

## 1. Pendahulua

### 1.1 Latar Belakang

Pada beberapa tahun belakangan ini penggunaan platform sosial media di Indonesia terus meningkat dalam rangka pemilihan presiden 2024, pada awal tahun 2022 tercatat bahwa pengguna sosial media di Indonesia telah mencapai sekitar 191 juta pengguna yang dimana telah mengalami peningkatan 21 juta pengguna atau 12.6% dari tahun sebelumnya [1]. Sosial media seperti Facebook, Twitter dan Instagram sudah umum digunakan untuk mengikuti perkembangan berita politik terbaru [2]. Penggunaan sosial media di Indonesia yang sebesar itu terdapat kerugian dan keuntungan, keuntungannya adalah berita-berita mudah diakses oleh siapapun, biayanya yang rendah dan penyebaran informasi yang cepat [3]. Adapun kerugiannya adalah memungkinkannya berita palsu yang mudah tersebar. Dampak dari penyebaran berita *hoax* benar-benar terjadi di masyarakat, dimana mereka mempercayai berita *hoax* yang tersebar sehingga mengakibatkan kerugian bagi orang lain yang dianggap sebagai objek isu tersebut. Berita *hoax* atau berita palsu adalah artikel berita menyesatkan yang sengaja dibuat untuk tujuan tertentu [4]. Terdapat dua motivasi utama alasan terbuatnya berita palsu. Pertama adalah uang, artikel berita seperti menjadi virus di media sosial yang dapat menarik pendapatan dari adanya iklan saat pengguna mengeklik situs berita tersebut. Motivasi kedua adalah ideologis, beberapa pembuat berita palsu memiliki misi khusus, seperti memajukan kandidat yang mereka sukai [5].

Pada riset yang dilakukan untuk mendeteksi *hoax* pada bahasa Indonesia dengan menggunakan metode *Long Short-Term Memory* (LSTM) menunjukkan bahwa rata-rata hasil nilai *precision*, *recall* dan *f1-Score* sebesar 0.819, 0.809, 0.807 dari beberapa hasil percobaannya [6]. Dengan menggunakan model *Word2Vec* diperoleh nilai rata-rata tertinggi dari *confusion matrix*. Pendeteksian berita *hoax* pada media berita politik Indonesia di Twitter ini akan menggunakan LSTM. LSTM merupakan salah satu model dari *Recurrent Neural Network* (RNN) yang sengaja dibuat untuk mengatasi keterbatasan RNN yang tidak bisa menangkap long-term dependencies atau dengan kata lain, LSTM mampu untuk mengingat informasi jangka panjang dan cukup baik untuk diterapkan dalam kasus seperti analisis sentimen dan pendeteksian berita *hoax* [7].

Metode LSTM tidak dapat melakukan penilaian terhadap *input* data jika data tersebut berupa teks atau *string*, maka dari itu jika input yang diterima berupa teks atau *string*, maka diperlukannya proses *feature extraction* yang akan mengubah teks atau *string* menjadi vektor numerik yang masing-masing mewakili kata. Proses ini disebut *word embedding* [8]. Selain *Word2Vec*, terdapat metode *word embedding* lainnya seperti *GloVe* (*Global Vectors for Word Representation*) dan *FastText*. *GloVe* memanfaatkan statistik global dari korpus kata dan menghasilkan vektor dengan meminimalkan loss dari hubungan kata dalam sebuah matriks co-occurrence, sementara *FastText* memperluas pendekatan *Word2Vec* dengan mempertimbangkan sub-kata sehingga lebih efektif dalam menangani kata-kata yang jarang muncul atau kata-kata baru [9].

Pada penelitian ini, *Word2Vec* dipilih karena keunggulannya dalam menangkap hubungan semantik antar kata dengan mengandalkan informasi lokal dari bahasa. Semantik yang dipelajari dari sebuah kata sangat dipengaruhi oleh kata-kata sekitarnya, yang membuatnya cocok untuk mendeteksi pola dalam teks seperti berita *hoax*. Sedangkan Metode LSTM dipilih dalam penelitian ini karena kemampuannya untuk menangkap dependensi jangka panjang yang sangat penting dalam menganalisis teks berita, terutama dalam konteks berita *hoax* yang sering kali membutuhkan pemahaman tentang hubungan antar kata dalam kalimat dan konteksnya. Namun, untuk membandingkan performa *Word2Vec*, digunakan juga metode *TF-IDF* sebagai pembanding. Perbandingan ini dilakukan untuk mengidentifikasi apakah model *Word2Vec* dapat menghasilkan performa yang lebih unggul dibandingkan metode lainnya dalam pendeteksian berita *hoax* pada media berita politik Indonesia di Twitter.

### 1.2 Topik dan Batasannya

Dalam penelitian ini, penulis merancang dan mengembangkan sebuah sistem untuk melakukan deteksi berita politik *hoax* berdasarkan cuitan berita politik di Indonesia Menggunakan model LSTM dan *Word2Vec*. Penelitian ini memiliki beberapa Batasan, yakni: (1) Data yang dianalisis cuitan berita politik di platform x; (2) Cuitan tersebut diklasifikasikan ke dalam dua kategori yaitu *Hoax* dan *Non-Hoax*

### 1.3 Tujuan

Tujuan utama dari penelitian ini adalah mengetahui performa sistem untuk mendeteksi berita *hoax* dengan menggunakan *deep learning* model LSTM dan *Word2Vec*.