BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan pesat dalam komunikasi digital telah mengubah secara fundamental cara informasi diproduksi, disebarluaskan, dan dikonsumsi. Transformasi ini tidak hanya mempermudah akses terhadap berita dan wacana publik, tetapi juga membuka peluang bagi penyebaran informasi palsu secara luas, yang dikenal sebagai berita hoaks atau *fake news*. Proliferasi konten semacam ini telah menimbulkan ancaman serius terhadap kepercayaan publik, stabilitas demokrasi, dan keamanan siber di tingkat nasional maupun global [1, 2]. Di kawasan Asia Tenggara, khususnya Indonesia dan Malaysia, tantangan ini diperparah oleh tingginya penetrasi media sosial, keberagaman linguistik, serta belum meratanya tingkat literasi digital [3, 4]. Faktor-faktor tersebut berinteraksi untuk menghasilkan sebuah ekosistem informasi yang memiliki kelemahan, yang memungkinkan penyebaran berita hoaks dengan cepat dan menyeluruh. Hal ini terjadi terutama selama periode-periode penting seperti saat pemilihan umum atau ketika menghadapi krisis kesehatan masyarakat.

Di Malaysia, sebanyak 3.288 aduan berita palsu tercatat dari tahun 2020 hingga 31 Mei 2022 [5]. Sementara itu, di Indonesia, Kementerian Komunikasi dan Informatika (Kominfo) mengidentifikasi sebanyak 11.642 konten hoaks sepanjang Agustus 2018 hingga Mei 2023 [6]. Angka ini menunjukkan bahwa penyebaran hoaks di kawasan Asia Tenggara masih menjadi tantangan serius yang dapat mengganggu ketertiban publik, memperlemah kepercayaan masyarakat terhadap informasi daring, serta mengancam stabilitas sosial-politik dan ketahanan siber nasional.

Perluasan definisi keamanan siber kini mencakup perlindungan terhadap ancaman non-teknis seperti disinformasi terstruktur [7, 8]. Hal ini mendorong integrasi teknologi *Artificial Intelligence* (AI) dan *Natural Language Processing* (NLP) sebagai solusi strategis. Selama sepuluh tahun terakhir, penggunaan *deep learning* khususnya arsitektur Bidirectional Long Short-Term Memory (Bi-LSTM) terbukti unggul menangkap nuansa linguistik yang kompleks dalam deteksi hoaks [9–11]. Kelebihan utama dari Bi-LSTM terletak pada kemampuannya untuk memproses urutan teks dalam dua arah secara bersamaan (*forward dan backward*).

Hal ini memungkinkan sistem untuk secara efektif mengidentifikasi pola semantik tersembunyi dan menggambarkan ketergantungan antar-kata yang menjadi ciri konten yang bersifat menyesatkan [12].

Namun, mayoritas riset Bi-LSTM masih terkonsentrasi pada tiga keterbatasan utama, pertama penggunaan *single-language dataset* (terutama bahasa Inggris), kedua ketergantungan pada data dari wilayah Barat, dan yang ketiga pengembangan model terpisah untuk setiap bahasa [13]. Padahal, karakteristik bahasa Indonesia dan Melayu yang memiliki kesamaan leksikal namun perbedaan kontekstual seperti variasi istilah politik dan pola *code-mixing* yang berbeda menuntut pendekatan terpadu.

Berdasarkan tantangan tersebut, penelitian ini mengusulkan model Bi-LSTM multibahasa yang dirancang untuk mendeteksi hoaks dengan memproses secara bersamaan dataset dari Indonesia dan Malaysia. Inovasi signifikan dari penelitian ini adalah integrasi dari dua korpus linguistik yang bersumber dari media formal di Indonesia dan konten berbasis web di Malaysia ke dalam satu kerangka arsitektur yang terintegrasi. Pendekatan tersebut tidak hanya mengisi celah penelitian dalam konteks bahasa yang digunakan di kawasan ASEAN, tetapi juga menyediakan kerangka metodologis yang dapat diterapkan secara umum pada kombinasi bahasa serumpun lainnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, permasalahan utama dalam penelitian ini dirumuskan sebagai berikut:

1. Bagaimana cara mengimplementasikan model Bi-LSTM (*Bidirectional Long Short-Term Memory*) untuk mendeteksi berita hoaks dalam bahasa Indonesia dan Melayu?
2. Berapa nilai akurasi, presisi, *recall*, dan *F1-score* yang dihasilkan oleh model Bi-LSTM dalam mendeteksi berita hoaks multibahasa?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan dari penelitian ini adalah sebagai berikut:

1. Mengimplementasikan model Bi-LSTM (*Bidirectional Long Short-Term Memory*) untuk melakukan deteksi berita hoaks dalam bahasa Indonesia dan Melayu.

2. Mengukur nilai akurasi, presisi, *recall*, dan *F1-score* yang dihasilkan oleh model Bi-LSTM dalam mendeteksi berita hoaks multibahasa?

1.4 Batasan Masalah

Penelitian ini dibatasi pada ruang lingkup berikut:

1. Bahasa yang digunakan terbatas pada bahasa Indonesia dan Melayu Malaysia dalam bentuk formal, tanpa mencakup dialek daerah atau campuran bahasa ekstrem.

2. Model yang digunakan adalah arsitektur dasar *Bidirectional Long Short-Term Memory* (Bi-LSTM), tanpa integrasi dengan metode lain seperti CNN, Attention, atau Transformer.

3. Data yang digunakan bersumber dari korpus berita daring berbahasa Indonesia dan Melayu yang telah dilabelkan sebagai hoaks atau fakta, serta memiliki struktur naratif yang utuh.

4. Penelitian ini hanya melakukan klasifikasi biner (hoaks dan real) tanpa membedakan jenis misinformasi secara lebih spesifik.

5. Evaluasi model dibatasi pada metrik akurasi, presisi, *recall*, dan *F1-score*.

1.5 Urgensi Penelitian

Penyebaran hoaks di Asia Tenggara, khususnya Indonesia dan Malaysia, terus meningkat akibat pesatnya penggunaan media sosial dan rendahnya literasi digital. Situasi ini diperumit oleh keragaman bahasa dan konteks sosial-politik yang kompleks. Penelitian ini penting karena menawarkan solusi deteksi otomatis berbasis *Bidirectional Long Short-Term Memory* (Bi-LSTM) multibahasa, yang mampu memproses berita hoaks berbahasa Indonesia dan Melayu secara bersamaan. Pendekatan ini berkontribusi pada penguatan ketahanan informasi serta mengisi celah riset terkait bahasa-bahasa di kawasan ASEAN.

1.6 Manfaat Penelitian

Penelitian ini diharapkan memberikan kontribusi dalam dua aspek utama. Secara teoretis, penelitian ini dapat memperkaya kajian mengenai penerapan model *Bidirectional Long Short-Term Memory* (Bi-LSTM) untuk bahasa serumpun yang tergolong rendah sumber daya. Secara praktis, hasil penelitian ini berpotensi mendukung pengembangan sistem deteksi hoaks multibahasa yang dapat digunakan untuk memperkuat literasi digital dan mitigasi penyebaran disinformasi, khususnya di wilayah Asia Tenggara.

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA