

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Sejumlah penelitian sebelumnya telah mengkaji analisis sentimen dengan memanfaatkan data dari media sosial melalui berbagai pendekatan dan model. Beberapa studi menitikberatkan pada peningkatan kinerja model dalam proses klasifikasi sentimen, sedangkan penelitian lainnya lebih berfokus pada eksplorasi opini publik terhadap berbagai isu yang berkembang di masyarakat. Dalam hal ini, media sosial menjadi sumber data yang penting karena mampu mencerminkan pengalaman serta pandangan pengguna secara langsung dan aktual.

Sehingga penelitian yang secara khusus mengkaji perbandingan performa model *Bi-LSTM* dan *IndoBERT* terhadap opini publik mengenai *work-life balance*, terutama yang bersumber dari platform TikTok, masih relatif terbatas. Selain itu, integrasi analisis sentimen dengan konteks kondisi ekonomi juga belum banyak dieksplorasi [14]. Oleh karena itu, penelitian ini bertujuan untuk mengisi kekosongan tersebut melalui analisis komparatif kedua model dalam memahami opini publik terkait *work-life balance* di TikTok. Berikut tabel 2.1 penelitian terdahulu.



Tabel 2.1 Penelitian Terdahulu

Tahun	2022
Judul	<i>Sentiment analysis of Indonesian reviews using finetuning IndoBERT and R-CNN</i>
Permasalahan	Perkembangan platform digital menyebabkan jumlah ulasan (reviews) terhadap produk dan layanan semakin meningkat. Ulasan tersebut dapat dimanfaatkan untuk mengetahui opini masyarakat maupun sebagai bahan evaluasi bagi perusahaan. Namun, jumlah data ulasan yang sangat besar membuat proses analisis secara manual menjadi sulit, memerlukan waktu lama, dan tidak efisien. Oleh karena itu, diperlukan metode analisis sentimen otomatis yang mampu mengidentifikasi polaritas sentimen positif, netral, dan negatif secara akurat .
Metode	<i>IndoBERT, R-CNN</i>
Hasil	Penelitian membuktikan bahwa fine-tuning IndoBERT menggunakan RCNN menghasilkan performa yang lebih baik dibandingkan IndoBERT Base dalam analisis sentimen ulasan berbahasa Indonesia. Penggabungan RCNN mampu meningkatkan kemampuan model dalam menangkap konteks kalimat dan fitur penting pada teks sehingga menghasilkan akurasi sebesar 95,16% dan F1-Score sebesar 93,27%. Dengan demikian, penggunaan arsitektur hybrid antara pretrained language model (IndoBERT) dan deep learning classifier (RCNN) efektif untuk meningkatkan performa analisis sentimen bahasa Indonesia [15].

Relevansi dengan Penelitian yang dilakukan	Sama-sama membahas analisis sentimen pada data teks berbahasa Indonesia menggunakan model berbasis deep learning, khususnya IndoBERT. Penelitian tersebut menunjukkan bahwa IndoBERT mampu menghasilkan performa klasifikasi sentimen yang sangat baik dengan akurasi mencapai 95,16% setelah dikombinasikan dengan RCNN. Hasil tersebut membuktikan bahwa model berbasis transformer efektif dalam memahami konteks bahasa Indonesia dan mengidentifikasi polaritas sentimen positif, netral, dan negatif. Temuan ini mendukung penggunaan IndoBERT dalam penelitian ini untuk menganalisis sentimen. Selain itu, penelitian tersebut menjadi referensi penting dalam mengevaluasi kemampuan model deep learning pada tugas klasifikasi sentimen bahasa Indonesia.
--	---

Tahun	2025
Judul	<i>Twitter Social Media-Based Sentiment Analysis Using Bi-LSTM Method With Genetic Algorithm Optimization</i>
Permasalahan	Banyaknya opini masyarakat mengenai Pemilihan Presiden Indonesia 2024 di Twitter menyebabkan sulitnya mengidentifikasi sentimen secara manual. Selain itu, akurasi model <i>Bi-LSTM</i> dasar masih perlu ditingkatkan agar mampu mengklasifikasikan sentimen positif dan negatif dengan lebih baik. Penelitian ini juga berupaya mencari kombinasi metode ekstraksi fitur dan optimasi yang dapat meningkatkan performa klasifikasi sentimen.
Metode	<i>Bi-LSTM + Genetic Algorithm</i>

Hasil	Model <i>Bi-LSTM</i> dasar memperoleh akurasi terbaik sebesar 75,02%. Setelah ditambahkan TF-IDF, akurasi meningkat menjadi 81,72%. Penambahan <i>Word2Vec</i> meningkatkan akurasi menjadi 82,41%. Optimasi menggunakan <i>Genetic Algorithm</i> menghasilkan akurasi tertinggi sebesar 83%, meningkat 7,98% dibandingkan model baseline <i>Bi-LSTM</i> [15].
Relevansi dengan Penelitian yang dilakukan	Penelitian terdahulu menunjukkan bahwa <i>Bi-LSTM</i> efektif digunakan untuk analisis sentimen media sosial dan performanya dapat ditingkatkan melalui optimasi fitur dan parameter model. Penelitian ini relevan sebagai pembandingan penggunaan <i>Bi-LSTM</i> pada analisis sentimen komentar TikTok terkait <i>work-life balance</i> .

Tahun	2025
Judul	Analisis Sentimen Media Sosial Menggunakan Algoritma <i>BERT</i> dan <i>LSTM</i>
Permasalahan	Penelitian ini dilatarbelakangi oleh semakin banyaknya opini masyarakat yang disampaikan melalui media sosial. Data yang dihasilkan sangat besar, tidak terstruktur, dan mengandung konteks bahasa yang kompleks sehingga sulit dianalisis secara manual. Selain itu, metode analisis sentimen konvensional sering mengalami kesulitan dalam memahami makna kontekstual suatu kalimat dan hubungan antar kata dalam teks media sosial. Oleh karena itu, diperlukan model yang mampu memahami konteks bahasa secara lebih mendalam sekaligus menangkap ketergantungan antar kata dalam urutan teks .
Metode	<i>BERT</i> + <i>LSTM</i>

Hasil	Hasil penelitian menunjukkan bahwa kombinasi <i>BERT</i> dan <i>LSTM</i> menghasilkan performa yang lebih baik dibandingkan penggunaan satu model saja. Model hybrid <i>BERT-LSTM</i> mampu mencapai akurasi di atas 90% dalam klasifikasi sentimen media sosial berbahasa Indonesia. Penelitian ini membuktikan bahwa <i>BERT</i> efektif dalam memahami konteks semantik teks, sedangkan <i>LSTM</i> membantu menangkap hubungan antar kata dalam urutan kalimat sehingga meningkatkan kualitas prediksi sentimen [16].
Relevansi dengan Penelitian yang dilakukan	Penelitian terdahulu ini menggunakan pendekatan berbasis <i>Transformer</i> (<i>BERT</i>) dan <i>LSTM</i> untuk klasifikasi sentimen teks media sosial. Sedangkan, penelitian yang dilakukan yaitu membandingkan performa <i>Bi-LSTM</i> dan <i>IndoBERT</i> pada komentar TikTok bertopik <i>work-life balance</i> , sedangkan jurnal ini menggunakan model hybrid <i>BERT-LSTM</i> pada data media sosial secara umum.
Tahun	2025
Judul	<i>Use of Natural Language Processing in Social Media Text Analysis</i>

Permasalahan	Penelitian ini membahas penerapan Natural Language Processing (NLP) untuk menganalisis teks media sosial yang memiliki karakteristik tidak terstruktur, banyak menggunakan bahasa informal, singkatan, slang, emotikon, serta kesalahan tata bahasa. Kondisi tersebut menyebabkan analisis teks media sosial menjadi lebih kompleks dibandingkan teks formal. Oleh karena itu, penelitian membandingkan model machine learning klasik, yaitu Naïve Bayes dan Support Vector Machine (SVM), dengan model deep learning berbasis transformer, yaitu BERT, untuk mengetahui kemampuan masing-masing model dalam memahami konteks bahasa pada media sosial. Hasil penelitian menunjukkan bahwa model deep learning memiliki kemampuan yang lebih baik dalam memahami hubungan antar kata secara kontekstual sehingga mampu menangani variasi bahasa informal yang umum ditemukan pada media sosial. Penelitian juga menegaskan bahwa meskipun model deep learning membutuhkan sumber daya komputasi yang lebih besar, model tersebut memberikan hasil yang lebih akurat dibandingkan pendekatan machine learning tradisional dalam berbagai tugas NLP seperti analisis sentimen, klasifikasi teks, dan Named Entity Recognition (NER).
Metode	NLP

Hasil	Pada tugas analisis sentimen, model Naïve Bayes memperoleh akurasi sebesar 68%, SVM memperoleh akurasi sebesar 75%, sedangkan BERT mencapai akurasi sebesar 89%. Hasil ini menunjukkan bahwa BERT lebih unggul dalam mengenali pola emosi dan opini pada teks media sosial yang mengandung bahasa tidak baku, singkatan, maupun slang. Selain itu, pada tugas klasifikasi teks, kombinasi TF-IDF dan SVM menghasilkan akurasi sebesar 77%, LSTM mencapai 85%, sedangkan BERT memperoleh akurasi tertinggi sebesar 92% setelah proses fine-tuning. Pada tugas Named Entity Recognition (NER), model berbasis BERT kembali menunjukkan performa terbaik dengan akurasi 88%, lebih tinggi dibandingkan CRF yang memperoleh 65% dan model spaCy yang memperoleh 78%. Berdasarkan keseluruhan hasil pengujian, penelitian menyimpulkan bahwa BERT secara konsisten menghasilkan performa terbaik karena mampu memahami konteks kata secara lebih mendalam dibandingkan model machine learning maupun deep learning konvensional [17].
-------	---



Relevansi dengan Penelitian yang dilakukan	Penelitian ini relevan dengan penelitian yang dilakukan karena sama sama membahas analisis teks media sosial menggunakan pendekatan Natural Language Processing (NLP) serta menunjukkan keunggulan model berbasis BERT dalam memahami konteks bahasa yang kompleks dan tidak terstruktur. Hasil penelitian membuktikan bahwa BERT mampu menghasilkan performa yang lebih baik dibandingkan model machine learning klasik maupun model deep learning lainnya dalam analisis sentimen dan klasifikasi teks media sosial. Temuan tersebut mendukung penggunaan IndoBERT dalam penelitian, karena komentar TikTok juga memiliki karakteristik yang serupa dengan data media sosial pada penelitian ini, yaitu menggunakan bahasa informal, singkatan, ekspresi emosional, dan variasi bahasa yang tinggi. Selain itu, penelitian ini menjadi landasan teoritis bahwa model transformer seperti BERT memiliki kemampuan yang lebih baik dalam memahami konteks bahasa dibandingkan model sekuensial tradisional, sehingga relevan sebagai dasar perbandingan performa antara Bi-LSTM dan IndoBERT yang dilakukan dalam penelitian.
Tahun	2025
Judul	Implementasi <i>Bi-Directional Long Short Term Memory</i> Terhadap Klasifikasi Sentimen Di Twitter Pada Dataset Terbatas

Permasalahan	Twitter menghasilkan data opini dalam jumlah besar yang tidak terstruktur sehingga sulit dianalisis secara manual. Selain itu, metode klasifikasi sentimen konvensional sering mengalami kesulitan dalam memahami konteks kalimat karena hanya memproses teks dari satu arah. Oleh karena itu, diperlukan metode <i>Bi-LSTM</i> yang mampu memahami konteks kata dari arah <i>forward</i> dan <i>backward</i> untuk meningkatkan akurasi klasifikasi sentimen.
Metode	<i>Bi-LSTM</i>
Hasil	<i>Bi-LSTM</i> mampu menangkap konteks kalimat secara dua arah dan menghasilkan performa yang lebih baik dibandingkan <i>LSTM</i> standar. Pada berbagai penelitian Twitter, model <i>Bi-LSTM</i> menunjukkan peningkatan nilai <i>F1-Score</i> dan akurasi karena mampu memahami hubungan antar kata secara lebih baik. Model menghasilkan peningkatan <i>F1-score</i> hingga 67.77% [17].
Relevansi dengan Penelitian yang dilakukan	Penelitian ini relevan karena sama-sama menggunakan algoritma <i>Bi-LSTM</i> untuk melakukan klasifikasi sentimen pada data media sosial. Perbedaanannya, penelitian tersebut menggunakan data Twitter, sedangkan penelitian yang dilakukan menggunakan komentar TikTok mengenai work-life balance serta membandingkan performa <i>Bi-LSTM</i> dengan <i>IndoBERT</i> .

Tahun	2025
Judul	<i>Analysis of Public Sentiment Regarding the 2024 Jakarta Election on Platform X Using Deep Learning</i>
Permasalahan	Penelitian ini dilakukan karena analisis sentimen pada media sosial masih menghadapi berbagai tantangan, seperti penggunaan bahasa informal, teks yang tidak terstruktur, ambiguitas makna, serta ketidakseimbangan distribusi kelas sentimen. Selain itu, penelitian mengenai opini publik terkait Pilkada Jakarta masih terbatas sehingga diperlukan model yang mampu memahami konteks politik lokal secara lebih baik. Oleh karena itu, penelitian membandingkan empat model deep learning untuk mengetahui model yang paling efektif dalam mengklasifikasikan sentimen masyarakat terhadap Pilkada Jakarta 2024. Hasil pengujian menunjukkan bahwa setiap model memiliki karakteristik yang berbeda. IndoBERT unggul dalam memahami konteks kata karena menggunakan arsitektur transformer, LSTM mampu menangkap hubungan jangka panjang dalam teks, CNN efektif mengekstraksi pola kata dan frasa penting, sedangkan GRU memiliki kemampuan yang baik dalam mengelola informasi sekuensial dengan kompleksitas yang lebih rendah dibanding LSTM. Namun, seluruh model masih mengalami kesulitan dalam mengklasifikasikan sentimen netral akibat dominasi jumlah data netral yang mencapai 70,30% dari keseluruhan dataset.
Metode	<i>IndoBERT, Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), dan Gated Recurrent Unit (GRU)</i>

Hasil	Berdasarkan hasil evaluasi, model CNN memperoleh performa terbaik dengan akurasi sebesar 83,37%, diikuti oleh LSTM sebesar 82,61%, GRU sebesar 82,30%, dan IndoBERT sebesar 80,77%. Pada kelas sentimen negatif, GRU menghasilkan precision tertinggi sebesar 95,94%, sedangkan CNN memperoleh F1-Score tertinggi sebesar 90,51%. Pada kelas sentimen positif, GRU kembali menunjukkan performa terbaik dengan recall sebesar 82,45% dan F1-Score sebesar 76,35%. Sementara itu, pada kelas sentimen netral seluruh model menunjukkan performa yang lebih rendah dibandingkan kelas lainnya karena adanya ketidakseimbangan data dan tingginya ambiguitas ekspresi netral. Secara keseluruhan, penelitian menyimpulkan bahwa CNN merupakan model yang paling efektif untuk analisis sentimen Pilkada Jakarta 2024 karena mampu mengidentifikasi pola kata dan frasa penting secara lebih baik dibanding model lainnya [18].
Relevansi dengan Penelitian yang dilakukan	Penelitian ini relevan dengan penelitian karena sama-sama melakukan analisis sentimen media sosial menggunakan pendekatan deep learning dan klasifikasi tiga kelas sentimen (positif, netral, negatif). Selain itu, penelitian ini juga menggunakan IndoBERT dan LSTM, yang merupakan model yang sejenis dengan model yang digunakan, yaitu IndoBERT dan Bi-LSTM. Hasil penelitian ini menunjukkan bahwa performa model deep learning dapat berbeda tergantung karakteristik dataset dan topik yang dianalisis.

Tahun	2023
Judul	<i>Sentiment Analysis of Indonesian TikTok Review Using LSTM and IndoBERTweet Algorithm</i>
Permasalahan	TikTok merupakan salah satu platform media sosial yang memiliki jutaan pengguna aktif dan menghasilkan ulasan dalam jumlah besar setiap harinya. Banyaknya ulasan tersebut menyulitkan perusahaan maupun peneliti untuk memahami persepsi pengguna secara manual. Selain itu, ulasan TikTok umumnya menggunakan bahasa informal, singkatan, dan bahasa gaul yang membuat proses klasifikasi sentimen menjadi lebih kompleks. Oleh karena itu, diperlukan metode analisis sentimen otomatis yang mampu memahami karakteristik bahasa Indonesia yang digunakan dalam media sosial.
Metode	<i>LSTM, IndoBERTweet</i>
Hasil	Penelitian membandingkan performa algoritma <i>LSTM</i> dan <i>IndoBERTweet</i> dalam melakukan klasifikasi sentimen terhadap ulasan TikTok. Hasil penelitian menunjukkan bahwa <i>IndoBERTweet</i> menghasilkan akurasi yang lebih tinggi dibandingkan <i>LSTM</i> karena memiliki kemampuan memahami konteks bahasa Indonesia secara lebih mendalam melalui proses pre-training pada korpus bahasa Indonesia. Sementara itu, <i>LSTM</i> masih mampu melakukan klasifikasi sentimen dengan baik, namun memiliki keterbatasan dalam memahami konteks yang kompleks dibandingkan model berbasis <i>Transformer</i> . Hasil penelitian membuktikan bahwa model <i>Transformer</i> lebih unggul dalam tugas analisis sentimen media sosial berbahasa Indonesia [19].

Relevansi dengan Penelitian yang dilakukan	Penelitian ini sangat relevan karena menggunakan data TikTok dan membandingkan model berbasis <i>LSTM</i> dengan model berbasis <i>BERT</i> . Penelitian tersebut menjadi referensi penting dalam pemilihan model <i>IndoBERT</i> dan <i>Bi-LSTM</i> pada penelitian yang dilakukan.
--	--

Tahun	2025
Judul	<i>Implementation of BiLSTM and IndoBERT for Sentiment Analysis of TikTok Reviews</i>
Permasalahan	TikTok menjadi salah satu aplikasi media sosial dengan jumlah pengguna yang sangat besar sehingga menghasilkan ulasan dan komentar dalam jumlah yang terus meningkat. Banyaknya ulasan tersebut menyulitkan perusahaan dalam memahami persepsi pengguna terhadap layanan yang diberikan. Analisis secara manual membutuhkan waktu yang lama dan rentan terhadap subjektivitas. Selain itu, karakteristik bahasa pada ulasan TikTok yang tidak terstruktur dan sering menggunakan bahasa informal menyebabkan proses klasifikasi sentimen menjadi lebih kompleks. Oleh karena itu, diperlukan metode yang mampu mengklasifikasikan sentimen secara otomatis dengan tingkat akurasi yang tinggi.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Hasil	Penelitian mengimplementasikan <i>Bi-LSTM</i> dan <i>IndoBERT</i> untuk melakukan klasifikasi sentimen terhadap ulasan TikTok. Hasil penelitian menunjukkan bahwa kedua model mampu melakukan klasifikasi sentimen dengan baik, namun kombinasi fitur yang dihasilkan oleh <i>IndoBERT</i> dan kemampuan <i>Bi-LSTM</i> dalam mempelajari hubungan sekuensial menghasilkan performa terbaik. Model memperoleh rata-rata akurasi <i>cross-validation</i> sekitar 92% serta menunjukkan kemampuan yang baik dalam membedakan sentimen positif, netral, dan negatif. Penelitian menyimpulkan bahwa integrasi model <i>Transformer</i> dan <i>Recurrent Neural Network</i> (RNN) dapat meningkatkan kualitas klasifikasi sentimen pada data media sosial [20].
Relevansi dengan Penelitian yang dilakukan	Penelitian ini menggunakan platform TikTok serta model <i>Bi-LSTM</i> dan <i>IndoBERT</i>. Perbedaan terletak pada objek penelitian yang dilakukan yang berfokus pada komentar TikTok terkait <i>work-life balance</i> sehingga dapat mengisi celah penelitian pada topik tersebut.

Tahun	2026
Judul	<i>IndoBERT-Sentiment: Context-Conditioned Sentiment Classification for Indonesian Text</i>

Permasalahan	Sebagian besar model analisis sentimen hanya berfokus pada kata-kata yang muncul dalam suatu kalimat tanpa memperhatikan konteks pembahasan secara menyeluruh. Padahal, dalam bahasa Indonesia, satu kata atau kalimat dapat memiliki makna sentimen yang berbeda tergantung pada konteks topik yang sedang dibahas. Kondisi ini menyebabkan banyak model analisis sentimen menghasilkan klasifikasi yang kurang akurat ketika menghadapi teks yang ambigu atau memiliki makna kontekstual yang kuat. Oleh karena itu, diperlukan model yang mampu memahami hubungan antara konteks pembahasan dan sentimen yang terkandung dalam suatu teks.
Metode	<i>IndoBERT</i>
Hasil	Penelitian menunjukkan bahwa <i>IndoBERT</i> yang dikondisikan dengan konteks pembahasan mampu meningkatkan performa klasifikasi sentimen dibandingkan pendekatan yang hanya memanfaatkan informasi teks secara langsung. Model dapat memahami makna kalimat yang ambigu serta menyesuaikan interpretasi sentimen berdasarkan topik yang sedang dibahas. Hasil evaluasi menunjukkan peningkatan akurasi, <i>precision</i> , <i>recall</i> , dan <i>F1-score</i> dibandingkan beberapa model analisis sentimen bahasa Indonesia lainnya. Dengan memanfaatkan representasi kontekstual yang lebih kaya, model mampu menghasilkan klasifikasi sentimen yang lebih konsisten dan akurat [21].

Relevansi dengan Penelitian yang dilakukan	Penelitian ini relevan karena menunjukkan keunggulan <i>IndoBERT</i> dalam memahami konteks bahasa Indonesia yang kompleks. Kemampuan tersebut penting dalam penelitian analisis sentimen komentar TikTok mengenai <i>work-life balance</i> karena banyak komentar mengandung opini yang bersifat subjektif dan kontekstual. Penelitian ini mendukung penggunaan <i>IndoBERT</i> sebagai salah satu model yang dibandingkan dalam penelitian.
--	--

Tahun	2025
Judul	Analisis Sentimen Ulasan Aplikasi TikTok menggunakan <i>Convolutional Neural Network</i>
Permasalahan	permasalahan banyaknya ulasan pengguna aplikasi TikTok di Google Play Store yang sulit dianalisis secara manual karena jumlahnya sangat besar. Sebanyak 19.569 ulasan berhasil dikumpulkan dan setelah melalui tahapan preprocessing berupa cleaning text, stopword removal, tokenisasi, padding, dan representasi fitur, diperoleh 11.368 data yang siap digunakan untuk proses klasifikasi sentimen. Penelitian membandingkan empat metode klasifikasi, yaitu Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Support Vector Machine (SVM), dan Logistic Regression (LR), guna mengetahui model yang paling efektif dalam mengidentifikasi sentimen pengguna TikTok. Hasil analisis menunjukkan bahwa pendekatan deep learning, khususnya CNN, memiliki kemampuan yang lebih baik dalam memahami pola lokal antar kata dan karakteristik teks pendek yang umum ditemukan pada ulasan media sosial dibandingkan metode machine learning tradisional.

Metode	CNN
Hasil	Hasil evaluasi menunjukkan bahwa model CNN memperoleh performa terbaik dibandingkan model lainnya dengan nilai Accuracy sebesar 98,95%, Macro-F1 sebesar 94,07%, dan Weighted-F1 sebesar 98,97%. Sementara itu, model LSTM memperoleh Accuracy sebesar 98,47%, SVM sebesar 97,82%, dan Logistic Regression sebesar 94,69%. Selain memiliki akurasi tertinggi, CNN juga menunjukkan kemampuan yang lebih baik dalam menangani distribusi kelas yang tidak seimbang serta mampu mengklasifikasikan berbagai kategori ulasan secara lebih konsisten. Berdasarkan classification report, CNN menghasilkan nilai precision, recall, dan F1-score yang sangat tinggi pada hampir seluruh kategori ulasan, khususnya pada kategori User Experience, Feature Request, dan Performance. Temuan ini membuktikan bahwa arsitektur CNN sangat efektif dalam mengekstraksi fitur lokal dari teks pendek dan informal yang banyak ditemukan pada ulasan pengguna TikTok [22].

Relevansi dengan Penelitian yang dilakukan	sama-sama membahas analisis sentimen pada data yang berasal dari platform TikTok dan menggunakan pendekatan deep learning untuk mengklasifikasikan opini pengguna. Penelitian tersebut menunjukkan bahwa model deep learning mampu menghasilkan performa yang sangat baik dalam memahami sentimen pada teks pendek dan informal yang menjadi karakteristik utama data media sosial. Selain itu, penelitian ini juga menggunakan klasifikasi sentimen positif, negatif, dan netral yang sejalan dengan penelitian saya. Perbedaannya, penelitian ini berfokus pada ulasan aplikasi TikTok di Google Play Store dengan model CNN, LSTM, SVM, dan Logistic Regression. Oleh karena itu, penelitian ini menjadi referensi penting yang menunjukkan efektivitas model deep learning dalam tugas analisis sentimen pada data TikTok, sekaligus menjadi dasar untuk mengevaluasi apakah model yang lebih modern seperti IndoBERT dapat memberikan performa yang lebih baik dibandingkan pendekatan deep learning lainnya.
--	---

Tahun	2025
Judul	Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma <i>IndoBERT</i> pada Platform Media Sosial

Permasalahan	Kemunculan berbagai isu publik di media sosial sering memunculkan beragam tanggapan masyarakat yang mengandung sentimen positif, negatif, maupun netral. Jumlah data yang sangat besar membuat proses analisis opini publik secara manual menjadi tidak efisien dan membutuhkan waktu yang lama. Selain itu, opini masyarakat yang disampaikan melalui media sosial sering kali menggunakan bahasa informal dan memiliki konteks yang beragam sehingga menyulitkan proses klasifikasi. Oleh karena itu, diperlukan metode berbasis <i>NLP</i> dan <i>Deep Learning</i> yang mampu melakukan analisis sentimen secara otomatis terhadap data media sosial.
Metode	<i>IndoBERT</i>
Hasil	Penelitian berhasil menerapkan <i>IndoBERT</i> untuk mengklasifikasikan sentimen publik terkait isu Danantara pada berbagai platform media sosial. Hasil penelitian menunjukkan bahwa <i>IndoBERT</i> mampu memahami konteks bahasa Indonesia dengan baik dan menghasilkan performa klasifikasi yang tinggi. Analisis sentimen yang dilakukan mampu menggambarkan kecenderungan opini masyarakat terhadap isu yang sedang berkembang, baik dalam bentuk dukungan, kritik, maupun sikap netral. Hasil tersebut membuktikan bahwa model Transformer seperti <i>IndoBERT</i> sangat efektif digunakan untuk analisis sentimen pada data media sosial berbahasa Indonesia [23].

Relevansi dengan Penelitian yang dilakukan	Penelitian ini relevan karena sama-sama menggunakan <i>IndoBERT</i> untuk menganalisis sentimen masyarakat pada media sosial. Meskipun topik yang diteliti berbeda, penelitian ini menunjukkan efektivitas <i>IndoBERT</i> dalam memahami opini publik berbahasa Indonesia. Temuan tersebut dapat menjadi dasar teoritis dalam penggunaan <i>IndoBERT</i> pada penelitian analisis sentimen komentar TikTok mengenai <i>work-life balance</i> .
--	---

Data pada Tabel 2.1 menunjukkan bahwa penelitian terdahulu telah banyak menerapkan algoritma *Bi-LSTM* maupun *IndoBERT* pada berbagai topik analisis sentimen di media sosial, seperti kebijakan publik, layanan digital, dan fenomena sosial. Namun, penelitian yang secara khusus membahas sentimen komentar TikTok mengenai *work-life balance* masih terbatas. Selain itu, sebagian besar penelitian hanya berfokus pada penggunaan satu model klasifikasi sehingga belum memberikan gambaran yang komprehensif mengenai kemampuan *Bi-LSTM* dan *IndoBERT* dalam memahami opini masyarakat terkait *work-life balance* [24]. Oleh karena itu, penelitian ini dilakukan untuk menganalisis sentimen komentar TikTok tentang *work-life balance* menggunakan algoritma *Bi-LSTM* dan *IndoBERT* guna mengetahui kemampuan kedua model dalam mengklasifikasikan sentimen pada data berbahasa Indonesia.

Penelitian yang mengkaji topik keseimbangan kerja-hidup di media sosial tetap lebih banyak menekankan pada pengidentifikasian persepsi atau eksplorasi tema diskusi tanpa melakukan analisis sentimen secara mendalam dengan pendekatan *deep learning*. Sebagai hasilnya, data tentang kecenderungan sentimen publik terhadap topik *work-life balance* di platform media sosial, terutama TikTok, masih terbatas.

Penerapan analisis sentimen dengan algoritma *Bi-LSTM* dan *IndoBERT* telah dilakukan pada berbagai objek penelitian, termasuk pendapat publik di Twitter, ulasan aplikasi, isu kebijakan publik, dan berbagai fenomena sosial yang muncul di media sosial. Beberapa studi menunjukkan bahwa *Bi-LSTM* efektif dalam

mempelajari hubungan sekuensial antar kata, sementara *IndoBERT* lebih unggul dalam memahami konteks bahasa Indonesia dengan pendekatan *Transformer*. Temuan dari studi sebelumnya juga menunjukkan bahwa kedua algoritma ini dapat memberikan hasil klasifikasi sentimen yang baik pada teks berbahasa Indonesia.

Sebagian besar penelitian sebelumnya menggunakan data dari Twitter, ulasan aplikasi, atau media sosial secara umum, sementara penelitian yang secara khusus memanfaatkan komentar TikTok sebagai sumber data masih cukup jarang. TikTok adalah salah satu platform media sosial dengan pengguna yang sangat banyak di Indonesia dan menjadi tempat utama bagi masyarakat untuk mengungkapkan pendapat tentang berbagai masalah, termasuk masalah ketenagakerjaan dan keseimbangan kerja-hidup.

Sebagian besar juga studi sebelumnya hanya memanfaatkan satu model klasifikasi atau membandingkan model dengan arsitektur yang berbeda tanpa secara khusus menilai kinerja *Bi-LSTM* dan *IndoBERT* dalam konteks *work-life balance*. Walaupun, kedua model itu memiliki sifat yang berbeda dalam menganalisis teks. *Bi-LSTM* lebih efektif dalam menangkap keterkaitan sekuensial antara kata-kata, sementara *IndoBERT* memiliki kemampuan untuk memahami konteks bahasa Indonesia yang lebih rumit melalui mekanisme perhatian mandiri. Oleh karena itu, masih belum jelas model mana yang lebih efisien dalam mengklasifikasikan sentimen komentar TikTok mengenai keseimbangan kerja-hidup.

2.2 Teori yang berkaitan

2.2.1 Analisis Sentimen

Analisis sentimen adalah teknik yang digunakan untuk menganalisis dan memahami opini, perasaan, atau sentimen yang terkandung dalam teks yang dihasilkan oleh pengguna di berbagai platform digital, seperti media sosial, forum, blog, atau ulasan produk [25]. Tujuan dari analisis sentimen adalah untuk mengekstrak opini yang terkandung dalam teks tersebut, yang kemudian dapat digunakan untuk membuat keputusan yang lebih baik dalam berbagai bidang, seperti bisnis, pemasaran, kebijakan pemerintah, atau analisis sosial [26]. Proses ini berfungsi untuk mengolah data teks yang tidak terstruktur menjadi informasi yang

lebih terorganisir dan mudah dipahami. Proses ini juga sering disebut sebagai opinion mining, yang merupakan salah satu cabang dalam studi ini [13].

Metode ini digunakan untuk menilai pandangan dan emosi individu yang tersirat dalam tulisan. Meskipun kedua istilah ini memiliki tujuan yang serupa, terdapat perbedaan di antara keduanya. *Opinion mining* lebih berfokus pada pandangan konkret seseorang terhadap suatu hal, yang melibatkan pemahaman terhadap satu pendapat atau sekelompok pendapat. Sementara itu, analisis sentimen lebih menekankan eksplorasi perasaan, emosi, opini yang terdapat dalam teks, serta mengklasifikasikan apakah emosi tersebut bersifat positif, negatif, atau netral. Proses ini melibatkan pengumpulan komentar/opini dari TikTok [27]. Dengan demikian, analisis sentimen tidak hanya membantu dalam mengidentifikasi apakah teks tersebut mengandung opini, tetapi juga menggali lebih dalam terkait konteks dan sentimen yang terkandung di dalamnya .

Dalam melakukan analisis sentimen ini memanfaatkan teknik pengolahan bahasa alami untuk mengenali pola-pola dalam teks yang menunjukkan sentimen positif, negatif, atau netral. Proses ini melibatkan analisis terhadap beberapa komponen penting, yaitu subjek (topik yang dibicarakan), polaritas (apakah sentimen tersebut positif, negatif, atau netral). Analisis sentimen tidak hanya membantu dalam mengidentifikasi apakah teks tersebut mengandung opini, tetapi juga menggali lebih dalam terkait aspek dan sentimen yang terkandung di dalamnya.

2.2.2 Media Sosial



Gambar 2.1 Media Sosial

Media sosial adalah bagian dari perkembangan internet yang berfungsi sebagai platform digital bagi penggunanya untuk berinteraksi, memperluas jaringan pertemanan, mencari dan berbagi informasi, mengunggah konten dalam format teks, video pendek, dan juga foto, serta menyampaikan opini dan mengekspresikan pandangan terhadap isu yang terkait [28]. Salah satu media sosial yang berkembang secara signifikan yang dapat digunakan untuk menyampaikan opini adalah TikTok [29].

2.2.3 Work Life Balance



Gambar 2.2 Work Life Balance

Gambar 2.3 *Work-life balance* merupakan suatu kondisi yang menunjukkan kemampuan individu dalam mengelola dan menyeimbangkan tanggung jawab pekerjaan dengan kehidupan pribadinya secara seimbang. Keseimbangan tersebut memungkinkan seseorang untuk menjalankan kedua peran tanpa mengalami konflik yang berlebihan antara tuntutan pekerjaan dan kebutuhan pribadi. Konsep *work-life balance* semakin mendapat perhatian seiring dengan perkembangan teknologi, perubahan sistem kerja, serta meningkatnya tuntutan profesional yang membuat batas antara kehidupan kerja dan kehidupan pribadi menjadi semakin sulit dipisahkan. Apabila keseimbangan tersebut tidak dapat tercapai, individu berisiko mengalami berbagai dampak negatif, seperti menurunnya kesehatan fisik dan mental, berkurangnya produktivitas kerja, serta menurunnya kualitas hidup secara keseluruhan.

Menurut Greenhaus dan Allen (2011), *work-life balance* adalah tingkat dimana seseorang dapat menjalankan peran dalam pekerjaan dan kehidupan pribadi secara efektif serta memperoleh tingkat kepuasan yang seimbang pada kedua peran tersebut. Definisi ini menekankan bahwa *work-life balance* tidak hanya berkaitan dengan pembagian waktu antara pekerjaan dan kehidupan pribadi, tetapi juga mencakup keterlibatan psikologis, komitmen, dan kepuasan individu terhadap kedua aspek kehidupan tersebut. Dengan kata lain, seseorang dikatakan memiliki *work-life balance* yang baik apabila mampu memenuhi tanggung jawab pekerjaan tanpa mengabaikan kebutuhan pribadi, keluarga, maupun kehidupan sosialnya. [30].

2.2.4 Tiktok

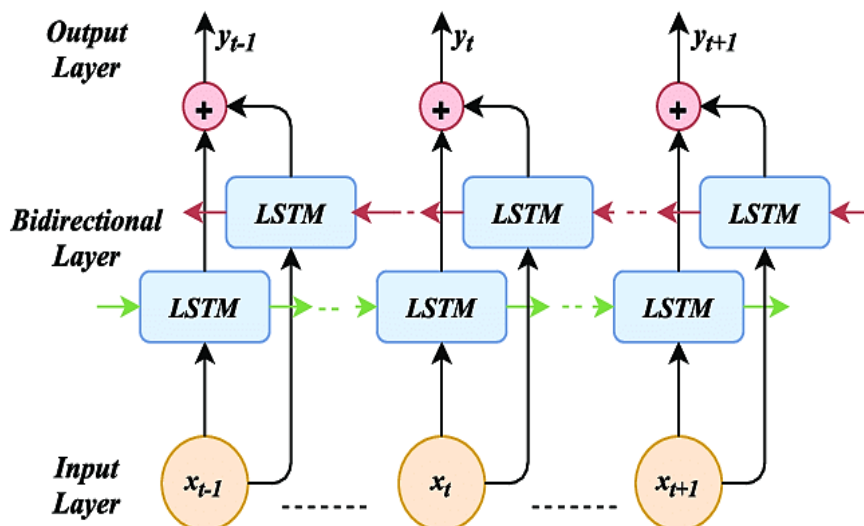


Gambar 2.3 TikTok

Gambar 2.1 Tiktok adalah aplikasi media sosial yang berasal dari China, dirilis pada September 2016, dengan nama awal yaitu Douyin diciptakan oleh Zhang Yiming. Seiring berjalannya waktu, untuk memudahkan pengguna dalam pelafalan, Douyin akhirnya berubah menjadi Tiktok. Tiktok merupakan media sosial berbasis video pendek, yang memungkinkan penggunanya untuk mengekspresikan diri melalui konten video, gambar, teks seperti ; *daily vlog*, promosi produk, *quotes* berbentuk teks, video motivasi, video prank, video *trend* Tiktok memiliki fitur utama seperti video pendek, musik dan efek, filter, duet atau *stitch video*, *explore*, dan *live streaming*, *shopping*. Yang masing masing memiliki kegunaan untuk kreativitas, pemasaran, kolaborasi, kampanye sosial [8].

2.3 Framework/Algoritma yang digunakan

2.3.1 Bidirectional Long Short-Term Memory (Bi-LSTM)



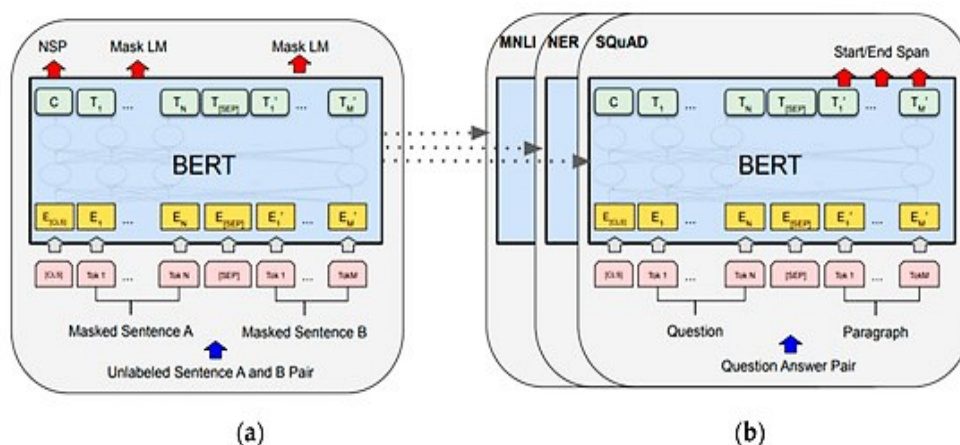
Gambar 2.4 Bidirectional Long Short-Term Memory (BiLSTM)

Gambar 2.4 *Bi-Directional Long Short-Term Memory* (Bi-LSTM) adalah sebuah inovasi dari arsitektur *Long Short-Term Memory* (LSTM) yang berada di dalam kelompok *Recurrent Neural Network* (RNN) [31]. Tidak seperti *LSTM* tradisional yang hanya menangani informasi dalam satu arah (maju), *Bi-LSTM* dapat memproses data sekuensial dalam dua arah secara bersamaan, yaitu maju (*forward*) dan mundur (*backward*). Kemampuan ini memberi model akses pada informasi konteks yang lebih menyeluruh, sehingga setiap kata dapat dimaknai berdasarkan kata-kata yang ada sebelum dan setelahnya dalam satu kalimat [32].

Dalam tugas *Natural Language Processing* (NLP), pemahaman konteks sangat krusial karena arti suatu kata sering kali dipengaruhi oleh kata-kata di sekitarnya. Karena itu, *Bi-LSTM* sering dimanfaatkan dalam berbagai tugas *Natural Language Processing* (NLP) seperti analisis sentimen, klasifikasi teks, pengenalan entitas (*Named Entity Recognition*), penerjemahan mesin, dan berbagai aplikasi pengolahan bahasa lainnya [9]. Dengan menggunakan pemrosesan informasi secara dua arah, *Bi-LSTM* dapat menciptakan representasi teks yang lebih mendalam sehingga meningkatkan akurasi model dalam memahami arti sebuah kalimat.. Berikut keunggulan dari *Bi-LSTM* :

1. Mampu menangkap hubungan antar kata dalam teks
2. Efektif untuk data sekuensial seperti teks
3. Banyak digunakan dalam klasifikasi teks

2.3.2 Indonesian Bidirectional Encoder Representations from Transformers (IndoBERT)



Gambar 2.5 Indonesian Bidirectional Encoder Representations from Transformers

Gambar 2.5 *IndoBERT* adalah model bahasa yang sudah dilatih sebelumnya dan dirancang khusus untuk menganalisis serta memahami teks dengan bahasa Indonesia. *IndoBERT* dikembangkan menggunakan arsitektur *Bidirectional Encoder Representations from Transformers* (BERT) yang memanfaatkan mekanisme *Transformer* untuk mendalami hubungan antar kata dalam sebuah kalimat [33]. Tidak seperti model sekuensial seperti *LSTM* atau *Bi-LSTM*, *IndoBERT* dapat memproses seluruh kata dalam kalimat sekaligus, sehingga lebih efektif dalam memahami konteks kata dari berbagai posisi [34].

IndoBERT telah menjalani pelatihan awal (*pre-training*) dengan menggunakan sejumlah besar data teks berbahasa Indonesia. Proses *pre-training* ini membantu model memahami struktur bahasa, tata bahasa, dan hubungan semantik antar kata sebelum diterapkan pada tugas tertentu. Selanjutnya, model dapat disesuaikan (*fine-tuning*) untuk berbagai tugas *Natural Language Processing* (NLP) seperti analisis sentimen, klasifikasi teks, deteksi emosi, penjawaban pertanyaan, *Named Entity Recognition* (NER), dan berbagai tugas pemrosesan

bahasa lainnya. Jenis model dari *IndoBERT* yaitu *Transformer/Pre-trained Language Model*. Berikut keunggulan dari *IndoBERT*:

1. Memahami konteks bahasa secara mendalam
2. Memiliki performa tinggi dalam berbagai tugas NLP
3. Lebih akurat dalam menangani bahasa Indonesia

2.4 Tools/software yang digunakan

2.4.1 Python



Gambar 2.6 Python

Gambar 2.6 Python merupakan bahasa pemrograman yang berorientasi objek dan memiliki struktur data tingkat tinggi. Bahasa ini terkenal, serbaguna, dan banyak digunakan. Diciptakan oleh Guido van Rossum, Python pertama kali dirilis pada tahun 1991 dengan penekanan pada kemudahan penggunaan, dan juga kejelasan, kemudahan dalam pemahaman kode, mengintegrasikan kemampuan, dan kejelasan sintaks [35]. Dengan tujuan menyederhanakan tugas programmer. Dan juga Python memiliki kelebihan yaitu ; efisiensi waktu dalam pembuatan program, kemudahan dalam pengembangan. Python memiliki fitur yang dimana masing masing fitur seperti ; Sintaks yang mudah dipelajari membuat mudah untuk dipelajari oleh pemula. Bahasa yang diinterpretasikan yang dimana kode dieksekusi baris per baris, mudah dalam pengujian. *Library* dukungan seperti ; *Pandas*, *Numpy*, *Tensorflow*. Python juga memiliki kegunaan yang lain yaitu seperti pada *Data Science* yaitu *library* seperti *Pandas*, *Numpy*, *Matplotlib*, dan juga *Scipy*. Pada *Machine Learning*, *library* meliputi *Tensorflow*, *Keras*, *Scikit learn* [36].

2.4.2 Visual Studio Code



Gambar 2.7 Visual Studio Code

Gambar 2.7 merupakan Visual Studio Code merupakan editor kode sumber yang dikembangkan oleh Microsoft dan dapat digunakan pada bermacam platform, seperti Windows, Linux, dan macOS. VS Code menyediakan bermacam-macam fungsi yang memudahkan pengembangan perangkat lunak, meliputi penyorotan sintaks, penyelesaian kode (*IntelliSense*), penelusuran kesalahan, pengelolaan versi dengan Git, serta sokongan untuk bermacam bahasa pemrograman lewat tambahan ekstensi [37].

2.4.3 Apify



Gambar 2.7 Apify

Gambar 2.7 Apify adalah platform *web scraping* otomatis yang sudah menyediakan *scraper* siap pakai yang digunakan untuk melakukan pengumpulan dan analisis data untuk kebutuhan bisnis. Data dikumpulkan dari media sosial TikTok, Instagram, dan X berdasarkan kata kunci terkait *work-life balance* [38] .