakukannya reduksi kelas sentimen netral untuk menghindari kerancuan hasil, maka dihasilkan sebaran sentimen ulasan dengan 85 atau 37.6% ulasan berlabel positif, dan 141 atau 62.4% ulasan berlabel negatif. Dengan banyaknya ulasan pelanggan yang bersentimen negatif (62.4%), maka analisis perbaikan berdasarkan keluhan pelanggan diperlukan.
2. Dari kelas / label ulasan yang telah direduksi, maka selanjutnya data ulasan melalui tahapan klasifikasi. Pembangunan model klasifikasi dalam penelitian ini menggunakan algoritma *Naive bayes classifier*, dengan 3 skenario uji yakni *dataset splitting* 90:10, 80:20, dan 70:30. Dari ketiga skenario uji tersebut, skenario kedua dengan perbandingan *dataset splitting* 80:20 memiliki hasil uji yang terbaik diantara ketiganya, dengan *accuracy* sebesar 84.8%, *precision* sebesar 84.9%, dan *recall* sebesar 84.8%.
3. Dari tahapan *association rule* yang telah dilakukan, dengan parameter uji yakni *support* 0.04, dan *confidence* 0.2, terbentuk 180 *rule.* Dari *rule* tersebut, dapat disortir berdasarkan top 10 *frequent words* dalam ulasan negatif, dimana asosiasi dari masing-masing kata pada *frequent words* divisualisasikan dengan *tree*. Dari *tree* yang terbentuk, didapatkan kesimpulan bahwa secara garis besar, pelanggan mengeluhkan terkait kualitas produk. Terkhusus pada rasa kuah atau saus, tingkat kepedasan(terlalu pedas / kurang pedas), rasa *corndog*, dan porsi sajian.
4. Permasalahan yang terjadi pada Grobbak Korea *outlet* Pandeyan mayoritas terjadi pada aspek produk, diikuti dengan aspek *process,* dan *people* jika ditinjau dari konsep *marketing mix 6P*. Dari permasalahan tersebut, diusulkan penambahan penyelesaian permasalahan dengan pembuatan prosedur operasional yang mencakup aspek

pembuatan produk, manajemen & pengadaan bahan baku, serta kebersihan & sanitasi, dan *self supervising* guna perbaikan berkelanjutan. Dokumen SOP sendiri dicantumkan pada bagian lampiran laporan ini.

6.2 Saran

1. Bagi pihak Grobbak Korea

   Diharapkan bagi pihak Grobbak Korea dapat meninjau usulan-usulan yang penulis berikan, yang didasarkan pada ulasan pelanggan dan observasi. Penulis harap usulan tersebut dapat dijadikan masukan guna meningkatkan pelayanan Grobbak Korea kearah yang lebih baik.

2. Bagi peneliti selanjutnya

   Diharapkan bagi peneliti selanjutnya dapat mengumpulkan *dataset* dengan ukuran yang lebih besar, sehingga metode-metode *machine learning* selain *naïve bayes* dapat digunakan dan berjalan secara optimal dalam pembuatan model klasifikasi, serta metriks hasil pembuatan model klasifikasi dapat lebih optimal. Lalu dalam penelitian ini terdapat kendala dalam melakukan *scrapping* data, dimana data ulasan dari *platform GrabFood*, dan *ShopeeFood* belum bisa penulis *scrape.* Sehingga penulis berharap peneliti selanjutnya dapat mengambil / *scrape dataset* dari kedua *platform* tersebut, untuk memperbesar ukuran *dataset*.

DAFTAR PUSTAKA

Aleqabie, H. J., Sfoq, M. S., Albeer, R. A., & Abd, E. H. (2024). A Review Of Text Mining Techniques: Trends, and Applications In Various Domains. In *Iraqi Journal for Computer Science and Mathematics* (Vol. 5, Issue 1, pp. 125–141). College of Education, Al-Iraqia University. https://doi.org/10.52866/ijcsm.2024.05.01.009

Andika Johan, R., Himilda, R., & Auliza, N. (2019). Penerapan Metode Association Rule Untuk Strategi Penjualan Menggunakan Algoritma Apriori. *Jurnal Teknik Informatika*, *2*, 2654–2617. https://doi.org/10.52046/j-tifa.v2i2.268

Anggina, S., Setiawan, N. Y., & Bachtiar, F. A. (2022). Analisis Ulasan Pelanggan Menggunakan Multinomial Naïve Bayes Classifier dengan Lexicon-Based dan *TF-IDF* Pada Formaggio Coffee and Resto. *Is The Best Accounting Information Systems and Information Technology Business Enterprise*, *7*(1), 76–90. https://doi.org/10.34010/aisthebest.v7i1.7072

Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. *American Journal of Epidemiology*, *188*(12), 2222–2239. https://doi.org/10.1093/aje/kwz189

Chowdhury, S., & Alzarrad, A. (2023). Applications of Text Mining in the Transportation Infrastructure Sector: A Review. In *Information (Switzerland)* (Vol. 14, Issue 4). MDPI. https://doi.org/10.3390/info14040201

Cios, K. J., & Kurgan, L. A. (2005). Trends in Data Mining and Knowledge Discovery. In Nikhil R. Pal & Lakhmi Jain (Eds.), *Advanced Techniques in Knowledge Discovery and Data Mining* (1st ed.). Springer London. https://doi.org/10.1007/1-84628-183-0

Cynthia, G., & Qoiriah, A. (2024). Klasifikasi Tingkat Kesiapan Kerja Mahasiswa Universitas Negeri Surabaya Menggunakan Algoritma C5.0. *Journal of Informatics and Computer Science*, *06*, 383–396. https://doi.org/10.26740/jinacs.v6n02.p383-396

G. Torres, E., Fonseca-Larios, R., & Ramírez-Portilla, A. (2024). *Innovating R&D Practices to Adapt Swiftly to Market Trends: Insights from the Food and Beverage Industry*. https://doi.org/10.5772/intechopen.1006672

Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques* (Vol. 12). Springer Science & Business Media. https://doi.org/10.1007/978-3-642-19721-5

Gunawan, L., Anggreainy, M. S., Wihan, L., Santy, Lesmana, G. Y., & Yusuf, S. (2022). Support vector machine based emotional analysis of restaurant reviews. *Procedia Computer Science*, *216*, 479–484. https://doi.org/10.1016/j.procs.2022.12.160

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)* (3rd ed.). Morgan Kaufmann Publishers. https://doi.org/10.1016/C2009-0-61819-5

Han, J., Pei, J., & Tong, H. (2024). *Data mining: concepts and techniques* (H. Jiawei, P. Jian, & T. Hanghang, Eds.; 4th ed.). Morgan kaufmann. https://doi.org/10.1016/C2013-0-18660-6

Handayani, F., Feddy, D., & Pribadi, S. (2015). Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110. *Jurnal Teknik Elektro*, *7*(1), 19–24.

Jakkula, V. (2006). Tutorial on support vector machine (svm). *School of EECS, Washington State University*, *37*(2.5), 3. https://course.khoury.northeastern.edu/cs5100f11/resources/jakkula.pdf

Kay, Judy., Thomas, Paul., & Trotman, Andrew. (2009). *ADCS 2009 : proceedings of the fourteenth Australasian Document Computing Symposium, 4 December 2009* (J. kay, P.

Thomas, & A. Trotman, Eds.). School of Information Technology, University of Sydney. http://es.csiro.au/adcs2009

Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing and Its Applications*, *13*(3), 144–168. https://doi.org/10.15849/ijasca.211128.11

Kwekha-Rashid, A. S., Abduljabbar, H. N., & Alhayani, B. (2023). Coronavirus disease (COVID-19) cases analysis using machine-learning applications. *Applied Nanoscience (Switzerland)*, *13*(3), 2013–2025. https://doi.org/10.1007/s13204-021-01868-7

Laghbi, Y. A., & Al Dhoayan, M. (2024). Examining how customers perceive community pharmacies based on Google maps reviews: Multivariable and sentiment analysis. *Exploratory Research in Clinical and Social Pharmacy*, *15*. https://doi.org/10.1016/j.rcsop.2024.100498

Langgeni, D. P., Abdurahman Baizal, Z. K., & Firdaus, Y. (2010). Clustering Artikel Berita Berbahasa Indonesia Menggunakan Unsupervised Feature Selection. *Seminar Nasional Informatika*.

Li, L., Yang, L., & Zeng, Y. (2021). Improving sentiment classification of restaurant reviews with attention-based bi-gru neural network. *Symmetry*, *13*(8). https://doi.org/10.3390/sym13081517

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Megawati, S., & Lawi, A. (2021). Pengembangan Sistem Teknologi Internet of Things Yang Perlu Dikembangkan Negara Indonesia. *Journal of Information Engineering and Educational Technology*, *5*, 19–26. https://doi.org/10.26740/jieet.v5n1.p19-26

Muslim, S. N. S., Nurdiyansyah, F., & Rahman, A. Y. (2024). Perbandingan Algoritma Naive Bayes Dan Knn Dalam Analisis Sentimen Ulasan Pengguna Aplikasi Capcut. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(3S1). https://doi.org/10.23960/jitet.v12i3s1.5156

Nguyen, Q. H., Ly, H. B., Ho, L. S., Al-Ansari, N., Van Le, H., Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering*, *2021*. https://doi.org/10.1155/2021/4832864

Nofandi, A., Setiawan, N. Y., & Brata, D. W. (2023). Analisis Sentimen Ulasan Pelanggan dengan Metode Support Vector Machine (SVM) untuk Peningkatan Kualitas Layanan pada Restoran Warung Wareg. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *7*(1), 458–466. http://j-ptiik.ub.ac.id

Nur Ikhsan, A., Subarkah, P., Dafa Iftinani, A., & Fadilah, A. N. (2025). *Analisis Sentimen Kenaikan PPN menggunakan Algoritma Naïve Bayes dan Support Vector Machine*. *Infotekmesin, 16*(01). https://doi.org/10.35970/infotekmesin.v16i1.2518

Ode, W., Ahmad, A., Semolowaru, J., 45, N., Surabaya, S., & Timur Indonesia, J. (2024). Sentimen Analisis Menggunakan Naïve Bayes Untuk Rekomendasi Restoran Wakatobi. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 4).

OLUWAJINMI, B., & LONGE, O. (2024). Continuous Process Improvement and Customer Satisfaction of Selected Food and Beverage Companies in South-West, Nigeria. *IIARD International Journal of Economics and Business Management*, *9*(10), 94–109. https://doi.org/10.56201/ijebm.v9.no10.2023.pg94.109

Patil, R. N., Singh, Y. P., Rawandale, S. A., & Singh, S. (2024). Improving Sentiment Classification on Restaurant Reviews Using Deep Learning Models. *Procedia Computer Science*, *235*, 3246–3256. https://doi.org/10.1016/j.procs.2024.04.307

Peretz, O., Koren, M., & Koren, O. (2024). Naive Bayes classifier – An ensemble procedure for recall and precision enrichment. *Engineering Applications of Artificial Intelligence*, *136*. https://doi.org/10.1016/j.engappai.2024.108972

Qi, C., Fourie, A., Chen, Q., & Zhang, Q. (2018). A strength prediction model using artificial intelligence for recycling waste tailings as cemented paste backfill. *Journal of Cleaner Production*, *183*, 566–578. https://doi.org/10.1016/j.jclepro.2018.02.154

Roidar khan, Shehzad khan, Naseemullah, Aasim Ullah, & Atif khan. (2025). Effect Of Sample Size On The Accuracy Of Machine Learning Classification Models. *Spectrum of Engineering Sciences*, *3*(7), 826–834. https://sesjournal.com/index.php/1/article/view/659

Rokhva, S., Teimourpour, B., & Babaei, R. (2025). Enhanced Sentiment Analysis of Iranian Restaurant Reviews Utilizing Sentiment Intensity Analyzer & Fuzzy Logic. *Food and Humanity*, 100658. https://doi.org/10.1016/j.foohum.2025.100658

Saleh, M. A. S., & AlShafeey, M. (2025). Examining the synergies between industry 4.0 and sustainability dimensions using text mining, sentiment analysis, and association rules. *Sustainable Futures*, *9*. https://doi.org/10.1016/j.sftr.2024.100423

Saputra, A. M. A., Kharisma, L. P. I., Rizal, A. A., Burhan, M. I., & Purnawati, N. W. (2023). *Teknologi Informasi: Peranan TI dalam berbagai bidang* (Sepriano & Y. Agusdi, Eds.). PT. Sonpedia Publishing Indonesia.

Selimovic, M., Almisreb, A. A., & Amanzholova, S. (2024). The Role of Sentiment Analysis in Brand Management and Marketing: A Comparative Study. *Procedia Computer Science*, *251*, 579–584. https://doi.org/10.1016/j.procs.2024.11.152

Sifa Amalia, B., Umaidah, Y., Mayasari, R. (2021). *Analisis Sentimen Review Pelanggan Restoran Menggunakan Algoritma Support Vector Machine Dan K-Nearest Neighbor*. *Jurnal Sains, Teknologi dan Industri*, *19*(1), 28–34.

Simarmata, J. E., Weber, G.-W., & Chrisinta, D. (2024). Performance Evaluation of Classification Methods on Big Data: Decision Trees, Naive Bayes, K-Nearest Neighbors, and Support Vector Machines. *Jurnal Matematika, Statistika Dan Komputasi*, *20*(3), 623–638. https://doi.org/10.20956/j.v20i3.32970

Singgalen, Y. A. (2022). Sentiment Analysis on Customer Perception towards Products and Services of Restaurant in Labuan Bajo. *Journal of Information Systems and Informatics*, *4*(3). http://journal-isi.org/index.php/isi. https://doi.org/10.51519/journalisi.v4i3.276

Somvanshi, M., Chavan, P., Tambade, S., & Shinde, S. V. (2016). A review of machine learning techniques using decision tree and support vector machine. *2016 International Conference on Computing Communication Control and Automation (ICCUBEA)*, 1–7. https://doi.org/10.1109/ICCUBEA.2016.7860040

Sukwadi, R., Magdalena Silitonga, R., Maria Magdalena Wahyuni, Octavian, F., Jou, Y.-T., & Thu, N. T. B. (2025). Peningkatan Kualitas Layanan Jaringan Restoran Cepat Saji Indonesia: Analisis Sentimen dan Emosi Berbasis Aspek. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *12*(2), 359–368. https://doi.org/10.25126/jtiik.2025129416

Suryana, A., Purnamasari, A. I., & Ali, I. (2024). Mengoptimalkan Kepuasan Pengguna: Analisis Sentimen Review Aplikasi Grab Di Indonesia. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 3).

Taboada, M. (2016). Sentiment Analysis: An Overview from Linguistics. In *Annual Review of Linguistics* (Vol. 2, pp. 325–347). Annual Reviews Inc. https://doi.org/10.1146/annurev-linguistics-011415-040518

Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Informatika Dan Sistem Informasi*, *9*(2), 785–795. https://doi.org/10.35957/jatisi.v9i2.1835

Tarigan, M. Y. B., Simanjuntak, M. A. S., Sitanggang, M. A. L., & Sitanggang, D. (2023). Implementasi Data Mining Pengelompokan Tingkat Kepuasan Pelanggan Pada Pt. Bisa Group Menggunakan Fuzzy C-Means. *Jurnal Tekinkom (Teknik Informasi Dan Komputer)*, *6*(2), 688–695.

Vijay Gaikwad, S. (2014). Text Mining Methods and Techniques. *International Journal of Computer Applications*, *85*(17), 975–8887. https://doi.org/10.5120/14937-3507

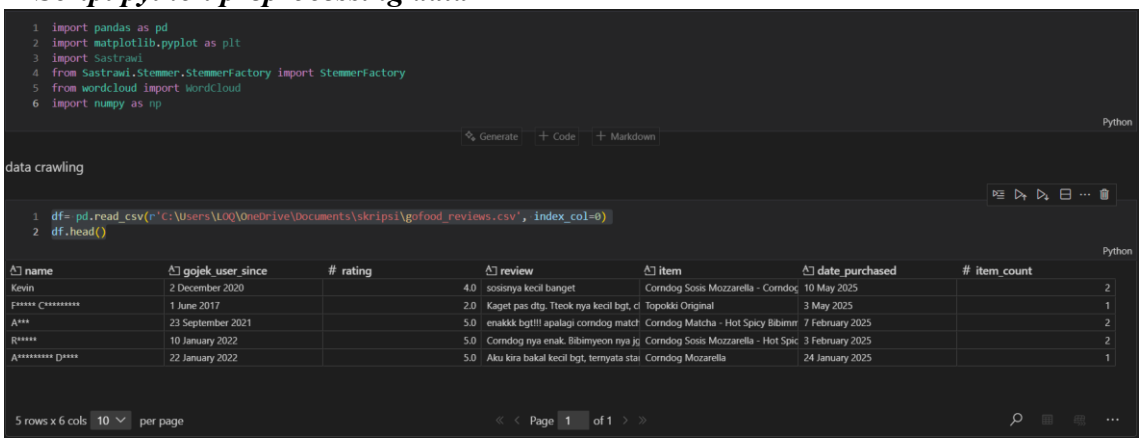
Yoseva Simanjuntak, A., & Septian Salomo Simatupang, I. (2022). Implementasi Data Mining Menggunakan Metode Naïve Bayes Classifier Untuk Data Kenaikan Pangkat Dinas Ketenagakerjaan Kota Medan. *Journal of Science and Social Research*, *1*, 85–91. https://doi.org/10.54314/jssr.v5i1.804

LAMPIRAN

A-*Dataset*


```
name,gojek_user_since,rating,review,item,date_purchased,item_count
Kevin,2 December 2020,4.0,sosisnya kecil banget,Corndog Sosis Mozzarella - Corndog Mozarella,10 May 2025,2
F***** C*********,1 June 2017,2.0,"Kaget pas dtg. Tteok nya kecil bgt, chikuwa nya jg tipis bgt potongannya. Saosnya super encer dan terlalu banyak, ga sebanding sm isiannya. Saos tteopokki yg bener harusnya kental pedas. Kalau ini malah kyk kuah sup b
A***,23 September 2021,5.0,enakkk bgt!!! apalagi corndog matchanya the best!!,Corndog Matcha - Hot Spicy Bibimmyeon,7 February 2025,2
R*****,10 January 2022,5.0,Corndog nya enak. Bibimyeon nya jg enak pedes chikuwa sama odengnya banyak,Corndog Sosis Mozzarella - Hot Spicy Bibimmyeon,3 February 2025,2
A********* D****,22 January 2022,5.0,"Aku kira bakal kecil bgt, ternyata standarnya corndoglah",Corndog Mozarella,24 January 2025,1
dian tri,17 November 2019,5.0,ter enak di jogja,Pakokki - Corndog Sosis Mozzarella,15 January 2025,2
Vinsa Athalia,24 June 2018,3.0,saus topokki & odeng terlalu cair,Corndog Sosis Mozzarella - Odeng Saus Korea - Topokki Original,4 January 2025,3
J**** i*****,6 February 2022,4.0,luar nya sedikit gosong sementara dalam nya sedikit kurang matang. diperbaiki lagi yaðŸ',Corndog Choco - Corndog Sosis Mozzarella,17 December 2024,2
S****,5 March 2016,4.0,"Corndog nya okelah 3,95/5, tapi tteokbokki nyaâ€¦â€¦ mending jualan corndog aja sih klo kata ku",Corndog Choco - Topokki Original,1 November 2024,2
Dahlia,7 July 2022,4.0,lumayan kalo mnurutku,Topokki Extra Odeng - Odeng Fish Soup - Corndog Choco,17 October 2024,3
mila,16 October 2016,5.0,saus terpisah selalu ada disetiap toko ya Chef,Corndog Mozarella,25 October 2024,1
F***** H*****,13 October 2019,4.0,"pertama kali coba order disini, enak banget bakal repeat order si tapi gatau kenapa pas mandu aku dateng agak ancur isiannya keluar-keluar gitu",Corndog Sosis Mozzarella - Topokki Extra Odeng Mozzarella - Mandu Fis
F***** H*****,13 October 2019,4.0,"ini orderan kedua akuuu disinii, sofar makanannnyaa enakk dan nagih tapi maaf banget kak dari orderan pertama aku ngerasa kurang di mandu, untuk kali ini mandunya keasinan mungkin bisa dikoreksi lagi buat rasanya b
M***** K*****,22 April 2021,3.0,"Corndognya rada gosong jadi sedikit pait sosisnya seperempat dari ukuran corndog, pakokkinya juga encer kayak kuah seblak",Pakokki - Corndog Sosis Mozzarella,21 September 2024,2
S***,22 April 2020,1.0,"Untuk mbk/mas yang jualan tolong di teliti lagi kalo ada orderan,,kasihan mas gojek nya harus balik lagi ngambil makanannya,,saya lihat di ulasan ternyata bnyk juga yg bilang kalo pesanan sering kurang tepat,,apa memang di sengaja,,
clarintaa,11 May 2021,4.0,enakk tp kurg nyoklat hehe,Corndog Choco,14 July 2024,1
sasa,9 June 2021,5.0,"enakkk, ini keduakalinya aku nyobain makanan korea lagi karna sebelumnya aku kurang suka... tapi yang ini enak banget. apa lagi kuah/sausnya! makasih udah masak ini ðŸ˜Š,Topokki Original - Rabokki Extra Mozzarella,2
N****,2 July 2017,4.0,"buat rasa si oke bgtt aku suka,tp di mienya terlalu bnya minyaakkkk sbnere uda enak bgttt sm kulit corndognya aga krg mateng dikitt sj tp ttep mantappppðŸ˜™,â€¢I,ðŸ'ðŸ¾ðŸ'ðŸ¾ðŸ'ðŸ¾ðŸ'ðŸ¾ðŸ«°ðŸ',Corndog Choco - Hot Spicy Bibimn
Sri Rezeki Amalia,27 January 2022,2.0,pesan extra chikuwa 8 potong dtg malah gaada,Corndog Choco - Odeng Fish Soup,27 June 2024,2
U***,12 March 2022,3.0,untuk rasa lebih ke pedas manis dan kuahnya terlalu encer padahal bisa untuk lebih kental,Bukokki - Chikuwa Saus Korea,3 June 2024,2
R****,24 April 2024,5.0,"enak bangtttt coklatnya ga pelit dan ga terlalu manis jadi ga bikin eneg
wajib repeat order lagi sih iniiiii",Corndog Sosis Mozzarella - Corndog Choco,23 May 2024,2
T****,6 January 2017,5.0,enak bgtttt woiiiiiiejsosjaelaksksnkaan,Corndog Choco,16 April 2024,1
```


B-*Script python preprocessing data*



```python
import pandas as pd
import matplotlib.pyplot as plt
import Sastrawi
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from wordcloud import WordCloud
import numpy as np
```

Python

Generate + Code + Markdown

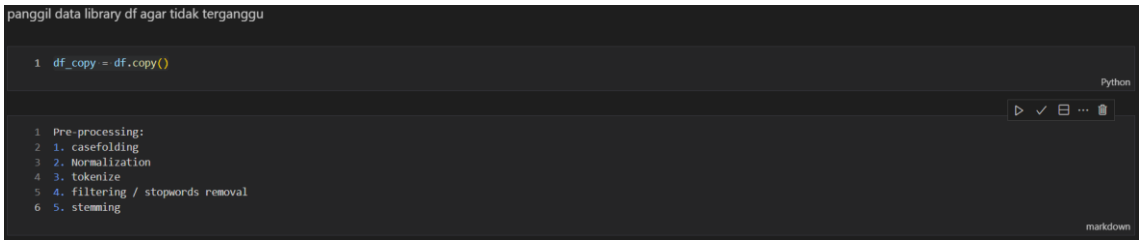
data crawling

```python
df= pd.read_csv(r'C:\Users\LOQ\OneDrive\Documents\skripsi\gofood_reviews.csv', index_col=0)
df.head()
```

Python

name	gojek_user_since	# rating	review	item	date_purchased	# item_count
Kevin	2 December 2020	4.0	sosisnya kecil banget	Corndog Sosis Mozzarella - Corndoc	10 May 2025	2
F***** C*********	1 June 2017	2.0	Kaget pas dtg. Tteok nya kecil bgt, cl	Topokki Original	3 May 2025	1
A***	23 September 2021	5.0	enakkk bgt!!! apalagi corndog matcl	Corndog Matcha - Hot Spicy Bibimm	7 February 2025	2
R*****	10 January 2022	5.0	Corndog nya enak. Bibimyeon nya jg	Corndog Sosis Mozzarella - Hot Spic	3 February 2025	2
A********* D****	22 January 2022	5.0	Aku kira bakal kecil bgt, ternyata stai	Corndog Mozarella	24 January 2025	1

5 rows x 6 cols 10 per page Page 1 of 1



panggil data library df agar tidak terganggu

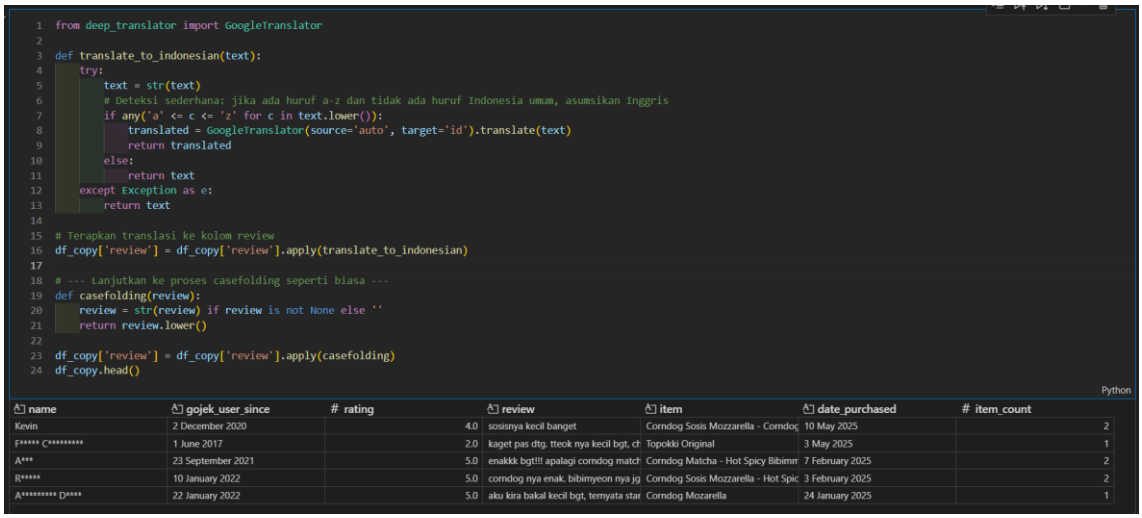
```python
df_copy = df.copy()
```

Python

```
Pre-processing:
1. casefolding
2. Normalization
3. tokenize
4. filtering / stopwords removal
5. stemming
```

markdown

- ***Casefolding***



```python
from deep_translator import GoogleTranslator

def translate_to_indonesian(text):
    try:
        text = str(text)
        # Deteksi sederhana: jika ada huruf a-z dan tidak ada huruf Indonesia umum, asumsikan Inggris
        if any('a' <= c <= 'z' for c in text.lower()):
            translated = GoogleTranslator(source='auto', target='id').translate(text)
            return translated
        else:
            return text
    except Exception as e:
        return text

# Terapkan translasi ke kolom review
df_copy['review'] = df_copy['review'].apply(translate_to_indonesian)

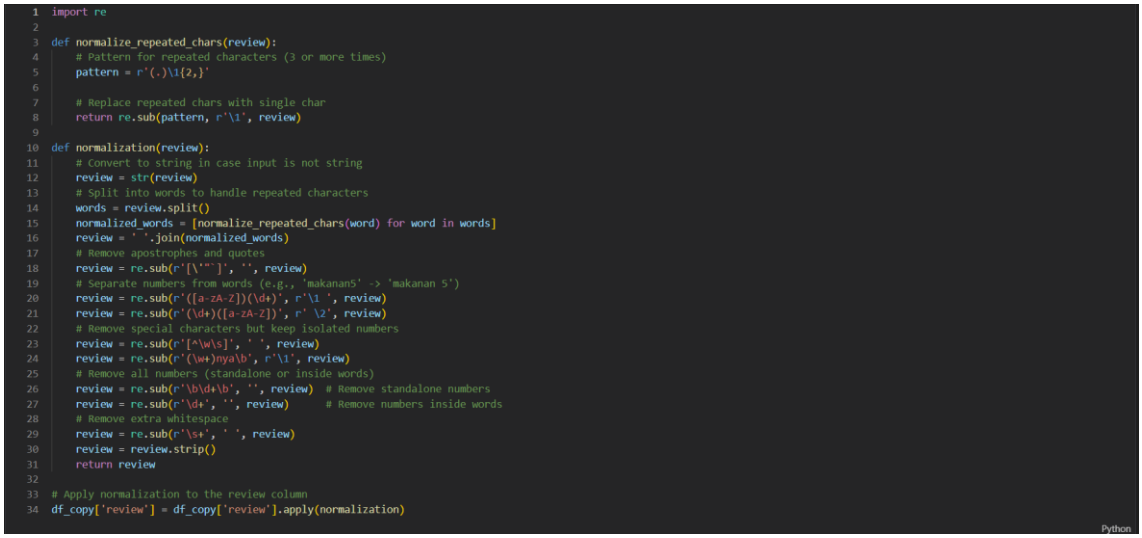
# --- Lanjutkan ke proses casefolding seperti biasa ---
def casefolding(review):
    review = str(review) if review is not None else ''
    return review.lower()

df_copy['review'] = df_copy['review'].apply(casefolding)
df_copy.head()
```

Python

name	gojek_user_since	# rating	review	item	date_purchased	# item_count
Kevin	2 December 2020	4.0	sosisnya kecil banget	Corndog Sosis Mozzarella - Corndoc	10 May 2025	2
F***** C*********	1 June 2017	2.0	kaget pas dtg. tteok nya kecil bgt, ch	Topokki Original	3 May 2025	1
A***	23 September 2021	5.0	enakkk bgt!!! apalagi corndog matcl	Corndog Matcha - Hot Spicy Bibimm	7 February 2025	2
R*****	10 January 2022	5.0	corndog nya enak. bibimyeon nya jg	Corndog Sosis Mozzarella - Hot Spic	3 February 2025	2
A********* D****	22 January 2022	5.0	aku kira bakal kecil bgt, ternyata stai	Corndog Mozarella	24 January 2025	1

- ***Normalization***



```python
import re

def normalize_repeated_chars(review):
    # Pattern for repeated characters (3 or more times)
    pattern = r'(.)\1{2,}'

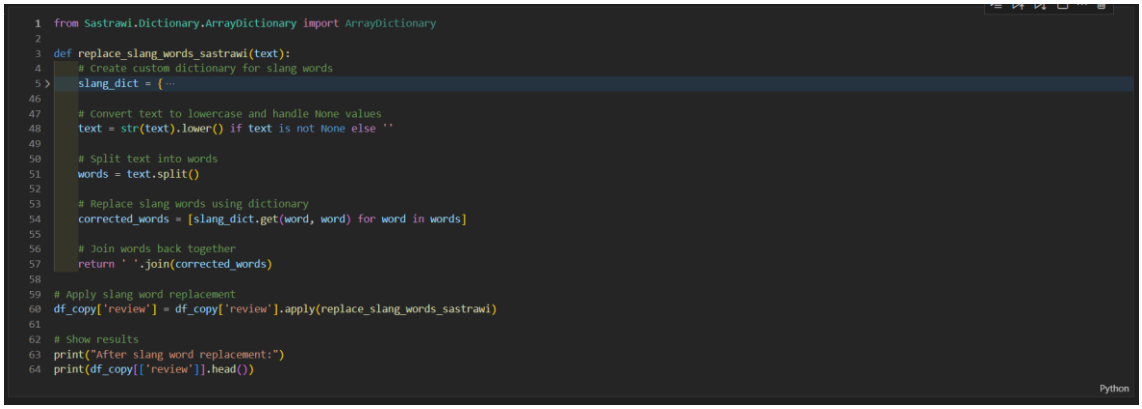
    # Replace repeated chars with single char
    return re.sub(pattern, r'\1', review)

def normalization(review):
    # Convert to string in case input is not string
    review = str(review)
    # Split into words to handle repeated characters
    words = review.split()
    normalized_words = [normalize_repeated_chars(word) for word in words]
    review = ' '.join(normalized_words)
    # Remove apostrophes and quotes
    review = re.sub(r'[\'"`]', '', review)
    # Separate numbers from words (e.g., 'makanan5' -> 'makanan 5')
    review = re.sub(r'([a-zA-Z])(\d+)', r'\1 ', review)
    review = re.sub(r'(\d+)([a-zA-Z])', r' \2', review)
    # Remove special characters but keep isolated numbers
    review = re.sub(r'[^\w\s]', ' ', review)
    review = re.sub(r'(\w+)nya\b', r'\1', review)
    # Remove all numbers (standalone or inside words)
    review = re.sub(r'\b\d+\b', '', review)  # Remove standalone numbers
    review = re.sub(r'\d+', '', review)      # Remove numbers inside words
    # Remove extra whitespace
    review = re.sub(r'\s+', ' ', review)
    review = review.strip()
    return review

# Apply normalization to the review column
df_copy['review'] = df_copy['review'].apply(normalization)
```

Python

- ***Slang words replacement***



```python
from Sastrawi.Dictionary.ArrayDictionary import ArrayDictionary

def replace_slang_words_sastrawi(text):
    # Create custom dictionary for slang words
    slang_dict = {...

    # Convert text to lowercase and handle None values
    text = str(text).lower() if text is not None else ''

    # Split text into words
    words = text.split()

    # Replace slang words using dictionary
    corrected_words = [slang_dict.get(word, word) for word in words]

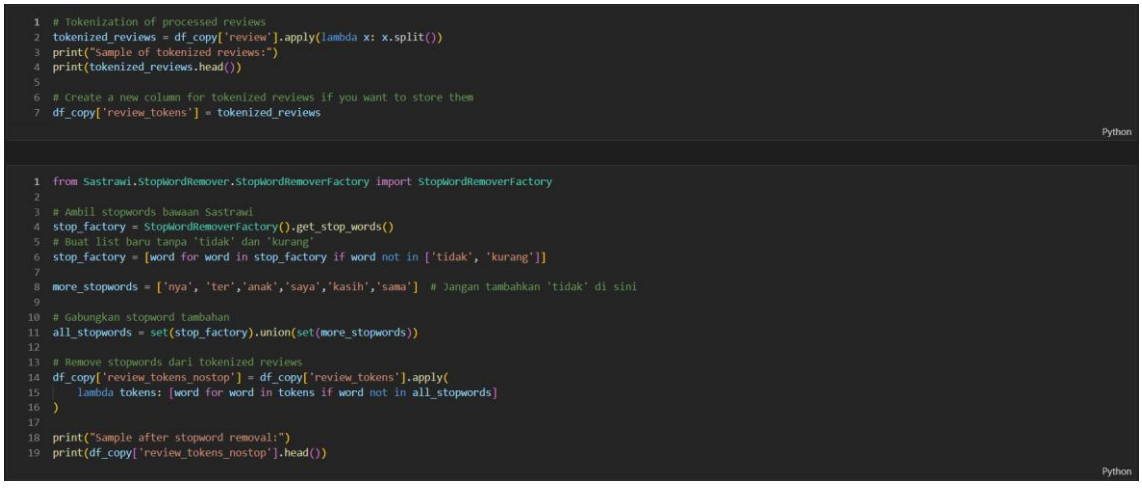
    # Join words back together
    return ' '.join(corrected_words)

# Apply slang word replacement
df_copy['review'] = df_copy['review'].apply(replace_slang_words_sastrawi)

# Show results
print("After slang word replacement:")
print(df_copy[['review']].head())
```

Python

- ***Tokenization***



```python
# Tokenization of processed reviews
tokenized_reviews = df_copy['review'].apply(lambda x: x.split())
print("Sample of tokenized reviews:")
print(tokenized_reviews.head())

# Create a new column for tokenized reviews if you want to store them
df_copy['review_tokens'] = tokenized_reviews
```

Python

```python
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory

# Ambil stopwords bawaan Sastrawi
stop_factory = StopWordRemoverFactory().get_stop_words()
# Buat list baru tanpa 'tidak' dan 'kurang'
stop_factory = [word for word in stop_factory if word not in ['tidak', 'kurang']]

more_stopwords = ['nya', 'ter','anak','saya','kasih','sama']  # Jangan tambahkan 'tidak' di sini

# Gabungkan stopword tambahan
all_stopwords = set(stop_factory).union(set(more_stopwords))

# Remove stopwords dari tokenized reviews
df_copy['review_tokens_nostop'] = df_copy['review_tokens'].apply(
    lambda tokens: [word for word in tokens if word not in all_stopwords]
)

print("Sample after stopword removal:")
print(df_copy['review_tokens_nostop'].head())
```

Python

- ***Stemming***

```python
# Initialize Sastrawi stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()

def stem_tokens(tokens):
    """ ...
    # Dictionary of words to preserve
    preserve_words = {...

    processed_words = []
    for token in tokens:
        token = str(token).lower()
        # Check if word should be preserved
        if token in preserve_words:
            processed_words.append(preserve_words[token])
        else:
            # Apply normal stemming
            processed_words.append(stemmer.stem(token))

    # Join back into text
    return ' '.join(processed_words)

# Apply stemming to review_tokens_nostop and create review_stemmed column
df_copy['review_stemmed'] = df_copy['review_tokens_nostop'].apply(stem_tokens)

# Display sample results
print("Sample of stemmed reviews:")
comparison_df = pd.DataFrame({
    'Original Tokens (No Stopword)': df_copy['review_tokens_nostop'],
    'Stemmed Text': df_copy['review_stemmed']
})
print(comparison_df.head())
```

Python

C-*Labelling data*

- ***Lexicon***

```python
# Kamus lexicon positif dan negatif (lebih lengkap)
lexicon_positif = [...

lexicon_negatif = [...

def lexicon_label(text):
    tokens = text.split()
    pos = sum(token in lexicon_positif for token in tokens)
    neg = sum(token in lexicon_negatif for token in tokens)
    if pos > neg:
        return 'positif'
    elif neg > pos:
        return 'negatif'
    else:
        return 'netral'

# Terapkan ke kolom review_stemmed
df_copy['label_lexicon'] = df_copy['review_stemmed'].apply(lexicon_label)
# Lihat distribusi label lexicon
print(df_copy['label_lexicon'].value_counts())
```

Python

```
label_lexicon
positif    85
netral     80
negatif    61
Name: count, dtype: int64
```

```python
# Fungsi pelabelan ulang untuk review netral, memaksa semua data punya polarity
def relabel_netral_force(text):
    tokens = text.split()
    pos = sum(token in lexicon_positif for token in tokens)
    neg = sum(token in lexicon_negatif for token in tokens)
    if pos > neg:
        return 'positif'
    elif neg > pos:
        return 'negatif'
    else:
        if 'tidak' in tokens or 'kurang' in tokens:
            return 'negatif'
        elif 'baik' in tokens or 'bagus' in tokens:
            return 'positif'
        else:
            return 'negatif'

# Kolom baru untuk label polarity
df_copy['label_polarity'] = df_copy['review_stemmed'].apply(lexicon_label)
print("Jumlah data netral sebelum relabel:", (df_copy['label_polarity'] == 'netral').sum())

mask_netral = df_copy['label_polarity'] == 'netral'
df_copy.loc[mask_netral, 'label_polarity'] = df_copy.loc[mask_netral, 'review_stemmed'].apply(relabel_netral_force)

print("Jumlah data netral setelah relabel:", (df_copy['label_polarity'] == 'netral').sum())
print(df_copy['label_polarity'].value_counts())
```

Python

```
Jumlah data netral sebelum relabel: 80
Jumlah data netral setelah relabel: 0
label_polarity
negatif    141
positif     85
Name: count, dtype: int64
```

- ***TF-IDF***

```python
from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer

# Filter data yang sudah terlabeli polarity
df_fix = df_copy[df_copy['label_polarity'].isin(['positif', 'netral', 'negatif'])].copy()
df_fix['review_stemmed'] = df_fix['review_stemmed'].fillna('tidak ada ulasan')

# Gunakan kolom hasil pre-processing dan pelabelan
Ulasan = df_fix['review_stemmed']

cv = CountVectorizer()
term_fit = cv.fit(Ulasan)

print(len(term_fit.vocabulary_))
print(term_fit.vocabulary_)
```

Python

```python
#kolom pertama mengartikan jumlah dokumen
#Kolom kedua mengartikan letak kata dalam dokumen
#kolom ketiga mengartikan hasil dari tf

term_frequency_all = term_fit.transform(Ulasan)
print(term_frequency_all)

# TF-IDF
tfidf_transformer = TfidfTransformer().fit(term_frequency_all)
tfidf_all = tfidf_transformer.transform(term_frequency_all)
print(tfidf_all)

# Contoh: TF dan TF-IDF untuk satu dokumen (misal dokumen ke-1)
ulasan_tf = Ulasan.iloc[1]
print("Ulasan ke-1:", ulasan_tf)
term_frequency = term_fit.transform([ulasan_tf])
print("TF dokumen ke-1:\n", term_frequency)
tfidf = tfidf_transformer.transform(term_frequency)
print("TF-IDF dokumen ke-1:\n", tfidf)
```

Python

```python
# Gunakan Ulasan dari df_fix['review_stemmed']
dokumen = term_fit.transform(Ulasan)
tfidf_transformer = TfidfTransformer().fit(dokumen)
print(tfidf_transformer.idf_)

# Contoh: TF-IDF untuk satu dokumen (misal dokumen ke-1)
term_frequency = term_fit.transform([Ulasan.iloc[1]])
tfidf = tfidf_transformer.transform(term_frequency)
print(tfidf)
```

Python

D-*Script python* klasifikasi

- ***Dataset split 90:10***

```python
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.model_selection import GridSearchCV, StratifiedKFold, train_test_split
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
from imblearn.pipeline import Pipeline
import numpy as np

# Gunakan label hasil polarity
df_fix = df_copy[df_copy['label_polarity'].isin(['positif','negatif'])].copy()
df_fix['review_stemmed'] = df_fix['review_stemmed'].fillna('tidak ada ulasan')

X = df_fix['review_stemmed']
y = df_fix['label_polarity']

# Split data manual: 90% train, 20% test
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.1, stratify=y, random_state=42
)

# Pipeline: TF-IDF ->  MultinomialNB
pipeline = Pipeline([
    ('tfidf', TfidfVectorizer(
        decode_error='replace',
        encoding='utf-8',
```

```
        ngram_range=(1,2),
        min_df=2,
        sublinear_tf=True
    )),
    ('clf', MultinomialNB())
])

# Hyperparameter grid
param_grid = {
    'clf__alpha': [0.01, 0.05, 0.1, 0.5, 1.0, 2.0, 5.0, 10.0],
    'clf__fit_prior': [True, False]
}

cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

# GridSearchCV
grid = GridSearchCV(
    pipeline,
    param_grid,
    cv=cv,
    scoring='accuracy',
    n_jobs=-1
)
grid.fit(X_train, y_train)
y_pred = grid.predict(X_test)

print("Best Params:", grid.best_params_)
print("Best CV Score:", grid.best_score_)

# Prediksi pada seluruh data (atau bisa split manual untuk test set)
y_pred = grid.predict(X_test)
print("Best Params:", grid.best_params_)
print("Best CV Score:", f"{grid.best_score_:.3f}")
print("\nConfusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred, digits=3))
print(f"\nAccuracy: {accuracy_score(y_test, y_pred):.3f}")
```

- ***Dataset split 80:20***

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.model_selection import GridSearchCV, StratifiedKFold, train_test_split
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
from imblearn.pipeline import Pipeline
import numpy as np

# Gunakan label hasil polarity
df_fix = df_copy[df_copy['label_polarity'].isin(['positif','negatif'])].copy()
```

```
df_fix['review_stemmed'] = df_fix['review_stemmed'].fillna('tidak ada ulasan')

X = df_fix['review_stemmed']
y = df_fix['label_polarity']

# Split data manual: 80% train, 20% test
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, stratify=y, random_state=42
)

# Pipeline: TF-IDF ->  MultinomialNB
pipeline = Pipeline([
    ('tfidf', TfidfVectorizer(
        decode_error='replace',
        encoding='utf-8',
        ngram_range=(1,2),
        min_df=2,
        sublinear_tf=True
    )),
    ('clf', MultinomialNB())
])

# Hyperparameter grid
param_grid = {
    'clf__alpha': [0.01, 0.05, 0.1, 0.5, 1.0, 2.0, 5.0, 10.0],
    'clf__fit_prior': [True, False]
}

cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

# GridSearchCV
grid = GridSearchCV(
    pipeline,
    param_grid,
    cv=cv,
    scoring='accuracy',
    n_jobs=-1
)
grid.fit(X_train, y_train)
y_pred = grid.predict(X_test)

print("Best Params:", grid.best_params_)
print("Best CV Score:", grid.best_score_)

# Prediksi pada seluruh data (atau bisa split manual untuk test set)
y_pred = grid.predict(X_test)
print("Best Params:", grid.best_params_)
print("Best CV Score:", f"{grid.best_score_:.3f}")
```

```
print("\nConfusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred, digits=3))
print(f"\nAccuracy: {accuracy_score(y_test, y_pred):.3f}")
```

- ***Dataset split 70:30***

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.model_selection import GridSearchCV, StratifiedKFold, train_test_split
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
from imblearn.pipeline import Pipeline
import numpy as np

# Gunakan label hasil polarity
df_fix = df_copy[df_copy['label_polarity'].isin(['positif','negatif'])].copy()
df_fix['review_stemmed'] = df_fix['review_stemmed'].fillna('tidak ada ulasan')

X = df_fix['review_stemmed']
y = df_fix['label_polarity']

# Split data manual: 70% train, 30% test
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.3, stratify=y, random_state=42
)

# Pipeline: TF-IDF ->  MultinomialNB
pipeline = Pipeline([
    ('tfidf', TfidfVectorizer(
        decode_error='replace',
        encoding='utf-8',
        ngram_range=(1,2),
        min_df=2,
        sublinear_tf=True
    )),
    ('clf', MultinomialNB())
])

# Hyperparameter grid
param_grid = {
    'clf__alpha': [0.01, 0.05, 0.1, 0.5, 1.0, 2.0, 5.0, 10.0],
    'clf__fit_prior': [True, False]
}

cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

# GridSearchCV
grid = GridSearchCV(
    pipeline,
```

```
    param_grid,
    cv=cv,
    scoring='accuracy',
    n_jobs=-1
)
grid.fit(X_train, y_train)
y_pred = grid.predict(X_test)

print("Best Params:", grid.best_params_)
print("Best CV Score:", grid.best_score_)

# Prediksi pada seluruh data (atau bisa split manual untuk test set)
y_pred = grid.predict(X_test)
print("Best Params:", grid.best_params_)
print("Best CV Score:", f"{grid.best_score_:.3f}")
print("\nConfusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred, digits=3))
print(f"\nAccuracy: {accuracy_score(y_test, y_pred):.3f}")
```

E-*Script python association rule*

- ***Association rule***

```
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori, association_rules
import pandas as pd
from collections import Counter

# Ambil review negatif berdasarkan label_polarity terbaru
df_negatif = df_copy[df_copy['label_polarity'] == 'negatif']

# Tampilkan jumlah review negatif yang dipakai sebagai input asosiasi
print(f"Jumlah review negatif (label_polarity): {len(df_negatif)}")

# 1. Siapkan transaksi dari review negatif
transactions = df_negatif['review_stemmed'].dropna().apply(str.split).tolist()
print(f"Jumlah transaksi (setelah drop NA): {len(transactions)}")

# 2. Hitung frekuensi semua kata
flat_words = [word for review in transactions for word in review]
word_counts = Counter(flat_words)
min_word_freq = 3  # bisa diatur sesuai kebutuhan

# 3. Hapus kata yang terlalu jarang muncul dan review yang terlalu pendek
filtered_transactions = [
    [word for word in review if word_counts[word] >= min_word_freq]
    for review in transactions
]
```

```
filtered_transactions = [review for review in filtered_transactions if len(review) >= 2]
print(f"Jumlah transaksi setelah filter: {len(filtered_transactions)}")

# 4. Encode transaksi
te = TransactionEncoder()
te_ary = te.fit(filtered_transactions).transform(filtered_transactions)
df_trans = pd.DataFrame(te_ary.astype(bool), columns=te.columns_)

# 5. Apriori (ganti fpgrowth)
frequent_itemsets = apriori(
    df_trans,
    min_support=0.04,  # bisa diubah sesuai kebutuhan
    use_colnames=True
)

# 6. Buat aturan asosiasi
rules = association_rules(
    frequent_itemsets,
    metric="confidence",
    min_threshold=0.1
)

# 7. Filter rules kuat dan urutkan
rules = rules[rules['lift'] > 1].sort_values(by='lift', ascending=False)

# 8. Output
print(f"Jumlah aturan asosiasi yang terbentuk: {len(rules)}")
print(rules[['antecedents', 'consequents', 'support', 'confidence', 'lift']].head(10))
```

- **Visualisasi**

```
# Visualisasi pohon asosiasi berdasarkan top 10 kata dari wordcloud (frekuensi tertinggi pada
review negatif)
from collections import Counter
import matplotlib.pyplot as plt
from matplotlib.patches import FancyBboxPatch
import numpy as np

# Ambil top 10 kata dari visualisasi wordcloud pada cell sebelumnya
words = text_negatif.split()
word_counts = Counter(words)
top_n = 10
most_common = word_counts.most_common(top_n)
top_words = [w for w, c in most_common]
```

```
# Siapkan tree_dict: parent = kata top 10, child = kata yang sering muncul bersama (dari
association rules)
max_child = 5  # maksimum child node
tree_dict = {}
for word in top_words:
    # Cari aturan asosiasi dengan antecedent = word (hanya satu kata) dan consequent satu kata
    matched_rules = rules[
        (rules['antecedents'].apply(lambda x: len(x) == 1 and word in x)) &
        (rules['consequents'].apply(lambda x: len(x) == 1))
    ].sort_values(by='confidence', ascending=False)  # sort by confidence
    children = []
    for _, row in matched_rules.head(max_child).iterrows():
        child = list(row['consequents'])[0]
        conf = row['confidence']
        lift = row['lift']
        children.append((child, conf, lift))
    if children:
        tree_dict[word] = children

roots = list(tree_dict.keys())

# Daftar warna pastel untuk node root
pastel_colors = [
    "#FFB5A7", "#FFD6A5", "#FDFFB6", "#CAFFBF", "#9BF6FF",
    "#A0C4FF", "#BDB2FF", "#FFC6FF", "#FFFFFC", "#E0FBFC"
]

def draw_node(ax, x, y, text, color, fontsize=13, boxstyle="round,pad=0.4"):
    ax.text(
        x, y, text, ha='center', va='center', fontsize=fontsize,
        bbox=dict(facecolor=color, boxstyle=boxstyle, edgecolor='gray', linewidth=1.5)
    )

def draw_tree(ax, root, children, color):
    draw_node(ax, 0.5, 0.80, root, color, fontsize=15, boxstyle="round,pad=0.7")  # dari 0.88 ke
0.80
    n_child = len(children)
    if n_child == 1:
        y_positions = [0.65]  # dari 0.60 ke 0.65
    else:
        y_top = 0.70          # dari 0.72 ke 0.70
        y_bottom = 0.50       # dari 0.40 ke 0.50
        y_positions = [y_top - i * (y_top - y_bottom) / (n_child - 1) for i in range(n_child)]
    for i, ((child, conf, lift), y_child) in enumerate(zip(children, y_positions)):
        draw_node(ax, 0.5, y_child, f"{child}\n(conf: {conf:.2f}, lift: {lift:.2f})", "#f5f5f5",
fontsize=13)
        ax.plot([0.5, 0.5], [0.78, y_child + 0.07], color='black', lw=1.5)  # dari 0.85 ke 0.78
    ax.axis('off')
```

```
# Layout grid otomatis, misal 4 kolom per baris
n_cols = 4
n_rows = (len(roots) + n_cols - 1) // n_cols
# Perpendek tinggi figure
fig, axes = plt.subplots(n_rows, n_cols, figsize=(n_cols*5.5, n_rows*5),
constrained_layout=False)  # dari n_rows*8 ke n_rows*5
# ...existing code...
axes = axes.flatten() if n_rows > 1 else [axes]

for idx, root in enumerate(roots):
    ax = axes[idx]
    color = pastel_colors[idx % len(pastel_colors)]
    children = tree_dict[root]
    draw_tree(ax, root, children, color)

for idx in range(len(roots), n_rows * n_cols):
    axes[idx].axis('off')

plt.subplots_adjust(hspace=0.7, top=0.92)
plt.suptitle("Visualisasi Pohon Aturan Asosiasi (Top 10 Kata Wordcloud Negatif)", fontsize=20,
fontweight='bold', y=0.97)
```

F-Dokumen *Standard Operation Procedure*

STANDARD OPERATING PROCEDURE (SOP)

Outlet Grobbak Korea — Pandeyan

Dokumen ini berisi pedoman operasional pembuatan produk, pengelolaan bahan baku, dan prosedur kebersihan. Semua ukuran/angka yang harus diisi outlet ditandai sebagai (....) agar dapat disesuaikan dengan standar operasional outlet.

DAFTAR ISI

DAFTAR ISI 2
1. Pendahuluan 3
2. Ruang Lingkup 3
3. Definisi Singkat (dipakai dalam SOP) 3
4. PEMBUATAN PRODUK 4
4.1. Corndog (Sosis / Mozzarella / Mix) 4
4.2. Tteokbokki saus korea 4
4.3. Chikuwa saus korea 5
4.4. Odeng fish soup 6
4.5. Rabokki (Tteokbokki + Ramyeon) 6
4.6. Mandu (Fried / Steam) 7
4.7. Catatan Penyajian 7
5. MANAJEMEN & PENGADAAN BAHAN BAKU 8
5.1. Pemesanan & penerimaan bahan baku 8
5.2. Penyimpanan 8
6. KEBERSIHAN DAN SANITASI 9
7. SELF SUPERVISING 10

1. Pendahuluan

Dokumen **Standard Operating Procedure (SOP)** ini disusun sebagai pedoman resmi bagi seluruh staf di **Outlet Grobbak Korea Pandeyan** dalam melaksanakan aktivitas operasional sehari-hari. Tujuan utama dari penyusunan SOP adalah untuk memastikan bahwa setiap proses kerja berjalan sesuai standar yang telah ditetapkan, sehingga produk yang dihasilkan memiliki mutu yang konsisten, aman dikonsumsi, serta sesuai dengan harapan pelanggan. Dengan adanya SOP ini, diharapkan seluruh aktivitas kerja di outlet akan terdokumentasi dengan baik, mudah diajarkan kepada staf baru, serta dapat dievaluasi secara berkala untuk perbaikan berkelanjutan.

2. Ruang Lingkup

SOP berlaku untuk seluruh staf dapur dan manajemen yang terlibat dalam:

1. Pembuatan produk: Corndog, Tteokbokki, Chikuwa, Odeng, Rabokki, Mandu.
2. Manajemen & pengadaan bahan baku
3. Kebersihan dan sanitasi

3. Definisi Singkat (dipakai dalam SOP)

- pcs : Buah/untuk item utuh (mis. 1 sosis = 1 pcs).
- sdm / sdt : Sendok makan / sendok teh (untuk takaran saus, tepung, dsb).
- menit : Waktu sebagai patokan pemanasan/masak.
- (....) : Placeholder yang harus diisi standard outlet.

4. PEMBUATAN PRODUK

4.1. Corndog (Sosis / Mozzarella / Mix)

Persiapan

1. Siapkan bahan: tepung adonan (.... sdm), sosis (.... pcs), mozzarella stick (.... pcs), tepung roti (.... sdm), saus Korea (.... sdm).
2. Siapkan peralatan: tusuk, wadah adonan, sendok takar, saringan/talenan, penggorengan, kertas minyak.

Proses

1. Tusuk 1 isi (sosis / mozzarella / gabungan) pada tusuk.
2. Celupkan isi kedalam adonan (ambil adonan dengan (.... sdm) agar lapisan merata).
3. Gulung/tepuk pada tepung roti sebanyak (.... sdm) agar menempel rata.
4. Goreng hingga berwarna keemasan selama (.... menit). (Catat: gunakan pengamatan visual; jika permukaan matang dan berwarna keemasan, angkat.)
5. Tiriskan selama (.... menit) pada rak peniris agar minyak berlebih hilang.

Penyajian

1. Bungkus 1 pcs corndog pada kemasan standar.
2. Quality check: warna cokelat keemasan, isi meleleh jika menu mengandung mozzarella.

4.2. Tteokbokki saus korea

Persiapan

1. Siapkan: tteok (.... pcs), odeng (.... pcs), sayuran (.... sdm), saus Korea (.... sdm), mozzarella opsional (.... sdm).
2. Siapkan panci, sendok sayur, peniris.

Proses

1. Rebus tteok sebanyak (.... pcs) selama (.... menit) hingga kenyal & tepung berubah warna. Tiriskan.

2. Panaskan wajan; masukkan saus Korea sebanyak (.... sdm) dan masak selama (.... menit).
3. Masukkan odeng/chikuwa opsional(.... pcs), bubuk cabai sesuai tingkat kepedasan(….sdt), aduk (.... menit).
4. Masukkan tteok, aduk bersama saus selama (.... menit) sampai saus mengental dan melapisi tteok.
5. Jika menu cheese: taburkan mozzarella (.... sdm) di permukaan, diamkan (.... menit) sampai meleleh.

Penyajian:

1. Sajikan dalam bowl.
2. Taburi wijen (.... sdt) dan daun bawang (.... sdt).
3. Quality check: tingkat kepedasan sesuai level pesanan, tteok empuk kenyal, porsi tteok = (.... pcs).

4.3. Chikuwa saus korea

Persiapan

4. Siapkan chikuwa (.... pcs) dan saus Korea (.... sdm).
5. Siapkan papan potong, pisau, panic, dan wajan.

Proses

1. Potong chikuwa tiap (.... cm) per potong sesuai standar, rebus chikuwa hingga matang selama (…. menit)
2. Panaskan wajan, tuang saus (.... sdm), masukkan bubuk cabai sesuai tingkat kepedasan (…. Sdt), panaskan selama (.... menit).
3. Masukkan chikuwa (.... pcs) dan aduk rata selama (.... menit) hingga saus melapisi tiap potong.

Penyajian

1. Sajikan (.... pcs) chikuwa didalam bowl.
2. Quality check: saus merata, chikuwa hangat.

4.4. Odeng fish soup

Persiapan

1. Siapkan odeng (.... pcs), kaldu (.... sdm), sayuran (.... sdm).
2. Siapkan panci.

Proses

1. Panaskan kaldu sampai panas; masukkan odeng (.... pcs) dan rebus selama (.... menit).
2. Tambahkan sayuran (.... sdm) dan masak tambahan (.... menit).

Penyajian

1. Sajikan dalam bowl, taburi dengan daun bawang (.... sdt).
2. Quality check: kuah hangat, rasa kaldu seimbang.

4.5. Rabokki (Tteokbokki + Ramyeon)

Persiapan:

Siapkan: tteok (.... pcs), ramyeon (.... pcs/bungkus), saus Korea (.... sdm), mozzarella (opsional) (.... sdm).

Proses:

1. Rebus ramyeon selama (.... menit) sampai setengah matang, tiriskan.
2. Rebus tteok (.... pcs) selama (.... menit), tiriskan.\
3. Panaskan saus (.... sdm) di wajan, tambahkan odeng (.... pcs) dan masak (.... menit).
4. Masukkan ramyeon & tteok, aduk rata selama (.... menit) hingga saus meresap.
5. Untuk cheese: tabur mozzarella (.... sdm), diamkan (.... menit) sampai meleleh.

Penyajian:

1. Sajikan dalam mangkuk besar (.... ml).
2. Quality check: tekstur mie & tteok pas, rasa pedas sesuai pilihan.

4.6. Mandu (Fried / Steam)

Persiapan:

Siapkan mandu (.... pcs per porsi) dan saus pendamping (.... sdm).

Proses:

- Fried: Goreng mandu (.... pcs) selama (.... menit) sampai kulit renyah.

- Steam: Kukus mandu (.... pcs) selama (.... menit) sampai matang.

Penyajian:

1. Sajikan (.... pcs) mandu pada bowl; tambahkan saus pendamping (.... sdm).
2. Quality check: isi panas, kulit matang, tidak bau.

4.7. Catatan Penyajian

1. Jika menu dibawa pulang/takeaway/online order, maka menu topokki dan chikuwa diolah secara terpisah dengan saus
2. Jika menu dibawa pulang/takeaway/online order, maka seluruh kondimen menu topokki dan chikuwa disajikan secara terpisah dengan kuah/saus(kuah dikemas dengan plastik secara terpisah)

5. MANAJEMEN & PENGADAAN BAHAN BAKU

5.1. Pemesanan & penerimaan bahan baku

1. Jadwal pemesanan rutin: (....) kali/…..
2. Reorder point: ketika stok mencapai (.... pcs) atau bahan diperkirakan habis dalam (....) hari → lakukan order.
3. Cek visual dan fisik produk saat penerimaan di outlet(kondisi kemasan, tanggal kadaluarsa,dll)
4. Beri label pada tiap kemasan produk bahan baku, catat tanggal penerimaan & tanggal kadaluarsa pada label.

5.2. Penyimpanan

1. Pastikan bahan baku yang tergolong kedalam *frozen foods* disimpan sesuai petunjuk penyimpanan produk.
2. Terapkan FIFO(*first in first out)*: keluarkan barang yang paling lama masuk terlebih dahulu.
3. Pemeriksaan harian stok: petugas memeriksa stok sisa & kondisi, update log stok.

6. KEBERSIHAN DAN SANITASI

1. Cuci tangan wajib: sebelum shift, sebelum dan setelah pegang bahan mentah, setelah buang sampah.
2. Membersihkan area kerja & peralatan: jadwal harian (shift end).
3. Gunakan sarung tangan saat handling menu.
4. Sampah dibuang sesuai kategori: organik / non-organik.

7. SELF SUPERVISING

1. Gunakan *checklist* kontrol harian (bahan lengkap, kebersihan terjaga) dan laporkan kepada pihak terkait
2. Pelaporan kondisi kebersihan *stall* dan area kerja sesaat sebelum toko ditutup kepada pihak terkait