

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pengaruh teknologi informasi telah masuk ke berbagai bidang aktivitas manusia dan memberikan manfaat yang besar dalam berbagai bidang yaitu dalam bidang pendidikan, kesehatan, politik, ekonomi, teknologi komunikasi dan informasi. Beberapa alat teknologi yang dimanfaatkan untuk komunikasi dan informasi yaitu *smartphone*, *email*, televisi, internet dan media lainnya. Penggunaan internet merupakan wadah yang tepat sebagai media komunikasi dan juga berperan sebagai media penyebaran informasi. Penyebaran informasi dengan mudah merupakan hal yang *positive* namun tidak semua informasi yang disebar di internet berupa fakta, banyaknya berita beredar di internet adalah berita yang tidak benar atau disebut dengan *hoax*.

Fenomena berita/informasi palsu atau berita *hoax* sangat mempengaruhi kehidupan sosial kita. Karena informasi membentuk cara kita memandang alam semesta, berita *hoax* dan fakta menyesatkan lainnya dapat memberikan sudut pandang yang berbeda. Penelitian di bidang intelijen menunjukkan bahwa rumor yang menyebar di media sosial meninggalkan dampak yang bertahan lama pada orang yang kurang pintar dan membuat mereka tidak mengambil keputusan yang tepat. Berita *hoax* digunakan untuk membangun ketakutan orang, memunculkan ide-ide rasis, dan mengarah pada penindasan dan kekerasan terhadap orang yang tidak bersalah.

Media sosial adalah platform yang digunakan oleh semua orang untuk berbagi informasi. Orang beralih ke media sosial digunakan untuk tempat komunikasi. Media sosial telah meledak sebagai kategori wacana online di mana orang bisa membuat dan berbagi informasi. Begitu pula dengan berita *hoax* yang juga berkembang di kalangan pengguna media sosial.

Latar belakang penelitian ini berfokus pada fenomena penyebaran berita *hoax* yang semakin marak di era digital, khususnya di Indonesia, di mana informasi dapat dengan cepat menyebar melalui berbagai platform media sosial.

Dalam konteks ini, kemampuan untuk mengidentifikasi dan mengklasifikasikan teks berita menjadi krusial untuk memerangi informasi yang tidak akurat. Oleh karena itu, penelitian ini mengusulkan penggunaan metode Word2Vec untuk mengubah teks berita menjadi representasi vektor yang lebih dapat dipahami oleh mesin, serta algoritma K Nearest Neighbor (KNN) untuk melakukan klasifikasi terhadap teks tersebut. Dengan menggabungkan kedua teknik ini, diharapkan dapat meningkatkan akurasi dan efisiensi dalam mengidentifikasi berita hoax, serta memberikan kontribusi yang signifikan terhadap upaya literasi digital dan penguatan masyarakat dalam menghadapi tantangan informasi yang salah. Penelitian pertama mengenai *Word2Vec* dilakukan oleh Mikolov menghasilkan sebuah model dari representasi kata dan frasa. Oleh karena itu, pada penelitian ini penulis akan menambahkan metode *Word2Vec* yang dikombinasikan dengan metode *K-Nearest Neighbor*.

Klasifikasi teks merupakan suatu metode pemisahan dan pengelompokan teks berdasarkan kategori-kategori yang telah didefinisikan sebelumnya. Klasifikasi teks telah diterapkan dan dikembangkan pada jenis penelitian data mining atau text mining, basis data, machine learning, dan aplikasi yang meliputi diagnosis medis, gambar, organisasi dokumen, dan lainnya (Septiani & Budi, 2022). Pengklasifikasian teks sangat dibutuhkan dalam berbagai macam aplikasi, terutama aplikasi yang jumlah dokumennya bertambah dengan cepat. Ada dua cara dalam penggolongan teks, yaitu clustering teks dan klasifikasi teks. Clustering teks berhubungan dengan menemukan sebuah struktur kelompok yang belum kelihatan (tak terpandu atau unsupervised) dari sekumpulan dokumen. Sedangkan pengklasifikasian teks dapat dianggap sebagai proses untuk membentuk golongan-golongan (kelas-kelas) dari dokumen berdasarkan pada kelas kelompok yang sudah diketahui sebelumnya (terpandu atau supervised).

Ada banyak algoritma yang umum digunakan untuk klasifikasi, salah satunya adalah *K-Nearest Neighbor*. Berdasarkan penelitian yang dilakukan oleh (Ramadhan, 2021) mengenai Indonesian Online News Topics Classification using *Word2Vec* dan *K-Nearest Neighbor*, pada penelitian ini dengan

menggunakan metode *Word2Vec* dan *K-Nearest Neighbor* mampu mengklasifikasikan beberapa topik berita berbahasa Indonesia untuk mengubah kata menjadi vektor yang bertujuan untuk memudahkan proses klasifikasi.

Algoritma *Word2Vec* merupakan jaringan saraf tiruan yang digunakan untuk memetakan kata ke variabel target yang dapat berupa kata maupun sekumpulan kata. Selain itu, teknik ini memberikan bobot kata yang direpresentasikan dalam bentuk vektor kata (Sabrila dkk., 2021). Metode ini menangkap hubungan semantic antar kata, sehingga bermanfaat untuk berbagai tugas pemrosesan bahasa alami seperti klasifikasi teks, analisis sentimen, dan terjemahan bahasa.

K-Nearest Neighbor (KNN) adalah algoritma yang digunakan untuk proses klasifikasi. Klasifikasi data dilakukan dengan berdasarkan mayoritas jumlah tetangga (neighbor) terdekat pada data yang akan diprediksi berdasarkan jumlah k yang telah ditentukan (Sabrila dkk., 2021). Algoritma KNN dapat digunakan untuk klasifikasi karena kesederhanaannya dalam penerapan. KNN merupakan algoritma berbasis instance yang sederhana, sehingga tidak memerlukan proses pelatihan yang rumit. Oleh Karena itu, algoritma ini cocok untuk diterapkan pada data pelatihan yang terbatas dan mampu memberikan hasil yang baik karena berfokus pada informasi spesifik.

Berdasarkan uraian di atas, penulis berinisiatif untuk melakukan penelitian dengan judul “*Klasifikasi teks berbahasa Indonesia menggunakan Word2Vec dan K-Nearest Neighbor untuk identifikasi berita hoax*”. Penulis berharap dengan adanya penelitian ini dapat mengetahui proses klasifikasi teks untuk identifikasi berita *hoax* serta mengetahui kinerja metode *Word2Vec* dan *K-Nearest Neighbor* dalam analisis berita *hoax*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang diuraikan di atas, dapat ditarik rumusan masalah sebagai berikut:

1. Bagaimana proses klasifikasi teks berbahasa Indonesia menggunakan metode *Word2vec* dan *K-Nearest Neighbor* untuk identifikasi berita *hoax*?

2. Bagaimana tingkat *accuracy* dalam pengklasifikasian teks berbahasa Indonesia dengan menggunakan metode *Word2Vec* dan *K-Nearest Neighbor* untuk identifikasi berita *hoax*?

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini dilakukan hanya untuk mengklasifikasikan teks berbahasa Indonesia untuk identifikasi berita *hoax*.
2. Untuk pengklasifikasian teks berbahasa Indonesia untuk identifikasi berita *hoax* yaitu dengan menggunakan algoritma *Word2Vec* dan *K-Nearest Neighbor*.
3. Data yang digunakan adalah berita berbahasa Indonesia, yang diperoleh dari website: *www.detik.com* dan *turnbackhoax.id* sebanyak 100 berita.
4. Kelas klasifikasi yang digunakan hanya kelas *hoax* dan *non hoax*.

1.4 Tujuan Penelitian

Adapun tujuan penelitian pada tugas akhir ini, yaitu:

1. Mengetahui proses klasifikasi teks berbahasa Indonesia menggunakan *Word2Vec* dan *K-Nearest Neighbor* untuk identifikasi berita *hoax*.
2. Mengetahui tingkat *accuracy* pengklasifikasian teks berbahasa Indonesia menggunakan *Word2Vec* dan *K-Nearest Neighbor* untuk identifikasi berita *hoax*.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

- a. Bagi Penulis
 1. Memahami cara kerja dari algoritma *Word2Vec* dan *K-Nearest Neighbor* dan hasil implementasi pada sistemnya.
 2. Sebagai portopolio untuk peneliti yang berguna untuk masa yang akan datang.

b. Bagi Universitas

1. Sebagai bahan referensi untuk penelitian di masa yang akan datang.
2. Sebagai bahan evaluasi bagi Universitas dalam mengembangkan keilmuan, dalam hal ini yang berkaitan dengan klasifikasi teks menggunakan *Word2Vec* dan *K-Nearest Neighbor*.