ANALISIS SENTIMEN TWEET POLITIK PEMILU INDONESIA MENGGUNAKAN RoBERTa-BiLSTM

TESIS

Oleh :
DAFID MIZTA CHULLOH
NIM. 240605210006

PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026

MOTTO

"Ya Tuhanku, tambahkanlah kepadaku ilmu pengetahuan." — QS. Taha [20]: 114.

"Barang siapa keluar untuk menuntut ilmu, maka ia berada di jalan Allah hingga ia kembali." — HR. At-Tirmidzi No. 2647.

“Tiada kekayaan yang lebih utama daripada akal. Tiada keadaan yang lebih menyedihkan daripada kebodohan. Dan tiada warisan yang lebih baik daripada pendidikan.” — Ali Bin Abi Thalib R.A.

"Ilmu itu bukan untuk berdebat atau menunjukkan kehebatan, tapi untuk memperbaiki diri, dan memberi manfaat kepada orang lain." — Syekh Nawawi Al-Bantani.

“Kami mungkin tidak bisa mewariskan uang, tanah, atau harta benda yang berlimpah. Namun, kami akan mewariskan sesuatu yang jauh lebih berharga: ilmu, pendidikan, dan kemampuan untuk meraih masa depanmu sendiri.” — Ibu/Bapak (Kasan Pariyo/Sudariyah).

“Memang, kuliah hingga S2 bukanlah jaminan kesuksesan. Namun, aku percaya bahwa setiap usaha yang sungguh-sungguh tidak akan sia-sia. Pendidikan yang sedang aku perjuangkan ini merupakan bagian dari ikhtiarku untuk mempersiapkan masa depan. Aku sadar bahwa yang bisa aku kendalikan adalah usaha, proses, dan kesungguhanku; sedangkan hasil akhirnya bukan berada dalam kendaliku, melainkan menjadi ketetapan Allah. Karena itu, tugasku adalah terus berusaha sebaik mungkin, sementara hasilnya aku serahkan kepada Allah.” — Dafid Mizta Chulloh

HALAMAN PERSEMBAHAN

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dengan mengucapkan syukur yang tak terhingga ke hadirat Allah SWT., atas setiap rahmat, kasih sayang, dan pertolongan-Nya yang senantiasa mengiringi setiap langkah, karya sederhana ini penulis persembahkan dengan penuh hormat dan ketulusan kepada:

1. Bapak Kasan Pariyo dan Ibu Sudariyah

   Dua insan luar biasa yang menjadi rumah pertama bagi setiap doa, harapan, dan perjuangan penulis. Terima kasih atas kasih sayang yang tidak pernah mengenal batas, kesabaran yang tidak pernah habis, serta pengorbanan yang mungkin tidak akan pernah mampu terbalaskan. Setiap pencapaian yang penulis raih hari ini adalah buah dari doa-doa yang Bapak dan Ibu panjatkan dalam setiap sujud. Semoga karya ini menjadi seberkas kebahagiaan dan bukti bahwa segala pengorbanan yang telah diberikan tidak pernah sia-sia.

2. Keluarga Besar Tercinta

   Terima kasih atas cinta, perhatian, dan doa yang selalu menguatkan. Kehadiran keluarga menjadi tempat penulis kembali menemukan semangat ketika lelah, serta menjadi pengingat bahwa setiap perjuangan akan selalu memiliki alasan untuk diselesaikan.

3. Untuk diri saya sendiri, Dafid Mizta Chulloh

   Terima kasih telah bertahan hingga sejauh ini. Terima kasih karena memilih untuk tetap berjalan ketika langkah terasa berat, memilih bangkit setiap kali harapan hampir padam, dan memilih percaya kepada Allah SWT. saat segala sesuatu terasa begitu sulit untuk dipahami. Tidak semua perjuangan terlihat oleh mata manusia, tetapi Allah mengetahui setiap sebuah usaha, setiap doa yang dipanjatkan dalam diam, dan setiap ikhtiar yang dilakukan dengan penuh kesungguhan. Hari ini bukanlah akhir dari sebuah perjalanan, melainkan awal dari langkah-langkah baru yang akan terus diiringi harapan. Semoga diri ini senantiasa diberi hati yang rendah, langkah yang teguh, serta keberanian untuk terus menggapai mimpi dengan ridha Allah SWT.

Akhirnya, semoga karya ini tidak hanya menjadi syarat akademik untuk meraih gelar, tetapi juga menjadi amal baik yang membawa manfaat bagi banyak orang serta menjadi pengingat bahwa setiap perjalanan yang dilandasi kesabaran, doa, dan keikhlasan akan selalu menemukan akhir yang indah sesuai dengan ketetapan-Nya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur kehadirat Allah SWT atas segala limpahan rahmat, nikmat, hidayah, serta kasih sayang-Nya yang senantiasa menyertai setiap langkah penulis. Berkat izin dan pertolongan-Nya, penulis dapat menyelesaikan pendidikan pada Program Studi Magister Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang, sekaligus menyelesaikan tesis ini sebagai salah satu syarat untuk memperoleh gelar Magister Komputer.

Shalawat serta salam semoga senantiasa tercurah kepada junjungan Nabi Muhammad SAW, teladan terbaik sepanjang masa yang telah membawa umat manusia dari zaman penuh kebodohan menuju kehidupan yang diterangi oleh iman, ilmu pengetahuan, dan akhlak yang mulia.

Penyusunan tesis ini merupakan proses yang penuh dengan pembelajaran, tantangan, serta pengalaman berharga. Penulis menyadari bahwa keberhasilan dalam menyelesaikan karya ilmiah ini tidak terlepas dari doa, dukungan, bimbingan, motivasi, serta bantuan dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat dan ketulusan, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. H. Agus Mulyono, S.Pd., M.Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

3. Prof. Dr. Muhammad Faisal, M.T., selaku Ketua Program Studi Magister Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Prof. Dr. Muhammad Faisal, M.T., selaku Dosen Pembimbing I yang dengan penuh kesabaran telah meluangkan waktu, tenaga, dan pemikiran untuk memberikan bimbingan, arahan, masukan, serta motivasi kepada penulis selama proses penyusunan tesis ini.
5. Dr. Totok Chamidy, M.Kom., selaku Dosen Pembimbing II yang telah memberikan berbagai masukan, ketelitian dalam setiap proses evaluasi, serta saran-saran yang berharga sehingga kualitas tesis ini menjadi lebih baik.
6. Dr. Ir. Fresy Nugroho, S.T., M.T., IPM., ASEAN Eng., selaku Dosen Penguji I yang telah mencermati tesis ini dengan saksama serta memberikan berbagai koreksi, masukan, dan rekomendasi yang sangat bermanfaat bagi penyempurnaan penelitian ini.
7. Dr. Ir. Yunifa Miftachul Arif, S.ST., M.T., SMIEEE., selaku Dosen Penguji II yang telah memberikan tanggapan, evaluasi, dan berbagai masukan yang berarti sebagai penyempurna penyusunan tesis ini.
8. Seluruh rekan mahasiswa Program Studi Magister Informatika, sahabat, serta berbagai pihak yang tidak dapat penulis sebutkan satu per satu, atas kebersamaan, dukungan, bantuan, dan doa yang telah diberikan selama menempuh pendidikan hingga terselesaikannya tesis ini.

Penulis menyadari bahwa tesis ini masih memiliki berbagai keterbatasan. Oleh karena itu, penulis sangat mengharapkan kritik, saran, dan masukan yang

bersifat membangun sebagai bahan evaluasi demi penyempurnaan penelitian di masa yang akan datang.

Penulis berharap tesis ini dapat memberikan manfaat bagi pembaca, memperkaya khazanah keilmuan di bidang informatika, serta menjadi salah satu referensi bagi penelitian selanjutnya, khususnya yang berkaitan dengan analisis sentimen, pemrosesan bahasa alami (*Natural Language Processing*), dan penerapan model pembelajaran mendalam (*Deep Learning*) pada data media sosial.

Akhir kata, penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah memberikan doa, dukungan, bimbingan, bantuan, serta motivasi selama proses menempuh pendidikan hingga terselesaikannya tesis ini. Semoga segala kebaikan dan ketulusan yang telah diberikan memperoleh balasan yang berlipat ganda serta senantiasa mendapat keberkahan dari Allah SWT. Aamiin Ya Rabbal 'Alamin.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

DAFTAR ISI

HALAMAN PENGAJUAN ..... **i**
HALAMAN JUDUL ..... **ii**
HALAMAN PERSETUJUAN……..... **iii**
HALAMAN PENGESAHAN..... **iv**
PERNYATAAN KEASLIAN TULISAN..... **v**
MOTTO ..... **vi**
HALAMAN PERSEMBAHAN ..... **vii**
KATA PENGANTAR..... **ix**
DAFTAR ISI..... **xii**
DAFTAR GAMBAR..... **xv**
DAFTAR TABEL ..... **xvi**
Abstrak..... **xvii**
Abstract..... **xviii**
الملخص ..... **xix**
BAB I PENDAHULUAN..... **1**
1.1 Latar Belakang..... 1
1.2 Pernyataan Masalah..... 6
1.3 Tujuan Penelitian..... 6
1.4 Batasan Masalah..... 6
1.5 Manfaat Penelitian..... 7
BAB II STUDI PUSTAKA **8**
2.1 Analisis Tweet Politik Pemilu 8
2.2 Kuantifikasi Performa Multidimensional 22
2.3 Theoretical Framework..... 28
BAB III METODOLOGI **30**
3.1 Kerangka Konsep..... 30
3.2 Desain Sistem 32
3.3 Desain Penelitian 33
3.4 Pengumpulan Data..... 34
3.5 *Preprocessing* 36

3.5.1 *Text Normalization* 36
3.5.2 *Text Cleaning* 38
3.5.3 *Selective Punctuation Preservation* 40
3.5.4 *Deduplication* 42
3.5.5 *Text Length Filtering* 44
3.6 RoBERTa 47
3.6.1 Perhitungan Numerik Ekstraksi Fitur RoBERTa 47
3.7 BiLSTM 49
3.8 Evaluasi 51
3.8.1 Pengawasan Konvergensi via Learning Curves 51
3.8.2 Mekanisme Kontrol dan Retensi Model Terbaik 52
BAB IV HASIL DAN PEMBAHASAN 54
4.1 Data Penelitian 54
4.2 Hasil Pra-pemrosesan Data 54
4.3 Hasil Ekstraksi Fitur RoBERTa 56
4.4 Evaluasi Proses Pelatihan (*Learning Curves*) 59
4.5 Matriks Kebingungan (*Confusion Matrix*) 63
4.6 Performa Model RoBERTa-BiLSTM 67
4.7 Hasil Ekstraksi Fitur Word2Vec 69
4.8 Evaluasi Proses Pelatihan Word2Vec-BiLSTM (*Learning Curves*) 72
4.9 Matriks Kebingungan (*Confusion Matrix*) Word2Vec-BiLSTM 76
4.10 Performa Model Word2Vec-BiLSTM 80
4.11 Hasil Ekstraksi Fitur FastText 82
4.12 Evaluasi Proses Pelatihan FastText-BiLSTM (Learning Curves) 85
4.13 Matriks Kebingungan (*Confusion Matrix*) FastText-BiLSTM 90
4.14 Performa Model FastText-BiLSTM 93
4.15 Integrasi Nilai-Nilai Islam 98
4.15.1 Hubungan dengan Allah (Hablum Minallah) 99
4.15.2 Hubungan dengan Manusia (Hablum Minannas) 102
4.15.3 Hubungan dengan Lingkungan (Hablum Minal 'Alam) 105
BAB V PENUTUP 109

5.1 Kesimpulan 109
5.2 Saran 112
DAFTAR PUSTAKA 114

DAFTAR GAMBAR

Gambar 2.1 Theoretical Framework 29
Gambar 3.1 Kerangka Konsep 31
Gambar 3.2 Desain Sistem 32
Gambar 3.3 Desain Penelitian 34
Gambar 3.4 Proses Pengumpulan Data 35
Gambar 4.1 Kurva Training dan Validation Loss 61
Gambar 4.2 Raw Confusion Matrix Model RoBERTa 64
Gambar 4.3 Normalized Confusion Matrix Model RoBERTa 65
Gambar 4.4 Kurva Training dan Validation Loss 73
Gambar 4.5 Raw Confusion Matrix Model RoBERTa 77
Gambar 4.6 Normalized Confusion Matrix Model RoBERTa 78
Gambar 4.7 Kurva Training dan Validation Loss 87
Gambar 4.8 Raw Confusion Matrix Model RoBERTa 91
Gambar 4.9 Normalized Confusion Matrix Model FastText 92

DAFTAR TABEL

Tabel 3.1 Hasil *Text Normalization* 37
Tabel 3.2 Hasil *Text Cleaning* 40
Tabel 3.3 Hasil *selective punctuation preservation* 42
Tabel 3.4 Hasil *deduplication* 44
Tabel 3.5 Hasil *Text Length Filtering* 46
Tabel 3.6 Nilai *contextual embedding* Token pada Dimensi ke-1 Hasil Ekstraksi Fitur RoBERTa 48
Tabel 4.1 Perbandingan Data Sebelum dan Sesudah Pra-pemrosesan 55
Tabel 4.2 Perbandingan Data Sebelum dan Sesudah Ekstraksi Fitur RoBERTa . 58
Tabel 4.3 Hasil Evaluasi Model RoBERTa-BiLSTM 68
Tabel 4.4 Perbandingan Data Sebelum dan Sesudah Ekstraksi Fitur RoBERTa . 70
Tabel 4.5 Hasil Evaluasi Model Word2Vec-BiLSTM 82
Tabel 4.6 Perbandingan Data Sebelum dan Sesudah Ekstraksi Fitur FastText 83
Tabel 4.7 Hasil Evaluasi Model FastText-BiLSTM 94
Tabel 4.8 Hasil Evaluasi Keseluruhan Model Klasifikasi Sentimen 96

Abstrak

Dafid Mizta Chulloh 2026. **Analisis Sentimen Tweet Politik Pemilu Indonesia Menggunakan RoBERTa-BiLSTM.** Tesis. Program Study Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing (1) Prof. Dr. Muhammad Faisal, M.T., Pembimbing (II) Dr. Totok Chamidy, M.Kom.

Kata kunci: Analisis Sentimen, *Deep Learning*, Arsitektur Hibrida, RoBERTa-BiLSTM, Pemilu Indonesia, X.

Media sosial, khususnya Twitter, menjadi ruang penting dalam pembentukan dan pertukaran opini politik. Analisis sentimen dapat digunakan untuk memahami persepsi publik terhadap isu politik, termasuk Pemilu Indonesia. Namun, penggunaan bahasa yang informal, kompleks, dan ambigu serta dinamika politik menuntut model yang mampu menangkap konteks secara mendalam. Oleh karena itu, penelitian ini membandingkan kinerja beberapa metode deep learning dalam analisis sentimen data Twitter terkait politik. Data yang digunakan dalam penelitian ini merupakan data yang telah dikumpulkan dari media sosial X. Proses pengumpulan data pada X menggunakan tool x-harvest yang dijalankan melalui perintah npx berbasis Node.js. Tool ini melakukan harvesting data X dari tab LATEST berdasarkan kata kunci pencarian tertentu, dengan parameter bahasa yang diinginkan dan jumlah maksimal tweet yang diambil sesuai dengan kebutuhan. Proses pengumpulan data ini biasa disebut dengan scraping data. Metode yang digunakan adalah RoBERTa-BiLSTM serta sebagai pembanding Word2Vec-BiLSTM dan FastText-BiLSTM. Hasil penelitian menunjukkan bahwa penggunaan representasi bahasa berbasis konteks mampu meningkatkan performa klasifikasi dibandingkan pendekatan embedding. Model Word2Vec-BiLSTM memperoleh nilai Precision sebesar 0.5922, Recall sebesar 0.6926, F1-Score sebesar 0.6289, serta Accuracy sebesar 0.8178. Model FastText-BiLSTM memperoleh nilai Precision sebesar 0.9096, Recall sebesar 0.5628, F1-Score sebesar 0.6484, meskipun Accuracy model masih tergolong tinggi, yaitu sebesar 0.8893. Berbeda dengan kedua model sebelumnya, RoBERTa-BiLSTM menunjukkan performa yang paling optimal pada sebagian besar metrik evaluasi. Model ini berhasil memperoleh nilai Accuracy tertinggi sebesar 0.9255, disertai nilai Recall sebesar 0.8261 dan F1-Score sebesar 0.8248, yang juga merupakan capaian terbaik dibandingkan model lainnya. Selain itu, nilai Precision sebesar 0.8279 menunjukkan bahwa model memiliki tingkat keseimbangan performa yang sangat baik dalam proses klasifikasi sentimen. Hasil tersebut mengindikasikan bahwa integrasi RoBERTa dengan BiLSTM mampu menghasilkan representasi fitur yang lebih kaya dan kontekstual untuk memahami konteks tweet politik Indonesia serta meningkatkan kemampuan klasifikasi sentimen.

Abstract

Dafid Mizta Chulloh 2026. **Sentiment Analysis of Indonesian Election-Related Political Tweets Using RoBERTa-BiLSTM.** Thesis. Master of Informatics Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Supervisors (1) Prof. Dr. Muhammad Faisal, M.T., Supervisors (II) Dr. Totok Chamidy, M.Kom.

Kata kunci: Sentiment Analysis, Deep Learning, Hybrid Architecture, RoBERTa-BiLSTM, Indonesian Election, X.

Social media, particularly Twitter, has become an important space for the formation and exchange of political opinions. Sentiment analysis can be used to understand public perceptions of political issues, including the Indonesian general election. However, the use of informal, complex, and ambiguous language, along with political dynamics, requires models capable of capturing context in depth. Therefore, this study compares the performance of several deep learning methods for sentiment analysis of political data from Twitter. The data used in this study were collected from the social media platform X. The data collection process on X used the *x-harvest* tool, which was executed through a Node.js-based *npx* command. This tool harvests data from X's "LATEST" tab based on specific search keywords, with parameters for the desired language and the maximum number of tweets to be retrieved according to the research requirements. This data collection process is commonly referred to as data scraping. The method used is RoBERTa-BiLSTM, with Word2Vec-BiLSTM and FastText-BiLSTM serving as comparison models. The results show that the use of context-based language representations can improve classification performance compared with embedding approaches. The Word2Vec-BiLSTM model achieved a Precision of 0.5922, a Recall of 0.6926, an F1-Score of 0.6289, and an Accuracy of 0.8178. The FastText-BiLSTM model achieved a Precision of 0.9096, a Recall of 0.5628, and an F1-Score of 0.6484, although the model's Accuracy remained relatively high at 0.8893. In contrast to the two previous models, RoBERTa-BiLSTM demonstrated the most optimal performance across most evaluation metrics. This model achieved the highest Accuracy of 0.9255, along with a Recall of 0.8261 and an F1-Score of 0.8248, which were also the best results compared with the other models. In addition, the Precision of 0.8279 indicates that the model has a very good level of performance balance in the sentiment classification process. These results indicate that integrating RoBERTa with BiLSTM can produce richer and more contextual feature representations for understanding the context of Indonesian political tweets and improving sentiment classification capabilities.

الملخص

دآفِد مِيزْتَا جُولُوه، 2026. **تحليل المشاعر للغريدات السياسية في الانتخابات الإندونيسية باستخدام RoBERTa-BiLSTM**. أطروحة الماجستير. قسم ماجستير الحاسوبيات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية بمالانج. المشرف الأول: أ.د. محمد فيصل، الماجستير في التكنولوجيا؛ المشرف الثاني: د. طوطوق تشاميدي، الماجستير في الحاسوب.

الكلمات المفتاحية: الانتخابات، RoBERTa-BiLSTM، تحليل المشاعر، التعلّم العميق، البنية الهجينة الإندونيسية، منصة X.

تعد وسائل التواصل الاجتماعي، ولا سيما تويتر، فضاء مهما في تشكيل الآراء السياسية وتبادلها. ويمكن استخدام تحليل المشاعر لفهم تصورات الجمهور تجاه القضايا السياسية، بما في ذلك الانتخابات الإندونيسية. ومع ذلك، فإن استخدام اللغة غير الرسمية والمعقدة والملتبسة، إلى جانب ديناميكيات السياسة، يتطلب نموذجا قادرا على التقاط السياق بصورة عميقة. ولذلك، تقارن هذه الدراسة أداء عدد من أساليب التعلم العميق في تحليل مشاعر بيانات تويتر المتعلقة بالسياسة. وقد استخدمت في هذه الدراسة بيانات جمعت من وسائل التواصل الاجتماعي X. وتمت عملية جمع البيانات من منصة X باستخدام أداة x-harvest التي تشغل من استنادا LATEST من تبويب X وتقوم هذه الأداة بحصاد بيانات Node.js. القائم على npx خلال الأمر إلى كلمات مفتاحية محددة للبحث، مع تحديد معلمة اللغة المطلوبة والحد الأقصى لعدد التغريدات التي يتم جمعها وفقا للاحتياجات. وتعرف عملية جمع البيانات هذه عادة باسم كشط البيانات (Data Scraping). والطريقة المستخدمة هي RoBERTa-BiLSTM، مع استخدام Word2Vec-BiLSTM وFastText-BiLSTM كنماذج للمقارنة. وأظهرت نتائج الدراسة أن استخدام تمثيل لغوي قائم على السياق قادر على تحسين أداء التصنيف مقارنة بنهج التضمين (Embedding). وقد حقق نموذج BiLSTM Word2Vec-BiLSTM بلغت 0.6289، بالإضافة إلى F1-Score بلغت 0.6926، و Recall بلغت 0.5922، و Precision قيمة بلغت Precision قيمة FastText-BiLSTM بلغت 0.8178. في حين حقق نموذج Accuracy Accuracy بلغت 0.6484، على الرغم من أن قيمة F1-Score بلغت 0.5628، و Recall و 0.9096، RoBERTa-BiLSTM للنموذج ظلت مرتفعة، حيث بلغت 0.8893. وعلى خلاف النموذجين السابقين، أظهر نموذج BiLSTM أداء أمثل في معظم مقاييس التقييم. فقد حقق هذا النموذج أعلى قيمة Accuracy 0.9255، بلغت 0.8248، وهما أيضا أفضل النتائج مقارنة F1-Score بلغت 0.8261 وقيمة Recall إلى جانب قيمة البالغة 0.8279 أن النموذج يتمتع بتوازن Precision بالنماذج الأخرى. بالإضافة إلى ذلك، أظهرت قيمة BiLSTM مع RoBERTa جيد جدا في الأداء أثناء عملية تصنيف المشاعر. وتشير هذه النتائج إلى أن دمج قادر على إنتاج تمثيل للخصائص أكثر ثراء وسياقية لفهم سياق التغريدات السياسية الإندونيسية، فضلا عن تحسين قدرة تصنيف المشاعر.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Informasi merupakan bagian dari pengetahuan yang terus mengalami transformasi seiring dengan perkembangan teknologi. Jika dahulu informasi hanya tersedia dalam bentuk cetak, seiring dengan perkembangan teknologi informasi, penyajiannya telah beralih ke bentuk digital melalui berbagai media, termasuk media sosial(Tanoto et al., 2024). Di antara berbagai media sosial, Twitter memiliki karakteristik unik karena sifatnya yang real-time, terbuka, dan padat informasi. Hal ini membuat Twitter menjadi salah satu kanal paling aktif dalam menyuarakan pendapat, terutama dalam politik pemilu Indonesia(Firdaus et al., 2024). Fenomena ini melahirkan kebutuhan akan metode analisis otomatis, salah satunya melalui Natural Language Processing (NLP).

Natural Language Processing (NLP) adalah bidang dalam kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa manusia. Melalui NLP, sistem komputer dapat memproses, memahami, serta menafsirkan teks serupa dengan komunikasi manusia(Susanti & Khu, 2025). Teknologi ini tidak hanya memungkinkan analisis terhadap data dalam skala besar, tetapi juga membuka peluang untuk mengungkap pola dan makna yang tersembunyi di balik bahasa(Supriyadi & Makatita, 2025). Namun demikian, pengembangan NLP masih menghadapi tantangan kompleks, seperti keragaman struktur bahasa, ambiguitas makna, serta kebutuhan komputasi yang tinggi(Nathan et al., 2025).

Analisis sentimen merupakan cabang penerapan dari pemrosesan bahasa alami NLP, yang bertujuan untuk mengidentifikasi, menafsirkan, dan mengklasifikasikan opini atau kecenderungan subjektif yang terkandung dalam suatu teks(Wu et al., 2025). Dalam konteks media sosial seperti Twitter, analisis ini berperan penting untuk memahami persepsi dan sikap publik terhadap isu tertentu, termasuk politik pemilu Indonesia(Asokere et al., 2025). Analisis sentimen dapat dipahami sebagai suatu metode komputasional yang digunakan untuk mengekstraksi dan mengklasifikasikan kecenderungan subjektif dalam teks, sehingga opini publik dapat dipetakan secara sistematis ke dalam kategori positif, negatif, maupun netral(Martinez-Espana et al., 2024). Dalam beberapa kasus, suatu sentimen dapat mengandung kata-kata yang bersifat kontradiktif, sehingga kecenderungan maknanya dapat digolongkan sebagai positif maupun negatif. Sementara itu, apabila tidak ditemukan indikasi keberpihakan yang jelas, sentimen tersebut dikategorikan sebagai netral(Cicekyurt & Bakal, 2025).

Meskipun analisis sentimen telah banyak diterapkan pada data media sosial, tantangan linguistik yang ditandai dengan penggunaan bahasa kompleks, informal, dan ambigu serta konteks politik yang dinamis menuntut pendekatan pemodelan yang mampu menangkap representasi semantik secara lebih mendalam. Dengan mempertimbangkan pentingnya akurasi dalam analisis sentimen, penelitian terkini di bidang ini berupaya untuk mengeksplorasi dan mengevaluasi kinerja komparatif berbagai model deep learning dalam analisis sentimen terhadap data twitter, khususnya yang berkaitan dengan konteks politik. Untuk mencapai tujuan tersebut, penelitian ini membandingkan beberapa metode.

Tujuan komparasi ini menjadi sangat relevan mengingat bahwa media sosial, khususnya Twitter telah mengalami transformasi menjadi tempat utama pembentukan, diseminasi, dan pertukaran opini politik di ruang publik digital. Sejalan dengan upaya peningkatan akurasi tersebut, penelitian awal ini mengkombinasikan kekuatan deep learning, natural language processing (NLP), dan geographic information systems (GIS), penelitian ini melakukan klasifikasi dan sentimen dalam percakapan politik di Twitter, serta keterkaitannya dengan misinformasi, ekstremisme, dan hasil Pemilu Presiden Amerika Serikat 2020. Penelitian ini menunjukkan bahwa model Long Short-Term Memory (LSTM), sebagai varian unggulan dari Recurrent Neural Network (RNN), mampu mencapai akurasi sebesar 67% dalam analisis sentimen. Pencapaian ini diperoleh melalui pemanfaatan dataset berskala besar yang terdiri atas 1.349.373 tweet asli berbahasa Inggris, yang dikumpulkan secara real time dari Twitter selama periode April 2020 hingga Januari 2021 dengan kata kunci ‘Trump’, ‘Biden’, ‘Democrats’, dan ‘Republicans’(Hashemi, 2023).

Berdasarkan temuan awal tersebut, analisis kemudian diperluas dengan pendekatan hybrid Convolutional Neural Network (CNN) dan Recurrent Neural Network (RNN) untuk mengevaluasi performa masing-masing model. Penelitian ini diarahkan pada pemanfaatan arsitektur hybrid model yang secara konseptual menggabungkan keunggulan mekanisme ekstraksi fitur lokal dengan kemampuan pemodelan ketergantungan jangka panjang dalam data tekstual. Pendekatan ini dipilih karena memiliki kapasitas adaptasi yang tinggi terhadap karakteristik data berskala besar, kompleks, dan bersifat real-time, seperti konten berita daring

maupun percakapan di media sosial yang berhubungan dengan isu-isu publik. Untuk menguji kinerja model, penelitian ini memanfaatkan himpunan data publik yang berisi sekitar 45.000 artikel berita politik dan dunia. Hasil eksperimen menunjukkan bahwa Pendekatan hybrid Convolutional Neural Network (CNN) dan Recurrent Neural Network (RNN) mencapai akurasi 60%, yang menegaskan potensi pendekatan ini dalam menganalisis data teks berskala besar, kompleks, dan dinamis(Nasir et al., 2021).

Meskipun metodologi-metodologi sebelumnya menegaskan potensi *deep learning* dalam analisis data yang kompleks, tingkat akurasi yang relatif masih rendah menunjukkan adanya kesenjangan penelitian yang signifikan. Dengan menerapkan kerangka kerja komparatif, penelitian ini berupaya meningkatkan kinerja analisis sentimen terhadap cuitan terkait pemilu, dengan menargetkan akurasi yang jauh lebih tinggi melalui pemodelan kontekstual yang lebih canggih. Selain meningkatkan kinerja teknis, pendekatan ini juga diharapkan mampu memperkuat kemampuan model dalam menghasilkan analisis sentimen yang lebih tepat, objektif, dan akurat pada tweet politik pemilu. Prinsip ketepatan, objektivitas, dan akurasi dalam analisis tersebut sejalan dengan nilai-nilai keadilan dan kejujuran yang tercantum dalam Al-Qur'an, sebagaimana disebutkan dalam QS. An-Nisa' [4]:135:

يَا أَيُّهَا الَّذِينَ آمَنُوا كُونُوا قَوَّامِينَ بِالْقِسْطِ شُهَدَاءَ لِلَّهِ وَلَوْ عَلَىٰ أَنفُسِكُمْ أَوِ الْوَالِدَيْنِ وَالْأَقْرَبِينَ ۚ إِن يَكُنْ غَنِيًّا أَوْ فَقِيرًا فَاللَّهُ أَوْلَىٰ بِهِمَا ۖ فَلَا تَتَّبِعُوا الْهَوَىٰ أَن تَعْدِلُوا ۚ وَإِن تَلْوُوا أَوْ تُعْرِضُوا فَإِنَّ اللَّهَ كَانَ بِمَا تَعْمَلُونَ خَبِيرًا

Artinya:

"Wahai orang-orang yang beriman! Jadilah kamu penegak keadilan, menjadi saksi karena Allah, walaupun terhadap dirimu sendiri atau terhadap ibu bapak dan kaum kerabatmu. Jika dia (yang didakwa) kaya ataupun miskin, maka Allah lebih tahu kemaslahatan keduanya. Maka janganlah kamu mengikuti hawa nafsu, karena ingin menyimpang dari kebenaran. Dan jika kamu memutarbalikkan kata-kata atau enggan menjadi saksi, maka sesungguhnya Allah Maha Mengetahui apa yang kamu kerjakan."

Dalam konteks penelitian ilmiah, ketepatan dan objektivitas merupakan pondasi utama yang menjamin integritas hasil analisis. Nilai-nilai tersebut sejalan dengan ajaran Islam sebagaimana termaktub dalam QS. An-Nisa [4]:135, yang menurut tafsir Ibnu Katsir memerintahkan kaum beriman untuk menegakkan keadilan (qawwāmīn bil-qisṭ), Penegakan keadilan dalam konteks ini adalah kejujuran ilmiah (al – amānah) (al –' ilmiyyah) dan keberpihakan pada kebenaran data, Dalam konteks analisis sentimen berbasis kecerdasan buatan, prinsip-prinsip tersebut dapat dipandang sebagai landasan etis untuk membangun model yang tidak bias, mampu menafsirkan opini publik secara akurat, dan mengedepankan transparansi metodologis. Dengan demikian, penerapan model *deep learning* seperti RoBERTa-BiLSTM pada analisis sentimen tweet politik Pemilu Indonesia tidak hanya berorientasi pada peningkatan akurasi dan keandalan teknis, tetapi juga mencerminkan komitmen ilmiah terhadap nilai-nilai ketepatan dan objektivitas sebagaimana digariskan dalam ajaran Islam.

1.2 Pernyataan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, dapat disimpulkan bahwa terdapat sejumlah permasalahan mendasar yang perlu diteliti lebih lanjut yaitu.

1. Bagaimana penerapan model hybrid RoBERTa-BiLSTM dalam memprediksi sentimen tweet politik Pemilu di Indonesia?
2. Bagaimana performa model hybrid RoBERTa-BiLSTM dalam memprediksi sentimen tweet politik Pemilu di Indonesia?

1.3 Tujuan Penelitian

Sejalan dengan pernyataan masalah yang telah diidentifikasi, tujuan dari penelitian ini adalah untuk mencapai hal-hal sebagai berikut.

1. Mengetahui penerapan model hybrid RoBERTa-BiLSTM dalam memprediksi sentimen tweet politik Pemilu di Indonesia?
2. Mengetahui performa model hybrid RoBERTa-BiLSTM dalam memprediksi sentimen tweet politik Pemilu di Indonesia?

1.4 Batasan Masalah

Dalam rangka menjaga fokus dan konsistensi penelitian agar sesuai dengan tujuan yang telah ditetapkan, ditentukan sejumlah batasan. Batasan ini bertujuan agar analisis dapat dilakukan secara lebih terarah dan mendalam. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Penelitian ini hanya menggunakan data dari media sosial X yang mengandung kata kunci atau tagar bertema politik Indonesia pada rentang waktu tertentu, sehingga tidak mencakup periode waktu secara keseluruhan.

2. Data yang digunakan adalah data publik yang diperoleh dari harvest melalui X API dengan tetap memperhatikan etika dan privasi pengguna sesuai ketentuan platform.

1.5 Manfaat Penelitian

Berdasarkan ruang lingkup dan tujuan penelitian, manfaat yang dapat diambil dari studi ini meliputi aspek-aspek berikut:

1. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang pemrosesan bahasa alami (Natural Language Processing), khususnya dalam penerapan model hybrid yang menggabungkan RoBERTa dan BiLSTM untuk analisis sentimen serta penelitian ini juga dapat menjadi referensi ilmiah bagi penelitian-penelitian selanjutnya.

BAB II

STUDI PUSTAKA

2.1 Analisis Tweet Politik Pemilu

Pemrosesan Bahasa Alami atau *Natural Language Processing* (NLP) dalam konteks bahasa Indonesia menghadirkan kompleksitas unik yang melampaui hambatan sintaksis tradisional. Tantangan utama berakar pada karakteristik morfologis bahasa Indonesia yang kaya akan afiksasi, serta evolusi penggunaannya di ruang digital. Khususnya pada platform media sosial, ujaran sering kali menyimpang dari Pedoman Umum Ejaan Bahasa Indonesia (PUEBI). Ekstraksi makna menjadi sangat menantang akibat maraknya penggunaan bahasa tidak baku (slang), singkatan yang tidak standar, serta dialek regional yang bercampur baur. Dalam konteks wacana politik, hambatan ini semakin tereskalasi dengan tingginya frekuensi penggunaan sarkasme dan ironi. Sebuah model komputasi dituntut untuk tidak hanya memetakan struktur leksikal, tetapi juga harus mampu meraba intensi implisit di balik frasa-frasa satir yang sering kali secara harfiah bernada positif namun secara kontekstual bertujuan menyerang atau mengkritik entitas politik tertentu. Untuk menafsirkan dan mengurai polaritas emosi dari kompleksitas bahasa linguistik tersebut, disiplin NLP mengandalkan pendekatan spesifik yang dikenal sebagai analisis sentimen(Carrasco & Dias, 2024).

Sebagai turunan langsung dari Pemrosesan Bahasa Alami (NLP), analisis sentimen telah diadopsi secara luas sebagai alat komputasional untuk mengukur dan menafsirkan persepsi publik di ruang digital. Dikenal pula melalui terminologi opinion mining (Ravi & Ravi, 2015), metode ini bekerja dengan cara membedah

kecenderungan psikologis atau sikap masyarakat terhadap suatu isu, peristiwa, maupun individu. Proses sistematis ini pada umumnya digunakan untuk mengategorikan menjadi tiga kelompok dasar, yaitu positif, negatif, dan netral. Dinamika pemrosesan teks seringkali dihadapkan pada redaksi kalimat yang ambivalen atau memuat polaritas ganda; dalam kondisi ini, algoritma bertugas memilah dominasi keberpihakan sentimen tersebut. Sementara itu, teks yang berkarakteristik objektif, informatif, dan tidak mengindikasikan bias emosional secara otomatis akan direduksi ke dalam kategori netral.

Dalam konteks eskalasi politik seperti pemilihan umum, implementasi metode ini menjadi sangat esensial untuk memetakan konstruksi opini publik yang terbangun di platform Twitter. Medium ini memfasilitasi interaksi asimetris yang unik, dimana wacana politik dikonstruksi secara masif melalui agregasi tagar, interpelasi antar pengguna melalui fitur sebutan (mention), dan sering kali memicu polarisasi pandangan yang dikotomis. Konstelasi data tekstual yang diproduksi secara instan ini memungkinkan perekaman fluktuasi persepsi massa tanpa jeda waktu (real-time). Oleh karena itu, kemampuan komputasional dalam menyaring kebisingan data (noise) menjadi konklusi kecenderungan sentimen memberikan proyeksi yang sangat komprehensif bagi para akademisi dan praktisi politik guna merespons pergeseran iklim elektoral.

RoBERTa merupakan arsitektur Deep Learning berbasis Transformer yang dikembangkan oleh Facebook AI Research (FAIR) sebagai bentuk penyempurnaan dari model BERT (Bidirectional Encoder Representations from Transformers)(Liu et al., 2019). Dalam penelitian ini, model bahasa dieksekusi menggunakan varian

pra-latih cahya/roberta-base-indonesian-522M yang secara spesifik dirancang untuk memahami morfologi dan sintaksis bahasa Indonesia. Perbedaan fundamental RoBERTa dibandingkan pendahulunya terletak pada skema pelatihannya: RoBERTa menghilangkan tugas Next Sentence Prediction (NSP), menggunakan batch size yang jauh lebih masif, serta mengimplementasikan Dynamic Masking yang mengubah pola penyembunyian token (masking) pada setiap epoch, sehingga model memiliki kemampuan generalisasi semantik yang lebih tangguh terhadap teks noisy seperti opini politik di media sosial.

RoBERTa sepenuhnya dibangun di atas blok Encoder dari arsitektur Transformer. Komponen matematis utama yang menggerakkan kemampuan pemahaman konteksnya adalah mekanisme Scaled Dot-Product Attention. Mekanisme ini memproyeksikan representasi kata ke dalam tiga matriks berdimensi tinggi: *Query* (*Q*), *Key* (*K*), dan *Value* (*V*).

Perhitungan bobot atensi (*attention weights*) diformulasikan sebagai berikut:

$$\text{Attention}\ (Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \qquad (1)$$

Keterangan :

$\text{Attention}\ (Q, K, V)$ = (Q) Query (Pertanyaan), (K) Key (Kunci), dan (V) Value (Nilai). Proses untuk menentukan informasi mana yang paling relevan dan memberikan bobot pada informasi tersebut.

K^T = transpose matriks K, agar perkalian matriks QK^T menghasilkan skor hubungan antara kata-kata.

d_k = jumlah dimensi vektor Key.

$\sqrt{d_k}$ = faktor pembagi untuk menjaga skor agar tidak terlalu besar.

softmax = mengubah skor menjadi bobot yang jumlahnya 1 untuk setiap query.

Dalam persamaan tersebut, matriks Q dikalikan secara *dot-product* dengan transposisi dari matriks K untuk menghitung seberapa besar korelasi (afinitas) antara satu token dengan token lainnya dalam satu kalimat. Hasilnya dibagi dengan $\sqrt{d_k}$ (akar kuadrat dari dimensi matriks *Key*) untuk mencegah terjadinya ledakan gradien (*gradient explosion*) sebelum dimasukkan ke dalam fungsi aktivasi *Softmax*.

Untuk memungkinkan model memproses representasi semantik dari berbagai sub-ruang (subspaces) yang berbeda secara paralel misalnya satu head fokus pada subjek kalimat, sementara head lain fokus pada kata sifat atau sarkasme RoBERTa mengimplementasikan Multi-Head Attention:

$$\text{MultiHead} = (Q, K, V) = \text{Concat}(Head_1 \ldots, Head_2) W^O \quad (2)$$

Keterangan :

$\text{MultiHead} = (Q, K, V)$ = (Q) Query (Pertanyaan), (K) Key (Kunci), dan (V) Value (Nilai). Hasil akhir dari proses Multi-Head Attention setelah seluruh hasil attention head digabungkan dan dikalikan dengan matriks bobot W^O.

$Head_1 \ldots, Head_2$ = hasil dari beberapa mekanisme attention yang bekerja secara paralel.

h = jumlah attention head yang digunakan dalam proses Multi-Head Attention.

Concat = operasi penggabungan seluruh hasil attention head menjadi satu representasi.

W^O = matriks bobot yang digunakan untuk mengubah hasil penggabungan agar sesuai dengan dimensi keluaran yang diinginkan.

Di mana setiap $Head_1$ dihitung secara independen:

$$head_i = Attention\ (QW_i^Q, KW_i^K, VW_i^V) \quad (3)$$

Keterangan :

$head_i$ = hasil attention pada head ke-i

$Attention$ = proses untuk menentukan informasi yang relevan.

Q, K, V = Query (informasi yang dicari), Key (informasi untuk mencocokkan relevansi), dan Value (informasi yang diambil).

QW_i^Q = matriks bobot untuk Query.

KW_i^K = matriks bobot untuk Key.

VW_i^V = matriks bobot untuk Value.

i = indeks atau urutan head.

Matriks QW_i^Q, KW_i^K , VW_i^V dan W^O merupakan matriks bobot parameter yang diperbarui selama masa pra-pelatihan model.

Pada arsitektur sistem ini, RoBERTa tidak di-fine-tune secara menyeluruh, melainkan dibekukan (frozen) dan difungsikan sebagai Feature Extractor. Setiap tweet yang masuk akan dibatasi dan dipotong (truncated atau padded) menjadi panjang maksimum 128 token.RoBERTa memetakan input diskrit tersebut menjadi representasi ruang kontinu berdimensi tiga yang disebut last_hidden_state. Jika didefinisikan secara matematis, keluaran tensor dari RoBERTa memiliki bentuk spasial (B, L, H), di mana:

1. B = Batch size
2. L = Sequence Length (128 token)
3. H = Hidden dimension (768 fitur)

Representasi sekuensial berukuran 128×768 inilah yang menyimpan jejak memori setiap kata berdasarkan konteks kalimat utuhnya, yang selanjutnya disuntikkan sebagai input dasar bagi arsitektur BiLSTM.

Untuk kebutuhan kompresi representasi semantik menjadi vektor tunggal tingkat kalimat (seperti dalam visualisasi ruang vektor logis), mekanisme Mean Pooling diimplementasikan. Berbeda dengan pendekatan tradisional yang hanya mengambil token awalan [CLS], mean pooling menghitung rata-rata dari seluruh hidden state dengan secara eksplisit mengabaikan token padding melalui Attention Mask (*M*):

$$v_{sentence} = \frac{\sum_{i=1}^{L}(h_i \cdot M_i)}{\sum_{i=1}^{L} M_i} \quad (4)$$

Keterangan :

$v_{sentence}$ = vektor representasi kalimat yang diperoleh dari penggabungan informasi vektor setiap kata.

L = jumlah kata dalam kalimat.

h_i = vektor representasi kata ke- i yang menggambarkan informasi semantik atau makna kata.

M_i = bobot atau nilai mask untuk kata ke- i, yang menentukan apakah kata tersebut diperhitungkan dalam proses perhitungan.

$\sum_{i=1}^{L}(h_i \cdot M_i)$ = penjumlahan vektor setiap kata yang telah dikalikan dengan bobot atau mask masing-masing.

$\sum_{i=1}^{L} M_i$ = jumlah seluruh bobot atau mask yang digunakan sebagai pembagi untuk memperoleh rata-rata.

Di mana $h_i \epsilon \mathbb{R}^{128}$ adalah vektor representasi kata pada posisi ke-i, dan $M_i \epsilon \{0,1\}$ adalah nilai *mask* yang memastikan elemen *padding* bernilai 0 tidak merusak rata-rata matematis vektor.

BiLSTM adalah arsitektur turunan tingkat lanjut dari Recurrent Neural Network (RNN). RNN konvensional memiliki kelemahan fatal berupa vanishing gradient (gradien menghilang) ketika harus memproses deret waktu (time-series) atau teks yang panjang, di mana model "melupakan" informasi di awal kalimat saat mencapai akhir kalimat. LSTM memecahkan masalah ini menggunakan struktur Cell State (status memori selular), dan BiLSTM menyempurnakannya dengan membaca sekuens teks dari dua arah secara simultan(Chi et al., 2025).

Proses pemfilteran dan penyimpanan memori pada setiap langkah waktu (*t*) di dalam sel LSTM dikendalikan oleh tiga mekanisme gerbang (gating) yang beroperasi secara diferensiabel. Gerbang-gerbang ini menerima input ganda: vektor kata pada waktu saat ini (x_t) dan status tersembunyi dari langkah sebelumnya h_{t-1}.

1. **Gerbang Lupa (Forget Gate - f_t:**

   Menentukan persentase informasi historis dari cell state sebelumnya C_{t-1} yang harus dibuang atau dipertahankan. Fungsi sigmoid (σ) menekan nilai keluaran ke rentang $[0,1]$.

$$f_t = \sigma\big(W_f \cdot [h_{t-1}, x_t] + b_f\big) \quad (5)$$

Keterangan :

f_t = forget gate pada waktu ke- t, yang menentukan informasi dari memori sebelumnya yang akan dipertahankan atau dilupakan.

σ = fungsi aktivasi sigmoid yang menghasilkan nilai antara 0 dan 1.

W_f = matriks bobot (weight matrix) pada forget gate.

h_{t-1} = hidden state dari waktu sebelumnya, yaitu $t-1$

x_t = input pada waktu ke t

h_{t-1}, x_t = penggabungan (concatenation) antara hidden state sebelumnya dan input saat ini.

b_f = vektor bias (bias vector) pada forget gate.

t = indeks waktu atau urutan pemrosesan data.

2. Gerbang Masukan (Input Gate - i_t) dan Kandidat Sel ($\widetilde{C}_t$):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (6)$$

Keterangan :

i_t = input gate pada waktu ke- t, yang menentukan seberapa banyak informasi baru yang akan dimasukkan ke dalam memori sel.

σ = fungsi aktivasi sigmoid yang menghasilkan nilai antara 0 dan 1.

W_i = matriks bobot (weight matrix) pada input gate.

h_{t-1} = hidden state dari waktu sebelumnya.

x_t = input pada waktu ke- t

h_{t-1}, x_t = penggabungan (concatenation) hidden state sebelumnya dan input saat ini.

b_i = vektor bias (bias vector) pada input gate.

t = indeks waktu atau urutan pemrosesan data.

Keterangan :

$\tilde{C}_t$ = candidate cell state, yaitu kandidat informasi baru yang akan digunakan untuk memperbarui memori sel.

$tanh$ = fungsi aktivasi hyperbolic tangent yang menghasilkan nilai antara -1 dan 1.

W_C = matriks bobot untuk menghasilkan kandidat informasi baru pada memori sel.

h_{t-1} = hidden state dari waktu sebelumnya.

x_t = input pada waktu ke- t

h_{t-1}, x_t = penggabungan (concatenation) hidden state sebelumnya dan input saat ini.

b_C = vektor bias untuk kandidat cell state.

t = indeks waktu atau urutan pemrosesan data.

3. Pembaruan Status Sel Utama (Cell State Update - C_t)

Ini adalah inti memori jangka panjang dari LSTM. Status sel baru diformulasikan dengan mengalikan status lama dengan gerbang lupa, lalu menjumlahkannya dengan informasi kandidat baru yang telah difilter oleh gerbang masukan.

$$C_t = f_t * C_{t-1} + i_t * \tilde{b}_C \quad (7)$$

Keterangan :

C_t= cell state pada waktu ke- t, yaitu memori sel yang telah diperbarui.

f_t = forget gate yang menentukan seberapa banyak informasi dari memori sebelumnya yang dipertahankan.

C_{t-1} = cell state pada waktu sebelumnya, yaitu memori sel yang telah terbentuk pada langkah sebelumnya.

i_t = input gate yang mengatur seberapa banyak informasi baru yang dimasukkan ke dalam memori sel.

$\tilde{b}_C$ = perlu diperiksa kembali notasinya. Dalam rumus LSTM standar, simbol yang biasanya digunakan adalah C_t, yaitu candidate cell state atau kandidat informasi baru yang akan disimpan dalam memori sel.

$*$ = operasi perkalian. Dalam LSTM, operasi ini umumnya merupakan perkalian elemen demi elemen (element-wise multiplication).

4. Gerbang Keluaran (Output Gate - o_t) dan Status Tersembunyi (h_t):

Langkah terakhir adalah menentukan *hidden state* baru yang akan diteruskan ke waktu berikutnya atau ke lapisan (*layer*) berikutnya.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * tanh(C_t) \quad (8)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * tanh(C_t)$$

Keterangan :

o_t= nilai output gate pada waktu ke- t, yang menentukan informasi dari memori sel yang akan dikeluarkan.

σ = fungsi aktivasi sigmoid yang menghasilkan nilai antara 0 dan 1.

W_o = matriks bobot untuk menghitung output gate.

h_{t-1} = hidden state dari waktu sebelumnya.

x_t = input pada waktu ke- t

h_{t-1}, x_t = penggabungan (concatenation) hidden state sebelumnya dan input saat ini.

b_o = bias pada output gate.

t = indeks waktu atau urutan pemrosesan data.

Keterangan :

h_t= hidden state pada waktu ke-t, yaitu representasi informasi yang dikeluarkan oleh LSTM.

o_t= nilai output gate pada waktu ke-t, yang menentukan informasi dari memori sel yang akan dikeluarkan.

$tanh$ = fungsi aktivasi hyperbolic tangent yang menghasilkan nilai antara -1 dan 1.

C_t= cell state pada waktu ke- t, yaitu memori sel yang telah diperbarui.

$*$ = operasi perkalian. Dalam LSTM, operasi ini umumnya merupakan perkalian elemen demi elemen (element-wise multiplication).

Pada pemrosesan opini politik, sentimen sering kali dikunci oleh kata di akhir kalimat yang membalikkan makna di awal kalimat (misalnya pada kalimat

sarkasme). Arsitektur Bidirectional menyelesaikan masalah ini dengan menduplikasi lapisan LSTM dasar.

1. **Forward LSTM ($\vec{h}$) :** Membaca matriks berukuran 128×768 dari RoBERTa dari token x_1 hingga x_{128}.
2. **Forward LSTM ($\overleftarrow{h}$) :** Membaca matriks yang sama secara terbalik dari token x_{128} menuju x_1.

Keluaran pada setiap langkah waktu adalah hasil penggabungan (*concatenation*) dimensi spasial dari kedua arah tersebut:

$$H_t = [\overrightarrow{h_t} \oplus \overleftarrow{h_t}] \quad (9)$$

Keterangan :

H_t = representasi atau vektor keluaran pada waktu ke-\(t\) yang menggabungkan informasi dari dua arah.

$\overrightarrow{h_t}$ = hidden state dari proses maju (forward), yang memproses urutan data dari awal ke akhir.

$\overleftarrow{h_t}$ = hidden state dari proses mundur (backward), yang memproses urutan data dari akhir ke awal.

$\oplus$ = operasi penggabungan (concatenation) dua vektor menjadi satu vektor.

t = indeks waktu atau urutan pemrosesan data.

Sebagai contoh komputasi arsitektur, jika sebuah lapisan BiLSTM memiliki 128 unit neuron tersembunyi (*hidden units*), maka hasil konkatenasi akan melipat gandakan ruang fitur menjadi matriks keluaran berukuran 128×256 fitur.

Untuk mengekstraksi hierarki pola sintaksis yang lebih dalam (*deep learning*), sistem mengimplementasikan topologi *Stacked BiLSTM* (BiLSTM bertumpuk lapis ganda).

1. **Lapisan Pertama:** Berfungsi sebagai Sequence Returner. Lapisan ini menerima sequence embedding RoBERTa dan memetakan kembali matriks 3 Dimensi berurutan untuk diteruskan secara temporal ke lapisan berikutnya.
2. **Lapisan Kedua:** Berfungsi sebagai Feature Aggregator. Lapisan ini mereduksi seluruh rentetan waktu menjadi satu representasi vektor 1 Dimensi padat (vektor sentimen) yang akan disuntikkan ke dalam Classification Head (jaringan Dense dan Softmax).

Untuk menanggulangi kompleksitas parameter dimensi tinggi yang sangat rentan terhadap overfitting, regularisasi Dropout disisipkan pada komputasi seluler. Dropout secara matematis menonaktifkan probabilitas aktivasi neuron tertentu sebesar p $(misalnya\ p\ = 0.2)$ secara acak pada setiap iterasi pelatihan.

$$r_j^{(I)} \sim Bernoulli(1-p)$$
$$\tilde{y}^{(I)} = r^I * y^I \tag{10}$$

Keterangan :

$r_j^{(I)}$ = nilai dropout mask untuk neuron ke- j pada lapisan ke- I.

$\sim$ = menyatakan bahwa suatu variabel diambil berdasarkan distribusi probabilitas tertentu.

$Bernoulli(1-p)$ = distribusi Bernoulli dengan probabilitas menghasilkan nilai 1 sebesar $1-p$ dan nilai 0 sebesar p.

p = probabilitas neuron dinonaktifkan (dropout rate).

j = indeks neuron.

I = indeks lapisan (layer).

Keterangan :

$\tilde{y}^{(I)}$ = keluaran lapisan ke- I setelah penerapan dropout.

r^I = vektor dropout mask pada lapisan ke- I.

y^I = keluaran asli lapisan ke- I sebelum penerapan dropout.

* = perkalian elemen demi elemen (element-wise multiplication).

I = indeks lapisan (layer).

Mekanisme ini memaksa jaringan BiLSTM untuk tidak bergantung pada satu representasi kata tertentu saja, melainkan mendistribusikan bobot pembelajarannya ke seluruh konteks kalimat, sehingga model menjadi tangguh dalam memprediksi data uji di lingkungan dunia nyata.

Kompleksitas arsitektur BiLSTM yang mampu memodelkan dependensi kontekstual dua arah menjadikan pendekatan ini banyak diadopsi dalam berbagai penelitian analisis sentimen berbasis *deep learning*. Seiring dengan perkembangan model representasi bahasa berbasis *transformer*, integrasi BiLSTM dengan contextual embedding mulai menunjukkan peningkatan performa yang signifikan dibandingkan pendekatan embedding. Beberapa penelitian terdahulu telah membuktikan bahwa kombinasi tersebut mampu meningkatkan kemampuan model dalam memahami hubungan semantik dan konteks linguistik secara lebih mendalam. Meskipun berbagai penelitian menunjukkan bahwa integrasi BiLSTM dengan *contextual embedding* dapat meningkatkan kemampuan model dalam

memahami hubungan semantik dan konteks bahasa, peningkatan tersebut tetap perlu dibuktikan melalui evaluasi. Oleh karena itu, performa model tidak hanya diukur berdasarkan nilai akurasi, tetapi juga dianalisis menggunakan beberapa metrik evaluasi yang dapat memberikan gambaran lebih lengkap mengenai kualitas hasil klasifikasi pada setiap kelas sentimen, terutama ketika distribusi data antar kelas tidak seimbang.

2.2 Kuantifikasi Performa Multidimensional

Pada tahap akhir konklusi, keandalan model tidak dipukul rata sekadar pada nilai akurasi mentah, melainkan dibedah secara granular melalui metrik evaluasi yang lebih representatif terhadap kondisi ketidakseimbangan kelas politik(Hussein & Lakizadeh, 2025).

1. **Accuracy:** Merupakan metrik evaluasi global yang mengukur proporsi total prediksi benar (baik positif, negatif, maupun netral) dibandingkan dengan seluruh total observasi dalam dataset uji. Akurasi memberikan gambaran umum mengenai efektivitas model dalam mengenali pola sentimen secara keseluruhan.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

Keterangan :

TP (*True Positive*)

TN (*True Negative*)

FP (*False Positive*)

FN (*False Negative*)

2. **Precision:** Mengukur ketajaman prediksi model dalam meminimalisasi kesalahan positif palsu (false positive) pada setiap kelas sentimen.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (12)$$

3. **Recall:** Mengevaluasi daya jangkau model dalam mengidentifikasi seluruh sampel aktual dari masing-masing kategori polaritas.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (13)$$

4. **F1-Score:** Rata-rata harmonik yang mensintesis Precision dan Recall menjadi satu nilai tunggal guna memvalidasi efektivitas model pada distribusi data yang fluktuatif.

$$F1 - Score = 2x\frac{(\text{Precision x Recall})}{(\text{Precision } + \text{ Recall})} \quad (14)$$

Seluruh hasil komputasi kemudian diproyeksikan melalui *Normalized Confusion Matrix* berbasis persentase baris. Visualisasi ini krusial untuk menyingkap letak titik buta (*blind spots*) klasifikasi, yakni posisi di mana model mengalami kebingungan semantik dalam membedakan antara sentimen netral dengan polaritas politik yang ekstrem.

Salah satu penelitian yang relevan dilakukan oleh (Marutho et al., 2024) yang memperkenalkan pendekatan inovatif dalam analisis sentimen berbasis aspek menggunakan embedding kalimat melalui model transformer SBERT (*Sentence-BERT*). Pendekatan ini bertujuan untuk meningkatkan akurasi klasifikasi sentimen dengan memanfaatkan representasi kalimat yang lebih kontekstual. Meskipun tidak secara spesifik membahas konteks politik Indonesia, metodologi yang digunakan dapat diadaptasi untuk analisis sentimen politik di Twitter. Pendekatan ini

menunjukkan peningkatan kinerja dibandingkan metode sebelumnya, dengan akurasi mencapai 94,38% dan F1-score sebesar 83,24%.

Sebagai perbandingan, Penelitian lain melakukan eksperimen untuk mendeteksi komentar yang mengandung kalimat toksik di media sosial di Indonesia. Penelitian ini melibatkan pelatihan model klasifikasi multilabel dan mengevaluasi hasil klasifikasi yang dihasilkan oleh model Multilingual BERT (MBERT), IndoBERT, dan Indo Roberta Small. Meskipun menggunakan BERT, pendekatan dan temuan dalam studi ini dapat memberikan dasar untuk adaptasi menggunakan RoBERTa dalam konteks serupa. Hasil optimal dari penelitian ini adalah penggunaan model IndoBERT dengan skor F1 sebesar 0,8897(Nabiilah et al., 2022).

Selain itu, Penelitian lain mengeksplorasi analisis sentimen yang mempertimbangkan lokasi geografis dalam kampanye pemilu. Penelitian ini menggunakan pendekatan analisis sentimen berbasis lokasi yang disebut LEPM (*Location-Based Election Prediction Model*) untuk memprediksi hasil pemilu di berbagai negara bagian di Amerika Serikat. Teknik analisis sentimen yang digunakan dalam studi ini mencakup analisis sentimen berbasis aturan, *Valence Aware Dictionary and Sentiment Reasoner* (VADER), serta analisis sentimen berbasis transformer seperti *Bidirectional Encoder Representations from Transformers* (BERT), BERTweet, dan *Election-based BERT* (ElecBERT). Meskipun tidak spesifik pada Indonesia, pendekatan ini dapat diadaptasi untuk memahami sentimen politik di berbagai wilayah. Dengan menerapkan model LEPM pada pemilu tersebut, Menunjukkan kemampuannya dalam memprediksi

hasil dengan akurasi di 41 negara bagian, mencapai tingkat akurasi sebesar 0,84 di tingkat negara bagian(Khan et al., 2023).

Selanjutnya, terdapat penelitian yang mengeksplorasi pengaruh konten media sosial terhadap sikap dan tindakan masyarakat selama peristiwa kritis. Pendekatan ini melibatkan analisis interaksi digital dan klasifikasi sentimen untuk memahami dinamika sosial. Meskipun tidak secara langsung membahas politik, temuan ini dapat memberikan wawasan tentang dinamika sentimen di media sosial. Metodologi yang digunakan adalah melalui penerapan jaringan saraf *Deep Random Vector Functional Link* (D-RVFL), yang menunjukkan kinerja lebih unggul dibandingkan model-model tradisional seperti *Support Vector Machines* (SVM), *naive Bayes*, dan jaringan saraf *back propagation* (BP), dengan rata-rata akurasi keseluruhan sebesar 78,30%(Henríquez & Alessandri, 2024).

Di penelitian lain, penelitian lain berfokus pada penerapan algoritma *deep learning* untuk melakukan analisis sentimen terhadap data Twitter yang berkaitan dengan pandemi COVID-19. Tujuan utama penelitian ini adalah untuk mengkategorikan sentimen masyarakat menjadi tiga kelas, yaitu positif, negatif, dan netral. Penelitian ini menekankan pentingnya teknik preprocessing data teks serta membandingkan kinerja berbagai model *deep learning* dalam mengklasifikasikan sentimen dari teks pendek seperti tweet. Metode yang digunakan antara lain *Recurrent Neural Network* (RNN), *Long Short-Term Memory* (LSTM), *Convolutional Neural Network* (CNN), dan *Gated Recurrent Unit* (GRU). Dengan menggunakan tersebut LSTM hasilnya menunjukkan tingkat akurasi yang mengesankan sebesar 96% sebaliknya *Artificial Neural Network* (ANN), hanya

mencapai tingkat akurasi yang jauh lebih rendah yaitu 76%(Valarmathi et al., 2024).

Melengkapi berbagai pendekatan tersebut, Penelitian penelitian terbaru memperkenalkan sebuah model hybrid baru untuk konteks analisis sentimen, bertujuan untuk meningkatkan performa analisis sentimen, khususnya untuk teks pendek seperti tweet dan review online, yang sering kali mengandung informasi emosional yang kompleksagar bisa memahami konteks sentimen dalam teks dengan lebih akurat. Metode yang digunakan yaitu kombinasi antara RoBERTa (sebuah transformer pre-trained language model) dan LSTM (Long Short-Term Memory network). Hasil eksperimen menunjukkan bahwa model hybrid yang diusulkan melampaui metode-metode mutakhir dengan mencapai skor F1 masing-masing sebesar 93%, 91%, dan 90% pada dataset IMDb, dataset Twitter US Airline Sentiment, dan dataset Sentiment140(Tan et al., 2022).

Sementara itu, penerapan analisis sentimen dalam konteks lokal diperlihatkan melalui penelitian ini menyelidiki penerimaan konsumen terhadap kendaraan listrik di Indonesia melalui percakapan di Twitter. Meskipun fokusnya bukan pada politik, studi ini menunjukkan penerapan analisis sentimen dalam konteks Indonesia, yang dapat memberikan wawasan metodologis untuk penelitian. Penelitian ini menggunakan pendekatan zero-shot sentiment analysis dengan model klasifikasi sentimen Indo-RoBERTa. Analisis penelitian ini menunjukkan bahwa sentimen publik terhadap kendaraan listrik di Indonesia sebagian besar baik, dengan beberapa kekhawatiran terkait infrastruktur pengisian daya dan harga kendaraan.

Hal tersebut didominasi oleh sentimen netral (41,5%) dan sentimen negatif (31,1%), sedangkan sentimen positif hanya 27,4%(Nugroho & Widianto, 2024).

Selaras dengan upaya menciptakan ruang digital yang lebih sehat, penelitian lain membahas cara mendeteksi dan mengkategorikan toksisitas online (seperti ujaran kebencian, penghinaan, ancaman, bahasa tidak senonoh) dalam bahasa Indonesia, khususnya dari data Twitter, untuk menciptakan lingkungan digital yang lebih sehat di Indonesia. Dalam penelitian ini pengklasifikasiannya dengan memanfaatkan model IndoBERT dan IndoRoBERTa. IndoBERTweet menunjukkan kinerja yang lebih unggul dibandingkan Indonesian RoBERTa pada seluruh tahap deteksi, dengan pencapaian nilai Precision, Recall, dan F1-Score yang lebih tinggi. Secara khusus, nilai macro average F1-Score IndoBERTweet pada setiap tahap adalah sebagai berikut: 0,93 untuk deteksi tipe toksisitas, 0,86 untuk identifikasi target audiens, dan 0,84 untuk klasifikasi topik toksisitas(Alamsyah & Sagama, 2024).

Adapun dalam ranah keberlanjutan, Penelitian ini melakukan analisis sentimen terhadap data Twitter yang membahas aksi iklim dan perubahan iklim, dengan tujuan memahami opini publik terkait target Sustainable Development Goals (SDGs) seperti energi bersih, kota berkelanjutan, dan aksi iklim. Dalam penelitian ini, tiga metode utama digunakan untuk menganalisis sentimen: VADER (Valence Aware Dictionary for Sentiment Reasoning), TextBlob, dan model berbasis deep learning, yaitu BERT (Bidirectional Encoder Representations from Transformers). Hasil analisis menunjukkan bahwa model berbasis transformers menghasilkan akurasi yang lebih tinggi dibandingkan model lain. Model BERT,

mencapai akurasi sebesar 69%, lebih unggul dibandingkan VADER dengan akurasi 56% dan TextBlob dengan akurasi 46%(Rosenberg et al., 2023).

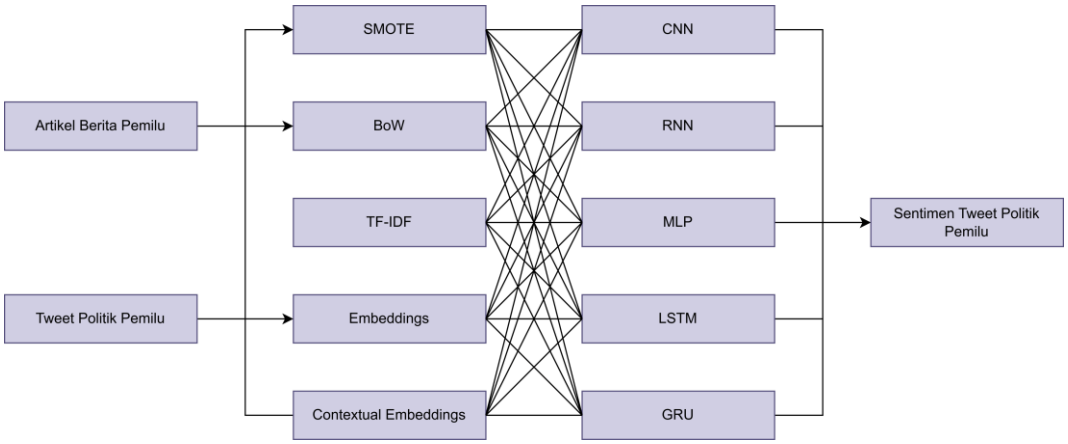
Terakhir, penelitian ini membahas analisis sentimen berbasis aspek (*Aspect-Based Sentiment Analysis*/ABSA) terhadap tweet para influencer Twitter untuk memprediksi tren harga tujuh *cryptocurrency* (Bitcoin, Ethereum, Binance, Ripple, Dogecoin, Cardano, dan Solana). Fokus utama adalah menghubungkan sentimen publik spesifik terhadap aspek-aspek tertentu dari *cryptocurrency* dengan prediksi pergerakan harga mereka dalam beberapa hari ke depan. Penelitian ini menggunakan model hybrid yang merupakan kombinasi dari RoBERTa (pre-trained), BiGRU, dan Attention Layer, digunakan untuk analisis sentimen. Hasil penelitian menunjukkan bahwa model yang diusulkan mencapai akurasi sebesar 5,94% serta peningkatan F1-score sebesar 9,9% pada dataset SemEval 2015 dan peningkatan 2,61% pada dataset SemEval 2016 dibandingkan dengan model state-of-the-art(Jahanbin & Chahooki, 2023).

2.3 Theoretical Framework

Kerangka teori (*theoretical framework*) dalam penelitian ini disusun sebagai landasan konseptual yang mendukung perumusan masalah, tujuan penelitian, serta pendekatan metodologis yang digunakan. Kerangka teori tersebut berfungsi untuk menjelaskan hubungan antar konsep dan teori yang relevan dengan penelitian mengenai analisis sentimen tweet politik Pemilu Indonesia menggunakan metode RoBERTa-BiLSTM.

Dengan adanya kerangka teori, penelitian dapat dilakukan secara lebih terstruktur serta memiliki dasar ilmiah yang jelas dalam menjelaskan proses

maupun hasil penelitian. Hubungan antar konsep dan tahapan penelitian tersebut digambarkan secara sistematis pada Gambar 2.2.



Gambar 2.1 ***Theoretical Framework***

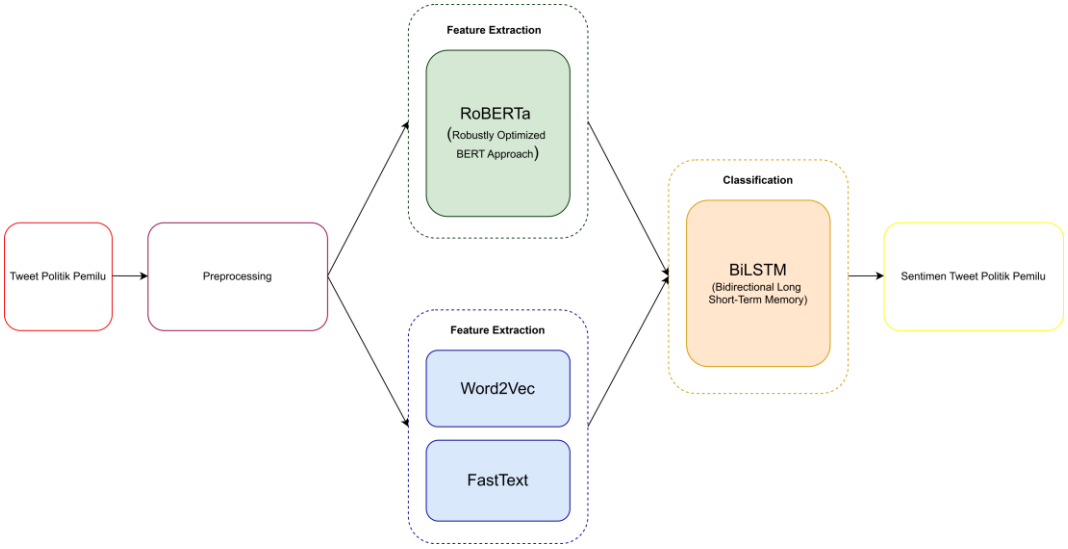
BAB III

METODOLOGI

Bab ini menjelaskan metodologi yang digunakan dalam penelitian, mulai dari tahapan pengumpulan data, proses *preprocessing*, ekstraksi fitur, perancangan model, hingga proses evaluasi performa model klasifikasi sentimen. Metodologi penelitian disusun secara sistematis untuk memastikan setiap tahapan penelitian berjalan terarah dan sesuai dengan tujuan penelitian yang telah ditetapkan. Selain itu, bab ini juga memaparkan alur implementasi model RoBERTa-BiLSTM dalam melakukan analisis sentimen terhadap data tweet politik pemilu yang diperoleh dari media sosial X.

3.1 Kerangka Konsep

Berlandaskan sintesis berbagai kajian literatur serta landasan teoritis yang telah dipaparkan sebelumnya, penelitian ini memerlukan suatu batasan konseptual dan alur pemikiran yang sistematis sebagai pedoman dalam pelaksanaan penelitian. Atas dasar tersebut, bagian ini menyajikan kerangka konsep yang berfungsi sebagai panduan operasional penelitian dengan menggambarkan keterkaitan antar gagasan utama secara terstruktur untuk menjawab rumusan masalah yang telah ditetapkan. Kerangka konsep penelitian ini dapat dilihat pada Gambar 3.1.



Gambar 3.1 Kerangka Konsep

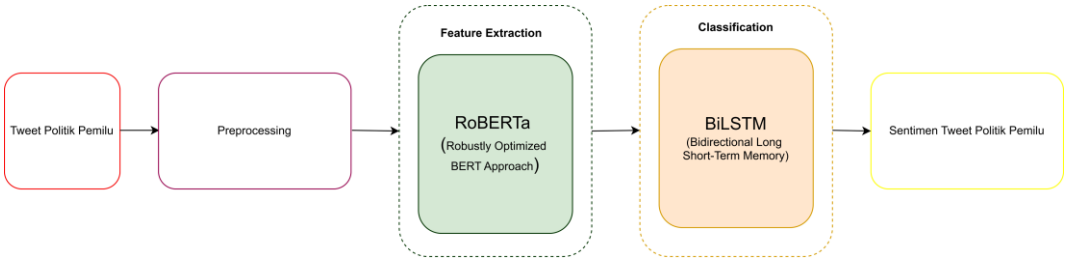
Berdasarkan kerangka konsep pada Gambar 3,1, Penelitian ini diawali dengan pengumpulan data berupa tweet politik pemilu yang selanjutnya diproses melalui tahapan *preprocessing*. Setelah proses prapemrosesan selesai dilakukan, data kemudian direpresentasikan ke dalam bentuk fitur atau yang biasa disebut *feature extraction* dalam penelitian ini, yaitu berbasis transformer melalui RoBERTa serta sebagai metode pembanding, penelitian ini turut menerapkan pendekatan representasi berbasis *embedding* menggunakan Word2Vec dan FastText.

Representasi fitur yang dihasilkan selanjutnya digunakan sebagai masukan pada model klasifikasi BiLSTM (*Bidirectional Long Short-Term Memory*) untuk melakukan analisis sentimen terhadap tweet politik pemilu. Melalui alur tersebut, penelitian ini berupaya mengevaluasi kemampuan masing-masing pendekatan representasi teks dalam meningkatkan performa klasifikasi sentimen, khususnya

dalam memahami konteks linguistik, variasi morfologis, serta hubungan semantik pada data teks berbahasa Indonesia.

3.2 Desain Sistem

Desain sistem pada penelitian ini dirancang untuk menggambarkan keseluruhan proses analisis sentimen tweet politik pemilu mulai dari tahap pengumpulan data hingga proses klasifikasi sentimen. Setiap tahapan dalam sistem disusun secara berurutan guna memastikan proses pengolahan data berjalan secara sistematis serta mampu menghasilkan keluaran yang sesuai dengan tujuan penelitian. Adapun alur desain sistem penelitian ini disajikan pada Gambar 3.2.



Gambar 3.2 Desain Sistem

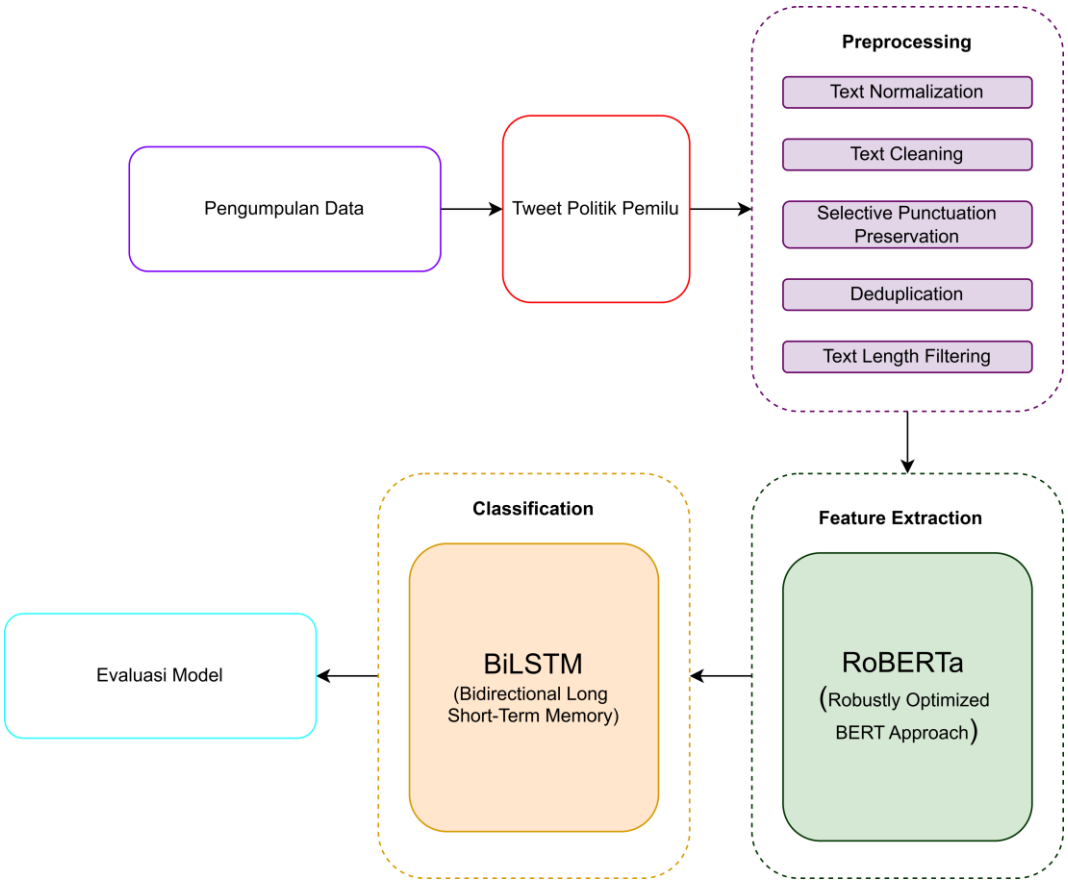
Berdasarkan desain sistem pada Gambar 3.2, proses penelitian diawali dengan pengumpulan data berupa tweet politik pemilu yang selanjutnya melalui tahapan *preprocessing*. Setelah tahap prapemrosesan selesai dilakukan, data kemudian memasuki tahap *feature extraction* menggunakan model RoBERTa (*Robustly Optimized BERT Approach*). Pada tahap ini, setiap teks direpresentasikan ke dalam bentuk vektor kontekstual yang mampu menangkap hubungan semantik antar kata berdasarkan konteks kalimat. Representasi fitur yang dihasilkan selanjutnya digunakan sebagai masukan pada model BiLSTM (*Bidirectional Long*

Short-Term Memory) untuk melakukan proses klasifikasi sentimen terhadap tweet politik pemilu.

Melalui rangkaian proses tersebut, sistem diharapkan mampu menghasilkan klasifikasi sentimen yang lebih optimal dalam memahami konteks bahasa alami, khususnya pada data media sosial yang memiliki karakteristik teks tidak terstruktur dan dinamis.

3.3 Desain Penelitian

Desain penelitian pada penelitian ini menggambarkan alur keseluruhan proses penelitian secara terstruktur mulai dari tahap pengumpulan data hingga interpretasi hasil pengujian model. Setiap tahapan dirancang untuk mendukung proses analisis sentimen tweet politik pemilu menggunakan pendekatan RoBERTa dan BiLSTM sehingga diperoleh hasil klasifikasi yang optimal dan dapat dievaluasi secara komprehensif. Desain penelitian tersebut disajikan pada Gambar 3.3.



Gambar 3.3 Desain Penelitian

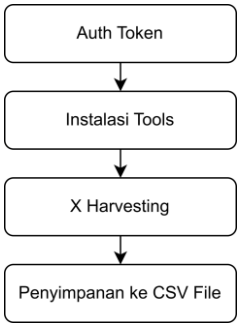
3.4 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data yang telah dikumpulkan dari media sosial X. Media sosial ini dipilih karena merupakan salah satu media sosial yang paling menonjol dalam diskursus politik, terutama di Indonesia. Aktivitas pengguna yang tinggi serta karakteristiknya yang mendukung penyebaran opini secara real-time menjadikan X sebagai sumber data yang relevan dan representatif dalam analisis sentiment.

Proses pengumpulan data pada X menggunakan tool x-harvest yang dijalankan melalui perintah npx berbasis Node.js. Tool ini melakukan harvesting data X dari tab LATEST berdasarkan kata kunci pencarian tertentu, dengan

parameter bahasa yang diinginkan dan jumlah maksimal tweet yang diambil sebanyak kebutuhan. Proses ini juga membutuhkan token autentikasi X agar dapat mengakses data publik secara real-time.

Data mentah dikumpulkan menggunakan tool x-harvest, Selanjutnya, data yang telah berhasil dikumpulkan akan disimpan dalam format .csv dan digunakan sebagai dasar untuk proses analisis lebih lanjut. Adapun alur proses pengumpulan data dapat dilihat pada Gambar 3.4.



Gambar 3.4 Proses Pengumpulan Data

Berdasarkan proses tersebut, data yang diperoleh dalam penelitian ini dapat dikategorikan sebagai data sekunder hasil harvesting dari sumber yang relevan dengan topik penelitian. Dataset yang terkumpul disimpan dalam format *comma-separated values* (CSV) guna memudahkan tahapan pengolahan dan analisis lanjutan. Secara keseluruhan, jumlah data yang berhasil dihimpun mencapai 26.000 entri, dengan periode pengambilan data dimulai dari Januari 2025 hingga November 2025. Pemilihan rentang waktu tersebut bertujuan untuk merepresentasikan dinamika percakapan publik secara lebih komprehensif selama periode penelitian berlangsung. Dengan demikian, hasil analisis yang diperoleh diharapkan mampu memberikan gambaran yang lebih akurat, kontekstual, dan

relevan terhadap fenomena sentimen politik yang berkembang di media sosial X/Twitter selama pelaksanaan Pemilu Indonesia.

3.5 *Preprocessing*

Data hasil proses *harvesting* terlebih dahulu diproses melalui tahap prapemrosesan (*preprocessing*) sebelum digunakan dalam pelatihan model. Tahap *preprocessing* pada penelitian ini dilakukan untuk meningkatkan kualitas data teks sebelum memasuki tahap ekstraksi fitur dan klasifikasi. Proses ini bertujuan untuk mengurangi noise, menyeragamkan format teks, serta mempertahankan informasi kontekstual yang relevan sehingga data dapat diproses secara lebih optimal oleh model yang digunakan. Adapun tahapan *preprocessing* yang diterapkan dalam penelitian ini meliputi *text normalization*, *text cleaning*, *selective punctuation preservation*, *deduplication*, dan *text length filtering*.

3.5.1 *Text Normalization*

Normalisasi teks merupakan tahapan awal dalam proses *preprocessing* yang bertujuan untuk menyeragamkan representasi data teks agar dapat diproses secara lebih optimal oleh model komputasi. Tahapan ini berperan penting dalam mengurangi inkonsistensi penulisan yang berpotensi memengaruhi kualitas representasi fitur pada proses analisis sentimen(Cui, 2025). Pada penelitian ini, normalisasi teks dilakukan melalui dua tahapan utama, yaitu *lowercasing* dan *whitespace normalization*.

Tahap pertama adalah lowercasing, yaitu proses mengubah seluruh karakter alfabet dalam teks menjadi huruf kecil (lowercase). Proses ini dilakukan untuk menghilangkan perbedaan representasi token yang disebabkan oleh variasi

kapitalisasi, meskipun secara semantik memiliki makna yang sama. Melalui transformasi tersebut, teks seperti “FUFUFAFA PIRANG STOP LEWAT TL GUE PLIS” akan diubah menjadi “fufufafa pirang stop lewat tl gue plis”. Proses ini turut berkontribusi terhadap reduksi ukuran kosakata (vocabulary size), karena token yang sebelumnya dianggap berbeda akibat penggunaan huruf kapital, seperti “STOP”, “Stop”, dan “stop”, akan direpresentasikan sebagai satu token yang identik.

Tahap kedua adalah whitespace normalization, yaitu proses normalisasi spasi untuk menjaga konsistensi struktur teks. Pada data media sosial, teks sering kali mengandung spasi berlebih, karakter newline, maupun tabulasi yang dapat mengganggu proses tokenisasi. Oleh karena itu, dilakukan transformasi berupa penghapusan spasi berlebih serta penyeragaman delimiter menjadi satu spasi tunggal. Tahapan normalisasi menghasilkan data teks dengan struktur yang lebih konsisten dan terstandarisasi. Kondisi ini memungkinkan proses ekstraksi fitur dan analisis lanjutan dilakukan secara lebih efektif. Hasil dari proses text normalization tersebut ditampilkan pada Tabel 3.1.

Tabel 3.1 Hasil *Text Normalization*

No	Sebelum Text Normalization	Sesudah Text Normalization
1	Ultranasionalis diorang pun boleh kata sokong buta-buta negara tengah huru-hara dengan prabowo & fufufafa pun boleh lagi nak urus-urus negara kita. Tapi nak sangat tengok penutur dua bahasa ni boleh bersatu ala-ala Skandinavia sana. Tulis lain sebut ikut tempat masing-masing.	ultranasionalis diorang pun boleh kata sokong buta-buta negara tengah huru-hara dengan prabowo & fufufafa pun boleh lagi nak urus-urus negara kita. tapi nak sangat tengok penutur dua bahasa ni boleh bersatu ala-ala skandinavia sana. tulis lain sebut ikut tempat masing-masing.
2	FUFUFAFA PIRANG STOP LEWAT TL GUE PLIS	fufufafa pirang stop lewat tl gue plis

No	Sebelum Text Normalization	Sesudah Text Normalization
3	@tham878 Jadi ingat proyek aglomerasi yg di pimpin si fufufafa proyek yg merampas hak² rakyat u di jual ke cina seseorang bajingan pengkhianat di negeri ini berlagak malaikat yg sok memperhatikan rakyat.	@tham878 jadi ingat proyek aglomerasi yg di pimpin si fufufafa proyek yg merampas hak² rakyat u di jual ke cina seseorang bajingan pengkhianat di negeri ini berlagak malaikat yg sok memperhatikan rakyat.
4	Udah mah ga enak tambah ga enak lagi grgr liat komuk fufufafa	udah mah ga enak tambah ga enak lagi grgr liat komuk fufufafa
5	@itsme_yennnn @tanyakanrl Jawa y dia maksud kan pulau isinya bukan cuma suku Jawa. Kenapa lu malah rasis idiot	@itsme_yennnn @tanyakanrl jawa y dia maksud kan pulau isinya bukan cuma suku jawa. kenapa lu malah rasis idiot

3.5.2 *Text Cleaning*

Tahap *text cleaning* merupakan bagian penting dalam proses *preprocessing* yang bertujuan untuk meningkatkan kualitas data teks sebelum dilakukan analisis lebih lanjut, khususnya pada analisis sentimen. Pada tahap ini, data teks mentah yang masih mengandung berbagai elemen tidak relevan (*noise*) dibersihkan sehingga menghasilkan representasi teks yang lebih terstruktur, konsisten, dan informatif(Jim et al., 2024).

Proses pembersihan teks dilakukan dengan menghapus elemen-elemen yang tidak memberikan kontribusi signifikan terhadap pembentukan makna sentimen, seperti *emoji*, *user mention* (@username), URL, karakter non-ASCII, serta simbol khusus lainnya. Keberadaan elemen-elemen tersebut cenderung meningkatkan kompleksitas data tanpa memberikan informasi semantik yang relevan terhadap proses klasifikasi sentimen, sehingga perlu dieliminasi untuk meningkatkan kualitas representasi teks.

Secara teknis, proses *text cleaning* dilakukan menggunakan pendekatan berbasis pola (*pattern matching*) melalui *regular expression* (*regex*) dan fungsi

manipulasi string. Setiap teks dipindai untuk mengidentifikasi pola tertentu, seperti *@username*, tautan URL, maupun karakter khusus, yang kemudian dihapus atau disesuaikan berdasarkan kebutuhan penelitian. Pendekatan ini memungkinkan proses pembersihan dilakukan secara sistematis dan konsisten terhadap seluruh data.

Dari perspektif konseptual, *text cleaning* didasarkan pada prinsip seleksi informasi, yaitu mempertahankan komponen teks yang relevan terhadap analisis sentimen serta menghilangkan komponen yang tidak memberikan kontribusi terhadap penentuan polaritas sentimen. Dengan demikian, elemen yang tidak memiliki nilai semantik terhadap konteks analisis diperlakukan sebagai *noise* dan dieliminasi dari data teks.

Transformasi text cleaning dilakukan terhadap teks mentah untuk menghasilkan teks yang telah dibersihkan dari elemen-elemen yang tidak relevan. Proses ini menghilangkan berbagai komponen noise seperti emoji, user mention, URL, serta simbol maupun karakter khusus lainnya.

Sebagai contoh, kalimat:

"@itsme_yennnn Jawa y dia maksud kan pulau isinya bukan cuma suku Jawa Kenapa lu malah rasis idiot"

setelah diproses melalui tahapan text cleaning akan menjadi:

"Jawa y dia maksud kan pulau isinya bukan cuma suku Jawa Kenapa lu malah rasis idiot"

Penghapusan elemen @itsme_yennnn dilakukan karena komponen tersebut tidak berkontribusi secara relevan terhadap analisis sentimen yang dilakukan.

Berdasarkan uraian tersebut, tahapan text cleaning berkontribusi signifikan dalam meningkatkan kualitas data teks dengan mengurangi noise serta mempertahankan informasi yang relevan. Hal ini diharapkan dapat meningkatkan kinerja model pada tahap ekstraksi fitur dan klasifikasi sentimen. Hasil dari proses text normalization tersebut ditampilkan pada Tabel 3.2.

Tabel 3.2 Hasil *Text Cleaning*

No	Sebelum Text Cleaning	Sesudah Text Cleaning
1	Ultranasionalis diorang pun boleh kata sokong buta-buta negara tengah huru-hara dengan prabowo & fufufafa pun boleh lagi nak urus-urus negara kita. Tapi nak sangat tengok penutur dua bahasa ni boleh bersatu ala-ala Skandinavia sana. Tulis lain sebut ikut tempat masing-masing.	Ultranasionalis diorang pun boleh kata sokong buta-buta negara tengah huru-hara dengan prabowo fufufafa pun boleh lagi nak urus-urus negara kita Tapi nak sangat tengok penutur dua bahasa ni boleh bersatu ala-ala Skandinavia sana Tulis lain sebut ikut tempat masing-masing
2	FUFUFAFA PIRANG STOP LEWAT TL GUE PLIS	FUFUFAFA PIRANG STOP LEWAT TL GUE PLIS
3	@tham878 Jadi ingat proyek aglomerasi yg di pimpin si fufufafa proyek yg merampas hak² rakyat u di jual ke cina seseorang bajingan pengkhianat di negeri ini berlagak malaikat yg sok memperhatikan rakyat.	Jadi ingat proyek aglomerasi yg di pimpin si fufufafa proyek yg merampas hak rakyat u di jual ke cina seseorang bajingan pengkhianat di negeri ini berlagak malaikat yg sok memperhatikan rakyat
4	Udah mah ga enak tambah ga enak lagi grgr liat komuk fufufafa	Udah mah ga enak tambah ga enak lagi grgr liat komuk fufufafa
5	@itsme_yennnn @tanyakanrl Jawa y dia maksud kan pulau isinya bukan cuma suku Jawa. Kenapa lu malah rasis idiot	Jawa y dia maksud kan pulau isinya bukan cuma suku Jawa Kenapa lu malah rasis idiot

3.5.3 *Selective Punctuation Preservation*

Selective punctuation preservation merupakan salah satu tahap dalam *text preprocessing* yang bertujuan untuk mempertahankan tanda baca tertentu yang dianggap memiliki kontribusi terhadap pemaknaan sentimen dalam teks. Berbeda

dengan pendekatan pembersihan teks secara umum yang cenderung menghapus seluruh tanda baca, pada tahap ini dilakukan seleksi terhadap tanda baca yang relevan secara semantik.

Beberapa tanda baca, seperti tanda seru (!) dan koma (,), tetap dipertahankan karena berpotensi merepresentasikan intensitas emosi dan ekspresi sentimen penulis. Tanda seru, misalnya, sering digunakan untuk menunjukkan penekanan atau emosi yang kuat, seperti kemarahan, antusiasme, atau kejutan. Sementara itu, koma dapat membantu mempertahankan struktur kalimat yang memengaruhi interpretasi makna secara keseluruhan.

Secara konseptual, pendekatan ini didasarkan pada asumsi bahwa tidak semua tanda baca bersifat *noise*. Dalam konteks analisis sentimen, beberapa tanda baca justru mengandung informasi tambahan yang dapat meningkatkan akurasi dalam mendeteksi polaritas dan intensitas sentimen(Siino et al., 2024). Oleh karena itu, dilakukan proses seleksi untuk mempertahankan tanda baca yang informatif dan menghapus tanda baca yang tidak relevan.

Proses tersebut dilakukan sebagai suatu fungsi transformasi selektif terhadap himpunan tanda baca. Dalam proses ini, tidak seluruh tanda baca dihapus, melainkan hanya mengeliminasi tanda baca di luar himpunan tanda baca yang dipertahankan.

Namun, tanda baca lain yang tidak relevan dapat dihapus sesuai dengan kebutuhan analisis.

Berdasarkan uraian tersebut, tahapan selective punctuation preservation berperan dalam menyeimbangkan proses pengurangan noise dengan pelestarian

informasi yang relevan. Hal ini berimplikasi pada peningkatan kualitas representasi teks dalam analisis sentimen. Hasil dari proses tersebut ditampilkan pada Tabel 3.2.

Tabel 3.3 Hasil *selective punctuation preservation*

No	Sebelum Selective Punctuation Preservation	Sesudah Selective Punctuation Preservation
1	Ultranasionalis diorang pun boleh kata sokong buta-buta negara tengah huru-hara dengan prabowo & fufufafa pun boleh lagi nak urus-urus negara kita. Tapi nak sangat tengok penutur dua bahasa ni boleh bersatu ala-ala Skandinavia sana. Tulis lain sebut ikut tempat masing-masing.	Ultranasionalis diorang pun boleh kata sokong buta-buta negara tengah huru-hara dengan prabowo fufufafa pun boleh lagi nak urus-urus negara kita Tapi nak sangat tengok penutur dua bahasa ni boleh bersatu ala-ala Skandinavia sana Tulis lain sebut ikut tempat masing-masing
2	FUFUFAFA PIRANG STOP LEWAT TL GUE PLIS	FUFUFAFA PIRANG STOP LEWAT TL GUE PLIS
3	@tham878 Jadi ingat proyek aglomerasi yg di pimpin si fufufafa proyek yg merampas hak² rakyat u di jual ke cina seseorang bajingan pengkhianat di negeri ini berlagak malaikat yg sok memperhatikan rakyat.	@tham878 Jadi ingat proyek aglomerasi yg di pimpin si fufufafa proyek yg merampas hak² rakyat u di jual ke cina seseorang bajingan pengkhianat di negeri ini berlagak malaikat yg sok memperhatikan rakyat
4	Udah mah ga enak tambah ga enak lagi grgr liat komuk fufufafa	Udah mah ga enak tambah ga enak lagi grgr liat komuk fufufafa
5	@itsme_yennnn @tanyakanrl Jawa y dia maksud kan pulau isinya bukan cuma suku Jawa. Kenapa lu malah rasis idiot	@itsme_yennnn @tanyakanrl Jawa y dia maksud kan pulau isinya bukan cuma suku Jawa Kenapa lu malah rasis idiot

3.5.4 *Deduplication*

Setelah melalui tahap *text cleaning* dan normalisasi, dilakukan proses *deduplication* untuk mengidentifikasi dan menghapus data teks yang memiliki kemiripan atau kesamaan secara identik. Tahap ini bertujuan untuk menghindari redundansi data yang berpotensi memengaruhi kinerja model analisis, khususnya dalam hal bias distribusi dan over-representasi terhadap pola tertentu.

Data duplikat dalam konteks analisis teks merujuk pada kemunculan lebih dari satu entri dengan isi yang sama atau sangat mirip setelah proses normalisasi dilakukan. Keberadaan data semacam ini dapat menyebabkan model pembelajaran mesin menjadi cenderung "menghafal" pola tertentu, sehingga menurunkan kemampuan generalisasi terhadap data baru(Ghassabi & Pahlevani, 2024).

Secara teknis, proses *deduplication* dilakukan dengan membandingkan setiap entri teks berdasarkan representasi hasil preprocessing. Jika dua atau lebih teks memiliki representasi yang identik, maka hanya satu yang dipertahankan, sementara sisanya dihapus. Proses ini umumnya dilakukan menggunakan operasi himpunan (*set operation*) atau fungsi penghapusan duplikasi pada struktur data, seperti *hashing* atau pencocokan string secara langsung.

Proses tersebut dilakukan untuk menghasilkan himpunan data tanpa duplikasi. Dengan demikian, setiap data yang dipertahankan bersifat unik dan tidak memiliki kemunculan berulang.

Kalimat yang memiliki struktur dan isi yang sama setelah normalisasi akan dianggap sebagai duplikat. Sistem akan mempertahankan satu representasi teks sebagai data unik, sementara entri lain yang identik akan dihapus dari dataset. Dengan demikian, tahap deduplication berperan penting dalam meningkatkan kualitas dataset dengan mengurangi redundansi, menjaga distribusi data tetap seimbang, serta meningkatkan efisiensi dan performa model pada tahap analisis selanjutnya. Adapun hasil proses tersebut ditampilkan pada Tabel 3.3.

Tabel 3.4 Hasil *deduplication*

No	Sebelum Deduplication	Sesudah Deduplication	Status
1	Ultranasionalis diorang pun boleh kata sokong buta-buta negara tengah huru-hara dengan prabowo & fufufafa pun boleh lagi nak urus-urus negara kita. Tapi nak sangat tengok penutur dua bahasa ni boleh bersatu ala-ala Skandinavia sana. Tulis lain sebut ikut tempat masing-masing.	Ultranasionalis diorang pun boleh kata sokong buta-buta negara tengah huru-hara dengan prabowo fufufafa pun boleh lagi nak urus-urus negara kita Tapi nak sangat tengok penutur dua bahasa ni boleh bersatu ala-ala Skandinavia sana Tulis lain sebut ikut tempat masing-masing	Dipertahankan (Unik)
2	FUFUFAFA PIRANG STOP LEWAT TL GUE PLIS	FUFUFAFA PIRANG STOP LEWAT TL GUE PLIS	Dipertahankan (Unik)
3	@tham878 Jadi ingat proyek aglomerasi yg di pimpin si fufufafa proyek yg merampas hak² rakyat u di jual ke cina seseorang bajingan pengkhianat di negeri ini berlagak malaikat yg sok memperhatikan rakyat.	@tham878 Jadi ingat proyek aglomerasi yg di pimpin si fufufafa proyek yg merampas hak² rakyat u di jual ke cina seseorang bajingan pengkhianat di negeri ini berlagak malaikat yg sok memperhatikan rakyat	Dipertahankan (Unik)
4	Udah mah ga enak tambah ga enak lagi grgr liat komuk fufufafa	Udah mah ga enak tambah ga enak lagi grgr liat komuk fufufafa	Dipertahankan (Unik)
5	@itsme_yennnn @tanyakanrl Jawa y dia maksud kan pulau isinya bukan cuma suku Jawa. Kenapa lu malah rasis idiot	@itsme_yennnn @tanyakanrl Jawa y dia maksud kan pulau isinya bukan cuma suku Jawa Kenapa lu malah rasis idiot	Dipertahankan (Unik)

3.5.5 *Text Length Filtering*

Tahap *text length filtering* dilakukan untuk memastikan bahwa data teks yang digunakan dalam analisis memiliki panjang yang memadai dan informatif.

Pada tahap ini, dilakukan penghitungan jumlah kata (*token count*) pada setiap entri teks, khususnya pada data tweet.

Penghitungan jumlah kata dilakukan dengan memecah teks menjadi unit-unit kata (*tokenization*), kemudian menghitung total token dalam setiap dokumen. Berdasarkan hasil penghitungan tersebut, data kemudian difilter dengan hanya mempertahankan teks yang memiliki jumlah kata lebih dari empat. Teks dengan panjang kurang dari atau sama dengan empat kata dianggap memiliki informasi yang terbatas dan berpotensi menjadi *noise* dalam proses analisis.

Secara konseptual, pendekatan ini didasarkan pada asumsi bahwa teks yang terlalu pendek cenderung tidak mengandung konteks yang cukup untuk menentukan polaritas sentimen secara akurat. Oleh karena itu, penghapusan teks pendek diharapkan dapat meningkatkan kualitas data latih serta membantu model dalam menangkap pola linguistik yang lebih representatif(Xiao & Luo, 2024).

Proses tersebut selanjutnya dilakukan sebagai fungsi seleksi terhadap panjang teks. Dengan demikian, hanya teks yang memiliki jumlah token lebih dari empat yang akan dipertahankan dalam dataset. Sebuah teks dengan jumlah kata sebanyak 40 token memenuhi kriteria panjang minimum, sehingga data tersebut dipertahankan. Sebaliknya, teks dengan jumlah kata yang berada di bawah ambang batas akan dieliminasi dari dataset. Berdasarkan uraian tersebut, tahapan text length filtering berkontribusi dalam meningkatkan kualitas dataset dengan mengeliminasi data yang kurang informatif. Hal ini berdampak pada peningkatan kinerja model dalam proses analisis sentimen. Hasil dari proses tersebut ditampilkan pada Tabel 3.4.

Tabel 3.5 Hasil *Text Length Filtering*

No	Sebelum Text Length Filtering	Jumlah Token	Status
1	Ultranasionalis diorang pun boleh kata sokong buta-buta negara tengah huru-hara dengan prabowo & fufufafa pun boleh lagi nak urus-urus negara kita. Tapi nak sangat tengok penutur dua bahasa ni boleh bersatu ala-ala Skandinavia sana. Tulis lain sebut ikut tempat masing-masing.	40 Kata	Dipertahankan
2	FUFUFAFA PIRANG STOP LEWAT TL GUE PLIS	7 Kata	Dipertahankan
3	@tham878 Jadi ingat proyek aglomerasi yg di pimpin si fufufafa proyek yg merampas hak² rakyat u di jual ke cina seseorang bajingan pengkhianat di negeri ini berlagak malaikat yg sok memperhatikan rakyat.	32 Kata	Dipertahankan
4	Udah mah ga enak tambah ga enak lagi grgr liat komuk fufufafa	12 Kata	Dipertahankan
5	@itsme_yennnn @tanyakanrl Jawa y dia maksud kan pulau isinya bukan cuma suku Jawa. Kenapa lu malah rasis idiot	18 Kata	Dipertahankan

3.6 RoBERTa

Penelitian ini memproyeksikan teks ke dalam vektor padat (*dense vector*) berdimensi tinggi. Model ini bertindak secara pasif sebagai pengekstraksi fitur murni (*feature extractor*), di mana pembaruan gradiennya dibekukan (*frozen*).

Setiap entri cuitan dipotong atau ditambahkan nilai semu (*padding*) hingga mencapai panjang statis 128 token. Ekstraksi kemudian menargetkan lapisan penutup dari arsitektur (*last hidden state*), menghasilkan keluaran berupa matriks spasial tiga dimensi (128×768) untuk setiap kumpulan pemrosesan (*batch*). Matriks ini tidak dikompresi, melainkan dipertahankan wujud sekuensialnya agar urutan kejadian kata tetap utuh saat diestafetkan ke arsitektur berikutnya yaitu BiLSTM.

3.6.1 Perhitungan Numerik Ekstraksi Fitur RoBERTa

Dalam memperjelas proses transformasi teks menjadi representasi numerik, pada bagian ini dijelaskan proses perhitungan yang didasarkan pada data aktual hasil penelitian. Representasi numerik yang dihasilkan oleh model *RoBERTa* dinyatakan dalam bentuk vektor berdimensi tinggi, sebagaimana ditunjukkan pada contoh berikut:

“mau ngurut apa ituu coba kita tanya ke fufufafa”

Hasil representasi *contextual embedding* dari teks tersebut dapat dinyatakan sebagai berikut:

“[0.943, -0.790, 1.216, ..., 0.741]”

Nilai-nilai pada vektor di atas diperoleh melalui proses agregasi contextual embedding setiap token dalam kalimat menggunakan metode rata-rata (mean

pooling). Dalam menjelaskan proses tersebut secara matematis berbasis angka nyata, digunakan contoh pada data pertama. Sebagaimana dapat dilihat pada Tabel 3.2.

Tabel 3.6 Nilai *contextual embedding* Token pada Dimensi ke-1 Hasil Ekstraksi Fitur RoBERTa

Nilai Contextual Embedding Dimensi ke-1	
Token	**Nilai**
mau	0.90
ngurut	1.05
apa	0.82
ituu	0.91
coba	1.18
kita	0.86
tanya	0.95
ke	0.89
fufufafa	0.93

Langkah Perhitungan

$$0.90+1.05=1.95$$

$$1.95+0.82=2.77$$

$$2.77+0.91=3.68$$

$$3.68+1.18=4.86$$

$$4.86+0.86=5.72$$

$$5.72+0.95=6.67$$

$$6.67+0.89=7.56$$

$$7.56+0.93=8.49$$

Kemudian dihitung rata-rata menggunakan persamaan mean pooling

$$\frac{8.49}{9} = 0.943$$

Sehingga diperoleh nilai contextual embedding pada dimensi ke-1 sebesar: 0.943

Nilai tersebut kemudian menjadi bagian dari sentence contextual embedding hasil ekstraksi fitur RoBERTa:

[0.943, -0.790, 1.216, ..., 0.741]

Pendekatan mean pooling digunakan untuk merepresentasikan keseluruhan informasi semantik kalimat berdasarkan rata-rata representasi setiap token pada dimensi contextual embedding tertentu.

3.7 BiLSTM

Integrasi model utama dalam penelitian ini dikembangkan menggunakan kerangka kerja (*framework*) TensorFlow dengan mengombinasikan kemampuan representasi kontekstual dari RoBERTa dan kemampuan pemrosesan data sekuensial dari BiLSTM. Representasi *contextual embedding* hasil ekstraksi fitur oleh RoBERTa selanjutnya diteruskan ke arsitektur *stacked* BiLSTM untuk mempelajari hubungan kontekstual antartoken secara lebih mendalam. Struktur arsitektur model yang diterapkan pada penelitian ini dijelaskan sebagai berikut:

1. **Lapisan BiLSTM Pertama**

   Lapisan pertama terdiri atas 128 unit neuron tersembunyi yang bekerja secara dua arah (bidirectional), sehingga model mampu memahami hubungan konteks baik dari urutan token sebelumnya maupun token sesudahnya dalam suatu kalimat. Konfigurasi return sequences=True diterapkan agar seluruh keluaran sekuens dari setiap token tetap diteruskan ke lapisan berikutnya. Pendekatan ini memungkinkan informasi kontekstual pada setiap posisi token tetap dipertahankan selama proses pembelajaran berlangsung.

2. Lapisan BiLSTM Kedua

Lapisan kedua menggunakan 64 unit neuron tersembunyi yang juga bekerja secara bidirectional. Berbeda dengan lapisan sebelumnya, konfigurasi *return sequences=False* diterapkan sehingga hanya representasi keluaran akhir yang digunakan sebagai ringkasan fitur sekuens. Proses ini bertujuan menghasilkan representasi fitur yang lebih ringkas, namun tetap mampu mempertahankan informasi penting yang relevan terhadap proses klasifikasi sentimen.

3. Lapisan Klasifikasi (*Classification Head*)

Representasi fitur yang telah dihasilkan kemudian diteruskan ke lapisan Dense dengan 64 neuron dan fungsi aktivasi ReLU (Rectified Linear Unit). Lapisan ini berfungsi untuk membangun hubungan nonlinier antarfitur sehingga model dapat mempelajari pola klasifikasi yang lebih kompleks sebelum memasuki tahap prediksi akhir.

4. Mekanisme Regularisasi

Untuk meminimalkan risiko overfitting akibat ketergantungan antarneuron (co-adaptation), penelitian ini menerapkan teknik regularisasi dropout pada beberapa bagian arsitektur model. Nilai dropout sebesar 20% diterapkan pada kedua lapisan BiLSTM, sedangkan dropout sebesar 30% diterapkan setelah lapisan Dense sebelum proses klasifikasi akhir dilakukan. Penerapan mekanisme ini bertujuan untuk meningkatkan kemampuan generalisasi model terhadap data baru yang belum pernah diproses sebelumnya.

Pada tahap akhir, lapisan keluaran (*output layer*) menggunakan tiga neuron yang merepresentasikan masing-masing kelas sentimen, yaitu negatif, netral, dan positif. Fungsi aktivasi Softmax diterapkan untuk mengubah keluaran model menjadi distribusi probabilitas, sehingga model dapat menentukan kategori sentimen berdasarkan nilai probabilitas prediksi tertinggi.

3.8 Evaluasi

Proses iterasi pembaruan bobot dalam penelitian ini dikendalikan oleh algoritma pengoptimal stokastik Adaptive Moment Estimation (Adam) dengan laju pembelajaran konstan. Jarak antara distribusi prediksi model dengan kebenaran dasar dikuantifikasi melalui kalkulasi Categorical Crossentropy Loss. Guna mengefisiensikan manajemen sumber daya pasca-ekstraksi fitur dari GPU, fase pelatihan relokasi secara taktis ke unit komputasi sentral (CPU) untuk menghindari limitasi memori (VRAM bottleneck).

3.8.1 Pengawasan Konvergensi via Learning Curves

Keandalan proses pembelajaran model tidak hanya dievaluasi berdasarkan hasil akhir pengujian, tetapi juga dianalisis secara berkelanjutan melalui pemanfaatan *learning curves*. Riwayat proses pelatihan (*training history*) diekstraksi untuk menggambarkan pola pembelajaran model pada setiap *epoch* menggunakan dua indikator utama sebagai berikut:

1. **Kurva Akurasi (*Accuracy Curve*)**

   Kurva akurasi digunakan untuk membandingkan kemampuan prediksi model pada data pelatihan (training set) dan data validasi (validation set). Pola pergerakan kurva ini memberikan gambaran mengenai perkembangan

kemampuan model selama proses pelatihan berlangsung, sekaligus membantu mengidentifikasi kondisi ketika model mulai mencapai titik konvergensi atau mengalami penurunan performa akibat ketidakstabilan proses pembelajaran.

2. **Kurva Kerugian (*Loss Curve*)**

   Kurva kerugian digunakan untuk memantau perubahan nilai fungsi objektif Categorical Crossentropy pada setiap epoch. Analisis terhadap selisih antara training loss dan validation loss dilakukan untuk mengevaluasi stabilitas model selama proses pelatihan. Semakin kecil perbedaan kedua nilai tersebut, maka semakin baik kemampuan generalisasi model terhadap data baru, sehingga risiko terjadinya overfitting dapat diminimalkan.

3.8.2 Mekanisme Kontrol dan Retensi Model Terbaik

Dalam menjaga integritas hasil eksperimen, prosedur pelatihan dilindungi oleh fungsi pengawas *Early Stopping* yang dikalibrasi untuk menghentikan iterasi secara prematur apabila grafik kerugian pada data validasi tidak menunjukkan depresiasi selama lima siklus berturut-turut. Intervensi ini menjamin bahwa arsitektur final tidak kehilangan kemampuan generalisasinya akibat iterasi berlebih. Selanjutnya, titik koordinat bobot optimal selama masa transisi tersebut diabadikan menggunakan instruksi *Model Checkpoint* berbasis nilai akurasi validasi tertinggi. Evaluasi performa model dilakukan menggunakan pendekatan multidimensional agar kemampuan klasifikasi dapat diukur secara lebih komprehensif, terutama pada kondisi distribusi kelas yang tidak seimbang. Pengukuran kinerja model dilakukan

menggunakan metrik Accuracy, Precision, Recall, dan F1-Score sebagai indikator utama dalam mengevaluasi hasil klasifikasi sentimen.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Data Penelitian

Data yang digunakan dalam penelitian ini bersumber dari media sosial X (sebelumnya Twitter), yang secara spesifik berfokus pada percakapan publik terkait isu politik Pemilihan Umum (Pemilu) di Indonesia. Pengumpulan data dilakukan menggunakan teknik crawling dengan memanfaatkan tool x-harvest yang dijalankan melalui lingkungan Node.js untuk memperoleh data tweet dari media sosial X. Pengambilan data dilakukan pada rentang waktu Januari 2025 hingga November 2025, dengan menggunakan kata kunci yang berkaitan dengan isu politik dalam Pemilu Indonesia. Berdasarkan parameter pencarian yang menggunakan kata kunci seperti "pemilu", "pilpres", dan "politik Indonesia", penelitian ini berhasil mengumpulkan sebanyak 26.000 tweet mentah (raw data) dari media sosial X yang selanjutnya digunakan sebagai dataset pada tahap pengolahan dan analisis sentimen.

4.2 Hasil Pra-pemrosesan Data

Tahap pra-pemrosesan data merupakan proses awal yang memiliki peranan penting dalam meningkatkan kualitas data teks sebelum digunakan pada tahapan pemodelan menggunakan metode RoBERTa-BiLSTM. Proses ini menjadi krusial karena data yang digunakan dalam penelitian berupa tweet politik terkait Pemilu Indonesia yang memiliki karakteristik tidak terstruktur, serta mengandung berbagai bentuk noise, seperti penggunaan bahasa tidak baku, singkatan, emoji, simbol, tautan, dan variasi penulisan yang beragam. Kondisi tersebut berpotensi

memengaruhi kualitas representasi teks serta menurunkan kemampuan model dalam memahami konteks sentimen secara optimal.

Dalam mengatasi permasalahan tersebut, diterapkan serangkaian tahapan pra-pemrosesan yang meliputi text normalization, text cleaning, selective punctuation preservation, deduplication, serta text length filtering. Setiap tahapan dirancang untuk meningkatkan kualitas data, mulai dari penghilangan elemen yang tidak relevan, pelestarian informasi penting, penyeragaman bentuk kata, hingga pengurangan redundansi dan eliminasi data yang kurang informatif.

Penerapan tahapan pra-pemrosesan tersebut bertujuan untuk menghasilkan data teks yang lebih bersih, terstruktur, dan representatif, sehingga dapat meningkatkan efektivitas model RoBERTa-BiLSTM dalam mempelajari pola sentimen pada tweet politik pemilu Indonesia. Selain itu, proses ini juga berperan dalam meminimalkan ambiguitas makna yang kerap muncul pada data media sosial akibat penggunaan bahasa yang bersifat informal. Perbandingan hasil sebelum dan sesudah tahap pra-pemrosesan data ditampilkan pada Tabel 4.1.

Tabel 4.1 Perbandingan Data Sebelum dan Sesudah Pra-pemrosesan

Sebelum Pra-pemrosesan Data	Sesudah Pra-pemrosesan Data
Ultranasionalis diorang pun boleh kata sokong buta-buta negara tengah huru-hara dengan prabowo & fufufafa pun boleh lagi nak urus-urus negara kita. Tapi nak sangat tengok penutur dua bahasa ni boleh bersatu ala-ala Skandinavia sana. Tulis lain sebut ikut tempat masing-masing.	ultranasionalis diorang pun boleh kata sokong butabuta negara tengah huruhara dengan prabowo amp fufufafa pun boleh lagi nak urusurus negara kita tapi nak sangat tengok penutur dua bahasa ni boleh bersatu alaala skandinavia sana tulis lain sebut ikut tempat masingmasing
FUFUFAFA PIRANG STOP LEWAT TL GUE PLIS	fufufafa pirang stop lewat tl gue plis
@tham878 Jadi ingat proyek aglomerasi yg di pimpin si fufufafa proyek yg merampas hak² rakyat u di	jadi ingat proyek aglomerasi yg di pimpin si fufufafa proyek yg merampas hak rakyat u di jual ke cina seseorang

Sebelum Pra-pemrosesan Data	Sesudah Pra-pemrosesan Data
jual ke cina seseorang bajingan pengkhianat di negeri ini berlagak malaikat yg sok memperhatikan rakyat.	bajingan pengkhianat di negeri ini berlagak malaikat yg sok memperhatikan rakyat
Udah mah ga enak tambah ga enak lagi grgr liat komuk fufufafa	udah mah ga enak tambah ga enak lagi grgr liat komuk fufufafa
@itsme_yennnn @tanyakanrl Jawa y dia maksud kan pulau isinya bukan cuma suku Jawa. Kenapa lu malah rasis idiot	jawa y dia maksud kan pulau isinya bukan cuma suku jawa kenapa lu malah rasis idiot

Berdasarkan Tabel 4.1, terlihat bahwa teks hasil pra-pemrosesan mengalami penyederhanaan struktur melalui proses normalisasi kata serta penghilangan karakter dan elemen yang tidak relevan. Transformasi ini menghasilkan bentuk teks yang lebih terstandarisasi dan konsisten dibandingkan dengan kondisi awalnya.

Selain itu, tahapan pra-pemrosesan terbukti efektif dalam mengurangi noise yang umum ditemukan pada data tweet politik pemilu, seperti penggunaan bahasa informal, pengulangan kata, serta variasi penulisan yang tidak sesuai kaidah. Berkurangnya elemen-elemen tersebut menyebabkan representasi teks menjadi lebih terfokus pada informasi semantik utama yang relevan dengan analisis sentimen.

Kondisi ini memberikan manfaat positif terhadap tahap pemodelan, khususnya dalam meningkatkan kemampuan model RoBERTa-BiLSTM dalam menangkap konteks dan pola linguistik pada data. Dengan demikian, data yang telah melalui tahap pra-pemrosesan menjadi lebih optimal untuk digunakan dalam proses pelatihan maupun evaluasi model.

4.3 Hasil Ekstraksi Fitur RoBERTa

Tahap ekstraksi fitur dilakukan untuk mengubah data teks hasil pra-pemrosesan menjadi representasi numerik yang dapat diproses oleh model *deep*

learning. Pada penelitian ini, ekstraksi fitur memanfaatkan model Indo-RoBERTa yang mampu menghasilkan representasi kontekstual berbasis arsitektur *transformer*.

Proses ekstraksi diawali dengan tokenisasi menggunakan tokenizer RoBERTa, di mana setiap teks dipecah menjadi token berbentuk *subword*. Selanjutnya, setiap token direpresentasikan ke dalam vektor contextual embedding berdimensi 768 yang diperoleh dari keluaran *last_hidden_state* model RoBERTa. Representasi ini mencerminkan informasi kontekstual dari setiap token dalam kalimat.

Berbeda dengan metode *word embedding* yang bersifat, contextual embedding yang dihasilkan RoBERTa bersifat kontekstual, sehingga makna suatu kata dapat berubah sesuai dengan konteks penggunaannya dalam kalimat. Hal ini memungkinkan model untuk menangkap hubungan semantik dan pola bahasa secara lebih mendalam.

Dalam penelitian ini, representasi yang digunakan tidak terbatas pada token khusus di awal kalimat, melainkan mencakup seluruh *sequence embedding* dengan ukuran (128, 768). Pendekatan ini memungkinkan model BiLSTM untuk mempelajari dependensi antar token secara lebih komprehensif dalam suatu urutan teks.

Selain sebagai masukan utama bagi model BiLSTM, representasi contextual embedding juga diolah menggunakan teknik *mean pooling* untuk keperluan visualisasi. Proses ini dilakukan dengan menghitung rata-rata contextual embedding token berdasarkan *attention mask*, sehingga representasi akhir tetap

mempertahankan informasi kontekstual utama dari kalimat. Perbandingan representasi data sebelum dan sesudah proses ekstraksi fitur RoBERTa ditampilkan pada Tabel 4.2.

Tabel 4.2 Perbandingan Data Sebelum dan Sesudah Ekstraksi Fitur RoBERTa

Sebelum Ekstraksi Fitur RoBERTa	Sesudah Ekstraksi Fitur RoBERTa
mau ngurut apa ituu coba kita tanya ke fufufafa	[0.943, -0.790, 1.216, ..., 0.741]
ini arahan megawati untuk 3214 anggota dprd fraksi pdip	[-0.315, 0.021, 0.470, ..., -0.321]
dalam rangka memperingati hari ulang tahun ke78 ketua umum pdi perjuangan megawati soekarnoputri dpc pdi perjuangan kabupaten semarang menggelar kegiatan berbagi kebahagiaan dengan masyarakat sekitar kamis 2312025 sebagai wujud syukur dan penghormatan atas dedikasi ibu	[0.985, 0.376, 0.588, ..., 0.380]
lalu siapa lagi yg ngisi polling ntuk nyangkutin masalah kesana kaum 58 mah udah pada move on dngn pilpres haya bani 16 aja yg masih ngotot amp nywrag jokowi sampe2 hasto nyamperin mantan kaesang ke singapore lhookoo ngihahajanaa	[0.915, -0.129, 1.050, ..., 0.528]
menteri esdm selaku jubir jokowi	[0.299, 0.009, 0.852, ..., -0.829]

Berdasarkan Tabel 4.2, terlihat bahwa setiap sampel teks yang semula berbentuk kalimat dalam bahasa alami telah berhasil ditransformasikan ke dalam representasi numerik berdimensi tinggi. Representasi tersebut berupa vektor contextual embedding yang dihasilkan melalui proses ekstraksi fitur menggunakan model Indo-RoBERTa, di mana nilai-nilai numerik mencerminkan informasi kontekstual dari token-token penyusun kalimat yang selanjutnya digabungkan menggunakan teknik *mean pooling*.

Berbeda dengan metode representasi teks konvensional yang menghasilkan embedding, pendekatan berbasis RoBERTa mampu menghasilkan representasi yang bersifat kontekstual. Dengan demikian, makna suatu token tidak dipahami secara terpisah, melainkan dalam keterkaitannya dengan token lain dalam struktur kalimat. Hal ini menjadi penting dalam analisis sentimen tweet politik Pemilu Indonesia yang umumnya mengandung bahasa informal, singkatan, serta variasi penulisan yang dinamis.

Hasil ekstraksi fitur menunjukkan bahwa model Indo-RoBERTa mampu memetakan informasi semantik ke dalam representasi numerik yang lebih komprehensif, tidak hanya mencakup aspek leksikal, tetapi juga hubungan kontekstual antar token yang berperan dalam pembentukan makna sentimen.

Representasi contextual embedding tersebut selanjutnya digunakan sebagai masukan pada model BiLSTM untuk mempelajari pola dependensi sekuensial antar token. Dengan kualitas representasi fitur yang lebih baik, model diharapkan mampu mengidentifikasi pola sentimen secara lebih akurat serta adaptif terhadap karakteristik bahasa pada media sosial.

4.4 Evaluasi Proses Pelatihan (*Learning Curves*)

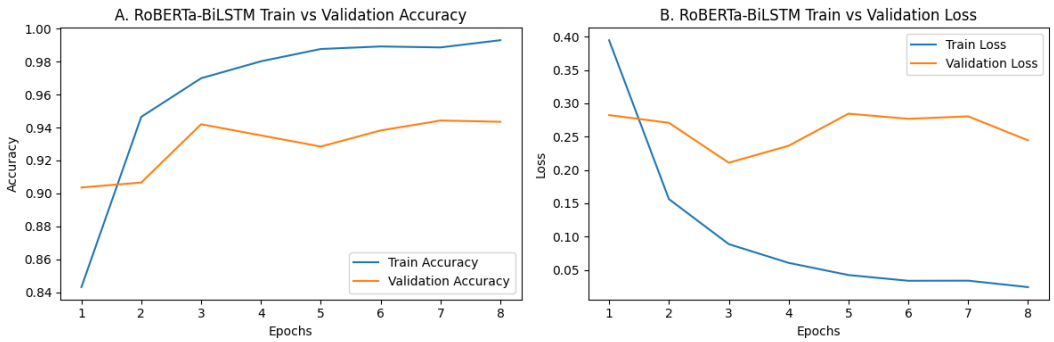
Evaluasi terhadap proses pelatihan model dilakukan untuk menganalisis dinamika pembelajaran pada model RoBERTa-BiLSTM selama tahapan *training* berlangsung. Analisis ini bertujuan untuk mengetahui kemampuan model dalam mempelajari pola dan karakteristik sentimen dari data tweet politik Pemilu Indonesia, sekaligus mengevaluasi kemampuan generalisasi model terhadap data validasi. Oleh karena itu, evaluasi tidak hanya difokuskan pada hasil performa

akhir, tetapi juga mencakup kualitas, stabilitas, dan konsistensi proses pembelajaran model selama pelatihan.

Pada penelitian ini, proses evaluasi divisualisasikan menggunakan *learning curves* yang menampilkan perkembangan nilai *accuracy* dan *loss* pada data *training* serta *validation* di setiap epoch. Visualisasi tersebut memberikan gambaran yang lebih jelas mengenai perilaku model selama proses pelatihan berlangsung. Melalui kurva pembelajaran tersebut, dapat diamati pola konvergensi model, tingkat kestabilan proses pembelajaran, serta indikasi terjadinya *overfitting* maupun *underfitting*.

Selain itu, perubahan nilai *accuracy* dan *loss* pada setiap epoch juga digunakan untuk mengevaluasi efektivitas arsitektur RoBERTa-BiLSTM dalam memahami representasi sentimen pada data teks yang bersifat dinamis, tidak terstruktur, dan kontekstual seperti tweet politik. Analisis ini menjadi penting karena karakteristik bahasa pada media sosial umumnya mengandung variasi penulisan, singkatan, bahasa informal, serta konteks opini yang kompleks. Dengan demikian, evaluasi proses pelatihan tidak hanya menunjukkan tingkat performa model, tetapi juga memberikan pemahaman mengenai kemampuan model dalam mempertahankan keseimbangan antara proses pembelajaran pada data latih dan kemampuan generalisasi terhadap data baru.

Hasil visualisasi proses pelatihan model RoBERTa-BiLSTM ditampilkan pada Gambar 4.1.



Gambar 4.1 Kurva Training dan Validation Loss

Berdasarkan hasil pelatihan model pada Gambar 4.1 A, terlihat bahwa nilai *training accuracy* mengalami peningkatan yang cukup signifikan pada beberapa epoch awal, kemudian meningkat secara lebih perlahan hingga mencapai sekitar 0,994 pada epoch ke-8. Kondisi ini menunjukkan bahwa model RoBERTa-BiLSTM mampu mempelajari pola pada data latih dengan baik. Sejalan dengan peningkatan *training accuracy*, nilai *training loss* mengalami penurunan secara konsisten dari sekitar 0,40 pada epoch pertama hingga mencapai sekitar 0,03 pada epoch ke-7 dan sekitar 0,02 pada epoch ke-8. Pola tersebut menunjukkan bahwa proses optimisasi model berjalan dengan baik pada data pelatihan.

Namun demikian, pola yang berbeda terlihat pada *validation accuracy*. Pada awal pelatihan, nilai akurasi validasi meningkat dari sekitar 0,903 pada epoch pertama menjadi sekitar 0,942 pada epoch ke-3. Setelah itu, nilai *validation accuracy* mengalami fluktuasi, yaitu menurun hingga sekitar 0,928 pada epoch ke-5, kemudian kembali meningkat hingga mencapai nilai tertinggi sekitar 0,945 pada epoch ke-7. Pada epoch ke-8, nilai *validation accuracy* sedikit menurun menjadi sekitar 0,944. Kondisi tersebut menunjukkan bahwa peningkatan performa pada data latih tidak sepenuhnya diikuti oleh peningkatan performa pada data validasi.

Pola tersebut juga terlihat pada grafik *loss* di bagian kanan Gambar 4.1 B. Nilai *training loss* terus mengalami penurunan pada setiap epoch, sedangkan *validation loss* menunjukkan pola yang lebih fluktuatif. Nilai *validation loss* menurun dari sekitar 0,28 pada epoch pertama hingga mencapai nilai terendah sekitar 0,21 pada epoch ke-3. Setelah itu, *validation loss* kembali meningkat pada epoch ke-4 hingga sekitar 0,24 dan mencapai sekitar 0,28 pada epoch ke-5. Nilai tersebut kemudian relatif berfluktuasi pada epoch berikutnya dan menurun kembali menjadi sekitar 0,25 pada epoch ke-8. Perbedaan pola antara *training loss* yang terus menurun dan *validation loss* yang cenderung berfluktuasi menunjukkan adanya indikasi *overfitting* selama proses pelatihan.

Fenomena *overfitting* tersebut menunjukkan bahwa model semakin baik dalam menyesuaikan parameter terhadap data latih, tetapi peningkatan tersebut tidak selalu menghasilkan peningkatan performa yang sebanding pada data validasi. Hal ini terlihat dari semakin tingginya *training accuracy* hingga sekitar 0,994 pada epoch ke-8, sementara *validation accuracy* cenderung berada pada kisaran 0,93–0,95. Dengan demikian, terdapat kesenjangan antara performa model pada data pelatihan dan data validasi yang perlu diperhatikan dalam menentukan epoch optimal.

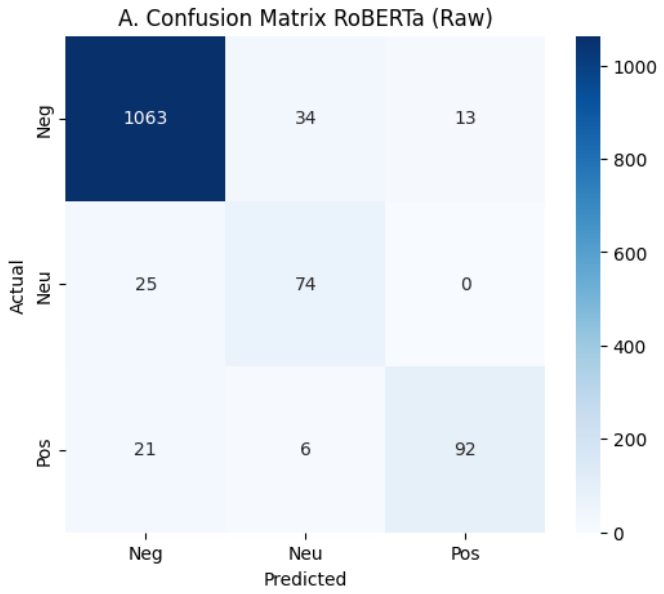
Berdasarkan hasil evaluasi, epoch ke-7 dapat dianggap sebagai titik optimal apabila pemilihan model didasarkan pada *validation accuracy*, karena pada epoch tersebut diperoleh nilai *validation accuracy* tertinggi, yaitu sekitar 0,945. Sementara itu, berdasarkan *validation loss*, nilai terendah diperoleh pada epoch ke-3 dengan nilai sekitar 0,21. Perbedaan kedua indikator tersebut menunjukkan

bahwa pemilihan epoch terbaik perlu disesuaikan dengan metrik yang digunakan sebagai dasar pemantauan selama proses *training*. Apabila penelitian menggunakan *EarlyStopping* dengan pemantauan *validation accuracy*, maka epoch ke-7 dapat digunakan sebagai acuan *best epoch*. Sebaliknya, apabila yang dipantau adalah *validation loss*, maka epoch ke-3 merupakan titik dengan performa validasi terbaik berdasarkan metrik tersebut.

Dengan demikian, model RoBERTa-BiLSTM menunjukkan kemampuan yang baik dalam mempelajari data pelatihan dan menghasilkan akurasi validasi yang relatif tinggi. Namun, adanya perbedaan pola antara metrik pelatihan dan validasi menunjukkan indikasi *overfitting*, sehingga penggunaan mekanisme *EarlyStopping* dan pemilihan bobot terbaik berdasarkan metrik validasi diperlukan untuk menjaga kemampuan generalisasi model terhadap data tweet politik Pemilu Indonesia.

4.5 Matriks Kebingungan (*Confusion Matrix*)

Hasil evaluasi model RoBERTa-BiLSTM terhadap data uji divisualisasikan menggunakan *confusion matrix* dalam dua bentuk, yaitu representasi absolut (*raw confusion matrix*) dan representasi proporsional (*normalized confusion matrix*). Penggunaan kedua representasi tersebut bertujuan untuk memberikan gambaran yang lebih menyeluruh mengenai kemampuan model dalam mengklasifikasikan setiap kelas sentimen, baik berdasarkan jumlah prediksi aktual maupun proporsi keberhasilan klasifikasi pada masing-masing kategori.



Gambar 4.2 Raw Confusion Matrix Model RoBERTa

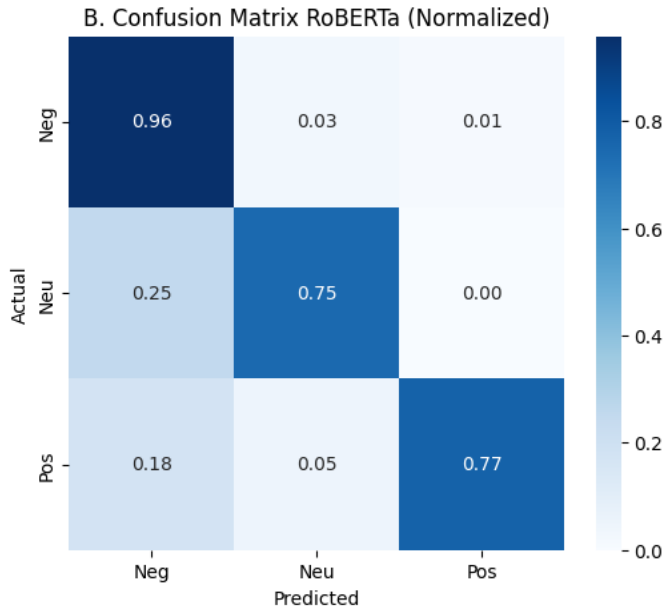
Berdasarkan raw confusion matrix yang ditampilkan pada Gambar A 4.2, model RoBERTa-BiLSTM menunjukkan performa klasifikasi yang baik pada seluruh kelas sentimen. Pada kelas negatif, model berhasil mengklasifikasikan 1063 data secara tepat sebagai negatif, sementara hanya 34 data salah diprediksi sebagai netral dan 13 data sebagai positif. Hasil ini menunjukkan tingkat klasifikasi yang tinggi pada kelas negatif dengan jumlah kesalahan prediksi yang relatif rendah dibandingkan total data yang diklasifikasikan.

Pada kelas netral, model mampu mengidentifikasi 74 data secara benar, dengan kesalahan prediksi sebanyak 25 data ke kelas negatif dan 0 data ke kelas positif. Meskipun kelas netral memiliki jumlah prediksi benar yang lebih rendah dibandingkan kelas negatif dan positif, model RoBERTa-BiLSTM tetap mampu mengklasifikasikan sebagian besar data netral dengan benar.

Sementara itu, pada kelas positif, model berhasil mengklasifikasikan 92 data dengan benar, sedangkan kesalahan prediksi tercatat sebanyak 21 data ke kelas

negatif dan 6 data ke kelas netral. Jumlah prediksi benar yang lebih dominan dibandingkan kesalahan menunjukkan bahwa model memiliki kemampuan yang baik dalam mengklasifikasikan sentimen positif pada data tweet politik.

Secara keseluruhan, distribusi pada raw confusion matrix memperlihatkan bahwa mayoritas data berhasil diprediksi secara tepat pada kelas masing-masing. Kesalahan klasifikasi yang muncul juga relatif lebih sedikit dibandingkan dengan jumlah prediksi yang benar. Hal ini menunjukkan bahwa model RoBERTa-BiLSTM mampu melakukan klasifikasi sentimen dengan baik pada data teks politik pemilu Indonesia.



Gambar 4.3 Normalized Confusion Matrix Model RoBERTa

Untuk memperoleh interpretasi performa yang lebih proporsional, dilakukan normalized confusion matrix sebagaimana ditunjukkan pada Gambar B 4.3. Pada kelas negatif, model memperoleh nilai klasifikasi benar sebesar 0,96, sedangkan kesalahan prediksi sebesar 0,03 ke kelas netral dan 0,01 ke kelas positif.

Nilai tersebut menunjukkan bahwa hampir seluruh data negatif mampu dikenali dengan sangat baik oleh model.

Pada kelas netral, model menghasilkan tingkat klasifikasi benar sebesar 0,75, dengan kesalahan sebesar 0,25 diprediksi sebagai negatif dan 0,00 sebagai positif. Hasil ini menunjukkan bahwa model tetap mampu mempertahankan stabilitas performa pada kelas netral yang secara karakteristik merupakan kelas dengan tingkat kompleksitas interpretasi tertinggi.

Adapun pada kelas positif, model juga memperoleh nilai prediksi benar sebesar 0,77, dengan kesalahan sebesar 0,18 ke kelas negatif dan 0,05 ke kelas netral. Konsistensi nilai performa pada kelas netral dan positif menunjukkan bahwa model mampu memahami distribusi konteks sentimen secara lebih seimbang tanpa mengalami dominasi prediksi pada kelas tertentu.

Hasil evaluasi tersebut menegaskan bahwa model RoBERTa-BiLSTM memiliki performa klasifikasi yang sangat unggul dibandingkan pendekatan berbasis representasi statis maupun model konvensional. Integrasi RoBERTa sebagai contextual embedding dengan BiLSTM sebagai mekanisme pemodelan dependensi sekuensial memungkinkan model memahami hubungan antar kata secara lebih mendalam dan kontekstual. RoBERTa mampu menghasilkan representasi semantik yang kaya berdasarkan konteks kalimat secara keseluruhan, sedangkan BiLSTM memperkuat kemampuan model dalam menangkap keterkaitan informasi dari arah maju dan mundur dalam urutan teks.

Kombinasi kedua pendekatan tersebut menjadikan model lebih adaptif dalam menangani karakteristik bahasa pada media sosial, khususnya tweet politik

yang sering mengandung ambiguitas makna, sindiran, campuran opini, serta penggunaan bahasa tidak baku. Hal ini tercermin dari tingginya nilai klasifikasi benar pada seluruh kelas sentimen dan rendahnya tingkat kesalahan prediksi antar kelas.

Dengan demikian, berdasarkan hasil confusion matrix baik dalam bentuk absolut maupun normalisasi, model RoBERTa-BiLSTM dapat dinyatakan sebagai model dengan performa terbaik dalam penelitian ini. Model tidak hanya mampu mencapai tingkat akurasi klasifikasi yang tinggi, tetapi juga menunjukkan stabilitas dan keseimbangan performa pada seluruh kelas sentimen, sehingga lebih efektif dalam merepresentasikan konteks semantik pada analisis sentimen data tweet politik. Dengan demikian, pendekatan ini memiliki potensi sebagai metode state-of-the-art untuk analisis sentimen teks berbahasa Indonesia.

4.6 Performa Model RoBERTa-BiLSTM

Berdasarkan hasil evaluasi, model RoBERTa-BiLSTM menunjukkan performa yang sangat baik dalam tugas analisis sentimen tweet politik Pemilu Indonesia. Model mencapai nilai accuracy sebesar 0.9255, dengan precision 0.8279, recall 0.8261, dan F1-score 0.8248.

Tingginya nilai accuracy menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan tepat. Selain itu, keseimbangan antara precision dan recall mengindikasikan bahwa model tidak hanya mampu menghasilkan prediksi yang akurat, tetapi juga efektif dalam mengidentifikasi data yang relevan pada masing-masing kelas sentimen.

Nilai precision yang sedikit lebih tinggi dibandingkan recall menunjukkan bahwa model memiliki kemampuan yang baik dalam menghasilkan prediksi positif yang tepat. Namun demikian, perbedaan yang relatif kecil antara precision dan recall menunjukkan bahwa model memiliki performa yang cukup seimbang dalam mengidentifikasi variasi sentimen. Kondisi ini juga mengindikasikan bahwa model mampu mempertahankan keseimbangan antara ketepatan prediksi dan kemampuan dalam menemukan data yang relevan pada masing-masing kelas sentimen, meskipun performanya dapat dipengaruhi oleh distribusi data yang tidak seimbang serta kompleksitas karakteristik bahasa pada media sosial.

Performa tersebut menegaskan bahwa kombinasi model RoBERTa-BiLSTM mampu menghasilkan representasi fitur yang lebih komprehensif. RoBERTa berperan dalam membangun contextual embedding yang mampu menangkap makna kata berdasarkan konteks, sedangkan BiLSTM memperkuat pemodelan melalui penangkapan dependensi sekuens dalam teks.

Integrasi kedua pendekatan ini memungkinkan model memahami pola bahasa pada tweet politik secara lebih mendalam, termasuk penggunaan bahasa informal, singkatan, dan struktur kalimat yang dinamis. Dengan demikian, model RoBERTa-BiLSTM dalam penelitian ini menunjukkan potensi yang kuat sebagai pendekatan berbasis deep learning modern untuk analisis sentimen teks berbahasa Indonesia. Hasil evaluasi model berdasarkan nilai precision, recall, F1-score, dan accuracy ditampilkan pada Tabel 4.3.

Tabel 4.3 Hasil Evaluasi Model RoBERTa-BiLSTM

Model	Precision	Recall	F1-score	Accuracy

RoBERTa-BiLSTM	0.8279	0.8261	0.8248	0.9255

Meskipun menunjukkan performa yang tinggi, evaluasi model tidak cukup hanya didasarkan pada satu pendekatan. Oleh karena itu, untuk menilai efektivitas arsitektur yang diusulkan secara lebih objektif, dilakukan analisis komparatif dengan model atau metode lain yaitu Word2Vec-BiLSTM dan FastText-BiLSTM. Melalui perbandingan tersebut, penelitian ini bertujuan untuk mengidentifikasi kontribusi dalam meningkatkan performa klasifikasi sentimen, khususnya pada data tweet politik pemilu di Indonesia

4.7 Hasil Ekstraksi Fitur Word2Vec

Sama dengan yang atas setelah melalui tahapan pra-pemrosesan, data teks dikonversi ke dalam representasi numerik menggunakan model Word2Vec yang dilatih pada kumpulan data tweet dalam penelitian ini. Setiap kata direpresentasikan sebagai vektor berdimensi 100 yang bersifat statis, sehingga memiliki nilai representasi yang tetap pada berbagai konteks kalimat.

Model Word2Vec mempelajari pola kemunculan kata dalam korpus untuk menangkap hubungan distribusional antar kata. Kata-kata yang sering muncul dalam konteks serupa akan memiliki kedekatan dalam ruang *embedding*, sehingga mampu merepresentasikan hubungan semantik dasar. Representasi ini selanjutnya dimanfaatkan oleh model BiLSTM untuk mempelajari pola sekuens dalam proses klasifikasi.

Namun demikian, karena bersifat statis, Word2Vec tidak mampu menangkap perubahan makna kata berdasarkan konteks kalimat. Keterbatasan ini menjadi signifikan pada teks media sosial yang sering mengandung ambiguitas,

ironi, serta penggunaan bahasa informal, sehingga berpotensi memengaruhi performa model dalam memahami sentimen. Keterbatasan ini menjadi salah satu alasan penggunaan pendekatan berbasis contextual embedding seperti RoBERTa dalam penelitian ini. Perbandingan data sebelum dan sesudah proses ekstraksi fitur Word2Vec ditampilkan pada Tabel 4.4.

Tabel 4.4 Perbandingan Data Sebelum dan Sesudah Ekstraksi Fitur RoBERTa

Sebelum Ekstraksi Fitur Word2Vec	Sesudah Ekstraksi Fitur Word2Vec
selamat ulang tahun ke78 prof dr hc megawati soekarnoputri ketua umum pdi perjuangan semoga senantiasa diberikan kesehatan kekuatan dan kebijaksanaan untuk terus menginspirasi serta memimpin pdi perjuangan dalam mewujudkan citacita indonesia raya	[0.107, 0.503, -0.354, ..., -0.493]
potong kuping saya jika ada pejabat wakanda tidak ada korupsi apalagi jokowi yg nafsu nya kemarin nepotisme dan kolusi iyu termasuk korup kali ah jengan begok njing buzzerp pejabat anjg	[0.226, 0.373, -0.372, ..., 0.109]
dasi nya cerah ya mas	[0.111, 0.246, -0.280, ..., 0.513]
gara gara ini mantan napi ini benci jokowi	[0.112, 0.244, -0.295, ..., 0.125]
wts wtb jual beli aplikasi premium netflix disney canva vidio capcut amazon prime getcontact youtube iqiyi viu bstation wetv loklok chatgpt youku order ke wa t lolly p diddy siang kamis nikmir fufufafa bali ariel tatum zonauang 26	[0.409, 0.216, -0.296, ..., 0.193]

Berdasarkan Tabel 4.4, setiap teks tweet yang telah melalui tahap pra-pemrosesan direpresentasikan ke dalam vektor numerik menggunakan model Word2Vec berdimensi 100. Representasi ini diperoleh melalui pembelajaran

distribusi kata pada korpus pelatihan, sehingga setiap kata memiliki vektor yang mencerminkan karakteristik semantik dasar dalam ruang *embedding*.

Hasil ekstraksi fitur menunjukkan bahwa setiap kalimat menghasilkan urutan vektor dengan panjang yang bervariasi, bergantung pada jumlah token dalam teks. Oleh karena itu, dilakukan proses *padding* untuk menyeragamkan panjang urutan menjadi ukuran tetap, sehingga data dapat diproses secara optimal oleh model BiLSTM.

Nilai-nilai numerik pada vektor *embedding*, seperti 0,107, 0,503, dan -0,354, merepresentasikan bobot hasil pembelajaran terhadap hubungan antar kata. Meskipun tidak memiliki makna linguistik secara langsung, nilai tersebut berfungsi sebagai representasi matematis yang memungkinkan model mengenali pola hubungan semantik selama proses klasifikasi.

Namun demikian, representasi Word2Vec bersifat statis, sehingga satu kata memiliki vektor yang sama pada berbagai konteks kalimat. Keterbatasan ini berbeda dengan pendekatan berbasis transformer seperti RoBERTa yang mampu menghasilkan representasi kontekstual sesuai dengan penggunaan kata dalam kalimat.

Implikasinya, Word2Vec memiliki keterbatasan dalam menangkap ambiguitas makna, ironi, singkatan, serta variasi bahasa informal yang umum ditemukan pada data media sosial. Dalam konteks ini, lapisan BiLSTM berperan penting karena mampu memodelkan dependensi sekuensial dari dua arah, sehingga sebagian informasi kontekstual tetap dapat ditangkap meskipun menggunakan *embedding*.

Dengan demikian, kombinasi Word2Vec-BiLSTM memungkinkan model untuk tetap mengenali pola sentimen secara efektif, khususnya dalam memahami hubungan antar kata dalam urutan teks, meskipun belum sepenuhnya mampu menangkap dinamika makna kontekstual secara mendalam. Keterbatasan ini selanjutnya menjadi dasar penggunaan model berbasis *contextual embedding* pada eksperimen berikutnya untuk meningkatkan kemampuan pemahaman konteks secara lebih komprehensif.

4.8 Evaluasi Proses Pelatihan Word2Vec-BiLSTM (*Learning Curves*)

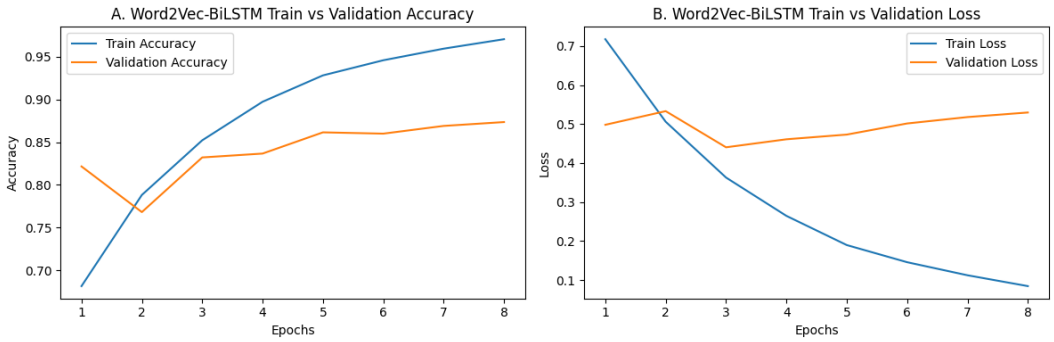
Evaluasi terhadap proses pelatihan model Word2Vec-BiLSTM dilakukan untuk menganalisis kemampuan model dalam mempelajari pola sentimen pada data tweet politik selama proses *training* berlangsung. Analisis ini bertujuan untuk mengetahui sejauh mana representasi *embedding* dari Word2Vec mampu menangkap karakteristik sentimen pada data latih serta mempertahankan kemampuan generalisasi terhadap data validasi. Selain itu, evaluasi ini juga digunakan untuk menilai stabilitas proses pembelajaran model dalam menghadapi karakteristik data teks media sosial yang cenderung dinamis, tidak terstruktur, dan memiliki variasi bahasa yang cukup kompleks.

Pada penelitian ini, proses pelatihan Word2Vec-BiLSTM divisualisasikan melalui *learning curves* yang menampilkan perkembangan nilai *accuracy* dan *loss* pada data *training* maupun *validation* di setiap epoch. Visualisasi tersebut memberikan gambaran yang lebih komprehensif mengenai dinamika pembelajaran model selama proses pelatihan berlangsung. Melalui kurva pembelajaran tersebut,

dapat diamati tingkat konvergensi model, kestabilan proses pembelajaran, serta indikasi terjadinya *overfitting* ataupun *underfitting*.

Analisis terhadap perubahan nilai *accuracy* dan *loss* juga digunakan untuk mengevaluasi efektivitas Word2Vec sebagai representasi kata dalam mendukung proses klasifikasi sentimen. Berbeda dengan pendekatan *contextual embedding* pada RoBERTa, Word2Vec menghasilkan representasi kata yang bersifat statis sehingga makna kata tidak berubah sesuai konteks kalimat. Keterbatasan tersebut memengaruhi kemampuan model dalam memahami hubungan semantik dan konteks opini pada tweet politik yang umumnya mengandung bahasa informal, singkatan, serta variasi ekspresi pengguna media sosial.

Dengan demikian, evaluasi proses pelatihan Word2Vec-BiLSTM tidak hanya bertujuan untuk melihat tingkat performa model, tetapi juga untuk memahami sejauh mana model mampu mempertahankan keseimbangan antara pembelajaran data latih dan kemampuan generalisasi terhadap data baru. Hasil visualisasi proses pelatihan model Word2Vec-BiLSTM ditampilkan pada Gambar 4.4.



Gambar 4.4 Kurva Training dan Validation Loss

Berdasarkan hasil pelatihan model pada Gambar 4.2 A, terlihat bahwa nilai *training accuracy* pada model Word2Vec-BiLSTM mengalami peningkatan secara bertahap selama proses pelatihan. Akurasi data latih meningkat dari sekitar 0,67 pada epoch pertama hingga mencapai sekitar 0,97 pada epoch ke-8. Peningkatan tersebut menunjukkan bahwa model mampu mempelajari pola dari data latih dengan semakin baik seiring bertambahnya epoch pelatihan.

Sementara itu, *validation accuracy* pada epoch pertama berada di sekitar 0,82, kemudian mengalami penurunan menjadi sekitar 0,77 pada epoch kedua. Setelah itu, nilai *validation accuracy* kembali meningkat secara bertahap hingga mencapai sekitar 0,88 pada epoch ke-8. Meskipun menunjukkan kecenderungan meningkat, peningkatan *validation accuracy* relatif lebih lambat dibandingkan *training accuracy*. Pada beberapa epoch, khususnya epoch ke-5 hingga epoch ke-8, nilai *validation accuracy* juga cenderung mengalami perubahan yang relatif kecil. Kondisi tersebut menunjukkan bahwa peningkatan kemampuan model dalam mempelajari data latih belum sepenuhnya diikuti oleh peningkatan performa pada data validasi.

Berdasarkan grafik *loss* pada bagian kanan Gambar 4.2 B, terlihat bahwa *training loss* mengalami penurunan secara konsisten dari sekitar 0,72 pada epoch pertama hingga mendekati 0,08 pada epoch ke-8. Penurunan tersebut menunjukkan bahwa proses optimisasi model berjalan dengan baik pada data pelatihan, sehingga kesalahan prediksi terhadap data latih semakin kecil seiring bertambahnya epoch.

Berbeda dengan *training loss*, pola *validation loss* menunjukkan perubahan yang tidak konsisten. *Validation loss* meningkat dari sekitar 0,50 pada epoch

pertama menjadi sekitar 0,54 pada epoch kedua, kemudian mengalami penurunan hingga mencapai nilai terendah sekitar 0,44 pada epoch ketiga. Setelah epoch ketiga, *validation loss* kembali meningkat secara bertahap hingga mencapai sekitar 0,53 pada epoch ke-8. Peningkatan *validation loss* setelah epoch ketiga menunjukkan bahwa model mulai mengalami penurunan kemampuan dalam melakukan generalisasi terhadap data yang tidak digunakan dalam proses pelatihan.

Perbedaan pola antara *training loss* dan *validation loss* tersebut menunjukkan adanya indikasi *overfitting*. Model semakin baik dalam mengenali pola pada data latih, yang ditunjukkan oleh meningkatnya *training accuracy* dan menurunnya *training loss*. Namun, kondisi tersebut tidak diikuti dengan peningkatan performa validasi yang sebanding. Gejala *overfitting* mulai terlihat setelah epoch ketiga, ketika *validation loss* mulai meningkat sementara *training loss* terus mengalami penurunan. Dengan demikian, semakin bertambahnya epoch pelatihan justru menyebabkan kesenjangan antara performa pada data latih dan data validasi semakin besar.

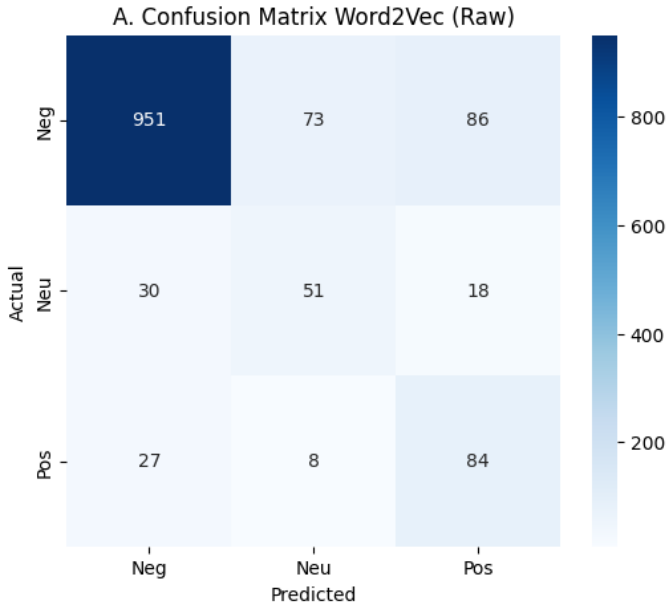
Jika dibandingkan dengan model RoBERTa-BiLSTM pada pengujian sebelumnya, model Word2Vec-BiLSTM menunjukkan performa validasi yang relatif lebih rendah dan pola generalisasi yang kurang optimal. Model RoBERTa-BiLSTM memiliki keunggulan dalam merepresentasikan konteks bahasa karena menggunakan *contextual embedding*, yaitu representasi kata yang mempertimbangkan konteks penggunaannya dalam suatu kalimat. Sebaliknya, Word2Vec menghasilkan representasi kata yang bersifat statis, sehingga satu kata cenderung memiliki representasi yang sama meskipun digunakan dalam konteks

yang berbeda. Perbedaan karakteristik representasi tersebut dapat memengaruhi kemampuan kedua model dalam memahami konteks sentimen pada data tweet politik Pemilu Indonesia.

Dengan demikian, berdasarkan hasil pelatihan pada Gambar 4.2, model Word2Vec-BiLSTM mampu mempelajari pola pada data latih dengan baik, yang ditunjukkan oleh peningkatan *training accuracy* hingga sekitar 0,97 dan penurunan *training loss* hingga sekitar 0,08 pada epoch ke-8. Namun, peningkatan tersebut tidak sepenuhnya diikuti oleh peningkatan performa pada data validasi. *Validation accuracy* hanya mencapai sekitar 0,88, sedangkan *validation loss* kembali meningkat setelah mencapai nilai terendah pada epoch ketiga. Kondisi tersebut menunjukkan adanya indikasi *overfitting* dan kemampuan generalisasi yang masih belum optimal. Berdasarkan perbandingan dengan model RoBERTa-BiLSTM pada pengujian sebelumnya, Word2Vec-BiLSTM memiliki keterbatasan dalam menangkap konteks bahasa secara dinamis pada data tweet politik, sehingga performanya secara keseluruhan masih berada di bawah model RoBERTa-BiLSTM.

4.9 Matriks Kebingungan (*Confusion Matrix*) Word2Vec-BiLSTM

Hasil evaluasi model Word2Vec-BiLSTM divisualisasikan menggunakan *confusion matrix* dalam dua bentuk, yaitu representasi absolut (*raw confusion matrix*) dan representasi proporsional (*normalized confusion matrix*). Pendekatan ini digunakan untuk memberikan gambaran yang lebih komprehensif mengenai distribusi hasil prediksi model serta kemampuan klasifikasi pada masing-masing kelas sentimen.



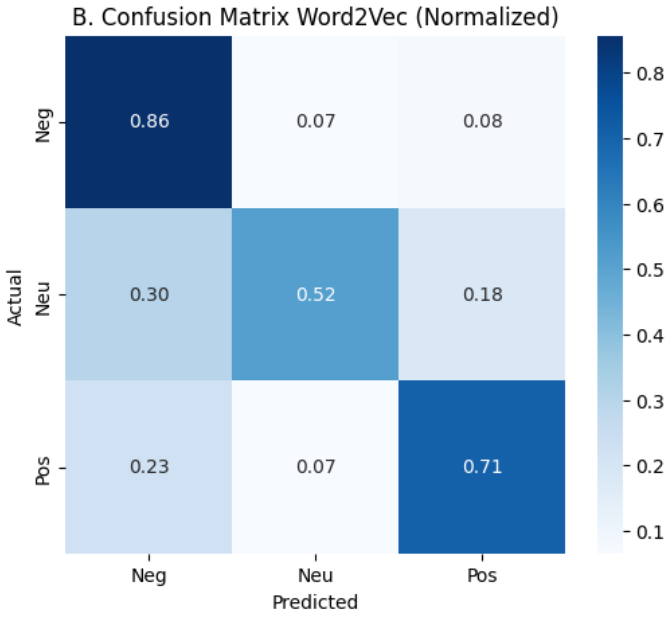
Gambar 4.5 Raw Confusion Matrix Model RoBERTa

Berdasarkan raw confusion matrix yang ditampilkan pada Gambar A 4.5, model Word2Vec-BiLSTM menunjukkan performa yang cukup baik dalam mengklasifikasikan kelas negatif. Dari keseluruhan data negatif, sebanyak 951 data berhasil diprediksi dengan benar sebagai negatif, sedangkan 73 data salah diprediksi sebagai netral dan 86 data diprediksi sebagai positif. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang cukup tinggi dalam mengenali pola sentimen negatif dibandingkan kelas lainnya.

Pada kelas netral, model berhasil mengklasifikasikan 51 data secara benar, sementara 30 data salah diprediksi sebagai negatif dan 18 data sebagai positif. Jumlah prediksi benar pada kelas netral masih lebih rendah dibandingkan kelas negatif, yang menunjukkan bahwa model mulai mengalami kesulitan dalam membedakan karakteristik sentimen netral yang cenderung ambigu dan memiliki kemiripan konteks dengan kelas lain.

Sementara itu, pada kelas positif, model mampu mengklasifikasikan 84 data secara tepat, dengan kesalahan sebanyak 27 data diprediksi sebagai negatif dan 8 data sebagai netral. Distribusi kesalahan tersebut memperlihatkan bahwa sebagian data positif masih cenderung diklasifikasikan ke kelas negatif, sehingga mengindikasikan adanya kecenderungan model dalam memusatkan prediksi pada kelas dominan tertentu.

Secara keseluruhan, distribusi pada raw confusion matrix menunjukkan bahwa performa model masih didominasi oleh kemampuan klasifikasi pada kelas negatif. Meskipun jumlah prediksi benar pada masing-masing kelas relatif lebih tinggi dibandingkan kesalahan klasifikasi, ketidakseimbangan distribusi kesalahan antar kelas menunjukkan bahwa model belum sepenuhnya mampu memahami konteks sentimen secara merata.



Gambar 4.6 Normalized Confusion Matrix Model RoBERTa

Untuk memperoleh interpretasi performa yang lebih proporsional, dilakukan normalized confusion matrix sebagaimana ditunjukkan pada Gambar B 4.6. Pada kelas negatif, model memperoleh nilai klasifikasi benar sebesar 0,86, sedangkan kesalahan prediksi sebesar 0,07 terjadi pada kelas netral dan 0,08 pada kelas positif. Nilai tersebut menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali sentimen negatif.

Pada kelas netral, model hanya memperoleh nilai prediksi benar sebesar 0,52, dengan kesalahan sebesar 0,30 diprediksi sebagai negatif dan 0,18 sebagai positif. Hasil ini menunjukkan bahwa sekitar sepertiga data netral masih mengalami kesalahan klasifikasi, terutama ke arah kelas negatif.

Adapun pada kelas positif, model juga menghasilkan nilai klasifikasi benar sebesar 0,71, dengan kesalahan sebesar 0,23 diprediksi sebagai negatif dan 0,07 sebagai netral. Kondisi tersebut memperlihatkan bahwa model masih mengalami kesulitan dalam membedakan sentimen positif dan netral terhadap sentimen negatif, khususnya pada data tweet politik yang memiliki kompleksitas bahasa tinggi.

Hasil normalisasi tersebut menegaskan bahwa model Word2Vec-BiLSTM memiliki performa yang cukup baik pada kelas negatif, namun performanya mengalami penurunan pada kelas netral dan positif. Tingginya proporsi kesalahan prediksi dari kedua kelas tersebut ke kelas negatif menunjukkan bahwa model masih memiliki kecenderungan bias terhadap kelas dominan.

Fenomena ini dapat dijelaskan melalui karakteristik Word2Vec yang menggunakan representasi embedding, sehingga makna kata direpresentasikan secara tetap tanpa mempertimbangkan perubahan konteks dalam kalimat. Pada data

tweet politik, satu kata dapat memiliki makna berbeda bergantung pada konteks penggunaan, terutama pada kalimat yang mengandung sindiran, ironi, maupun opini implisit. Dalam kondisi tersebut, Word2Vec cenderung lebih sulit memahami hubungan semantik secara mendalam dibandingkan pendekatan berbasis *contextual embedding*.

Meskipun integrasi BiLSTM mampu membantu model dalam menangkap hubungan dependensi sekuensial antar kata, keterbatasan representasi Word2Vec dalam memahami konteks menyebabkan performa klasifikasi antar kelas belum sepenuhnya seimbang. Hal ini tercermin dari masih tingginya kesalahan klasifikasi pada kelas netral dan positif.

Dengan demikian, hasil confusion matrix menunjukkan bahwa model Word2Vec-BiLSTM telah mampu melakukan klasifikasi sentimen dengan cukup baik, terutama pada kelas negatif. Namun, dibandingkan model RoBERTa-BiLSTM, performa model ini masih lebih rendah dalam memahami kompleksitas semantik pada data tweet politik berbahasa Indonesia.

4.10 Performa Model Word2Vec-BiLSTM

Berdasarkan hasil evaluasi, model Word2Vec-BiLSTM menunjukkan performa yang cukup baik dalam analisis sentimen tweet politik Pemilu Indonesia. Model mencapai nilai accuracy sebesar 0.8178, dengan precision 0.5922, recall 0.6926, dan F1-score 0.6289.

Dalam nilai accuracy menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan tepat. Selain itu, keseimbangan antara precision dan recall mengindikasikan bahwa model tidak hanya mampu menghasilkan prediksi yang

akurat, tetapi juga efektif dalam mengidentifikasi data yang relevan pada masing-masing kelas sentimen.

Nilai recall yang lebih tinggi dibandingkan precision menunjukkan bahwa model memiliki sensitivitas yang baik dalam mengenali variasi sentimen. Namun demikian, kondisi ini juga mengindikasikan adanya kecenderungan model untuk menghasilkan prediksi yang lebih inklusif terhadap kelas tertentu, yang dapat dipengaruhi oleh distribusi data yang tidak seimbang serta kompleksitas karakteristik bahasa pada media sosial.

Performa tersebut menegaskan bahwa kombinasi Word2Vec dan BiLSTM mampu menghasilkan representasi fitur yang cukup komprehensif. Word2Vec berperan dalam membangun representasi kata berdasarkan kemiripan semantik, sedangkan BiLSTM memperkuat pemodelan melalui penangkapan dependensi sekuens dalam teks.

Integrasi kedua pendekatan ini memungkinkan model memahami pola bahasa pada tweet politik secara lebih mendalam, termasuk penggunaan bahasa informal, singkatan, dan struktur kalimat yang dinamis. Namun demikian, performa model ini masih berada di bawah arsitektur hybrid RoBERTa-BiLSTM yang mampu menghasilkan nilai accuracy, precision, recall, dan F1-score yang lebih tinggi. Hal tersebut menunjukkan bahwa penggunaan contextual embedding pada RoBERTa lebih efektif dalam memahami konteks bahasa dibandingkan pendekatan embedding Word2Vec, sehingga model RoBERTa-BiLSTM dapat dianggap sebagai model terbaik dalam penelitian ini untuk analisis sentimen tweet politik

Pemilu Indonesia. Hasil evaluasi model berdasarkan nilai precision, recall, F1-score, dan accuracy ditampilkan pada Tabel 4.5.

Tabel 4.5 Hasil Evaluasi Model Word2Vec-BiLSTM

Model	Precision	Recall	F1-score	Accuracy
Word2Vec-BiLSTM	0.5922	0.6926	0.6289	0.8178

Integrasi kedua pendekatan ini memungkinkan model memahami pola bahasa pada tweet politik secara lebih mendalam, termasuk penggunaan bahasa informal, singkatan, dan struktur kalimat yang dinamis. Namun demikian, performa model ini masih berada di bawah arsitektur hybrid RoBERTa-BiLSTM yang mampu menghasilkan nilai accuracy, precision, recall, dan F1-score yang lebih tinggi. Hal tersebut menunjukkan bahwa penggunaan contextual embedding pada RoBERTa lebih efektif dalam memahami konteks bahasa dibandingkan pendekatan embedding Word2Vec, sehingga model RoBERTa-BiLSTM dapat dianggap sebagai model terbaik dalam penelitian ini untuk analisis sentimen tweet politik Pemilu Indonesia.

Pendekatan ini selanjutnya dievaluasi untuk mengkaji sejauh mana peningkatan performa yang dihasilkan dibandingkan dengan Word2Vec dalam konteks klasifikasi sentimen berbasis data media sosial.

4.11 Hasil Ekstraksi Fitur FastText

Setelah melalui tahap pra-pemrosesan data tweet dikonversi ke dalam representasi numerik menggunakan model FastText. Pada pendekatan ini, setiap

kata direpresentasikan sebagai vektor berdimensi 100 yang dihasilkan melalui mekanisme pembelajaran berbasis sub-kata (*subword n-grams*).

Berbeda dengan Word2Vec yang mempelajari representasi pada tingkat kata secara utuh, FastText mempertimbangkan struktur karakter penyusun kata. Pendekatan ini memungkinkan model menghasilkan representasi yang lebih adaptif terhadap variasi morfologi, kesalahan penulisan, singkatan, serta kata di luar kosakata (*out-of-vocabulary*). Karakteristik tersebut menjadikan FastText lebih sesuai untuk data media sosial seperti Twitter yang memiliki tingkat variasi bahasa yang tinggi.

Dalam implementasinya, setiap tweet direpresentasikan sebagai urutan vektor kata (*sequence embedding*) yang digunakan sebagai masukan pada model BiLSTM. Untuk kebutuhan visualisasi pada penelitian ini, dilakukan proses agregasi menggunakan metode *mean pooling* guna merangkum seluruh vektor kata dalam satu tweet menjadi satu vektor representatif berdimensi 100.

Proses agregasi tersebut hanya digunakan untuk keperluan penyajian hasil ekstraksi fitur pada tabel dan tidak memengaruhi mekanisme utama pembelajaran model, yang tetap memanfaatkan urutan vektor kata secara sekuensial. Perbandingan representasi data sebelum dan sesudah proses ekstraksi fitur FastText ditampilkan pada Tabel 4.6.

Tabel 4.6 Perbandingan Data Sebelum dan Sesudah Ekstraksi Fitur FastText

Sebelum Ekstraksi Fitur FastText	Sesudah Ekstraksi Fitur FastText
mau ngurut apa ituu coba kita tanya ke fufufafa	[0.0085, -0.0027, 0.0128, ..., 0.1097]
ini arahan megawati untuk 3214 anggota dprd fraksi pdip	[0.0166, 0.0120, 0.0173, ..., 0.0329]

Sebelum Ekstraksi Fitur FastText	Sesudah Ekstraksi Fitur FastText
dalam rangka memperingati hari ulang tahun ke78 ketua umum pdi perjuangan megawati soekarnoputri dpc pdi perjuangan kabupaten semarang menggelar kegiatan berbagi kebahagiaan dengan masyarakat sekitar kamis 2312025 sebagai wujud syukur dan penghormatan atas dedikasi ibu	[-0.0017, -0.0042, -0.0015, ..., 0.0420]
lalu siapa lagi yg ngisi polling ntuk nyangkutin masalah kesana kaum 58 mah udah pada move on dngn pilpres haya bani 16 aja yg masih ngotot amp nywrag jokowi sampe2 hasto nyamperin mantan kaesang ke singapore lhookoo ngihahajanaa	[0.0203, -0.0155, 0.0186, ..., 0.0262
menteri esdm selaku jubir jokowi	[-0.0034, -0.0190, -0.0008, ..., 0.0171]

Berdasarkan Tabel 4.6, setiap teks tweet berhasil ditransformasikan ke dalam representasi numerik berupa vektor berdimensi 100. Setiap dimensi pada vektor tersebut merepresentasikan fitur laten yang mengandung informasi leksikal dan semantik dari kata-kata penyusun teks. Representasi ini memungkinkan data teks yang semula tidak terstruktur diolah menjadi format numerik yang dapat diproses pada tahap pembelajaran mendalam (*deep learning*).

Pada konteks data tweet politik yang umumnya mengandung *noise*, singkatan, variasi bahasa informal, serta ketidakkonsistenan penulisan, pendekatan FastText menunjukkan keunggulan dibandingkan metode embedding. Hal ini disebabkan FastText tidak hanya mempelajari hubungan antar kata secara utuh, tetapi juga memanfaatkan informasi pada tingkat sub-kata (*subword*). Dengan mekanisme tersebut, model tetap mampu membentuk representasi yang relevan meskipun terdapat variasi morfologi, kesalahan ejaan, maupun kata di luar kosakata (*out-of-vocabulary*).

Perlu dicatat bahwa vektor tunggal yang ditampilkan pada tabel merupakan hasil agregasi menggunakan metode *mean pooling* yang digunakan khusus untuk keperluan visualisasi. Pada proses pelatihan, representasi FastText tetap digunakan dalam bentuk urutan vektor kata (*sequence embedding*) dan diberikan sebagai masukan ke lapisan BiLSTM, sehingga hubungan kontekstual dan dependensi sekuensial antar token tetap dapat dipertahankan.

Selain itu, lapisan embedding dikonfigurasi dalam kondisi *trainable*, sehingga bobot representasi FastText dapat diperbarui selama proses pelatihan (*fine-tuning*). Pendekatan ini memungkinkan representasi kata menyesuaikan karakteristik domain data tweet politik yang digunakan. Kombinasi antara representasi berbasis sub-kata dari FastText dan kemampuan BiLSTM dalam memodelkan dependensi dua arah antar token diharapkan mampu meningkatkan pemahaman konteks sentimen secara lebih efektif.

Secara keseluruhan, pendekatan FastText-BiLSTM mampu memberikan representasi tekstual yang lebih adaptif terhadap kompleksitas bahasa pada media sosial, khususnya pada data tweet politik yang memiliki tingkat variasi linguistik dan ketidakbakuan bahasa yang tinggi.

4.12 Evaluasi Proses Pelatihan FastText-BiLSTM (Learning Curves)

Evaluasi terhadap proses pelatihan model FastText-BiLSTM dilakukan untuk menganalisis efektivitas representasi *embedding* berbasis *subword* dalam mempelajari pola sentimen pada data tweet politik selama proses *training* berlangsung. Analisis ini bertujuan untuk mengevaluasi kemampuan model dalam memahami karakteristik data latih, terutama pada penggunaan kata tidak baku,

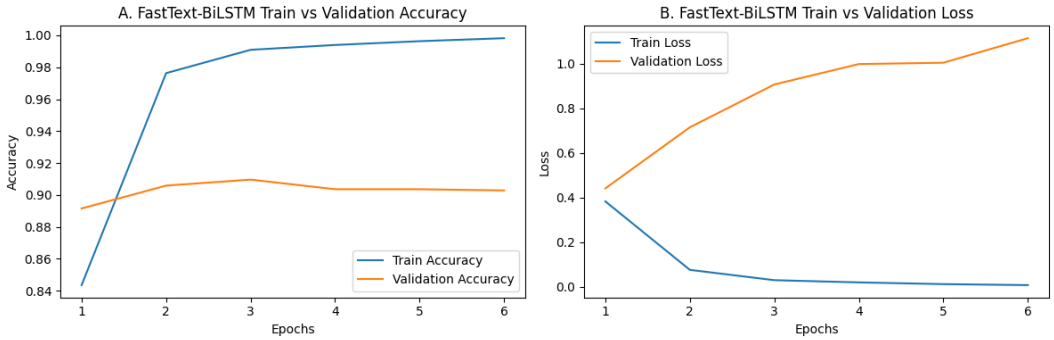
singkatan, kesalahan penulisan, serta variasi morfologi yang umum ditemukan pada teks media sosial. Selain itu, evaluasi ini juga dilakukan untuk menilai kemampuan generalisasi model terhadap data validasi selama proses pelatihan berlangsung.

Pada penelitian ini, proses pelatihan FastText-BiLSTM divisualisasikan menggunakan *learning curves* yang menampilkan perkembangan nilai *accuracy* dan *loss* pada data *training* maupun *validation* di setiap epoch. Visualisasi tersebut memberikan gambaran yang lebih jelas mengenai dinamika pembelajaran model, tingkat konvergensi, serta stabilitas performa model selama proses pelatihan. Melalui analisis kurva pembelajaran tersebut, dapat diamati indikasi terjadinya *overfitting* maupun *underfitting*, sekaligus mengevaluasi efektivitas pendekatan FastText dalam menangani kompleksitas bahasa pada tweet politik.

Berbeda dengan Word2Vec yang menggunakan representasi kata statis secara utuh, FastText memanfaatkan pendekatan berbasis *subword* sehingga model mampu memahami hubungan antar karakter dan potongan kata (*n-gram character*). Pendekatan ini memungkinkan model mengenali variasi kata, kata tidak baku, maupun istilah baru yang sering muncul pada media sosial. Oleh karena itu, FastText secara teoritis memiliki kemampuan representasi yang lebih baik dibandingkan *embedding* dalam menangani data teks yang bersifat informal dan dinamis.

Namun demikian, evaluasi proses pelatihan tetap diperlukan untuk mengetahui apakah peningkatan kemampuan representasi tersebut mampu menghasilkan generalisasi model yang lebih baik terhadap data validasi. Dengan demikian, analisis terhadap kurva *accuracy* dan *loss* menjadi penting untuk

memahami keseimbangan antara kemampuan model dalam mempelajari data latih dan mempertahankan performa pada data yang belum pernah dipelajari sebelumnya. Hasil visualisasi proses pelatihan model FastText-BiLSTM ditampilkan pada Gambar 4.7.



Gambar 4.7 Kurva Training dan Validation Loss

Berdasarkan hasil pelatihan model pada Gambar 4.7 A, terlihat bahwa nilai *training accuracy* pada model FastText-BiLSTM mengalami peningkatan yang sangat cepat sejak awal proses pelatihan. Akurasi data latih meningkat dari sekitar 0,84 pada epoch pertama hingga mendekati 1,00 pada epoch ke-6. Peningkatan tersebut menunjukkan bahwa model mampu mempelajari pola pada data latih dengan sangat baik dalam jumlah epoch yang relatif sedikit.

Namun, pola yang berbeda terlihat pada *validation accuracy*. Nilai akurasi validasi berada di sekitar 0,89 pada epoch pertama, kemudian meningkat hingga mencapai sekitar 0,91 pada epoch ketiga. Setelah itu, *validation accuracy* cenderung mengalami penurunan tipis dan relatif stagnan hingga epoch ke-6, dengan nilai akhir berada di sekitar 0,90. Kondisi tersebut menunjukkan bahwa peningkatan performa pada data latih tidak diikuti oleh peningkatan kemampuan generalisasi model terhadap data validasi. Dengan kata lain, model semakin baik

dalam mempelajari data latih, tetapi performanya pada data validasi tidak mengalami peningkatan yang sebanding.

Indikasi tersebut semakin diperkuat oleh grafik *loss* pada bagian kanan Gambar 4.7 B. Terlihat bahwa *training loss* mengalami penurunan secara drastis dari sekitar 0,38 pada epoch pertama hingga mendekati 0 pada epoch ke-6. Penurunan tersebut menunjukkan bahwa model semakin mampu menyesuaikan parameter terhadap data pelatihan sehingga kesalahan prediksi pada data latih menjadi sangat kecil.

Sebaliknya, *validation loss* menunjukkan pola yang berlawanan. Nilainya berada di sekitar 0,45 pada epoch pertama, kemudian meningkat menjadi sekitar 0,72 pada epoch kedua dan terus mengalami peningkatan hingga mencapai sekitar 1,11 pada epoch ke-6. Peningkatan *validation loss* yang berlangsung secara konsisten, sementara *training loss* terus menurun, menunjukkan bahwa model semakin menyesuaikan diri terhadap data latih tetapi semakin kurang mampu melakukan generalisasi terhadap data validasi.

Perbedaan pola antara *training loss* dan *validation loss* tersebut menunjukkan adanya indikasi *overfitting* yang cukup kuat. Gejala ini terlihat dari peningkatan *training accuracy* yang hampir mencapai 100% dan penurunan *training loss* hingga mendekati nol, sementara *validation accuracy* cenderung stagnan di sekitar 0,90 dan *validation loss* justru terus meningkat. Kondisi tersebut menunjukkan bahwa model mulai terlalu menyesuaikan diri dengan karakteristik data latih sehingga kemampuan untuk memprediksi data yang belum pernah dilihat sebelumnya menjadi kurang optimal.

Jika dibandingkan dengan model Word2Vec-BiLSTM dan RoBERTa-BiLSTM, model FastText-BiLSTM menunjukkan indikasi *overfitting* yang lebih kuat. Meskipun *training accuracy* mampu mencapai hampir 100%, peningkatan tersebut tidak diikuti oleh peningkatan *validation accuracy* yang sebanding. Selain itu, kenaikan *validation loss* yang terus berlangsung menunjukkan bahwa model kurang mampu mempertahankan performanya pada data validasi. Hal ini menunjukkan adanya kesenjangan yang cukup besar antara kemampuan model dalam mempelajari data latih dan kemampuan model dalam melakukan generalisasi.

Sementara itu, model RoBERTa-BiLSTM mampu mempertahankan keseimbangan yang lebih baik antara proses pembelajaran dan kemampuan generalisasi berdasarkan hasil pengujian sebelumnya. Keunggulan tersebut berkaitan dengan kemampuan RoBERTa dalam menghasilkan representasi kata yang bersifat *contextual embedding*, sehingga makna suatu kata dapat direpresentasikan berdasarkan konteks penggunaannya dalam kalimat. Sebaliknya, FastText menghasilkan representasi kata berdasarkan informasi subkata (*subword*) yang membantu menangani variasi bentuk kata, tetapi tidak memiliki kemampuan pemodelan konteks secara dinamis seperti model berbasis RoBERTa.

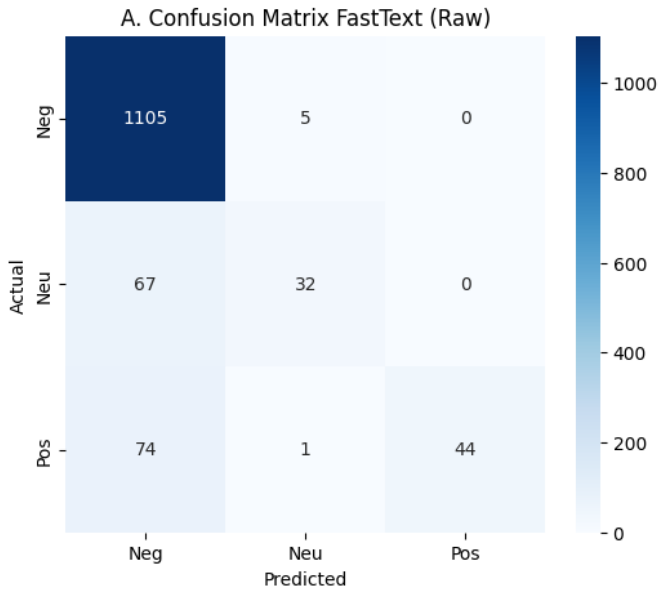
Dengan demikian, hasil pengujian menunjukkan bahwa FastText-BiLSTM mampu mempelajari data latih dengan sangat baik, tetapi mengalami *overfitting* yang cukup kuat. Hal tersebut ditunjukkan oleh *training accuracy* yang meningkat hingga mendekati 1,00 dan *training loss* yang menurun hingga mendekati nol, sedangkan *validation accuracy* hanya mencapai sekitar 0,91 dan kemudian

cenderung stagnan. Pada saat yang sama, *validation loss* terus meningkat hingga sekitar 1,11 pada epoch ke-6. Oleh karena itu, meskipun FastText-BiLSTM memiliki kemampuan pembelajaran yang cepat, kemampuan generalisasinya terhadap data validasi masih kurang optimal dibandingkan dengan model yang memiliki representasi konteks bahasa yang lebih kuat.

4.13 Matriks Kebingungan (*Confusion Matrix*) FastText-BiLSTM

Hasil evaluasi prediksi model FastText-BiLSTM divisualisasikan menggunakan *confusion matrix* dalam dua bentuk, yaitu representasi absolut (*raw confusion matrix*) dan representasi proporsional (*normalized confusion matrix*). Pendekatan ini digunakan untuk memberikan gambaran yang lebih komprehensif mengenai performa model, baik dari sisi jumlah prediksi maupun kemampuan klasifikasi pada masing-masing kelas.

Hasil evaluasi model FastText-BiLSTM terhadap data uji divisualisasikan menggunakan *confusion matrix* dalam dua bentuk, yaitu representasi absolut (*raw confusion matrix*) dan representasi proporsional (*normalized confusion matrix*). Pendekatan ini digunakan untuk memberikan gambaran yang komprehensif mengenai distribusi prediksi serta kemampuan klasifikasi model pada masing-masing kelas sentimen.

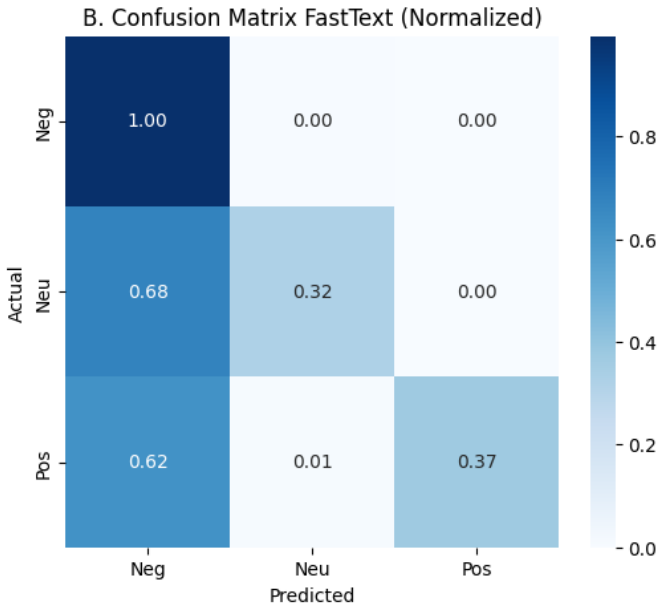


Gambar 4.8 Raw Confusion Matrix Model RoBERTa

Berdasarkan raw confusion matrix yang ditampilkan pada Gambar A 4.8, Berdasarkan *raw confusion matrix* yang ditampilkan pada Gambar A 4.8, untuk kelas negatif, model berhasil mengklasifikasikan 1.105 data secara benar, dengan 5 data salah diprediksi sebagai netral dan tidak terdapat data yang diprediksi sebagai positif. Pada kelas netral, model berhasil mengklasifikasikan 32 data secara benar, sementara 67 data salah diprediksi sebagai negatif dan tidak terdapat data yang diprediksi sebagai positif. Adapun pada kelas positif, model berhasil mengklasifikasikan 44 data secara benar, dengan 74 data salah diprediksi sebagai negatif dan 1 data salah diprediksi sebagai netral.

Distribusi tersebut menunjukkan bahwa kesalahan prediksi terbesar terjadi pada kelas netral dan positif yang masing-masing cenderung diklasifikasikan sebagai negatif. Kondisi ini menunjukkan adanya kecenderungan model dalam memprediksi data ke kelas negatif, yang dapat dipengaruhi oleh distribusi data

antarkelas yang tidak seimbang, di mana jumlah data pada kelas negatif jauh lebih besar dibandingkan kelas netral dan positif.



Gambar 4.9 Normalized Confusion Matrix Model FastText

Untuk memperoleh interpretasi performa yang lebih proporsional, dilakukan normalized confusion matrix sebagaimana ditunjukkan pada Gambar B 4.9. Pada kelas negatif, sebesar 1,00 data berhasil diklasifikasikan dengan benar, sementara 0,00 diprediksi sebagai netral dan 0,00 sebagai positif. Pada kelas netral, hanya 0,32 data yang diklasifikasikan dengan benar, sedangkan 0,68 salah diprediksi sebagai negatif dan 0,00 sebagai positif. Sementara itu, pada kelas positif, model menghasilkan 0,37 prediksi benar, dengan kesalahan sebesar 0,62 ke kelas negatif dan 0,01 ke kelas netral.

Hasil ini menegaskan bahwa model memiliki kemampuan yang sangat baik dalam mengenali kelas negatif, namun mengalami penurunan performa yang cukup signifikan pada kelas netral dan positif. Tingginya proporsi kesalahan dari kedua

kelas tersebut ke kelas negatif menunjukkan bahwa model cenderung menggeneralisasi prediksi ke kelas dominan.

Fenomena ini dapat dijelaskan melalui karakteristik linguistik pada data tweet politik yang sarat dengan ambiguitas semantik. Teks netral sering kali mengandung kata-kata berkonotasi negatif, sementara teks positif juga dapat memuat kritik terhadap pihak tertentu. Dalam kondisi ini, model FastText yang masih bersifat statis cenderung lebih sensitif terhadap kemunculan kata-kata berpolaritas negatif tanpa mempertimbangkan konteks secara menyeluruh.

Dengan demikian, meskipun FastText mampu menangani variasi morfologi dan kata di luar kosakata melalui pendekatan subword, model masih memiliki keterbatasan dalam memahami konteks semantik secara mendalam. Keterbatasan ini tercermin pada ketidakseimbangan performa antar kelas dan menjadi dasar perlunya pendekatan berbasis representasi kontekstual yang lebih adaptif.

4.14 Performa Model FastText-BiLSTM

Berdasarkan hasil evaluasi, model FastText-BiLSTM menunjukkan performa yang cukup baik dalam analisis sentimen tweet politik Pemilu Indonesia. Model mencapai nilai accuracy sebesar 0.8893, dengan precision 0.9096, recall 0.5628, dan F1-score 0.6484.

Nilai accuracy menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan tepat. Selain itu, nilai precision yang cukup tinggi mengindikasikan bahwa model mampu menghasilkan prediksi sentimen yang relatif akurat pada data yang teridentifikasi positif oleh model. Namun, nilai recall

yang lebih rendah menunjukkan bahwa masih terdapat sejumlah data sentimen yang belum berhasil dikenali dengan optimal.

Perbedaan antara precision dan recall menunjukkan bahwa model cenderung lebih selektif dalam melakukan klasifikasi sentimen. Kondisi ini dapat dipengaruhi oleh karakteristik distribusi data, variasi penggunaan bahasa pada media sosial, serta keterbatasan representasi embedding dalam memahami konteks kalimat secara menyeluruh.

Performa tersebut menegaskan bahwa kombinasi FastText dan BiLSTM mampu menghasilkan representasi fitur yang cukup baik. FastText berperan dalam merepresentasikan kata hingga tingkat subword sehingga mampu menangani kata-kata tidak baku atau variasi penulisan, sedangkan BiLSTM memperkuat pemodelan melalui penangkapan dependensi sekuens dalam teks. Hasil evaluasi model berdasarkan nilai precision, recall, F1-score, dan accuracy ditampilkan pada Tabel 4.8.

Tabel 4.7 Hasil Evaluasi Model FastText-BiLSTM

Model	Precision	Recall	F1-score	Accuracy
FastText-BiLSTM	0.9096	0.5628	0.6484	0.8893

Integrasi kedua pendekatan ini memungkinkan model memahami pola bahasa pada tweet politik secara lebih mendalam, termasuk penggunaan bahasa informal, singkatan, dan struktur kalimat yang dinamis. Namun demikian, performa model ini masih berada di bawah model RoBERTa-BiLSTM yang mampu menghasilkan nilai accuracy, precision, recall, dan F1-score yang lebih tinggi. Hal tersebut menunjukkan bahwa penggunaan contextual embedding pada RoBERTa

lebih efektif dalam memahami konteks bahasa dibandingkan pendekatan embedding konvensional FastText, sehingga model RoBERTa-BiLSTM dapat dianggap sebagai model terbaik dalam penelitian ini untuk analisis sentimen tweet politik Pemilu Indonesia.

Berdasarkan Berdasarkan hasil perbandingan keseluruhan model, arsitektur hybrid RoBERTa-BiLSTM menunjukkan performa terbaik dibandingkan model Word2Vec-BiLSTM dan FastText-BiLSTM dalam tugas analisis sentimen tweet politik Pemilu Indonesia. Model RoBERTa-BiLSTM berhasil memperoleh nilai accuracy sebesar 0.9255, precision 0.8279, recall 0.8261, dan F1-score 0.8248, dengan nilai accuracy, recall, dan F1-score yang lebih tinggi dibandingkan model lainnya.

Sementara itu, model Word2Vec-BiLSTM memperoleh accuracy sebesar 0.8178 dengan precision 0.5922, recall 0.6926, dan F1-score 0.6289, sedangkan FastText-BiLSTM memperoleh accuracy sebesar 0.8893 dengan precision 0.9096, recall 0.5628, dan F1-score 0.6484. Meskipun FastText-BiLSTM memiliki nilai precision yang cukup tinggi, nilai recall dan F1-score yang lebih rendah menunjukkan bahwa model tersebut belum mampu mengenali keseluruhan variasi sentimen secara optimal. Di sisi lain, Word2Vec-BiLSTM menunjukkan performa yang lebih stabil, namun masih belum mampu melampaui kemampuan contextual embedding yang dimiliki RoBERTa.

Hasil tersebut menunjukkan bahwa penggunaan contextual embedding pada RoBERTa memberikan kemampuan pemahaman konteks bahasa yang lebih baik dibandingkan embedding seperti Word2Vec dan FastText. Kombinasi model

RoBERTa-BiLSTM mampu menangkap makna kata berdasarkan konteks kalimat sekaligus memahami dependensi sekuens dalam teks secara lebih efektif.

Dengan demikian, model RoBERTa-BiLSTM dapat dinyatakan sebagai model terbaik dalam penelitian ini karena mampu memberikan performa paling unggul dan konsisten pada sebagian besar metrik evaluasi untuk analisis sentimen tweet politik Pemilu Indonesia. Berdasarkan hasil pengujian yang telah dilakukan, setiap model menunjukkan tingkat performa yang berbeda pada masing-masing metrik evaluasi. Perbandingan hasil evaluasi model Word2Vec-BiLSTM, FastText-BiLSTM, dan RoBERTa-BiLSTM berdasarkan nilai precision, recall, F1-score, dan accuracy dapat dilihat pada Tabel 4.8.

Tabel 4.8 Hasil Evaluasi Keseluruhan Model Klasifikasi Sentimen

Model	Precision	Recall	F1-Score	Accuracy
RoBERTa-BiLSTM	0.8279	0.8261	0.8248	0.9255
Word2Vec-BiLSTM	0.5922	0.6926	0.6289	0.8178
FastText-BiLSTM	0.9096	0.5628	0.6484	0.8893

Tingginya tingkat akurasi yang dihasilkan oleh model RoBERTa-BiLSTM dalam melakukan klasifikasi sentimen menunjukkan bahwa model dalam penelitian ini memberikan manfaat secara langsung, khususnya di tengah meningkatnya intensitas penyebaran informasi di media sosial. Media sosial saat ini tidak lagi hanya berfungsi sebagai sarana komunikasi dan pertukaran informasi, tetapi telah berkembang menjadi ruang publik digital yang memiliki pengaruh signifikan dalam membentuk opini, persepsi, dan perilaku politik penggunanya. Dalam konteks

Pemilu, berbagai isu politik dapat menyebar dengan sangat cepat dan menjangkau masyarakat secara luas melalui platform media sosial, sehingga respons publik terhadap isu yang berkembang menjadi semakin dinamis, masif, dan sering kali berlangsung tanpa proses penyaringan yang memadai.

Tingginya arus pertukaran informasi tersebut menghadirkan dua konsekuensi yang saling berlawanan. Di satu sisi, keterbukaan akses informasi mampu mendorong transparansi, meningkatkan partisipasi politik, serta memperluas ruang diskusi publik secara demokratis. Namun, di sisi lain, tingginya intensitas penyebaran informasi juga berpotensi memunculkan berbagai bentuk distorsi opini akibat bercampurnya narasi positif, kritik, provokasi, maupun sentimen negatif dalam satu ruang komunikasi digital. Kondisi ini semakin diperkuat oleh kecenderungan sebagian pengguna media sosial yang merespons berbagai konten yang beredar tanpa melakukan verifikasi yang memadai terhadap sumber maupun validitas informasi yang diterima. Akibatnya, polarisasi opini publik dapat berkembang lebih cepat dan memperlebar perbedaan pandangan di tengah masyarakat.

Dalam konteks tersebut, keberadaan sistem klasifikasi sentimen berbasis RoBERTa-BiLSTM menjadi penting sebagai instrumen analisis data untuk memahami dinamika opini publik secara lebih objektif. Model ini tidak hanya mampu mengidentifikasi polaritas sentimen masyarakat terhadap isu politik tertentu, tetapi juga dapat digunakan untuk memetakan kecenderungan persepsi publik secara sistematis dan terstruktur. Dengan kemampuan tersebut, hasil analisis sentimen berpotensi dimanfaatkan sebagai pendukung pengambilan keputusan,

evaluasi terhadap respons publik atas kebijakan maupun isu politik tertentu, serta sebagai sarana untuk meningkatkan literasi digital dalam menyikapi informasi politik secara lebih kritis dan rasional.

Selain mengidentifikasi polaritas sentimen, hasil analisis juga memungkinkan pemetaan distribusi sentimen masyarakat terhadap berbagai isu politik yang berkembang di media sosial. Informasi tersebut dapat memberikan gambaran mengenai potensi terbentuknya bias informasi dan pola penyebaran opini di ruang digital. Temuan ini menjadi penting untuk memahami bagaimana sentimen negatif maupun provokatif dapat berkembang dan mempengaruhi stabilitas diskursus publik di media sosial. Oleh karena itu, selain penguatan dari sisi teknologi analisis data, diperlukan pula peningkatan kesadaran kritis masyarakat dalam mengonsumsi serta menyebarkan informasi digital. Kesadaran tersebut tidak hanya dapat dibangun melalui pendekatan teknologi, tetapi juga melalui penguatan nilai-nilai etika dan moral yang mengarahkan masyarakat untuk menggunakan media sosial secara bertanggung jawab.

4.15 Integrasi Nilai-Nilai Islam

Dalam perspektif Islam, aktivitas menerima, mengolah, dan menyebarkan informasi merupakan bagian dari tanggung jawab moral yang harus dilakukan secara bijaksana. Islam mengajarkan pentingnya bersikap kritis, objektif, dan berhati-hati dalam menyikapi setiap informasi agar tidak menimbulkan kesalahpahaman, fitnah, maupun konflik di tengah masyarakat. Nilai-nilai tersebut semakin penting pada era media sosial, ketika informasi dapat diproduksi dan disebarkan dengan sangat cepat tanpa melalui proses verifikasi yang memadai.

Hasil penelitian menunjukkan bahwa model RoBERTa-BiLSTM mampu mengklasifikasikan sentimen masyarakat terhadap isu politik dengan tingkat akurasi yang baik. Kemampuan model ini dapat dimanfaatkan untuk membantu memahami dinamika opini publik secara lebih objektif. Hal tersebut sejalan dengan nilai-nilai Islam yang menekankan pentingnya kejujuran, keadilan, dan amanah dalam menyampaikan maupun mengolah informasi. Dalam penelitian ilmiah, nilai-nilai tersebut diwujudkan melalui upaya menghasilkan model yang mampu mengurangi bias, meningkatkan ketepatan klasifikasi, dan menyajikan hasil yang dapat dipertanggungjawabkan.

Dengan demikian, penerapan model RoBERTa-BiLSTM pada analisis sentimen tweet politik Pemilu Indonesia tidak hanya berfokus pada peningkatan performa model, tetapi juga mencerminkan penerapan nilai-nilai Islam dalam pengelolaan informasi. Integrasi tersebut dapat dilihat melalui tiga aspek, yaitu hubungan manusia dengan Allah (Hablum Minallah), hubungan manusia dengan sesama (Hablum Minannas), dan hubungan manusia dengan lingkungan (Hablum Minal 'Alam).

4.15.1 Hubungan dengan Allah (Hablum Minallah)

Dalam Islam, setiap aktivitas manusia, termasuk menerima, mengolah, menganalisis, dan menyebarkan informasi merupakan amanah yang akan dipertanggungjawabkan kepada Allah SWT. Oleh karena itu, penggunaan teknologi kecerdasan buatan dalam penelitian ini tidak hanya bertujuan memperoleh hasil klasifikasi yang akurat, tetapi juga menjadi bentuk tanggung jawab dalam menghasilkan informasi yang benar, objektif, dan dapat dipertanggungjawabkan.

Hasil penelitian menunjukkan bahwa model RoBERTa-BiLSTM mampu memetakan sentimen masyarakat terhadap isu politik secara lebih objektif. Kemampuan model dalam menghasilkan klasifikasi sentimen yang akurat sejalan dengan nilai-nilai Islam yang menekankan kejujuran, keadilan, dan amanah dalam menyampaikan informasi. Dalam konteks penelitian ilmiah, prinsip-prinsip tersebut terlihat dari upaya peneliti untuk menghasilkan model yang mampu mengurangi bias, meningkatkan ketepatan klasifikasi, serta menyajikan hasil analisis yang dapat dipertanggungjawabkan secara ilmiah.

Selain itu, Islam mengajarkan agar setiap informasi yang diterima tidak langsung dipercaya begitu saja tanpa didasarkan pada pengetahuan yang benar. Sikap tersebut penting diterapkan, terutama pada era media sosial ketika informasi dapat menyebar dengan sangat cepat dan belum tentu sesuai dengan fakta. Oleh karena itu, setiap muslim dituntut untuk bersikap hati-hati dan menggunakan akal sebelum menerima maupun menyebarkan suatu informasi.

Hal tersebut dijelaskan dalam firman Allah SWT pada QS. Al-Isra' ayat 36 sebagai berikut.

"Dan janganlah kamu mengikuti apa yang kamu tidak mempunyai pengetahuan tentangnya. Sesungguhnya pendengaran, penglihatan dan hati, semuanya itu akan diminta pertanggungjawabannya." (QS. Al-Isra' [17]:36).

Ayat tersebut menjelaskan bahwa setiap manusia tidak boleh mengikuti atau menyampaikan sesuatu tanpa memiliki pengetahuan yang benar. Pendengaran, penglihatan, dan hati yang disebutkan dalam ayat tersebut menunjukkan bahwa seluruh informasi yang diterima melalui pancaindra harus diproses dengan

pertimbangan yang baik sebelum diyakini ataupun disampaikan kepada orang lain. Dengan kata lain, setiap informasi harus didasarkan pada pengetahuan, bukti, dan proses verifikasi yang benar.

Nilai tersebut memiliki hubungan yang erat dengan penelitian ini. Penggunaan model RoBERTa-BiLSTM bertujuan membantu menghasilkan analisis sentimen yang lebih objektif melalui proses pengolahan data yang sistematis. Model tidak dibangun hanya untuk memperoleh akurasi yang tinggi, tetapi juga sebagai upaya menghasilkan informasi yang lebih dapat dipercaya sehingga mampu memberikan gambaran opini publik berdasarkan data yang tersedia.

Selain itu, penelitian ini juga menunjukkan bahwa kualitas data sangat memengaruhi hasil klasifikasi sentimen. Data yang bersih dan representatif akan menghasilkan model yang lebih baik dibandingkan data yang mengandung banyak bias maupun informasi yang tidak relevan. Oleh karena itu, proses pengolahan data, pembersihan data (*preprocessing*), hingga evaluasi model merupakan bentuk tanggung jawab peneliti dalam menjaga kualitas penelitian. Sikap tersebut sejalan dengan nilai amanah dalam Islam, yaitu melaksanakan pekerjaan secara jujur, teliti, dan bertanggung jawab.

Dengan demikian, hubungan manusia dengan Allah dalam penelitian ini diwujudkan melalui sikap amanah dalam mengelola data, kejujuran dalam melakukan penelitian, serta tanggung jawab dalam menyajikan hasil analisis yang objektif. Nilai-nilai tersebut menunjukkan bahwa pemanfaatan kecerdasan buatan tidak hanya berorientasi pada perkembangan teknologi, tetapi juga menjadi bagian

dari ikhtiar untuk menjalankan amanah ilmu pengetahuan sesuai dengan ajaran Islam.

4.15.2 Hubungan dengan Manusia (Hablum Minannas)

Islam mengajarkan bahwa hubungan antarmanusia harus didasarkan pada sikap jujur, adil, saling menghormati, dan tidak merugikan orang lain. Nilai-nilai tersebut juga perlu diterapkan dalam penggunaan media sosial, karena setiap informasi yang dibagikan dapat memengaruhi cara berpikir dan perilaku masyarakat. Hasil penelitian ini menunjukkan bahwa penyebaran hoaks, ujaran kebencian, kampanye hitam, serta polarisasi opini merupakan tantangan yang dapat menurunkan kualitas komunikasi di ruang digital.

Salah satu persoalan yang ditemukan dalam interaksi pengguna media sosial adalah penyebaran informasi tanpa melalui proses verifikasi yang memadai. Banyak pengguna yang langsung membagikan kembali (retweet atau share) informasi politik tanpa memastikan kebenaran sumbernya. Akibatnya, informasi yang belum tentu benar dapat menyebar dengan cepat dan memengaruhi opini masyarakat.

Penyebaran informasi secara impulsif tersebut sering menjadi penyebab meningkatnya sentimen negatif dan penyebaran hoaks selama masa pemilihan umum. Selain mempercepat penyebaran informasi yang belum terverifikasi, kondisi tersebut juga dapat membentuk persepsi masyarakat berdasarkan informasi yang belum tentu sesuai dengan fakta. Jika terus terjadi, masyarakat akan semakin sulit membedakan antara informasi yang benar dan informasi yang menyesatkan.

Dalam Islam, sikap tersebut bertentangan dengan prinsip tabayyun, yaitu memeriksa dan memastikan kebenaran suatu informasi sebelum mempercayai atau menyebarkannya. Prinsip ini dijelaskan dalam firman Allah SWT pada QS. Al-Hujurat ayat 6.

"Hai orang-orang yang beriman, jika datang kepadamu orang fasik membawa suatu berita, maka periksalah dengan teliti (tabayyun), agar kamu tidak menimpakan suatu musibah kepada suatu kaum tanpa mengetahui keadaannya yang menyebabkan kamu menyesal atas perbuatanmu itu." (QS. Al-Hujurat [49]:6).

Ayat tersebut mengajarkan bahwa setiap informasi harus diverifikasi terlebih dahulu agar tidak menimbulkan kerugian bagi orang lain. Dalam konteks media sosial, prinsip tabayyun dapat diterapkan dengan melakukan *fact-checking*, membandingkan informasi dari beberapa sumber yang terpercaya, dan tidak terburu-buru membagikan informasi yang belum jelas kebenarannya. Dengan demikian, penyebaran hoaks dan misinformasi dapat diminimalkan sehingga komunikasi digital menjadi lebih sehat.

Selain penyebaran hoaks, penelitian ini juga menunjukkan adanya fenomena echo chamber dan confirmation bias pada media sosial. Fenomena ini terjadi karena algoritma media sosial cenderung menampilkan informasi yang sesuai dengan minat dan pandangan pengguna. Akibatnya, seseorang lebih sering menerima informasi yang mendukung pendapatnya sendiri dibandingkan informasi yang memiliki sudut pandang berbeda.

Dalam jangka panjang, kondisi tersebut dapat memperkuat polarisasi masyarakat. Setiap kelompok cenderung menganggap pendapatnya sebagai yang

paling benar, sementara pendapat dari kelompok lain dianggap salah. Akibatnya, ruang diskusi menjadi kurang terbuka dan masyarakat semakin sulit menerima perbedaan pendapat.

Islam mengajarkan agar seseorang mau mendengarkan berbagai pendapat sebelum mengambil kesimpulan. Hal tersebut dijelaskan dalam QS. Az-Zumar ayat 18.

"Yang mendengarkan perkataan lalu mengikuti apa yang paling baik di antaranya. Mereka itulah orang-orang yang telah diberi petunjuk oleh Allah dan mereka itulah orang-orang yang mempunyai akal." (QS. Az-Zumar [39]:18).

Ayat tersebut menunjukkan bahwa seorang muslim hendaknya terbuka terhadap berbagai informasi, kemudian memilih pendapat yang paling baik berdasarkan pertimbangan yang rasional. Dalam konteks media sosial, nilai tersebut dapat diterapkan dengan tidak hanya membaca informasi yang sesuai dengan keyakinan pribadi, tetapi juga mempertimbangkan sudut pandang lain agar penilaian terhadap suatu isu menjadi lebih objektif.

Penelitian ini juga menemukan bahwa percakapan politik di media sosial masih didominasi oleh ujaran kebencian, penghinaan, serta kampanye hitam (*black campaign*). Islam telah memberikan pedoman yang jelas mengenai etika berkomunikasi. Allah SWT berfirman dalam QS. Al-Hujurat ayat 11.

"Hai orang-orang yang beriman, janganlah suatu kaum mengolok-olok kaum yang lain, boleh jadi mereka yang diperolok lebih baik daripada mereka. Dan janganlah kamu saling mencela dan janganlah kamu saling memanggil dengan gelar-gelar yang buruk...." (QS. Al-Hujurat [49]:11).

Ayat tersebut menegaskan bahwa seorang muslim tidak diperbolehkan menghina, mencela, ataupun memberikan gelar yang buruk kepada orang lain. Larangan tersebut sangat relevan dengan kondisi media sosial saat ini, di mana penghinaan dan pelabelan negatif sering muncul dalam percakapan politik. Apabila komunikasi dilakukan dengan bahasa yang santun dan menghargai orang lain, maka kualitas diskusi akan menjadi lebih baik.

Dari sisi penelitian, penggunaan bahasa yang lebih sopan dan berbasis argumentasi juga akan meningkatkan kualitas data yang digunakan dalam analisis sentimen. Data yang lebih bersih akan membantu model RoBERTa-BiLSTM mempelajari pola bahasa secara lebih baik sehingga hasil klasifikasi menjadi lebih akurat. Dengan demikian, etika komunikasi yang diajarkan dalam Islam tidak hanya memberikan manfaat bagi hubungan antarmanusia, tetapi juga mendukung kualitas pengolahan data pada penelitian ini.

Secara keseluruhan, hubungan manusia dengan sesama dalam penelitian ini diwujudkan melalui penerapan prinsip tabayyun, sikap terbuka terhadap berbagai pandangan, menghindari ujaran kebencian, serta membangun komunikasi digital yang jujur, santun, dan bertanggung jawab. Nilai-nilai tersebut diharapkan dapat menciptakan ruang diskusi yang lebih sehat sekaligus mendukung pengembangan teknologi kecerdasan buatan yang memberikan manfaat bagi masyarakat.

4.15.3 Hubungan dengan Lingkungan (Hablum Minal 'Alam)

Dalam perspektif Islam, hubungan manusia dengan lingkungan (Hablum Minal 'Alam) tidak hanya berkaitan dengan menjaga lingkungan fisik, tetapi juga mencakup lingkungan sosial dan lingkungan digital. Saat ini, media sosial telah

menjadi bagian dari kehidupan masyarakat sehingga ruang digital juga perlu dijaga agar tetap menjadi tempat yang sehat, aman, dan memberikan manfaat bagi semua orang. Oleh karena itu, penggunaan teknologi, termasuk kecerdasan buatan, hendaknya diarahkan untuk menciptakan kemaslahatan dan mengurangi dampak negatif yang dapat merugikan masyarakat.

Hasil penelitian menunjukkan bahwa penyebaran misinformasi, hoaks, polarisasi opini, dan fenomena *echo chamber* dapat menurunkan kualitas ruang publik digital. Informasi yang tidak benar dapat menyebar dengan cepat dan memengaruhi cara masyarakat memahami suatu peristiwa. Kondisi tersebut tidak hanya menimbulkan perbedaan pendapat, tetapi juga dapat memicu konflik sosial dan memperburuk kualitas diskusi di media sosial. Akibatnya, ruang digital yang seharusnya menjadi sarana berbagi informasi justru dipenuhi oleh informasi yang menyesatkan dan memecah belah masyarakat.

Selain itu, penelitian ini juga menunjukkan bahwa dinamika polarisasi di media sosial tidak hanya dipengaruhi oleh penyebaran informasi yang salah, tetapi juga oleh narasi yang sengaja dirancang untuk memengaruhi emosi masyarakat. Narasi tersebut sering kali memunculkan rasa marah, takut, atau cemas sehingga pengguna media sosial lebih mudah bereaksi tanpa terlebih dahulu memeriksa kebenaran informasi yang diterima. Fenomena ini dikenal sebagai computational propaganda, yaitu upaya memanfaatkan teknologi digital untuk membentuk opini publik melalui penyebaran informasi yang bersifat manipulatif.

Kondisi tersebut menunjukkan bahwa salah satu tantangan dalam menjaga kualitas lingkungan digital adalah kemampuan masyarakat untuk berpikir kritis.

Pengguna media sosial tidak cukup hanya mampu mengakses informasi, tetapi juga harus mampu menilai apakah informasi tersebut berasal dari sumber yang terpercaya, sesuai dengan fakta, dan memiliki konteks yang jelas. Dengan demikian, masyarakat tidak mudah terprovokasi oleh isu-isu yang sedang ramai diperbincangkan.

Islam mengajarkan agar manusia tidak mengikuti sesuatu tanpa memiliki pengetahuan yang benar. Nilai tersebut sebagaimana dijelaskan dalam QS. Al-Isra' ayat 36 menjadi pedoman bagi setiap muslim untuk menggunakan akal dan pertimbangannya sebelum mempercayai atau menyebarkan suatu informasi. Oleh karena itu, kemampuan berpikir kritis dan bersikap bijaksana dalam menggunakan media sosial merupakan bagian dari penerapan nilai-nilai Islam dalam kehidupan sehari-hari.

Dalam penelitian ini, kemampuan model RoBERTa-BiLSTM untuk mengidentifikasi sentimen masyarakat dapat membantu memberikan gambaran opini publik secara lebih objektif berdasarkan data yang tersedia. Informasi yang dihasilkan dari model tersebut dapat menjadi bahan pertimbangan dalam memahami dinamika sentimen masyarakat sehingga tidak hanya bergantung pada opini yang berkembang di media sosial. Dengan demikian, penggunaan model ini tidak hanya memberikan manfaat dari sisi teknis, tetapi juga mendukung terciptanya lingkungan digital yang lebih sehat, transparan, dan bertanggung jawab.

Selain itu, penelitian ini menunjukkan bahwa kualitas data memiliki peran penting dalam menghasilkan analisis sentimen yang akurat. Semakin bersih dan representatif data yang digunakan, semakin baik pula hasil klasifikasi yang

dihasilkan oleh model. Oleh karena itu, upaya mengurangi hoaks, ujaran kebencian, serta berbagai bentuk informasi yang menyesatkan tidak hanya bermanfaat bagi masyarakat, tetapi juga mendukung terciptanya data yang lebih berkualitas untuk pengembangan teknologi kecerdasan buatan.

Menjaga kualitas lingkungan digital merupakan salah satu bentuk ikhtiar dalam mewujudkan kemaslahatan (maslahah) bagi masyarakat. Lingkungan digital yang sehat akan mendorong terciptanya komunikasi yang lebih santun, penyebaran informasi yang lebih bertanggung jawab, serta pengambilan keputusan yang didasarkan pada informasi yang benar. Sebaliknya, apabila ruang digital dipenuhi oleh hoaks, ujaran kebencian, dan propaganda, maka dampaknya tidak hanya dirasakan oleh individu, tetapi juga dapat mengganggu kehidupan sosial secara lebih luas.

Dengan demikian, hubungan manusia dengan lingkungan dalam penelitian ini diwujudkan melalui upaya menjaga kualitas ekosistem informasi digital, meningkatkan literasi digital masyarakat, serta memanfaatkan teknologi kecerdasan buatan untuk menghasilkan informasi yang lebih objektif dan bermanfaat. Pemanfaatan model RoBERTa-BiLSTM tidak hanya bertujuan meningkatkan performa analisis sentimen, tetapi juga menjadi salah satu upaya untuk mendukung terciptanya ruang digital yang sehat, transparan, dan memberikan kemaslahatan bagi masyarakat sesuai dengan nilai-nilai Islam.

BAB V

PENUTUP

5.1 Kesimpulan

Penelitian ini telah mengkaji secara komprehensif efektivitas berbagai teknik representasi teks yang diintegrasikan ke dalam arsitektur Bidirectional Long Short-Term Memory (BiLSTM) pada tugas analisis sentimen data tweet politik Pemilu. Hasil penelitian menunjukkan bahwa penggunaan representasi bahasa berbasis konteks mampu meningkatkan performa klasifikasi secara signifikan dibandingkan pendekatan embedding. Peningkatan tersebut terlihat pada sebagian besar metrik evaluasi, yaitu Recall, F1-Score, dan Accuracy, sehingga menunjukkan bahwa pemahaman konteks kalimat memiliki peran penting dalam proses klasifikasi sentimen pada data media sosial yang kompleks dan tidak terstruktur.

Dalam komparasi antar model, pendekatan berbasis embedding masih menunjukkan keterbatasan dalam memahami konteks linguistik pada data tweet politik. Model Word2Vec-BiLSTM memperoleh nilai Precision sebesar 0.5922, Recall sebesar 0.6926, F1-Score sebesar 0.6289, serta Accuracy sebesar 0.8178. Hasil tersebut menunjukkan bahwa model belum mampu menghasilkan performa klasifikasi yang optimal dan seimbang pada seluruh kategori sentimen. Meskipun nilai Recall yang diperoleh relatif lebih tinggi dibandingkan FastText-BiLSTM, representasi embedding pada Word2Vec masih memiliki keterbatasan dalam menangkap hubungan semantik dan makna kontekstual antar kata. Kondisi ini

menjadi tantangan tersendiri, terutama pada data tweet politik yang banyak mengandung bahasa informal, singkatan, serta variasi konteks yang dinamis.

Sementara itu, model FastText-BiLSTM menunjukkan karakteristik performa yang berbeda dibandingkan Word2Vec-BiLSTM. Model ini memperoleh nilai Precision sebesar 0.9096, yang mengindikasikan kemampuan model dalam menghasilkan prediksi positif dengan tingkat ketepatan yang cukup baik. Namun demikian, nilai Recall yang dihasilkan hanya sebesar 0.5628, sehingga menjadi nilai terendah dibandingkan model lainnya. Ketimpangan antara Precision dan Recall tersebut berdampak pada nilai F1-Score yang hanya mencapai 0.6484, meskipun Accuracy model masih tergolong tinggi, yaitu sebesar 0.8893. Hasil tersebut menunjukkan bahwa pendekatan FastText masih belum mampu mengidentifikasi keseluruhan distribusi kelas sentimen secara optimal. Walaupun pendekatan subword pada FastText mampu membantu dalam menangani variasi kata dan istilah tidak baku pada media sosial, model ini masih memiliki keterbatasan dalam memahami konteks kalimat secara lebih mendalam dan dinamis.

Berbeda dengan kedua model sebelumnya, RoBERTa-BiLSTM menunjukkan performa yang paling optimal pada sebagian besar metrik evaluasi. Model ini berhasil memperoleh nilai Accuracy tertinggi sebesar 0.9255, disertai nilai Recall sebesar 0.8261 dan F1-Score sebesar 0.8248, yang juga merupakan capaian terbaik dibandingkan model lainnya. Selain itu, nilai Precision sebesar 0.8279 menunjukkan bahwa model memiliki tingkat keseimbangan performa yang sangat baik dalam proses klasifikasi sentimen. Hasil tersebut mengindikasikan

bahwa integrasi RoBERTa dengan BiLSTM mampu menghasilkan representasi fitur yang lebih kaya dan kontekstual, sehingga model dapat memahami makna kata berdasarkan konteks penggunaannya dalam kalimat secara lebih akurat.

Secara konseptual, keunggulan RoBERTa-BiLSTM menegaskan bahwa pendekatan berbasis pre-trained transformer lebih efektif dalam menangani kompleksitas bahasa alami dibandingkan pendekatan embedding seperti Word2Vec dan FastText. Representasi kontekstual yang dihasilkan RoBERTa memungkinkan model memahami perubahan makna kata sesuai konteks kalimat, sehingga proses klasifikasi sentimen dapat dilakukan secara lebih akurat dan stabil. Kemampuan tersebut menjadi sangat penting dalam analisis sentimen pada media sosial yang memiliki karakteristik bahasa informal, dinamis, serta sering dipengaruhi oleh konteks sosial dan politik yang terus berkembang.

Secara keseluruhan, penelitian ini memberikan bukti empiris bahwa pemanfaatan model pre-trained transformer, khususnya RoBERTa, mampu mengatasi keterbatasan pendekatan embedding dalam tugas klasifikasi teks. Temuan ini tidak hanya memberikan kontribusi akademik dalam pengembangan bidang Natural Language Processing (NLP), tetapi juga memiliki implikasi praktis dalam pengembangan sistem analisis sentimen yang lebih akurat, adaptif, dan relevan terhadap dinamika opini publik di media sosial. Dengan performa yang lebih stabil dan seimbang, model RoBERTa-BiLSTM berpotensi menjadi pendekatan yang efektif dalam mendukung pemetaan opini masyarakat, analisis isu politik, serta mitigasi polarisasi informasi di ruang digital.

5.2 Saran

Pertama, beranjak dari temuan komprehensif yang telah diuraikan, terdapat beberapa area krusial yang menuntut eksplorasi lebih mendalam bagi pengembangan riset di masa mendatang. Pertama, dari dimensi arsitektur komputasi, meskipun integrasi RoBERTa dan BiLSTM terbukti menghasilkan akurasi yang superior, kompleksitas parameter yang dilibatkan relatif masif. Oleh karena itu, penelitian lanjutan sangat direkomendasikan untuk menguji efisiensi varian transformer yang lebih ringan, seperti DistilRoBERTa atau ALBERT, guna menekan beban inferensi komputasi tanpa memicu degradasi performa yang signifikan. Di samping itu, penyetelan hyperparameter tingkat lanjut atau injeksi mekanisme Attention tambahan pada lapisan sekuensial berpotensi mengoptimalkan lebih jauh metrik F1-Score pada kelas klasifikasi yang bersinggungan.

Kedua, dari perspektif representasi data, validasi terhadap ketahanan model dalam menghadapi pergeseran domain linguistik (domain shift) menjadi esensial untuk diinvestigasi. Model tiga kategori klasifikasi yang telah dibangun perlu diuji silang menggunakan korpus teks yang lebih heterogen, baik lintas platform maupun lintas sektor industri. Penerapan teknik augmentasi data mutakhir berbasis generative language model juga patut dipertimbangkan untuk memperkaya variabilitas leksikal dan memastikan stabilitas pelatihan jangka panjang terhadap data yang tidak seimbang (imbalanced data).

Ketiga, pada ranah implementasi praktis dan operasionalisasi model, transisi dari fase pemodelan evaluatif menuju kerangka analisis sentimen real-time berbasis

streaming data akan memberikan nilai kontribusi yang jauh lebih substansial. Agenda riset ke depan dapat merancang arsitektur integrasi pemantauan performa prediktif secara langsung melalui visualisasi dinamis. Transformasi pipeline ini menjadi antarmuka yang mampu mendeteksi fluktuasi sentimen secara presisi di lingkungan produksi tidak hanya akan menyempurnakan landasan teoretis yang telah dicapai dalam studi ini, tetapi juga menghadirkan utilitas industri yang terukur dan berdampak nyata.

DAFTAR PUSTAKA

Alamsyah, A., & Sagama, Y. (2024). Empowering Indonesian internet users: An approach to counter online toxicity and enhance digital well-being. *Intelligent Systems with Applications*, *22*. https://doi.org/10.1016/j.iswa.2024.200394

Asokere, M., Wusu, A., & Olabanjo, O. (2025). Twitter (X) as an electoral barometer: systematic evidence from sentiment analysis of Twitter data. *International Journal of Information Technology (Singapore)*. https://doi.org/10.1007/s41870-025-03039-1

Carrasco, P., & Dias, S. (2024). Enhancing Restaurant Management through Aspect-Based Sentiment Analysis and NLP Techniques. *Procedia Computer Science*, *237*, 129–137. https://doi.org/10.1016/j.procs.2024.05.088

Chi, D., Huang, T., Jia, Z., & Zhang, S. (2025). Research on sentiment analysis of hotel review text based on BERT-TCN-BiLSTM-attention model. *Array*, *25*. https://doi.org/10.1016/j.array.2025.100378

Cicekyurt, E., & Bakal, G. (2025). Enhancing Sentiment Analysis in Stock Market Tweets Through BERT-Based Knowledge Transfer. *Computational Economics*. https://doi.org/10.1007/s10614-025-10901-8

Cui, L. (2025). Construction and implementation of knowledge enhancement pre-trained language model for text sentiment analysis. *Systems and Soft Computing*, *7*. https://doi.org/10.1016/j.sasc.2025.200293

Firdaus, A. A., Yudhana, A., Riadi, I., & Mahsun. (2024). Indonesian presidential election sentiment: Dataset of response public before 2024. *Data in Brief*, *52*. https://doi.org/10.1016/j.dib.2023.109993

Ghassabi, K., & Pahlevani, P. (2024). DEDUCT: A Secure Deduplication of Textual Data in Cloud Environments. *IEEE Access*, *12*, 70743–70758. https://doi.org/10.1109/ACCESS.2024.3402544

Hashemi, M. (2023). Geographical visualization of tweets, misinformation, and extremism during the USA 2020 presidential election using LSTM, NLP, and GIS. *Journal of Big Data*, *10*(1). https://doi.org/10.1186/s40537-023-00797-2

Henríquez, P. A., & Alessandri, F. (2024). Analyzing digital societal interactions and sentiment classification in Twitter (X) during critical events in Chile. *Heliyon*, *10*(12). https://doi.org/10.1016/j.heliyon.2024.e32572

Hussein, H. H., & Lakizadeh, A. (2025). A systematic assessment of sentiment analysis models on iraqi dialect-based texts. *Systems and Soft Computing*, *7*. https://doi.org/10.1016/j.sasc.2025.200203

Jahanbin, K., & Chahooki, M. A. Z. (2023). Aspect-Based Sentiment Analysis of Twitter Influencers to Predict the Trend of Cryptocurrencies Based on Hybrid Deep Transfer Learning Models. *IEEE Access*, *11*, 121656–121670. https://doi.org/10.1109/ACCESS.2023.3327060

Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. In *Natural Language Processing Journal* (Vol. 6). Elsevier Ltd. https://doi.org/10.1016/j.nlp.2024.100059

Khan, A., Boudjellal, N., Zhang, H., Ahmad, A., & Khan, M. (2023). From Social Media to Ballot Box: Leveraging Location-Aware Sentiment Analysis for Election Predictions. *Computers, Materials and Continua*, *77*(3), 3037–3055. https://doi.org/10.32604/cmc.2023.044403

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. http://arxiv.org/abs/1907.11692

Martinez-Espana, R., Fernandez-Pedauye, J., Lucia, J. G. P. de, Rojo-Martinez, J. M., Bakdid-Albane, K., & Garcia-Escribano, J. J. (2024). Methodology for Measuring Individual Affective Polarization Using Sentiment Analysis in Social Networks. *IEEE Access*, *12*, 102035–102049. https://doi.org/10.1109/ACCESS.2024.3431999

Marutho, D., Muljono, Rustad, S., & Purwanto. (2024). Optimizing aspect-based sentiment analysis using sentence embedding transformer, bayesian search clustering, and sparse attention mechanism. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(1). https://doi.org/10.1016/j.joitmc.2024.100211

Nabiilah, G. Z., Prasetyo, S. Y., Izdihar, Z. N., & Girsang, A. S. (2022). BERT base model for toxic comment analysis on Indonesian social media. *Procedia Computer Science*, *216*, 714–721. https://doi.org/10.1016/j.procs.2022.12.188

Nasir, J. A., Khan, O. S., & Varlamis, I. (2021). Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights*, *1*(1). https://doi.org/10.1016/j.jjimei.2020.100007

Nathan, F., Jacknius, Atha Rabbaanii, M. N., Hasani, M. F., & Maulina, A. (2025). Tracking Political Bias in U.S. News Media: A Daily Sentiment Analysis of the 2024 Presidential Election Using RoBERTa. *Procedia Computer Science*, *269*, 289–298. https://doi.org/10.1016/j.procs.2025.08.281

Nugroho, S. A., & Widianto, S. (2024). Exploring electric vehicle adoption in Indonesia using zero-shot aspect-based sentiment analysis. *Sustainable*

Operations and Computers, *5*, 191–205. https://doi.org/10.1016/j.susoc.2024.08.002

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, *89*, 14–46. https://doi.org/10.1016/j.knosys.2015.06.015

Rosenberg, E., Tarazona, C., Mallor, F., Eivazi, H., Pastor-Escuredo, D., Fuso-Nerini, F., & Vinuesa, R. (2023). Sentiment analysis on Twitter data towards climate action. *Results in Engineering*, *19*. https://doi.org/10.1016/j.rineng.2023.101287

Siino, M., Tinnirello, I., & La Cascia, M. (2024). Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers. *Information Systems*, *121*. https://doi.org/10.1016/j.is.2023.102342

Supriyadi, E., & Makatita, P. N. (2025). Sentiment Analysis of TikTok User Comments on QRIS Adoption in Indonesia Using IndoBERT. *Procedia Computer Science*, *269*, 121–130. https://doi.org/10.1016/j.procs.2025.08.265

Susanti, M. H., & Khu, S. (2025). The 2024 Indonesian presidential election in the accounts of millennials: A case study of Prabowo Subianto and Gibran Rakabuming Raka voters in Semarang, Central Java. *Social Sciences and Humanities Open*, *11*. https://doi.org/10.1016/j.ssaho.2025.101629

Tan, K. L., Lee, C. P., Anbananthen, K. S. M., & Lim, K. M. (2022). RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network. *IEEE Access*, *10*, 21517–21525. https://doi.org/10.1109/ACCESS.2022.3152828

Tanoto, K., Gunawan, A. A. S., Suhartono, D., Mursitama, T. N., Rahayu, A., & Ariff, M. I. M. (2024). Investigation of challenges in aspect-based sentiment analysis enhanced using softmax function on twitter during the 2024 Indonesian presidential election. *Procedia Computer Science*, *245*, 989–997. https://doi.org/10.1016/j.procs.2024.10.327

Valarmathi, B., Gupta, N. S., Karthick, V., Chellatamilan, T., Santhi, K., & Chalicheemala, D. (2024). Sentiment Analysis of Covid-19 Twitter Data using Deep Learning Algorithm. *Procedia Computer Science*, *235*, 3397–3407. https://doi.org/10.1016/j.procs.2024.04.320

Wu, H., Duan, C., & Wang, H. (2025). Research on Social Network Sentiment Analysis Based on Big Data. *Procedia Computer Science*, *261*, 135–140. https://doi.org/10.1016/j.procs.2025.04.181

Xiao, H., & Luo, L. (2024). An Automatic Sentiment Analysis Method for Short Texts Based on Transformer-BERT Hybrid Model. *IEEE Access*, *12*, 93305–93317. https://doi.org/10.1109/ACCESS.2024.3422268