

Peningkatan Akurasi Analisis Sentimen Publik tentang kehadiran *Artificial Intelligence* menggunakan Model *Hybrid Deep Learning* berbasis NLP

Improving the Accuracy of Public Sentiment Analysis on the Presence of Artificial Intelligence using an NLP-based Hybrid Deep Learning Model

¹Daniel Sintong Pardamean Simanjuntak*, ²FA Bambang Sukoco, ³Ahmad Husaein,
⁴Muhammad Zidan

^{1,4}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dinamika Bangsa

^{2,3}Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dinamika Bangsa

^{1,2,3,4}Jl. Jendral Sudirman, Thehok, Jambi Selatan, Jambi, Indonesia, 36138

*e-mail: daniel.sintong.87@gmail.com

(received: 13 July 2026, revised: 28 September 2026, accepted: 29 September 2026)

Abstrak

Perkembangan Artificial Intelligence (AI) telah memunculkan beragam respons publik yang terekspressi secara luas melalui media sosial, seperti Twitter/X, TikTok, dan YouTube. Opini publik terhadap AI menjadi penting untuk dipahami karena dapat memberikan gambaran mengenai tingkat penerimaan, kekhawatiran, dan persepsi masyarakat terhadap perkembangan teknologi tersebut. Namun, analisis sentimen pada teks media sosial berbahasa Indonesia masih menghadapi tantangan akibat karakteristik bahasa yang informal, tidak terstruktur, mengandung slang, typo, serta variasi morfologis yang tinggi. Penelitian ini bertujuan mengembangkan model analisis sentimen berbasis hybrid Bidirectional Long Short-Term Memory–Conditional Random Fields (BiLSTM–CRF) dengan integrasi Word2Vec dan FastText untuk meningkatkan performa klasifikasi sentimen publik terhadap Artificial Intelligence. Dataset penelitian dikumpulkan dari tiga platform media sosial, yaitu Twitter/X, TikTok, dan YouTube, dengan total 30.550 komentar setelah proses preprocessing dan pelabelan ke dalam kategori positif, netral, dan negatif. Eksperimen dilakukan dengan membandingkan model yang diusulkan terhadap model baseline CRF, LSTM, dan BiLSTM menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model BiLSTM–CRF dengan FastText memberikan performa terbaik dengan accuracy 93.12%, precision 92.74%, recall 92.31%, dan F1-score 92.52%, melampaui BiLSTM–CRF dengan Word2Vec (91.38%), BiLSTM (87.24%), LSTM (84.15%), dan CRF (78.42%). Temuan ini menunjukkan bahwa integrasi BiLSTM–CRF dengan representasi subword FastText efektif dalam meningkatkan akurasi analisis sentimen pada data media sosial berbahasa Indonesia. Penelitian ini berkontribusi pada pengembangan metode analisis sentimen yang lebih robust, adaptif, dan representatif untuk memahami opini publik terhadap Artificial Intelligence.

Kata kunci: artificial intelligence, BiLSTM–CRF, fasttext, media sosial, analisis sentimen.

Abstract

The development of Artificial Intelligence (AI) has generated diverse public responses, which are widely expressed through social media platforms such as Twitter/X, TikTok, and YouTube. Understanding public sentiment toward AI is important because it provides insights into public acceptance, concerns, and perceptions regarding the development of this technology. However, sentiment analysis of Indonesian-language social media text remains challenging due to the informal and unstructured nature of the language, including slang, typographical errors, and substantial morphological variations. This study aims to develop a hybrid sentiment analysis model based on Bidirectional Long Short-Term Memory–Conditional Random Fields (BiLSTM–CRF), integrating Word2Vec and FastText to improve the performance of public sentiment classification toward Artificial Intelligence. The dataset was collected from three social media platforms—Twitter/X, TikTok, and YouTube—and comprised 30,550 comments after preprocessing and labeling into three

<http://sistemasi.ftik.unisi.ac.id>

sentiment categories: positive, neutral, and negative. The experiments compared the proposed models with the baseline CRF, LSTM, and BiLSTM models using accuracy, precision, recall, and F1-score as evaluation metrics. The results show that the BiLSTM-CRF model with FastText achieved the best performance, with an accuracy of 93.12%, precision of 92.74%, recall of 92.31%, and F1-score of 92.52%. This performance exceeded that of BiLSTM-CRF with Word2Vec (91.38%), BiLSTM (87.24%), LSTM (84.15%), and CRF (78.42%). These findings demonstrate that integrating BiLSTM-CRF with FastText's subword-based representation is effective in improving sentiment analysis performance on Indonesian-language social media data. This study contributes to the development of more robust, adaptive, and representative sentiment analysis methods for understanding public opinion toward Artificial Intelligence.

Keywords: *artificial intelligence, BiLSTM-CRF, fasttext, social media, sentiment analysis.*

1 Pendahuluan

Perkembangan Artificial Intelligence (AI) dalam beberapa tahun terakhir telah membawa transformasi signifikan di berbagai sektor, termasuk industri, pemerintahan, pendidikan, layanan publik, dan ekonomi digital [1][2][3]. Teknologi AI telah diterapkan dalam berbagai bentuk, seperti chatbot, sistem rekomendasi, predictive analytics, computer vision, dan generative AI, yang bertujuan meningkatkan efisiensi operasional, akurasi pengambilan keputusan, dan produktivitas kerja [1]. Di sisi lain, kehadiran AI juga menimbulkan berbagai kekhawatiran di masyarakat, seperti ancaman terhadap privasi data, bias algoritmik, keamanan informasi, hingga potensi hilangnya pekerjaan akibat otomatisasi [1][4]. Kondisi ini menunjukkan bahwa persepsi publik terhadap AI tidak bersifat seragam, melainkan dinamis dan dipengaruhi oleh berbagai faktor sosial, teknologi, dan pengalaman pengguna.

Pemahaman terhadap persepsi publik menjadi aspek penting dalam pengembangan dan implementasi teknologi AI yang bertanggung jawab. Persepsi masyarakat dapat memengaruhi tingkat penerimaan teknologi, efektivitas kebijakan digital, serta arah inovasi teknologi di masa depan. Dalam konteks transformasi digital, opini publik tidak hanya menjadi indikator penerimaan sosial, tetapi juga dapat digunakan sebagai dasar pengambilan keputusan strategis bagi pemerintah, industri, dan akademisi. Oleh karena itu, diperlukan pendekatan analitis yang mampu mengidentifikasi, mengukur, dan memahami kecenderungan opini masyarakat terhadap perkembangan AI secara sistematis.

Media sosial telah menjadi salah satu platform utama dalam mengekspresikan opini publik terhadap isu-isu teknologi secara real-time. Platform seperti Twitter/X, TikTok, dan YouTube memungkinkan pengguna menyampaikan pandangan, tanggapan, dan pengalaman secara spontan melalui komentar, diskusi, maupun interaksi digital lainnya [5]. Twitter/X umumnya digunakan untuk menyampaikan opini singkat dan cepat terhadap isu terkini, TikTok memunculkan pola interaksi berbasis komentar terhadap konten visual yang sedang tren, sedangkan YouTube menyediakan ruang diskusi yang lebih panjang dan argumentatif. Perbedaan karakteristik ini menjadikan media sosial sebagai sumber data yang kaya untuk memahami opini publik secara lebih komprehensif. Dalam konteks ini, analisis sentimen menjadi salah satu pendekatan yang efektif untuk mengidentifikasi kecenderungan opini publik ke dalam kategori positif, negatif, atau netral [6].

Meskipun analisis sentimen telah banyak digunakan dalam berbagai domain, implementasinya pada data teks media sosial berbahasa Indonesia masih menghadapi tantangan yang kompleks. Bahasa yang digunakan pada media sosial Indonesia cenderung informal, tidak baku, mengandung singkatan, kata slang, campuran bahasa asing, kesalahan penulisan (*typo*), serta variasi morfologis yang tinggi [6][5][9], [10]. Karakteristik tersebut menyebabkan proses ekstraksi makna dari teks menjadi lebih sulit dibandingkan teks formal. Pendekatan machine learning klasik seperti Naive Bayes, Support Vector Machine (SVM), Logistic Regression, dan metode berbasis frekuensi kata seperti Bag-of-Words atau TF-IDF sering mengalami keterbatasan dalam menangkap hubungan semantik antar kata maupun konteks kalimat secara menyeluruh [6][7].

Seiring perkembangan teknologi komputasi, pendekatan deep learning mulai banyak digunakan untuk meningkatkan performa analisis sentimen. Model berbasis Recurrent Neural Network (RNN), khususnya Long Short-Term Memory (LSTM), memiliki kemampuan untuk menangani data sekuensial dan memahami hubungan antar kata dalam suatu kalimat [7]. Pengembangan lebih lanjut melalui Bidirectional Long Short-Term Memory (BiLSTM) memungkinkan model memahami

konteks teks dari dua arah, yaitu berdasarkan kata sebelumnya dan kata sesudahnya, sehingga menghasilkan representasi kontekstual yang lebih baik. Berbagai penelitian menunjukkan bahwa pendekatan deep learning mampu memberikan performa klasifikasi yang lebih tinggi dibanding machine learning konvensional dalam berbagai tugas analisis sentimen [7][8].

Namun demikian, model BiLSTM masih memiliki keterbatasan dalam menjaga konsistensi hubungan antar label prediksi pada data teks sekuensial yang kompleks. Pada data media sosial, struktur kalimat yang tidak teratur, ambigu, dan penuh noise dapat menyebabkan model menghasilkan prediksi yang tidak stabil. Untuk mengatasi permasalahan tersebut, Conditional Random Fields (CRF) dapat digunakan sebagai lapisan sequence modeling yang mempertimbangkan dependensi antar label prediksi [9][7]. Integrasi antara BiLSTM dan CRF menghasilkan model hybrid yang mampu menggabungkan kekuatan representasi kontekstual dari BiLSTM dengan kemampuan CRF dalam menjaga konsistensi prediksi sekuensial [10][8]. Pendekatan BiLSTM-CRF telah menunjukkan performa yang baik pada berbagai tugas NLP seperti Named Entity Recognition (NER) dan Part-of-Speech Tagging, namun implementasinya pada analisis sentimen berbahasa Indonesia masih relatif terbatas [9][10].

Selain arsitektur model, kualitas representasi fitur teks juga memegang peranan penting dalam menentukan performa klasifikasi. Representasi fitur tradisional seperti TF-IDF hanya mengandalkan frekuensi kemunculan kata tanpa memahami hubungan makna antar kata[11]. Untuk mengatasi keterbatasan tersebut, pendekatan word embedding seperti Word2Vec dan FastText dikembangkan untuk merepresentasikan kata dalam bentuk vektor kontinu yang mampu menangkap hubungan semantik[12]. Word2Vec efektif dalam memodelkan kemiripan makna berdasarkan konteks penggunaan kata, sedangkan FastText memiliki keunggulan tambahan karena memanfaatkan informasi subword, sehingga lebih robust terhadap kata tidak baku, typo, dan istilah slang yang umum ditemukan pada media sosial Indonesia [15]. Penelitian sebelumnya menunjukkan bahwa penggunaan word embedding mampu meningkatkan performa model analisis sentimen dibanding pendekatan berbasis fitur tradisional [13][14].

Beberapa penelitian terdahulu telah mengembangkan model analisis sentimen berbasis deep learning. Khan et al. [7] menunjukkan bahwa model CNN-LSTM mampu meningkatkan performa analisis sentimen dibanding metode machine learning klasik. Sorour et al. [8] mengembangkan pendekatan hybrid deep learning untuk sentiment classification dengan hasil yang lebih stabil pada customer review. Putri [15] menerapkan model BERT dan RoBERTa untuk menganalisis sentimen publik terhadap Artificial Intelligence di Indonesia dengan hasil akurasi yang tinggi. Namun demikian, sebagian besar penelitian tersebut masih memiliki beberapa keterbatasan, seperti penggunaan single-platform dataset, fokus pada data formal, atau kebutuhan komputasi yang tinggi pada model transformer.

Berdasarkan kajian tersebut, terdapat beberapa research gap yang masih terbuka. Pertama, sebagian besar penelitian analisis sentimen di Indonesia masih menggunakan data dari satu platform media sosial, sehingga belum mampu merepresentasikan opini publik secara lebih luas[5]. Kedua, implementasi BiLSTM-CRF untuk analisis sentimen bahasa Indonesia masih terbatas, karena model ini lebih banyak digunakan pada tugas sequence labeling seperti NER [9][10]. Ketiga, kombinasi dual word embedding (Word2Vec dan FastText) dalam model hybrid untuk menangani teks informal media sosial Indonesia masih jarang dieksplorasi [13][14]. Keempat, penelitian yang secara khusus membahas sentimen publik terhadap perkembangan Artificial Intelligence berbasis multi-platform media sosial masih relatif sedikit[15].

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan model hybrid BiLSTM-CRF yang dikombinasikan dengan Word2Vec dan FastText untuk analisis sentimen opini publik terhadap Artificial Intelligence menggunakan data dari Twitter/X, TikTok, dan YouTube. Pendekatan ini diharapkan mampu menghasilkan model analisis sentimen yang lebih akurat, robust, dan adaptif terhadap karakteristik teks media sosial berbahasa Indonesia. Kontribusi penelitian ini meliputi pengembangan model hybrid sequence learning untuk sentiment analysis, pemanfaatan dual word embedding untuk meningkatkan representasi semantik teks informal, serta integrasi data multi-platform guna memperoleh gambaran opini publik yang lebih komprehensif terhadap perkembangan AI di Indonesia.

2 Tinjauan Literatur

Penelitian mengenai analisis sentimen telah berkembang pesat seiring meningkatnya penggunaan media sosial sebagai sarana komunikasi publik. Berbagai pendekatan telah digunakan, mulai dari machine learning klasik hingga deep learning modern, untuk meningkatkan akurasi klasifikasi opini berbasis teks. Sejumlah penelitian terdahulu menunjukkan bahwa performa model analisis sentimen sangat dipengaruhi oleh arsitektur model, representasi fitur, karakteristik dataset, serta konteks domain yang dianalisis.

Khan *et al.* [7] mengembangkan model CNN-LSTM untuk analisis sentimen pada data media sosial berbahasa Inggris dan Roman Urdu. Penelitian tersebut menunjukkan bahwa kombinasi lapisan convolutional dan recurrent mampu meningkatkan performa klasifikasi dibanding pendekatan machine learning tradisional. CNN berperan dalam mengekstraksi fitur lokal dari teks, sedangkan LSTM menangani dependensi sekuensial antar kata. Meskipun hasil penelitian menunjukkan peningkatan akurasi yang signifikan, model tersebut masih memiliki keterbatasan dalam menjaga konsistensi hubungan antar label prediksi pada struktur teks yang kompleks.

Sorour *et al.* [8][6] mengusulkan pendekatan hybrid deep learning untuk klasifikasi sentimen pada customer reviews. Penelitian tersebut menunjukkan bahwa kombinasi beberapa lapisan deep learning mampu meningkatkan stabilitas klasifikasi, precision, recall, dan F1-score dibanding model tunggal. Namun, data yang digunakan berupa customer review formal dengan struktur kalimat relatif teratur, sehingga belum sepenuhnya mewakili karakteristik data media sosial yang umumnya lebih informal, noisy, dan tidak terstruktur.

Dalam konteks sequence modeling, Khan *et al.* [9] menunjukkan bahwa Conditional Random Fields (CRF) efektif dalam tugas sequence labeling karena mampu mempertimbangkan hubungan antar label prediksi. Demikian pula, Qiu *et al.* [10] menunjukkan bahwa kombinasi BiLSTM dan CRF mampu meningkatkan performa entity recognition melalui pemodelan konteks dua arah dan sequence dependency. Meskipun pendekatan tersebut terbukti efektif pada tugas NLP seperti Named Entity Recognition (NER), penerapannya pada analisis sentimen, khususnya untuk teks media sosial berbahasa Indonesia, masih relatif terbatas.

Pada konteks Indonesia, Fitroh [6] melalui systematic literature review menunjukkan bahwa mayoritas penelitian analisis sentimen masih menggunakan pendekatan Naive Bayes, Support Vector Machine (SVM), Random Forest, CNN, dan LSTM. Hasil kajian menunjukkan bahwa deep learning umumnya memberikan performa lebih baik dibanding machine learning klasik, terutama pada dataset dengan volume besar. Namun demikian, tantangan utama tetap terletak pada kualitas preprocessing, representasi fitur, serta karakteristik bahasa informal Indonesia.

Setiawan [5] juga menegaskan bahwa implementasi analisis sentimen di Indonesia masih didominasi oleh penggunaan single-platform dataset, terutama Twitter. Pendekatan tersebut cukup efektif untuk studi awal, tetapi kurang mampu menggambarkan opini publik secara komprehensif karena setiap platform media sosial memiliki pola interaksi yang berbeda. Twitter/X cenderung berisi opini singkat dan spontan, TikTok didominasi komentar berbasis tren visual, sedangkan YouTube lebih banyak memuat diskusi panjang dan argumentatif.

Pada domain Artificial Intelligence, Putri [15] menerapkan model transformer berbasis BERT dan RoBERTa untuk analisis sentimen publik terhadap perkembangan AI di Indonesia. Penelitian tersebut menunjukkan bahwa model transformer memberikan akurasi klasifikasi yang tinggi karena mampu memahami konteks bahasa secara mendalam. Namun, pendekatan transformer memiliki kebutuhan komputasi yang besar, waktu pelatihan yang lebih lama, serta kompleksitas implementasi yang lebih tinggi, sehingga kurang efisien untuk skenario penelitian dengan fokus optimasi model sequence learning yang lebih ringan.

Dalam aspek representasi fitur, Jasmir *et al.* [13] menunjukkan bahwa penggunaan Word2Vec mampu meningkatkan performa machine learning pada analisis sentimen komentar platform digital dibanding pendekatan berbasis TF-IDF. Word embedding memungkinkan model memahami hubungan semantik antar kata sehingga meningkatkan kualitas representasi fitur teks. Namun, Word2Vec masih memiliki kelemahan dalam menangani kata yang tidak ditemukan dalam vocabulary (*out-of-vocabulary*), yang sering muncul pada data media sosial informal.

Untuk mengatasi keterbatasan tersebut, Haziq *et al.* [14] menerapkan FastText pada analisis sentimen aplikasi digital berbasis deep learning. Hasil penelitian menunjukkan bahwa FastText

memberikan peningkatan performa karena mampu memanfaatkan informasi subword sehingga lebih robust terhadap typo, singkatan, kata slang, dan variasi morfologis. Karakteristik ini sangat relevan untuk analisis sentimen media sosial berbahasa Indonesia. Namun demikian, penelitian tersebut hanya menggunakan satu sumber data sehingga validitas generalisasi terhadap opini publik multi-platform masih terbatas.

Secara konseptual, Word2Vec yang diperkenalkan oleh Mikolov *et al.* [16] menjadi salah satu teknik representasi kata yang paling banyak digunakan dalam NLP karena mampu memetakan kata ke ruang vektor kontinu berdasarkan konteks penggunaannya. Pengembangan lebih lanjut melalui FastText oleh Bojanowski *et al.* [17] menambahkan representasi subword sehingga meningkatkan kemampuan model dalam menangani kata tidak baku. Kedua pendekatan ini menjadi fondasi penting dalam pengembangan model NLP modern berbasis embedding semantik.

Berdasarkan analisis terhadap penelitian terdahulu, dapat diidentifikasi beberapa kesenjangan penelitian yang masih terbuka. Pertama, sebagian besar penelitian analisis sentimen di Indonesia masih menggunakan single-platform dataset, sehingga representasi opini publik belum komprehensif [5][10]. Kedua, implementasi BiLSTM-CRF untuk sentiment analysis bahasa Indonesia masih sangat terbatas, karena model tersebut lebih sering diterapkan pada tugas sequence labeling [9][10][7], [8]. Ketiga, kombinasi dual word embedding (Word2Vec dan FastText) untuk meningkatkan robustness analisis sentimen media sosial Indonesia masih jarang dieksplorasi [13][14][12], [13]. Keempat, penelitian khusus mengenai opini publik terhadap Artificial Intelligence berbasis multi-platform media sosial masih relatif sedikit [15][11].

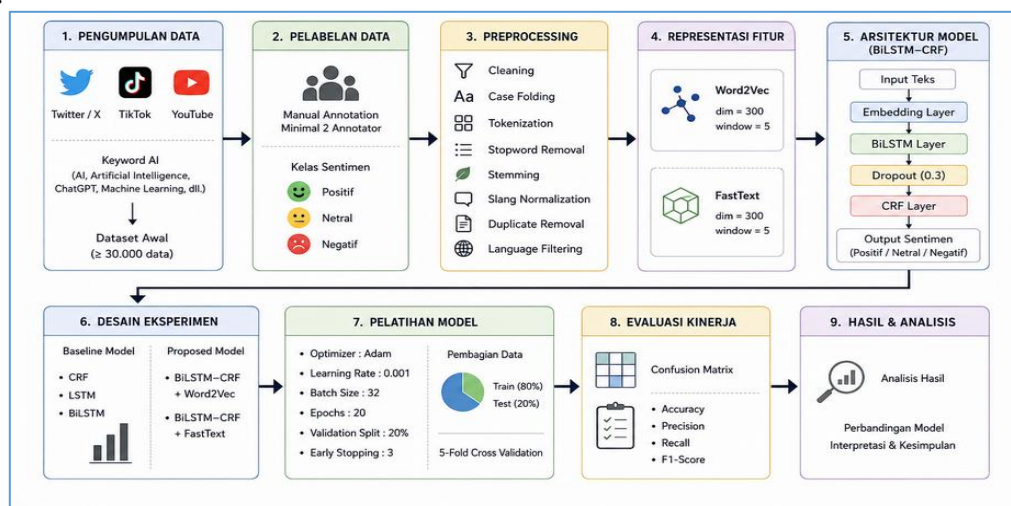
Dengan demikian, penelitian ini menawarkan pendekatan baru berupa integrasi BiLSTM-CRF dengan Word2Vec dan FastText untuk menganalisis sentimen publik terhadap Artificial Intelligence menggunakan data dari Twitter/X, TikTok, dan YouTube, sehingga diharapkan mampu memberikan hasil klasifikasi yang lebih akurat, stabil, dan representatif.

3 Metode Penelitian

3.1 Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental untuk mengembangkan dan mengevaluasi model analisis sentimen berbasis deep learning terhadap opini publik mengenai Artificial Intelligence (AI) pada media sosial. Fokus utama penelitian adalah membandingkan performa model hybrid BiLSTM-CRF dengan beberapa model baseline dalam mengklasifikasikan sentimen publik ke dalam kategori positif, negatif, dan netral.

Tahapan penelitian meliputi: pengumpulan data, preprocessing teks, pelabelan data, pembentukan representasi fitur menggunakan Word2Vec dan FastText, pembangunan model baseline, pengembangan model hybrid BiLSTM-CRF, pelatihan model, evaluasi performa, serta analisis hasil eksperimen. Alur penelitian dirancang untuk menghasilkan model klasifikasi sentimen yang akurat, robust, dan adaptif terhadap karakteristik teks media sosial berbahasa Indonesia seperti terlihat pada gambar 1.



Gambar 1 Kerangka kerja penelitian

3.2 Pengumpulan Data

Dataset penelitian diperoleh dari tiga platform media sosial, yaitu Twitter/X, TikTok, dan YouTube, yang dipilih untuk merepresentasikan karakteristik komunikasi publik yang berbeda terhadap isu Artificial Intelligence. Pengumpulan data dilakukan pada periode Oktober 2025 sampai Maret 2026 menggunakan teknik web scraping dan Application Programming Interface (API) sesuai dengan mekanisme akses data pada masing-masing platform.

Proses pengumpulan menggunakan kata kunci yang berkaitan dengan Artificial Intelligence, yaitu *Artificial Intelligence*, *AI*, *Kecerdasan Buatan*, *ChatGPT*, *Generative AI*, *Machine Learning*, *Deep Learning*, *Otomasi AI*, dan *Teknologi AI*. Komentar dimasukkan ke dalam dataset apabila memenuhi kriteria: (1) ditulis dalam Bahasa Indonesia atau didominasi Bahasa Indonesia; (2) memiliki keterkaitan langsung dengan topik Artificial Intelligence; (3) berupa komentar tekstual yang dapat diidentifikasi; dan (4) bukan merupakan komentar duplikat, spam, iklan, atau konten yang tidak relevan. Identitas personal pengguna tidak digunakan sebagai fitur dalam proses pemodelan.

Pada tahap awal diperoleh sebanyak 35.842 komentar. Setelah dilakukan penghapusan duplikasi, penyaringan Bahasa Indonesia, pembersihan *noise*, dan eliminasi komentar yang tidak relevan, diperoleh dataset akhir sebanyak 30.550 komentar, terdiri atas 10.420 komentar Twitter/X, 9.865 komentar TikTok, dan 10.265 komentar YouTube.

Penggunaan data multi-platform bertujuan meningkatkan keragaman karakteristik linguistik, tetapi tetap memiliki potensi bias. Twitter/X cenderung merepresentasikan opini yang singkat dan responsif terhadap isu aktual, TikTok lebih dipengaruhi oleh konteks konten visual dan tren, sedangkan YouTube cenderung memuat komentar yang lebih panjang dan argumentatif. Selain itu, pengguna ketiga platform tidak dapat dianggap merepresentasikan seluruh populasi masyarakat Indonesia. Oleh karena itu, hasil penelitian ini diinterpretasikan sebagai sentimen pengguna pada data media sosial yang dianalisis, bukan sebagai representasi statistik seluruh masyarakat Indonesia.

3.3 Pelabelan Data

Pelabelan data dilakukan untuk mengklasifikasikan opini publik ke dalam tiga kategori sentimen: Positif, Negatif dan Netral. Proses pelabelan dilakukan menggunakan pendekatan manual annotation oleh minimal dua annotator independen untuk menjaga konsistensi interpretasi sentimen. Jika terdapat perbedaan label antar annotator, maka dilakukan diskusi atau adjudikasi untuk menentukan label final. Untuk mengukur konsistensi pelabelan, digunakan Cohen's Kappa Score, yang dirumuskan sebagaimana persamaan (1):

$$K = \frac{P_o - P_e}{1 - P_e} \quad (1)$$

di mana:

- P_o = observed agreement
- P_e = expected agreement by chance

Nilai Kappa > 0.75 menunjukkan konsistensi pelabelan yang baik.

3.4 Preprocessing Data

Tahapan preprocessing dilakukan untuk meningkatkan kualitas data teks sebelum memasuki proses pembelajaran model. Mengingat data penelitian berasal dari media sosial, data yang diperoleh umumnya mengandung berbagai bentuk noise, seperti penggunaan bahasa informal, singkatan, slang, emotikon, URL, hashtag, serta karakter khusus yang dapat memengaruhi kinerja model klasifikasi. Oleh karena itu, preprocessing menjadi tahap penting untuk menghasilkan data yang lebih bersih, terstruktur, dan representatif.

Tahap pertama adalah case folding, yaitu mengubah seluruh karakter huruf menjadi huruf kecil (lowercase) untuk menghindari perbedaan representasi kata yang disebabkan oleh penggunaan huruf kapital. Selanjutnya dilakukan URL removal, yaitu menghapus seluruh tautan atau alamat situs web yang tidak memiliki kontribusi terhadap informasi sentimen. Pada tahap yang sama, dilakukan pula mention and hashtag removal untuk menghilangkan simbol mention (@) dan hashtag (#) yang umumnya tidak memberikan makna sentimen secara langsung.

Tahap berikutnya adalah number and punctuation removal, yaitu menghapus angka, tanda baca, dan karakter khusus yang tidak relevan dengan proses analisis sentimen. Setelah itu dilakukan

tokenization, yaitu memecah kalimat menjadi unit-unit kata (token) sehingga memudahkan proses representasi teks pada tahap selanjutnya. Token yang dihasilkan kemudian diproses menggunakan stopword removal untuk menghilangkan kata-kata umum yang memiliki frekuensi kemunculan tinggi tetapi tidak memberikan kontribusi signifikan terhadap penentuan sentimen, seperti “yang”, “dan”, “di”, “ke”, serta “dari”.

Selanjutnya dilakukan stemming untuk mengubah kata berimbuhan menjadi bentuk kata dasar menggunakan stemmer Bahasa Indonesia. Proses ini bertujuan untuk mengurangi variasi bentuk kata sehingga kata-kata yang memiliki makna sama dapat direpresentasikan secara konsisten. Selain itu, dilakukan slang normalization untuk mengonversi kata-kata tidak baku yang umum digunakan pada media sosial menjadi bentuk baku. Sebagai contoh, kata “gk” diubah menjadi “tidak”, “bgt” menjadi “banget”, dan “udh” menjadi “sudah”. Tahap ini penting karena data media sosial Indonesia sangat kaya akan penggunaan bahasa informal dan singkatan.

Untuk menghindari bias akibat data yang berulang, dilakukan duplicate removal guna menghapus komentar yang memiliki isi identik atau sangat mirip. Tahap terakhir adalah language filtering, yaitu menyaring dan mempertahankan hanya komentar berbahasa Indonesia agar sesuai dengan ruang lingkup penelitian. Hasil dari seluruh tahapan preprocessing ini adalah dataset yang lebih bersih, konsisten, dan siap digunakan pada proses pembentukan fitur menggunakan Word2Vec dan FastText serta pelatihan model hybrid BiLSTM-CRF.

3.5 Representasi Fitur

Representasi fitur merupakan tahapan penting dalam analisis sentimen karena menentukan bagaimana informasi semantik dari teks direpresentasikan dan dipahami oleh model klasifikasi. Pada penelitian ini, representasi teks dilakukan menggunakan dua pendekatan word embedding, yaitu Word2Vec dan FastText. Kedua metode dipilih karena mampu menghasilkan representasi kata dalam bentuk vektor numerik yang dapat menangkap hubungan semantik dan kontekstual antar kata, sehingga lebih efektif dibandingkan representasi tradisional seperti Bag-of-Words atau TF-IDF.

Metode pertama yang digunakan adalah Word2Vec, yaitu teknik word embedding yang mempelajari representasi vektor kata berdasarkan konteks kemunculannya dalam korpus teks. Penelitian ini menggunakan pendekatan Skip-Gram, yang bertujuan memprediksi kata-kata di sekitar suatu kata target sehingga mampu menghasilkan representasi semantik yang lebih baik, terutama untuk kata-kata yang jarang muncul. Word2Vec digunakan untuk menangkap hubungan makna antar kata dalam komentar media sosial terkait Artificial Intelligence. Parameter yang digunakan pada proses pelatihan Word2Vec meliputi embedding dimension sebesar 300, window size sebesar 5, minimum word frequency (min_count) sebesar 3, dan jumlah pelatihan (epochs) sebanyak 20 iterasi. Konfigurasi ini dipilih untuk menghasilkan representasi kata yang kaya secara semantik tanpa meningkatkan kompleksitas komputasi secara berlebihan.

Selain Word2Vec, penelitian ini juga memanfaatkan FastText sebagai metode representasi fitur kedua. FastText merupakan pengembangan dari Word2Vec yang tidak hanya mempelajari representasi kata secara utuh, tetapi juga mempertimbangkan informasi subword atau karakter penyusun kata. Pendekatan ini memungkinkan FastText untuk menangani kata yang tidak terdapat dalam kosakata (*out-of-vocabulary*), kesalahan penulisan (*typo*), singkatan, maupun kata-kata slang yang umum ditemukan pada media sosial. Karakteristik tersebut menjadikan FastText sangat sesuai untuk memproses data teks berbahasa Indonesia yang bersifat informal dan tidak terstruktur. Parameter yang digunakan pada pelatihan FastText sama dengan Word2Vec, yaitu embedding dimension 300, window size 5, min_count 3, dan epochs 20, sehingga perbandingan performa kedua metode dapat dilakukan secara lebih objektif.

Representasi vektor yang dihasilkan oleh Word2Vec dan FastText selanjutnya digunakan sebagai input pada model hybrid BiLSTM-CRF. Dengan memanfaatkan kedua pendekatan embedding tersebut, penelitian ini bertujuan mengevaluasi kemampuan masing-masing metode dalam merepresentasikan makna kata serta pengaruhnya terhadap peningkatan performa klasifikasi sentimen pada data media sosial berbahasa Indonesia.

3.6 Arsitektur Model

Penelitian ini menggunakan dua kelompok model, yaitu model baseline dan model yang diusulkan. Model baseline digunakan sebagai pembanding untuk mengevaluasi efektivitas pendekatan

<http://sistemasi.ftik.unisi.ac.id>

yang dikembangkan dalam penelitian ini. Sementara itu, model yang diusulkan dirancang untuk meningkatkan kemampuan klasifikasi sentimen pada data media sosial berbahasa Indonesia yang memiliki karakteristik bahasa informal, tidak terstruktur, dan kaya akan variasi morfologis.

Model baseline yang digunakan terdiri atas Conditional Random Fields (CRF), Long Short-Term Memory (LSTM), dan Bidirectional Long Short-Term Memory (BiLSTM). CRF dipilih karena merupakan salah satu metode sequence labeling yang mampu memodelkan hubungan antar label dalam suatu urutan data. Namun, CRF memiliki keterbatasan dalam melakukan ekstraksi fitur secara otomatis karena masih bergantung pada representasi fitur yang diberikan. Selanjutnya, LSTM digunakan sebagai model baseline deep learning karena memiliki kemampuan dalam menangkap dependensi jangka panjang pada data sekuensial. Meskipun demikian, LSTM hanya memproses informasi dari satu arah sehingga informasi konteks yang diperoleh masih terbatas. Untuk mengatasi kelemahan tersebut, digunakan BiLSTM yang mampu memproses informasi dari dua arah sekaligus, yaitu forward dan backward, sehingga representasi konteks kalimat menjadi lebih lengkap.

Sebagai solusi yang diusulkan, penelitian ini mengembangkan model hybrid BiLSTM–CRF yang dikombinasikan dengan dua metode word embedding, yaitu Word2Vec dan FastText. Pada arsitektur ini, Word2Vec dan FastText berfungsi sebagai lapisan representasi fitur yang mengubah teks menjadi vektor numerik yang merepresentasikan hubungan semantik antar kata. Selanjutnya, lapisan BiLSTM digunakan untuk mempelajari konteks linguistik dari dua arah sehingga mampu menangkap dependensi kata yang lebih kompleks. Output dari BiLSTM kemudian diteruskan ke lapisan CRF yang berfungsi sebagai mekanisme decoding untuk menghasilkan prediksi label sentimen yang lebih konsisten dengan mempertimbangkan hubungan antar label dalam suatu sekuens.

Dua variasi model yang diusulkan yang dievaluasi dalam penelitian ini adalah BiLSTM–CRF dengan Word2Vec dan BiLSTM–CRF dengan FastText. Penggunaan dua jenis embedding tersebut bertujuan untuk menganalisis pengaruh representasi semantik terhadap performa klasifikasi sentimen. Melalui perbandingan dengan model baseline, penelitian ini berupaya mengidentifikasi arsitektur yang paling efektif dalam mengklasifikasikan sentimen publik terhadap Artificial Intelligence pada data media sosial multi-platform.

3.7 Arsitektur Hybrid BiLSTM–CRF

Model yang diusulkan dalam penelitian ini menggunakan arsitektur Hybrid Bidirectional Long Short-Term Memory–Conditional Random Fields (BiLSTM–CRF) yang dirancang untuk menggabungkan kemampuan representasi konteks sekuensial dari BiLSTM dengan kemampuan sequence decoding dari CRF. Arsitektur ini dipilih karena mampu menangkap hubungan semantik antar kata sekaligus mempertahankan konsistensi prediksi label sentimen pada data teks media sosial yang bersifat dinamis dan tidak terstruktur.

Lapisan pertama pada model adalah Embedding Layer, yang berfungsi mengubah teks hasil preprocessing menjadi representasi vektor numerik menggunakan metode Word2Vec atau FastText. Pada tahap ini, setiap kata direpresentasikan dalam ruang vektor berdimensi 300 sehingga hubungan semantik antar kata dapat dipelajari oleh model. Representasi embedding ini menjadi input utama bagi lapisan pembelajaran sekuensial pada tahap berikutnya.

Selanjutnya, data diproses oleh Bidirectional Long Short-Term Memory (BiLSTM) Layer. Berbeda dengan LSTM konvensional yang hanya memproses informasi dari satu arah, BiLSTM memproses urutan kata dari arah maju (*forward*) dan mundur (*backward*) secara simultan. Pendekatan ini memungkinkan model memahami konteks suatu kata berdasarkan kata-kata sebelum dan sesudahnya, sehingga menghasilkan representasi kontekstual yang lebih lengkap. Pada penelitian ini, lapisan BiLSTM menggunakan 128 hidden units, dengan dropout sebesar 0,3 dan recurrent dropout sebesar 0,3 untuk mengurangi risiko overfitting selama proses pelatihan model.

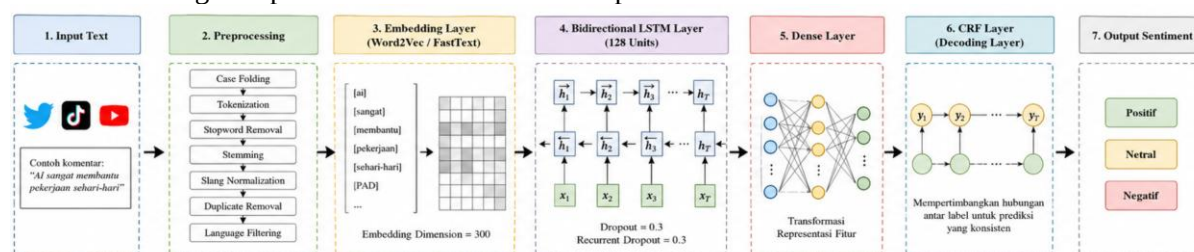
Output dari lapisan BiLSTM kemudian diteruskan ke Dense Layer yang berfungsi melakukan transformasi representasi fitur menjadi bentuk yang lebih sesuai untuk proses klasifikasi. Dense layer membantu mengekstraksi informasi penting dari representasi kontekstual yang dihasilkan BiLSTM sebelum dilakukan proses prediksi akhir. Lapisan ini juga berperan dalam meningkatkan kemampuan model untuk membedakan karakteristik sentimen positif, negatif, dan netral.

Tahap akhir pada arsitektur model adalah Conditional Random Fields (CRF) Layer. Berbeda dengan fungsi aktivasi Softmax yang melakukan prediksi secara independen pada setiap keluaran, CRF mempertimbangkan hubungan antar label prediksi dalam suatu urutan sehingga menghasilkan

<http://sistemasi.ftik.unisi.ac.id>

keputusan klasifikasi yang lebih konsisten. Dengan demikian, CRF berfungsi sebagai mekanisme decoding yang mampu memperbaiki prediksi label berdasarkan pola transisi antar label yang dipelajari selama proses pelatihan. Integrasi BiLSTM dan CRF memungkinkan model memanfaatkan informasi kontekstual yang kaya sekaligus menjaga konsistensi prediksi sentimen.

Secara keseluruhan, alur kerja model yang diusulkan seperti pada Gambar 2 dimulai dari Input Text, kemudian diproses melalui Embedding Layer (Word2Vec/FastText), dilanjutkan ke BiLSTM Layer, Dense Layer, dan akhirnya CRF Layer untuk menghasilkan prediksi sentimen akhir berupa kategori positif, netral, atau negatif. Arsitektur ini diharapkan mampu memberikan performa klasifikasi yang lebih baik dibandingkan model baseline dalam menganalisis opini publik terhadap Artificial Intelligence pada data media sosial multi-platform.



Gambar 2 Arsitektur Model hybrid BiLSTM-CRF

3.8 Pelatihan Model

Proses pelatihan model dilakukan untuk mempelajari pola sentimen yang terdapat pada data opini publik terkait Artificial Intelligence. Pada penelitian ini, seluruh model dilatih menggunakan algoritma optimasi Adaptive Moment Estimation (Adam) karena memiliki kemampuan konvergensi yang cepat dan stabil dalam proses pembelajaran jaringan saraf. Nilai learning rate ditetapkan sebesar 0,001 untuk menjaga keseimbangan antara kecepatan pembelajaran dan stabilitas proses optimasi. Selain itu, pelatihan model dilakukan menggunakan batch size sebesar 32 dan jumlah epoch sebanyak 20 iterasi, yang dinilai cukup untuk memungkinkan model mempelajari representasi fitur tanpa meningkatkan risiko overfitting secara berlebihan.

Untuk memantau performa model selama proses pelatihan, digunakan validation split sebesar 20% dari data pelatihan. Data validasi digunakan untuk mengevaluasi kemampuan generalisasi model pada setiap epoch sehingga proses pembelajaran dapat dikontrol dengan lebih baik. Selain itu, diterapkan mekanisme early stopping dengan nilai patience sebesar 3, yang berfungsi menghentikan proses pelatihan secara otomatis apabila tidak terjadi peningkatan performa pada data validasi dalam tiga epoch berturut-turut. Strategi ini bertujuan untuk mengurangi risiko overfitting dan meningkatkan efisiensi waktu pelatihan.

Sebelum proses pelatihan dilakukan, dataset dibagi menjadi dua bagian, yaitu 80% sebagai data pelatihan (training set) dan 20% sebagai data pengujian (testing set). Data pelatihan digunakan untuk membangun model, sedangkan data pengujian digunakan untuk mengevaluasi performa akhir model terhadap data yang belum pernah dilihat sebelumnya. Pembagian data ini dilakukan secara proporsional untuk menjaga distribusi kelas sentimen pada setiap subset data.

Untuk meningkatkan reliabilitas dan validitas hasil eksperimen, penelitian ini menerapkan Stratified 5-Fold Cross Validation. Metode ini membagi dataset menjadi lima subset dengan mempertahankan proporsi distribusi kelas sentimen pada setiap fold. Pada setiap iterasi, empat fold digunakan sebagai data pelatihan dan satu fold digunakan sebagai data pengujian, hingga seluruh fold memperoleh kesempatan menjadi data pengujian. Hasil evaluasi dari lima iterasi kemudian dirata-ratakan untuk memperoleh estimasi performa model yang lebih stabil dan representatif. Penggunaan Stratified 5-Fold Cross Validation diharapkan mampu mengurangi bias akibat pembagian data tunggal serta meningkatkan keandalan hasil evaluasi model yang diusulkan.

3.9 Evaluasi Model

Evaluasi model dilakukan untuk mengukur kemampuan model dalam mengklasifikasikan sentimen publik terhadap Artificial Intelligence secara akurat dan konsisten. Pada penelitian ini, proses evaluasi dilakukan menggunakan confusion matrix dan beberapa metrik evaluasi yang umum digunakan pada tugas klasifikasi, yaitu accuracy, precision, recall, dan F1-score. Penggunaan beberapa metrik evaluasi bertujuan untuk memberikan gambaran yang lebih komprehensif mengenai

performa model, terutama dalam menangani distribusi data yang berpotensi tidak seimbang antar kelas sentimen.

Accuracy digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data yang dievaluasi, sebagaimana dihitung menggunakan Persamaan (2). Metrik ini memberikan gambaran umum mengenai tingkat keberhasilan model dalam melakukan klasifikasi sentimen. Namun, accuracy saja belum cukup untuk menggambarkan kualitas model secara menyeluruh, terutama ketika distribusi kelas tidak seimbang.

Untuk melengkapi evaluasi, precision dihitung menggunakan Persamaan (3). Precision mengukur tingkat ketepatan prediksi model terhadap suatu kelas sentimen, yaitu proporsi data yang diprediksi sebagai suatu kelas dan benar-benar termasuk dalam kelas tersebut. Sementara itu, recall dihitung menggunakan Persamaan (4) untuk mengukur kemampuan model dalam mengidentifikasi seluruh data yang sebenarnya termasuk ke dalam suatu kelas sentimen. Nilai recall yang tinggi menunjukkan bahwa model mampu mengenali sebagian besar data yang relevan.

Selain precision dan recall, penelitian ini menggunakan F1-score yang dihitung menggunakan Persamaan (5). F1-score merupakan rata-rata harmonis antara precision dan recall sehingga dapat memberikan ukuran yang lebih seimbang antara ketepatan prediksi dan kemampuan model dalam mendeteksi data pada setiap kelas sentimen. Metrik ini terutama berguna ketika distribusi data antar kelas tidak sepenuhnya seimbang.

Untuk memperoleh analisis yang lebih rinci, digunakan confusion matrix yang menyajikan informasi mengenai jumlah prediksi yang benar maupun salah pada setiap kategori sentimen. Melalui confusion matrix dapat diketahui nilai True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) yang dihasilkan oleh model. Informasi ini memungkinkan peneliti untuk mengidentifikasi pola kesalahan klasifikasi, mengevaluasi kemampuan model dalam membedakan sentimen positif, netral, dan negatif, serta memberikan dasar yang lebih kuat dalam membandingkan performa antara model baseline dan model yang diusulkan.

Hasil evaluasi yang diperoleh dari seluruh metrik tersebut selanjutnya digunakan sebagai dasar untuk menentukan model terbaik serta menganalisis efektivitas integrasi Word2Vec dan FastText pada arsitektur hybrid BiLSTM-CRF dalam meningkatkan performa analisis sentimen pada data media sosial berbahasa Indonesia.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

3.10 Lingkungan Implementasi

Seluruh eksperimen dalam penelitian ini diimplementasikan menggunakan bahasa pemrograman Python versi 3.10 karena memiliki dukungan yang luas terhadap berbagai pustaka (*library*) yang digunakan dalam pengembangan model machine learning dan deep learning. Proses pembentukan, pelatihan, dan evaluasi model deep learning dilakukan menggunakan TensorFlow dengan antarmuka Keras, yang menyediakan framework komputasi yang fleksibel dan efisien untuk membangun arsitektur hybrid BiLSTM-CRF. Selain itu, pustaka scikit-learn digunakan untuk mendukung proses preprocessing data, pembagian dataset, validasi silang, serta perhitungan berbagai metrik evaluasi seperti accuracy, precision, recall, dan F1-score.

Pada tahap representasi fitur, penelitian ini memanfaatkan pustaka Gensim untuk membangun model word embedding berbasis Word2Vec dan FastText. Sementara itu, implementasi model Conditional Random Fields (CRF) dilakukan menggunakan pustaka sklearn-crfsuite, yang menyediakan berbagai fungsi untuk pelatihan dan evaluasi model sequence labeling. Untuk pengolahan data dan manipulasi dataset, digunakan pustaka Pandas, sedangkan NumPy digunakan untuk mendukung operasi komputasi numerik yang diperlukan selama proses pelatihan dan pengujian model.

Eksperimen dijalankan pada lingkungan komputasi yang menggunakan prosesor Intel Core i7 atau AMD Ryzen 7, dengan kapasitas memori utama (RAM) sebesar 16 GB. Konfigurasi perangkat keras ini dinilai memadai untuk menangani proses preprocessing, pembentukan word embedding, pelatihan model deep learning, serta evaluasi performa pada dataset berukuran besar. Untuk mempercepat proses pelatihan model, penelitian ini juga mendukung penggunaan Graphics Processing Unit (GPU) NVIDIA RTX Series sebagai akselerator komputasi. Pemanfaatan GPU memungkinkan proses pelatihan model BiLSTM-CRF dilakukan dengan waktu yang lebih efisien dibandingkan penggunaan CPU secara penuh, terutama pada eksperimen yang melibatkan data dalam jumlah besar dan proses validasi berulang.

Kombinasi perangkat lunak dan perangkat keras yang digunakan pada penelitian ini dirancang untuk mendukung proses eksperimen secara optimal, sehingga model yang diusulkan dapat dikembangkan, dilatih, dan dievaluasi secara efektif serta menghasilkan performa yang dapat direproduksi pada lingkungan penelitian yang serupa.

4 Hasil dan Pembahasan

4.1 Deskripsi Dataset

Dataset penelitian diperoleh dari tiga platform media sosial, yaitu Twitter/X, TikTok, dan YouTube, yang dipilih karena merepresentasikan karakteristik komunikasi publik yang berbeda terhadap isu Artificial Intelligence. Proses pengumpulan data dilakukan menggunakan teknik scraping dan API berdasarkan kata kunci yang relevan, seperti *Artificial Intelligence*, *AI*, *Kecerdasan Buatan*, *ChatGPT*, *Machine Learning*, dan *Generative AI*. Pada tahap awal, diperoleh sebanyak 35.842 komentar dari seluruh platform. Setelah dilakukan proses pembersihan data, penghapusan duplikasi, penyaringan bahasa, dan eliminasi komentar yang tidak relevan, jumlah dataset final yang digunakan dalam eksperimen sebanyak 30.550 data teks.

Meskipun distribusi jumlah data antar platform relatif seimbang, keseimbangan jumlah komentar tidak sepenuhnya menghilangkan *platform bias*. Perbedaan karakteristik pengguna, pola interaksi, panjang komentar, tren konten, dan mekanisme rekomendasi pada masing-masing platform dapat memengaruhi distribusi opini yang diperoleh. Oleh karena itu, hasil penelitian dibatasi pada karakteristik dataset yang dikumpulkan dan tidak dimaksudkan untuk menggeneralisasi opini seluruh populasi Indonesia. Distribusi dataset berdasarkan platform ditunjukkan pada Tabel 1.

Tabel 1 Distribusi dataset berdasarkan platform

Platform	Jumlah Data	Persentase
Twitter/X	10.420	34.11%
TikTok	9.865	32.29%
YouTube	10.265	33.60%
Total	30.550	100%

Distribusi ini menunjukkan bahwa dataset yang digunakan cukup seimbang antar platform, sehingga memungkinkan model untuk mempelajari pola opini publik dari berbagai karakteristik media sosial. Twitter/X cenderung menyajikan opini singkat dan spontan, TikTok berisi komentar informal yang dipengaruhi tren digital, sedangkan YouTube mengandung diskusi yang lebih panjang dan argumentatif.

Setelah proses pelabelan manual oleh dua annotator independen, dataset diklasifikasikan ke dalam tiga kategori sentimen: positif, negatif, dan netral. Tingkat kesepakatan antar annotator diukur menggunakan Cohen's Kappa Score dan menghasilkan nilai 0.82, yang menunjukkan tingkat konsistensi pelabelan yang sangat baik. Distribusi label sentimen ditampilkan pada Tabel 2.

Tabel 2 Distribusi label sentimen

Label	Jumlah	Persentase
Positif	11.980	39.21%
Netral	8.745	28.63%
Negatif	9.825	32.16%
Total	30.550	100%

Distribusi label menunjukkan bahwa opini publik terhadap Artificial Intelligence cenderung didominasi oleh sentimen positif, meskipun sentimen negatif juga cukup signifikan. Hal ini mengindikasikan bahwa AI dipersepsikan sebagai teknologi yang menjanjikan, tetapi tetap memunculkan kekhawatiran tertentu di masyarakat.

4.2 Hasil Preprocessing Data

Karena data berasal dari media sosial, preprocessing menjadi tahapan krusial untuk meningkatkan kualitas data sebelum digunakan dalam proses pembelajaran model. Tahapan preprocessing meliputi case folding, penghapusan URL, mention, hashtag, tanda baca, tokenisasi, stopword removal, stemming, slang normalization, duplicate removal, dan filtering Bahasa Indonesia. Ringkasan hasil preprocessing ditampilkan pada Tabel 3.

Tabel 3 Hasil preprocessing	
Tahapan	Jumlah Data
Data awal	35.842
Penghapusan duplikasi	33.415
Filtering Bahasa Indonesia	31.880
Noise cleaning	30.550

Hasil preprocessing menunjukkan bahwa sekitar 14.76% data dieliminasi karena duplikasi, noise, atau tidak relevan. Proses ini penting untuk meningkatkan kualitas representasi fitur dan mengurangi bias selama pelatihan model.

4.3 Analisis Representasi Word Embedding

Penelitian ini menggunakan dua pendekatan word embedding, yaitu Word2Vec dan FastText, untuk membangun representasi semantik teks. Word2Vec menghasilkan representasi kata berdasarkan konteks kemunculannya dalam korpus, sedangkan FastText menambahkan informasi subword sehingga lebih adaptif terhadap kata tidak baku.

Hasil analisis menunjukkan bahwa Word2Vec cukup efektif dalam menangkap hubungan semantik antar kata. Sebagai contoh, kata *AI* memiliki kedekatan semantik dengan *machine learning*, *chatbot*, *otomasi*, dan *teknologi*. Namun, pada kata informal seperti *chatgptnya*, *gpt banget*, atau *chatgpt-an*, Word2Vec mengalami keterbatasan karena kata-kata tersebut tidak selalu terdapat dalam vocabulary model.

Sebaliknya, FastText menunjukkan performa yang lebih baik dalam menangani variasi kata informal, typo, dan slang, karena menggunakan representasi subword. Hal ini menjadikan FastText lebih sesuai untuk data media sosial Indonesia yang kaya dengan bentuk bahasa tidak baku.

4.4 Hasil Perbandingan Model

Eksperimen dilakukan dengan membandingkan model yang diusulkan dengan beberapa model baseline, yaitu CRF, LSTM, dan BiLSTM. Evaluasi dilakukan menggunakan accuracy, precision, recall, dan F1-score. Hasil perbandingan dapat dilihat pada Tabel 4.

Tabel 4 Hasil perbandingan performa model				
Model	Accuracy	Precision	Recall	F1-Score
CRF	78.42%	77.81%	76.95%	77.38%
LSTM	84.15%	83.67%	82.91%	83.29%
BiLSTM	87.24%	86.80%	86.15%	86.47%
BiLSTM-CRF	+	91.38%	90.94%	90.72%
Word2Vec				
BiLSTM-CRF + FastText	93.12%	92.74%	92.31%	92.52%

Hasil eksperimen menunjukkan bahwa model yang diusulkan BiLSTM-CRF + FastText memberikan performa terbaik pada seluruh metrik evaluasi. Dibanding model baseline CRF,

peningkatan accuracy mencapai 14.70%, sedangkan dibanding BiLSTM murni terjadi peningkatan sebesar 5.88%. Peningkatan ini menunjukkan bahwa integrasi sequence modeling menggunakan CRF dengan representasi subword dari FastText memberikan kontribusi signifikan terhadap kualitas klasifikasi sentimen.

Tabel 5 Hasil stratified 5-fold cross validation model BiLSTM–CRF + FASTTEXT

Fold	Accuracy	Precision	Recall	F1-Score
Fold 1	93.08%	92.69%	92.25%	92.47%
Fold 2	93.21%	92.83%	92.39%	92.61%
Fold 3	92.97%	92.58%	92.17%	92.37%
Fold 4	93.18%	92.81%	92.36%	92.58%
Fold 5	93.16%	92.79%	92.38%	92.58%
Mean	93.12%	92.74%	92.31%	92.52%
SD	0.10%	0.10%	0.10%	0.10%
Mean ± SD	93.12 ± 0.10%	92.74 ± 0.10%	92.31 ± 0.10%	92.52 ± 0.10%

Pada Tabel 5, hasil Stratified 5-Fold Cross Validation menunjukkan bahwa model BiLSTM–CRF dengan FastText menghasilkan performa yang relatif konsisten pada seluruh fold. Rata-rata accuracy yang diperoleh adalah $93,12 \pm 0,10\%$, sedangkan precision, recall, dan F1-score masing-masing mencapai $92,74 \pm 0,10\%$, $92,31 \pm 0,10\%$, dan $92,52 \pm 0,10\%$. Nilai standar deviasi yang rendah menunjukkan bahwa variasi performa antar-fold relatif kecil. Dengan demikian, performa model tidak bergantung secara kuat pada satu pembagian data tertentu dan menunjukkan kestabilan yang baik terhadap variasi data pengujian.

Pada accuracy, misalnya, hasil antar-fold hanya berada pada rentang 92,97%–93,21%. Pola serupa terlihat pada precision, recall, dan F1-score. Hal ini memperkuat hasil evaluasi bahwa BiLSTM–CRF + FastText tidak hanya menghasilkan nilai performa yang tinggi, tetapi juga menunjukkan hasil yang konsisten pada lima skenario pembagian data.

4.5 Analisis Confusion Matrix

Confusion matrix untuk model terbaik (BiLSTM–CRF + FastText) ditampilkan pada Tabel 6.

Tabel 6 Confusion matrix model yang diusulkan

Actual / Predicted	Positif	Netral	Negatif
Positif	2265	98	72
Netral	121	1598	94
Negatif	84	112	1666

Berdasarkan confusion matrix, model BiLSTM–CRF + FastText menunjukkan kemampuan klasifikasi yang baik pada ketiga kelas sentimen, tetapi masih terdapat beberapa pola kesalahan yang perlu diperhatikan. Pada kelas positif, sebanyak 2.265 komentar berhasil diklasifikasikan dengan benar, sedangkan 98 komentar diprediksi sebagai netral dan 72 komentar sebagai negatif. Kesalahan positif–netral menunjukkan bahwa beberapa komentar mengandung ekspresi positif yang lemah atau tidak eksplisit sehingga sulit dibedakan dari pernyataan informatif.

Pada kelas netral, sebanyak 1.598 komentar berhasil diklasifikasikan dengan benar, sementara 121 komentar diprediksi sebagai positif dan 94 komentar sebagai negatif. Kesalahan pada kelas netral menunjukkan bahwa kelas ini merupakan salah satu kelas yang relatif sulit dibedakan karena komentar netral dapat mengandung kata-kata bernuansa evaluatif tanpa secara eksplisit menunjukkan sikap positif atau negatif.

Pada kelas negatif, sebanyak 1.666 komentar berhasil diklasifikasikan dengan benar, sedangkan 84 komentar diprediksi sebagai positif dan 112 komentar sebagai netral. Kesalahan negatif–netral dapat terjadi pada komentar yang menyampaikan kritik secara implisit, menggunakan ironi, sarkasme,

negasi, atau ekspresi kontekstual yang tidak dapat sepenuhnya direpresentasikan melalui informasi leksikal.

Secara keseluruhan, pola kesalahan menunjukkan bahwa tantangan utama model bukan hanya pada representasi kata informal, tetapi juga pada ambiguitas semantik, intensitas sentimen yang rendah, negasi, serta ekspresi implisit. Meskipun FastText mampu mengurangi masalah *out-of-vocabulary*, typo, dan variasi morfologis melalui representasi subword, kemampuan tersebut belum sepenuhnya menyelesaikan masalah pemahaman konteks pragmatis seperti sarkasme dan ironi. Temuan ini menjadi salah satu arah pengembangan penelitian selanjutnya melalui penggunaan contextual embedding atau model transformer Bahasa Indonesia.

4.6 Pembahasan

Hasil penelitian menunjukkan bahwa model hybrid BiLSTM–CRF secara konsisten mengungguli seluruh model baseline. Model CRF menghasilkan performa terendah karena keterbatasannya dalam feature representation. CRF hanya efektif dalam sequence dependency tetapi tidak memiliki kemampuan pembelajaran representasi fitur yang mendalam. LSTM menunjukkan peningkatan performa karena kemampuannya menangkap dependensi sekuensial, namun masih terbatas pada pemrosesan satu arah. BiLSTM memberikan hasil yang lebih baik karena memahami konteks dua arah, yang sangat penting dalam interpretasi sentimen teks. Peningkatan paling signifikan diperoleh melalui integrasi CRF pada BiLSTM. CRF membantu menjaga konsistensi hubungan antar label prediksi sehingga mengurangi kesalahan klasifikasi yang disebabkan oleh struktur kalimat tidak teratur. FastText menghasilkan performa terbaik karena kemampuannya menangani kata tidak baku, typo, dan slang yang umum ditemukan pada media sosial Indonesia. Karakteristik ini sangat penting karena komentar media sosial jarang menggunakan bahasa formal. Representasi subword memungkinkan model memahami variasi kata yang tidak dikenal oleh Word2Vec. Penggunaan dataset multi-platform meningkatkan keberagaman karakteristik linguistik dan pola komunikasi yang dipelajari model. Namun, kemampuan generalisasi lintas platform secara khusus masih memerlukan pengujian lebih lanjut melalui skenario *cross-platform validation*.

Tabel 7 Perbandingan dengan penelitian terdahulu

Penelitian	Metode	Accuracy
Khan et al.	CNN-LSTM	85.20%
Putri	BERT/RoBERTa	90.10%
Jasmir et al.	Word2Vec + ML	87.00%
Penelitian ini	BiLSTM-CRF + FastText	93.12%

Hasil pada Tabel 7 menunjukkan bahwa model BiLSTM–CRF dengan FastText yang diusulkan memperoleh akurasi tertinggi sebesar 93,12%, melampaui penelitian Khan et al. (85,20%), Putri (90,10%), dan Jasmir et al. (87,00%). Peningkatan performa ini menunjukkan bahwa kombinasi FastText dan BiLSTM–CRF mampu menangkap hubungan semantik serta konteks dua arah secara lebih efektif, sekaligus mempertahankan konsistensi prediksi label. Selain itu, penggunaan dataset multi-platform dari Twitter/X, TikTok, dan YouTube memberikan keragaman data yang lebih tinggi sehingga meningkatkan kemampuan generalisasi model. Hasil tersebut menunjukkan bahwa pendekatan yang diusulkan efektif untuk analisis sentimen publik terhadap Artificial Intelligence pada media sosial berbahasa Indonesia. Hasil ini menunjukkan bahwa model yang diusulkan memberikan performa kompetitif bahkan melampaui beberapa pendekatan terdahulu. Meskipun transformer memiliki performa tinggi, model yang diusulkan menawarkan keseimbangan antara akurasi dan efisiensi komputasi.

5 Kesimpulan

Penelitian ini berhasil mengembangkan model analisis sentimen publik terhadap Artificial Intelligence (AI) berbasis hybrid BiLSTM–CRF dengan integrasi Word2Vec dan FastText menggunakan dataset multi-platform yang berasal dari Twitter/X, TikTok, dan YouTube. Pendekatan yang diusulkan dirancang untuk mengatasi tantangan analisis sentimen pada teks media sosial

berbahasa Indonesia yang memiliki karakteristik informal, tidak terstruktur, kaya slang, typo, dan variasi morfologis. Hasil eksperimen menunjukkan bahwa model hybrid yang diusulkan secara konsisten memberikan performa yang lebih baik dibanding model baseline, yaitu CRF, LSTM, dan BiLSTM. Berdasarkan hasil evaluasi, model BiLSTM–CRF dengan FastText menghasilkan performa terbaik dengan accuracy sebesar 93.12%, precision 92.74%, recall 92.31%, dan F1-score 92.52%, melampaui model BiLSTM–CRF dengan Word2Vec yang memperoleh accuracy 91.38%, serta model baseline BiLSTM (87.24%), LSTM (84.15%), dan CRF (78.42%). Peningkatan performa ini menunjukkan bahwa kombinasi sequence modeling menggunakan CRF dengan kemampuan contextual learning dari BiLSTM serta representasi subword dari FastText efektif dalam meningkatkan kualitas klasifikasi sentimen pada data media sosial Indonesia. Kontribusi utama penelitian ini terletak pada penerapan hybrid BiLSTM–CRF untuk analisis sentimen Bahasa Indonesia, pemanfaatan dual word embedding (Word2Vec dan FastText), serta penggunaan dataset multi-platform yang memberikan representasi opini publik yang lebih komprehensif dibanding pendekatan single-platform. Secara praktis, model yang dihasilkan berpotensi dimanfaatkan sebagai alat bantu pemantauan persepsi publik terhadap perkembangan Artificial Intelligence oleh pemerintah, industri, akademisi, maupun pemangku kebijakan dalam mendukung implementasi teknologi AI yang lebih adaptif, inklusif, dan bertanggung jawab. Meskipun demikian, penelitian ini masih memiliki keterbatasan karena hanya memanfaatkan data teks tanpa mempertimbangkan informasi multimodal dari platform berbasis video seperti TikTok dan YouTube. Oleh karena itu, penelitian selanjutnya dapat mengembangkan pendekatan multimodal sentiment analysis, membandingkan performa dengan model transformer seperti IndoBERT, serta mengintegrasikan teknik Explainable Artificial Intelligence (XAI) untuk meningkatkan interpretabilitas model.

Ucapan Terima Kasih

Kami mengucapkan terima kasih kepada Kementerian Pendidikan Tinggi Sain dan teknologi, atas bantuan Hibah penelitian skema Pemula dengan Nomor Kontrak Induk: 289/C3/DT.05.00/PL-BARU/2026 dan Nomor Kontrak Turunan : 039/DST/LL10/AL.04.03/PL-BARU/2026,033/MOU-R/LPPM/UNAMA/VI/2026

Referensi

- [1] J. Philippe and D. Thomas, “Artificial Intelligence and Work : A Critical Review of Recent Research from the Social Sciences,” *AI Soc.*, Vol. 39, No. 2, pp. 675–691, 2024, DOI: 10.1007/s00146-022-01496-x.
- [2] M. Rigou, “From E-Government to AI E-Government : A Systematic Review of Citizen Attitudes,” *Informatics*, 2025.
- [3] J. Garzón and E. Patiño, “Systematic Review of Artificial Intelligence in Education : Trends , Benefits , and Challenges,” *Multimodal Technol. Interact.*, pp. 1–19, 2025.
- [4] Z. Chen, “Ethics and Discrimination in Artificial Intelligence- Enabled Recruitment Practices,” *Humanit. Soc. SCI. Commun.*, No. 2019, 2023, DOI: 10.1057/s41599-023-02079-x.
- [5] B. Setiawan, “A Review of Sentiment Analysis Applications in Indonesia Between 2023-2024,” *J. Inf. Eng. Educ. Technol.*, Vol. 08, pp. 71–83, 2024.
- [6] F. H. Fitroh, “Systematic Literature Review : Analisis Sentimen berbasis Deep Learning,” *J. Nas. Teknol. dan Sist. Inf.*, Vol. 02, pp. 132–140, 2023.
- [7] L. Khan, A. Amjad, K. M. Afaq, and H. Chang, “Deep Sentiment Analysis using CNN-LSTM Architecture of English and Roman Urdu Text Shared in Social Media,” *Appl. SCI.*, 2022.
- [8] S. E. Sorour, A. Alojail, A. El-shora, and A. E. Amin, “A Hybrid Deep Learning Approach for Enhanced Sentiment Classification and Consistency Analysis in Customer Reviews,” *Mathematics*, pp. 1–38, 2024.
- [9] W. Khan, A. Daud, K. Shahzad, T. Amjad, and A. Banjar, “Named Entity Recognition using Conditional Random Fields,” *Appl. SCI.*, pp. 1–18, 2022.
- [10] Y. Qiu, L. Dong, W. Zhang, H. Xing, and J. Huang, “A Diffusion Enhanced CRF and BiLSTM Framework for Accurate Entity Recognition,” *SCI. Rep.*, pp. 1–25, 2025.
- [11] K. M. Suryaningrum, “Comparison of the TF-IDF Method with the Count Vectorizer to

<http://sistemasi.ftik.unisi.ac.id>

- Classify Hate Speech,” Eng. Math. Comput. Sci. J.*, Vol. 5, No. 2, pp. 79–83, 2023, DOI: 10.21512/emacsjournal.v5i2.9978.
- [12] Siti Khomsah, Rima Dias Ramadhani, and Sena Wijaya, “*The Accuracy Comparison between Word2Vec and FastText on Sentiment Analysis of Hotel Reviews*,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, Vol. 6, No. 3, pp. 352–358, 2022, DOI: 10.29207/resti.v6i3.3711.
- [13] J. Jasmir, E. Rohaini, M. R. Pahlevi, D. Sintong, and P. Simanjuntak, “*Word Embedding Feature for Improvement Machine Learning Performance in Sentiment Analysis Disney Plus Hotstar Comments*,” *J. Ilm. Tek. Elektro Komput. dan Inform.*, Vol. 10, No. 2, pp. 290–301, 2024, DOI: 10.26555/jiteki.v10i2.28799.
- [14] M. R. Haziq, Y. Sibaroni, and S. S. Prasetyowati, “*Word Embedding Optimization in Sentiment Analysis of Reviews on Mytelkomsel App using Long Short-Term Memory and Synthetic Minority Over-Sampling Technique*,” *J. Tek. Inform.*, Vol. 5, No. 6, pp. 1581–1589, 2024.
- [15] A. Nisrina Akbar Rizky Putri, “*Analisis Sentimen terhadap Kemajuan Kecerdasan Buatan di Indonesia menggunakan BERT dan RoBERTa*,” *J. Sains dan Inform.*, Vol. 9, No. November, pp. 136–145, 2023, DOI: 10.34128/jsi.v9i2.649.
- [16] T. Mikolov, G. Corrado, K. Chen, and J. Dean, “*Efficient Estimation of Word Representations in Vector Space*,” *ArVix*, pp. 1–12, 2013.
- [17] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “*Enriching Word Vectors with Subword Information*,” *Assoc. Comput. Linguist.*, Vol. 5, pp. 135–146, 2017.