

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Di era digital saat ini, media sosial dan forum daring telah berkembang pesat dalam menyebarkan jumlah informasi dan opini publik yang tersedia secara daring. Untuk memahami dan memanfaatkan data tersebut, analisis sentimen berperan penting dalam menganalisis konten pengguna seperti postingan dan komentar. Analisis sentimen merupakan sebuah teknik berbasis data teks yang memanfaatkan *Natural Language Processing* (NLP) dan *Machine Learning* untuk secara otomatis mengevaluasi emosi yang disampaikan penulis dalam teks seperti positif atau negatif (Prakash & Aloysius, 2021). Analisis sentimen dapat membantu bisnis, pemerintah dan organisasi dalam merancang strategi, mengenali tren pasar dan merespons tantangan yang timbul dari opini publik. (Zhao et al., 2023).

Penggunaan model *machine learning* dalam analisis sentimen berguna untuk mengidentifikasi pola kompleks dan mengklasifikasikan sentimen dalam teks berbahasa dan konteks beragam sehingga memberikan wawasan penting mengenai opini publik (Delimayanti et al., 2021). Seiring dengan meningkatnya jumlah data teks yang tidak terstruktur dan kompleks, terutama dari media sosial, muncul tantangan dalam mengidentifikasi pola sentimen secara akurat dan menyeluruh. Model *machine learning* konvensional yang umum digunakan dalam analisis sentimen seringkali memiliki keterbatasan dalam menangani keragaman dan kompleksitas bahasa dalam data media sosial.

Sejalan dengan hal tersebut, penelitian menunjukkan bahwa model *machine learning* konvensional cenderung memproses kalimat hanya sebagai sekumpulan kata tanpa mempertimbangkan urutan dan hubungan antar kata, sehingga kurang efektif dalam menganalisis sentimen secara akurat (Sofia & Supriyadi, 2021). Selain itu, model konvensional biasanya bergantung pada proses rekayasa fitur manual yang rumit dan memerlukan keahlian khusus untuk menentukan fitur yang

relevan, sehingga membatasi kemampuannya dalam menangkap makna kontekstual secara mendalam dari data teks yang beragam (Kansara & Sawant, 2020). Karakteristik bahasa di media sosial seperti penggunaan singkatan serta struktur kalimat yang tidak baku menjadi tantangan tambahan bagi model konvensional yang hanya mengandalkan pola statistik sederhana (Mariel et al., 2024). Oleh karena itu, dibutuhkan eksplorasi model yang lebih canggih agar mampu mengatasi kompleksitas tersebut secara lebih efektif.

Pemilihan model *machine learning* yang tepat sangat penting dalam analisis sentimen karena berpengaruh terhadap kualitas hasil yang diperoleh (Naquitasia et al., 2021). Salah satu sub-bidang dari *machine learning* yang telah menunjukkan pencapaian signifikan dalam beberapa tahun terakhir adalah *deep learning* (Setiawan, 2021). Dibandingkan dengan model *machine learning* konvensional, *deep learning* memiliki keunggulan dalam menangani data yang tidak terstruktur dan beragam dengan menggunakan jaringan saraf tiruan berlapis yang memungkinkan sistem dapat mempelajari representasi data yang kompleks (Khotimah et al., 2024). Pendekatan yang digunakan dalam *deep learning* melibatkan transformasi data ke dalam bentuk matriks numerik yang kemudian digunakan untuk melatih model secara lebih optimal. (Naquitasia et al., 2021). Beberapa model *deep learning* yang dapat digunakan dalam analisis sentimen antara lain *Long Short-Term Memory* (LSTM) dan *Indonesia Bidirectional Encoder Representations from Transformers* (IndoBERT).

Model LSTM merupakan jenis arsitektur jaringan saraf dalam *Deep Learning* dan merupakan pengembangan dari *Recurrent Neural Network* (RNN) yang memiliki keunggulan utama dalam menyimpan informasi dari urutan data yang panjang serta mengatasi masalah *vanishing* atau *exploding gradien* melalui mekanisme *gate*. (Witanto et al., 2022). Mekanisme *gate* ini terdiri dari *input gate*, *forget gate* dan *output gate* yang memungkinkan LSTM mempertahankan dan mengenali pola dan hubungan kata yang panjang dalam teks, sehingga sangat efektif untuk analisis sentimen. (Pradana et al., 2023). Sejalan dengan keunggulan tersebut, beberapa penelitian telah menunjukkan efektivitas model LSTM dalam analisis sentimen teks berbahasa Indonesia. Berdasarkan penelitian oleh Mudding

& Arifin (2022), akurasi model LSTM mencapai 97%. Sementara itu, penelitian oleh Ramadhan dan Hendrastuty (2024) menunjukkan bahwa model LSTM memiliki kinerja yang lebih baik dibandingkan *Naïve Bayes*, dengan akurasi LSTM mencapai 91.85% dan *Naïve Bayes* 83%. Hasil tersebut semakin menegaskan kapabilitas model LSTM dalam mengolah data teks berbahasa Indonesia untuk analisis sentimen.

Model IndoBERT merupakan model *deep learning* berbasis *transformer* yang mampu memanfaatkan *transfer learning* secara efektif dengan model yang telah dilatih sebelumnya pada kumpulan besar teks berbahasa Indonesia. (Kusnadi et al., 2022). Model ini menggunakan pendekatan pembelajaran dua arah (*bidirectional learning*), yaitu mempelajari konteks kata dari kiri ke kanan dan sebaliknya sehingga mampu menangkap makna teks secara lebih menyeluruh dan menghasilkan representasi yang lebih akurat. (Syah et al., 2025). Keunggulan tersebut membuat IndoBERT banyak digunakan dalam tugas-tugas pemrosesan bahasa alami, termasuk analisis sentimen. Penelitian oleh Pradana (2024) menunjukkan bahwa IndoBERT mampu mencapai akurasi sebesar 90,5% dalam klasifikasi sentimen. Sementara itu, penelitian oleh Jayadianti (2022) menunjukkan akurasi yang lebih tinggi yakni sebesar 95,16%. Temuan ini memperkuat bahwa IndoBERT merupakan salah satu model yang potensial dan efektif untuk digunakan dalam analisis sentimen teks berbahasa Indonesia.

Meskipun model *deep learning* telah terbukti memiliki performa yang baik berdasarkan penelitian sebelumnya dalam analisis sentimen, masih terdapat keterbatasan studi yang secara langsung membandingkan performa model *deep learning* seperti LSTM dan IndoBERT, khususnya dalam analisis sentimen berbahasa Indonesia. Perbandingan ini relevan mengingat penelitian-penelitian sebelumnya yang mengkaji analisis sentimen terhadap program MSIB masih terbatas pada penggunaan model *machine learning* konvensional seperti *Naive Bayes*, yang menunjukkan akurasi beragam seperti pada penelitian yang dilakukan oleh (Adam, 2023) menunjukkan hasil akurasi model sebesar 78%. Lalu penelitian oleh (Cahya, 2023) sebesar 65%, serta sebesar 85% untuk hasil akurasi dari penelitian (Sefuloh, 2023; Husni, 2025).

Berdasarkan hal tersebut, penelitian ini bertujuan untuk membandingkan performa model LSTM dan IndoBERT dalam melakukan analisis sentimen terhadap opini publik. Adapun, Program MSIB dipilih sebagai studi kasus karena menyediakan opini publik dari media sosial X untuk digunakan sebagai data sentimen dalam penelitian ini. Evaluasi performa model dilakukan menggunakan metrik evaluasi klasifikasi seperti *accuracy*, *precision*, *recall* dan *f1-score* sehingga hasil penelitian ini diharapkan tidak hanya memberikan rekomendasi model yang lebih optimal tetapi juga turut berkontribusi dalam pemilihan model analisis sentimen berbahasa Indonesia yang lebih akurat untuk kebijakan publik.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, penelitian ini berfokus pada analisis sentimen terhadap Program MSIB menggunakan model *deep learning*. Oleh karena itu, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana karakteristik data opini publik program MSIB yang dikumpulkan dan dilakukan pra-pemrosesan untuk analisis sentimen?
2. Bagaimana hasil performa model LSTM dalam melakukan analisis sentimen terhadap opini publik mengenai Program MSIB berdasarkan metrik evaluasi seperti *accuracy*, *precision*, *recall* dan *F1-score*?
3. Bagaimana hasil performa model IndoBERT dalam melakukan analisis sentimen terhadap opini publik mengenai Program MSIB berdasarkan metrik evaluasi seperti *accuracy*, *precision*, *recall* dan *F1-score*?
4. Model mana yang memiliki performa lebih baik dalam analisis sentimen terhadap opini publik mengenai Program MSIB berdasarkan hasil perbandingan *accuracy*, *precision*, *recall* dan *F1-score*?

1.3 Ruang Lingkup Penelitian

Agar penelitian tetap berada pada konteks yang relevan, maka peneliti membatasi permasalahan dalam penelitian ini untuk memfokuskan pembahasan pada inti permasalahan dan tidak terlalu meluas. Berikut ruang lingkup penelitiannya yaitu:

1. Penelitian ini berfokus pada perbandingan performa model *deep learning* LSTM dan IndoBERT dalam analisis sentimen terhadap opini publik di media sosial, dengan Program MSIB sebagai studi kasus. Analisis sentimen dilakukan dengan mengklasifikasikan data ke dalam dua kategori utama yaitu sentimen positif dan negatif.
2. Data yang digunakan dalam penelitian ini berasal dari opini publik pada media sosial X (Twitter) dalam rentang waktu 17 Oktober 2023 – 31 Januari 2025.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dikemukakan, penelitian ini bertujuan untuk:

1. Mengetahui karakteristik data opini publik Program MSIB yang telah dikumpulkan dan dilakukan pra-pemrosesan untuk analisis sentimen.
2. Mengetahui bagaimana performa LSTM dalam melakukan klasifikasi sentimen terhadap opini publik mengenai Program MSIB berdasarkan metrik evaluasi seperti *accuracy*, *precision*, *recall* dan *F1-score*.
3. Mengetahui bagaimana performa IndoBERT dalam melakukan klasifikasi sentimen terhadap opini publik mengenai Program MSIB menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall* dan *F1-score*.
4. Membandingkan hasil analisis sentimen antara LSTM dan IndoBERT untuk menentukan model yang lebih unggul dalam mengklasifikasikan opini publik terhadap Program MSIB berdasarkan metrik *accuracy*, *precision*, *recall* dan *F1-score*.

1.5 Manfaat penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik dari segi teoritis maupun praktis di antaranya:

1. Manfaat Teoritis

Secara teoritis penelitian ini dapat bermanfaat sebagai informasi pada masyarakat mengenai sentimen publik terhadap program MSIB baik berupa sentimen positif dan negatif. Serta menjadi referensi untuk penelitian selanjutnya yang relevan dengan topik analisis sentimen.

2. Manfaat Praktis

Secara praktis penelitian ini dapat memberikan gambaran dan referensi untuk melakukan analisis sentimen serta evaluasi pemodelan pada topik atau masalah tertentu dengan memanfaatkan model *machine learning* seperti LSTM atau IndoBERT.