Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 3, Maret 2026
e-ISSN 2685-998X
DOI 10.30865/json.v7i3.9486

Analisis Sentimen Ulasan Pengguna LinkedIn Menggunakan Naïve Bayes dan K-Nearest Neighbor

Ricko Muhammad Firdaus*, Eko Darmanto, Rhoedy Setiawan
Sistem Informasi, Fakultas Teknik, Universitas Muria Kudus, Kudus, Indonesia
Email: [1,*]202153091@std.umk.ac.id, [2]eko.darmanto@umk.ac.id, [3,]rhoedy.setiawan@umk.ac.id
Email Penulis Korespondensi: 202153091@std.umk.ac.id *
Submitted: **27/01/2026**; Accepted: **19/02/2026**; Published: **31/03/2026**

Abstrak– Perkembangan aplikasi profesional seperti LinkedIn mendorong meningkatnya jumlah ulasan pengguna, namun volume data yang besar menyulitkan proses analisis sentimen secara manual. Permasalahan utama dalam penelitian ini adalah bagaimana mengklasifikasikan sentimen ulasan pengguna secara akurat untuk mengetahui persepsi pengguna terhadap aplikasi LinkedIn. Penelitian ini bertujuan untuk membandingkan kinerja algoritma Naïve Bayes dan K-Nearest Neighbor (KNN) dalam analisis sentimen ulasan pengguna LinkedIn. Metode penelitian yang digunakan adalah CRISP-DM, dengan tahapan *preprocessing* teks .Ekstraksi fitur dilakukan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Dataset yang digunakan terdiri dari 3.500 ulasan pengguna dengan pembagian data latih dan data uji sebesar 80% dan 20%. Hasil evaluasi menunjukkan bahwa algoritma Naïve Bayes memperoleh performa yang lebih baik dengan *accuracy* sebesar 88%, nilai AUC sebesar 0,9424, sedangkan KNN menghasilkan *accuracy* sebesar 79% dan nilai *AUC* sebesar 0,8557. Hasil penelitian ini menunjukkan bahwa Naïve Bayes lebih efektif dalam mengklasifikasikan sentimen ulasan aplikasi LinkedIn dan dapat diimplementasikan ke dalam aplikasi berbasis web sebagai sistem pendukung analisis sentimen.

Kata Kunci: Sentimen; LinkedIn; Naïve Bayes; CRISP-DM; KNN

Abstract– The rapid growth of professional mobile applications such as LinkedIn has led to a large volume of user reviews, making manual sentiment analysis increasingly difficult. The main problem addressed in this study is how to accurately classify user sentiment to understand public perception of the LinkedIn application. This study aims to compare the performance of Naïve Bayes and K-Nearest Neighbor (KNN) algorithms in sentiment analysis of LinkedIn user reviews. The research methodology follows the CRISP-DM framework, including text preprocessing stages. Feature extraction is performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The dataset consists of 3,500 user reviews, divided into 80% training data and 20% testing data. Evaluation results show that the Naïve Bayes algorithm outperforms KNN, achieving an accuracy of 88% and an Area Under the Curve (AUC) value of 0.9424, while KNN achieves an accuracy of 79% and an AUC value of 0.8557. These findings indicate that Naïve Bayes is more effective for classifying sentiment in LinkedIn application reviews and can be implemented in a web-based system to support sentiment analysis.

Keywords: Sentiment; LinkedIn; Naïve Bayes; CRISP-DM; KNN

1. PENDAHULUAN

Pesatnya perkembangan aplikasi mobile profesional seperti LinkedIn telah menciptakan volume ulasan pengguna yang masif di platform Google Play Store, namun tantangan utama terletak pada ketidakmampuan mengklasifikasikan sentimen secara akurat untuk mengidentifikasi pain points pengguna terkait fitur *networking*, keamanan data, dan rekomendasi pekerjaan [1]. Ulasan ini sering kali mengandung bahasa campuran teknis-formal yang sulit dianalisis secara manual, menyebabkan keterlambatan dalam perbaikan aplikasi dan hilangnya retensi pengguna di tengah persaingan ketat dengan platform seperti Indeed atau Glassdoor. Solusi yang diharapkan dari penelitian ini adalah pengembangan model komparatif antara algoritma Naive Bayes dan K-Nearest Neighbor (KNN) yang dioptimalkan dengan *preprocessing* teks dan TF-IDF.

Penelitian terkait pertama yang sebanding dilakukan pada analisis sentimen ulasan aplikasi Threads di Google Play Store, di mana Naive Bayes dibandingkan dengan KNN menggunakan rasio data latih-uji 80:20, menghasilkan keunggulan Naive Bayes dengan *accuracy* 11% lebih tinggi berkat penanganan yang lebih baik terhadap distribusi kelas tidak seimbang [2]. Studi ini menyoroti efektivitas Naive Bayes pada data ulasan aplikasi sosial, meskipun terbatas pada satu platform dan tidak mengeksplorasi variasi parameter KNN seperti metrik jarak cosine.

Penelitian kedua fokus pada sentimen publik terkait PHK di Twitter menggunakan Naive Bayes dan SVM, dengan pengujian pada tiga skenario pembagian data (80:20, 70:30, 90:10) yang menunjukkan SVM sedikit unggul, namun Naive Bayes tetap kompetitif pada dataset besar 3.458 tweet berbahasa Indonesia [3]. Pendekatan ini berguna untuk data media sosial real-time, tetapi tidak membahas KNN dan kurang menangani ulasan aplikasi mobile yang lebih struktural dibandingkan *tweet* pendek.

Penelitian ketiga menganalisis sentimen terhadap kenaikan PPN 12% di media sosial X, Instagram, dan TikTok menggunakan IndoBERT fine-tuned, mencapai *accuracy* 84,94% pada 2.581 sampel setelah pra-pemrosesan dan tokenisasi [4]. Model transformer ini unggul dalam menangkap konteks bahasa Indonesia, namun memerlukan sumber daya komputasi tinggi dan tidak komparatif dengan metode probabilistik sederhana seperti Naive Bayes atau KNN, sehingga kurang praktis untuk penelitian skala kecil.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 3, Maret 2026
e-ISSN 2685-998X
DOI 10.30865/json.v7i3.9486

Penelitian keempat melakukan bibliometrik tren analisis sentimen dari Scopus (2020-2025), mengidentifikasi dominasi BERT, LSTM, dan lexicon-based di domain teknologi serta bisnis, dengan peningkatan signifikan pada *aspect-based sentiment analysis*[5]. Temuan ini mengonfirmasi relevansi analisis sentimen di era digital, tetapi bersifat *review* literatur tanpa implementasi empiris pada ulasan aplikasi spesifik seperti LinkedIn.

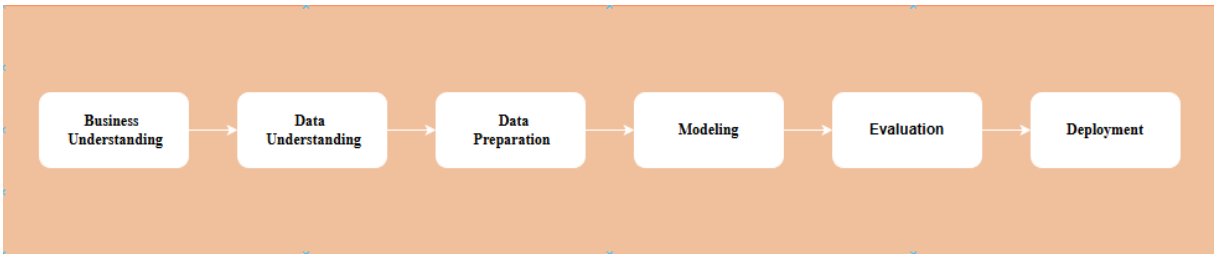
Penelitian kelima mengeksplorasi sentimen Twitter tentang Ibu Kota Nusantara menggunakan LSTM dan lexicon-based pada 5.112 *tweet*, menghasilkan distribusi 44,8% positif, 36,2% negatif, dan 19% netral, dengan performa model dipengaruhi ukuran dataset[6] . Pendekatan *deep learning* ini efektif untuk teks panjang, namun tidak membandingkan dengan algoritma klasik dan mengabaikan konteks ulasan aplikasi profesional.

Berdasarkan penelitian-penelitian sebelumnya, sebagian besar studi hanya menerapkan satu algoritma klasifikasi atau berfokus pada pendekatan *deep learning* tanpa melakukan perbandingan komprehensif terhadap metode klasik pada domain aplikasi profesional. Selain itu, belum ditemukan penelitian yang secara spesifik mengintegrasikan perbandingan Naïve Bayes dan K-Nearest Neighbor dengan implementasi sistem berbasis web pada analisis sentimen ulasan LinkedIn. Oleh karena itu, penelitian ini memberikan kontribusi berupa evaluasi komparatif dua algoritma klasifikasi tradisional serta implementasinya ke dalam sistem nyata yang dapat digunakan untuk analisis sentimen secara praktis.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada penelitian ini metode yang digunakan yaitu CRISP-DM (Cross Industri Standard Process for Data Mining)[7]. Metode CRISP-DM secara khusus berfokus pada pemahaman konteks data serta tujuan dari proses pengolahan data. Dalam penerapannya, CRISP-DM memberikan perhatian yang jelas terhadap penentuan sumber data dan tahapan persiapan data. Karakteristik metode ini yang bersifat fleksibel dan terukur membantu analis data dalam memilih sumber data yang valid dan memiliki tingkat kredibilitas yang tinggi[8]. Tahapan-tahapan dalam metode CRISP-DM dapat dilihat pada gambar 1.



Gambar 1. Tahapan CRISP-DM

2.2 *Business Understanding*

Tahap awal dalam metodologi CRISP-DM berfokus pada pemahaman terhadap tujuan penelitian serta konteks permasalahan yang akan dianalisis. Dalam penelitian ini, tujuan utama yang ingin dicapai adalah melakukan analisis sentimen terhadap ulasan pengguna aplikasi LinkedIn. Ulasan pengguna tersebut merepresentasikan pengalaman, penilaian, serta kepuasan pengguna terhadap fitur dan layanan yang disediakan oleh aplikasi LinkedIn. Melalui analisis sentimen terhadap data ulasan tersebut, penelitian ini diharapkan mampu memberikan gambaran umum mengenai persepsi pengguna, apakah cenderung bersifat positif maupun negatif terhadap aplikasi LinkedIn[9].

2.3 *Data Understanding*

Tahapan ini diawali dengan proses pengumpulan data menggunakan teknik *scraping*[10]. Data yang diambil berasal dari ulasan pengguna aplikasi LinkedIn pada Google Play Store sebanyak 3.500 ulasan dari tanggal 28 Maret 2024 sampai 5 Mei 2025 . Seluruh data yang diperoleh kemudian disatukan ke dalam satu dataset terstruktur untuk memudahkan proses analisis selanjutnya. Dataset tersebut selanjutnya ditelaah secara menyeluruh guna mengevaluasi karakteristik data serta mengidentifikasi permasalahan yang muncul, seperti ketidakseimbangan kelas, data tidak lengkap, maupun pola sentimen yang terkandung di dalamnya. Tahapan ini bertujuan untuk memperoleh pemahaman awal terhadap kondisi data sebagai dasar dalam menentukan metode pengolahan dan analisis yang tepat[11].

2.4 *Data Preparation*

Pada tahap ini, proses pengolahan data teks melibatkan beberapa teknik *preprocessing*, antara lain *case folding, cleaning, normalization ,tokenizing, stopword removal* dan *stemming*[12]. Tahap *case folding* dilakukan untuk mengubah seluruh huruf pada teks menjadi huruf kecil agar format kata menjadi seragam. Selain itu, tahap *cleaning* diterapkan untuk menghilangkan elemen-elemen yang tidak relevan dalam teks, seperti tanda baca,

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 3, Maret 2026
e-ISSN 2685-998X
DOI 10.30865/json.v7i3.9486

angka, simbol, tautan URL, serta *username* Selanjutnya, proses *normalization* bertujuan untuk memperbaiki kata-kata tidak baku serta mengonversi singkatan menjadi bentuk kata yang sesuai dengan kaidah bahasa[13]. Pada tahap *tokenization*, teks diuraikan menjadi kumpulan token kata sehingga dapat diproses secara individual. Langkah ini mempermudah analisis lanjutan berbasis kata. Selanjutnya, proses *stopword removal* diterapkan untuk menghilangkan kata-kata umum yang tidak memiliki pengaruh besar terhadap makna teks[14]. Proses *stemming* dilakukan untuk mengembalikan kata ke bentuk dasarnya dengan menggunakan *library* yang dijadikan sebagai referensi. Tahapan ini bertujuan untuk menyeragamkan kata-kata yang memiliki makna sama sehingga dapat meningkatkan efektivitas analisis dan pengolahan data teks[15]. Setelah tahap *preprocessing* selesai, proses dilanjutkan dengan penerapan metode TF-IDF, yaitu teknik pembobotan kata yang mempertimbangkan frekuensi kemunculan kata dalam suatu dokumen serta tingkat kelangkaannya pada keseluruhan kumpulan dokumen[16].

2.5 *Modeling*

Tahap berikutnya adalah *Modeling*, yaitu proses pemilihan dan penerapan metode *modeling* yang sesuai terhadap data yang telah melalui tahap persiapan [17]. Pada tahap *modeling*, data yang telah melalui proses *preprocessing* dan ekstraksi fitur menggunakan TF-IDF dibagi menjadi data latih dan data uji. Penelitian ini menggunakan dua algoritma klasifikasi, yaitu *Multinomial Naïve Bayes dan K-Nearest Neighbor (KNN).*

2.5.1 Multinomial Naïve Bayes

Algoritma yang digunakan adalah Multinomial Naïve Bayes karena sesuai untuk data teks berbasis frekuensi kata hasil pembobotan TF-IDF. Model diimplementasikan menggunakan pustaka *scikit-learn* dengan parameter alpha (α) sebesar 1.0 sebagai *Laplace smoothing* untuk menghindari probabilitas nol pada fitur yang jarang muncul. Nilai tersebut merupakan nilai default pada implementasi scikit-learn.Model ini bekerja berdasarkan Teorema Bayes yang dirumuskan sebagai berikut:

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (1)$$

Keterangan:

$P(C_i|X)$ adalah probabilitas posterior bahwa data X termasuk dalam kelas C_i
$P(X|C_i)$ adalah *likelihood* atau probabilitas kemunculan data X pada kelas C_i
$P(C_i)$ adalah *prior probability* dari kelas C_i
$P(X)$ adalah *evidence* atau probabilitas total data

2.5.2 K-Nearest Neighbor (KNN)

Algoritma K-Nearest Neighbor digunakan sebagai metode pembanding. Model diimplementasikan menggunakan parameter n_neighbors = 5. Nilai k dipilih sebagai bilangan ganjil untuk menghindari kemungkinan hasil voting yang imbang.Metrik jarak yang digunakan adalah Minkowski distance dengan parameter p = 2, yang secara matematis ekuivalen dengan Euclidean distance. Perhitungan jarak antar dua vektor dinyatakan sebagai:

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

Keterangan:

$d\ (x, y)$ adalah jarak antara vektor x dan y
x_i dan y_i adalah nilai atribut ke-i
n adalah jumlah atribut atau dimensi data

Proses klasifikasi dilakukan dengan menghitung jarak antara data uji dan seluruh data latih, kemudian memilih 5 tetangga terdekat dan menentukan kelas berdasarkan *majority voting.*

2.6 *Evaluation*

Tahap *Evaluation* bertujuan untuk mengukur tingkat kinerja model yang telah dibangun dalam melakukan klasifikasi data secara tepat. Pada tahap ini, performa model dievaluasi menggunakan beberapa metrik pengujian, seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil evaluasi tersebut menjadi dasar untuk menilai apakah model yang dihasilkan telah memenuhi tujuan penelitian atau masih memerlukan perbaikan lebih lanjut[18].Adapun perhitungan *evalutaion matrix* yang digunakan dalam penelitian ini dapat dirumuskan sebagai berikut:

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 3, Maret 2026
e-ISSN 2685-998X
DOI 10.30865/json.v7i3.9486

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1\ Score = 2x \frac{recall\ x\ precision}{recall + precision} \quad (6)$$

2.7 *Deployment*

Tahap *deployment* merupakan tahapan penutup dalam metode CRISP-DM yang berfokus pada implementasi hasil analisis ke dalam sebuah aplikasi atau sistem. Pada tahap ini, model dan data yang telah melalui proses pengolahan pada tahapan sebelumnya dimanfaatkan sebagai dasar dalam pengembangan sistem[19]. Model analisis sentimen yang telah dilatih kemudian diintegrasikan ke dalam aplikasi berbasis web yang dirancang untuk menerima data ulasan pengguna, melakukan proses *preprocessing* secara otomatis, serta menampilkan hasil klasifikasi sentimen beserta evaluasi performanya.

3. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil pengujian dan analisis terhadap model analisis sentimen yang telah dibangun pada tahapan sebelumnya. Pembahasan difokuskan pada evaluasi performa model berdasarkan metrik pengujian yang digunakan, serta interpretasi hasil klasifikasi sentimen terhadap data ulasan pengguna.

3.1 Tahap *Preprocessing*

Proses *preprocessing* data dilakukan secara bertahap untuk menghasilkan data teks yang bersih dan siap digunakan pada proses analisis selanjutnya. Tahapan *preprocessing* yang diterapkan meliputi *case folding*, *cleaning*, normalisasi*(normalization)*, *tokenizing*, *stopword removal*, dan *stemming*. Hasil dari setiap tahapan *preprocessing* ditampilkan pada Gambar 2.

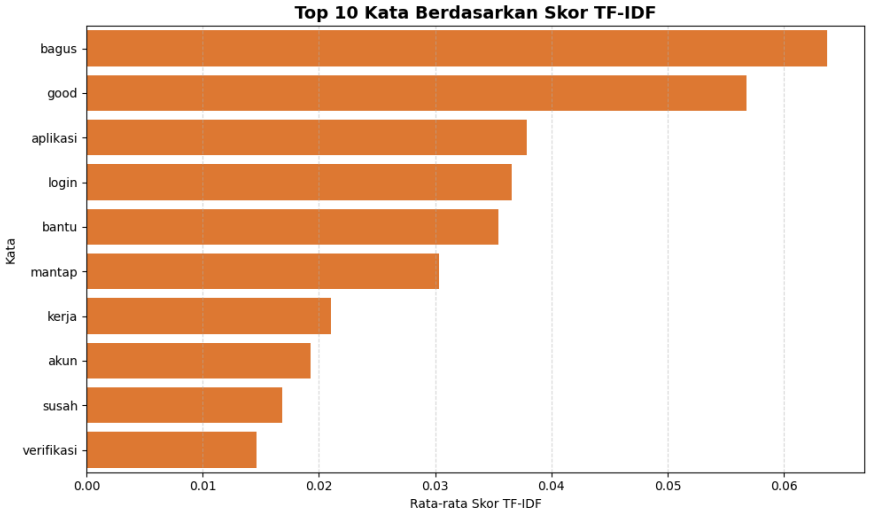
	userName	score	at	content	case_folding	cleaning	normalisasi	tokenizing	stopword_removal	stemming
0	Pengguna Google	1	2025-05-05 13:37:40	aplikasi ini bikin kita rugi, sudah di download mala nggak bisa di buka, yg benar lah kalau niat mau bantu jangan merugikan	aplikasi ini bikin kita rugi, sudah di download mala nggak bisa di buka, yg benar lah kalau niat mau bantu jangan merugikan	aplikasi ini bikin kita rugi sudah di download mala nggak bisa di buka yg benar lah kalau niat mau bantu jangan merugikan	aplikasi ini bikin kita rugi sudah di download malah tidak bisa di buka yang benar kalau niat mau bantu jangan merugikan	[aplikasi, ini, bikin, kita, rugi, sudah, di, download, malah, tidak, bisa, di, buka, yang, benar, kalau, niat, mau, bantu, jangan, merugikan]	[aplikasi, bikin, rugi, download, buka, benar, niat, bantu, jangan, merugikan]	[aplikasi, bikin, rugi, download, buka, benar, niat, bantu, jangan, rugi]
1	Pengguna Google	5	2025-04-30 07:57:11	tetap terhubung dengan grup family pribadi saya ini seluruh nya dan semakin lengkap dan sangat memuaskan.	tetap terhubung dengan grup family pribadi saya ini seluruh nya dan semakin lengkap dan sangat memuaskan.	tetap terhubung dengan grup family pribadi saya ini seluruh nya dan semakin lengkap dan sangat memuaskan	tetap terhubung dengan grup family pribadi saya ini seluruh nya dan semakin lengkap dan sangat memuaskan	[tetap, terhubung, dengan, grup, family, pribadi, saya, ini, seluruh, nya, dan, semakin, lengkap, dan, sangat, memuaskan]	[terhubung, grup, family, pribadi, semakin, lengkap, memuaskan]	[hubung, grup, family, pribadi, makin, lengkap, muas]
2	Pengguna Google	5	2025-04-30 03:04:59	sangat berguna kerana menambah insights	sangat berguna kerana menambah insights	sangat berguna kerana menambah insights	sangat berguna karena menambah insights	[sangat, berguna, karena, menambah, insights]	[berguna, menambah, insights]	[guna, tambah, insights]

Gambar 2. Hasil *Preprocessing*

Berdasarkan Gambar 2, dapat dilihat bahwa setiap tahapan *preprocessing* memberikan perubahan pada teks ulasan. Proses *case folding* mengubah seluruh huruf menjadi huruf kecil, kemudian tahap *cleaning* menghilangkan karakter yang tidak diperlukan. Selanjutnya, normalisasi*(normalization)* dilakukan untuk memperbaiki kata tidak baku. Tahap *tokenizing* memecah teks menjadi kumpulan kata, diikuti dengan *stopword removal* untuk menghilangkan kata-kata yang kurang informatif. Proses *stemming* kemudian mengubah kata menjadi bentuk dasarnya sehingga data menjadi lebih seragam dan siap digunakan pada tahap pembobotan fitur.

Setelah data teks melalui seluruh tahapan preprocessing, langkah selanjutnya adalah melakukan ekstraksi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Metode ini digunakan untuk mengubah data teks menjadi representasi numerik agar dapat diproses oleh algoritma *machine learning*. Hasil pembobotan TF-IDF ditampilkan pada Gambar 3.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 3, Maret 2026
e-ISSN 2685-998X
DOI 10.30865/json.v7i3.9486



Gambar 3. Representasi Data Menggunakan TF-IDF

Berdasarkan Gambar 3, dapat dilihat bahwa setiap kata direpresentasikan dalam bentuk bobot numerik berdasarkan tingkat kepentingannya dalam dokumen. Kata yang sering muncul pada satu dokumen namun jarang muncul pada dokumen lain akan memiliki bobot TF-IDF yang lebih tinggi. Representasi ini memungkinkan model untuk membedakan kata-kata yang bersifat informatif dan relevan dalam proses analisis sentimen.

3.2 *Modeling*

Setelah data ulasan melalui tahap *preprocessing* dan diekstraksi menggunakan metode TF-IDF, tahapan selanjutnya adalah modeling. Pada tahap ini, data yang telah direpresentasikan dalam bentuk numerik digunakan untuk membangun model klasifikasi sentimen. Proses *modeling* bertujuan untuk menghasilkan model yang mampu mengenali pola sentimen pada data ulasan pengguna dan melakukan klasifikasi sentimen secara akurat terhadap data baru. Tahapan ini merupakan bagian penting dalam proses analisis karena kualitas model yang dihasilkan sangat bergantung pada data yang telah dipersiapkan sebelumnya serta metode pemodelan yang digunakan.

3.2.1 *Data Splitting*

Sebelum proses pelatihan model dilakukan, dataset yang telah melalui tahap *preprocessing* dan ekstraksi fitur dibagi menjadi dua bagian, yaitu data latih *(training* data) dan data uji (*testing* data). Pembagian data ini bertujuan untuk mengukur kemampuan model dalam melakukan klasifikasi terhadap data yang belum pernah digunakan pada proses pelatihan. Dengan adanya pemisahan data, performa model dapat dievaluasi secara objektif dan tidak hanya berdasarkan kemampuan menghafal data latih.

Pada penelitian ini, pembagian data dilakukan dengan perbandingan 80% data latih dan 20% data uji. Proporsi ini dipilih karena dianggap mampu memberikan jumlah data yang cukup bagi model untuk mempelajari pola sentimen, sekaligus menyediakan data uji yang memadai untuk mengevaluasi performa model. Data latih digunakan dalam proses pembelajaran model, sedangkan data uji digunakan untuk mengukur kemampuan generalisasi model terhadap data baru.

3.2.2 *Data Balancing Using* SMOTE

Pada permasalahan klasifikasi sentimen, sering kali ditemukan kondisi ketidakseimbangan jumlah data antar kelas, di mana salah satu kelas memiliki jumlah data yang lebih dominan dibandingkan kelas lainnya. Kondisi ini dapat menyebabkan model cenderung lebih memprediksi kelas mayoritas dan mengabaikan kelas minoritas, sehingga hasil klasifikasi menjadi kurang optimal.

Untuk mengatasi permasalahan tersebut, pada penelitian ini diterapkan Teknik *Synthetic Minority Over-sampling Technique (SMOTE)*. *SMOTE* bekerja dengan cara menghasilkan data sintetis baru pada kelas minoritas berdasarkan karakteristik data yang sudah ada. Teknik ini diterapkan setelah proses pembagian data, khususnya pada data latih, dengan tujuan untuk menyeimbangkan distribusi kelas sebelum proses pelatihan model dilakukan. Dengan data latih yang lebih seimbang, model diharapkan dapat mempelajari pola sentimen dari setiap kelas secara lebih adil dan menghasilkan performa klasifikasi yang lebih baik.

3.2.3 *Model Deployment*

Setelah data latih berada dalam kondisi seimbang, tahap selanjutnya adalah model deployment. Pada tahap ini, algoritma machine learning diterapkan untuk mempelajari hubungan antara fitur-fitur hasil ekstraksi TF-

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 3, Maret 2026
e-ISSN 2685-998X
DOI 10.30865/json.v7i3.9486

IDF dan label sentimen yang telah ditentukan. Proses pelatihan model dilakukan menggunakan data latih, di mana model akan menyesuaikan parameter internalnya untuk meminimalkan kesalahan klasifikasi.

Model yang dibangun pada tahap ini bertujuan untuk mengenali pola-pola tertentu dalam data teks yang merepresentasikan sentimen positif dan negatif. Setelah proses pelatihan selesai, model yang dihasilkan siap digunakan untuk melakukan prediksi terhadap data uji maupun data baru. Tahapan modeling ini menjadi dasar bagi proses evaluasi model yang akan dibahas pada tahap selanjutnya.

Pada penelitian ini, algoritma Naïve Bayes yang digunakan adalah *Multinomial Naïve Bayes* karena sesuai untuk data teks berbasis frekuensi kata hasil pembobotan TF-IDF. Sementara itu, algoritma K-Nearest Neighbor menggunakan nilai k = 5 dengan metrik jarak cosine similarity dalam mengukur kedekatan antar dokumen. Pemilihan nilai k dilakukan melalui pengujian awal untuk memperoleh performa klasifikasi yang optimal terhadap data latih.

3.3 *Evaluation and Visualization*

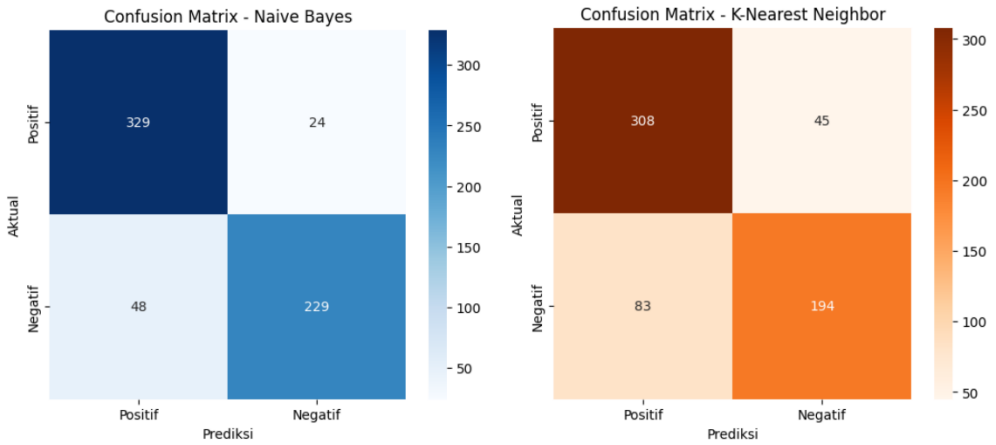
Tahap evaluation and visualization bertujuan untuk menilai performa model klasifikasi yang telah dibangun serta menyajikan hasil analisis dalam bentuk visual agar lebih mudah dipahami. Evaluasi dilakukan menggunakan berbagai metrik pengukuran untuk mengetahui tingkat keberhasilan model dalam mengklasifikasikan sentimen ulasan pengguna. Sementara itu, visualisasi digunakan untuk membantu interpretasi hasil klasifikasi dan memperlihatkan pola sentimen yang muncul pada data.

3.3.1 *Classification Model Evaluation*

Model evaluation dilakukan dengan menggunakan data uji yang sebelumnya telah dipisahkan dari data latih. Pada tahap ini, hasil prediksi model dibandingkan dengan label aktual untuk mengetahui tingkat ketepatan klasifikasi. Beberapa metrik evaluasi digunakan untuk memberikan gambaran performa model secara menyeluruh, tidak hanya dari sisi ketepatan prediksi, tetapi juga dari kemampuan model dalam mengenali setiap kelas sentimen.

a. *Confusion Matrix*

Untuk mengetahui performa model dalam mengklasifikasikan sentimen ulasan pengguna, dilakukan evaluasi menggunakan *confusion matrix*. *Confusion matrix* digunakan untuk menggambarkan perbandingan antara label aktual dan hasil prediksi yang dihasilkan oleh model, sehingga dapat diketahui jumlah data yang diklasifikasikan dengan benar maupun yang mengalami kesalahan klasifikasi pada masing-masing kelas sentimen. Pada penelitian ini, confusion matrix disajikan untuk dua model klasifikasi yang digunakan, yaitu Naive Bayes dan K-Nearest Neighbor, dengan dua kelas sentimen, yakni positif dan negatif.



Gambar 4. *Confusion Matrix* Naïve Bayes dan KNN

Berdasarkan hasil *confusion matrix* yang ditampilkan, model Naive Bayes menunjukkan performa yang lebih baik dibandingkan dengan K-Nearest Neighbor. Pada model Naive Bayes, sebagian besar data sentimen positif dan negatif berhasil diklasifikasikan dengan benar, dengan jumlah kesalahan prediksi yang relatif rendah. Sementara itu, model K-Nearest Neighbor masih mampu mengklasifikasikan sentimen dengan cukup baik, namun menghasilkan jumlah kesalahan klasifikasi yang lebih besar, khususnya pada kelas sentimen negatif. Hasil ini menunjukkan bahwa model Naive Bayes lebih konsisten dalam membedakan sentimen positif dan negatif pada dataset yang digunakan, sehingga memberikan performa klasifikasi yang lebih optimal.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 3, Maret 2026
e-ISSN 2685-998X
DOI 10.30865/json.v7i3.9486

b. *Classification Report*

Selain menggunakan *confusion matrix*, evaluasi performa model juga dilakukan dengan menggunakan *classification report*. *Classification report* memberikan informasi yang lebih rinci mengenai kinerja model dalam mengklasifikasikan data ke dalam masing-masing kelas sentimen. Metrik yang digunakan meliputi *accuracy*, *precision*, *recall*, *dan f1-score*, yang masing-masing berperan dalam menggambarkan tingkat ketepatan prediksi, kemampuan model dalam mengenali data aktual, serta keseimbangan antara *precision dan recall*. Gambar berikut menampilkan *classification report* untuk masing-masing model yang digunakan.

	precision	recall	f1-score	support
Negatif	0.9051	0.8267	0.8642	277.0000
Positif	0.8727	0.9320	0.9014	353.0000
accuracy	0.8857	0.8857	0.8857	0.8857
macro avg	0.8889	0.8794	0.8828	630.0000
weighted avg	0.8870	0.8857	0.8850	630.0000

Gambar 5. *Classification Report* Naïve Bayes

Berdasarkan hasil classification report pada model Naïve Bayes, terlihat bahwa model mampu memberikan performa yang cukup baik dalam mengklasifikasikan sentimen positif maupun negatif. Untuk kelas negatif, model memperoleh nilai *precision* sebesar 0,9051, *recall* 0,8267, dan *f1-score* 0,8642, yang menunjukkan bahwa sebagian besar data negatif dapat diklasifikasikan dengan tepat. Sementara itu, pada kelas positif, nilai *recall* yang mencapai 0,9320 menunjukkan bahwa model sangat baik dalam mengenali ulasan positif, meskipun nilai *precision* sedikit lebih rendah dibandingkan kelas negatif. Secara keseluruhan, model ini menghasilkan nilai *accuracy* sebesar 0,8857 atau 88%, yang mengindikasikan bahwa mayoritas data uji berhasil diklasifikasikan dengan benar.

	precision	recall	f1-score	support
Negatif	0.8117	0.7004	0.7519	277.0000
Positif	0.7877	0.8725	0.8280	353.0000
accuracy	0.7968	0.7968	0.7968	0.7968
macro avg	0.7997	0.7864	0.7899	630.0000
weighted avg	0.7983	0.7968	0.7945	630.0000

Gambar 6. *Classification Report* KNN

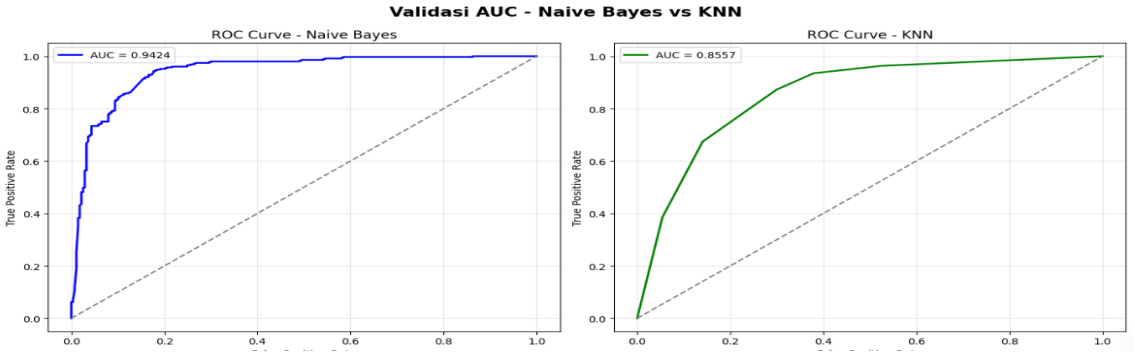
Pada model KNN, hasil *classification report* menunjukkan performa yang relatif lebih rendah dibandingkan model sebelumnya. Untuk kelas negatif, nilai *precision* sebesar 0,8117 dan *recall* 0,7004 mengindikasikan bahwa masih terdapat cukup banyak data negatif yang tidak terklasifikasi dengan tepat. Pada kelas positif, model memperoleh nilai *recall* 0,8725, yang menunjukkan kemampuan cukup baik dalam mengenali ulasan positif, meskipun nilai *precision* masih berada di bawah mode lNaive Bayes. Secara keseluruhan, model ini menghasilkan nilai *accuracy* sebesar 0,7968 atau 79% yang menandakan bahwa tingkat ketepatan prediksi masih lebih rendah jika dibandingkan dengan model sebelumnya.

Berdasarkan hasil evaluasi menggunakan *classification report*, dapat disimpulkan bahwa model Naïve Bayes menunjukkan performa yang lebih unggul dibandingkan KNN dalam mengklasifikasikan sentimen ulasan. Hal ini terlihat dari nilai *accuracy*, *precision*, *recall*, *dan f1-score* yang secara konsisten lebih tinggi. Dengan demikian, model pertama dinilai lebih stabil dan efektif dalam menangani data ulasan pengguna, khususnya dalam mengidentifikasi sentimen positif dan negatif secara seimbang.

c. *ROC Curve dan AUC*

Selain menggunakan *confusion matrix dan classification* report, evaluasi performa model juga dilakukan dengan menggunakan *Receiver Operating Characteristic (ROC) Curve* dan nilai *Area Under the Curve (AUC). ROC Curve* digunakan untuk menggambarkan hubungan antara *True Positive Rate (TPR)* dan *False Positive Rate (FPR)* pada berbagai ambang batas klasifikasi. Sementara itu, nilai *AUC* digunakan sebagai indikator kemampuan model dalam membedakan kelas positif dan negatif, di mana nilai *AUC* yang semakin mendekati 1 menunjukkan performa klasifikasi yang semakin baik. Gambar berikut menampilkan perbandingan *ROC Curve* dan nilai *AUC* dari masing-masing model yang digunakan.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 3, Maret 2026
e-ISSN 2685-998X
DOI 10.30865/json.v7i3.9486



Gambar 7. Perbandingan *ROC Curve* dan *AUC* Naive Bayes dan KNN

Berdasarkan hasil visualisasi *ROC Curve*, model Naïve Bayes menunjukkan kurva yang lebih mendekati sudut kiri atas grafik dibandingkan model KNN. Hal ini tercermin dari nilai AUC sebesar 0,9424, yang mengindikasikan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antara sentimen positif dan negatif. Nilai *AUC* yang tinggi menunjukkan bahwa model mampu mempertahankan tingkat *true positive* yang tinggi dengan tingkat *false positive* yang rendah pada berbagai ambang batas klasifikasi.

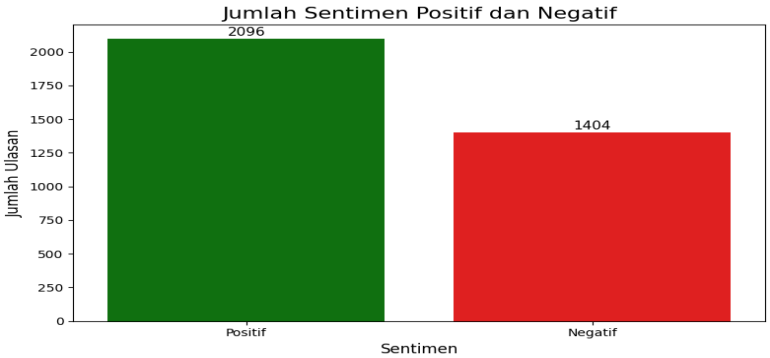
Sebaliknya, model KNN menghasilkan nilai *AUC* sebesar 0,8557, yang meskipun masih tergolong baik, namun berada di bawah performa model Naïve Bayes. *ROC Curve* pada model ini menunjukkan peningkatan *true positive rate* yang lebih lambat, sehingga kemampuan pemisahan antar kelas tidak seoptimal model Naïve Bayes. Perbedaan nilai *AUC* ini memperkuat hasil evaluasi sebelumnya, di mana model Naïve Bayes secara konsisten menunjukkan performa yang lebih unggul dalam proses klasifikasi sentimen.

3.3.2 Visualisasi Hasil Klasifikasi

Selain evaluasi numerik, hasil klasifikasi juga disajikan dalam bentuk visualisasi untuk memudahkan analisis dan interpretasi data.

a. Visualisasi Distribusi Sentimen

Pada tahap visualisasi hasil klasifikasi, dilakukan analisis terhadap distribusi sentimen ulasan pengguna untuk mengetahui kecenderungan persepsi pengguna terhadap aplikasi yang diteliti. Visualisasi distribusi sentimen bertujuan untuk memberikan gambaran proporsi antara sentimen positif dan sentimen negatif berdasarkan hasil pelabelan data. Dalam penelitian ini, proses pelabelan sentimen dilakukan berdasarkan nilai rating ulasan, di mana ulasan dengan rating 1 hingga 3 dikategorikan sebagai sentimen negatif, sedangkan ulasan dengan rating 4 dan 5 dikategorikan sebagai sentimen positif.



Gambar 8. Diagram Batang Distribusi Sentimen

Berdasarkan hasil visualisasi distribusi sentimen, diperoleh sebanyak 2.096 ulasan termasuk ke dalam kategori sentimen positif, sedangkan 1.404 ulasan tergolong sebagai sentimen negatif. Hasil ini menunjukkan bahwa mayoritas pengguna memberikan tanggapan positif terhadap aplikasi, yang mengindikasikan tingkat kepuasan pengguna yang relatif tinggi. Dominasi sentimen positif mencerminkan bahwa fitur, layanan, dan pengalaman penggunaan aplikasi dinilai baik oleh sebagian besar pengguna.

Di sisi lain, jumlah sentimen negatif yang masih cukup signifikan menunjukkan adanya aspek-aspek tertentu yang perlu diperhatikan dan dievaluasi lebih lanjut oleh pengembang aplikasi. Sentimen negatif tersebut dapat menjadi sumber informasi penting untuk mengidentifikasi permasalahan, kekurangan fitur, atau kendala teknis yang dirasakan oleh pengguna. Dengan demikian, visualisasi distribusi sentimen tidak hanya memberikan gambaran umum persepsi pengguna, tetapi juga dapat dimanfaatkan sebagai dasar pengambilan keputusan untuk peningkatan kualitas aplikasi di masa mendatang.

b. *Worldcloud*

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 3, Maret 2026
e-ISSN 2685-998X
DOI 10.30865/json.v7i3.9486

Visualisasi *WordCloud* digunakan untuk menampilkan kata-kata yang paling sering muncul pada ulasan pengguna berdasarkan hasil klasifikasi sentimen. Ukuran kata dalam *WordCloud* merepresentasikan tingkat frekuensi kemunculannya, di mana semakin sering suatu kata muncul dalam data ulasan, maka semakin besar ukuran kata tersebut pada visualisasi. Dengan demikian, *WordCloud* memberikan gambaran awal mengenai kata kunci yang dominan dalam setiap kategori sentimen secara intuitif dan mudah dipahami.

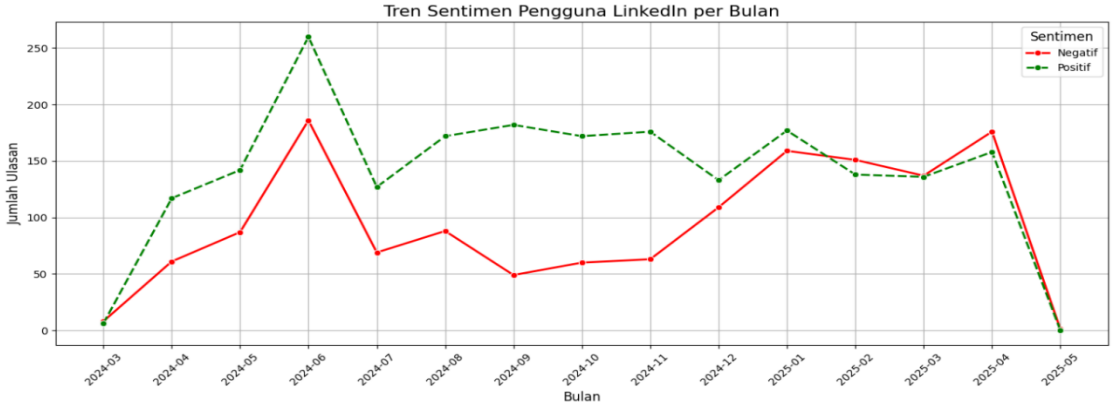
Pada *WordCloud* sentimen positif, kata-kata yang muncul umumnya berkaitan dengan pengalaman pengguna yang baik terhadap aplikasi, seperti kemudahan penggunaan, kelengkapan fitur, serta manfaat yang dirasakan oleh pengguna. Hal ini menunjukkan bahwa sebagian besar ulasan positif mencerminkan kepuasan pengguna terhadap fungsionalitas aplikasi dan nilai tambah yang diberikan dalam mendukung aktivitas profesional mereka.

Gambar 9. Visualisasi *WorldCloud* Ulasan Negatif dan Positif

Sementara itu, *WordCloud* pada sentimen negatif menampilkan kata-kata yang berkaitan dengan permasalahan yang dialami pengguna, seperti kendala teknis, kesulitan dalam mengakses aplikasi, atau fitur yang tidak berjalan sesuai harapan. Kata-kata dominan pada sentimen negatif ini dapat menjadi indikator adanya aspek yang perlu diperbaiki oleh pengembang aplikasi agar kualitas layanan dapat ditingkatkan. Secara keseluruhan, visualisasi *WordCloud* berperan sebagai pelengkap hasil evaluasi model klasifikasi sentimen dengan memberikan interpretasi kualitatif terhadap data ulasan. Informasi yang diperoleh dari *WordCloud* dapat dimanfaatkan sebagai bahan analisis tambahan untuk memahami persepsi pengguna secara lebih mendalam, baik dari sisi kepuasan maupun keluhan yang disampaikan.

c. Visualisasi Tren Sentimen Pengguna

Untuk memahami dinamika perubahan sentimen pengguna dari waktu ke waktu, dilakukan visualisasi distribusi sentimen positif dan negatif berdasarkan periode bulanan. Visualisasi ini bertujuan untuk menunjukkan pola tren ulasan pengguna terhadap aplikasi LinkedIn dalam rentang waktu tertentu, sehingga dapat diketahui periode peningkatan maupun penurunan sentimen pengguna.



Gambar 10. Grafik Sentimen Pengguna Per Bulan

Gambar tersebut menampilkan tren jumlah ulasan sentimen positif dan negatif pengguna aplikasi LinkedIn per bulan selama periode Maret 2024 hingga Mei 2025. Secara umum, jumlah ulasan dengan sentimen positif cenderung lebih tinggi dibandingkan ulasan negatif pada hampir seluruh periode pengamatan, yang menunjukkan bahwa persepsi pengguna terhadap aplikasi LinkedIn didominasi oleh sentimen positif. Meskipun demikian, kedua jenis sentimen memperlihatkan fluktuasi yang cukup signifikan dari bulan ke bulan, yang mencerminkan adanya dinamika tingkat kepuasan pengguna terhadap aplikasi.

Puncak jumlah ulasan positif terjadi pada Juni 2024 dengan jumlah ulasan yang melebihi 250 ulasan, bersamaan dengan peningkatan ulasan negatif yang mendekati 190 ulasan pada bulan yang sama. Kondisi ini menjadikan Juni 2024 sebagai periode dengan aktivitas ulasan tertinggi. Setelah periode tersebut, baik ulasan positif maupun negatif mengalami penurunan yang cukup tajam pada Juli 2024, sebelum kembali menunjukkan tren peningkatan secara bertahap pada bulan Agustus hingga November 2024.

Pada periode akhir pengamatan, khususnya antara Desember 2024 hingga April 2025, jumlah ulasan negatif menunjukkan kecenderungan meningkat. Pada beberapa bulan, seperti Februari hingga April 2025, jumlah

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 3, Maret 2026
e-ISSN 2685-998X
DOI 10.30865/json.v7i3.9486

ulasan negatif hampir menyamai bahkan sedikit melampaui jumlah ulasan positif. Hal ini mengindikasikan adanya potensi penurunan kepuasan pengguna.

3.4 *Deployment*

Tahap *deployment* merupakan tahapan akhir dalam metode CRISP-DM yang berfokus pada penerapan hasil analisis ke dalam bentuk sistem yang dapat digunakan secara langsung. Pada tahap ini, model analisis sentimen yang telah melalui proses *preprocessing*, *modeling*, dan *evaluation* diimplementasikan ke dalam sebuah aplikasi berbasis web.

Gambar 11. Tampilan Aplikasi Analisis Sentimen

Berdasarkan gambar tersebut, aplikasi analisis sentimen dikembangkan dengan antarmuka berbasis web yang memungkinkan pengguna untuk melakukan analisis sentimen secara interaktif. Aplikasi ini menyediakan fitur utama seperti pengambilan data ulasan, *preprocessing* otomatis, klasifikasi sentimen, evaluasi performa model, serta visualisasi hasil analisis. Pengguna dapat mengunggah atau mengambil data ulasan baru, kemudian sistem akan secara otomatis memproses data tersebut dan menampilkan hasil klasifikasi sentimen dalam bentuk teks maupun visualisasi grafik.

Selain itu, sistem dirancang agar model yang telah dibangun dapat digunakan kembali tanpa perlu melakukan proses pelatihan ulang. Dengan demikian, aplikasi mampu menganalisis data ulasan baru secara efisien dan konsisten sesuai dengan model yang telah dievaluasi sebelumnya. Implementasi tahap *deployment* ini menunjukkan bahwa model analisis sentimen yang dihasilkan dapat diintegrasikan ke dalam sistem nyata, sehingga mendukung pengambilan keputusan berbasis data dan meningkatkan nilai praktis dari hasil penelitian yang dilakukan.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa analisis sentimen terhadap ulasan pengguna aplikasi LinkedIn pada berhasil memberikan gambaran yang jelas mengenai persepsi pengguna terhadap aplikasi tersebut. Melalui penerapan metode CRISP-DM, seluruh tahapan penelitian mulai dari *business understanding* , *data understanding*, *data preparation*, *modeling*, *evaluation*, hingga *deployment* dapat dijalankan secara sistematis. Proses *preprocessing* teks terbukti mampu menghasilkan data teks yang lebih bersih dan representatif, sehingga meningkatkan kualitas fitur yang dihasilkan melalui metode TF-IDF. Hasil modeling menunjukkan bahwa algoritma Naïve Bayes memiliki performa yang lebih unggul dibandingkan K-Nearest Neighbor dalam mengklasifikasikan sentimen ulasan pengguna. Hal ini ditunjukkan oleh nilai *accuracy*, *precision*, *recall*, *F1-score*, serta *AUC* yang secara konsisten lebih tinggi. Visualisasi hasil klasifikasi, baik melalui distribusi sentimen, *WordCloud*, maupun tren sentimen per bulan, turut memperkaya interpretasi hasil dengan menampilkan pola dan dinamika sentimen pengguna secara lebih intuitif. Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan, antara lain penggunaan data yang hanya bersumber dari satu platform serta pembagian sentimen yang didasarkan pada rating ulasan. Selain itu, penelitian ini belum mengeksplorasi penggunaan algoritma lain atau pendekatan berbasis *deep learning* yang berpotensi memberikan performa lebih tinggi. Oleh karena itu, penelitian selanjutnya diharapkan dapat memperluas sumber data, menambahkan kategori sentimen yang lebih beragam, serta membandingkan model klasik dengan model pembelajaran mendalam. Dengan demikian, hasil analisis sentimen yang diperoleh dapat menjadi semakin komprehensif dan akurat dalam mendukung pengambilan keputusan berbasis data.

REFERENCES

[1] L. Kusneti, A. , Ratu, and A. , Wijaya, “Analisis Sentimen Ulasan Aplikasi LinkedIn Dalam Google Play Store Dengan Model Naive Bayes,” *Djtechno J. Teknol. Inf.*, vol. 4, no. 2, pp. 374–385, 2023, doi: 10.46576/djtechno.

[2] F. T. Berton *et al.*, “Perbandingan Naïve Bayes Dan K-Nearest Neighbor Untuk Analisis Sentimen Terhadap Ulasan Aplikasi Threads,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 1, pp. 1–7, 2023.

[3] A. M. Alhakim, P. D. Atika, H. Herlawati, and T. Fax, “Analisis Sentimen Masyarakat Terhadap PHK di Indonesia Pada Twitter Menggunakan Naïve Bayes dan Support Vector Machine (SVM),” *JSRCS J. Students Res. Comput.*

Sci., vol. 6, no. 1, pp. 113–124, 2025.

[4] M. R. Manoppo, I. C. Kolang, D. N. Fiat, R. Michelly, and C. Mawara, "Analisis Sentimen Publik Di Media Sosial Terhadap Kenaikan PPN 12 % Di Indonesia Menggunakan Indobert," *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 4, no. 2, pp. 152–163, 2025.

[5] A. F. Zein, S. J. Putra, and Q. Aini, "Bibliometrik Analisis Sentimen : Tren Metode Penelitian dan Domain Implementasi," *JUSIFOR J. Sist. Inf. dan Inform.*, vol. 4, no. 1, pp. 120–128, 2025.

[6] F. Sains, U. Islam, N. Syarif, and H. Jakarta, "Sentimen Analisis Twitter Ibu Kota Negara Nusantara Menggunakan Long Short-Term Memory dan Lexicon Based," *J. Manaj. Sist. Inf. dan Teknol.*, vol. 8, no. 200, 2022.

[7] F. Adelia Irawan, A. Rialdy Atmadja, and A. Wahana, "Analisis Sentimen Ulasan Aplikasi Bank Digital Menggunakan Algoritma Naïve Bayes," *J. Comput. Sci. Inf. Technol.*, vol. 4, no. 2, pp. 60–68, 2024, doi: 10.47065/explorer.v4i2.1181.

[8] Y. A. Singgalen, "Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT)," *J. Media Inform. Budidarma*, vol. 7, no. 3, p. 1551, 2023, doi: 10.30865/mib.v7i3.6461.

[9] D. Pramudita, Y. Akbar, and T. Wahyudi, "Analisis Sentimen Terhadap Program Kartu Indonesia Pintar Kuliah pada Media Sosial X Menggunakan Algoritma Naive Bayes," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1420–1430, 2024, doi: 10.57152/malcom.v4i4.1565.

[10] D. Kurniawan and M. Yasir, "Optimization Sentimen Analysis using CRISP-DM and Naive Bayes Methods Implemented on Social Media," *Cybersp. J. Pendidik. Teknol. Inf.*, vol. 6, no. 2, p. 74, 2022, doi: 10.22373/cj.v6i2.12793.

[11] B. Sifa Amalia, Y. Umaidah, and R. Mayasari, "Analisis Sentimen Review Pelanggan Restoran Menggunakan Algoritma Support Vector Machine Dan K-Nearest Neighbor," *SITEKIN J. Sains, Teknol. dan Ind.*, vol. 19, no. 1, pp. 28–34, 2021.

[12] M. Jamil, R. Rosihan, and M. S Hanafi, "Analisis Sentimen Calon Kepala Daerah Maluku Utara dengan Metode CRISP-DM," *bit-Tech*, vol. 7, no. 3, pp. 995–1003, 2025, doi: 10.32877/bt.v7i3.2285.

[13] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.

[14] I. H. Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 302–307, 2023, doi: 10.30591/jpit.v8i3.5734.

[15] N. Q. Rizkina and F. N. Hasan, "Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes," *J. Inf. Syst. Res.*, vol. 4, no. 4, pp. 1136–1144, 2023, doi: 10.47065/josh.v4i4.3803.

[16] M. Hafizh Mahendra, D. Triantoro Murdiansyah, and K. Muslim Lhaksmana, "Analisis Sentimen Tweet COVID-19 menggunakan K-Nearest Neighbors dengan TF-IDF dan Ekstraksi Fitur CountVectorizer," *Dike J. Ilmu Multidisiplin* , vol. 1, no. 2, pp. 37–43, 2023, [Online]. Available: https://ejournal.cvrobema.com/index.php/dike/article/view/35

[17] Z. Purwanti and Sugiyono, "Pemodelan Text Mining untuk Analisis Sentimen Terhadap Program Makan Siang Gratis di Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM)," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 3, pp. 3065–3079, 2024, doi: 10.35870/jimik.v5i3.1001.

[18] A. A. Lailany and S. Lestari, "Analisis Sentimen Publik Terhadap Penurunan Jumlah Pernikahan di Indonesia menggunakan Algoritma K-Nearest Neighbors (KNN)," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 3, pp. 3043–3053, 2024, doi: 10.35870/jimik.v5i3.1000.

[19] H. Akbar Kosasih and A. Yuliana, "Analisis Sentimen Rencana Pendirian Pabrik Apple Inc Di Bandung Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 5, pp. 8396–8404, 2025, doi: 10.36040/jati.v9i5.14986.