

Analisis Sentimen Review ChatGPT di Play Store menggunakan Support Vector Machine dan K-Nearest Neighbor

Adji Surya Pamungkas¹, Nuri Cahyono^{1,*}

¹ Program Studi Informatika, Universitas Amikom Yogyakarta, Indonesia

* Correspondence: nuricahyono@amikom.ac.id

Copyright: © 2024 by the authors

Received: 17 November 2023 | Revised: 28 Januari 2024 | Accepted: 30 Februari 2024 | Published: 20 Juni 2024

Abstrak

Aplikasi ChatGPT untuk Android diluncurkan pada 25 Juli 2023, model bahasa dari OpenAI meraih rating 4.8 hingga awal 2024. Meskipun mayoritas ulasan positif, laporan pengguna yang menyatakan bahwa ChatGPT memberikan jawaban yang tidak akurat menimbulkan kekhawatiran terhadap keandalan aplikasi ini. Penelitian ini bertujuan membandingkan model algoritma *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN) dalam menganalisis sentimen ulasan aplikasi ChatGPT. Menggunakan metode *text mining* untuk menemukan informasi dari teks, data dikumpulkan dari ulasan *Google Play Store* menggunakan teknik *data scraping* dan dianalisis dengan algoritma *Support Vector Machine* dan *K-Nearest Neighbor*. Pengujian kedua algoritma tersebut dilakukan validasi silang 5 lipatan dan *split data* menggunakan data latih 80% dan data tes 20%. Hasil klasifikasi sentimen menggunakan algoritma *Support Vector Machine* memiliki rerata akurasi sebesar 80% dan *K-Nearest Neighbor* sebesar 71%. Algoritma SVM memiliki keunggulan karena kemampuannya mengatasi kelemahan KNN terhadap fitur - fitur yang kurang relevan dan tidak memberikan kontribusi signifikan dalam pembuatan prediksi. Melalui hasil penelitian ini, diharapkan para pengembang mampu mengetahui dan menanggapi umpan balik pengguna terkait keandalan ChatGPT.

Kata kunci : chatgpt; sentimen; validasi silang; svm; knn

Abstract

The ChatGPT application for Android was launched on July 25, 2023, and the language model from OpenAI achieved a rating of 4.8 until early 2024. Despite the majority of positive reviews, user reports stating that ChatGPT provides inaccurate answers raise concerns about the reliability of this application. This research aims to compare the models of the *Support Vector Machine* (SVM) and *K-Nearest Neighbor* (KNN) algorithms in analyzing the sentiment of ChatGPT application reviews. Utilizing text mining methods to extract information from text, data was collected from *Google Play Store* reviews using data scraping techniques and analyzed with *Support Vector Machine* and *K-Nearest Neighbor* algorithms. Cross-validation with 5 folds and data split using 80% training and 20% testing data were applied to evaluate the performance of both algorithms. The sentiment classification results showed that the *Support Vector Machine* algorithm achieved an average accuracy of 80%, while *K-Nearest Neighbor* reached 71%. SVM excels due to its ability to overcome KNN's limitations regarding less relevant features that do not significantly contribute to predictions. The findings of this study are expected to help developers understand and respond to user feedback regarding the reliability of ChatGPT.

Keywords: chatgpt; sentiment; cross validation; svm; knn



PENDAHULUAN

Kemajuan kecerdasan buatan (AI) terbaru, yaitu ChatGPT, dikembangkan oleh OpenAI pada November 2022 (Dowling & Lucey, 2023). ChatGPT merespons interaksi manusia dan mesin dengan model jaringan saraf tiruan mendalam berparameter 175 miliar, fokus pada pemrosesan bahasa alami (Leippold, 2023). ChatGPT memiliki keunggulan utama dalam hal fleksibilitas dan kemudahan penggunaan (Hassani & Silva, 2023). Setiap aplikasi memiliki aspek positif dan negatif yang dapat menimbulkan berbagai tanggapan dari pengguna, seperti rasa puas atau kekecewaan terhadap aplikasi tersebut (Giovani et al., 2020).

Hingga awal 2024, ChatGPT meraih rating 4.8 di Google Play Store dengan banyak ulasan positif, tetapi juga mendapat kritik negatif terkait jawaban tidak akurat dan dugaan kebohongan. Menurut penelitian (Gravel et al., 2023), saat diminta menjawab beberapa pertanyaan medis, ChatGPT memberikan jawaban dengan kualitas yang terbatas untuk keperluan publikasi ilmiah. Yang lebih penting, ChatGPT memberikan referensi yang tampak nyata. Penting bagi pengguna ChatGPT untuk memeriksa referensi yang diberikan sebelum menggabungkannya ke dalam naskah (Gravel et al., 2023).

Masalah jawaban tidak akurat pada ChatGPT menjadi fokus analisis sentimen, untuk mengevaluasi kesesuaian model bahasa tersebut (Shen et al., 2023). Penerapan analisis sentimen dengan pendekatan *machine learning* diusulkan sebagai solusi efisien untuk mengatasi masalah ini (Alaei et al., 2017). Analisis sentimen, atau penambangan opini, merupakan studi komputasional tentang pendapat, perasaan, dan emosi dalam teks (Liu, 2012). Pemanfaatan analisis sentimen dapat meramalkan dan menganalisis mood publik serta ekspresi perasaan pengguna internet dan layanan teknologi (Amrullah et al., 2020).

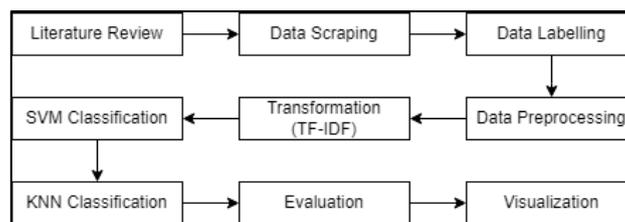
Ekspresi emosi dalam komentar pengguna aplikasi diungkapkan melalui tulisan, memicu penelitian ini menggunakan teknik *text mining* (Rahayu et al., 2022). *Text mining*, seperti yang dijelaskan oleh Sepri (2020) adalah proses ekstraksi informasi dari berbagai sumber teks. Penelitian ini menerapkan model *text mining* dengan algoritma SVM dan KNN untuk klasifikasi analisis sentimen, sesuai dengan beberapa studi sebelumnya yang relevan.

Penelitian yang dilakukan oleh Rifaldi et al. (2023) pada 301 tweet menggunakan metode Naive Bayes mencapai akurasi 80%, dengan precision 80,95% dan recall 89,47%. Analisis sentimen efektif mengindikasikan bahwa sentimen masyarakat terhadap aplikasi ChatGPT di Twitter cenderung positif, dengan presentase tweet positif sebesar 74,09%. Selain itu, Atmajaya et al. (2023) menggunakan SVM dan Naive Bayes. SVM mencapai akurasi tertinggi 59% dengan pembagian data testing 30% dan data training 70%. Pelabelan data dengan Vader dan RoBERTa terbukti efektif untuk mengklasifikasikan sentimen pada tweet, terutama label sentimen Vader. Sedangkan Yusuf & Masripah (2023) menggunakan 1729 atribut pada 2000 data untuk analisis sentimen ChatGPT. Dengan metode Naive Bayes yang dioptimalkan oleh PSO, ditemukan akurasi 69,23% dengan proporsi kelas positif 0,503 dan kelas negatif 0,497. Kurva AUC diperoleh sebesar 0,68 +/- 0,55. Sementara itu, Hidayatullah et al. (2023) menggunakan *machine learning* dengan algoritma Naive Bayes yang dioptimalkan oleh Information Gain dan SMOTE untuk analisis sentimen pada komentar pengguna aplikasi ChatGPT. Dari 10 model yang dievaluasi, model Opt2 mencapai kinerja terbaik dengan akurasi 87.20% dan F1-score 84.74%, menunjukkan peningkatan rata-rata akurasi 6.25% dan F1-score 25.44%.

Penelitian ini berfokus dengan menyajikan analisis sentimen ChatGPT berbasis ulasan pengguna aplikasi, dibandingkan dengan penelitian-penelitian sebelumnya yang menggunakan data tweet. Algoritma *Support Vector Machine* dan *K-Nearest Neighbor* digunakan dengan validasi silang 5 lipatan. Tujuan penelitian kami adalah untuk membandingkan model algoritma SVM dan KNN dalam menganalisis sentimen ulasan aplikasi ChatGPT.

METODE

Penelitian ini menggunakan 9 tahap dalam metode penelitian, meliputi *Literature Review*, *Data Scraping*, *Data Labelling*, *Data Preprocessing*, *Transformation (TF-IDF)*, *SVM Classification*, *KNN Classification*, *Evaluation*, dan *Visualization*. Tahapan penelitian dapat dilihat pada gambar 1. Tahap awal adalah melakukan *Literature Review* dengan mencari sumber pendukung untuk penelitian, termasuk teori-teori terkait analisis sentimen, teknik pengambilan data (*scraping*), pra-pemrosesan teks, algoritma SVM & KNN, dan lainnya. Selanjutnya, dalam tahap *Data Scraping*, data mentah yang digunakan dikumpulkan dari halaman aplikasi ChatGPT di Play Store menggunakan teknik *scraping* (Krotov et al., 2020). Teknik *scraping* yang digunakan adalah *library google-play-scraper* pada google colab. Tahap ketiga *Data Labelling*, sesuai penelitian (Lailiyah et al., 2017), peneliti memanfaatkan kamus *lexicon SentiWordNet* untuk mengatributkan setiap komentar ke dalam kelas yang sesuai. Data akan dikelompokkan ke dalam kelas positif jika skornya > 0 , ke dalam kelas negatif jika skornya < 0 , dan ke dalam kelas netral jika skornya $= 0$ (Rahayu et al., 2022).



Gambar 1. Tahapan Penelitian

Pada tahap *Data Preprocessing* teks dalam penelitian ini dilakukan proses secara bertahap, dimulai dengan tokenisasi, yang melibatkan pemisahan kata dari teks (Yusuf & Masripah, 2023). Selain itu, dilakukan juga langkah pembentukan huruf kecil, penghapusan kata-kata umum (*stopword removal*), dan pereduksian kata. Tujuan dari *Data Preprocessing* adalah untuk menghasilkan data yang relevan dengan memperoleh konten yang lebih mudah diproses dalam sistem (Ruslim et al., 2019). Tahap kelima, *Transformation(TF-IDF)*, sebelum masuk TF-IDF, data hasil dari preprocessing dibagi menjadi data latih dan data uji dengan rasio 80% :20%. TF-IDF menurut (Rofiqi et al., 2019), TF-IDF adalah perhitungan atau pembobotan kata melalui teknik tokenisasi, stopwords, dan stemming, dan frekuensi munculnya kata dalam dokumen.

Tahap *SVM Classification* dan *KNN Classification*, yaitu implementasi proses klasifikasi sentimen menggunakan algoritma SVM dan KNN untuk menentukan hasil klasifikasi pada kelas-kelas positif, negatif, dan netral. Proses klasifikasi dan regresi dapat diprediksi menggunakan teknik relatif baru dengan menggunakan algoritma SVM, yang dapat dikategorikan sebagai pengklasifikasi linear, non-probabilistik, dan biner (Handayani et al., 2020). Prinsip kerja KNN adalah mengklasifikasikan berdasarkan kedekatan (jarak) satu titik data terhadap yang lain (Prasetyo, 2013). Pada tahap klasifikasi, algoritma SVM dan KNN sama – sama menggunakan pengujian validasi silang sebanyak 5 lipatan (*cross validation=5*).

Pada tahap *Evaluation*, hasil klasifikasi sentimen dibuat *confusion matrix* untuk menganalisis hasil klasifikasi lebih rinci. *Confusion matrix* mencakup empat kategori: *True Positive*, *True Negative*, *False Positive*, dan *False Negative*. Ini membantu mengidentifikasi sejauh mana model dapat mengklasifikasikan ulasan dengan benar. Tahap akhir adalah *Visualization* data menggunakan pie chart dari *library matplotlib* untuk melihat jumlah keseluruhan ulasan mendapat sentimen positif, negatif, dan netral.

HASIL DAN PEMBAHASAN

Hasil

Hasil temuan kami pada tahap *Data Scraping*, dilakukan pengumpulan data ulasan aplikasi ChatGPT menggunakan library google-play-scraper di Google Colab pada tanggal 20 Januari 2024. Dataset yang berhasil diambil mencakup 4.712 ulasan, data ulasan dapat dilihat pada tabel 1.

Tabel 1. Data ulasan

Nama Pengguna	Tanggal	Ulasan
Fahmi Riyadi	18 Desember 2023	<i>The app is good, as I subscribed to the plus model, makes it even better. However, the aspect ratio for folding phone is still not great, it is just a scretched out layout when using the phone in its unfolded state.</i>
Ayu Gamon	13 November 2023	<i>ChatGPT is good enough for writing an full story. I've been using it for helping me writing my story, and the AI didn't forgot any piece of information I send it for a long time, incase I mentioned it again.</i>

Setelah memperoleh dataset ulasan, langkah selanjutnya adalah melakukan *Data Labelling* dengan menggunakan library SentiWordNet. Proses labelling ini dilakukan dengan membagi data ulasan ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral, berdasarkan skor sentimen dari SentiWordNet. Ulasan akan diklasifikasikan ke dalam kelas positif apabila skornya lebih besar dari 0, ke dalam kelas negatif apabila skornya kurang dari 0, dan ke dalam kelas netral apabila skornya sama dengan 0. Hasil pelabelan dapat dilihat pada tabel 2.

Tabel 2. Hasil pelabelan

Ulasan	Skor	Total Skor	Label
<i>It keeps crashing. I have a modern phone and a Plus subscription. It doesn't matter what I do, it just crashes when I prompt it.</i>	<i>It(0) keeps(-50) crashing(-325). I(0) have(0) a(0) modern(25) phone(0) and(0) a(0) Plus(0) matter(0) what(0) I(0) do(0), it(0) just(0) crashes(-50) when(0) I(0) prompt(0) it(0).</i>	-400	Negatif
<i>Loving the voice interface! Can't wait for the ability to voice chat at the same time as web browsing/using plugins.</i>	<i>Loving(100) the(0) voice(50) interface(0)! Can't(0) wait(25) for(0) the(0) ability(50) to(0) voice(50) chat(0) at(0) the(0) same(0) time(0) as(0) web(0) browsing(0)/using(0) plugins(0).</i>	275	Positif
<i>Its too slow.</i>	<i>Its(0) too(25) slow(-25).</i>	0	Netral

Selanjutnya, setelah dilakukan pelabelan data, didapatkan sebanyak 4.712 data ulasan yang telah mengandung sentimen positif, negatif, dan netral. Data tersebut dapat dilihat pada tabel 3. Data ulasan kemudian melalui empat tahap *Data Preprocessing*, dimulai dari *case folding*, *tokenize*, *remove stopword*, dan *stemming*. *Case folding* merupakan proses pengubahan teks menjadi huruf kecil. *Tokenize* atau biasa disebut tokenisasi merupakan proses pengubahan

teks menjadi token – token, proses ini dibantu *library word tokenize*. Selanjutnya, *remove stopwords* merupakan proses penghapusan kata- kata yang tidak penting, proses ini dibantu *library stopwords*. Lalu, *stemming* merupakan proses pengubahan kata berimbuhan menjadi kata dasar, proses ini dibantu dengan *library Porter Stemmer*. Hasil dari proses *preprocessing* dapat dilihat pada Tabel 4.

Tabel 3. Jumlah data

Kelas	Jumlah
Positif	2677
Negatif	1145
Netral	890
Total	4712

Tabel 4. Hasil *case folding* sampai *stemming*

Tahap	Hasil
Ulasan Awal	<i>The Chatgpt app is absolutely phenomenal</i>
<i>Case folding</i>	<i>the chatgpt app is absolutely phenomenal</i>
<i>Tokenize</i>	<i>[the, chatgpt, app, is, absolutely, phenomenal]</i>
<i>Remove Stopwords</i>	<i>[chatgpt, app, absolutely, phenomenal]</i>
<i>Stemming</i>	<i>[chatgpt, app, absolut, phenomen,]</i>

Selanjutnya, data hasil tahap *Data Preprocessing* diteruskan dengan *Transformation TF-IDF* setelah split data 80:20, menghasilkan 3769 data latih dan 943 data uji. Pembobotan TF-IDF dilakukan pada 3 dokumen dengan 3 kata pilihan (error, chatgpt, slow). Penghitungan menggunakan *library sklearn*, dan hasilnya tercatat dalam Tabel 5. Pada dokumen pertama (D1), kata "error" muncul satu kali dengan bobot 0,1494 menggunakan metode pembobotan TF-IDF.

Tabel 5. Hasil penghitungan TF-IDF

Term	Kemunculan Data			TF			TF.IDF		
	D1	D2	D3	D1	D2	D3	D1	D2	D3
<i>error</i>	1	0	0	0	0,1	0	0,1494	0	0
<i>chatgpt</i>	0	2	0	0	0,0876	0	0	0	0
<i>slow</i>	0	1	0	0	0,0876	0	0	0,08471	0

Pada *classification*, dilakukan klasifikasi sentimen menggunakan algoritma SVM dan KNN dimulai setelah proses *Transformation (TF-IDF)*. Proses ini menerapkan validasi silang selama 5 lipatan (*cross validation=5*) agar kinerja model dapat optimal. Setelah dilakukan klasifikasi untuk kedua algoritma menghasilkan nilai akurasi model tersebut. Hasil rata – rata akurasi pada percobaan selama 5 kali dapat dilihat pada tabel 6.

Berdasarkan tabel 6, algoritma SVM secara akurasi lebih unggul dalam klasifikasi sentimen ulasan dibandingkan algoritma KNN. Model klasifikasi ini menggunakan rasio 80%:20% dimana 80% untuk data latih dan 20% untuk data uji dengan validasi silang sebanyak 5 kali, secara keseluruhan rata – rata 80% untuk SVM dan 71% untuk KNN. Hasil performa algoritma SVM dapat dilihat pada tabel 7.

Tabel 6. Rata – rata akurasi

<i>Testing</i>	<i>SVM</i>	<i>KNN</i>
1	80%	70%
2	82%	71%
3	80%	71%
4	79%	72%
5	79%	70%
Rerata	80%	71%

Tabel 7. Hasil performa svm

<i>Testing</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-1 Score</i>
1	80%	80%	77%	77%
2	82%	80%	79%	79%
3	80%	80%	77%	77%
4	79%	81%	75%	75%
5	79%	76%	75%	75%
Rerata	80%	79%	77%	77%

Hasil pada tabel 7, klasifikasi SVM dilakukan melalui lima pengujian berbeda, dan hasilnya dicatat untuk setiap iterasi. Rata-rata akurasi dari kelima pengujian mencapai 80%, menunjukkan tingkat ketepatan prediksi yang baik oleh model SVM. *Precision*, yang mengukur sejauh mana hasil positif yang diprediksi benar, memiliki rata-rata sebesar 79%, sementara *recall*, yang mengukur sejauh mana model mampu mengidentifikasi semua instance positif, mencapai rata-rata sebesar 77%. *F-1 Score*, yang merupakan pengukuran gabungan dari *precision* dan *recall*, juga mencapai rata-rata sebesar 77%. Hasil ini memberikan indikasi bahwa model SVM menunjukkan kinerja yang konsisten dan baik dalam mengklasifikasikan data. Hasil performa algoritma KNN dapat dilihat pada tabel 8.

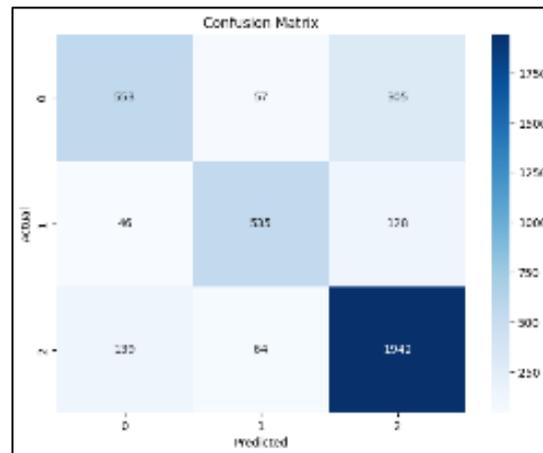
Tabel 8. Hasil performa knn

<i>Testing</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-1 Score</i>
1	70%	67%	70%	68%
2	71%	68%	72%	70%
3	71%	68%	72%	70%
4	71%	69%	72%	70%
5	70%	67%	71%	69%
Rerata	71%	68%	72%	69%

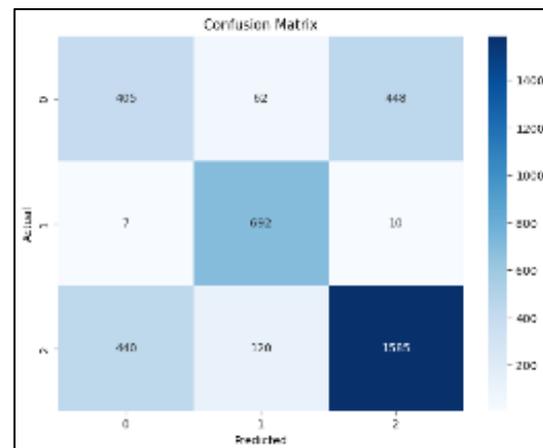
Klasifikasi KNN dilakukan melalui serangkaian pengujian yang terdiri dari lima iterasi, dan dicatat untuk setiap iterasi tersebut. Rata-rata akurasi dari kelima pengujian sebesar 71% yang disajikan pada tabel 8, menunjukkan tingkat ketepatan prediksi yang relatif baik oleh model KNN. *Precision*, yang mengukur sejauh mana hasil positif yang diprediksi benar, memiliki rata-rata sebesar 68%, sementara *recall*, yang mengukur sejauh mana model mampu mengidentifikasi semua instance positif, mencapai rata-rata sebesar 72%. *F-1 Score*, sebagai pengukuran kombinasi antara *precision* dan *recall*, juga mencapai rata-rata sebesar 69%. Meskipun terdapat variasi hasil antar iterasi, kinerja keseluruhan model KNN menunjukkan kemampuan yang relatif stabil dan cukup baik dalam mengklasifikasikan data.

Evaluasi dengan *Confusion Matrix* pada gambar 2 dan 3 menunjukkan sejumlah kata yang salah diklasifikasikan oleh algoritma SVM dan KNN. Dengan kunci *Actual* (kelas

sebenarnya), *Predicted* (prediksi dari algoritma SVM atau KNN), dan kelas 0, 1, 2 mewakili negatif, netral, dan positif.



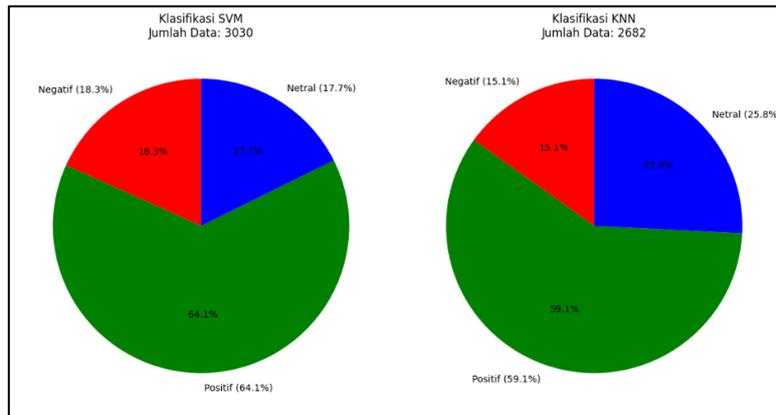
Gambar 2. *Confusion matrix svm*



Gambar 3. *Confusion matrix knn*

Pada gambar 2, terjadi True Negatives (TN) sebanyak 553 (kelas 0 yang benar diprediksi sebagai kelas 0), False Positives (FP) sebanyak 57 (kelas 1 yang salah diprediksi sebagai kelas 0) dan 128 (kelas 2 yang salah diprediksi sebagai kelas 0), serta False Negatives (FN) sebanyak 46 (kelas 0 yang salah diprediksi sebagai kelas 1) dan 64 (kelas 0 yang salah diprediksi sebagai kelas 2). Selain itu, terdapat True Positives (TP) sebanyak 535 (kelas 1 yang benar diprediksi sebagai kelas 1) dan 1942 (kelas 2 yang benar diprediksi sebagai kelas 2).

Selanjutnya pada gambar 3, ditemukan True Negatives (TN) sebanyak 405 (kelas 0 yang benar diprediksi sebagai kelas 0), serta False Positives (FP) sebanyak 62 (kelas 1 yang salah diprediksi sebagai kelas 0) dan 448 (kelas 2 yang salah diprediksi sebagai kelas 0). Selain itu, terdapat False Negatives (FN) sebanyak 7 (kelas 0 yang salah diprediksi sebagai kelas 1) dan 10 (kelas 1 yang salah diprediksi sebagai kelas 2). Di sisi lain, True Positives (TP) ditemukan sebanyak 692 (kelas 1 yang benar diprediksi sebagai kelas 1) dan 1585 (kelas 2 yang benar diprediksi sebagai kelas 2). Selanjutnya, untuk memudahkan memahami hasil, visualisasi data diperlukan. Pie chart digunakan untuk memperlihatkan keseluruhan data ulasan yang terklasifikasi sentimen positif, negatif, dan netral. Visualisasi data hasil klasifikasi benar penelitian dapat dilihat pada gambar 4.



Gambar 4. Hasil klasifikasi benar kedua algoritma

Pembahasan

Data yang digunakan pada penelitian ini berupa dataset ulasan berjumlah 4712 ulasan yang didapat dari tahap *data scraping*. Setelah melalui pelabelan dan praproses, data dibagi menjadi data latih dan data uji pada rasio 80%:20% dengan 3769 data latih dan 943 data uji. Pengujian analisis sentimen menggunakan algoritma SVM dan KNN dengan masing – masing validasi silang 5 lipatan.

Algoritma SVM menunjukkan kinerja konsisten dengan rata-rata akurasi 80%, precision 79%, recall 77%, dan F-1 Score 77% dalam lima pengujian. Evaluasi confusion matrix menunjukkan kecenderungan model untuk memprediksi kelas 0 dengan baik (TN: 553), namun terdapat kesalahan pada kelas 1 dan 2, terutama FP dan FN. Meskipun demikian, jumlah TP untuk kelas 1 dan 2 tinggi (masing-masing 535 dan 1942), menunjukkan kemampuan SVM dalam mengklasifikasikan kelas tertentu. Validasi silang dengan dataset lebih besar dapat meningkatkan hasil klasifikasi, terutama akurasi tertinggi dibandingkan penelitian sebelumnya (Atmajaya et al., 2023; Rifaldi et al., 2023) dengan akurasi 80% dan 59%.

Algoritma KNN menunjukkan kinerja relatif stabil dengan rata-rata akurasi 71%, precision 68%, recall 72%, dan F-1 Score 69% dalam lima pengujian. Evaluasi confusion matrix menunjukkan kemampuan KNN memprediksi kelas 0 dengan baik, meskipun terdapat kesalahan pada kelas 1 dan 2, terutama FP dan FN. Jumlah TP untuk kelas 1 dan 2 tinggi (masing-masing 692 dan 1585), menunjukkan kemampuan KNN dalam mengklasifikasikan data. Dengan melakukan validasi silang dan menggunakan dataset lebih besar, model KNN dapat meningkatkan hasil klasifikasi, terutama akurasi, dibandingkan penelitian sebelumnya (Yusuf & Masripah, 2023) yang mencapai akurasi 69,23%.

Hasil pengujian dan evaluasi menunjukkan bahwa terdapat kesalahan signifikan dalam klasifikasi data berlabel netral. Penyebab utamanya adalah ketidakandalan kamus kata pada lexicon SentiWordNet, terutama dalam pelabelan kelas netral. Kalimat dengan sentimen netral cenderung memiliki variasi kata yang lebih beragam dan kurang mengandung kata-kata khas, sehingga sulit diidentifikasi (Santoso et al., 2017). Oleh karena itu, metode SentiWordNet dianggap belum sepenuhnya akurat dalam memprediksi teks dengan sentimen netral.

Adapun hasil pengujian menunjukkan bahwa SVM algoritma terbaik dalam klasifikasi analisis sentimen, dengan rerata akurasi 80% dalam lima pengujian dan hasil klasifikasi benar sebanyak 3030 data. Keunggulan SVM terletak pada kemampuannya mengatasi kelemahan KNN terhadap fitur-fitur kurang relevan yang tidak berkontribusi secara signifikan dalam membuat prediksi (Lestari et al., 2022).

SIMPULAN

Temuan kami menunjukkan bahwa penggunaan validasi silang sebanyak lima kali menghasilkan kinerja yang cukup baik menggunakan algoritma SVM dan KNN pada dataset sejumlah 4712. Model klasifikasi sentimen dengan algoritma SVM mencapai rerata akurasi

yang lebih tinggi, yaitu sebesar 80%, sementara KNN memiliki rerata akurasi sebesar 71%. Meskipun demikian, terdapat data yang masih belum terklasifikasi dengan benar pada sentiment netral, yang diyakini disebabkan oleh ketidakandalan kamus kata pada lexicon SentiWordNet dalam melakukan pelabelan kelas netral. Dengan demikian, dapat diketahui bahwa, meskipun terdapat sejumlah kekhawatiran pengguna, sentimen terhadap aplikasi ChatGPT cenderung positif dilihat dari jumlah data ulasan terklasifikasi benar positif sebanyak 64%.

REFERENSI

- Alaei, A. R., Becken, S., & Stantic, B. (2017). Sentiment analysis in tourism: Capitalizing on big data. *Journal of Travel Research*, 58(2), 175–191. <https://doi.org/10.1177/0047287517747753>
- Amrullah, A. Z., Anas, A. S., & Hidayat, M. A. J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal BITE*, 2(1), 40–44. <https://doi.org/10.30812/bite.v2i1.804>
- Atmajaya, D., Febrianti, A., & Darwis, H. (2023). Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter. *Indonesian Journal of Computer Science*, 12(4). <https://doi.org/10.33022/ijcs.v12i4.3341>
- Dowling, M., & Lucey, B. (2023). ChatGPT for (Finance) research: The Bananarama Conjecture. *Finance Research Letters*, 53, 103662. <https://doi.org/10.1016/j.frl.2023.103662>
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>
- Gravel, J., D'Amours-Gravel, M., & Osmanliu, E. (2023). Learning to fake it: Limited responses and fabricated references provided by ChatGPT for medical questions. *Mayo Clinic Proceedings: Digital Health*, 1(3), 226–234. <https://doi.org/10.1016/j.mcpdig.2023.05.004>
- Handayani, A. S., Soim, S., Agusdi, T. E., Rumiasih, & Nurdin, A., (2020). Klasifikasi Kualitas Udara Dengan Metode Support Vector Machine. *Jurnal Informatika & Rekayasa Elektronika*, 3(2), pp. 187-199.
- Hassani, H., & Silva, E. S. (2023). The role of ChatGPT in data science: how ai-assisted conversational interfaces are revolutionizing the field. *Big data and cognitive computing*, 7(2), 62. <https://doi.org/10.3390/bdcc7020062>
- Hidayatullah, H., Purwantoro, P., & Umaidah, Y. (2023). Penerapan Naive Bayes Dengan Optimasi Information Gain Dan SMOTE Untuk Analisis Sentimen Pengguna Aplikasi ChatGPT. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(3), 1546–1553. <https://doi.org/10.36040/jati.v7i3.6887>
- Krotov, V., Johnson, L., & Silva, L. (2020). Legality and ethics of web scraping. *Communications of the Association for Information Systems*, 47, 539–563. <https://doi.org/10.17705/1cais.04724>
- Leippold, M. (2023). Thus spoke GPT-3: Interviewing a large-language model on climate finance. *Finance Research Letters*, 53(May 2023), 103617. <https://doi.org/10.1016/j.frl.2022.103617>
- Lestari, S., Mupaat, M., & Erfina, A. (2022). Analisis Sentimen Masyarakat Indonesia terhadap Pemindahan Ibu Kota Negara Indonesia pada Twitter. *JUSIFO (Jurnal Sistem Informasi)*, 8(1), 13–22. <https://doi.org/10.19109/jusifo.v8i1.12116>
- Lailiyah, M., Sumpeno, S., & Purnama, I. K. E. (2017, August). Sentiment analysis of public complaints using lexical resources between Indonesian sentiment lexicon and

- Sentiwordnet. *2017 International Seminar on Intelligent Technology and Its Applications (ISITIA)*. <http://dx.doi.org/10.1109/isitia.2017.8124100>
- Liu, B., (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Prasetyo, E. (2013). *Data Mining : Konsep Dan Aplikasi Menggunakan Matlab*. Yogyakarta: CV Andi Offset.
- Rahayu, S., Mz, Y., Bororing, J. E., & Hadiyat, R. (2022). Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 98–106. <https://doi.org/10.29408/edumatic.v6i1.5433>
- Rifaldi, M., Ramadhan, Y., & Jaelani, I. (2023). Analisis Sentimen Terhadap Aplikasi Chatgpt Pada Twitter Menggunakan Algoritma Naïve Bayes. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 7(2), 802-814. doi:<http://dx.doi.org/10.30645/j-sakti.v7i2.687>
- Rofiqi, M. A., Fauzan, Abd. C., Agustin, A. P., & Saputra, A. A. (2019). Implementasi term-frequency inverse document frequency (TF-IDF) untuk mencari relevansi dokumen berdasarkan query. *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, 1(2), 58–64. <https://doi.org/10.28926/ilkomnika.v1i2.18>
- Ruslim, K. I., Adikara, P. P., & Indriati, I. (2019). Analisis Sentimen Pada Ulasan Aplikasi Mobile Banking Menggunakan Metode Support Vector Machine dan Lexicon Based Features. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(7), 6694–6702. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/5792>
- Sepri, D. (2020). Penerapan Algoritma Naïve Bayes Untuk Analisis Kepuasan Penggunaan Aplikasi Bank. *Journal of Computer System and Informatics (JoSYC)*, 2(1), 135-139.
- Shen, X., Chen, Z., Backes, M., & Zhang, Y. (2023). *In chatgpt we trust? Measuring and characterizing the reliability of chatgpt*. arXiv.Org. <https://arxiv.org/abs/2304.08979>
- Santoso, V. I., Virginia, G., & Lukito, Y. (2017). Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode Support Vector Machine. *Jurnal Transformatika*, 14(2), 72–76. <https://doi.org/10.26623/transformatika.v14i2.439>
- Yusuf, L., & Masripah, S. (2023). Sentimen Analisis ChatGPT Dengan Algoritma Naive Bayes Dan Optimasi PSO. *INTI Nusa Mandiri*, 18(1), 59–64. <https://doi.org/10.33480/inti.v18i1.4230>