BAB II
KAJIAN LITERATUR

1.1 Tinjauan Pustaka

Pada bagian ini berisikan landasan teori dan akan dijelaskan tinjauan pustaka yang berkaitan dengan penelitian yang dilakukan.

1.1.1 Analisis Sentimen

Analisis sentiment merupakan salah satu dari *Natural Language Processing* (NLP) yang berfokus pada identifikasi dan kategorisasi dari opini yang disampaikan melalui teks [5]. Perkembangan media sosial yang sangat pesat membuat ketersediaan opini dan sentimen publik semakin meningkat. Hal ini membuat analisis sentimen menjadi alat penting untuk memahami sentimen publik terhadap suatu produk atau layanan. Dalam konteks *e-commerce*, analisis sentimen digunakan untuk mengetahui persepsi pengguna terhadap produk atau layanan. Analisis sentimen dapat dikategorikan kedalam tiga bagian berdasarkan tingkat granularitas, yaitu [6]:

1. **Tingkat kata**, dimana analisis menentukan polaritas berdasarkan suatu kata
2. **Tingkat kalimat**, dimana analisis menentukan polaritas berdasarkan suatu kalimat. Analisis ini biasa digunakan untuk analisis opini.
3. **Tingkat dokumen**, dimana analisis menentukan polaritas berdasarkan suatu dokumen.

Dalam analisis sentimen, berfokus terutama pada analisis orientasi sentimen terhadap kumpulan komentar, yang menunjukkan apakah kalimat yang disampaikan menunjukkan sentimen positif, negatif, atau netral terhadap suatu produk atau layanan [7]. Ada tiga aspek dalam menentukan sentimen publik. Yang pertama adalah menentukan subjektivitas dari teks, yaitu, apakah itu benar atau menunjukkan pendapat tentang topik tersebut. Yang kedua adalah menentukan polaritas dari teks, yaitu, menentukan apakah sebuah teks menunjukkan pendapat yang positif atau negatif tentang topik tersebut. Yang ketiga adalah menentukan kekuatan dari polaritas teks [8].

Saat ini, metode analisis sentimen berbasis teks dibagi menjadi tiga kategori: metode analisis sentimen berbasis *dictionary*, metode analisis sentimen berbasis *machine learning*, dan metode analisis sentimen berbasis *deep learning* [7].

Teknik *machine learing* dan *deep learning* sedang berkembang pesat dalam analisis data berbasis kecerdasan buatan (AI). Teknik-teknik ini membawa proses analisis data ke tingkat yang lebih tinggi, di mana model yang dibangun berbasis data (*data-driven*) daripada berbasis model (*model-driven*). Model pembelajaran terbaik dapat dilatih sesuai dengan domain aplikasi yang mendasari [6].

1.1.2 Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah bagian dari ilmu komputer dan kecerdasan buatan (AI) yang menjadi bahasa alami dalam interaksi antara manusia dan komputer. Dengan NLP, maka komputer dapat membaca, memahami, dan menghasilkan teks yang bisa dimengerti oleh manusia. Dengan perkembangan teknologi saat ini, dimana manusia hidup dalam era digital, NLP merupakan salah satu solusi untuk mengolah volume teks yang sangat besar yang tidak dapat dilakukan secara manual oleh manusia dalam waktu singkat. NLP mampu menangkap makna, konteks, ambiguitas, serta meniru pemrosesan bahasa manusia dalam memahami pengetahuan yang tersembunyi dalam teks [9].

Klasifikasi adalah salah satu proses implementasi dalam NLP. Dalam pemrosesan bahasa alami, klasifikasi mengacu pada proses memberikan teks kepada model komputer, yang kemudian mengelompokkannya ke dalam kategori atau label yang sesuai. Tugas klasifikasi ini adalah salah satu dari banyak tugas proses pengolahan bahasa yang sangat umum dan penting.

1.1.3 Deep Learning

Deep Learning merupakan bagian dari *neural network* konvensional, namun dianggap lebih baik dari pada metode *neural network* yang lain. *Deep Learning* juga membangun model pembelajaran berlapis-lapis secara bersamaan dengan menggunakan teknologi graf dan transformasi. *Deep Learning* bisa digunakan untuk berbagai aplikasi, seperti *visual data processing*, *Natural Language Processing* (NLP), dan termasuk *audio and speech processing* [9].

Beberapa tahun belakangan ini, banyak peneliti menerapkan metode dari *deep learning* kedalam proses analisis sentimen dan mendapatkan hasil yang baik. *Deep learning* menggunakan pembelajaran berbasis AI. Pada pembelajaran berbasis AI ini, *deep learning* dibangun berdasarkan *data-driven* daripada *model-driven*. Misalnya *Convolution-based Neural Networks* (CNN) yang cocok untuk analisis gambar, dan *Reccurent Neural Networks* (RNN) yang cocok untuk analisis data teks [6].

Ada beberapa model dari RNN, dan perbedaan dari kebanyakan model tersebut adalah dari cara menyimpan data. Pada dasarnya, RNN tidak memiliki kemampuan untuk menyimpan data masukan dan dinamakan dengan *feed forwarding-based learning*. Namun, ada model khusus dari RNN yang dinamakan *feedback-based model*, dimana model tersebut mampu untuk menyimpan data dan mempelajari data lama tersebut [6].

1.1.4 *Bidirectional Encoder Representations from Transformers* (BERT)

Bidirectional Encoder Representations from Transformers (BERT) merupakan salah satu terobosan besar dalam bidang pemrosesan bahasa alami (NLP). Model ini dikembangkan oleh Google pada tahun 2018 dan mengandalkan mekanisme transformer untuk memahami konteks kata secara dua arah, yaitu dari kiri ke kanan dan sebaliknya. Pendekatan ini memungkinkan BERT memahami makna kata secara lebih akurat dalam konteks kalimat, yang menjadikannya sangat unggul dalam berbagai tugas NLP seperti analisis sentimen, ekstraksi entitas, dan klasifikasi teks [10].

BERT dilatih dengan dua strategi utama, yaitu *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP). Pada MLM, sebagian kata dalam kalimat disembunyikan dan model diminta untuk memprediksi kata tersebut, melatih kemampuan pemahaman konteks. Sementara pada NSP, dua kalimat diberikan dan model harus menilai apakah kalimat kedua merupakan kelanjutan dari yang pertama. Kedua metode ini menjadikan BERT sangat unggul dalam memahami hubungan antar kalimat dan makna kontekstual [10].

Dalam berbagai studi, BERT terbukti memberikan performa unggul dalam analisis sentimen dibanding metode tradisional. Misalnya, penelitian oleh Singh et al. (2021) menunjukkan bahwa BERT mampu mengklasifikasikan sentimen dalam *tweet* terkait COVID-19 dengan akurasi hingga 94%, mengungguli metode berbasis *lexicon* atau *machine learning* seperti Naïve Bayes dan SVM. BERT juga digunakan untuk memahami intensitas emosi dan opini masyarakat secara global serta regional (khususnya India), menunjukkan fleksibilitasnya dalam menangani data dari berbagai bahasa dan konteks [11].

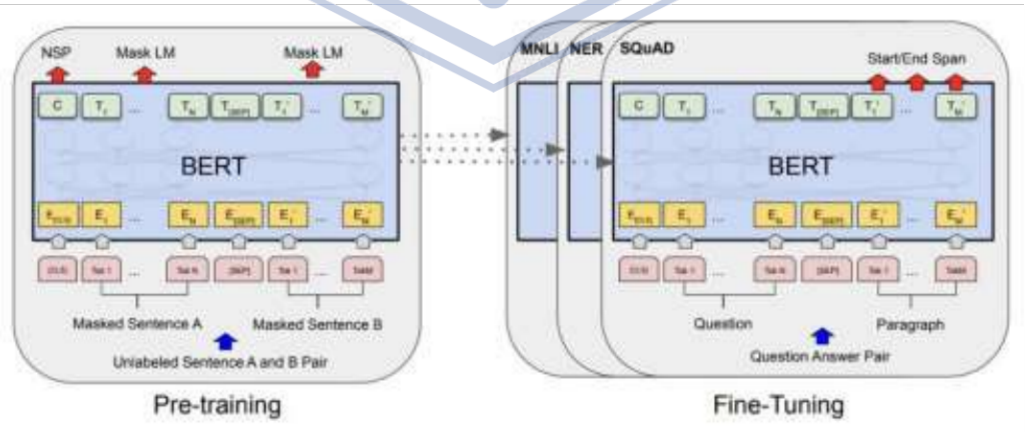
Dalam penelitian lain, BERT dikombinasikan dengan arsitektur *deep learning* seperti CNN, RNN, dan BiLSTM untuk meningkatkan performa klasifikasi sentimen. Studi oleh Bello et al. (2023) menunjukkan bahwa gabungan BERT dengan BiLSTM menghasilkan akurasi tertinggi (93%) dan F1-score 95% pada data tweet berbahasa Inggris, jauh mengungguli kombinasi Word2Vec + CNN maupun penggunaan BERT tanpa tambahan layer klasifikasi lanjutan [10].

1.1.5 Fine-Tuning IndoBERT

IndoBERT adalah model bahasa berbasis arsitektur BERT (Bidirectional Encoder Representations from Transformers) yang telah dilatih (pretrained) secara khusus menggunakan korpus besar berbahasa Indonesia yang terdiri dari miliaran kata. Model ini dikembangkan untuk menjawab keterbatasan model NLP umum seperti BERT standar dalam memahami struktur dan nuansa linguistik khas bahasa Indonesia. IndoBERT terbukti efektif dalam memahami nuansa bahasa Indonesia dan menghasilkan kinerja yang lebih tinggi dibanding model berbasis machine learning tradisional seperti Naïve Bayes atau SVM [12].

Fine-tuning adalah proses pelatihan lanjutan terhadap model pralatih (*pretrained model*) menggunakan *dataset* khusus sesuai dengan tugas tertentu. Fine-tuning dilakukan setelah tahap pretraining selesai, di mana model telah mempelajari representasi bahasa umum dan melatih ulang sebagian atau seluruh lapisan IndoBERT menggunakan dataset baru yang relevan [13].

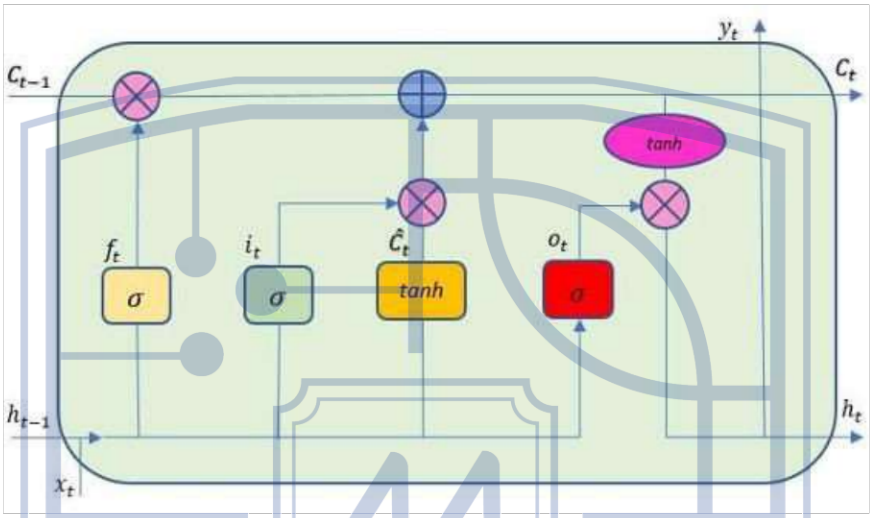
Fine-tuning dilakukan dengan menggunakan model IndoBERT base (12 layer, 768 hidden units). Ada dua pendekatan umum dalam proses fine-tuning, yaitu *Full Fine-Tuning* dimana seluruh parameter IndoBERT di-update selama pelatihan dan *Partial Fine-Tuning (Layer Freezing)* dimana hanya sebagian lapisan atas IndoBERT yang dilatih ulang, sisanya dibekukan (*frozen*). Fine-tuning IndoBERT melibatkan kombinasi teknik NLP modern seperti tokenisasi sub-word, pelatihan lanjutan dengan layer freezing, serta penambahan classifier sederhana di atas representasi [CLS]. Strategi ini memungkinkan model memahami konteks lokal (Bahasa Indonesia) dan domain khusus seperti wisata atau media sosial, sekaligus menjaga efisiensi komputasi [14]. Berikut ini adalah gambar proses *fine-tuning*:



Gambar 2.1 Diagram proses indoBERT [15]

1.1.6 *Long Short-Term Memory* (LSTM)

LSTM merupakan varian dari RNN karena mampu untuk menyimpan kata-kata sentimen dari suatu teks. LSTM menggunakan blok *memory-cell* yang terdiri dari *input gate, forget gate*, dan *output gate*. *Memory-cell* ini berguna untuk mengatasi masalah *vanishing gradient* pada model RNN lainnya. Dengan LSTM, informasi dapat disimpan dalam waktu panjang karena bisa mempelajari *long-term dependency* [3].



Gambar 2.2 Diagram LSTM [16]

Pada LSTM, terbagi tiga gerbang utama, yaitu *forget gate, input gate*, dan *output gate*. LSTM juga terdiri dari dua bagian yaitu *cell state* dan *hidden state*.

Pada *forget gate* terjadi proses pemilahan informasi yang ada pada *cell state* dengan persamaan:

$$f_t = \sigma(W_f \,.\, [h_{t-1}, X_t] + b_f) \quad \text{..... (1)}$$

Informasi akan dibuang dari *cell state* jika pada *forget state* bernilai 0, dan akan disimpan ke *cell state* jika *forget state* bernilai 1.

Pada *input gate*, informasi akan melewati 2 lapisan yaitu *sigmoid* (persamaan 2) dan *tanh* (persamaan 3). *Cell state* akan dihasilkan dari nilai *output* dari penggabungan *sigmoid* dan *tanh*.

$$i_t = \sigma(W_i \,.\, [h_{t-1}, X_t] + b_i) \quad \text{.....} \quad (2)$$

$$C_t = tanh(W_c \,.\, [h_{t-1}, X_t] + b_c) \quad \text{..... (3)}$$

Proses selanjutnya adalah perbaharuan nilai *cell state* (persamaan 4) dan kemudian akan melewati *output gate*, dimana nilai *output* dari *cell state* akan ditentukan. Pada lapisan *sigmoid* akan digunakan untuk memilih *output* berdasarkan *cell state* (persamaan 5) dan selanjutnya hasil tersebut akan diteruskan ke lapisan *tanh* (persamaan 6).

$$C_t = f_t \,.\, C_{t-1} + i_t \,.\, C_t$$

……...(4)

$$O_t = \sigma(W_o \,.\, [h_{t-1}, X_t] + b_o)$$ ……..……..

(5)

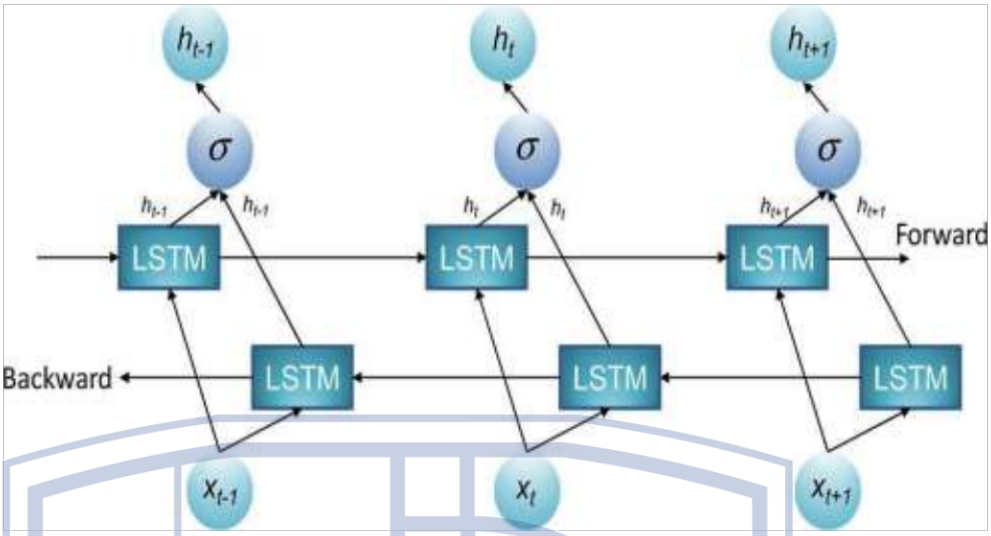
$$h_t = O_t \,.\, \tanh\,(C_t)$$ ……...……..

(6)

1.1.7 Bidirectional *Long Short-Term Memory* (BiLSTM)

Bidirectional Long Short-Term Memory (BiLSTM) merupakan pengembangan dari *Long Short-Term Memory* (LSTM). LSTM merupakan varian dari RNN karena mampu untuk menyimpan kata-kata sentimen dari suatu teks. Namun, model LSTM hanya mampu memproses data dalam satu arah yang mana hanya bisa menggunakan informasi dari masa lalu [3].

Oleh karena itu, dikembangkan model BiLSTM yang bertujuan untuk mengatasi kelemahan LSTM. Model BiLSTM mampu memproses data dalam dua arah yaitu maju (*forward*) dan mundur (*backward*) sehingga bisa menangkap informasi dari masa lalu dan masa depan secara berurutan. Pendekatan ini membuat BiLSTM bisa menghasilkan Tingkat akurasi yang tinggi [17]. Penggunaan jaringan ini sangat diperlukan dalam NLP. Hal ini dikarenakan, makna sebuah kata dalam suatu kalimat dapat bergantung pada kata-kata sebelumnya maupun sesudahnya. Oleh karena itu, sangat dibutuhkan untuk menelusuri kalimat dalam kedua arah guna memperoleh makna yang tepat. BiLSTM memiliki kemampuan ini, yang membedakannya dari jaringan LSTM biasa. Berbeda dengan model RNN lainnya seperti *Support Vector Machine* (SVM) dan *Naïve Bayes,* BiLSTM mampu secara efektif memodelkan ketergantungan berurutan antara kata dan kalimat dalam kedua arah suatu urutan, sehingga sangat cocok untuk menganalisis urutan masukan dalam kalimat yang bisa menghasilkan prediksi yang akurat [18].

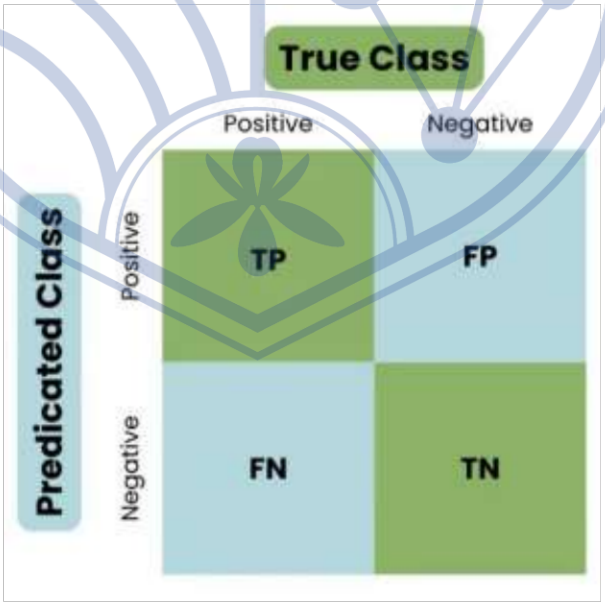


Gambar 2.3 Diagram BiLSTM [16]

Pada Gambar 2.3, menunjukkan proses dari BiLSTM yang terdiri dari *forward* dan *backward* dari LSTM, yang mana saat terjadi proses data *forward*, urutan masukkan dibalik dan dihitung Kembali kedalam proses data *backward* dimana hasil akhir adalah berupa tumpukan nilai *output* dari kedua proses LSTM tersebut.

1.1.8 *Confusion Matrix*

Confusion Matrix merupakan suatu metode evaluasi untuk kinerja sebuah model dan dihasilkan dalam bentuk tabel. Metode ini digunakan sebagai tolak ukur sistem untuk menunjukkan perbandingan antara data hasil prediksi dengan data sebenarnya [19].



Gambar 2.4 *Confusion Matrix* [19]

Keterangan:

a. *True Positive* (TP) : model memprediksikan hasil positif sesuai data sebenarnya.
b. *True Negative* (TN) : model memprediksikan hasil negatif sesuai data sebenarnya.
c. *False Positive* (FP) : model salah memprediksikan hasil positif.
d. *False Negative* (FN) : model salah memprediksikan hasil negatif.

Dengan *confusion matrix*, makan dapat mengevaluasi hasil dari proses klasifikasi dengan memperhitungkan nilai *accuracy, precision, recall,* dan *F1-score*.

1. *Accuracy*

   Accuracy adalah *metrics* paling umum yang digunakan untuk mengevaluasi tingkat akurasi model klasifikasi. Akurasi merupakan nilai rasio antara jumlah prediksi yang tepat terhadap total prediksi yang dibuat.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \text{ X } 100\% \quad \text{.....(7)}$$

2. *Precision*

   Precision mengukur tingkat ketepatan suatu model dalam melakukan klasifikasi. *Precision* mengukur rasio antara prediksi *true positive* dengan total prediksi positif (baik yang benar maupun salah) pada setiap label.

$$Precision = \frac{TP}{TP+FP} \quad \text{.....(8)}$$

3. *Recall*

   Recall mengukur seberapa baik model dalam mendeteksi semua label benar dari kelas tertentu. *Recall* mengukur rasio antara prediksi positif yang benar dengan total jumlah data yang sebenarnya positif.

$$Recall = \frac{TP}{TP+FN} \quad \text{.....(9)}$$

4. *F1-score*

F1-Score adalah nilai rata-rata antara *precision* dan *recall*. *F1-Score* mengukur seberapa baik model kita dalam mengklasifikasikan sentiment positif maupun negatif secara akurat.

$$F1 = 2 \text{ x } \frac{precision \text{ x } recall}{precision + recall} \quad \text{(10)}$$

2.2 Penelitian Terdahulu

Penelitian terkait analisis sentimen menggunakan model BiLSTM telah dilakukan oleh beberapa peneliti sebelumnya yang diterapkan pada beberapa bidang yang berbeda. Berikut ini adalah rangkuman beberapa penelitian yang telah dilakukan tentang analisis sentimen berbasis teks berdasarkan *literature review*.

Tabel 2.1 *Literature Review* Analisis Sentimen

No	Judul Penelitian	Tahun	Penulis	Objek Penelitian	Metode	Hasil Utama
1	Analisis Sentimen Kepuasan Pelanggan pada Jasa Ekspedisi Menggunakan BiLSTM dan BiGRU	2022	Salsabila Zahirah Pranida	Ekspedisi (JNE, JNT, Sicepat, Anteraja)	BiLSTM & BiGRU	Akurasi BiLSTM: 71.9%, BiGRU: 70.1%
2	Analisis Sentimen dan Pemodelan Topik Aplikasi Telemedicine pada Google Play Menggunakan	2023	Siti Mutmainah dkk.	Aplikasi Telemedicine (Alodokter, Halodoc, dll)	BiLSTM & LDA	Akurasi sentimen 95%, koherensi topik positif: 0.6437

	BiLSTM dan LDA					
3	Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia	2022	Dloifur Rohman Alghifari dkk.	Layanan Grab	BiLSTM	Akurasi: 91%, training loss: 28%, lebih baik daripada LSTM
4	BERT- and BiLSTM-Based Sentiment Analysis of Online Chinese Buzzwords	2022	Xinlu Li dkk.	Buzzwords Tiongkok	BERT + BiLSTM	Model outperform 7 state-of-the-art, unggul dalam F1-score dan recall
5	Analisis Sentimen Opini Publik Menggunakan Metode BiLSTM pada Media Sosial Twitter	2023	Ardian Nur Romadhan dkk.	Opini publik soal UU KPK di Twitter	BiLSTM + TF-IDF & Word2Vec	Akurasi TF-IDF: 84.84%, Word2Vec: 85.72%, F1-score: 85.92%
6	Sentiment Analysis of Comment Texts Based on BiLSTM	2019	Guixian Xu dkk.	Komentar online (umum)	BiLSTM + TF-IDF berbobot	Lebih akurat dibanding RNN, CNN, LSTM, NB; baik dalam precision, recall, F1

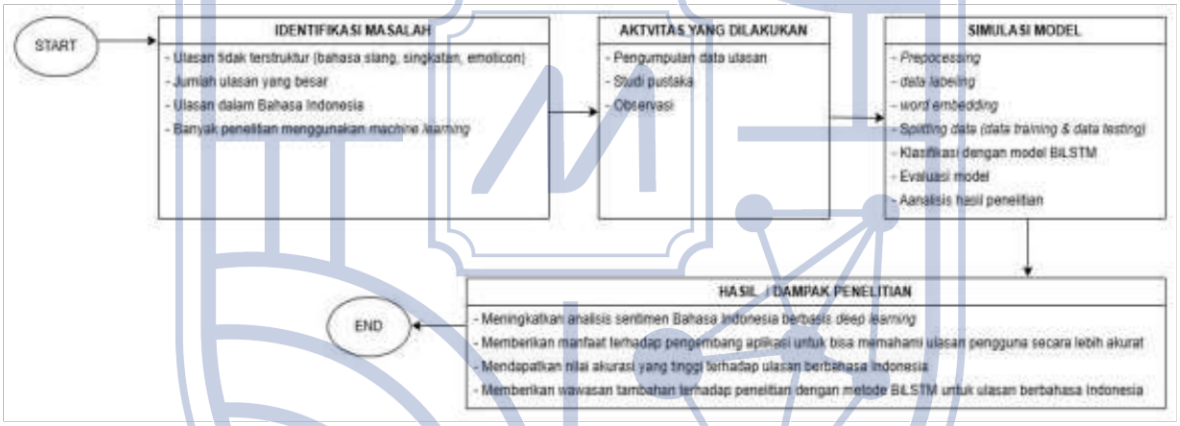
Berdasarkan uraian penelitian terdahulu, BiLSTM memiliki pencapaian akurasi yang sangat baik dibandingkan metode lainnya, dan masih belum dilakukan pada aplikasi Tokopedia. Penelitian terdahulu tentang analisis sentimen pada aplikasi Tokopedia umumnya menggunakan metode seperti LSTM, *Support Vector Machine* (SVM), Naïve Bayes, dan *Logistic Regression* untuk mengklasifikasikan sentimen pengguna. Pendekatan ini sering kali terbatas dalam menangkap konteks bahasa yang lebih kompleks dan hubungan antar kata dalam teks ulasan. Sebaliknya, penelitian yang akan dilakukan dengan BiLSTM menawarkan keunggulan dalam memahami konteks dari kedua arah yaitu *backward* dan *forward* sehingga dapat meningkatkan akurasi analisis sentimen. Dengan BiLSTM, model dapat lebih efektif dalam menangkap pola bahasa alami dan nuansa sentimen yang lebih halus dalam ulasan pengguna Tokopedia.

Pada penelitian terdahulu, rata-rata hanya menggunakan tiga label klasifikasi, yaitu negatif, netral, dan positif. Sedangkan di penelitian yang akan dilakukan, akan diklasifikasikan kedalam lima label, yaitu sangat negatif, negatif, netral, positif, dan sangat positif.

Dengan keunggulan dari penelitian ini diharapkan dapat memberikan kontribusi lebih dalam bidang *e-commerce* untuk memahami sentimen pengguna berdasarkan ulasan mereka secara lebih akurat, sehingga dapat bermanfaat bagi pihak Tokopedia dalam meningkatkan layanan mereka, kepuasan pelanggan, serta merancang strategi pemasaran yang lebih tepat sasaran sesuai dengan kebutuhan pelanggan.

2.3 Kerangka Pikir Pemecahan Masalah

Kerangka ini dirancang untuk menunjukkan alur penyelesaian masalah penelitian melalui Solusi yang diusulkan dan hasil yang dihasilkan pada penelitian ini.



Gambar 2.5 Kerangka konseptual

Pada gambar 2.4 merupakan kerangka alur pemikiran dalam melakukan penelitian ini. Tahap awal berfokus pada identifikasi masalah, mencakup tantangan seperti ulasan tidak terstruktur (bahasa slang, singkatan, emoticon), jumlah data yang besar, serta dominasi penelitian yang menggunakan pendekatan *machine learning*. Pemahaman ini penting untuk merancang metode yang sesuai dalam menangani ulasan berbahasa Indonesia agar bisa menghasilkan analisis sentimen yang lebih akurat dan bermakna.

Setelah permasalahan teridentifikasi, maka akan dilakukan tahapan aktivitas dan simulasi model, mulai dari pengumpulan data ulasan dari aplikasi Tokopedia sebanyak 30.000 data ulasan terbaru, studi pustaka, hingga dimasukkan kedalam proses pemodelan yang dimulai dari proses *preprocessing* seperti *casefolding, data cleaning, stopwords removal, stemming, tokenizer* dan *normalization*. Setelah itu akan dilanjutkan ke tahap *word*

embedding, *data splitting (training/testing)*, klasifikasi dengan model BiLSTM, dan evaluasi performa model. Pendekatan ini memungkinkan analisis sentimen dengan mempertimbangkan konteks kata dalam suatu ulasan, menangkap makna yang lebih kompleks dibandingkan metode tradisional.

Hasil dan dampak penelitian, yang diharapkan dapat meningkatkan analisis sentimen untuk ulasan berbahasa Indonesia berbasis *deep learning*. Selain mendapatkan tingkat akurasi yang lebih tinggi, penelitian ini juga memberikan manfaat praktis bagi pengembang aplikasi dalam memahami preferensi pengguna berdasarkan ulasan mereka. Dengan pendekatan ini, bisnis dapat mengambil keputusan yang lebih tepat, mengoptimalkan pengalaman pengguna, dan mengembangkan strategi pemasaran yang lebih efektif.