

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi dalam bidang komunikasi telah mengubah pola produksi, distribusi, dan konsumsi informasi secara signifikan. Perubahan ini tidak hanya memberikan kemudahan dalam mengakses berita serta informasi, namun juga menciptakan peluang untuk penyebaran informasi palsu secara luas, yang biasa dikenal sebagai *hoax* atau berita palsu. Konten *hoax* seperti ini telah menciptakan ancaman yang serius pada kepercayaan publik, stabilitas demokrasi, dan keamanan siber pada tingkat nasional maupun global [1]. Faktor tersebut menciptakan lingkungan yang rentan terkena *hoax* atau berita palsu yang dapat menyebar dengan cepat dan luas. Kerentanan ini terutama sering terlihat pada masa-masa krusial, seperti saat penyelenggaraan pemilihan umum atau masa demonstrasi masyarakat.

Di Malaysia, pada jangka waktu tahun 2020 hingga 31 Mei 2022, tercatat sebanyak 3.288 kasus atau laporan yang berkaitan dengan penyebaran berita palsu [2]. Sementara itu, Kementerian Komunikasi dan Informatika (Kominfo) Indonesia menemukan 11.642 konten *hoax* dalam rentang waktu Agustus 2018 sampai Mei 2023 [3]. Lebih lanjut, berdasarkan laporan terbaru dari Kementerian Komunikasi dan Digital (Komdigi), sepanjang tahun 2024 telah diidentifikasi 1.923 konten hoaks, dengan kategori terbanyak berupa penipuan sebanyak 890 konten, disusul politik sebanyak 237 konten, pemerintahan sebanyak 214 konten, kesehatan sebanyak 163 konten, kebencanaan sebanyak 145 konten, serta kategori lainnya dalam jumlah lebih kecil [4]. Data tersebut memperlihatkan bahwa peredaran *hoax* masih menjadi permasalahan serius yang berpotensi menurunkan tingkat kepercayaan masyarakat terhadap informasi daring, mengganggu ketertiban umum, serta mengancam stabilitas sosial-politik dan keamanan siber nasional di kawasan Asia Tenggara.

Selain itu, bahasa Inggris memiliki peran penting di kedua negara tersebut, karena digunakan secara luas dalam pendidikan, bisnis, dan komunikasi daring lintas negara di kawasan Asia Tenggara. Dalam konteks linguistik regional, bahasa Inggris berfungsi sebagai *lingua franca* untuk negara-negara asia tenggara [5].

Berdasarkan tantangan tersebut, penelitian ini mengusulkan model DistilBERT (*Distilled Multilingual Bidirectional Encoder Representations from*

Transformers) yang dilatih untuk mendeteksi berita *hoax* berbahasa Indonesia, Inggris, dan Melayu. Model ini dipilih karena mayoritas penelitian masih terkonsentrasi pada bahasa-bahasa Barat seperti Inggris sedangkan penerapannya untuk bahasa Indonesia maupun bahasa Melayu masih terbatas dan belum banyak dieksplorasi [6]. Kontribusi utama dari penelitian ini terletak pada penggabungan tiga korpus linguistik, yaitu konten berita berbasis web berbahasa Indonesia, Melayu, dan Inggris.

1.2 Rumusan Masalah

1. Bagaimana cara mengimplementasikan model DistilBERT dalam mendeteksi berita *hoax* berbahasa Indonesia, Inggris, dan Melayu
2. Berapa akurasi model DistilBERT dalam mendeteksi berita *hoax* berbahasa Indonesia, Inggris, dan Melayu berdasarkan *evaluation metrics*-nya (akurasi, presisi, *recall*, *F1 Score*)?

1.3 Batasan Permasalahan

Penelitian ini dibatasi pada ruang lingkup berikut:

1. Bahasa yang digunakan untuk penelitian ini dibatasi pada Indonesia, Inggris, dan Melayu.
2. Model yang digunakan adalah DistilBERT.
3. Penelitian ini hanya mencakup klasifikasi biner (*Hoax* dan *Real*) tanpa membedakan jenis misinformasi yang lebih spesifik.
4. Evaluasi model hanya terbatas pada *accuracy*, *precision*, *recall*, *F1 Score*.

1.4 Tujuan Penelitian

1. Melatih model DistilBERT untuk membuktikan kemampuan model ini dalam mendeteksi berita *hoax* berbahasa Indonesia, Inggris, dan Melayu.
2. Mengukur *evaluation metrics* (*accuracy*, *precision*, *recall*, *F1 Score*) dari model DistilBERT ini dalam mendeteksi berita *hoax* berbahasa Indonesia, Inggris, dan Melayu.

1.5 Manfaat Penelitian

Penelitian ini diharapkan mampu memberikan kontribusi seperti, memperluas pemahaman terkait pemanfaatan model DistilBERT pada bahasa yang masih tergolong rendah sumber daya. Penelitian ini juga diharapkan dapat membantu mengurangi penyebaran misinformasi serta diharapkan untuk dapat mendukung pengembangan sistem deteksi *hoax* multibahasa yang dapat dimanfaatkan untuk meningkatkan literasi digital, terutama di kawasan Asia Tenggara.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
Bab ini membahas mengenai latar belakang permasalahan penelitian, tujuan penelitian dilakukan, serta batasan pada penelitian. Selain itu, bab ini juga akan menjelaskan mengenai masalah yang akan diselesaikan, serta kontribusi penelitian.
- Bab 2 LANDASAN TEORI
Bab ini menjelaskan teori dan konsep dasar yang digunakan untuk penelitian yang meliputi berita *hoax*, *Natural Language Processing*, BERT, DistilBERT.
- Bab 3 METODOLOGI PENELITIAN
Bab ini membahas mengenai metodologi yang digunakan untuk penelitian ini, seperti pengumpulan data, data *pre-processing*, *model training*, serta evaluasi model terhadap data uji.
- Bab 4 HASIL DAN DISKUSI
Bab ini menjelaskan mengenai proses *model training*, evaluasi model, dan analisa performa model yang diuji.
- Bab 5 KESIMPULAN DAN SARAN
Bab ini berisi rangkuman kesimpulan dari hasil penelitian yang telah dilakukan sebagai jawaban atas pertanyaan penelitian. Selain itu, bab ini juga menyajikan saran untuk penelitian selanjutnya atau pengembangan lebih lanjut yang berkaitan dengan topik penelitian.