

BAB II LANDASAN TEORI

2.1 Analisis Sentimen

Analisis sentimen, yang juga dikenal sebagai penggalian opini, adalah proses untuk mengidentifikasi, memahami, dan mengklasifikasikan emosi atau opini yang disampaikan oleh pengguna terhadap berbagai layanan, seperti film, produk, acara, atau atribut lainnya. Proses ini bertujuan untuk menentukan apakah sentimen yang diungkapkan bersifat positif, negatif, atau netral. Data yang digunakan untuk analisis ini biasanya berasal dari berbagai saluran komunikasi sosial, seperti situs ulasan, forum diskusi, blog, mikroblog, Twitter, dan platform serupa. Bidang penelitian ini telah menjadi sangat populer karena semakin banyaknya data opini yang tersedia dalam format digital. Data ini sering kali mencakup ulasan atau pendapat yang bermanfaat bagi kehidupan sehari-hari, seperti ulasan produk atau layanan tertentu (Pandya & Mehta, 2020).

Dalam analisis sentimen, penggalian data digunakan untuk mengekstrak informasi opini yang relevan dan menghasilkan keluaran yang bermanfaat. Pendekatan ini sering kali melibatkan ringkasan berdasarkan emosi atau ekstraksi perasaan dan pandangan (Hasibuan et al., 2023). Analisis sentimen, yang juga disebut *opinion mining*, menggunakan teknologi *Natural Language Processing* (NLP) untuk mengidentifikasi emosi dan opini publik tentang topik tertentu, baik itu produk maupun layanan. Karena popularitasnya, analisis ini sangat berguna dalam berbagai konteks, seperti survei dan kampanye iklan, dengan membantu menilai keberhasilan produk atau layanan berdasarkan pendapat atau saran pengguna. Selain itu, analisis ini dapat memberikan wawasan tentang hal-hal yang disukai atau tidak disukai konsumen, sehingga perusahaan dapat memahami fitur produk mereka secara lebih mendalam.

Berbagai metode dapat digunakan untuk melakukan analisis sentimen pada media sosial seperti Twitter. Banyak penelitian telah mengembangkan metode yang sudah ada menjadi teknik-teknik baru

dengan tingkat akurasi yang lebih tinggi. Sebagai contoh, beberapa penelitian membandingkan metode seperti *Naïve Bayes*, *K-Nearest Neighbor* (KNN), dan kombinasi *K-Means* serta *Learning Vector Quantization (LVQ)* dalam mengklasifikasikan kategori buku berbahasa Indonesia. Penelitian tersebut menggunakan 200 buku sebagai data, dengan 150 buku sebagai data pelatihan dan 50 buku sebagai data pengujian. Hasil penelitian menunjukkan bahwa metode KNN mencapai akurasi sebesar 96%, sementara metode *Naïve Bayes* berhasil mencapai akurasi tertinggi, yaitu 98%. Sebaliknya, kombinasi *K-Means* dan LVQ menghasilkan akurasi sebesar 92,2%. Dari perbandingan ini, metode *Naïve Bayes* terbukti memiliki performa terbaik dalam klasifikasi data (Darujati & Bimo Gumelar, 2012).

2.2 *Text Mining*

Text mining adalah algoritma yang dirancang untuk menggali data guna memenuhi kebutuhan informasi dengan menerapkan metode pembelajaran mesin (*machine learning*), data mining, pemrosesan bahasa alami (*natural language processing*), manajemen pengetahuan, serta teknik pencarian informasi. Proses *text mining* melibatkan langkah-langkah awal seperti pengkategorian teks, ekstraksi informasi, dan pengambilan kata-kata penting. Teknik ini berguna untuk mengekstraksi informasi dari sumber data melalui identifikasi dan eksplorasi pola-pola yang relevan (Ridwansyah, 2022). *Text Mining* merupakan metode yang digunakan untuk menangani masalah seperti klasifikasi, *information retrieval*, *information extraction*, dan *clustering*. Teknik ini membantu menganalisis data teks yang tidak terstruktur untuk menemukan pola dan informasi penting. Dengan memanfaatkan teknologi *machine learning* dan pemrosesan bahasa alami, *text mining* memungkinkan pengolahan data dalam jumlah besar secara efisien, sehingga mendukung pengambilan keputusan berbasis data di berbagai bidang (Supriyatna & Fahrudin, 2024).

Secara umum, proses kerja *Text Mining* banyak mengadopsi pendekatan dari penelitian data mining. Namun, perbedaan mendasarnya terletak pada sumber polanya. *Text Mining* memanfaatkan pola yang diekstraksi dari teks dalam bahasa alami yang tidak terstruktur, seperti dokumen atau ulasan pengguna, sementara data mining menggunakan pola dari basis data yang terstruktur. Tahapan algoritma *Text Mining* biasanya dimulai dengan praproses teks, yang meliputi pembersihan data, penghapusan kata yang tidak relevan, dan normalisasi teks. Proses ini bertujuan untuk meningkatkan akurasi dan efisiensi analisis selanjutnya (Fudholi et al., 2024).

2.3 Google Play Store

Google Play Store adalah *platform* distribusi digital yang dibuat oleh *Google* untuk perangkat Android. Sebagai toko aplikasi resmi dalam ekosistem Android, Google Play Store menawarkan beragam konten digital, seperti aplikasi, game, buku, film, dan musik. Platform ini memungkinkan pengguna untuk mencari, mengunduh, dan memperbarui aplikasi dengan mudah dan aman, menjadikannya salah satu layanan distribusi aplikasi terbesar di dunia (Maniah et al., 2023). Google Play Store juga menjadi penghubung antara pengembang aplikasi dan pengguna. Pengembang dapat mempublikasikan aplikasi mereka untuk menjangkau audiens yang lebih luas, sementara pengguna dapat mengakses jutaan aplikasi yang dikategorikan berdasarkan fungsi dan kegunaan, baik yang gratis maupun berbayar. Google Play Store memastikan bahwa setiap aplikasi yang tersedia telah melalui proses verifikasi untuk menjaga keamanan dan kenyamanan pengguna.

Google Play Store adalah *platform* yang tidak hanya menyediakan layanan unduhan aplikasi, tetapi juga menjadi ekosistem interaktif melalui sistem penilaian dan ulasan pengguna. Ulasan dan rating aplikasi di Play store berfungsi sebagai *feedback* bagi pengembang sekaligus referensi bagi pengguna lain. Selain itu, Play store menawarkan fitur-fitur seperti pembaruan otomatis, penghapusan aplikasi, dan perlindungan perangkat

melalui *Play Protect*, yang membantu menjaga stabilitas dan performa perangkat. Dalam konteks penelitian, data ulasan dan rating dari Play store menjadi sumber informasi berharga untuk analisis sentimen, karena mencerminkan persepsi dan pengalaman pengguna, yang dapat dimanfaatkan untuk meningkatkan kualitas layanan aplikasi, termasuk dalam kasus analisis ulasan aplikasi TransJatim.

2.4 TransJatim

TransJatim adalah layanan transportasi umum berbasis bus raya terpadu di Jawa Timur yang menghubungkan antar kota/kabupaten, seperti Sidoarjo, Surabaya, dan Gresik. Diluncurkan pada 19 Agustus 2022, layanan ini menggunakan 22 unit bus *highdeck* pada rute sepanjang 75 km, menggantikan Trans Sidoarjo yang berhenti beroperasi sejak April 2020. TransJatim terintegrasi dengan berbagai moda transportasi lain, seperti Suroboyo Bus dan Trans Semanggi Suroboyo, sehingga memudahkan perpindahan antar moda. Aplikasi mobile TransJatim mendukung kemudahan akses informasi layanan bus, seperti rute, lokasi halte, estimasi waktu kedatangan, serta memungkinkan pengguna memberikan ulasan untuk evaluasi dan peningkatan kualitas layanan (Ajeng et al., 2023). Berikut logo dari aplikasi TransJatim-Ajaib



Gambar 2. 1 Logo Aplikasi TransJatim-Ajaib

2.5 Text Preprocessing

Text preprocessing merupakan langkah awal yang sangat penting dalam analisis teks, termasuk analisis sentimen. Tahapan ini bertujuan untuk membersihkan dan mempersiapkan teks agar dapat diproses menggunakan algoritma *machine learning*. Tahapan utama dalam *text preprocessing* meliputi *casefolding*, *cleansing*, *tokenizing*, *stopword removal*, dan *stemming* (Rahman Isnain et al., 2021).

Tahapan-tahapan *preprocessing* dijelaskan sebagai berikut:

1. *Casefolding*

Proses ini mengubah semua huruf dalam teks menjadi huruf kecil. Tujuannya adalah untuk memastikan bahwa kata-kata yang seharusnya dianggap sama tidak dibedakan hanya karena perbedaan huruf kapital. Sebagai contoh, kata “Ganda” dan “ganda” akan dianggap identik setelah melalui tahap ini.

2. *Cleansing*

Pada tahap ini, teks dibersihkan dari karakter atau elemen yang tidak relevan seperti angka, tanda baca, simbol, dan karakter khusus lainnya yang tidak dibutuhkan untuk proses analisis.

3. *Tokenizing*

Tokenizing adalah proses memecah teks menjadi unit-unit kecil seperti kata atau token. Setiap kata dalam kalimat akan dipisahkan menjadi elemen individual yang dapat dianalisis secara terpisah.

4. *Stopword Removal*

Proses ini menghapus kata-kata umum yang sering muncul dalam teks, tetapi tidak memiliki makna penting dalam analisis. Contohnya adalah kata-kata seperti “yang,” “dan,” “di,” atau “ke.” Penghapusan *stopword* bertujuan untuk fokus pada kata-kata yang lebih signifikan.

5. *Stemming*

Stemming adalah proses mengubah kata-kata ke bentuk dasarnya atau akar katanya. Proses ini memastikan kata-kata dengan variasi bentuk

yang sama dapat dikenali sebagai entitas yang serupa, misalnya “mendukung” menjadi “dukung.”

2.6 Pembobotan Kata *TF-IDF*

TF-IDF (Term Frequency-Inverse Document Frequency) adalah sebuah ukuran statistik yang digunakan untuk menentukan seberapa relevan sebuah kata dalam sebuah dokumen di dalam suatu kumpulan dokumen. Ini dihitung dengan mengalikan dua komponen utama: *Term Frequency (TF)*, yang mengukur seberapa sering kata tersebut muncul dalam dokumen, dan *Inverse Document Frequency (IDF)*, yang mengukur seberapa jarang kata tersebut muncul di seluruh dokumen dalam korpus. Kombinasi kedua nilai ini membantu untuk menilai kepentingan kata tersebut dalam konteks dokumen tertentu, sehingga kata yang sering muncul di dokumen namun jarang ditemukan di seluruh korpus akan mendapatkan nilai *TF-IDF* yang tinggi.

1. *Term Frequency (TF)*

Mengukur seberapa sering sebuah kata muncul dalam sebuah dokumen. Terdapat beberapa cara untuk menghitung *TF*, namun cara yang paling umum digunakan adalah:

$$TF(t, d) = \frac{f_{t,d}}{N_d} \quad (2.1)$$

Dimana:

- a. $TF(t, d)$ adalah *Term Frequency* untuk kata t dalam dokumen d .
- b. $f_{t,d}$ adalah Frekuensi kata t dalam dokumen d , yaitu jumlah kemunculan kata t dalam dokumen tersebut.
- c. N_d adalah Jumlah total kata dalam dokumen d .

2. *Inverse Document Frequency (IDF)*

Inverse Document Frequency (IDF) mengukur sejauh mana sebuah kata penting dalam sebuah korpus dengan mempertimbangkan seberapa jarang kata tersebut muncul dalam dokumen-dokumen lain. Semakin sedikit sebuah kata muncul di banyak dokumen, semakin tinggi nilai

IDF-nya, yang menandakan kata tersebut lebih unik atau penting untuk dokumen tertentu. Rumus umum yang digunakan untuk menghitung IDF adalah:

$$IDF(t) = \log\left(\frac{N}{df(t)}\right) \quad (2.2)$$

Dimana:

- a. N adalah total jumlah dokumen dalam korpus.
- b. $df(t)$ adalah jumlah dokumen yang mengandung kata t .

IDF berfungsi untuk menyeimbangkan TF dalam model TF-IDF, di mana kata yang sering muncul di satu dokumen tapi jarang muncul di seluruh korpus akan mendapatkan nilai IDF tinggi, menandakan kata tersebut lebih penting dalam konteks dokumen tersebut.

2.7 Naïve Bayes

Naïve Bayes adalah algoritma klasifikasi berbasis teorema Bayes dengan asumsi bahwa setiap fitur dalam data bersifat independen terhadap fitur lainnya, yang dikenal sebagai asumsi *naïve*. Metode ini sering digunakan dalam analisis teks, termasuk analisis sentimen, karena sederhana namun efektif dalam menangani data dengan dimensi tinggi seperti teks. Metode ini berdasarkan Teorema Bayes, yang dirumuskan sebagai:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (2.3)$$

Di mana X adalah data tuple hasil pengujian dari sebuah set data yang telah ditentukan untuk masuk ke dalam kelas tertentu. C adalah kelas yang menjadi target klasifikasi, sementara $P(X|C)$ adalah probabilitas posterior, yaitu peluang bahwa X termasuk ke dalam kelas C , berdasarkan bukti X yang diperoleh dari observasi. Dalam rumus diatas, $P(X|C)$ adalah probabilitas *likelihood*, yaitu peluang munculnya data X jika diketahui bahwa X berasal dari kelas C . Selanjutnya, $P(C)$ adalah probabilitas prior, yaitu probabilitas awal suatu kelas sebelum mempertimbangkan data X . Sementara itu, $P(X|C)$ adalah probabilitas *evidence*, yaitu peluang

keseluruhan data X tanpa memperhatikan kelas tertentu. Dengan pendekatan ini, *Naïve Bayes* memungkinkan analisis probabilistik untuk menentukan seberapa besar kemungkinan X termasuk ke dalam kelas C berdasarkan data yang ada.

Dalam penelitian ini, aturan Bayes akan diterapkan dengan studi kasus tertentu. Oleh karena itu, aturan Bayes dapat dinyatakan sebagai berikut:

$$P(C_j|X) = \frac{P(X|C_j).P(C_j)}{P(X)} \quad (2.4)$$

Di mana c_j adalah kategori teks yang akan diklasifikasikan, dan $P(c_j)$ adalah probabilitas prior dari kategori teks c_j . Sementara itu, d adalah dokumen teks yang direpresentasikan sebagai himpunan kata (W_1, W_2, \dots, W_n), di mana W_1 adalah kata pertama, W_2 adalah kata kedua, dan seterusnya. Dalam proses pengklasifikasian dokumen teks, pendekatan Bayes akan memilih kategori dengan probabilitas tertinggi (CMAP), yaitu:

$$C_{MAP} = \operatorname{argmax} \frac{P(C_j).P(X|C_j)}{P(X)} \quad (2.5)$$

Nilai $P(X)$ dapat diabaikan karena nilainya adalah konstan untuk semua c_j , sehingga persamaan (2.6) dapat dituliskan:

$$C_{MAP} = \operatorname{argmax} p(c_j) p(X|c_j) \quad (2.6)$$

Probabilitas $p(c_j)$ dapat diperkirakan dengan menghitung jumlah dokumen pelatihan dalam setiap kategori (c_j). Namun, menghitung distribusi $p(X|c_j)$ sulit karena jumlah term yang sangat besar, yaitu jumlah kombinasi posisi kata dikalikan dengan jumlah kategori yang akan diklasifikasikan. Dengan pendekatan *Naïve Bayes* yang mengasumsikan bahwa setiap kata dalam setiap kategori independen satu sama lain, perhitungan dapat disederhanakan sebagai berikut:

$$C_{MAX} = \operatorname{argmax} p(c_j) p(X|c_j) \quad (2.7)$$

Dengan menggunakan persamaan (2.4), maka persamaan (2.8) dapat dituliskan menjadi:

$$C_{MAX} = \operatorname{argmax} p(C_j) p(X|C_j) \quad (2.8)$$

Nilai $p(c_j)$ dan $P(W_i|c_j)$ dihitung selama proses pelatihan dengan menggunakan persamaan berikut:

$$p(c_j) = \frac{|docs\ j|}{|contoh|} \quad (2.9)$$

$$p(W_i|c_j) = \frac{1 + n_i}{|c| + n(kosakata)} \quad (2.10)$$

Dimana:

- $p(W_i|c_j)$ = Probabilitas kata W_i pada kategori c_j
- $|docs\ j|$ = Jumlah dokumen pada kategori j
- $|contoh|$ = Jumlah seluruh dokumen sampel yang digunakan dalam proses training
- n_i = Frekuensi kemunculan kata W_i pada kategori c_j
- $|c|$ = Jumlah semua kata pada kategori c_j

2.8 Evaluasi

Evaluasi model merupakan langkah penting dalam menilai kinerja suatu model klasifikasi. Salah satu metode evaluasi yang paling umum digunakan adalah *confusion matrix*. *Confusion matrix* adalah tabel yang membandingkan hasil prediksi model dengan label sebenarnya. Tabel ini terdiri atas empat komponen utama, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)* (Nilwanda et al., 2024).

Tabel 2.1 *Confusion Matrix*

True Class	Predicted Class		
	Negatif	Netral	Positif
Negatif	TN	FN	FN
Netral	FN	TN	FN
Positif	FP	FP	TP

Penjelasan singkat tentang elemen – elemen yang terdapat dalam *confusion matrix*:

- a. *TP (True Positive)*: Jumlah data yang benar-benar termasuk dalam kelas positif dan berhasil diprediksi sebagai positif.
- b. *TN (True Negative)*: Jumlah data yang benar-benar termasuk dalam kelas negatif dan berhasil diprediksi sebagai negatif.
- c. *TN (True Neutral)*: Jumlah data yang benar-benar termasuk dalam kelas netral dan berhasil diprediksi sebagai netral.
- d. *FP (False Positive)*: Jumlah data yang sebenarnya tidak termasuk dalam kelas tertentu, tetapi diprediksi sebagai kelas tersebut. Contohnya, data negatif diprediksi sebagai positif.
- e. *FN (False Negative)*: Jumlah data yang sebenarnya termasuk dalam kelas tertentu, tetapi diprediksi sebagai kelas yang berbeda. Contohnya, data positif diprediksi sebagai negatif.
- f. *FN (False Neutral)*: Jumlah data yang sebenarnya termasuk dalam kelas tertentu, tetapi diprediksi sebagai kelas lain, misalnya data positif diprediksi sebagai netral.

Akurasi merupakan metrik yang digunakan untuk menghitung persentase prediksi yang tepat dibandingkan dengan total keseluruhan prediksi, yang dirumuskan sebagai:

$$Akurasi = \frac{TP + TN + TN}{TP + TN + TN + FP + FN + FN} \quad (2.11)$$

Precision adalah metrik yang menunjukkan persentase prediksi positif yang benar-benar sesuai dengan kondisi positif sebenarnya, dirumuskan sebagai:

$$Precision = \frac{T(Class)}{T(Class) + F(Class)} \quad (2.12)$$

Recall adalah metrik yang menggambarkan persentase data positif yang berhasil terdeteksi dengan benar oleh model, dirumuskan sebagai:

$$Recall = \frac{T(Class)}{T(Class) + F(Class)} \quad (2.13)$$

F1 – score adalah metrik yang menggabungkan nilai *precision* dan *recall* menjadi satu angka untuk memberikan penilaian kinerja model secara seimbang, dirumuskan sebagai:

$$F1 - Score = 2 \cdot \frac{Precision(Class) \times Recall(Class)}{Precision(Class) + Recall(Class)} \quad (2.14)$$

Metrik-metrik ini memberikan pemahaman yang lebih menyeluruh mengenai kinerja model, terutama dalam situasi klasifikasi yang memiliki distribusi data yang tidak seimbang.

2.9 Penelitian Sejenis

Terdapat sejumlah penelitian sebelumnya yang relevan dan telah dilakukan oleh berbagai pihak. Penelitian-penelitian ini menjadi rujukan penting dan memberikan kontribusi sebagai dasar untuk penelitian saat ini. Berikut beberapa penelitian yang dapat dijadikan referensi dalam penelitian ini:

Tabel 2. 2 Penelitian Sejenis

No	Literatur	Metode	Permasalahan	Hasil Penelitian
1	Judul: Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes Penulis: Artanti Inez Tanggraeni Sitokdana,	Naïve Bayes	Menganalisis sentimen ulasan aplikasi e-Government untuk mengevaluasi layanan pemerintah berbasis digital.	Metode Naïve Bayes menunjukkan kinerja yang baik dalam pengelompokan sentimen dengan akurasi mencapai 87%. Hasil analisis membantu pemerintah dalam memahami

No	Literatur	Metode	Permasalahan	Hasil Penelitian
	Melkior N. N. Sitokdana Tahun: 2022			pengalaman dan kebutuhan pengguna aplikasi.
2	Judul: Analisis Sentimen Ulasan Aplikasi Capcut Menggunakan Metode Naïve Bayes Penulis: VF. Anindya, Andri Wijaya Tahun: 2023	Naïve Bayes	Menganalisis sentimen ulasan pengguna aplikasi Capcut di Google Play Store.	Analisis sentimen ulasan aplikasi CapCut dengan metode <i>Naïve Bayes</i> menunjukkan 49,7% ulasan positif dan 50,3% negatif. Hasil ini memberikan wawasan bagi pengembang untuk meningkatkan pengalaman pengguna, meskipun tahap <i>Preprocessing</i> data masih dapat diperbaiki.
3	Judul: Analisis Sentimen Ulasan Aplikasi Buzzbreak	Naïve Bayes	Menganalisis sentimen ulasan aplikasi Buzzbreak di	Analisis sentimen aplikasi Buzzbreak dengan Naïve Bayes

No	Literatur	Metode	Permasalahan	Hasil Penelitian
	<p>Menggunakan Metode Naïve Bayes Classifier pada Situs Google Play Store</p> <p>Penulis: Dinda Putri Santoso, Wahyu Wibowo</p> <p>Tahun: 2022</p>		<p>Google Play Store.</p>	<p>menunjukkan lebih banyak ulasan positif dibandingkan negatif, dengan akurasi klasifikasi sedang (AUC 76,52% hingga 79,34%).</p>
4	<p>Judul: Analisis Sentimen Data Ulasan Aplikasi Ruangguru Pada Situs Google Play Menggunakan Algoritma Naïve Bayes Classifier Dengan Normalisasi Kata Levenshtein Distance</p> <p>Penulis: Hindun</p>	<p>Naïve Bayes Classifier dengan Normalisasi Kata Levenshtein Distance</p>	<p>Menganalisis sentimen data ulasan aplikasi Ruangguru di Google Play dengan normalisasi kata menggunakan Levenshtein Distance.</p>	<p>Naïve Bayes Classifier dengan Levenshtein Distance meningkatkan akurasi sentimen ulasan Ruangguru, mencapai rata-rata 88,20% dan tertinggi 94%.</p>

No	Literatur	Metode	Permasalahan	Hasil Penelitian
	Habibatul Mubaroroh, Hasbi Yasin, Agus Rusgiyono Tahun: 2022			
5	Judul: Analisis Sentimen Ulasan Aplikasi Bank Digital Menggunakan Algoritma Naïve Bayes Penulis: Febby Adelia Irawan, Aldy Rialdy Atmadja, Agung Wahana Tahun: 2024	Naïve Bayes	Menganalisis sentimen ulasan pengguna pada aplikasi bank digital di Google Play Store.	Penerapan model klasifikasi dengan algoritma Naïve Bayes menghasilkan 46% ulasan positif dan 54% ulasan negatif. Nilai akurasi tertinggi mencapai 89% dengan pembagian data training dan testing 70:30.
6	Judul: Analisis Sentimen Ulasan Aplikasi Identitas Kependudukan Digital pada Play Store	Naïve Bayes	Menganalisis sentimen ulasan aplikasi Identitas Kependudukan Digital di Play Store.	Hasil analisis sentimen aplikasi Identitas Kependudukan Digital menunjukkan lebih banyak sentimen negatif.

No	Literatur	Metode	Permasalahan	Hasil Penelitian
	Menggunakan Metode Naïve Bayes Penulis: Ahmad Komarudin, Atiqah Meutia Hilda Tahun: 2024			Dengan SMOTE Upsampling, Naïve Bayes mencapai akurasi 85.06%, presisi 80.31%, dan recall 92.89%.
7	Judul: Analisis Sentimen pada Ulasan Aplikasi Tokopedia Menggunakan Algoritma Naïve Bayes Classifier Penulis: Nanda Aurelia Salsabila, Umi Sa'adah, Fatkhurohman Fauzi Tahun: 2024	Naïve Bayes	Menganalisis sentimen pada ulasan aplikasi Tokopedia.	Analisis sentimen ulasan aplikasi Tokopedia menggunakan Naïve Bayes menghasilkan akurasi 82,97%, dengan 338 ulasan positif dan 1481 ulasan negatif. Pembagian data training 80% dan testing 20% memberikan hasil akurasi terbaik,
8	Judul: Analisis Sentimen Pengguna Aplikasi	Support Vector Machine	Menganalisis sentimen pengguna aplikasi	Model SVM tanpa seleksi fitur menghasilkan akurasi 87,58%,

No	Literatur	Metode	Permasalahan	Hasil Penelitian
	<p>DANA</p> <p>Berdasarkan Ulasan pada Google Play</p> <p>Menggunakan Metode Support Vector Machine</p> <p>Penulis: Abitdavy Athallah Muhammad, Ermatita, Desta Sandya Prasvita</p> <p>Tahun: 2022</p>		<p>DANA</p> <p>berdasarkan ulasan di Google Play.</p>	<p>presisi 90,21%, dan recall 91,20%. Setelah menggunakan seleksi fitur Chi-Square, akurasi meningkat menjadi 89,41%, presisi 93,29%, dan recall sedikit menurun menjadi 90,76%.</p>
9	<p>Judul: Analisis Sentimen pada Ulasan Aplikasi Sirekap di Google Play Store</p> <p>Menggunakan Metode Naïve Bayes</p> <p>Penulis: Muhammad Rizky Hanafi,</p>	<p>Naïve Bayes</p>	<p>Menganalisis sentimen pada ulasan aplikasi Sirekap di Google Play Store.</p>	<p>Analisis <i>Naïve Bayes</i> menunjukkan 1397 ulasan negatif untuk SIREKAP, dengan akurasi model 90%. Ulasan negatif menyoroti kesulitan penggunaan aplikasi, terutama</p>

No	Literatur	Metode	Permasalahan	Hasil Penelitian
	Rakhmat Kurniawan R 2 Tahun: 2024			saat pemilu. Model ini menunjukkan kinerja baik dengan presisi 98% dan F1-score 89%, memberikan wawasan penting tentang persepsi masyarakat terhadap aplikasi.
10	Judul: Analisis Sentimen terhadap Kualitas Pelayanan Aplikasi Go-Jek Menggunakan Metode Naïve Bayes Penulis: Khofifah Diah Indarwati, Herny Februariyanti Tahun: 2023	Naïve Bayes	Menganalisis sentimen terhadap kualitas pelayanan aplikasi Go-Jek.	Analisis sentimen ulasan pengguna Gojek menunjukkan aplikasi ini berdampak besar pada pencarian transportasi umum. Kata teratas dari wordcloud adalah "mantap" (131.085) dan "bantu" (106.727). Pengguna lebih memilih aplikasi

No	Literatur	Metode	Permasalahan	Hasil Penelitian
				karena praktis, terpercaya, menyediakan data driver, dan informasi rute.

