

SENTIMENT ANALYSIS OF CHATGPT APP USER REVIEWS USING SVM AND CNN METHODS

Nurul Widaad^{*1}, Dina Anggraini²

^{1,2}Department of Information System, Technology and Engineering Faculty, Universitas Gunadarma, Indonesia
Email: ¹nurulwidaad@gmail.com, ²dina_anggraini@staff.gunadarma.ac.id

(Article received: October 17, 2024; Revision: November 21, 2024; published: December 29, 2024)

Abstract

The rapid development of Artificial Intelligence (AI) has significantly impacted various sectors, including user interactions with natural language-based applications such as ChatGPT. This study aims to analyze user sentiment towards ChatGPT amidst the emergence of alternative AI technologies like Gemini (Google Deep Mind), Claude (Anthropic AI), and LLaMA (Meta AI). ChatGPT was chosen as the research subject due to its role as a pioneer in public AI usage. The research focuses on uncovering user sentiments—positive, negative, or neutral. A total of 155,529 reviews from the Google Play Store were analyzed using Support Vector Machine (SVM) and Convolutional Neural Network (CNN) algorithms. The research process involved data collection (data scraping), preprocessing (emoji removal, case folding, punctuation removal, tokenization, stopword removal, stemming, and normalization), sentiment labeling, data splitting (80% training and 20% testing), and model evaluation using accuracy, precision, recall, and F1-score metrics. The results indicate that the SVM model achieved an accuracy of 85%, with a precision of 0.83, recall of 0.55, and an F1-score of 0.58. Meanwhile, the CNN model recorded an accuracy of 84%, with a precision of 0.68, recall of 0.59, and an F1-score of 0.62. Among the analyzed reviews, 75% expressed positive sentiment, 18.22% negative, and 6.71% neutral. The dominance of positive sentiment reaffirms ChatGPT's position as a preferred choice among users, although certain aspects need improvement to maintain its competitiveness amidst growing AI competition. This study provides valuable insights for developers to identify the strengths and weaknesses of ChatGPT based on user feedback, enabling them to optimize the application's features to create a more satisfying and relevant user experience in the future.

Keywords: Artificial Intelligence, Convolutional Neural Network, Sentiment Analysis, Support Vector Machine.

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI CHATGPT MENGGUNAKAN METODE SVM DAN CNN

Abstrak

Perkembangan pesat kecerdasan buatan (*Artificial Intelligence/AI*) telah memengaruhi berbagai sektor, termasuk interaksi pengguna dengan aplikasi berbasis bahasa alami, seperti ChatGPT. Penelitian ini bertujuan menganalisis sentimen pengguna terhadap ChatGPT di tengah munculnya alternatif teknologi AI seperti Gemini (Google Deep Mind), Claude (Anthropic AI), dan LLaMA (Meta AI). ChatGPT dipilih sebagai objek penelitian karena merupakan pionir AI untuk publik. Penelitian ini bertujuan mengungkap sentimen pengguna—positif, negatif, atau netral. Sebanyak 155.529 ulasan dari Google Play Store dianalisis menggunakan algoritma Support Vector Machine (SVM) dan Convolutional Neural Network (CNN). Proses penelitian mencakup pengumpulan data (*data scraping*), pra-pemrosesan (*emoji removal, case folding, punctuation removal, tokenisasi, stopword removal, stemming, dan normalisasi*), pelabelan sentimen, pembagian data (80% pelatihan dan 20% pengujian), serta evaluasi model dengan metrik akurasi, presisi, recall, dan *F1-score*. Hasil menunjukkan SVM mencapai akurasi 85%, dengan presisi 0,83, recall 0,55, dan *F1-score* 0,58. Sementara CNN memiliki akurasi 84%, presisi 0,68, recall 0,59, dan *F1-score* 0,62. Dari ulasan yang dianalisis, 75% termasuk sentimen positif, 18,22% negatif, dan 6,71% netral. Dominasi sentimen positif menegaskan posisi ChatGPT sebagai pilihan utama pengguna, meskipun terdapat aspek tertentu yang perlu ditingkatkan untuk menjaga daya saing di tengah persaingan AI yang semakin ketat. Penelitian ini memberikan wawasan berharga bagi pengembang dalam mengenali keunggulan dan kekurangan ChatGPT berdasarkan opini pengguna, sehingga dapat mengoptimalkan fitur aplikasi untuk menciptakan pengalaman yang lebih memuaskan dan relevan di masa mendatang.

Kata kunci: Analisis Sentimen, Artificial Intelligence, Convolutional Neural Network, Support Vector Machine.

1. PENDAHULUAN

Perkembangan pesat teknologi Kecerdasan Buatan (AI) telah memberikan dampak besar pada berbagai sektor [1], termasuk interaksi manusia dengan aplikasi berbasis bahasa alami seperti ChatGPT. Sebagai pionir AI untuk publik, ChatGPT menjadi objek penelitian yang relevan untuk memahami persepsi pengguna di tengah persaingan dengan teknologi AI alternatif, seperti Gemini, Claude, dan LLaMA. Penggunaannya yang luas dan cepat, mencapai 100 juta pengguna dalam dua bulan, menegaskan peran penting teknologi ini dalam kehidupan sehari-hari [2]. Namun, teknologi ini juga memiliki keterbatasan, seperti kesalahan faktual dan ketidakmampuan untuk menangani perintah ambigu. Meski demikian, banyak yang melihat ChatGPT sebagai langkah maju dalam evolusi AI [3].

Urgensi penelitian ini terletak pada adopsi cepat AI dan pentingnya memahami sentimen pengguna—positif, netral, atau negatif. Analisis sentimen penting dalam aplikasi AI untuk memahami emosi pengguna, meningkatkan pengalaman, dan menyesuaikan layanan [4]. Wawasan ini dapat membantu meningkatkan fungsionalitas, keadilan, dan etika aplikasi AI seperti ChatGPT agar lebih sesuai dengan harapan pengguna dan norma sosial.

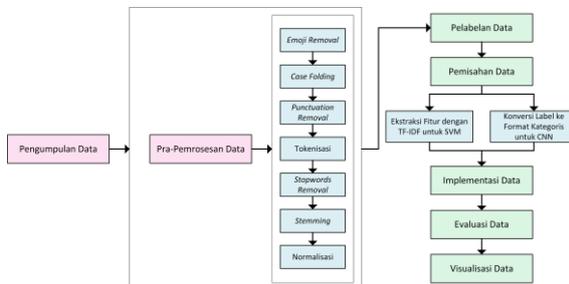
Penelitian terkait ChatGPT dengan kombinasi algoritma SVM dan CNN masih terbatas. Sejalan dengan penelitian sebelumnya [5], SVM dipilih karena kemampuannya dalam mengklasifikasikan sentimen negatif dengan akurasi yang lebih tinggi dibandingkan model lainnya. Sementara itu, CNN dipilih karena ulasan yang diteliti berbahasa Inggris dan umumnya terdiri dari kalimat atau frasa yang pendek.

Berdasarkan penelitian sebelumnya [6], CNN lebih unggul dalam mengenali pola lokal dalam teks singkat, menjadikannya lebih efisien untuk menganalisis ulasan Google Play Store. Dibandingkan model seperti IndoBERT yang cocok untuk dokumen panjang, CNN lebih cepat dalam pelatihan dan mampu menghindari *overfitting*, sambil tetap mempertahankan informasi penting untuk klasifikasi sentimen yang akurat. Penelitian ini menggunakan 155.529 ulasan dari Google Play Store, dengan evaluasi model berdasarkan akurasi, presisi, *recall*, dan *F1-score* untuk menilai kemampuan klasifikasinya secara komprehensif.

Penelitian ini diharapkan memberikan kontribusi bagi pengembangan ChatGPT melalui evaluasi sentimen pengguna, serta memberikan wawasan tentang kendala dan peluang yang dihadapi. Hasilnya dapat membantu pengembang, pelajar, dan pekerja mempersiapkan diri menghadapi dinamika teknologi AI, dengan fokus pada perbandingan performa SVM dan CNN dalam analisis sentimen.

2. METODE PENELITIAN

Penelitian ini menggunakan teknik NLP yang diimplementasikan melalui dua algoritma pembelajaran mesin SVM dan CNN.



Gambar 2.1. Hasil Pra-pemrosesan Data

Metodologi ini dirancang untuk menganalisis sentimen pengguna terhadap aplikasi ChatGPT melalui proses yang terdiri dari beberapa tahapan seperti terlihat pada Gambar 2.1, meliputi pengumpulan data, pra-pemrosesan data, pelabelan data, pemisahan data, ekstraksi fitur, pelatihan model, evaluasi kinerja, serta visualisasi hasil.

2.1. Alur Metodologi

Metodologi penelitian diawali dengan pengumpulan data dari *user review* aplikasi ChatGPT. Setelah data diperoleh, dilakukan tahapan pra-pemrosesan untuk memastikan bahwa data siap dianalisis. Proses ini mencakup langkah-langkah seperti *emoji removal*, *case folding*, *punctuation removal*, *stopword removal*, *stemming*, serta normalisasi teks.

Setelah data melalui tahap pemrosesan, dilakukan pelabelan sentimen di mana ulasan pengguna dikategorikan ke dalam sentimen positif, negatif, atau netral. Kemudian, data tersebut dibagi menjadi dua kelompok: set pelatihan (*training set*) dan set uji (*testing set*) dengan rasio 80:20.

Pada tahap selanjutnya, algoritma SVM diimplementasikan dengan teknik TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk melakukan ekstraksi teks, sedangkan CNN mengonversi label ke dalam format kategorikal dan mengolah teks menggunakan tokenisasi serta *padding*. Kedua model kemudian dilatih menggunakan set data yang telah diproses.

Tahap akhir penelitian ini melibatkan evaluasi performa kedua model menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil klasifikasi sentimen kemudian divisualisasikan melalui *Confusion Matrix*, Grafik Distribusi, dan *WordCloud* untuk memberikan pemahaman yang lebih mendalam mengenai distribusi sentimen serta kata-kata yang paling sering muncul dalam ulasan positif, negatif, dan netral.

2.2. Pengumpulan Data (*Data Collection*)

Penelitian ini memanfaatkan data sebanyak 155.529 ulasan aplikasi ChatGPT yang diperoleh dari Google Play Store. Proses pengumpulan data dilakukan dengan teknik *data scraping* menggunakan Google Colaboratory dan beberapa pustaka Python seperti *beautifulsoup* dan *selenium*. Data dikumpulkan dengan mempertimbangkan kriteria inklusi sebagai berikut:

- Ulasan harus mengandung opini atau penilaian terhadap aplikasi ChatGPT.
- Ulasan diambil dari kategori “*newest*” untuk memastikan bahwa data relevan dan terkini.

Proses *data scraping* dimulai dengan menginstal pustaka yang diperlukan, melakukan mounting ke Google Drive untuk menyimpan data, dan mengambil data ulasan dengan beberapa parameter seperti skor ulasan dan urutan ulasan. Data yang diperoleh disimpan dalam format CSV untuk keperluan analisis selanjutnya.

2.3. Pra-pemrosesan Data (*Data Pre-processing*)

Pada tahapan ini teks ulasan dirapikan atau di *cleaning* dan disiapkan untuk diolah oleh model *Machine Learning* (ML) dengan prosesnya meliputi:

1. *Emoji Removal*: Pada tahap ini, semua emoji yang terdapat dalam ulasan dihapus menggunakan pola *regular expression* (regex). Emoji dapat menyebabkan distorsi pada hasil analisis sentimen karena tidak selalu memiliki makna tekstual yang jelas bagi algoritma. Penghapusan emoji bertujuan untuk memastikan analisis lebih fokus pada teks.
2. *Case Folding*: Ditahap ini, seluruh teks yang ada diubah menjadi *lowercase* atau huruf kecil untuk memastikan konsistensi dalam analisis. *Case folding* membantu mengurangi variabilitas yang disebabkan oleh perbedaan kapitalisasi huruf dan memungkinkan model untuk mengolah data secara lebih optimal.
3. *Punctuation Removal*: Penghapusan tanda baca dilakukan agar model fokus pada kata-kata yang memiliki nilai semantik, bukan pada simbol-simbol yang tidak relevan. Proses ini membantu model untuk lebih tepat dalam menganalisis makna teks tanpa terganggu oleh tanda baca.
4. *Tokenisasi*: Setiap kalimat ulasan dipecah menjadi unit-unit kata (*token*) melalui proses tokenisasi. Hal ini memungkinkan algoritma untuk mengidentifikasi kata-kata individu dalam teks, yang nantinya digunakan sebagai input dalam model pembelajaran mesin.
5. *Stopword Removal*: *Stopword* adalah kata-kata seperti “*the*”, “*is*”, dan “*at*” yang seringkali muncul, tetapi tidak memiliki makna yang berarti terhadap teks secara keseluruhan. Dalam analisis sentimen, kata-kata ini dianggap tidak relevan karena tidak mempengaruhi interpretasi sentimen dari sebuah ulasan. Oleh sebab itu,

penghapusan *stopword* dilakukan untuk memastikan bahwa analisis hanya berfokus pada kata-kata yang lebih signifikan dan dapat membantu mengidentifikasi sentimen dengan lebih baik. Dengan menghilangkan *stopword*, proses analisis dapat menjadi lebih efisien, karena model hanya memproses kata-kata yang relevan yang memiliki dampak langsung terhadap hasil klasifikasi sentimen, sehingga mengurangi distraksi dari kata-kata yang tidak perlu dan memperkuat akurasi keseluruhan.

6. *Stemming*: Ini adalah proses pengurangan kata ke bentuk dasar. Misalnya, kata “*running*” akan diubah menjadi “*run*”. Proses ini membantu mengurangi variasi kata yang muncul, sehingga model dapat lebih fokus pada makna inti dari kata tersebut. Setelah *stemming*, teks ulasan dinormalisasi menjadi satu bentuk standar. Langkah ini memastikan bahwa semua kata telah diproses dengan cara yang seragam, yang mempermudah analisis lebih lanjut.
7. *Normalisasi*: Langkah terakhir adalah normalisasi, yang menyatukan teks yang sudah melalui proses pembersihan dan penyederhanaan menjadi bentuk yang siap untuk analisis lebih lanjut.

2.4. Pelabelan Data (*Data Labeling*)

Setelah data melewati proses pengolahan, setiap ulasan akan diberi label sentimen positif, negatif, ataupun netral. Proses pelabelan dilakukan secara semi-otomatis dengan mengombinasikan teknik pemrograman untuk mendeteksi kata kunci terkait sentimen dan pelabelan manual untuk meningkatkan akurasi data.

Untuk pelabelan otomatis, digunakan sebuah pendekatan berbasis daftar kata kunci di mana setiap kata dalam daftar dikaitkan dengan sentimen tertentu. Kata-kata kunci ini berfungsi sebagai indikator yang digunakan untuk mengidentifikasi dan menentukan sentimen sebuah ulasan berdasarkan frekuensi kemunculannya dalam teks. Namun, untuk memastikan tingkat akurasi yang lebih tinggi dan memperkaya variasi set data, sebagian data juga dilabeli secara manual oleh peneliti.

Pelabelan manual ini penting untuk menangani kasus-kasus yang tidak dapat terdeteksi dengan baik melalui teknik otomatis, terutama ketika ada konteks atau makna tersirat dalam teks yang mungkin tidak tertangkap oleh model berbasis kata kunci. Dengan demikian, kombinasi dari kedua pendekatan ini memberikan jaminan bahwa set data memiliki label sentimen yang lebih akurat dan representatif, yang pada akhirnya dapat meningkatkan kualitas hasil analisis sentimen secara keseluruhan.

2.5. Pemisahan Data (*Data Splitting*)

Setelah tahap pelabelan selesai, data lalu dipecah ke dalam 2 (dua) kelompok, yakni 80% untuk

training set data dan 20%-nya untuk *testing* set data. *Splitting* ini bertujuan untuk memastikan bahwa model tidak hanya belajar dari data pelatihan secara berlebihan, tetapi juga mampu menggeneralisasi pola yang ditemukan untuk diterapkan pada data baru. Hal ini sangat penting dalam rangka menghindari terjadinya *overfitting*, yaitu ketika model terlalu fokus pada detail spesifik dari data pelatihan sehingga kehilangan kemampuannya untuk beradaptasi dengan data baru yang tidak pernah dilihat sebelumnya. Dengan memisahkan data menjadi set pelatihan dan pengujian, model dapat diuji pada data yang tidak terlibat dalam pelatihan, yang memungkinkan kita menilai sejauh mana model mampu memprediksi atau mengelompokkan sentimen pada data yang bervariasi. Langkah ini sangat penting dalam memastikan model memiliki kemampuan generalisasi yang baik, sehingga kinerjanya tetap optimal saat digunakan pada set data yang lebih besar atau berbeda.

2.6. Ekstraksi Fitur dengan TF-IDF untuk Klasifikasi dengan SVM

Pada model SVM, fitur teks diambil menggunakan metode TF-IDF, yang memberikan *weight* pada masing-masing kata dilihat dari frekuensi kemunculannya di seluruh set data [7]. Teknik ini membantu model menitikberatkan analisis yang dilakukan pada kata-kata yang lebih relevan dan signifikan untuk klasifikasi sentimen, sehingga meningkatkan performa prediksi model dengan representasi teks yang lebih baik.

2.7. Konversi Label ke Format Kategorikal untuk Klasifikasi dengan CNN

Pada model CNN, label sentimen dikonversi ke format kategoris menggunakan *one-hot encoding*. Teknik ini mengubah variabel kategorikal menjadi variabel biner di mana setiap kategori direpresentasikan oleh nilai 1 untuk menunjukkan kehadiran kategori tersebut, dan nilai 0 untuk menunjukkan ketidakhadirannya. *One-hot encoding* tidak digunakan ketika data kategorikal tidak memiliki urutan yang inheren (nominal), sehingga menjadi metode yang sederhana untuk mengonversi data non-numerik menjadi format yang dapat digunakan oleh model ML [8].

2.8. Implementasi Model SVM dan CNN

Model SVM dilatih menggunakan set data pelatihan yang telah diekstraksi fiturnya dengan TF-IDF. Algoritma SVM menggunakan kernel linear untuk mengklasifikasikan sentimen. Proses pelatihan ini memungkinkan model untuk mempelajari pola-pola dalam data dan mengoptimalkan parameternya guna memprediksi sentimen dengan lebih akurat. Secara matematis, tujuan dari klasifikasi ini dinyatakan dengan rumus berikut:

$$(f(x) = w \cdot x + b \quad (1)$$

Di mana w adalah vektor bobot, x adalah vektor fitur *input* (teks ulasan), dan b adalah bias. SVM mengoptimalkan w dan b untuk memaksimalkan margin antara berbagai kelas sentimen. Persamaan (1) mewakili proses optimasi ini, yang bertujuan untuk menemukan *hyperplane* yang membedakan titik-titik data dari kategori sentimen yang berbeda dengan margin sebesar mungkin.

Model CNN dirancang dengan lapisan Embedding yang merepresentasikan kata-kata sebagai vektor berdimensi tetap, diikuti oleh lapisan Conv1D untuk mendeteksi fitur penting dalam teks. Pengaturan hyperparameter menjadi bagian penting untuk mengoptimalkan performa model. Berikut adalah hyperparameter utama yang digunakan dalam CNN:

1. Lapisan Embedding:

- Dimensi *Embedding*: 100, yang merepresentasikan kata dalam bentuk vektor berdimensi tetap untuk menangkap makna kata secara efektif.

2. Lapisan Konvolusi (Conv1D):

- Jumlah Filter: 128, untuk mendeteksi berbagai pola lokal dalam teks.
- Ukuran *Filter* (Kernel Size): 5, yang memungkinkan model menangkap informasi dari lima kata secara bersamaan.
- Fungsi Aktivasi: ReLU (*Rectified Linear Unit*), untuk mempercepat konvergensi model dan mengatasi masalah *vanishing gradient*.

3. Lapisan Pooling:

- Jenis *Pooling*: MaxPooling dengan ukuran pooling 2, untuk mereduksi dimensi fitur secara signifikan tanpa kehilangan informasi penting.

4. Lapisan Fully Connected:

- Dropout Rate: 0.5, untuk mengurangi risiko *overfitting* dengan menghapus neuron secara acak selama pelatihan.
- Fungsi Aktivasi: *Softmax*, untuk mengeluarkan probabilitas pada setiap kelas sentimen (positif, netral, negatif).

5. Proses Pelatihan:

- *Batch Size*: 32, untuk menyeimbangkan stabilitas pelatihan dan efisiensi komputasi.
- Jumlah *Epoch*: 10, dengan performa optimal tercapai pada *epoch* ke-4 (akurasi validasi 85.33% dan loss validasi 0.3778).
- *Optimizer*: Adam, dengan learning rate sebesar 0.001 untuk memastikan pembaruan bobot yang stabil dan cepat.
- Fungsi *Loss*: *Categorical Cross-Entropy*, digunakan karena cocok untuk tugas klasifikasi multi-kelas.

Proses pelatihan pada CNN dirancang untuk mengenali pola lokal dalam teks ulasan pendek dengan lebih efisien. Keluaran dari setiap lapisan konvolusional secara matematis dinyatakan sebagai:

$$h^{(1)} = f(W^{(1)} \cdot x^{(l-1)} + b^{(1)}) \quad (2)$$

Di mana $h^{(1)}$ adalah keluaran dari lapisan ke-1, $W^{(1)}$ mewakili bobot filter, $x^{(l-1)}$ adalah input dari lapisan sebelumnya, dan $b^{(1)}$ adalah nilai bias. Persamaan (2) menunjukkan proses konvolusi, di mana CNN menerapkan filter dan bias pada input untuk menghasilkan keluaran dari lapisan saat ini. Dengan pengaturan hyperparameter yang optimal, CNN mampu mengenali pola lokal dan menghasilkan klasifikasi sentimen yang akurat.

2.9. Evaluasi Model SVM dan CNN

Penilaian performa model dilakukan dengan memanfaatkan data uji guna mengukur efektivitas model dalam memprediksi sentimen dari ulasan yang belum pernah dianalisis sebelumnya. Berbagai metrik mulai dari akurasi, presisi, *recall*, hingga *F1-score* digunakan untuk memastikan bahwa evaluasi dilakukan secara komprehensif. Pada SVM, dilakukan evaluasi melalui laporan klasifikasi yang memetakan prediksi model terhadap label asli dari data pengujian. Pada model CNN, penilaian dilakukan dengan menghitung nilai *loss* dan akurasi dari data uji, yang mencerminkan efektivitas model dalam memprediksi sentimen dengan akurat. Metrik evaluasi didefinisikan [9] [10] sebagai berikut:

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

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

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

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Dimana TP (*True Positive*) bermakna ketika model memprediksi ulasan sebagai positif dan ulasan tersebut memang benar positif, TN (*True Negative*) adalah ketika model mengklasifikasikan ulasan sebagai negatif dan ulasan tersebut benar-benar negatif, FP (*False Positive*) terjadi ketika model mengklasifikasikan ulasan sebagai positif, padahal ulasan tersebut sebenarnya negatif, serta FN (*False Negative*) ketika model mengklasifikasikan ulasan sebagai negatif, padahal ulasan tersebut sebenarnya positif.

Persamaan (3) mendefinisikan Akurasi (*Accuracy*) sebagai rasio dari TP dan TN terhadap total sampel yang diklasifikasikan, termasuk FP dan FN. Ini mengukur akurasi keseluruhan dari model.

Persamaan (4) menyatakan presisi sebagai rasio TP terhadap jumlah TP dan FP, yang berfokus pada akurasi prediksi positif.

Persamaan (5) mendefinisikan *recall* sebagai rasio TP terhadap jumlah TP dan FN, yang menilai

seberapa efektif model dalam mendeteksi kasus positif yang sebenarnya.

Persamaan (6) menghitung *F1-score* sebagai rata-rata harmonis dari presisi dan *recall*, yang menyeimbangkan kedua metrik tersebut, terutama berguna untuk data yang tidak seimbang.

2.10. Visualisasi Hasil

Kinerja model dievaluasi dengan melalui beberapa metode, salah satunya adalah *Confusion Matrix* yang digunakan untuk menggambarkan perbandingan antara prediksi dan label sebenarnya, karena matriks ini secara khusus memvisualisasikan hasil klasifikasi dengan menunjukkan berapa banyak prediksi benar dan salah untuk setiap kelas [11].

Selain itu, *WordCloud* juga diterapkan untuk setiap kategori sentimen (positif, netral, dan negatif) untuk menampilkan visualisasi kata paling banyak dalam masing-masing kategori. Selanjutnya kata-kata disajikan dengan ukuran lebih besar, sementara warna atau intensitasnya menunjukkan frekuensi relatif dari kata-kata tersebut dibandingkan dengan yang lain [12]. Adapun akurasi dan *F1-score* juga divisualisasikan dengan grafik untuk mempermudah analisis terhadap kinerja model.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Pengumpulan Data

Dalam penelitian ini, peneliti memproses sebanyak 155.529 data yang diperoleh dari ulasan pengguna aplikasi ChatGPT di Google Play Store. Data tersebut dikumpulkan menggunakan teknik *scraping* dan kemudian disimpan dalam sebuah file di Google Drive.

Setiap ulasan yang dianalisis memuat berbagai informasi rinci, termasuk beberapa elemen seperti “reviewId”, yang merupakan kode identifikasi unik untuk setiap ulasan, “userName”, yang menunjukkan nama pengguna yang memberikan ulasan, serta “content”, yang berisi teks lengkap dari ulasan itu sendiri. Selain itu, terdapat juga atribut “score”, yaitu penilaian yang diberikan pengguna dalam bentuk angka, yang berada dalam rentang skala 1-5. Dengan rating tinggi (seperti 4 atau 5 bintang) menandakan kepuasan pengguna, sedangkan rating rendah (seperti 1 atau 2 bintang) menunjukkan ketidakpuasan mereka terhadap aplikasi [13].

3.2. Hasil Pra-pemrosesan Data

Proses pra-pemrosesan data dilakukan untuk mempersiapkan data ulasan yang akan dianalisis dalam klasifikasi sentimen, menggunakan model SVM dan CNN. Tahapan pra-pemrosesan dapat mengambil hingga 50%-80% dari keseluruhan proses klasifikasi, yang menunjukkan betapa pentingnya proses ini dalam menghasilkan model yang handal [14].

Setiap langkah dalam pra-pemrosesan ini krusial karena proses ini dapat membantu membersihkan dan menyederhanakan teks sehingga model ML dapat memprosesnya dengan lebih baik sehingga tingkat akurasi meningkat [15]. Umumnya tahap ini lebih sering diimplementasikan untuk Bahasa non-formal atau sehari-hari yang sulit dipahami dan kompleks sehingga membutuhkan ML untuk analisisnya [16].

```
Original Text: pretty good ui
Case Folding: pretty good ui
Tokenizes: ['pretty', 'good', 'ui']
Stemmed: ['pretti', 'good', 'ui']
Cleaned Content: pretti good ui
-----
Original Text: I've been using the ChatGPT Android app for a while now, and I must say, it's an impressive language model with excellent capabilities.
Case Folding: ive been using the chatgpt android app for a while now and i must say its an impressive language model with excellent ca
Tokenizes: ['ive', 'using', 'chatgpt', 'android', 'app', 'must', 'say', 'impressive', 'language', 'model', 'excellent', 'capabilities']
Stemmed: ['ive', 'use', 'chatgpt', 'android', 'app', 'must', 'say', 'impress', 'language', 'model', 'excel', 'capabl', 'ai', 'abil', 'c']
Cleaned Content: ive use chatgpt android app must say impress languag model excel capabl ai abil comprehend gener text trull remark ma
-----
Original Text: This artificial intelligence is very good. I do the assignment given by my school with this artificial intelligence but
Case Folding: this artificial intelligence is very good i do the assignment given by my school with this artificial intelligence but
Tokenizes: ['artificial', 'intelligence', 'good', 'assignment', 'given', 'school', 'artificial', 'intelligence', 'artificial', 'intell']
Stemmed: ['artifici', 'intellig', 'good', 'assign', 'given', 'school', 'artifici', 'intellig', 'artifici', 'intellig', 'problem', 'man']
Cleaned Content: artific intellig good assign given school artific intellig artific intellig problem mani time right ansser reequest
-----
Original Text: First installed...
Case Folding: first installed
Tokenizes: ['first', 'installed']
Stemmed: ['first', 'instal']
Cleaned Content: first instal
```

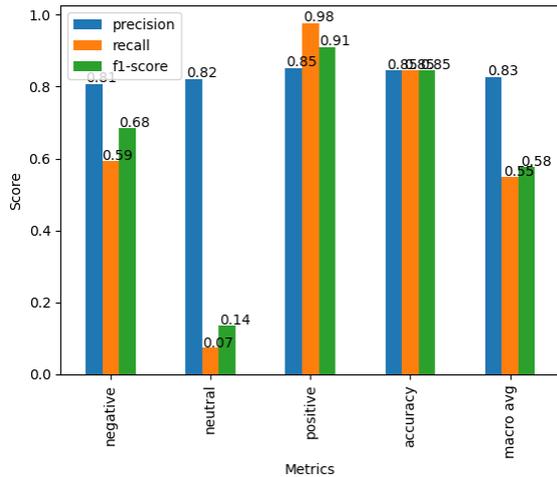
Gambar 3.2. Hasil Pra-pemrosesan Data

Hasil pra-pemrosesan ditunjukkan pada Gambar 3.1 dengan langkah-langkahnya dijabarkan berikut ini:

1. **Emoji Removal:** Emoji dihapus karena tidak relevan dalam analisis sentimen. Misalnya, teks “*First installed...*” dengan emoji senyum akan menjadi “*First installed*” setelah emoji dihapus.
2. **Case Folding:** Semua huruf diubah menjadi huruf kecil agar tidak ada perbedaan dalam kata yang sama. Contoh, teks “*What type this app*” diubah menjadi “*what type this app*”.
3. **Punctuation Removal:** Tanda baca seperti koma dan apostrof dihilangkan agar teks lebih sederhana. Misalnya, kalimat “*I’ve been using the ChatGPT Android app for a while now, and I must say, it’s an impressive language model with fantastic capabilities.*” diubah menjadi “*ive using ChatGPT android app must say impress languag model excel capabl ai*”.
4. **Tokenisasi:** Teks dipecah menjadi kata-kata agar bisa dianalisis per kata. Contohnya, “*This Artificial Intelligence is very good*” dipecah menjadi [‘artificial’, ‘intelligence’, ‘good’].
5. **Stopword Removal:** Kata-kata yang tidak penting seperti “*what*” dihapus. Misalnya, “*What type this app*” menjadi “*type app*”.
6. **Stemming:** Kata-kata diubah ke bentuk dasar. Misalnya, “*installed*” diubah menjadi “*instal*” untuk mengurangi variasi bentuk kata.
7. **Normalisasi:** Kata-kata yang sudah diproses digabung kembali menjadi teks yang lebih bersih. Contohnya, [‘first’, ‘instal’] menjadi “*first instal*” yang disebut sebagai teks ‘*cleaned content*’ siap untuk dianalisis.

3.3. Hasil Klasifikasi Sentimen dengan SVM

Model SVM dilatih dengan data yang telah diproses menggunakan teknik TF-IDF untuk mengekstrak fitur teks menjadi representasi numerik.

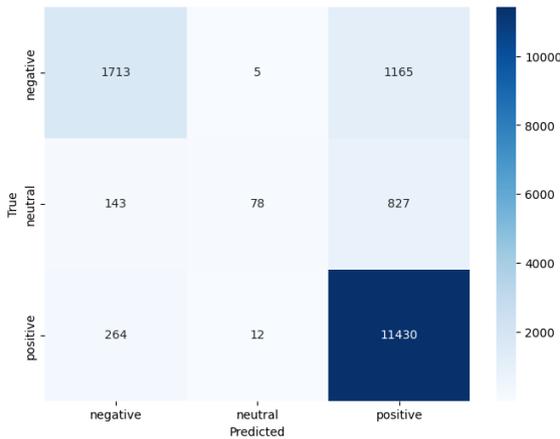


Gambar 3.3. Hasil Metrik Performa Analisis Sentimen dengan SVM

Hasil evaluasi model SVM ditampilkan pada Gambar 3.2, yang meliputi beberapa metrik evaluasi. Berdasarkan hasil tersebut, beberapa temuan utama meliputi:

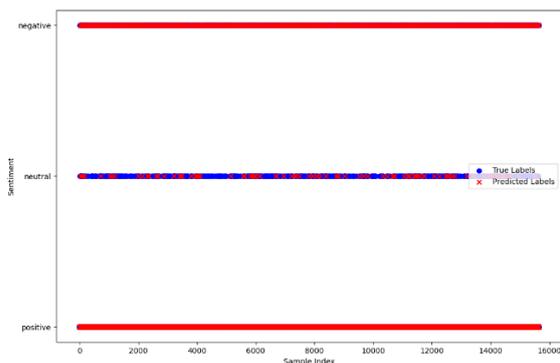
1. **Precisi:** Presisi untuk sentimen negatif mencapai 0.81, yang berarti dari semua ulasan yang diprediksi negatif, 81% memang benar-benar negatif. Untuk sentimen netral, presisi berada di angka 0.82, menunjukkan model cukup akurat dalam mengenali ulasan netral meskipun jumlah datanya lebih sedikit. Sementara itu, presisi tertinggi ada pada sentimen positif, yaitu 0.85, yang artinya 85% dari prediksi positif memang sesuai dengan ulasan yang sebenarnya.
2. **Recall:** Untuk *recall*, sentimen negatif hanya dikenali dengan benar sebesar 59%, menunjukkan bahwa banyak ulasan negatif yang tidak terdeteksi. Recall untuk sentimen netral lebih rendah lagi, hanya 0.07, yang artinya model kesulitan besar mengenali ulasan netral. Sebaliknya, *recall* untuk sentimen positif sangat tinggi, yaitu 0.98, yang menunjukkan hampir semua ulasan positif dikenali dengan benar oleh model.
3. **F1-score:** Nilai *F1-score* untuk sentimen negatif mencapai 0,68, mengindikasikan bahwa meskipun presisinya tergolong tinggi, rendahnya nilai *recall* berdampak pada penurunan kinerja keseluruhan dalam kategori ini. *F1-score* sentimen netral hanya 0.14, yang menunjukkan performa sangat rendah akibat *recall* yang buruk. Sementara itu, *F1-score* untuk sentimen positif mencapai 0.91, yang menunjukkan bahwa model sangat baik dalam mendeteksi ulasan positif.
4. **Support:** Dari segi jumlah data, terdapat 2883 ulasan negatif, 1048 ulasan netral, dan 11,706 ulasan positif. Ulasan positif mendominasi, sehingga model memiliki lebih banyak kesempatan untuk belajar dari data ini, yang berkontribusi pada tingginya *recall* dan *F1-score* untuk sentimen positif.

5. Akurasi: Akurasi keseluruhan model adalah 0.85, yang berarti 85% ulasan diklasifikasikan dengan benar. Namun, ketidakseimbangan data dapat mempengaruhi akurasi ini.
6. *Macro Average*: Presisi rata-rata adalah 0.83, tetapi *recall* rata-rata hanya 0.55, menunjukkan kesulitan model dalam mengenali ulasan netral. *F1-score* rata-rata adalah 0.58, yang menunjukkan bahwa model tidak seimbang dalam mengenali semua kategori sentimen dengan baik.



Gambar 3.4. Hasil *Confusion Matrix* Analisis Sentimen dengan SVM

Confusion matrix pada Gambar 3.3 menunjukkan kinerja model SVM dalam mengklasifikasikan sentimen negatif, netral, dan positif. Model bekerja cukup baik, tetapi masih mengalami kesulitan membedakan antara sentimen negatif dan netral. Untuk sentimen negatif, 1713 ulasan diprediksi dengan benar, namun 5 salah diklasifikasikan sebagai netral dan 1165 sebagai positif. Pada kategori netral, 143 ulasan salah diprediksi sebagai negatif dan 827 sebagai positif, hanya 78 ulasan yang benar-benar dikenali sebagai netral. Di sisi positif, model berhasil memprediksi dengan benar 11430 ulasan, tetapi 264 ulasan salah diklasifikasikan sebagai negatif dan 12 sebagai netral. Secara keseluruhan, model lebih cenderung mengklasifikasikan ulasan sebagai positif, namun masih kesulitan dalam membedakan sentimen negatif dan netral.



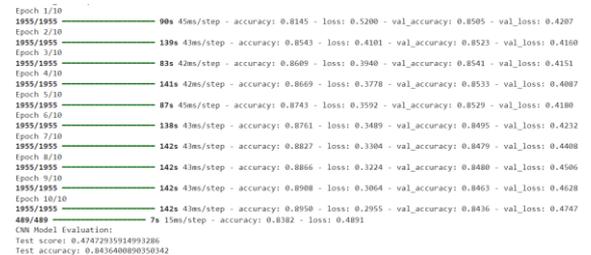
Gambar 3.5. Hasil Grafik Distribusi Analisis Sentimen dengan SVM

Gambar 3.4 menunjukkan perbandingan antara *true label* (lingkaran biru) dan *predicted label* (X merah) dalam analisis sentimen dengan model SVM. Untuk sentimen positif, hampir semua prediksi akurat dan tumpang tindih dengan label asli, menunjukkan akurasi yang sangat baik.

Pada sentimen negatif, meskipun ada beberapa kesalahan, prediksi model sebagian besar sesuai dengan label asli. Namun, untuk sentimen netral, banyak prediksi tidak sesuai dengan label asli, menandakan model kesulitan memprediksi sentimen netral secara akurat, kemungkinan karena data netral yang lebih sedikit. Secara keseluruhan, model SVM sangat baik untuk ulasan positif, cukup baik untuk negatif, tetapi kurang akurat dalam memprediksi ulasan netral.

3.4. Hasil Klasifikasi Sentimen dengan CNN

Hasil klasifikasi sentimen menggunakan model CNN yang ditampilkan pada Gambar 3.5 menunjukkan bagaimana model berkembang selama 10 *epoch*. Pada setiap *epoch*, metrik seperti akurasi (*accuracy*), kesalahan prediksi (*loss*), akurasi validasi (*validation accuracy*), dan kesalahan validasi (*validation loss*) dievaluasi untuk mengukur kinerja model.



Gambar 3.6. Hasil Klasifikasi Sentimen dengan CNN

Pada *epoch 1*, akurasi pelatihan dimulai di angka 81.45%, sedangkan akurasi validasi lebih tinggi, yaitu 85.05%. *Loss* pelatihan cukup tinggi di 0.5200, menunjukkan bahwa model baru mulai belajar. Pada *epoch 2*, akurasi pelatihan meningkat menjadi 85.43%, sementara akurasi validasi tetap stabil di 85.23%, menunjukkan bahwa model mulai menguasai pola dalam data.

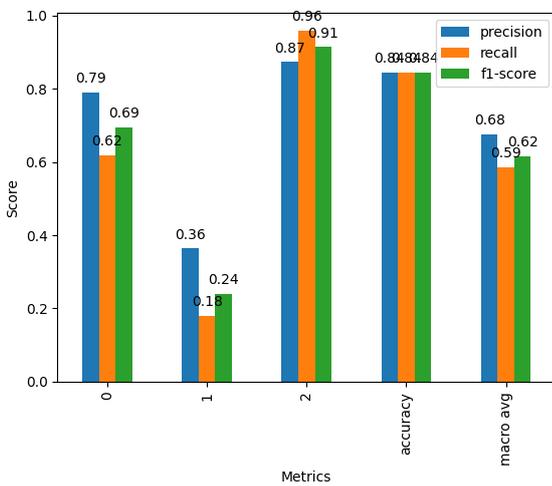
Selanjutnya, di *epoch 4*, akurasi pelatihan mencapai 86.69% dengan *loss* pelatihan 0.3778, dan akurasi validasi 85.33%. Ini dianggap sebagai *epoch* terbaik karena menunjukkan keseimbangan yang baik antara akurasi pelatihan dan validasi, dengan perbedaan minimal antara keduanya. Namun, mulai *epoch 5*, tanda-tanda *overfitting* muncul karena *loss* validasi mulai naik (0.4180), meskipun akurasi pelatihan terus meningkat.

Pada *epoch 7*, akurasi pelatihan mencapai 88.27%, namun akurasi validasi turun menjadi 84.79%, dan *loss* validasi naik hingga 0.4408, mengindikasikan *overfitting*. Hal ini semakin jelas pada *epoch 9*, di mana akurasi pelatihan meningkat menjadi 89.08%, tetapi akurasi validasi turun menjadi

84.63%, dan *loss* validasi terus meningkat hingga 0.4628.

Pada *epoch* terakhir (10), akurasi pelatihan mencapai puncaknya di 89.50%, namun akurasi validasi turun ke 84.36% dan *loss* validasi meningkat hingga 0.4747. Secara keseluruhan, *epoch* 4 adalah yang paling seimbang, sementara *epoch* 9 menunjukkan *overfitting* yang signifikan, di mana model hanya fokus pada data pelatihan tetapi kesulitan memprediksi data validasi.

Selama 10 *epoch*, akurasi pelatihan naik dari 81.45% menjadi 89.50%, sementara *loss* turun dari 0.5200 menjadi 0.2955. Akurasi validasi stabil di sekitar 84.6%-85%, namun *loss* validasi meningkat dari 0.4207 pada awal hingga 0.4747 pada akhir pelatihan. Evaluasi model dengan *test score* sebesar 0.47 dan *test accuracy* 84.36% menunjukkan bahwa model mampu memprediksi 84.36% ulasan dengan benar.



Gambar 3.6. Hasil Metrik Performa Analisis Sentimen dengan CNN

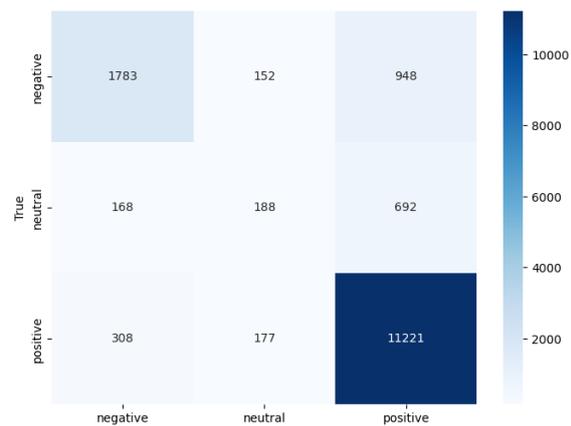
Pada model CNN data yang telah melewati proses pra-pemrosesan kemudian dioptimalkan menggunakan *categorical cross-entropy loss* dan *Adam optimizer*. Hasil evaluasi model CNN ditampilkan pada Gambar 3.6, dengan metrik evaluasi utama untuk setiap kategori sentimen: negatif (0), netral (1), dan positif (2) diuraikan sebagai berikut:

1. **Presisi:** Dalam hal presisi, model menunjukkan performa yang baik untuk sentimen negatif dengan nilai presisi sebesar 0,79, yang mengindikasikan bahwa 79% dari prediksi negatif tersebut akurat. Namun, presisi untuk sentimen netral hanya 0.36, mengindikasikan bahwa model tidak begitu akurat dalam mengenali ulasan netral. Di sisi lain, presisi tertinggi ditemukan pada sentimen positif dengan skor 0.87, yang menunjukkan bahwa prediksi model untuk ulasan positif umumnya tepat.
2. **Recall:** Dari segi *recall*, sentimen negatif memiliki *recall* 0.69, menunjukkan bahwa

model berhasil mengenali 69% dari ulasan negatif yang ada. Namun, *recall* untuk sentimen netral sangat rendah, hanya 0.18, yang menunjukkan kesulitan signifikan model dalam mendeteksi ulasan netral. Sebaliknya, *recall* untuk sentimen positif sangat tinggi, yaitu 0.96, menunjukkan bahwa hampir semua ulasan positif dikenali dengan benar.

3. **F1-score:** Disini, model menunjukkan performa yang cukup seimbang untuk sentimen negatif dengan skor 0.62, meskipun *recall* sedikit lebih rendah dari presisi. Namun, *F1-score* untuk sentimen netral sangat rendah, hanya 0.24, yang mempertegas kesulitan model dalam mendeteksi ulasan netral. Di sisi lain, *F1-score* untuk sentimen positif sangat baik, yaitu 0.91, yang menandakan bahwa model sangat efektif dalam mengklasifikasikan ulasan positif.
4. **Akurasi:** Akurasi keseluruhan model adalah 0.84, artinya model mampu memprediksi dengan benar 84% ulasan yang diuji.
5. **Macro Average:** Presisi rata-rata adalah 0.68, sedangkan *recall* rata-rata hanya 0.59, menunjukkan kesulitan dalam mengenali ulasan netral. *F1-score* rata-rata adalah 0.62, dengan kinerja kuat pada sentimen positif dan negatif, tetapi lemah untuk netral.

Secara keseluruhan, model memiliki akurasi tinggi (84%), namun ada ketidakseimbangan dalam kemampuannya mendeteksi semua kategori sentimen dengan baik. Model sangat andal dalam memprediksi sentimen positif, cukup baik untuk negatif, namun sangat lemah dalam mengenali sentimen netral. Tantangan terbesar model terletak pada peningkatan performa dalam klasifikasi sentimen netral.



Gambar 3.7. Hasil Confusion Matrix Analisis Sentimen dengan CNN

Confusion matrix yang ditampilkan pada Gambar 3.7 di atas menggambarkan kinerja model CNN dalam mengklasifikasikan sentimen negatif, netral, dan positif. Berdasarkan confusion matrix ini, terlihat bahwa model memiliki kinerja yang baik dalam beberapa aspek, namun juga menunjukkan kelemahan, terutama dalam membedakan sentimen netral dari sentimen lainnya.

Gambar *WordCloud* di atas (Gambar 3.10) memperlihatkan kata-kata yang paling sering muncul dalam ulasan positif mengenai aplikasi ChatGPT. Kata-kata seperti “good,” “use,” dan “ChatGPT” mendominasi, menunjukkan bahwa pengguna sangat puas dengan kegunaan dan kualitas aplikasi ini. Banyak pengguna merasa aplikasi ini membantu mereka, seperti terlihat dari kata “help,” dan beberapa bahkan mengucapkan terima kasih, tercermin dalam kata “thank.” Kata “best” juga sering muncul, menunjukkan bahwa banyak yang menganggap ChatGPT sebagai aplikasi terbaik. Secara keseluruhan, *WordCloud* ini menunjukkan bahwa pengguna menghargai aplikasi ChatGPT karena bermanfaat dan berkualitas tinggi.



Gambar 3.11. *WordCloud* Sentimen Negatif dengan SVM

WordCloud sentimen negatif pada Gambar 3.11 mengungkapkan berbagai keluhan utama yang disampaikan oleh pengguna terkait dengan kinerja aplikasi. Kata “app” dan “work” mendominasi, menunjukkan bahwa banyak pengguna mengeluhkan performa aplikasi yang tidak sesuai harapan. Kata “bad” dan “doesn’t work” juga sering muncul, menandakan pengalaman buruk dan masalah teknis yang dialami pengguna. Ulasan negatif juga banyak mencakup kata “answer,” yang menunjukkan ketidakpuasan terhadap jawaban yang diberikan aplikasi. Selain itu, pengguna sering menyebut kata “bug” dan meminta perbaikan (fix), mengindikasikan bahwa masalah teknis menjadi keluhan utama. Secara keseluruhan, *WordCloud* ini mencerminkan ketidakpuasan pengguna terkait kinerja, akurasi, dan kecepatan aplikasi.

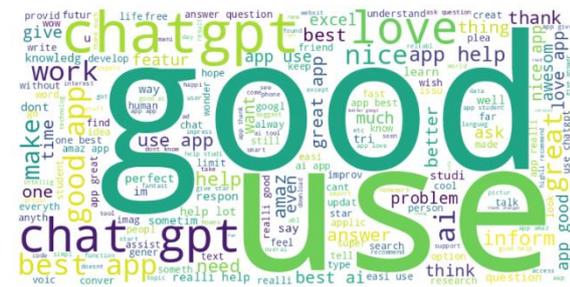


Gambar 3.12. *WordCloud* Sentimen Netral dengan SVM

Pada Gambar 3.12, *WordCloud* yang menunjukkan kata-kata yang sering muncul dalam ulasan netral tentang aplikasi ChatGPT. Meskipun sentimen bersifat netral, kata “good” masih

mendominasi, menandakan pengakuan terhadap kualitas aplikasi tanpa ekspresi yang sangat positif atau negatif. Kata “use” dan “work” sering muncul, menunjukkan bahwa aplikasi digunakan secara rutin oleh pengguna, namun tanpa pujian atau kritik yang kuat.

Pengguna juga menyebutkan “response” dan “better,” yang mengindikasikan bahwa ada harapan untuk peningkatan, terutama terkait waktu respons. Kata “okay” sering digunakan untuk menggambarkan bahwa aplikasi memadai, tetapi masih bisa diperbaiki. Secara keseluruhan, *WordCloud* ini mencerminkan ulasan yang deskriptif dan fungsional tanpa emosi yang ekstrem, baik positif maupun negatif.



Gambar 3.13. *WordCloud* Sentimen Positif dengan CNN

WordCloud positif yang dihasilkan oleh model CNN pada Gambar 3.13 menampilkan kata-kata seperti “good,” “use,” dan “work” yang mendominasi, menandakan bahwa banyak pengguna menilai aplikasi ChatGPT sangat baik dalam hal performa dan fungsionalitas. Kata “love” juga muncul, menunjukkan apresiasi emosional yang tinggi dari pengguna terhadap aplikasi. Selain itu, kata-kata seperti “help” dan “give” mengindikasikan bahwa pengguna merasa terbantu oleh aplikasi dalam berbagai situasi, terutama dalam memberikan jawaban atau solusi.

Dibandingkan dengan *WordCloud* positif dari model SVM, *WordCloud* CNN menunjukkan dominasi kata “love” dan “future,” yang mengindikasikan lebih banyak ulasan positif yang emosional dan harapan terhadap perkembangan masa depan aplikasi. Sementara itu, *WordCloud* dari model SVM lebih fokus pada kata “inform” dan “answer,” yang cenderung mengarah pada fungsi spesifik yang dihargai oleh pengguna.



Gambar 3.14. *WordCloud* Sentimen Negatif dengan CNN

Ketidakseimbangan data dalam penelitian ini sangat memengaruhi performa kedua model. Dengan dominasi ulasan positif (75%), model cenderung bias terhadap kategori ini, yang terlihat dari recall yang tinggi pada sentimen positif untuk kedua model. Ketidakseimbangan data ini mengurangi performa pada kategori netral, karena jumlah data yang terbatas membuat model sulit belajar pola yang representatif. Ketidakseimbangan juga menyebabkan CNN lebih sering salah mengklasifikasikan ulasan netral sebagai positif atau negatif, yang terlihat dari presisi rendah pada kategori ini. Adapun potensi solusi untuk mengatasi ketidakseimbangan data, yaitu:

1. **Oversampling atau Undersampling:**

- Oversampling dapat digunakan untuk meningkatkan jumlah data pada kategori netral, misalnya dengan teknik SMOTE (Synthetic Minority Oversampling Technique).
- Undersampling dapat diterapkan untuk mengurangi jumlah data pada kategori positif sehingga distribusi data menjadi lebih seimbang.

2. **Penggunaan Class Weights:**

- Memberikan bobot lebih besar pada kategori netral selama pelatihan model dapat membantu algoritma lebih fokus pada kategori ini, tanpa mengorbankan performa pada kategori lainnya.

3. **Augmentasi Data:**

- Augmentasi data teks, seperti sinonimisasi (penggantian kata dengan sinonim) atau pertukaran frasa, dapat digunakan untuk memperluas jumlah data pada kategori netral dan negatif.

4. **Model Hybrid:**

- Kombinasi pendekatan SVM dan CNN dapat dijelajahi untuk memanfaatkan keunggulan masing-masing model. Misalnya, menggunakan SVM untuk mengklasifikasikan kategori netral dan CNN untuk kategori negatif dan positif.

Secara keseluruhan, SVM lebih cocok untuk memberikan performa yang stabil di semua kategori sentimen, terutama pada sentimen netral yang sulit dikenali oleh CNN. Di sisi lain, CNN lebih unggul dalam mengenali sentimen negatif dan positif, terutama pada kategori positif, di mana CNN memiliki nilai presisi yang sedikit lebih tinggi. Oleh karena itu, SVM lebih disarankan jika dibutuhkan performa yang seimbang di seluruh kategori, terutama jika data tidak seimbang, sedangkan CNN bisa menjadi pilihan yang lebih baik untuk fokus pada klasifikasi ulasan negatif dan positif.

Temuan penelitian ini memberikan wawasan penting, khususnya bagi para pengembang, untuk lebih memahami kebutuhan pengguna dan mengoptimalkan kualitas aplikasi. Sehingga, pengembang dapat mengatasi tantangan teknis dan memastikan aplikasi berkembang sesuai harapan pengguna.

5. KESIMPULAN

Berdasarkan hasil analisis sentimen terhadap ulasan pengguna aplikasi ChatGPT di Google Play Store, dapat disimpulkan bahwa kedua model—SVM dan CNN—berhasil mengklasifikasikan ulasan ke dalam kategori sentimen positif, negatif, dan netral. Ulasan dengan sentimen positif mendominasi dataset, diikuti oleh sentimen negatif dan netral. Kedua model menunjukkan tingkat efektivitas yang berbeda dalam menangkap perbedaan sentimen.

Model SVM memiliki kinerja lebih stabil dan unggul dalam klasifikasi sentimen netral, dengan presisi sebesar 0.82, serta sedikit lebih baik dalam kategori sentimen negatif. Sebaliknya, CNN menunjukkan performa yang lebih baik dalam mengenali sentimen positif, dengan presisi 0.87, dan lebih efektif dalam mendeteksi lebih banyak ulasan negatif dengan recall sebesar 0.69. SVM lebih cocok digunakan dalam situasi di mana data tidak seimbang, sebagaimana terlihat dalam distribusi dataset penelitian ini, sedangkan CNN lebih efektif pada kategori sentimen positif yang dominan.

Penelitian ini memberikan wawasan penting bagi pengembang aplikasi ChatGPT untuk memahami area yang dinilai positif, negatif, atau membutuhkan perbaikan. Hasil analisis sentimen ini dapat digunakan sebagai dasar perencanaan pembaruan fitur atau perbaikan yang lebih sesuai dengan kebutuhan pengguna, terutama dalam menangani masalah yang sering dikeluhkan. Selain itu, penelitian ini memberikan manfaat bagi pelajar dan pekerja, dengan menunjukkan potensi ChatGPT sebagai alat yang efektif untuk mendukung berbagai tugas, seperti penulisan dan pemecahan masalah. Penelitian ini juga menjadi referensi berharga bagi peneliti lain yang ingin melanjutkan studi terkait perbandingan model NLP dalam analisis sentimen teks.

Namun, penelitian ini juga mengidentifikasi tantangan yang muncul akibat ketidakseimbangan data, di mana kategori netral memiliki jumlah data yang jauh lebih sedikit dibandingkan kategori positif dan negatif. Ketidakseimbangan ini memengaruhi kemampuan model dalam mengenali ulasan netral dengan akurasi yang memadai.

Sebagai langkah selanjutnya, direkomendasikan untuk mengeksplorasi penggunaan model NLP yang lebih canggih, seperti GPT-3 atau GPT-4, yang memiliki potensi untuk memahami konteks dan nuansa lebih kompleks, terutama dalam menangani klasifikasi sentimen netral. Selain itu, penelitian mendatang dapat mengadopsi pendekatan multidimensi dalam analisis sentimen, yang tidak hanya membedakan antara sentimen positif, negatif, dan netral, tetapi juga mencakup identifikasi emosi spesifik seperti kebahagiaan, frustrasi, atau kebingungan. Hal ini dapat diperkuat dengan integrasi metode seperti analisis topik (LDA) dan analisis emosi (*emotion detection*).

Untuk memperluas cakupan penelitian, studi mendatang juga disarankan untuk melakukan analisis lintas *platform*, seperti media sosial, forum *online*, atau situs ulasan lainnya, guna mendapatkan gambaran yang lebih komprehensif tentang persepsi pengguna. Dengan memanfaatkan data dari berbagai sumber, hasil penelitian akan menjadi lebih generalis dan valid dalam menangkap opini pengguna yang beragam.

Penelitian ini diharapkan dapat memberikan kontribusi tidak hanya bagi pengembang dalam meningkatkan kualitas aplikasi AI seperti ChatGPT, tetapi juga bagi peneliti dan praktisi NLP yang tertarik untuk memahami perbandingan model SVM dan CNN dalam analisis teks. Temuan ini juga menjadi langkah awal yang relevan dalam mengembangkan teknologi AI yang lebih responsif terhadap kebutuhan dan ekspektasi pengguna.

DAFTAR PUSTAKA

- [1] P. Kecerdasan Buatan Dalam Mendukung Pembelajaran dan B. Karyadi, "PEMANFAATAN KECERDASAN BUATAN DALAM MENDUKUNG PEMBELAJARAN MANDIRI," *Jurnal Teknologi Pendidikan*, vol. 8, no. 2, hlm. 253–258, Jul 2023, doi: 10.32832/educate.v8i02.14843.
- [2] J. Carvalko, J. R. Carvalko, dan J. Ieee, "GPT-A Paradigm Shift for the Twenty-First Century," Okt 2023, doi: 10.36227/techrxiv.23690874.v1.
- [3] B. Gordijn dan H. ten Have, "ChatGPT: evolution or revolution?," 1 Maret 2023, *Springer Science and Business Media B.V.* doi: 10.1007/s11019-023-10136-0.
- [4] P. Chakriswaran, D. R. Vincent, K. Srinivasan, V. Sharma, C. Y. Chang, dan D. G. Reina, "Emotion AI-driven sentiment analysis: A survey, future research directions, and open issues," 1 Desember 2019, *MDPI AG*. doi: 10.3390/app9245462.
- [5] M. Özel dan Ö. Çetinkaya Bozkurt, "Sentiment Analysis on GPT-4 with Comparative Models Using Twitter Data," *Acta Infologica*, vol. 8, no. 1, hlm. 23–33, Jun 2024, doi: 10.26650/acin.1418834.
- [6] L. Alzubaidi *dkk.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 53, Des 2021, doi: 10.1186/s40537-021-00444-8.
- [7] I. M. Karo Karo, J. A. Karo Karo, Y. Yudianto, H. Hariyanto, M. Falah, dan M. Ginting, "Analisis Sentimen Ulasan Aplikasi Info BMKG di Google Play Menggunakan TF-IDF dan Support Vector Machine," *Journal of Information System Research (JOSH)*, vol. 4, no. 4, hlm. 1423–1430, Jul 2023, doi: 10.47065/josh.v4i4.3943.
- [8] M. K. Dahouda dan I. Joe, "A Deep-Learned Embedding Technique for Categorical Features Encoding," *IEEE Access*, vol. 9, hlm. 114381–114391, 2021, doi: 10.1109/ACCESS.2021.3104357.
- [9] M. F. Fakhrezi, Adian Fatchur Rochim, dan Dinar Mutiara Kusomo Nugraheni, "Comparison of Sentiment Analysis Methods Based on Accuracy Value Case Study: Twitter Mentions of Academic Article," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 1, hlm. 161–167, Feb 2023, doi: 10.29207/resti.v7i1.4767.
- [10] G. M. Foody, "Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient," *PLoS One*, vol. 18, no. 10, hlm. 1–27, Okt 2023, doi: 10.1371/journal.pone.0291908.
- [11] D. Krstinić, M. Braović, L. Šerić, dan D. Božić-Štulić, "Multi-label Classifier Performance Evaluation with Confusion Matrix," Academy and Industry Research Collaboration Center (AIRCC), Jun 2020, hlm. 1–14. doi: 10.5121/csit.2020.100801.
- [12] Naruaki Ogasawara, "Visualization of Research Topics Using Word Cloud Technology, Characteristics of Research Trends in Article Titles and Keywords," *The Japanese Society of Internal Medicine (JSIM)*, Mei 2024, doi: 10.51094/jxiv.722.
- [13] F. Setya Ananto dan F. N. Hasan, "Implementasi Algoritma Naïve Bayes Terhadap Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store," *Jurnal ICT: Information Communication & Technology*, vol. 23, no. 1, hlm. 75–80, Jul 2023, [Daring]. Tersedia pada: <https://ejournal.ikmi.ac.id/index.php/jict-ikmi>
- [14] K. Maharana, S. Mondal, dan B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, hlm. 91–99, Jun 2022, doi: 10.1016/j.gltp.2022.04.020.
- [15] M. A. Rosid, A. S. Fitriani, I. R. I. Astutik, N. I. Mulloh, dan H. A. Gozali, "Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi," dalam *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Jul 2020. doi: 10.1088/1757-899X/874/1/012017.
- [16] Rianto, A. B. Mutiara, E. P. Wibowo, dan P.

I. Santosa, "Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation," *J Big Data*, vol. 8, no. 26, hlm. 1–16, Des 2021, doi: 10.1186/s40537-021-00413-1.