BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Analisis Sentimen

Analisis sentimen adalah bagian dari kajian *text mining* yang digunakan untuk mengidentifikasi dan mengelompokkan opini, penilaian, atau ekspresi emosi individu terhadap suatu objek melalui data berbentuk teks. Objek yang dianalisis dapat berupa produk, layanan, organisasi, atau peristiwa tertentu, sedangkan data teks umumnya berasal dari ulasan pelanggan, komentar media sosial, atau dokumen opini lainnya (R & T, 2023). Secara umum, analisis sentimen mengelompokkan teks ke dalam kategori sentimen seperti positif atau negatif (Nip & Berthelier, 2024; Yadav, 2023).

Analisis sentimen didefinisikan sebagai proses komputasional yang memanfaatkan teknik *Natural Language Processing* (NLP), *text mining*, dan *machine learning* untuk mengekstraksi serta menentukan polaritas sentimen dari suatu teks (Goswami dkk., 2025). Pendekatan ini dipilih karena dapat memproses ulasan dalam jumlah besar secara otomatis dengan waktu yang lebih efisien, sehingga membantu organisasi dalam memahami kepuasan maupun ketidakpuasan pengguna (Kumar & Upadhyay, 2023).

Berdasarkan tingkat kedalaman analisis, analisis sentimen diklasifikasikan ke dalam tiga level utama, yaitu analisis sentimen berbasis dokumen, berbasis kalimat, dan berbasis aspek (Sharma & Kumar, 2021; Yadav, 2023), yaitu:

1. Analisis sentimen berbasis dokumen

   Analisis sentimen berbasis dokumen mengklasifikasikan satu dokumen atau satu ulasan secara keseluruhan ke dalam satu kategori sentimen. Metode ini sederhana, namun kurang efektif ketika satu ulasan mengandung opini terhadap beberapa aspek layanan.

2. Analisis sentimen berbasis kalimat

   Analisis sentimen berbasis kalimat memberikan tingkat lebih detail yang lebih tinggi, tetapi masih belum mampu mengaitkan sentimen dengan aspek tertentu secara eksplisit.

3. Analisis sentimen berbasis aspek (*Aspect-Based Sentiment Analysis*)

   Analisis sentimen berbasis aspek dinilai lebih relevan untuk memahami opini pengguna secara komprehensif (Nip & Berthelier, 2024). Oleh karena itu, ABSA dianggap lebih informatif dan relevan untuk analisis ulasan pelanggan.

2.1.2 Analisis Sentimen Berbasis Aspek

Aspect-Based Sentiment Analysis (ABSA) merupakan pendekatan analisis sentimen yang digunakan untuk mengidentifikasi aspek atau atribut tertentu dalam sebuah teks sekaligus menentukan polaritas sentimen yang berkaitan dengan aspek tersebut (Yadav, 2023). Pendekatan ini memungkinkan satu ulasan mengandung lebih dari satu sentimen terhadap aspek layanan yang berbeda, sehingga menghasilkan analisis yang lebih rinci dan informatif (Nip & Berthelier, 2024).

Aspek dalam analisis sentimen dapat dipahami sebagai bagian, atribut, atau unsur tertentu dari produk maupun layanan yang menjadi fokus opini pengguna,

seperti tarif, kualitas pengemudi, performa aplikasi, atau keamanan layanan (Yusuf dkk., 2024). Identifikasi aspek dapat dilakukan secara manual atau otomatis. Penentuan aspek dalam penelitian ini dilakukan secara manual dengan mengacu pada jurnal acuan terkait analisis sentimen aplikasi *ride-hailing*. Aspek yang digunakan meliputi sentimen umum, pengemudi, layanan, aplikasi, dan harga/biaya, yang merupakan aspek yang sering muncul dan relevan dalam ulasan pelanggan aplikasi transportasi berbasis aplikasi (Iqrom dkk., 2025). Pendekatan ini dipilih untuk menjaga konsistensi analisis dan meningkatkan interpretasi dan perbandingan hasil penelitian.

Perbedaan utama antara analisis sentimen umum dan ABSA terletak pada tingkat analisis. Analisis sentimen umum hanya menghasilkan satu label sentimen untuk seluruh teks, sedangkan ABSA menghasilkan sentimen untuk setiap aspek yang teridentifikasi (Ghadage dkk., 2024). Dalam penelitian ini, ABSA digunakan untuk menganalisis sentimen pelanggan terhadap aspek-aspek layanan aplikasi Grab, Gojek, dan Maxim secara spesifik.

Menurut Liu (2012) analisis sentimen tingkat aspek memiliki dua tugas inti, yaitu:

1. *Aspect Extraction*, proses mengidentifikasi fitur atau aspek dari suatu entitas yang menjadi objek opini dalam teks.
2. *Aspect Sentiment Classification*, proses menentukan polaritas sentimen (positif, negatif, atau netral) yang diekspresikan terhadap setiap aspek yang telah diidentifikasi.

Selain dua tugas utama tersebut, hasil analisis dapat digunakan untuk *Opinion Summarization* yaitu proses merangkum hasil analisis opini terhadap berbagai aspek sehingga diperoleh gambaran umum sentimen pengguna terhadap suatu entitas. Menurut Liu (2012), *aspect-based opinion summarization* memiliki dua karakteristik utama. Pertama, peringkasan opini berfokus pada inti opini, yaitu target opini (entitas dan aspek-aspeknya) serta sentimen yang terkait dengan masing-masing aspek tersebut. Kedua, peringkasan opini bersifat kuantitatif, yang berarti menyajikan jumlah atau persentase opini positif dan negatif terhadap setiap entitas maupun aspek yang dianalisis.

2.1.3 *Text Mining*

Text mining adalah proses pengolahan data teks tidak terstruktur untuk menemukan informasi, pola, serta pengetahuan yang bermanfaat dengan bantuan teknik statistik, *machine learning*, dan NLP (R & T, 2023). Data teks yang dianalisis umumnya berupa ulasan pelanggan, komentar, atau dokumen opini yang tidak memiliki format terstruktur seperti data numerik (Goswami dkk., 2025) .

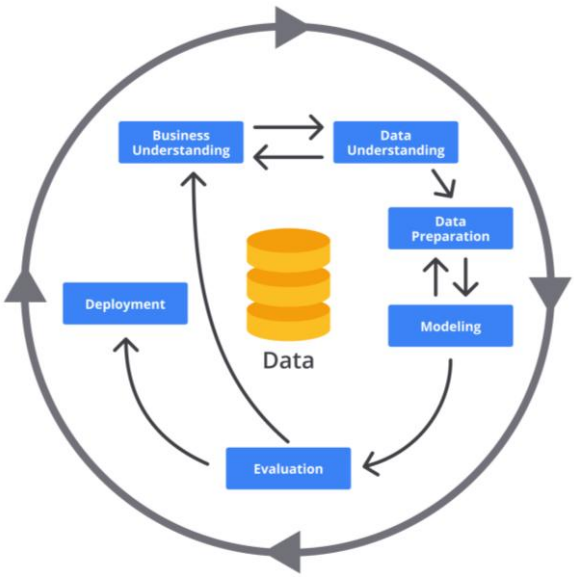
Perbedaan kunci antara *text mining* dan *data mining* terdapat pada tipe data yang dianalisis. *Data mining* berfokus pada data terstruktur, sedangkan *text mining* berfokus penanganan data tidak terstruktur seperti pra-pemrosesan dan ekstraksi fitur (Chen, 2025).

Dalam konteks analisis sentimen, *text mining* berperan penting sebagai kerangka utama yang mencakup pengumpulan data, pra-pemrosesan, ekstraksi fitur, dan klasifikasi sentimen (R & T, 2023). Pada penelitian ini, *text mining*

digunakan sebagai pendekatan utama untuk mengolah ulasan pelanggan aplikasi Grab, Gojek, dan Maxim sebelum dilakukan analisis sentimen berbasis aspek.

2.1.4 CRISP-DM

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) merupakan kerangka kerja standar yang banyak digunakan dalam penelitian dan pengembangan sistem *data mining*, termasuk pada analisis sentimen dan *text mining*. CRISP-DM menyediakan alur kerja yang sistematis dan iteratif untuk memastikan bahwa proses analisis data dilakukan secara terstruktur, mulai dari pemahaman permasalahan hingga evaluasi hasil model (Ayele, 2020).



Gambar 2. 1 Tahapan CRISP-DM

CRISP-DM terdiri dari enam tahap utama, yaitu:

1. *Business Understanding*

Tahap ini berfokus pada pemahaman tujuan penelitian dan permasalahan yang ingin diselesaikan. Dalam konteks penelitian ini, tahap ini mencakup identifikasi kebutuhan untuk memahami sentimen pelanggan terhadap layanan

Grab, Gojek, dan Maxim secara lebih rinci melalui analisis sentimen berbasis aspek.

2. *Data Understanding*

Tahap ini meliputi pengumpulan data awal, eksplorasi data, serta pemahaman karakteristik dataset. Pada penelitian ini, *data understanding* dilakukan dengan menganalisis ulasan pelanggan aplikasi *ride-hailing* dari Google Play Store yang diperoleh melalui Kaggle.

3. *Data Preparation*

Data preparation merupakan tahap yang paling memakan waktu, mencakup proses pembersihan data, pra-pemrosesan teks, normalisasi, serta pelabelan sentimen dan pengelompokan aspek. Tujuan dari tahapan tersebut adalah untuk memperoleh dataset yang telah terstruktur dan siap digunakan pada tahap pemodelan.

4. *Modeling*

Pada tahap *modeling*, teknik dan algoritma *data mining* diterapkan untuk membangun model. Dalam penelitian ini, tahap modeling mencakup ekstraksi fitur teks menggunakan TF-IDF serta pembangunan model klasifikasi sentimen menggunakan algoritma *Logistic Regression*.

5. *Evaluation*

Kelayakan model yang telah dibangun ditinjau melalui proses evaluasi untuk mengetahui sejauh mana model mampu menjalankan tugas klasifikasi sentimen sesuai dengan tujuan penelitian. Pengukuran performa dilakukan dengan

memanfaatkan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, *F1-score*, *Confusion Matrix*, serta *Stratified 5-Fold Cross Validation*.

6. *Deployment*

Tahap *deployment* berfokus pada pemanfaatan hasil analisis. Dalam konteks penelitian ini, tahap deployment diwujudkan dalam bentuk analisis dan visualisasi distribusi sentimen berbasis aspek yang dapat dimanfaatkan oleh penyedia layanan *ride-hailing* sebagai dasar dalam mengevaluasi dan meningkatkan kualitas layanan.

2.1.5 Pra-pemrosesan Teks (*Text Preprocessing*)

Pra-pemrosesan teks adalah proses awal yang dilakukan untuk mengolah data teks mentah menjadi data yang lebih bersih, terstruktur, dan siap dianalisis secara komputasional. Tahapan ini mencakup *data cleansing*, *case folding*, *tokenization*, *stopword removal*, *stemming*, dan normalisasi kata. Pra-pemrosesan diperlukan karena data ulasan pelanggan sering mengandung *noise* maupun kata tidak baku yang dapat memengaruhi kinerja model. Dalam penelitian ini, pra-pemrosesan dilakukan agar data teks memiliki bentuk yang lebih bersih dan representatif sebelum digunakan pada tahap ekstraksi fitur serta klasifikasi sentimen.

1. *Case Folding*

Case folding dilakukan dengan mengubah semua karakter huruf dalam teks menjadi huruf kecil (*lowercase*) supaya penulisan menjadi seragam dan konsisten. Proses ini membantu mengurangi variasi data yang tidak perlu dan meningkatkan

konsistensi data sebelum proses *tokenization* dan ekstraksi fitur TF-IDF (Palomino & Aider, 2022).

2. *Cleaning*

Cleaning bertujuan untuk menghilangkan karakter dan elemen yang tidak relevan bagi analisis (seperti nama pengguna, spasi berlebih, tanda baca, simbol) yang bersifat *noise*. Tahapan ini diperlukan karena ulasan pelanggan sering mengandung kesalahan penulisan dan format yang tidak seragam sehingga dapat menurunkan kualitas data. Pada penelitian ini, tahap *cleaning* dilakukan agar data ulasan menjadi lebih bersih dan sesuai untuk digunakan pada proses ekstraksi fitur serta klasifikasi sentimen.

3. *Normalization*

Normalization adalah proses mengubah kata-kata tidak baku atau *slang* menjadi bentuk baku agar variasi kata yang memiliki arti serupa dapat diperlakukan sebagai satu entitas. Proses ini penting karena data ulasan pelanggan sering kali memuat bahasa tidak baku dan singkatan. Contoh, kata “ga” diubah menjadi “tidak” (Shaik dkk., 2023; Siino dkk., 2024).

4. *Stopword Removal*

Stopword adalah kata-kata umum dalam suatu bahasa (seperti “dan”, “yang”, “di”, “ke”, dll) yang sering muncul dalam teks tetapi memiliki kontribusi makna yang rendah terhadap analisis (Palomino & Aider, 2022), Penghapusan *stopword* bertujuan untuk menyederhanakan teks dan membantu model lebih berfokus pada kata-kata yang bermakna (Chanda & Pal, 2023; Shaik dkk., 2023).

5. *Stemming* (Bahasa Indonesia)

Stemming bertujuan untuk menyederhanakan kata berimbuhan menjadi bentuk kata dasar dengan menghapus unsur imbuhan, seperti awalan, sisipan, maupun akhiran. (Swain & Nayak, 2018). Contoh, kata “pelayanan” akan diubah menjadi bentuk dasarnya “layan” (Albab dkk., 2023). *Stemming* diterapkan dalam penelitian ini untuk meningkatkan kualitas ekstraksi fitur.

2.1.6 *Inter*-Anotator *Agreement*

Inter-Anotator *Agreement* (IAA) merupakan metode yang digunakan untuk menilai sejauh mana dua atau lebih anotator sepakat dalam proses pelabelan data. Pengukuran ini penting untuk memastikan bahwa hasil anotasi tidak bersifat subjektif dan memiliki tingkat konsistensi yang baik. Dalam konteks analisis sentimen, IAA digunakan untuk mengevaluasi apakah label yang diberikan oleh anotator terhadap data teks memiliki kesamaan persepsi (Moreno-Ortiz dkk., 2019).

Salah satu cara yang sering dipakai untuk menilai IAA adalah Cohen’s Kappa. Metode ini dirancang untuk menghitung tingkat kesepakatan antara dua anotator dengan mempertimbangkan kemungkinan kesepakatan yang muncul secara kebetulan. Dengan demikian, Cohen’s Kappa memberikan hasil pengukuran yang lebih objektif dibandingkan dengan sekadar menghitung persentase kesepakatan sederhana (Moreno-Ortiz dkk., 2019).

Secara matematis, nilai Cohen's Kappa dirumuskan sebagai berikut:

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (1)$$

dengan:

κ : nilai Cohen's Kappa

P_o : *observed agreement* (proporsi kesepakatan aktual antar anotator)

P_e : *expected agreement* (proporsi kesepakatan yang terjadi secara kebetulan)

Nilai P_o diperoleh dari jumlah data yang memiliki label sama antara anotator dibagi dengan total data, sedangkan P_e dihitung berdasarkan probabilitas distribusi masing-masing kategori label yang diberikan oleh anotator.

Cohen's Kappa dinyatakan dalam skala -1 hingga 1. Nilai yang semakin tinggi dan mendekati 1 menggambarkan bahwa anotator memiliki konsistensi penilaian yang baik. Sementara itu, nilai yang berada di sekitar 0 mengindikasikan bahwa kesepakatan antar anotator belum kuat karena dapat terjadi akibat peluang kebetulan. Adapun interpretasi nilai Cohen's Kappa dapat dilihat pada Tabel 2. 1.

Tabel 2. 1 Interpretasi Kappa

Nilai Kappa	**Interpretasi**
< 0.00	Sangat buruk
0.00 – 0.20	Lemah
0.21 – 0.40	Cukup
0.41 – 0.60	Sedang
0.61 – 0.80	Baik
0.81 – 1.00	Sangat baik

Berdasarkan interpretasi Landis dan Koch (1977) tersebut, nilai Kappa yang berada pada kategori baik atau sangat baik menunjukkan bahwa hasil anotasi memiliki konsistensi yang baik dan dapat digunakan sebagai acuan dalam proses evaluasi model.

2.1.7 Ekstraksi Fitur TF-IDF

Data teks tidak dapat langsung digunakan oleh algoritma klasifikasi karena masih berbentuk bahasa alami. Oleh karena itu, diperlukan proses ekstraksi fitur untuk mengubah ulasan menjadi bentuk angka yang dapat dikenali oleh model pembelajaran mesin (Suhaidi dkk., 2021). Dalam penelitian ini, proses tersebut dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF).

Prinsip kerja TF-IDF tidak hanya melihat banyaknya suatu kata muncul, tetapi juga mempertimbangkan seberapa khas kata tersebut dalam kumpulan dokumen. Nilai *Term Frequency* (TF) merepresentasikan frekuensi kemunculan kata pada satu dokumen, sedangkan *Inverse Document Frequency* (IDF) digunakan untuk menurunkan pengaruh kata yang terlalu umum muncul di banyak dokumen (Ayash dkk., 2025). Dengan demikian, kata-kata yang lebih mencerminkan karakteristik suatu ulasan akan memperoleh bobot lebih tinggi dibandingkan kata umum yang kurang membedakan isi dokumen.

Term Frequency menunjukkan frekuensi kemunculan suatu *term* t_k dalam dokumen d_j. Semakin sering suatu kata muncul dalam dokumen, maka semakin besar nilai TF yang dimilikinya (Widianto dkk., 2024).

Secara matematis, *Term Frequency* dirumuskan sebagai berikut:

$$\mathrm{TF}(t_k, d_j) = f(t_k, d_j) \tag{2}$$

Dimana $f(t_k, d_j)$ menyatakan jumlah kemunculan *term* t_k pada dokumen d_j.

IDF merepresentasikan tingkat keunikan suatu *term* berdasarkan sebarannya dalam seluruh dokumen. *Term* yang muncul pada banyak dokumen akan memperoleh bobot lebih rendah karena dianggap terlalu umum dan kurang mampu merepresentasikan karakteristik khusus dari suatu dokumen (Widianto dkk., 2024). IDF dirumuskan sebagai berikut:

$$IDF(t_k) = log \frac{D}{df(t_k)} \tag{3}$$

dengan:

D : adalah jumlah seluruh dokumen dalam korpus,

$df(t_k)$: adalah jumlah dokumen yang mengandung *term* t_k.

Bobot TF-IDF dihitung dengan mengalikan nilai TF dan IDF untuk menunjukkan tingkat kepentingan suatu kata dalam dokumen terhadap keseluruhan korpus. Persamaannya dituliskan sebagai berikut:

$$\text{TF IDF } (t_k, d_j) = Tf(t_k, d_j)\ x\ IDF(t_k) \tag{4}$$

Bobot TF-IDF yang tinggi mengindikasikan bahwa suatu *term* memiliki kontribusi penting dalam suatu dokumen karena kemunculannya cukup dominan

pada dokumen tersebut, namun tidak bersifat umum pada dokumen lain dalam korpus. Dengan demikian, *term* tersebut dapat merepresentasikan karakteristik utama dari dokumen yang dianalisis.

Dalam penelitian ini, metode TF-IDF dimanfaatkan untuk merepresentasikan setiap ulasan pengguna Grab, Gojek, dan Maxim ke dalam bentuk vektor angka. Representasi tersebut kemudian digunakan sebagai data masukan pada proses klasifikasi sentimen.

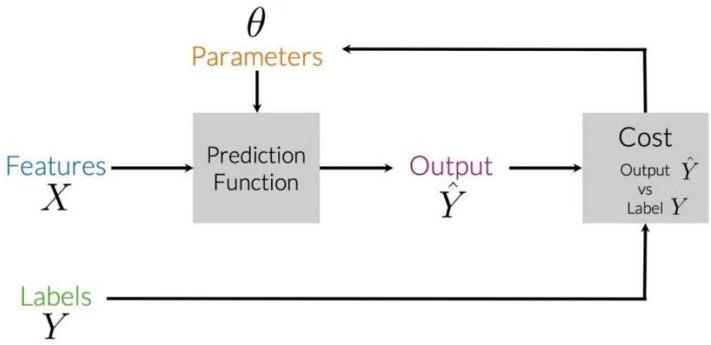
2.1.8 *Logistic Regression*

Logistic Regression bekerja dengan mengubah kombinasi fitur masukan menjadi nilai probabilistik yang merepresentasikan kecenderungan suatu data terhadap kelas tertentu. Dalam prosesnya, fitur-fitur yang telah direpresentasikan secara numerik dihitung secara linier, kemudian hasilnya dipetakan menggunakan fungsi logistik. Karena keluaran yang dihasilkan berada pada rentang 0 hingga 1, metode ini dapat digunakan untuk menentukan peluang keanggotaan data pada suatu kelas (Kaur & Himanshu, 2023; Kravets dkk., 2024).

Pada penelitian analisis sentimen, teks ulasan terlebih dahulu diubah menjadi representasi numerik melalui proses ekstraksi fitur menggunakan TF-IDF. Representasi tersebut kemudian digunakan sebagai masukan bagi model, sedangkan keluaran yang diprediksi berupa kategori sentimen, yaitu positif, negatif, atau netral. (Lu, 2025).

Pada Gambar 2. 2 ditunjukkan alur kerja *Logistic Regression*, yang menggambarkan proses transformasi fitur input X melalui kombinasi linier dengan

parameter θ, kemudian dipetakan menggunakan fungsi *sigmoid* untuk menghasilkan output berupa probabilitas kelas.



Gambar 2. 2 Alur Kerja *Logistic Regression*

Sumber: (Chadha & Aman, 2020)

Secara matematis, *Logistic Regression* diawali dengan perhitungan kombinasi linier dari fitur input sebagai berikut:

$$z = \theta_0 + \theta_{1X_1} + \theta_{2X_2} + \cdots + \theta_{nX_n} \tag{5}$$

Atau dalam bentuk vektor:

$$z = \theta^T X \tag{6}$$

dengan:

$X = (X_1,\ X_2, \dots ,\ X_n)$: merupakan vektor fitur hasil ekstraksi TF-IDF

$\theta = (\theta_1,\ \theta_2, \dots ,\ \theta_n)$: merupakan parameter atau bobot model

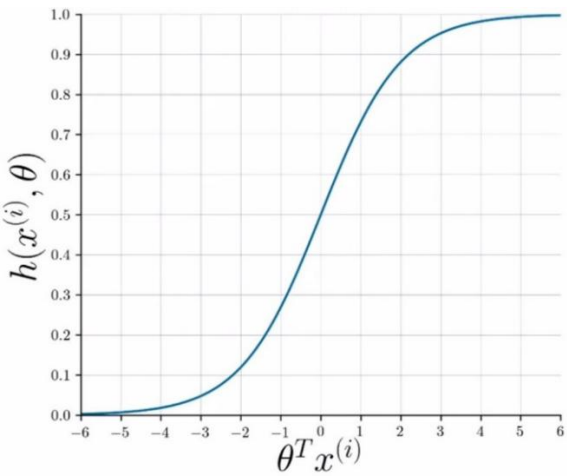
z : merupakan skor linier

Nilai *z* selanjutnya dipetakan menggunakan fungsi logistik (*sigmoid*) untuk menghasilkan probabilitas kelas, yang dirumuskan sebagai berikut:

$$h(X^{(i)};\theta) = \frac{1}{1 + e^{-\theta^T X^{(i)}}} \quad (7)$$

Fungsi *sigmoid* digunakan untuk mengubah hasil perhitungan linier menjadi nilai probabilitas dengan rentang antara 0 sampai 1. Nilai tersebut kemudian menunjukkan peluang suatu data termasuk ke dalam kelas tertentu. Semakin besar nilai $-\theta^T X$, maka probabilitas yang dihasilkan akan semakin mendekati 1, dan sebaliknya akan mendekati 0 apabila nilainya semakin kecil.

Bentuk kurva fungsi *sigmoid* ditunjukkan pada Gambar 2. 3, yang memperlihatkan bagaimana nilai input dari kombinasi linier fitur ditransformasikan menjadi probabilitas kelas.

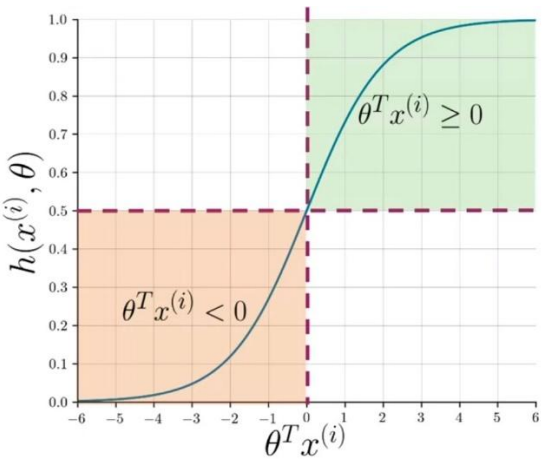


Gambar 2. 3 Kurva *Sigmoid*

Sumber: (Chadha & Aman, 2020)

Dalam klasifikasi biner, probabilitas hasil fungsi *sigmoid* dibandingkan dengan nilai ambang (*threshold*), yang umumnya ditetapkan sebesar 0,5 (Chadha & Aman, 2020). Jika nilai probabilitas memenuhi atau melampaui *threshold* tersebut, maka data diprediksi sebagai kelas positif, sedangkan jika lebih rendah

dari *threshold*, data diprediksi sebagai kelas negatif. Mekanisme klasifikasi ini ditunjukkan pada Gambar 2. 4.



Gambar 2. 4 Ilustrasi Aturan Klasifikasi

Sumber: (Chadha & Aman, 2020)

Dalam klasifikasi biner, probabilitas hasil fungsi *sigmoid* dibandingkan dengan nilai ambang tertentu. Jika probabilitas berada di atas ambang yang ditentukan, data diklasifikasikan ke salah satu kelas, sedangkan jika berada di bawah ambang tersebut, data diklasifikasikan ke kelas lainnya.

Pada permasalahan dengan jumlah kelas lebih dari dua, *Logistic Regression* dapat diterapkan menggunakan strategi *one-vs-rest* (OvR). Strategi ini bekerja dengan membangun beberapa model klasifikasi biner, di mana setiap kelas dibandingkan dengan gabungan kelas lainnya. Dengan pendekatan tersebut, model menghasilkan skor prediksi untuk masing-masing kelas, kemudian kelas dengan skor tertinggi dipilih sebagai label akhir. Pendekatan OvR umum digunakan pada implementasi *Logistic Regression* tertentu, salah satunya ketika menggunakan

solver liblinear, karena *solver* tersebut lebih sesuai untuk skema klasifikasi biner yang diperluas ke permasalahan multikelas.

Dalam penelitian ini, *Logistic Regression* digunakan untuk mengklasifikasikan sentimen ke dalam tiga kelas, yaitu negatif, netral, dan positif. Oleh karena itu, pendekatan *one-vs-rest* digunakan agar model dapat membedakan setiap kelas sentimen melalui beberapa proses klasifikasi biner.

Logistic Regression memiliki beberapa keunggulan yang menjadikannya sesuai untuk analisis sentimen berbasis teks, antara lain:

1. Mampu memproses data teks dengan jumlah fitur yang besar secara efisien, seperti representasi teks yang dihasilkan melalui TF-IDF.
2. Bersifat *interpretable*, sehingga kontribusi setiap fitur terhadap hasil klasifikasi dapat dianalisis.
3. Menunjukkan kinerja yang stabil dan kompetitif pada berbagai penelitian analisis sentimen.

Dengan mempertimbangkan efisiensi komputasi, kemudahan interpretasi, serta kemampuannya dalam menangani data teks berdimensi tinggi, *Logistic Regression* digunakan sebagai metode klasifikasi pada penelitian ini untuk mengidentifikasi sentimen berbasis aspek pada ulasan pengguna aplikasi Grab, Gojek, dan Maxim.

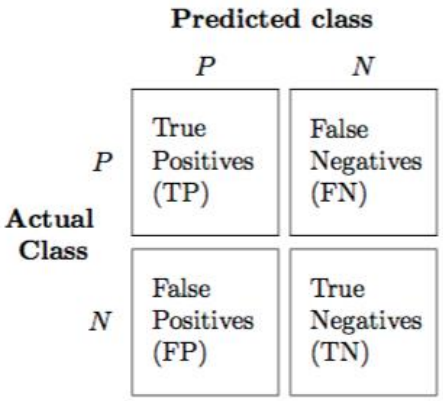
2.1.9 Evaluasi Kinerja Model

Tahap evaluasi bertujuan untuk mengukur seberapa efektif model dalam mengklasifikasikan sentimen berdasarkan label acuan yang telah ditentukan. Evaluasi pada penelitian ini dilakukan dengan menilai kemampuan model dalam

membedakan tiga kelas sentimen, yaitu positif, netral, dan negatif. Hasil evaluasi tersebut digunakan untuk membandingkan performa setiap skenario pemodelan, sehingga dapat dipilih skenario yang memberikan hasil klasifikasi paling baik.

1. *Confusion Matrix*

Confusion Matrix digunakan untuk mengevaluasi kesesuaian antara label sebenarnya dan hasil prediksi dari model. Matriks ini berisi empat jenis hasil klasifikasi, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Keempat elemen tersebut menjadi landasan untuk menilai kemampuan model dalam melakukan klasifikasi dengan lebih mendalam. Gambar 2. 5 menunjukkan Ilustrasi *Confusion Matrix* dengan dua kelas.



Gambar 2. 5 *Confusion Matrix*

True Positives (TP) : Kondisi ketika data yang memang termasuk kelas positif berhasil dikenali oleh model sebagai kelas positif.

True Negatives (TN) : Kondisi ketika data yang sebenarnya termasuk

kelas negatif berhasil diklasifikasikan oleh model sebagai kelas negatif.

False Positives (FP) : Kondisi ketika data yang sebenarnya berasal dari kelas negatif, tetapi keliru diprediksi oleh model sebagai kelas positif.

False Negatives (FN) : Kondisi ketika data yang sebenarnya termasuk kelas positif, tetapi keliru diklasifikasikan oleh model sebagai kelas negatif.

Nilai TP, TN, FP, dan FN pada *Confusion Matrix* menjadi dasar untuk menghitung metrik evaluasi, seperti *accuracy*, *precision*, *recall*, dan F1-*score*. Metrik ini dimanfaatkan untuk memahami kemampuan model dalam melakukan klasifikasi secara lebih menyeluruh.

2. *Accuracy*

Accuracy merepresentasikan persentase keberhasilan model dalam memberikan prediksi yang benar pada data *testing*. Semakin besar nilai *accuracy*, semakin baik performa model dalam menghasilkan prediksi yang sesuai dengan label sebenarnya pada seluruh data uji (Naidu dkk., 2023). Nilai *accuracy* dapat dihitung menggunakan persamaan berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

3. *Precision*

Precision menunjukkan seberapa tepat model dalam memprediksi kelas positif, yaitu dengan membandingkan prediksi positif yang benar terhadap seluruh data yang diprediksi positif (Naidu dkk., 2023). *Precision* sangat penting ketika kesalahan prediksi positif memiliki dampak yang signifikan. Nilai *precision* dapat dihitung menggunakan persamaan:

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

4. *Recall*

Recall menunjukkan sejauh mana model mampu mengidentifikasi data aktual pada suatu kelas secara tepat. (Burduk, 2023). Nilai *recall* yang semakin besar menunjukkan bahwa model memiliki kemampuan yang baik dalam mengenali data yang benar-benar termasuk ke dalam kelas positif. Perhitungan nilai *recall* ditunjukkan pada persamaan berikut:

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

5. F1-*Score*

F1-*score* merupakan ukuran evaluasi yang menghitung keseimbangan antara *precision* dan *recall*. Metrik ini digunakan untuk menilai performa model secara lebih proporsional, khususnya pada kondisi ketika jumlah data pada setiap kelas tidak seimbang (Naidu dkk., 2023). Nilai F1-*score* dapat dihitung menggunakan persamaan berikut:

$$F1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (11)$$

6. *Macro Precision, Macro Recall*, dan *Macro* F1-*Score*

Pada klasifikasi multikelas, nilai *precision*, *recall*, dan F1-*score* dapat dihitung menggunakan pendekatan *Macro average*. *Macro average* menghitung nilai metrik pada setiap kelas secara terpisah, kemudian mengambil rata-rata dari seluruh kelas tanpa mempertimbangkan jumlah data pada masing-masing kelas. Dengan demikian, setiap kelas memiliki bobot yang sama dalam perhitungan akhir. Secara matematis, *Macro average* dapat dirumuskan sebagai berikut:

$$Macro\ Precision = \frac{1}{\mathrm{K}}\sum_{\mathrm{i}=1}^{\mathrm{K}} \mathrm{Precision_i} \quad (12)$$

$$Macro\ Recall = \frac{1}{\mathrm{K}}\sum_{\mathrm{i}=1}^{\mathrm{K}} \mathrm{Recall_i}$$

$$Macro\ \mathrm{F1} = \frac{1}{\mathrm{K}}\sum_{\mathrm{i}=1}^{\mathrm{K}} \mathrm{F1_i}$$

dengan:

K : jumlah kelas

$(Precision_i), (Recall_i), dan\ (F1_i)$: nilai *precision*, *recall*, dan F1-*score* pada kelas ke-(i)

Macro precision digunakan untuk mengetahui rata-rata ketepatan prediksi model pada seluruh kelas, sedangkan *Macro recall* digunakan untuk mengetahui rata-rata kemampuan model dalam mengenali data dari setiap kelas. Sementara itu,

Macro F1-*score* merupakan rata-rata F1-*score* dari seluruh kelas yang menggambarkan keseimbangan antara *precision* dan *recall* pada masing-masing kelas.

Macro F1-*score* digunakan sebagai metrik utama dalam penelitian ini karena data sentimen memiliki distribusi kelas yang tidak seimbang. Pada kondisi tersebut, *accuracy* dapat terlihat tinggi apabila model lebih banyak benar dalam memprediksi kelas mayoritas, tetapi belum tentu mampu mengenali kelas minoritas dengan baik. *Macro* F1-*score* lebih sesuai digunakan karena memberikan bobot yang sama pada setiap kelas, baik kelas mayoritas maupun kelas minoritas. Dengan demikian, performa model pada kelas negatif, netral, dan positif dapat dievaluasi secara lebih seimbang.

7. *Weighted* F1-*Score*

Weighted F1-*score* merupakan rata-rata F1-*score* dari setiap kelas dengan mempertimbangkan jumlah data atau *support* pada masing-masing kelas. Kelas dengan jumlah data lebih besar akan memberikan pengaruh lebih besar terhadap nilai akhir. Oleh karena itu, *weighted* F1-*score* dapat menunjukkan performa model secara keseluruhan dengan memperhatikan distribusi data aktual.

Weighted F1-*score* dapat dirumuskan sebagai berikut:

$$Weighted\ \mathrm{F1} = \sum_{\mathrm{i}=1}^{\mathrm{K}} \frac{\mathrm{n_i}}{\mathrm{N}} \mathrm{F1_i} \tag{13}$$

dengan:

$\mathrm{n_i}$: jumlah data pada kelas ke-(i)

N : jumlah seluruh data

$F1_i$: nilai F1-*score* pada kelas ke-(i)

8. *Classification Report*

Classification report merupakan ringkasan hasil evaluasi model yang menampilkan nilai *precision, recall,* F1-*score*, dan *support* untuk setiap kelas. *Support* menunjukkan jumlah data sebenarnya pada masing-masing kelas dalam data pengujian. Melalui *classification report*, performa model dapat dianalisis secara lebih rinci karena hasil evaluasi tidak hanya ditampilkan dalam bentuk nilai rata-rata, tetapi juga berdasarkan setiap kelas.

Dalam penelitian ini, *classification report* digunakan untuk melihat kemampuan model dalam mengklasifikasikan sentimen negatif, netral, dan positif. Informasi ini penting karena setiap kelas memiliki karakteristik dan jumlah data yang berbeda, sehingga evaluasi per kelas diperlukan untuk mengetahui apakah model bekerja secara seimbang atau cenderung hanya baik pada kelas tertentu.

2.2 Penelitian Terkait

2.2.1 *State of the Art*

Bagian ini membahas beberapa penelitian terdahulu yang memiliki keterkaitan dengan topik penelitian, baik dari segi metode, objek, maupun pendekatan analisis yang digunakan. Ringkasan penelitian terdahulu tersebut disajikan pada Tabel 2. 2.

Tabel 2. 2 *State of the Art*

No	Peneliti	Judul	Permasalahan	Algoritma	Hasil Penelitian
1	(Iqrom dkk., 2025)	Sentiment Analysis of Gojek, Grab, and Maxim Applications Using *Support* Vector Machine Algorithm	Tingginya persaingan layanan transportasi *online* di Indonesia menuntut perusahaan untuk memahami persepsi pengguna secara spesifik. Diperlukan analisis terhadap ribuan ulasan yang mencakup aspek pengemudi, harga, dan aplikasi untuk menentukan strategi peningkatan layanan yang tepat bagi masing-masing platform.	SVM, TF-IDF	Penelitian ini melakukan analisis sentimen terhadap review pengguna aplikasi Gojek, Grab, dan Maxim yang bersumber dari Google Play Store. Temuan penelitian tersebut mengindikasikan bahwa Maxim mendapatkan proporsi sentimen positif paling tinggi, yaitu sebesar 42,45%, yang dipengaruhi oleh persepsi pengguna terhadap aspek harga. Sementara itu, dari sisi performa klasifikasi, Gojek memperoleh akurasi tertinggi sebesar 94%, diikuti oleh Maxim sebesar 90% dan Grab sebesar 87%.
2	(Yutika dkk., 2021)	Analisis Sentimen Berbasis Aspek pada Review Female Daily	Penelitian ini mengekstraksi informasi berharga berdasarkan aspek tertentu dari teks yang tidak terstruktur untuk	TF-IDF, Complement Naïve Bayes (CNB)	Penelitian tersebut menangani ketidakseimbangan data dengan menerapkan metode Complement Naïve Bayes. Hasil pengujian,

No	Peneliti	Judul	Permasalahan	Algoritma	Hasil Penelitian
			membantu konsumen dalam memilih produk kosmetik.		performa terbaik diperoleh pada scenario di mana data diterjemahkan terlebih dahulu ke bahasa Inggris, lalu kembali ke bahasa Indonesia, tanpa melalui tahap penghapusan stopword. Pada skenario tersebut, model menghasilkan nilai F1-*score* sebesar 62,81%.
3	(Ayub dkk., 2023)	Aspect Extraction Approach for Sentiment Analysis Using *Keyword*s	Sebagian besar pendekatan analisis sentimen hanya berfokus pada aspek eksplisit dan sering mengabaikan aspek implisit dalam ulasan konsumen. Hal ini menyebabkan kurangnya detail informasi bagi teknisi kebutuhan (*requirement engineers*) dalam memahami keinginan konsumen secara mendalam dari situs e-commerce.	*Keyword*s-Based Aspect Extraction, WordNet, SentiWordNet, TF-IDF (frequency-based), Naïve Bayes	Metode yang diusulkan mampu menangkap aspek eksplisit dan implisit secara bersamaan dengan bantuan kemiripan semantik. Pengujian pada delapan dataset menunjukkan performa yang lebih unggul dibandingkan algoritma standar seperti CRF dan Double Propagation dalam hal ekstraksi aspek yang akurat.

No	Peneliti	Judul	Permasalahan	Algoritma	Hasil Penelitian
4	(Pamungkas dkk., 2025)	Comparison of Machine Learning Models for Classifying Consumer Sentiment of Coffee Shops	Studi ini mengevaluasi sejumlah algoritma *machine learning* untuk menentukan metode yang memiliki performa paling konsisten dalam mengklasifikasikan sentimen dari data teks dari media sosial.	*Logistic Regression*, SVM, Naïve Bayes, TF-IDF	Hasil evaluasi mengindikasikan bahwa LR mencapai kinerja terbaik dengan tingkat akurasi 79%. Selain menghasilkan akurasi tertinggi, LR juga menunjukkan kinerja yang relatif stabil pada metrik *precision*, *recall*, dan F1-*score*. Performa tersebut lebih seimbang dibandingkan SVM yang memperoleh akurasi 78% dan *Naïve Bayes* sebesar 75% dalam klasifikasi sentimen multikelas.
5	(Ghadage dkk., 2024)	Utilizing Aspect-Based Sentiment Analysis to evaluate and enhance product quality	Analisis sentimen tradisional sering kali hanya memberikan gambaran umum tanpa merinci aspek spesifik produk yang disukai atau dikeluhkan pelanggan. Hal ini menyulitkan bisnis dalam melakukan peningkatan kualitas yang tepat sasaran pada fitur atau layanan tertentu.	*Logistic Regression*, Random Forest, Decision Tree	Pendekatan ABSA berhasil mendeteksi fitur utama produk dan sikap pelanggan terhadap setiap fitur tersebut secara spesifik. Hasilnya memungkinkan bisnis untuk melakukan pemasaran yang lebih akurat, meningkatkan layanan pelanggan, dan memahami kasus dukungan pelanggan potensial dengan lebih baik.

No	Peneliti	Judul	Permasalahan	Algoritma	Hasil Penelitian
6	(Maulana dkk., 2023)	Comparison of *Logistic Regression*, MultinomialNB, SVM, and K-NN Methods on Sentiment Analysis of Gojek App Reviews on The Google Play Store	Penelitian ini membandingkan berbagai metode klasifikasi guna menemukan algoritma yang paling presisi dalam memilah ulasan menjadi kategori positif atau negatif.	*Logistic Regression*, MultinomialNB, SVM, K-NN	Hasil penelitian menunjukkan bahwa LR menjadi metode dengan kinerja terbaik dibandingkan metode lainnya, dengan tingkat akurasi 82,45%, *recall* 82,49%, serta presisi 82,45%. Capaian tersebut menunjukkan bahwa LR mampu melakukan klasifikasi sentimen ulasan aplikasi secara stabil, sehingga bisa dijadikan salah satu metode yang relevan untuk analisis sentimen pada aplikasi *mobile*.
7	(Handani dkk., 2019)	Sentiment Analysis for Go-Jek on Google Play Store	Penelitian ini mengkaji bagaimana opini pengguna terhadap aplikasi Go-Jek dapat dianalisis secara otomatis dari ulasan daring.	Naïve Bayes Classifier (NBC)	Penelitian yang dilakukan pada ulasan akhir tahun 2017 menunjukkan dominasi sentimen negatif. Hasil ini menjadi bahan evaluasi bagi Go-Jek untuk memperbaiki kualitas layanan, meskipun catatan menunjukkan ketidakpuasan dapat dipengaruhi faktor eksternal di luar teknis aplikasi

No	Peneliti	Judul	Permasalahan	Algoritma	Hasil Penelitian
8	(Taparia & Bagla, 2020)	Sentiment Analysis: Predicting Product Reviews' *Rating*s using *Online* Customer Reviews	Penelitian ini membahas bagaimana memprediksi *rating* produk secara akurat dengan mengekstraksi fitur teks dan memahami korelasi antara sentimen tertulis dengan skor numerik yang diberikan pelanggan.	*Logistic Regression*, Multinomial Naïve Bayes, Linear SVC TF-IDF, N-gram	LR mengungguli model lainnya dengan akurasi 54,1%, menunjukkan peningkatan 170% dibandingkan tebakan acak (*Baseline* 20%) pada klasifikasi 5 kelas. Penelitian mengungkap bahwa polaritas teks dan panjang ulasan memiliki pengaruh signifikan terhadap akurasi prediksi *rating*.
9	(Kadir & Fairuzabadi, 2025)	Analisis Sentimen Ulasan Shopee di Google Play dengan TF-IDF dan *Logistic Regression*	Peningkatan penggunaan layanan belanja *online* menyebabkan ulasan pengguna Shopee di Google Play Store semakin banyak dan perlu dianalisis sebagai bahan evaluasi aplikasi. Tantangan utama penelitian tersebut adalah ketidakseimbangan distribusi data sentimen antara kelas positif, negatif, dan netral, yang dapat memengaruhi hasil klasifikasi.	*Logistic Regression*, TF-IDF	Penerapan *Random Oversampling* pada penelitian tersebut membantu *Logistic Regression* mencapai akurasi 85,11% dan *Macro*-F1 sebesar 0,58. Performa terbaik terlihat pada kelas positif dengan F1-*score* 0,92, sedangkan prediksi terhadap kelas netral masih belum optimal dan menjadi kendala utama dalam klasifikasi sentimen.

No	Peneliti	Judul	Permasalahan	Algoritma	Hasil Penelitian
10	(Deepa dkk., 2023)	Sentiment Analysis on Food-Reviews Dataset	Penelitian ini mengulas berbagai pendekatan analisis sentimen dan tantangan pemrosesan teks opini.	Multinomial Naïve Bayes, *Logistic Regression*	Studi ini menegaskan bahwa metode pembelajaran mesin klasik seperti LR tetap relevan karena efisiensi dan interpretabilitasnya
11	(Dhanalakshmi dkk., 2023)	Sentiment Analysis Using VADER and *Logistic Regression*	Penelitian ini membandingkan efektivitas metode berbasis leksikon dan algoritma pembelajaran mesin untuk menemukan pendekatan yang paling presisi dalam klasifikasi emosi teks.	VADER (*Lexicon-based*), *Logistic Regression*	Hasil evaluasi pada dataset tweet maskapai menunjukkan bahwa LR mengungguli VADER dengan tingkat akurasi sebesar 79%. Hal ini mengindikasikan bahwa pendekatan berbasis pelatihan data (machine learning) lebih efektif untuk konteks ulasan layanan maskapai di Twitter dibandingkan aturan leksikon yang kaku.
12	(Ekanata & Budi, 2018)	Mobile application review classification for the Indonesian language using machine learning approach	Pengembang aplikasi mobile di Indonesia seringkali kewalahan memilah ribuan ulasan pengguna yang tidak terstruktur untuk menemukan laporan bug atau permintaan fitur. Tantangan utamanya adalah belum adanya metode klasifikasi otomatis yang	*Logistic Regression*, Naïve Bayes, SVM, Decision Tree	*Logistic Regression* memberikan performa terbaik dengan F-measure mencapai 85% ketika dikombinasikan dengan fitur unigram, panjang kalimat, dan skor sentimen. Fitur unigram terbukti sebagai komponen paling prediktif, sedangkan fitur bigram dan *rating* bintang justru

No	Peneliti	Judul	Permasalahan	Algoritma	Hasil Penelitian
			optimal khusus untuk karakteristik bahasa Indonesia dalam ulasan aplikasi		memberikan dampak negatif pada kinerja klasifikasi.
13	(Zhikri & Istiono, 2024)	Handling Class *Imbalance* for Indonesian Twitter Sentiment Analysis A Comparative Study of Algorithms	Penelitian ini mengkaji masalah ketidakseimbangan kelas (class *imbalance*) yang ekstrem antara opini positif dan negatif. Permasalahan ini dapat menurunkan akurasi model jika tidak ditangani dengan teknik pengambilan sampel yang tepat sebelum proses klasifikasi.	*Logistic Regression* (LR), Multinomial Naïve Bayes (MNB), SMOTE	Hasil evaluasi menunjukkan bahwa LR memberikan performa terbaik dengan akurasi 86%, sementara MNB mencapai akurasi 74% pada skenario pembagian data 80:20. Penggunaan SMOTE mampu meningkatkan kinerja MNB secara cukup besar, yaitu hingga 14%. Namun, secara keseluruhan LR tetap menunjukkan performa yang lebih stabil dan unggul dalam klasifikasi sentimen pada data Twitter.
14	(Cahyani & Patasik, 2021)	Performance comparison of TF-IDF and Word2Vec models for emotion text classification	Penelitian ini membandingkan teknik ekstraksi fitur yang tepat antara TF-IDF (dimensi tinggi) dan Word2Vec (dimensi rendah) untuk klasifikasi emosi.	SVM, MNB, TF-IDF, Word2Vec (Skip-gram)	Model TF-IDF dapat menghasilkan akurasi yang tinggi meskipun pada kelas emosi minoritas (seperti sedih atau takut), sedangkan Word2Vec gagal mengenali emosi dengan jumlah data yang kecil.

No	Peneliti	Judul	Permasalahan	Algoritma	Hasil Penelitian
15	(Karo dkk., 2023)	Analisis Sentimen Ulasan Aplikasi Info BMKG di Google Play Menggunakan TF-IDF dan *Support* Vector Machine	Studi ini mengkaji bentuk keluhan yang disampaikan oleh pengguna secara sistematis guna memberikan rekomendasi perbaikan bagi pengembang aplikasi	*Support* Vector Machine (SVM), TF-IDF	Berdasarkan 2500 ulasan yang dikumpulkan, ditemukan bahwa 66% ulasan bersifat positif. Skenario data split 75%:25% menggunakan SVM menghasilkan akurasi sebesar 79%, dengan topik utama ulasan mencakup masalah aplikasi, pembaruan (update), dan informasi gempa/cuaca

2.2.2 Matriks Penelitian

Tabel 2. 3 Matriks Penelitian

No	Nama Penulis (Tahun)	Judul	Ruang Lingkup Penelitian							
			Framework/Metode				Tools		Tujuan	
			SVM	NB	LR	DL/Lain	Python	Lain	Umum	Aspek
1	(Iqrom dkk., 2025)	Sentiment Analysis of Gojek, Grab, and Maxim Applications Using *Support* Vector Machine Algorithm	√	-	-	-	√	-	-	√
2	(Yutika dkk., 2021)	Analisis Sentimen Berbasis Aspek pada Review Female Daily	-	√	-	-	√	-	-	√
3	(Ayub dkk., 2023)	Aspect Extraction Approach for Sentiment Analysis Using *Keyword*s	-	-	-	√	-	√	-	√
4	(Pamungkas dkk., 2025)	Comparison of Machine Learning Models for Classifying Consumer Sentiment of Coffee Shops	√	√	√	-	√	-	√	-
5	(Ghadage dkk., 2024)	Utilizing Aspect-Based Sentiment Analysis to evaluate and enhance product quality	-	-	√	√	√	-	-	√

No	Nama Penulis (Tahun)	Judul	Ruang Lingkup Penelitian							
			Framework/Metode				Tools		Tujuan	
			SVM	NB	LR	DL/Lain	Python	Lain	Umum	Aspek
6	(Maulana dkk., 2023)	Comparison of *Logistic Regression*, MultinomialNB, SVM, and K-NN Methods on Sentiment Analysis of Gojek App Reviews on The Google Play Store	√	√	√	√	√	-	√	-
7	(Handani dkk., 2019)	Sentiment Analysis for Go-Jek on Google Play Store	-	√	-	-	√	-	√	-
8	(Taparia & Bagla, 2020)	Sentiment Analysis: Predicting Product Reviews' *Rating*s using *Online* Customer Reviews	-	√	√	√	√	-	√	-
9	(Kadir & Fairuzabadi, 2025)	Analisis Sentimen Ulasan Shopee di Google Play dengan TF-IDF dan *Logistic Regression*	-	-	√	-	√	-	√	-
10	(Deepa dkk., 2023)	Sentiment Analysis on Food-Reviews Dataset	-	√	√	-	√	-	√	-
11	(Dhanalakshmi dkk., 2023)	Sentiment Analysis Using VADER and *Logistic Regression*	-	-	√	√	√	-	√	-

No	Nama Penulis (Tahun)	Judul	Ruang Lingkup Penelitian							
			Framework/Metode				Tools		Tujuan	
			SVM	NB	LR	DL/Lain	Python	Lain	Umum	Aspek
12	(Ekanata & Budi, 2018)	Mobile application review classification for the Indonesian language using machine learning approach	√	√	√	√	√	-	√	-
13	(Zhikri & Istiono, 2024)	Handling Class *Imbalance* for Indonesian Twitter Sentiment Analysis A Comparative Study of Algorithms	-	√	√	-	√	-	√	-
14	(Cahyani & Patasik, 2021)	Performance comparison of TF-IDF and Word2Vec models for emotion text classification	√	√	-	-	√	-	√	-
15	(Karo dkk., 2023)	Analisis Sentimen Ulasan Aplikasi Info BMKG di Google Play Menggunakan TF-IDF dan *Support* Vector Machine	√	-	-	-	√	-	√	-
16	Penelitian yang Akan Dilakukan	Analisis Sentimen Berbasis Aspek Review Pelanggan Pada Aplikasi Transportasi *Online* Menggunakan *Logistic Regression* Dengan Ekstraksi Fitur TF-IDF	-	-	√	-	√	-	-	√

Berdasarkan analisis terhadap sejumlah penelitian relevan, dapat diketahui bahwa sebagian besar penelitian analisis sentimen pada aplikasi *ride-hailing* masih berfokus pada klasifikasi sentimen secara umum tanpa memetakan sentimen ke dalam aspek layanan tertentu.

Penelitian oleh Yutika dkk. (2021) telah menggunakan pendekatan *Aspect-Based Sentiment Analysis* (ABSA), namun objek penelitian terbatas pada satu platform dan belum dilakukan secara komparatif antar aplikasi dalam konteks layanan *ride-hailing*. Selain itu, algoritma yang digunakan masih didominasi oleh *Naïve Bayes* atau metode klasifikasi lainnya, sehingga belum mengeksplorasi penggunaan *Logistic Regression* pada konteks ABSA untuk ulasan aplikasi transportasi berbasis aplikasi.

Dari sisi sumber data, beberapa penelitian sebelumnya juga cenderung menggunakan dataset terbatas atau hanya berfokus pada satu aplikasi, sehingga belum memberikan gambaran perbandingan persepsi pelanggan antarplatform secara menyeluruh. Padahal, perbedaan karakteristik layanan antara Grab, Gojek, dan Maxim berpotensi menghasilkan distribusi sentimen yang berbeda pada setiap aspek layanan.

Berdasarkan perbandingan tersebut, terdapat celah penelitian berupa keterbatasan studi yang mengombinasikan pendekatan analisis sentimen berbasis aspek dengan algoritma *Logistic Regression* pada konteks komparatif tiga aplikasi *ride-hailing* sekaligus. Berdasarkan celah penelitian tersebut, penelitian ini dilakukan dengan menerapkan analisis sentimen berbasis aspek pada ulasan pengguna aplikasi Grab, Gojek, dan Maxim. Proses klasifikasi dilakukan

menggunakan ekstraksi fitur TF-IDF dan algoritma *Logistic Regression* agar hasil analisis tidak hanya menunjukkan polaritas sentimen secara umum, tetapi juga mampu menggambarkan persepsi pengguna pada setiap aspek layanan secara lebih rinci dan komparatif.