KOMPARASI KINERJA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI SENTIMEN PENGGUNA APLIKASI DEEPSEEK AI

Fathoni, Dayana Khoiriyah Harahap, Eva Theresia Pardede,
Syakillah Nachwa, Muammar Ramadhani Maulizidan, Ali Ibrahim
Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya
Jalan Palembang - Prabumulih No.KM. 32, Indralaya Indah, Kec. Indralaya Kab. Ogan Ilir Sumatera Selatan
09031282227081@student.unsri.ac.id

ABSTRAK

Perkembangan *Artificial Intelligence* (AI), khususnya dalam *Generative Artificial Intelligence* (GenAI), telah menghasilkan berbagai model berbasis *Large Language Model* (LLM) yang mampu memahami dan menghasilkan teks menyerupai bahasa manusia. DeepSeek AI menjadi model yang baru-baru ini menarik perhatian publik. Penelitian ini menguji performa algoritma *Naïve Bayes* dan *K-Nearest Neighbors* (KNN) dalam analisis sentimen aplikasi DeepSeek AI berdasarkan *accuracy, precision, recall,* dan *F1-score*. Data dibagi menggunakan *split data* dengan rasio 70:30 dan 80:20, serta diterapkan *Stratified K-Fold Cross Validation* untuk memastikan generalisasi model yang optimal. Hasil terbaik diperoleh pada skema 70:30, di mana *Naïve Bayes* mencapai *accuracy* 81%, *precision* 87%, *recall* 50%, dan *F1-score* 64%, sedangkan KNN memperoleh *accuracy* 80%, *precision* 71%, *recall* 70%, dan *F1-score* 70%. Hasil *cross validation* menunjukkan bahwa rata-rata akurasi KNN mencapai 80.78%, sedangkan *Naïve Bayes* adalah 80.13%. Meskipun *Naïve Bayes* memiliki *precision* lebih tinggi, *recall* yang rendah mengurangi keseimbangan klasifikasi. Sebaliknya, KNN menunjukkan performa lebih stabil dengan keseimbangan *precision* dan *recall*, menjadikannya model yang lebih direkomendasikan untuk analisis sentimen ulasan aplikasi DeepSeek AI.

Kata kunci : *Analisis Sentimen, DeepSeek AI, K-Nearest Neighbor, Naïve Bayes, AI, Komparasi*

1. PENDAHULUAN

Berkembangnya kecerdasan buatan (AI) dalam beberapa tahun terakhir telah berkontribusi secara signifikan terhadap kemajuan teknologi, terutama dalam bidang *Generative Artificial Intelligence* (GenAI). Salah satu bentuk GenAI yang saat ini banyak digunakan adalah *Large Language Model* (LLM) yang berbasis jaringan saraf tiruan yang dapat menyerupai bahasa manusia. Inovasi teknologi terkait dengan perkembangan LLM ini, seperti Chat-GPT, Claude, dan Gemini [1].

Salah satu inovasi dalam penerapan LLM ini adalah DeepSeek AI yang merupakan salah satu model kecerdasan buatan terbaru yang menarik perhatian karena bersifat *open-source* dan dikembangkan di Tiongkok dengan pendekatan *Reinforcement Learning* (RL) untuk meningkatkan kemampuan penalaran tanpa bergantung pada *Supervised Fine-Tuning* (SFT) [2].

DeepSeek AI disebut-sebut mampu menyaingi kompetitor lainnya seperti ChatGPT. Model ini semakin mendapat sorotan setelah peluncurannya karena menunjukkan performa yang setara dengan model AI terkemuka, seperti GPT-4.0 dan Claude-3.5-Sonnet, tetapi dengan biaya yang lebih rendah [2].

Berdasarkan uji kemampuan, DeepSeek AI mampu bersaing dengan Chat-GPT di beberapa aspek, seperti pembuatan konten kreatif, DeepSeek AI mampu menghasilkan konsep cerita lengkap lebih cepat dibandingkan Chat-GPT yang hanya menawarkan beberapa ide singkat, sedangkan pada uji coba pemrograman, DeepSeek AI adalah pilihan yang menarik bagi para pengembang karena dapat mengatasi tantangan pemrograman yang rumit yang bahkan tidak dapat dikerjakan oleh Chat-GPT o1. Selain itu, dalam hal ringkasan informasi, DeepSeek AI dapat memberikan konteks yang lebih mendalam dibandingkan Chat-GPT [3].

Saat ini, pengguna dapat mengakses DeepSeek AI melalui situs web resminya di *https://chat.deepseek.com/* atau mengunduh aplikasi selulernya dari Google Play Store. Data terbaru yang didapat dari Google Playstore, menunjukkan bahwa aplikasi DeepSeek AI telah diunduh sebanyak 10 juta kali dengan total 94 ribu ulasan pengguna. Jumlah ulasan pengguna yang signifikan ini menjadikan analisis sentimen sebagai aspek krusial dalam memahami tanggapan publik terhadap model AI yang baru diluncurkan ini. Melalui analisis sentimen, pola opini pengguna, baik positif maupun negatif dapat diidentifikasi, sehingga pada kemudian hari dapat dimanfaatkan oleh pengembang untuk meningkatkan aplikasi mereka di masa yang akan datang [4].

Analisis sentimen merupakan salah satu pengimplementasian dari *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi maksud atau opini seseorang berdasarkan teks yang mereka tulis, seperti ulasan, komentar, atau umpan balik [5].

K-Nearest Neighbor (KNN) dan *Naïve Bayes* merupakan beberapa contoh model yang sering digunakan dalam analisis sentimen publik.

Beberapa studi telah membandingkan performa algoritma dalam analisis sentimen ulasan aplikasi kecerdasan buatan (AI). Widaad dan Anggraini [6] menemukan bahwa *Support Vector Machine* (SVM)

lebih unggul dibandingkan *Convolutional Neural Network* (CNN) dalam mengevaluasi sentimen ulasan ChatGPT di Google Play Store, dengan akurasi mencapai 85%. Studi lain oleh Ningsih dkk. [7] menemukan bahwa algoritma *Decision Tree C4.5* dapat mencapai akurasi 83,28% dalam analisis sentimen aplikasi Generative AI.

Penelitian lain menunjukkan variasi akurasi pada penggunaan *Multinomial Naïve Bayes Classifier* (NBC) dalam analisis sentimen ChatGPT di platform X. Rahayu dkk. [8] mencatat akurasi 63%, sementara Sari dkk. [9] mencapai 99%. Hidayatullah dkk. [10] menggunakan *Naïve Bayes* yang telah dioptimasi melalui Information Gain dan SMOTE menghasilkan akurasi 87,20%. Selain itu, penelitian oleh Hilda dan Elly [11] menunjukkan bahwa SVM mampu mencapai akurasi 89,44% dalam menganalisis sentimen aplikasi Bing: Chat with AI & GPT-4 di Google Play Store. Sabir dkk. [12] juga menemukan bahwa SVM dengan fitur TF-IDF menghasilkan akurasi tertinggi (96,41%) dibandingkan algoritma lain dalam analisis sentimen tweet tentang ChatGPT.

Studi mengenai analisis sentimen pada aplikasi kecerdasan buatan telah banyak dilakukan, namun studi yang secara spesifik membahas DeepSeek AI masih jarang ditemukan. Maka dari itu, kajian ini berfokus pada komparasi kinerja algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* dalam mengevaluasi sentimen pengguna terhadap aplikasi DeepSeek AI. Hasil dari studi ini diharapkan dapat menyajikan wawasan yang lebih komprehensif terkait efektivitas kedua algoritma yang diteliti, serta menjadi bahan pertimbangan bagi pengembang dalam meningkatkan pengalaman pengguna pada aplikasi AI.

2. TINJAUAN PUSTAKA

2.1. Komparasi

Komparasi adalah suatu cara dalam statistik yang diaplikasikan untuk mengevaluasi serta membandingkan kondisi, karakteristik, atau pola yang terdapat pada dua atau lebih kelompok. Metode ini bertujuan untuk mengidentifikasi perbedaan maupun kesamaan yang signifikan di antara kelompok-kelompok tersebut, baik dari segi variabel tertentu maupun faktor yang mempengaruhi kondisi yang sedang diteliti [13].

2.2. Analisis Sentimen

Analisis sentimen meneliti bagaimana bahasa tertulis merefleksikan informasi emosional dan mengembangkan teknik komputasi untuk interpretasinya. Antarmuka manusia-komputer masih sangat bergantung pada teks, sehingga teks dianggap sebagai medium penting untuk mengekstraksi data emosional [14].

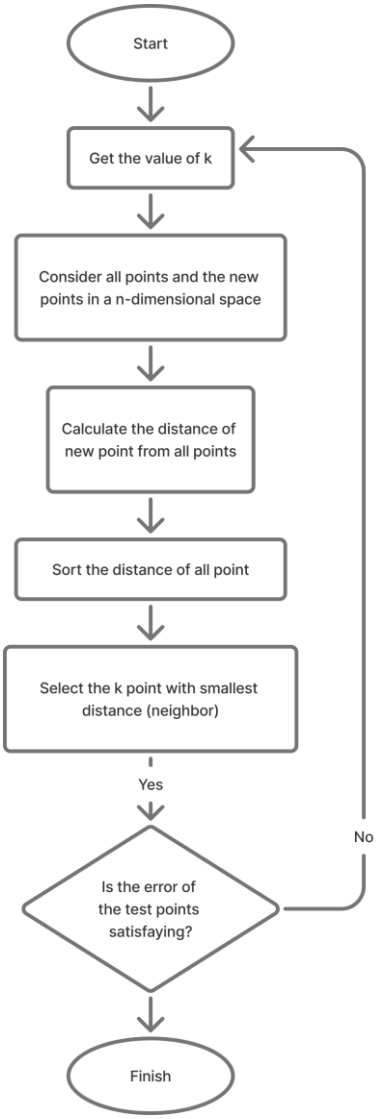
Sebelum analisis sentimen dilakukan, diperlukan tahap *preprocessing* untuk membersihkan data dari elemen tidak relevan agar lebih optimal dalam proses klasifikasi [15].

Tahapan tersebut di antaranya sebagai berikut.

a. *Cleansing* adalah tahapan mengeliminasi data yang tidak sesuai dan data duplikat guna mengurangi *noise* [16].
b. *Case Folding* merupakan tahap memodifikasi kata secara keseluruhan menjadi huruf non-kapital [17].
c. *Tokenisasi* adalah pemisahan teks menjadi elemen-elemen yang lebih kecil dan bermakna [18].
d. *Stopword removal* adalah tahapan mengeliminasi kata yang kurang relevan saat klasifikasi [19].
e. *Lemmatization* adalah mengubah sebuah kata menjadi bentuk kanoniknya, yaitu bentuk dasar yang sesuai dengan kamus [20][21].

2.3. Algoritma *K-Nearest Neighbor* (KNN)

Algoritma KNN merupakan model berbasis *supervised learning* yang menentukan kelas suatu data baru berdasarkan mayoritas kelas dari k tetangga terdekatnya [22].

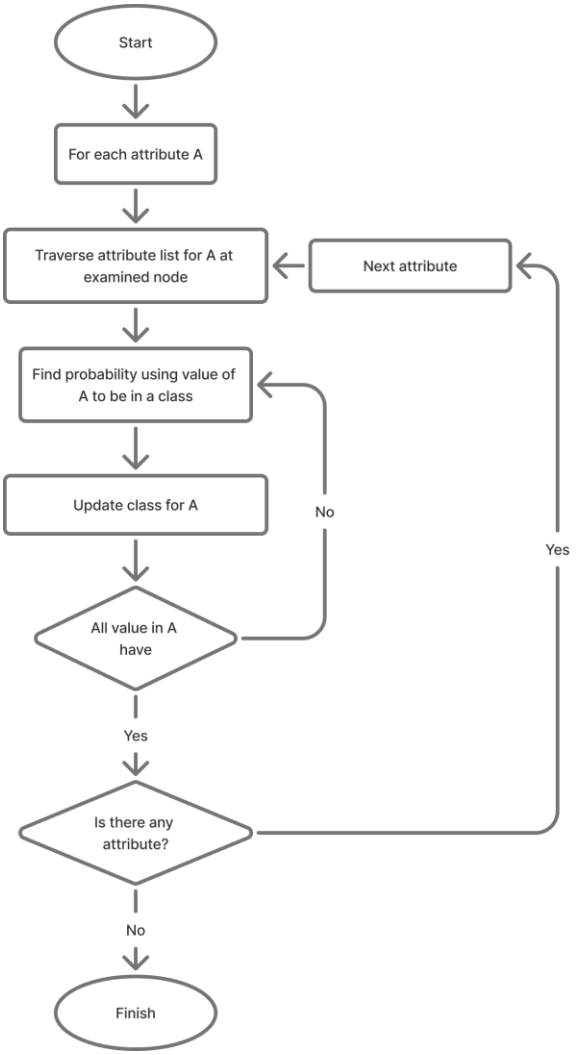


Gambar 1. Flowchart algoritma KNN

Gambar 1 menunjukkan alur kerja KNN dalam mengklasifikasikan data. Dimulai dengan menentukan nilai *k*, kemudian menghitung dan mengurutkan jarak data baru terhadap semua data yang ada. *K* tetangga terdekat dipilih untuk menentukan kelas berdasarkan mayoritas. Proses dievaluasi hingga error klasifikasi berada dalam batas yang dapat diterima.

2.4. Algoritma *Naïve Bayes*

Algoritma *Naïve Bayes* adalah sebuah model yang menelaah seluruh probabilitas dengan memadukan variasi nilai dan hitungan kemunculan yang bersumber dari data set [23].



Gambar 2. Flowchart algoritma *Naïve Bayes*

Diagram ini menjelaskan alur kerja Naive Bayes dalam mengklasifikasikan data. Setiap atribut dievaluasi untuk menghitung probabilitas keterkaitannya terhadap kelas tertentu. Probabilitas dihitung berdasarkan nilai dari atribut yang diamati, kemudian digunakan untuk memperbarui prediksi kelas. Proses ini diulang untuk semua atribut hingga seluruh nilai atribut dianalisis, dan klasifikasi akhir dapat ditentukan.

2.5. *Confusion Matrix*

Confusion matrix digunakan untuk mengevaluasi kesesuaian prediksi model dengan data aktual [24].

Matriks ini membantu menganalisis akurasi prediksi kelas positif dan negatif serta mengidentifikasi kesalahan prediksi [25].

Hasil prediksi dibagi menjadi empat kategori, yakni *True Positive* dan *True Negative* yang menunjukkan prediksi benar, serta *False Positive* dan *False Negative* yang mencerminkan kesalahan prediksi. [26].

$$F1 - score = 2 * \frac{Recall*Precision}{Recall+Precision} \quad (1)$$

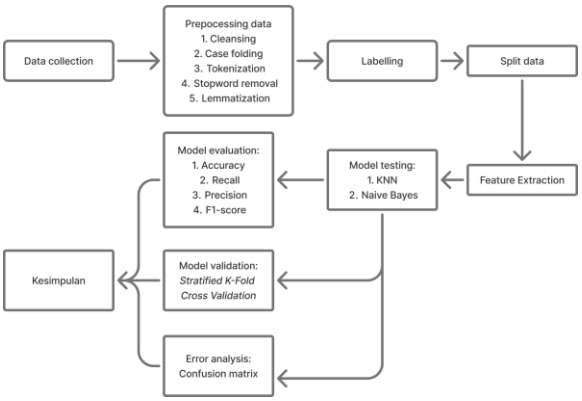
$$Accuracy = \frac{TP+TN}{Total\ Data} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

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

3. METODE PENELITIAN

Studi ini diawali dengan proses pengumpulan data, pra-pemrosesan data, pelabelan data, penyeimbangan distribusi data, pembagian data, ekstraksi fitur, evaluasi model, dan validasi model. Rangkaian tahapan penelitian ini akan diilustrasikan secara rinci pada Gambar 3.



Gambar 3. Alur penelitian

3.1. *Data Collection*

Data diperoleh melalui proses *scraping* ulasan aplikasi DeepSeek AI di Google Play Store menggunakan Google Play Scraper. Ulasan yang dikumpulkan berbahasa Inggris karena jumlahnya lebih banyak, sehingga menghasilkan dataset yang lebih kaya dan representatif untuk analisis sentimen. Dataset tersebut disimpan dalam format CSV.

3.2. *Preprocessing Data*

Tahapan *preprocessing data* bertujuan untuk membersihkan dan menyiapkan data sebelum klasifikasi. Tahapannya meliputi *cleansing*, *case folding*, *tokenization*, *stopword removal*, dan *lemmatization*.

3.3. *Data Labeling*

Dalam penelitian ini, pelabelan sentimen dilakukan untuk mengklasifikasikan muatan emosional teks ke dalam kategori positif atau negatif sebagai dasar untuk analisis lebih lanjut.

3.4. *Split Data*

Untuk mengevaluasi keandalan model, data dibagi menjadi dua kelompok, yaitu data latih dan data uji. Data latih digunakan sebagai dasar dalam membangun model, sementara data pengujian berfungsi untuk menilai kinerjanya. Pembagian data menggunakan dua skema, yaitu 70:30 dan 80:20.

3.5. *Feature Extraction* (TF-IDF)

TF-IDF diterapkan untuk menghitung tingkat keterulangan suatu kata dalam dokumen serta signifikansinya di seluruh kumpulan dokumen. Pendekatan ini berperan dalam mengurangi kompleksitas data dan meningkatkan efektivitas model.

3.6. Model *Testing*

Penelitian ini menerapkan algoritma KNN dan *Naïve Bayes* dalam proses klasifikasi. Hasil klasifikasi yang dibandingkan meliputi metrik evaluasi utama, yaitu *accuracy*, *precision*, *recall*, dan F1-*score*. Hasil dari kedua algoritma dikomparasikan berdasarkan metrik tersebut untuk menentukan model dengan performa terbaik.

3.7. Model Evaluasi dan Validasi

Evaluasi klasifikasi algoritma *Naïve Bayes* dan KNN dilakukan dengan *confusion matrix* untuk membandingkan performa algoritma tersebut. Untuk memvalidasi hasilnya dimanfaatkan teknik Model validasi menggunakan metode *Stratified K-Fold Cross Validation.*

4. HASIL DAN PEMBAHASAN

4.1. *Data Collection*

Sebanyak 10.000 ulasan pengguna dari Google Play Store dalam periode 14 Januari 2025 hingga 24 Februari 2025 terhadap aplikasi DeepSeek AI dikumpulkan sebagai dataset untuk analisis sentimen.

4.2. *Preprocessing Data*

Data yang diimplementasikan dalam tahapan ini sebanyak 9.952 ulasan setelah dilakukan penghapusan 48 duplikat. Rangkaian tahapan dalam prapemrosesan data disajikan dalam Tabel 1.

Tabel 1. Hasil tahapan *preprocessing*

Proses	Hasil
Ulasan asli	Very good 👍 I like this app but some time app is very slow in answering
Cleansing	Very good I like this app but some time app is very slow in answering
Case Folding	very good i like this app but some time app is very slow in answering
Tokenization	['very', 'good', 'i', 'like', 'this', 'app',
	'but', 'some', 'time', 'app', 'is', 'very', 'slow', 'in', 'answering']
Stopword Removal	['good', 'like', 'app', 'time', 'app', 'slow', 'answering']
Lemmatization	['good', 'like', 'app', 'time', 'app', 'slow', 'answer']

4.3. *Labeling*

Proses pelabelan dilakukan dengan menggunakan VADER, yakni menggabungkan kosakata yang telah diperkaya dengan nilai sentimen serta aturan-aturan linguistik yang mempertimbangkan elemen penulisan seperti huruf kapital, tanda baca, dan kata penguat (misalnya "sangat", "terlalu") yang dapat memengaruhi intensitas sentimen. Dalam penelitian ini, setiap ulasan dianalisis untuk memperoleh skor sentimen komposit yang berada dalam rentang -1 (sangat negatif) hingga +1 (sangat positif). Ulasan dengan skor ≥ 0.05 diklasifikasikan sebagai sentimen "Positif", sedangkan skor ≤ -0.05 dikategorikan sebagai "Negatif". Ulasan dengan nilai netral tidak digunakan dalam tahap analisis. Hasil dari proses pelabelan ini menghasilkan 6.710 ulasan berlabel "Positif" dan 3.242 ulasan berlabel "Negatif".

4.4. *Split Data*

Penelitian ini menggunakan dua skema pembagian data, yaitu 70:30 dan 80:20. Hasil pembagian data dapat dilihat pada Tabel 2.

Tabel 2. Hasil skema

Skema	Data Training	Data Testing
70:30	6.966	2.986
80:20	7.961	1.991

4.5. *Feature Extraction* (TF-IDF)

Pada proses TF-IDF ini dilakukan evaluasi frekuensi kemunculan kata dalam suatu kumpulan dokumen (*corpus*) guna mengidentifikasi pola distribusinya. Pendekatan ini memungkinkan penentuan kata yang paling dominan muncul digunakan dalam teks, sekaligus mempertimbangkan signifikansinya dalam berbagai dokumen yang dianalisis.

```
descartes : 9.319595547069767
descent : 10.012742727629712
describe : 8.62644836650982
describes : 9.319595547069767
description : 7.5278360778417115
descriptive : 10.012742727629712
desde : 9.319595547069767
desert : 9.319595547069767
deserve : 7.179529383573495
deserves : 7.710157634635666
desesperante : 9.319595547069767
design : 7.304692526527502
designer : 10.012742727629712
```

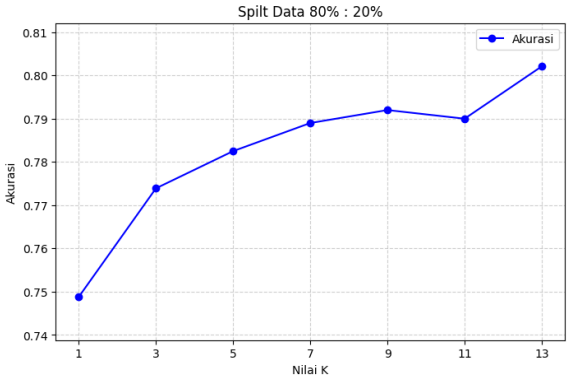
Gambar 4. Hasil pembobotan TF-IDF

4.6. *Model Testing*

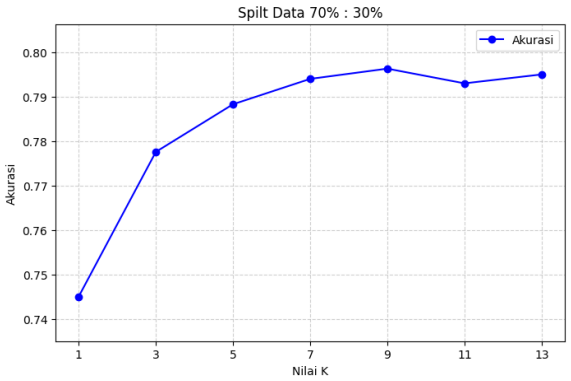
Tahapan ini mengevaluasi performa K-Nearest Neighbor dan Naïve Bayes pada skema data 70:30 dan 80:20 untuk membandingkan akurasi dan kestabilan klasifikasi sentimen.

4.7. KNN

Hasil evaluasi pada Gambar 5 menunjukkan bahwa split 80:20 dengan K=13 mencapai akurasi 0.8021, sedangkan split 70:30 pada Gambar 6 menunjukan dengan K=9 memperoleh akurasi 0.7963. Perbedaan ini mengindikasikan bahwa dengan jumlah data pelatihan yang lebih banyak pada split 80:20, model dapat mengenali pola lebih baik, sehingga berdampak pada peningkatan akurasi.



Gambar 5. Kelas terbaik berdasarkan *split data* 80:20



Gambar 6. Kelas terbaik berdasarkan *split data* 70:30

Evaluasi KNN menunjukkan performa terbaik pada split 80:20 dengan K=13 mencapai akurasi 80%, mencerminkan keseimbangan presisi dan sensitivitas. Sementara pada split 70:30 dengan K=9, akurasi sedikit lebih rendah yaitu 79%. Hasil ini menunjukkan bahwa skema 80:20 memberikan performa yang lebih stabil dan akurat.

Tabel 3. Hasil skema KNN

Spilt Data	*Accuracy*	*Precision*	*Recall*	*F1-score*
70:30 (K=9)	0.79	0.69	0.68	0.69
80:20 (K=13)	0.80	0.71	0.70	0.70

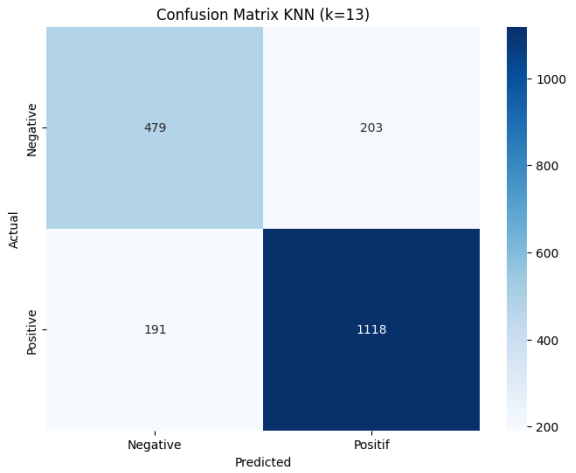
4.8. *Naïve Bayes*

Pengujian performa model Naïve Bayes dilakukan dengan dua skema pembagian data, yaitu 70:30 dan 80:20. Hasilnya menunjukkan akurasi yang konsisten sebesar 81% di kedua skema. Model yang diuji dengan pembagian data 70:30 menunjukkan *precision* sebesar 89%, sedikit lebih tinggi dibandingkan dengan skema 80:20 yang memiliki *precision* sebesar 87%. Hal ini menunjukkan bahwa model dalam skema 70:30 lebih efektif dalam menghindari kesalahan klasifikasi *false positive*. Namun, skema 80:20 memiliki *recall* yang lebih tinggi, yaitu 50% dibandingkan dengan 49% pada skema 70:30, yang mengindikasikan bahwa model lebih mampu mengenali kelas minoritas. Peningkatan *recall* pada skema 80:20 juga berdampak pada peningkatan nilai *F1-score*, yaitu 64% dibandingkan dengan 63% pada skema 70:30. Berdasarkan hasil tersebut, skema pembagian data 80:20 lebih optimal karena menghasilkan keseimbangan yang lebih baik antara *precision* dan *recall*, sehingga meningkatkan kemampuan model dalam mengenali kelas minoritas tanpa mengorbankan performa keseluruhan.

Tabel 4. Hasil skema *Naïve Bayes*

Split Data	*Accuracy*	*Precision*	*Recall*	*F1-score*
70:30	0.81	0.89	0.49	0.63
80:20	0.81	0.87	0.50	0.64

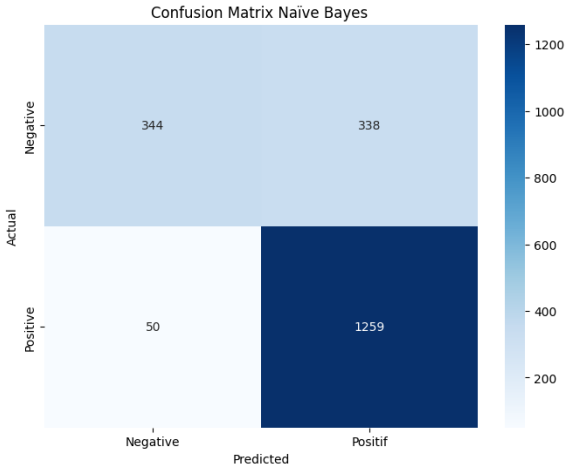
4.9. Hasil *Confusion Matrix* dan Visualisasi Kata



Gambar 7. *Confusion matrix* KNN

Hasil *confusion matrix* pada Gambar 7 dan Gambar 8 menunjukkan bahwa model *Naïve Bayes* berhasil mengklasifikasikan 344 data negatif dan 1.259 data positif dengan benar. Namun, model ini mengalami 338 kesalahan negatif sebagai positif dan 50 kesalahan positif sebagai negatif, yang menunjukkan kemampuannya dalam mengenali data positif lebih baik. Sementara itu, model KNN (K=13) mengklasifikasikan 479 data negatif dan 1.118 data positif secara akurat, tetapi menghasilkan 203 kesalahan negatif dan 191 kesalahan positif. Hal ini

menunjukkan bahwa KNN lebih unggul dalam mengidentifikasi data negatif dibandingkan Naïve Bayes. Secara keseluruhan, Naïve Bayes lebih efektif dalam mendeteksi data positif, sedangkan KNN lebih akurat dalam mengenali data negatif.



Gambar 8. *Confusion matrix Naïve Bayes*

Frekuensi kata yang muncul dalam ulasan dianalisis dan divisualisasikan berdasarkan label. Representasi visual per label disajikan dalam bentuk *word cloud* pada Gambar 9 dan Gambar 10.

Gambar 9. Hasil visualisasi kata positif

Gambar 10. Hasil visualisasi kata negatif

4.10. Hasil *Model Evaluation* dan *Validation*

Hasil evaluasi model menunjukkan bahwa *Naïve Bayes* dan KNN memiliki performa yang seimbang, tetapi dengan keunggulan pada aspek yang berbeda. *Naïve Bayes* mencapai akurasi sebesar 80.51% dengan *precision* tinggi 87.31%, yang menunjukkan kemampuannya dalam menghasilkan prediksi positif yang lebih akurat. Namun, *recall* yang hanya 50.44% menunjukkan keterbatasan dalam menangkap seluruh kelas positif, dengan *F1-score* sebesar 63.94%. Sementara itu, KNN memiliki akurasi yang hampir setara 80.21%, tetapi dengan keseimbangan yang lebih baik antara *precision* 71.49% dan *recall* 70.23%, sehingga menghasilkan *F1-score* yang lebih tinggi 70.86%. Dengan demikian, dalam kasus klasifikasi ulasan aplikasi, KNN lebih direkomendasikan karena memiliki *trade-off* yang lebih baik antara ketepatan dan sensitivitas dibandingkan *Naïve Bayes*.

Tabel 4. Model *evaluation*

Algoritma	***Accuracy***	***Precision***	***Recall***	***F1_Score***
KNN	0.80	0.71	0.70	0.70
Naïve Bayes	0.81	0.87	0.50	0.64

Berdasarkan hasil evaluasi validasi model menggunakan *Stratified K-Fold Cross Validation* dengan *5 fold*, model KNN menghasilkan rata-rata akurasi sebesar 0.8078 dengan akurasi pada tiap fold berkisar antara 0.7911 hingga 0.8155. Selain itu, model *Naïve Bayes* mencatatkan rata-rata akurasi 0.8013, dengan akurasi pada tiap fold berkisar antara 0.7925 hingga 0.8121.

Tabel 5. Model validation KNN dan Naïve Bayes

KNN					
n-*fold*	1	2	3	4	5
Accuracy	0.8128	0.8155	0.7911	0.8047	0.8148
Mean	0.8078				
Naïve Bayes					
n-*fold*	1	2	3	4	5
Accuracy	0.8121	0.7976	0.7925	0.8011	0.8033
Mean	0.8013				

5. KESIMPULAN DAN SARAN

Hasil studi ini menunjukkan bahwa dalam analisis sentimen ulasan pengguna terhadap DeepSeek AI, *K-Nearest Neighbor* (KNN) lebih disarankan dibandingkan *Naïve Bayes* karena memiliki keseimbangan yang lebih baik antara *precision* dan *recall*. KNN memiliki akurasi sebesar 80.21% dengan *precision* 71.49% dan *recall* 70.23%, menghasilkan *F1-score* tertinggi sebesar 70.86%. Sementara itu, *Naïve Bayes* memang memiliki *precision* lebih tinggi 87.31%, tetapi *recall* yang rendah 50.44% menyebabkan *F1-score* hanya mencapai 63.94%. Meskipun *Naïve Bayes* lebih akurat dalam memprediksi kelas positif, model ini kurang sensitif dalam menangkap keseluruhan data positif. Oleh karena itu, algoritma KNN direkomendasikan untuk analisis sentimen dikarenakan lebih stabil dalam

klasifikasi, sedangkan *Naïve Bayes* dapat dioptimalkan lebih lanjut atau dikombinasikan dengan metode lain untuk meningkatkan presisi dan keseimbangan performa model. Penelitian selanjutnya, disarankan untuk mengeksplorasi teknik lain, seperti penggunaan model berbasis *deep learning*, misalnya LSTM atau Transformer yang dapat menangkap konteks lebih baik dalam analisis sentimen. Selain itu, memperluas dan memperkaya dataset dapat digunakan untuk meningkatkan generalisasi model serta mengurangi bias klasifikasi.

DAFTAR PUSTAKA

[1] A. Borji and M. Mohammadian, "Battle of the Wordsmiths: Comparing ChatGPT, GPT-4, Claude, and Bard," *SSRN Electronic Journal*, Jun. 2023, doi: 10.2139/ssrn.4476855.

[2] DeepSeek-AI *et al.*, "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning," *ArXiv*, Jan. 2025, [Online]. Available: http://arxiv.org/abs/2501.12948

[3] G. Fraser, "DeepSeek vs ChatGPT-how do they compare?," https://www.bbc.com/news/articles/cqx9zn27700o. Accessed: Feb. 18, 2025. [Online]. Available: https://www.bbc.com/news/articles/cqx9zn27700o

[4] E. Apriani and N. Pratiwi, "Analisis Sentimen Game Show Clash of Champions Ruangguru dengan Algoritma KNN dan SVM," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 1, pp. 39–51, Jan. 2025, doi: 10.52436/1.jpti.556.

[5] E. Hokijuliandy, H. Napitupulu, and F. Firdaniza, "Analisis Sentimen Menggunakan Metode Klasifikasi Support Vector Machine (SVM) dan Seleksi Fitur Chi-Square," *Jurnal SisInfo (Sistem Informasi dan Informatika)*, vol. 5, no. 2, pp. 40–49, Aug. 2023, doi: https://doi.org/10.37278/sisinfo.

[6] N. Widaad and D. Anggraini, "SENTIMENT ANALYSIS OF CHATGPT APP USER REVIEWS USING SVM AND CNN METHODS," *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 6, pp. 1687–1700, Dec. 2024, doi: https://doi.org/10.52436/1.jutif.2024.5.6.4010.

[7] W. Ningsih, Rahmaddeni, S. Adrianto, F. Kurniawan, and B. Alfianda, "Analisis Sentimen Pengguna Aplikasi Gen Ai dari Appstore dan Googleplay Menggunakan Algoritma C45," *The Indonesian Journal of Computer Science*, vol. 13, no. 5, pp. 8153–8164, Sep. 2024, doi: 10.33022/ijcs.v13i5.4327.

[8] D. Sri Rahayu, R. Novita, T. Khairil Ahsyar, and Zarnelly, "Sentiment Analysis ChatGPT Using the Multinominal Naïve Bayes Classifier (NBC) Algorithm," *Jurnal Sistem Cerdas*, vol. 7, no. 1, pp. 66–74, Apr. 2024, doi: 10.37396/jsc.v7i1.388.

[9] Y. K. Sari, F. Rozi, S. Muhyiddin, and F. Sukmana, "SENTIMENT ANALYSIS OF PUBLIC OPINION ON APPLICATION X (TWITTER) IN INDONESIA AGAINST CHATGPT USING NAÏVE BAYES ALGORITHM," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 4, pp. 2473–2484, Dec. 2024, doi: 10.29100/jipi.v9i4.7052.

[10] H. Hidayatullah, P. Purwantoro, and Y. Umaidah, "PENERAPAN NAÏVE BAYES DENGAN OPTIMASI INFORMATION GAIN DAN SMOTE UNTUK ANALISIS SENTIMEN PENGGUNA APLIKASI CHATGPT," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 3, pp. 1546–1553, Oct. 2023, doi: 10.36040/jati.v7i3.6887.

[11] Muslih, A. M. Hilda, and M. J. Elly, "Metode Klasifikasi Support Vector Machine (SVM) Untuk Analisis Sentimen Aplikasi Bing: Chat with AI & GPT-4 Di Google Play Store," *PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika*, vol. 17, no. 1, pp. 68–76, Mar. 2024, doi: https://doi.org/10.33322/petir.v17i1.2283.

[12] A. Sabir, H. A. Ali, and M. A. Aljabery, "ChatGPT Tweets Sentiment Analysis Using Machine Learning and Data Classification," *Informatica*, vol. 48, no. 7, pp. 103–112, May 2024, doi: 10.31449/inf.v48i7.5535.

[13] M. Safrudin, M. Martanto, and U. Hayati, "PERBANDINGAN KINERJA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI SENTIMEN ULASAN GAME GENSHIN IMPACT," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 3182–3188, May 2024, doi: 10.36040/jati.v8i3.8415.

[14] E. Hauthal, D. Burghardt, C. Fish, and A. L. Griffin, "Sentiment Analysis," *International Encyclopedia of Human Geography, Second Edition*, pp. 169–177, Jan. 2020, doi: 10.1016/B978-0-08-102295-5.10593-1.

[15] A. F. Inayah, K. D. Tania, A. Wedhasmara, and A. Meiriza, "Knowledge Extraction Using Aspect-Based Sentiment Analysis and Classification Method," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 10, no. 3, pp. 378–388, Dec. 2024, doi: https://doi.org/10.26418/jp.v10i3.

[16] N. Putriani, F. R. Umbara, and P. N. Sabrina, "Analisis Sentimen pada Aplikasi PeduliLindungi dengan Menggunakan Metode Improved K-Nearest Neighbor dan Lexicon Based," *Jurnal Teknologi Informatika dan Komputer*, vol. 8, no. 1, pp. 350–364, Mar. 2022, doi: 10.37012/jtik.v8i1.1107.

[17] F. M. Sarimole and Kudrat, "Analisis Sentimen Terhadap Aplikasi Satu Sehat Pada Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine," *Jurnal Sains dan Teknologi*, vol. 5, no. 3, pp. 783–790, Feb. 2024,

doi: https://doi.org/10.55338/saintek.v5i1.2702.

[18] T. T. Mengistie and D. Kumar, "Deep Learning Based Sentiment Analysis On COVID-19 Public Reviews," in *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, IEEE, Apr. 2021, pp. 444–449. doi: 10.1109/ICAIIC51459.2021.9415191.

[19] N. A. Alabdulkarim, M. A. Haq, and J. Gyani, "Exploring Sentiment Analysis on Social Media Texts," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14442–14450, Jun. 2024, doi: 10.48084/etasr.7238.

[20] M. Horvat, G. Gledec, and F. Leontić, "Hybrid Natural Language Processing Model for Sentiment Analysis during Natural Crisis," *Electronics (Basel)*, vol. 13, no. 10, p. 1991, May 2024, doi: 10.3390/electronics13101991.

[21] O. Toporkov and R. Agerri, "On the Role of Morphological Information for Contextual Lemmatization," Feb. 2023.

[22] M. Suyal and P. Goyal, "A Review on Analysis of K-Nearest Neighbor Classification Machine Learning Algorithms based on Supervised Learning," *International Journal of Engineering Trends and Technology*, vol. 70, no. 7, pp. 43–48, Jul. 2022, doi: 10.14445/22315381/IJETT-V70I7P205.

[23] R. Rachman and R. N. Handayani, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *JURNAL INFORMATIKA*, vol. 8, no. 2, pp. 111–122, Sep. 2021, [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/ji

[24] Styawati, N. Hendrastuty, A. R. Isnain, and A. Y. Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 6, no. 3, pp. 150–155, Oct. 2021, doi: 10.30591/jpit.v6i3.2870.

[25] E. Eskiyaturrofikoh and R. R. Suryono, "ANALISIS SENTIMEN APLIKASI X PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE (SVM)," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 3, pp. 1408–1419, Aug. 2024, doi: 10.29100/jipi.v9i3.5392.

[26] M. R. Saputra and Parjito, "JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika) Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi ANALISIS SENTIMEN TWITTER TERHADAP KONFLIK DI PAPUA MENGGUNAKAN PERBANDINGAN NAIVE BAYES DAN SVM," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 1197–1208, Jun. 2025, doi: https://doi.org/10.29100/jipi.v10i3.6180