

ANALISIS SENTIMEN PADA PENGGUNA APLIKASI X TERHADAP PEMILIHAN UMUM PRESIDEN 2024 MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK (CNN)

Nurfauziyah, Ramaditia Dwiyanaputra, Santi Ika Murpratiwi, Arik Aranta

Teknik Informatika, Universitas Mataram
Jalan Majapahit 62, Mataram, Lombok NTB, Indonesia
nurfauziyahhs@gmail.com

ABSTRAK

Pemilihan Umum Presiden 2024 menjadi perhatian besar masyarakat Indonesia, terutama di media sosial seperti Aplikasi X. Media sosial menyediakan data yang kaya akan opini publik, namun analisis sentimen terhadap teks tidak terstruktur menghadapi tantangan, seperti pola bahasa informal dan campuran bahasa. Penelitian ini bertujuan menganalisis sentimen publik terkait Pemilu 2024 menggunakan *Convolutional Neural Network* (CNN). Dataset terdiri dari 10.250 tweet yang dikumpulkan melalui Twitter API selama Juni–September 2024. Data diproses melalui pembersihan, tokenisasi, stemming, pelabelan sentimen menggunakan VADER, dan pelatihan model CNN. Penelitian ini mencakup enam percobaan, termasuk penghapusan stopwords, penggunaan bobot TF-IDF, dan modifikasi arsitektur CNN. Hasil terbaik diperoleh dari kombinasi CNN dan TF-IDF dengan akurasi 84%, precision 83%, recall 83%, dan F1-score 83%. Distribusi sentimen menunjukkan 45,42% positif, 37,53% negatif, dan 17,04% netral. Penelitian ini mengonfirmasi efektivitas CNN dengan TF-IDF dalam menganalisis sentimen dari teks kompleks di media sosial.

Kata kunci : Analisis Sentimen, CNN, TF-IDF, Media Sosial, Pemilu Presiden 2024.

1. PENDAHULUAN

Seiring dengan perkembangan teknologi digital dan meningkatnya penggunaan media sosial, platform seperti Twitter telah menjadi ruang utama bagi masyarakat untuk mengekspresikan opini publik, termasuk dalam konteks politik. Pemilihan Umum Presiden 2024, sebagai salah satu agenda politik nasional yang penting, menarik perhatian besar dari masyarakat Indonesia. Diskusi politik yang ramai di media sosial, khususnya Twitter, memberikan gambaran tentang sentimen publik terhadap pemilu tersebut. Berdasarkan data, 77% pengguna Twitter di Indonesia merupakan pengguna aktif, menjadikan Indonesia salah satu negara dengan pengguna Twitter terbesar di dunia. Dengan tingginya interaksi dan diskusi terkait pemilu di Twitter, data dari platform ini dapat dimanfaatkan untuk menganalisis sentimen publik secara otomatis.

Tantangan muncul dalam mengolah volume data besar dan teks tidak terstruktur di Twitter. Pengguna sering menggunakan bahasa tidak formal, slang, singkatan, bahkan ironi, yang membuat analisis sentimen lebih kompleks. Selain itu, perbedaan penulisan antar pengguna, seperti kata tidak baku, tanda baca berlebihan, dan penggunaan emoji, menambah kesulitan dalam memahami konteks dan emosi dalam tweet. Oleh karena itu, dibutuhkan metode yang efektif untuk mengolah dan memahami data teks yang rumit tersebut.

Salah satu metode yang efektif untuk mengatasi tantangan ini adalah *Convolutional Neural Network* (CNN). Awalnya dikembangkan untuk pemrosesan gambar, CNN kini unggul dalam pengolahan teks dengan mengekstraksi fitur penting melalui lapisan

konvolusi. CNN mengubah teks menjadi representasi numerik dan menangkap pola berulang, seperti kombinasi kata atau frasa, untuk mengklasifikasikan sentimen. Kemampuan CNN dalam mengidentifikasi pola sentimen membuatnya sangat cocok untuk menganalisis data dari media sosial, seperti Twitter.

Selain itu, CNN juga memiliki keunggulan dalam hal kemampuan menangani data dalam jumlah besar. Dalam konteks analisis sentimen Pemilihan Umum Presiden 2024, terdapat ribuan hingga jutaan tweet yang dapat dianalisis dalam waktu singkat. CNN mampu melakukan proses pembelajaran dari data yang besar ini dengan lebih efisien dibandingkan metode tradisional, seperti *Support Vector Machine* (SVM) dan *Naive Bayes*, yang seringkali memerlukan praproses yang lebih kompleks.

Secara keseluruhan, penelitian ini memberikan kontribusi penting dalam memahami sentimen publik terhadap Pemilihan Umum Presiden 2024 melalui analisis data media sosial, serta menunjukkan potensi CNN dalam klasifikasi sentimen teks dengan akurasi yang tinggi. Peningkatan lebih lanjut dapat dilakukan dengan memperbesar dataset dan mempertimbangkan konteks temporal dari tweet yang dianalisis.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terkait

Beberapa penelitian sebelumnya telah menunjukkan hasil yang positif dengan menggunakan CNN dalam analisis sentimen media sosial. Putri et al. (2022) menggunakan *Convolutional Neural Network* (CNN) untuk analisis sentimen opini di Twitter tentang lingkungan hidup Kota Depok, dengan akurasi 86%. Program dibuat dengan Python dan Flask untuk

klasifikasi sentimen positif, negatif, atau netral [2]. Penelitian lainnya oleh Adit et al. (2023) menggunakan metode *Naive Bayes* dan *Convolutional Neural Network* (CNN). Hasilnya menunjukkan bahwa *Naive Bayes* memiliki akurasi sebesar 0,69, sedikit lebih tinggi dibandingkan CNN dengan 0,68, dan kedua metode dianggap cukup efektif untuk analisis sentimen dalam konteks ini [3].

Selanjutnya, Aziz et al. (2022) melakukan penelitian menganalisis sentimen Twitter terkait isu penundaan pemilu, menunjukkan sentimen positif (60.442) dominan, diikuti negatif (58.873) dan netral (32.223). Sentimen positif lebih mengarah pada pembelaan pemerintah [4]. Penelitian lainnya, oleh Yutika et al. (2021) menganalisis sentimen ulasan kecantikan di Female Daily dengan TF-IDF dan Complement *Naive Bayes*, mencapai F1-Score 62,81% melalui penerjemahan dan preprocessing teks [9].

Berdasarkan hasil penelitian-penelitian terdahulu ini, CNN terbukti sebagai metode yang efektif dalam analisis sentimen, terutama dalam mengolah data teks yang tidak terstruktur. Metode ini tidak hanya unggul dalam mengidentifikasi pola sentimen, tetapi juga dalam efisiensi pengolahan data dalam skala besar. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan metode tersebut dengan fokus pada Pemilihan Umum Presiden Indonesia 2024. Penelitian ini menggunakan data dari Twitter sebagai sumber utama, dengan tujuan untuk mengukur sentimen publik selama masa pemilu serta mengevaluasi performa CNN dalam mengklasifikasikan sentimen positif, netral, dan negatif dari tweet yang berkaitan dengan pemilu.

2.2. Analisis Sentimen

Analisis sentimen merupakan bidang studi yang menganalisis opini seseorang, merupakan proses otomatis untuk mengidentifikasi dan mengambil informasi terkait sentimen dalam teks, yang bertujuan untuk menentukan kecenderungan opini (positif, netral atau negatif) terhadap suatu objek atau topik [10].

2.3. Aplikasi X

Twitter adalah salah satu platform media sosial yang didirikan oleh Jack Dorsey, yang digunakan untuk mengirim pesan yang dikenal sebagai kicauan atau tweet. Di Indonesia, sekitar 59% pengguna mengakses Twitter, menjadikannya peringkat kelima media sosial yang paling sering digunakan pada tahun 2020. Berdasarkan data tersebut, Twitter memiliki pengaruh yang signifikan di kalangan pengguna di Indonesia [11].

2.4. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) merupakan pengembangan dari *Multilayer Perceptron* (MLP) yang dirancang khusus untuk mengolah data dua dimensi. Sebagai bagian dari kategori *Deep Neural Network*, CNN memiliki jaringan yang sangat mendalam. Teknologi ini biasanya diterapkan dalam

tugas klasifikasi gambar, video, atau data dua dimensi lainnya. Selain itu, CNN juga dapat digunakan untuk mengklasifikasikan data teks [4].

2.5. Lexicon-based Sentiment Analysis - VADER

Lexicon-based Sentiment Analysis dari VADER (*Valence Aware Dictionary and sEntiment Reasoner*). VADER adalah metode yang banyak digunakan dalam analisis sentimen karena kemampuannya dalam mengidentifikasi sentimen dari teks, terutama teks pendek yang biasanya terdapat di media sosial. VADER menggunakan kamus kata sentimen yang dirancang khusus untuk menangkap konteks sentimen dalam kalimat, termasuk intensitas dan pola umum dalam bahasa informal [7].

2.6. Evaluasi Model

Dalam analisis sentimen, evaluasi model merupakan langkah penting untuk mengukur performa algoritma yang digunakan. Beberapa metrik evaluasi yang umum digunakan meliputi [2]:

- Accuracy: Mengukur persentase prediksi yang benar dari keseluruhan data.

Didefinisikan sebagai:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- TP: True Positive
- TN: True Negative
- FP: False Positive
- FN: False Negative

- Precision: Menunjukkan proporsi prediksi positif yang benar dari semua prediksi positif:

$$Precision = \frac{TP}{TP + FP}$$

- Recall: Menunjukkan kemampuan model dalam mendeteksi data positif secara benar:

$$Recall = \frac{TP}{TP + FN}$$

- F1-Score: Merupakan rata-rata harmonik antara *precision* dan *recall*, memberikan keseimbangan antara keduanya:

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

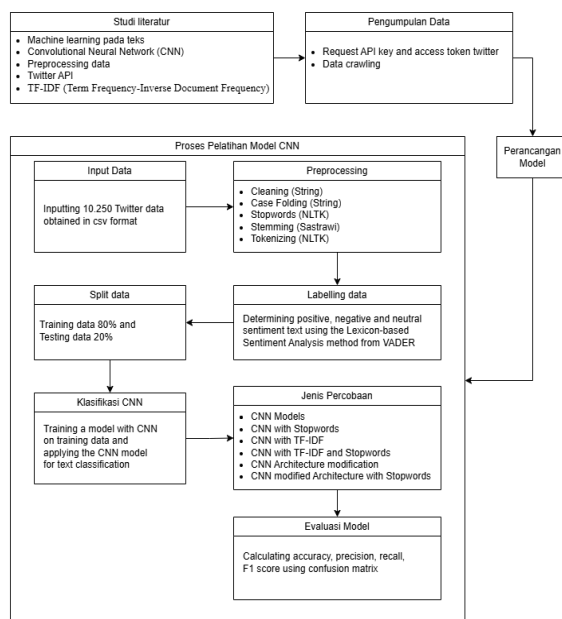
Penggunaan metrik-metrik tersebut memungkinkan peneliti untuk menilai keefektifan metode yang digunakan, seperti CNN atau metode berbasis *lexicon*, dalam menganalisis sentimen data teks.

3. METODE PENELITIAN

Penelitian mengenai analisis sentimen melalui media sosial telah berkembang pesat seiring dengan peningkatan penggunaan platform-platform ini sebagai sarana diskusi politik. Twitter, khususnya, sering digunakan sebagai sumber data untuk memahami opini publik dalam konteks politik, termasuk Pemilihan Umum Presiden Indonesia 2024. Pendekatan *quantitative research* dalam analisis sentimen menawarkan cara untuk mengkuantifikasi

sentimen positif, negatif, dan netral terhadap topik politik tertentu melalui pengolahan data dalam skala besar. Dalam konteks ini, metode *Convolutional Neural Network* (CNN) menjadi pilihan yang populer karena kemampuannya dalam menangani data teks yang kompleks.

CNN adalah salah satu jenis jaringan saraf tiruan yang banyak digunakan dalam analisis teks. Metode ini awalnya dikembangkan untuk pengolahan citra, tetapi seiring berjalannya waktu, CNN telah diadaptasi untuk berbagai tugas *Natural Language Processing* (NLP). CNN unggul dalam mengekstraksi fitur penting dari data teks dengan menggunakan lapisan konvolusi untuk menemukan pola-pola yang berguna dalam klasifikasi sentimen.



Gambar 1. Prosedur Penelitian

Beberapa tahapan dalam penelitian yaitu sebagai berikut:

3.1. Crawling Data

Dalam analisis sentimen politik, pengumpulan data dari Twitter menggunakan Twitter API adalah langkah yang umum dilakukan. Dalam konteks Pemilihan Umum Presiden 2024, tweet-tweet yang mengandung kata kunci seperti “pilpres 2024”, “calon presiden 2024”, “pemilu presiden 2024”, “presiden tahun 2024” dan “paslon presiden 2024” dikumpulkan. Periode pengumpulan data, yaitu dari Juni hingga September 2024, memastikan bahwa data yang dikumpulkan merupakan refleksi dari perkembangan opini publik yang terjadi dalam periode penting menjelang pemilihan. Pada studi sebelumnya, pengumpulan data dari Twitter terbukti dapat memberikan wawasan tentang preferensi politik dan persepsi masyarakat terhadap isu-isu yang berkembang [5].

3.2. Preprocessing Data

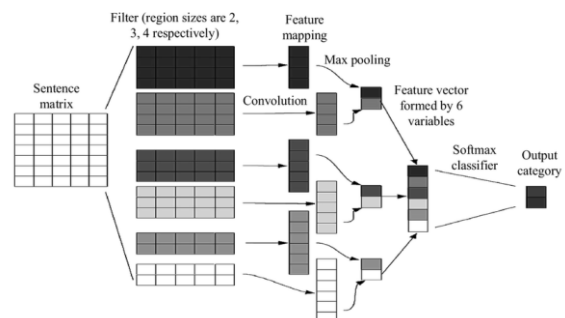
Tahap preprocessing merupakan langkah penting untuk mempersiapkan data agar dapat diproses lebih lanjut oleh model CNN. Tahapan preprocessing yang dilakukan dalam penelitian ini meliputi cleaning, case folding, stopwords, stemming, dan tokenizing. Hasil preprocessing ini kemudian diubah menjadi vektor numerik yang dapat diproses oleh model CNN.

3.3. Pelabelan Data

Proses pelabelan data adalah salah satu tahapan yang krusial dalam analisis sentimen. Pada penelitian ini, proses pelabelan data dilakukan dengan menggunakan metode *Lexicon-based Sentiment Analysis* dari VADER (Valence Aware Dictionary and sEntiment Reasoner). VADER adalah metode yang banyak digunakan dalam analisis sentimen karena kemampuannya dalam mengidentifikasi sentimen dari teks, terutama teks pendek yang biasanya terdapat di media sosial. VADER menggunakan kamus kata sentimen yang dirancang khusus untuk menangkap konteks sentimen dalam kalimat, termasuk intensitas dan pola umum dalam bahasa informal [8].

Metode ini dipilih karena VADER dapat dengan akurat menangani teks yang tidak terstruktur seperti tweet, yang sering kali mengandung elemen gaya bahasa khas seperti ironi, emoji, dan tanda baca berlebihan. Selain itu, karena analisis VADER tidak memerlukan pelatihan data, proses ini memungkinkan pelabelan data dilakukan lebih cepat dan efisien. Dengan demikian, penggunaan VADER dalam penelitian ini memungkinkan deteksi sentimen secara otomatis pada dataset Twitter dalam tiga kategori utama: positif, netral, dan negatif [8].

3.4. Pembangunan Model CNN



Gambar 2. Ilustrasi dari CNN Model [6]

Model CNN yang digunakan dalam penelitian ini terdiri dari beberapa lapisan utama:

- Embedding layer
- Convolutional layer
- Pooling layer
- Fully connected layer
- Output layer

Penelitian terdahulu menunjukkan bahwa CNN dapat bekerja dengan sangat baik dalam tugas-tugas klasifikasi teks, terutama ketika berhadapan dengan data dalam jumlah besar dan kompleksitas tinggi

seperti tweet [3]. Hal ini menjadikan CNN pilihan utama dalam penelitian analisis sentimen pada Pemilu Presiden Indonesia 2024.

3.5. Pelatihan dan Pengujian Model

Model CNN dilatih menggunakan data pelatihan yang dipersiapkan sebelumnya. Algoritma optimasi Adam digunakan untuk mempercepat proses pelatihan dengan menyesuaikan bobot model secara dinamis berdasarkan gradien dari fungsi kerugian *categorical cross-entropy*. Setelah proses pelatihan selesai, model kemudian diuji menggunakan data pengujian untuk menilai kemampuannya dalam melakukan prediksi.

3.6. Evaluasi Model

Untuk mengevaluasi kinerja model, beberapa metrik seperti accuracy, precision, recall, dan F1-score digunakan. Metrik-metrik ini memberikan gambaran yang lebih menyeluruh tentang seberapa baik model dalam mengklasifikasikan tweet ke dalam tiga kategori sentimen. Selain itu, *confusion matrix* digunakan untuk memberikan visualisasi yang lebih rinci tentang bagaimana model menangani klasifikasi yang salah atau benar.

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan Twitter API untuk mengumpulkan data terkait Pemilihan Umum Presiden Indonesia 2024, dengan kata kunci seperti “pemilihan presiden 2024”, “calon presiden 2024”, “pemilu presiden 2024”, “pemilu 2024”, “presiden tahun 2024” dan “paslon presiden 2024”. Proses pengumpulan data dilakukan dari Juni hingga September 2024. Pada rentang waktu ini, penggunaan media sosial Twitter mengalami peningkatan yang signifikan, terutama ketika berbagai debat dan kampanye berlangsung. Selain itu, peliputan media dan keterlibatan masyarakat dalam percakapan politik di platform ini juga mempengaruhi jumlah tweet yang relevan dengan topik pemilu.

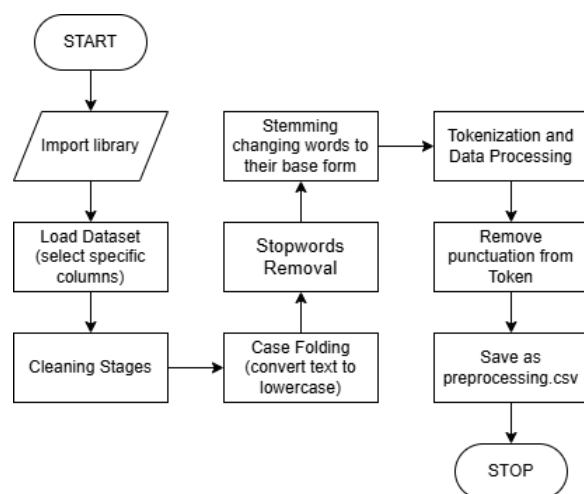
Dalam penelitian ini, jumlah total tweet yang dikumpulkan mencapai 10.250 tweet. Dataset ini mencerminkan keterlibatan publik dalam diskusi terkait Pemilihan Umum Presiden 2024, dengan variasi konten yang meliputi komentar tentang kandidat, proses pemilu, dan peristiwa politik yang relevan. Salah satu tantangan utama dalam pengumpulan data adalah memastikan bahwa tweet yang diambil benar-benar relevan dengan topik penelitian. Karena itu, seleksi kata kunci memainkan peran penting dalam memfilter tweet yang sesuai dengan isu pemilu dan calon presiden. Misalnya, kata kunci umum seperti “Pilpres 2024” dapat menangkap diskusi yang lebih luas, sementara kata kunci yang lebih spesifik seperti nama calon presiden dapat mempersempit cakupan percakapan.

Proses pengumpulan data juga mempertimbangkan aspek temporal dari tweet, yaitu bagaimana sentimen publik berubah dari waktu ke waktu. Pada bulan Juni, percakapan cenderung lebih

umum dan lebih fokus pada prediksi calon, sementara pada bulan September, ketika kampanye dan debat intensif berlangsung, percakapan berubah menjadi lebih terarah dan emosional. Variasi temporal ini menjadi penting untuk dianalisis lebih lanjut dalam konteks dinamika politik dan bagaimana sikap publik terhadap pemilu berkembang.

4.1. Preprocessing Data

Pada tahap preprocessing, data yang diperoleh dari Twitter dibersihkan dan diubah menjadi format yang dapat diproses oleh model CNN. Preprocessing merupakan langkah krusial dalam analisis sentimen karena data teks dari media sosial sering kali mengandung noise, seperti karakter-karakter khusus, emoji, atau kata-kata yang tidak relevan.



Gambar 3. Tahapan Preprocessing

Beberapa tahapan preprocessing yaitu sebagai berikut:

a. Proses Cleaning:

Pada penelitian ini pembersihan data meliputi pembersihan pola, tanda baca, nomor dan simbol.

Tabel 1. Data Hasil Proses Cleaning

Data Asli	Data Hasil Cleaning
Adapun KIM Plus mereka adalah perpanjangan dari koalisi <u>Prabowo-Gibran</u> dari Pemilu Presiden <u>2024</u> yang kini didukung oleh <u>13</u> partai politik.	Adapun KIM Plus mereka adalah perpanjangan dari koalisi Prabowo Gibran dari Pemilu Presiden yang kini didukung oleh partai politik

b. Proses Case Folding: Proses Case Folding dilakukan dengan mengubah semua huruf pada data menjadi huruf kecil. Hal ini dilakukan menggunakan metode lower dari library string, yang bekerja dengan mengubah setiap huruf kapital atau huruf besar dalam string menjadi huruf kecil.

Tabel 2. Data Hasil Proses Case Folding

Data Hasil Cleaning	Data Hasil Case Folding
Adapun KIM Plus mereka adalah perpanjangan dari koalisi Prabowo Gibran dari Pemilu Presiden yang kini didukung oleh partai politik	adapun kim plus mereka adalah perpanjangan dari koalisi prabowo gibran dari pemilu presiden yang kini didukung oleh partai politik

- c. Penghapusan stopwords: Stopwords adalah kata-kata umum yang tidak memberikan kontribusi signifikan terhadap analisis sentimen, seperti 'ada', 'adalah', 'adanya', 'adapun', 'agak', 'agakny', 'agar', 'akan', dan lainnya. Kata-kata ini dihapus untuk mengurangi noise dalam data dan fokus pada kata-kata yang lebih bermakna. Penghapusan stopwords penting untuk meningkatkan kinerja model karena kata-kata ini biasanya tidak membawa informasi penting terkait sentimen.

Tabel 3. Data Hasil Proses Stopwords

Data Hasil Stopwords	Data Hasil Stopwords
<u>adapun</u> kim plus <u>mereka</u> <u>adalah</u> perpanjangan <u>dari</u> koalisi prabowo gibran <u>dari</u> pemilu presiden <u>yang</u> <u>kini</u> didukung <u>oleh</u> partai politik	kim plus perpanjangan koalisi prabowo gibran pemilu presiden didukung partai politik

- d. Stemming: Stemming adalah proses mengubah setiap kata ke bentuk dasarnya. Misalnya, kata "mendukung" akan diubah menjadi "dukung", sehingga variasi kata yang sama dapat diidentifikasi oleh model sebagai entitas yang sama. Ini penting untuk mengurangi redundansi dalam data dan memastikan bahwa variasi morfologis dari kata yang sama tidak dianggap sebagai entitas yang berbeda oleh model.

Tabel 4. Data Hasil Proses Stemming

Data Hasil Stemming	Data Hasil Stemming
kim plus <u>perpanjangan</u> koalisi prabowo gibran pemilu presiden <u>didukung</u> partai politik	kim plus panjang koalisi prabowo gibran pemilu presiden dukung partai politik

- e. Proses Tokenisasi: Setiap tweet dipecah menjadi token-token kecil berupa kata atau simbol yang dapat diproses oleh mesin. Tokenisasi penting dalam pemrosesan bahasa alami karena model CNN bekerja dengan input numerik yang terstruktur. Dalam konteks ini, setiap kata dalam tweet diubah menjadi token yang mewakili kata tersebut dalam bentuk vektor numerik.

Tabel 5. Data Hasil Proses Tokenisasi

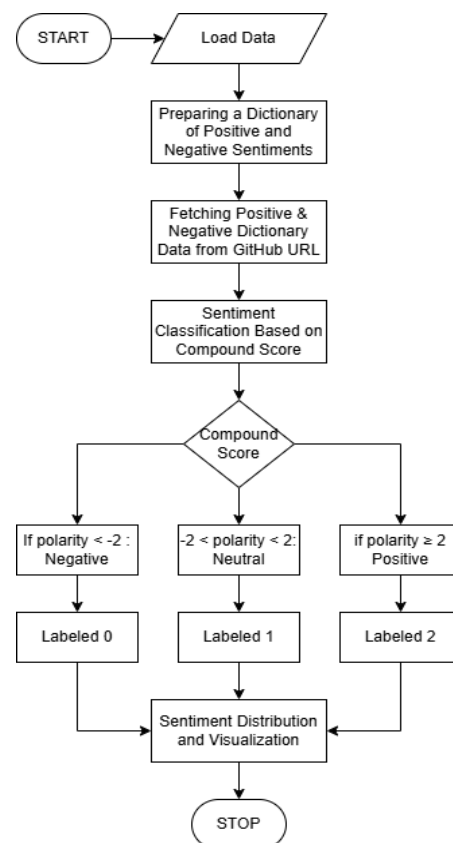
Data Hasil Stemming	Data Hasil Tokenisasi
kim plus panjang koalisi prabowo gibran pemilu presiden dukung partai politik	['kim', 'plus', 'panjang', 'koalisi', 'prabowo', 'gibran', 'pemilu', 'presiden', 'dukung', 'partai', 'politik']

- f. Padding: Padding dilakukan pada setiap tweet agar semua memiliki panjang yang sama, yaitu 100 token. Padding penting dalam konteks model CNN karena panjang input harus seragam agar dapat diproses secara konsisten oleh model.

Setelah preprocessing selesai, tweet yang awalnya dalam bentuk teks tidak terstruktur diubah menjadi data numerik yang siap digunakan dalam pelatihan model. Dari 10.250 tweet yang dikumpulkan, hasil preprocessing menghasilkan data yang siap latih dalam bentuk token numerik.

4.2. Labeling Data

Pada penelitian ini, proses pelabelan data dilakukan dengan menggunakan metode *Lexicon-based* yang diimplementasikan melalui *Lexicon-based Sentiment Analysis* dari VADER. Hasil dari pelabelan ini kemudian digunakan untuk menentukan sentimen dari tweet yang dikumpulkan, dan klasifikasi dilakukan secara otomatis ke dalam tiga kategori: positif, netral, dan negatif.



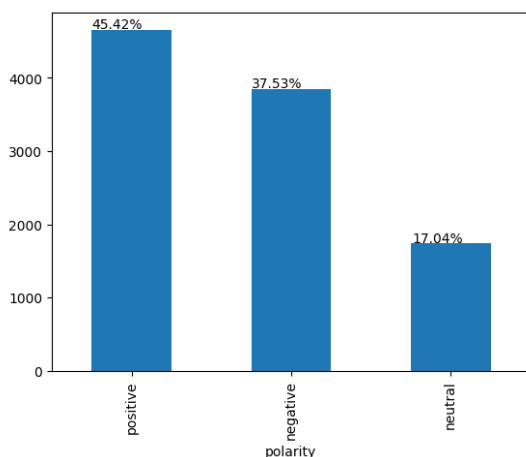
Gambar 4. Tahapan Labeling

Langkah-langkah utama serta hasil dari proses ini adalah sebagai berikut:

- a. Mempersiapkan Kamus Sentimen Positif dan Negatif: Dua kamus kosong (*lexicon_positive* dan *lexicon_negative*) disiapkan untuk menyimpan kata-kata positif dan negatif beserta skornya.

- b. Mengambil Data Kamus Positif dan Negatif dari URL github [8].
- c. Klasifikasi Sentimen Berdasarkan Skor Compound:
 - o Positif: jika skor lebih dari atau sama dengan 2.
 - o Negatif: jika skor kurang dari -2.
 - o Netral: jika skor berada di antara -2 dan 2.
 Hasil analisis menunjukkan bahwa sebagian besar tweet cenderung memiliki sentimen positif dan negatif, yang menunjukkan sentimen umum masyarakat terkait Pemilihan Umum Presiden 2024.
- d. Pemberian Label Numerik: Untuk mempermudah analisis lebih lanjut, setiap kategori sentimen dikonversi ke label numerik: “Negatif” dilabeli sebagai 0, “Netral” dilabel sebagai 1, dan “Positif” dilabeli sebagai 2. Label ini disimpan dalam kolom baru bernama label, yang digunakan untuk mempermudah pemrosesan data dalam model *Convolutional Neural Network* (CNN) pada tahap berikutnya.
- e. Distribusi dan Visualisasi Sentimen: Distribusi data dari hasil labeling ditampilkan dalam bentuk grafik batang untuk menunjukkan persentase masing-masing kategori sentimen (positif, netral, dan negatif) dalam keseluruhan dataset. Visualisasi ini memberikan gambaran tentang sikap umum masyarakat di Twitter terhadap Pemilihan Umum Presiden 2024.

Proses ini tidak hanya membantu dalam mengklasifikasikan sentimen secara efisien tetapi juga mengoptimalkan waktu dan sumber daya dalam mempersiapkan data untuk pelatihan model CNN. Penggunaan metode *Lexicon-based Sentiment Analysis* dari VADER menunjukkan hasil yang akurat dan mampu menangani data dengan volume besar dalam waktu singkat. Pendekatan ini memberikan fondasi yang kuat untuk klasifikasi sentimen selanjutnya, khususnya dalam mengidentifikasi sentimen secara lebih kompleks melalui model berbasis pembelajaran mendalam.



Gambar 5. Hasil Visualisasi Data Labeling

Distribusi sentimen dalam dataset menunjukkan 45,42% tweet positif, 37,53% negatif, dan 17,04% netral. Ketidakseimbangan ini mencerminkan bahwa mayoritas pengguna Twitter cenderung netral, yang mungkin menunjukkan sikap menunggu sebelum menentukan dukungan. Sentimen negatif sebesar 37,53% mengindikasikan adanya kritik signifikan terhadap proses pemilu atau kandidat, mencerminkan ketidakpuasan publik terhadap situasi politik.

4.3. Pembangunan Model CNN

Model *Convolutional Neural Network* (CNN) yang digunakan dalam penelitian ini dirancang untuk memproses data teks dengan beberapa lapisan penting, yaitu:

- a. Embedding Layer: Mengubah kata-kata dalam kalimat menjadi representasi vektor (angka), memungkinkan model untuk memahami hubungan antar kata berdasarkan data pelatihan.
- b. Convolutional Layer: Menerapkan filter untuk menemukan pola dalam urutan kata yang membantu dalam menentukan makna kalimat, seperti hubungan kata-kata yang mengarah pada sentimen positif, netral, atau negatif.
- c. Global Max Pooling Layer: Mengurangi dimensi data dengan memilih fitur paling relevan dan penting, mengurangi kompleksitas komputasi.
- d. Fully Connected Layer: Menggunakan informasi yang diproses untuk melakukan klasifikasi akhir dan mengidentifikasi label sentimen (positif, netral, atau negatif).

Tabel 6. Contoh Pemrosesan Data Model CNN

Layer	Proses Perubahan pada Data
Data	Berisi teks input untuk model. “Program kerja kandidat A sangat membantu masyarakat kecil, luar biasa”
Token IDs	Representasi tokenisasi teks input menjadi angka. [101, 1465, 7777, 5678, 10567, 4328, 1829, 8737, 10695, 8899, 102]
Attention Mask	Menunjukkan token yang aktif dalam input. [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Embedding Layer	Mengonversi token menjadi vektor representasi numerik. [[0.3456, 0.7612, 0.5432], [0.5678, 0.7891, 0.4321], [0.6789, 0.9876, 0.1234], ...]
Convolutional Layer	Menerapkan filter untuk menangkap fitur penting dalam teks. [[0.4567, 0.8765, 0.3456], [0.2345, 0.5678, 0.7890], [0.8765, 0.1234, 0.4321], ...]
Global Max Pooling Layer	Mengambil nilai maksimum dari setiap fitur hasil konvolusi. [0.8765, 0.9876, 0.7890]
Fully Connected Layer	Membangun keputusan akhir berdasarkan fitur penting. [0.3451, 0.6543, 0.8978]

Layer	Proses Perubahan pada Data
Logits	Output mentah dari model sebelum softmax. [2.1456, -1.0876]
Probabilities	Hasil normalisasi logits menggunakan softmax untuk menentukan probabilitas. [0.8904, 0.1096]
Predicted Label	Prediksi sentimen berdasarkan probabilitas tertinggi. "Positive"

Model CNN berhasil dibangun dan siap untuk dilatih. Model menggunakan optimizer Adam dan *loss function categorical cross-entropy*. Model ini dilatih dengan menggunakan 80% data latih, sementara 20% data digunakan sebagai data uji.

4.4. Pelatihan dan Pengujian Model

Model dilatih selama 20 epoch dan 15 epoch dengan batch size sebesar 32, menggunakan 80% data latih dan 20% data uji. Epoch mengacu pada satu kali pengolahan seluruh dataset oleh model, sementara batch size mengacu pada jumlah sampel yang diproses dalam satu kali iterasi. Penggunaan 80% data latih dan 20% data uji adalah praktik umum dalam machine learning, di mana data uji digunakan untuk menilai kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

Tabel 7. Data Hasil Akurasi Pada Enam Percobaan

Jenis Percobaan	Accuracy Epoch 20	Accuracy Epoch 15
CNN Models	82%	83%
CNN with Stopwords	77%	75%
CNN with TF-IDF	81%	84%
CNN with TF-IDF and Stopwords	78%	77%
CNN architecture modification	77%	78%
CNN architecture modification with Stopwords	76%	74%

Pada penelitian ini, dilakukan lima percobaan untuk menganalisis sentimen pengguna Twitter menggunakan metode CNN (Convolutional Neural Network) dengan variasi teknik pengolahan teks. Diantaranya sebagai berikut:

- Percobaan 1: Model CNN tanpa stopwords pada proses preprocessing, akurasi terbaik 82% (epoch ke-20) dan 83% (epoch ke-15).
- Percobaan 2: Dengan tambahan stopwords pada proses preprocessing, akurasi meningkat menjadi 77% (epoch ke-20) dan 75% (epoch ke-15).
- Percobaan 3: Menggunakan TF-IDF, akurasi 81% (epoch ke-20) dan 84% (epoch ke-15).
- Percobaan 4: Gabungan TF-IDF dan stopwords, akurasi 78% (epoch ke-20) namun sedikit turun menjadi 77% (epoch ke-15).
- Percobaan 5: Modifikasi arsitektur CNN, akurasi turun ke 77% (epoch ke-20) dan 78% (epoch ke-15).

- Percobaan 6: Modifikasi arsitektur CNN + stopwords, akurasi terendah 76% (epoch ke-20) dan 74% (epoch ke-15).

Pengujian menunjukkan bahwa kombinasi TF-IDF dan stopwords dengan arsitektur CNN standar memberikan performa terbaik, sedangkan modifikasi arsitektur tidak meningkatkan hasil secara signifikan. Pada percobaan ketiga, model CNN dengan TF-IDF mencapai akurasi tertinggi sebesar 84% pada epoch ke-15 dengan batch size 32 pada data uji.

Selain akurasi, metrik evaluasi lainnya mencakup precision, recall, dan F1-score, yang semuanya mencapai 83%. Precision mengindikasikan bahwa 83% prediksi model sesuai dengan label sebenarnya, sementara recall menunjukkan bahwa model berhasil mengidentifikasi 83% dari tweet yang benar-benar sesuai. F1-score sebesar 83% mencerminkan keseimbangan yang baik antara precision dan recall.

Tabel 8. Nilai Precision, Recall, dan F1-Score Evaluasi Model

Jenis Klasifikasi	Presisi	Recall	F1-Score
Negatif	86%	87%	86%
Netral	71%	65%	68%
Positif	86%	89%	88%

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil menerapkan model *Convolutional Neural Network* (CNN) untuk menganalisis sentimen tweet terkait Pemilihan Umum Presiden Indonesia 2024. Dengan akurasi sebesar 84% serta precision, recall, dan F1-score masing-masing 83%, model mampu mengklasifikasikan sentimen positif (45,42%), negatif (37,53%), dan netral (17,04%) secara efektif. Penyesuaian pembobotan kelas berhasil meningkatkan kinerja CNN meskipun data tidak seimbang, menunjukkan potensi CNN dalam memahami opini publik melalui media sosial. Penelitian dapat diperbaiki dengan memperluas dataset, menambahkan analisis temporal untuk memahami dinamika sentimen, dan mempertimbangkan model seperti BERT untuk akurasi yang lebih tinggi. Analisis multi-bahasa serta integrasi dengan analisis topik juga dapat memberikan wawasan yang lebih mendalam mengenai isu-isu utama terkait Pemilu.

DAFTAR PUSTAKA

- Twitter Indonesia. (2024). Twitter Active User Statistics in Indonesia. [Online] Available at: <https://twitter.com/indonesia> [Accessed on 20 October 2024].
- P. L. Parameswari and Prihandoko, "PENGUNAAN CONVOLUTIONAL NEURAL NETWORK UNTUK ANALISIS SENTIMEN OPINI LINGKUNGAN HIDUP KOTA DEPOK DI TWITTER," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 27, no. 1, pp. 29–42, 2022, doi: 10.35760/tr.2022.v27i1.4671.

- [3] A. P. Putra and A. F. Syafira, "Analisis Sentimen Data Twitter Topik Politik Dengan Metode Naive Bayes Dan *Convolutional Neural Networks* (Cnn)," Jurnal Ilmiah Wahana Pendidikan, Oktober, vol. 2023, no. 20, pp. 36–41, doi: 10.5281/zenodo.8396579.
- [4] A. Perdana, A. Hermawan, and D. Avianto, "Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Classifier," Jurnal Sisfokom (Sistem Informasi dan Komputer), vol. 11, no. 2, pp. 195–200, Jul. 2022, doi: 10.32736/sisfokom.v11i2.1412.
- [5] Aman, H., & Kahar, I. (2022). Twitter Sentiment Analysis on Political Elections Using Machine Learning Algorithms: A Case Study on Indonesian Presidential Election. *Journal of Data Science and Applications*, 10(2), 55-67.
- [6] R. A. Rudiyanto and E. B. Setiawan, "Sentiment Analysis Using *Convolutional Neural Network* (CNN) and Particle Swarm Optimization on Twitter," JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer), vol. 9, no. 2, pp. 188–195, Feb. 2024, doi: 10.33480/jitk.v9i2.5201.
- [7] A. H. Mulyadi and S. Lestari, "Terbit online pada laman web jurnal: <https://ejurnalunsam.id/index.php/jicom/> Analisis Sentimen Terhadap Sekolah Saat Covid-19 Pada Twitter Menggunakan Metode Lexicon Based", [Online]. Available: <https://ejurnalunsam.id/index.php/jicom/>.
- [8] Angelmetanosaa, "dataset," GitHub. Accessed: Dec. 02, 2024. [Online]. Available: <https://github.com/angelmetanosaa/dataset>.
- [9] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, "Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naive Bayes," JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 5, no. 2, p. 422, Apr. 2021, doi: 10.30865/mib.v5i2.2845.
- [10] R. Vindua and A. U. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python," JURIKOM (Jurnal Riset Komputer), vol. 10, no. 2, p. 479, Apr. 2023, doi: 10.30865/jurikom.v10i2.5945.
- [11] T. Krisdiyanto, E. Maricha, and O. Nurharyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naive Bayes Classifiers," Jurnal CoreIT, vol. 7, no. 1, 2021.