

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi informasi saat ini berkembang sangat pesat. Manusia dapat bersosialisasi satu sama lain tanpa harus bertemu secara langsung dengan menggunakan sarana media sosial. Keberadaan media sosial memungkinkan penggunaannya untuk dapat saling berbagi informasi, baik dalam bentuk teks, gambar, audio, dan video. Tak jarang, informasi berupa teks yang disebarluaskan merupakan pendapat pribadi dari penulis (pengguna). Namun, pengguna lain yang membaca informasi dari penulis kerap kali merasa kurang dapat memahami maksud sesungguhnya dari tulisan yang mereka baca. Informasi berupa teks seringkali membuat pembaca susah untuk merasakan perasaan aktual yang penulis ingin sampaikan. Hal tersebut berpeluang menimbulkan adanya kesalahpahaman atau miskomunikasi informasi yang juga bisa menyebabkan tersebarnya hoaks. Oleh karena itu, terdapat berbagai eksperimen menggunakan pembelajaran mesin yang luarannya dapat melakukan analisis maksud dari sebuah teks.

Suatu cara untuk memahami dan mengolah data berupa teks atau keilmuan yang berhubungan dengan komputasi opini, perasaan, serta kondisi hati disebut analisis sentimen (Budi, 2017). Proses analisis dapat dilakukan dengan bantuan seperti pembelajaran mesin. Terdapat berbagai jenis teknik pembelajaran mesin yang dapat digunakan untuk melakukan analisis sentimen, diantaranya Support Vector Machine, Random Forest, K-Nearest Neighbor, Decision Tree, serta Convolution Neural Network (Umarani et al, 2021). Saat analisis, terdapat proses yang dinamakan dengan ekstraksi fitur. Fungsi dari ekstraksi fitur adalah untuk menemukan ciri atau sebuah nilai yang khas dari suatu objek. Satu diantara jenis ekstraksi fitur adalah word embedding. Ketika melakukan word embedding, sebuah teks akan diubah menjadi bentuk vektor yang berisi kumpulan numerik. Metode yang dapat digunakan dalam word embedding, yaitu Word2Vec, GloVe, FastText, dan TF-IDF.

Media sosial tak hanya berperan sebagai wadah untuk bersosialisasi, tetapi juga bertindak sebagai tempat untuk bertukar opini publik. Segala topik perbincangan didiskusikan, salah satunya persoalan mengenai politik dan pemerintahan. Akhir-akhir ini, masyarakat Indonesia pro dan kontra terhadap keputusan Presiden Joko Widodo yang akan memindahkan ibu kota negara Indonesia yang awalnya berada di pulau Jawa ke pulau Kalimantan. Acapkali hal tersebut juga memicu perdebatan antar pengguna media sosial yang disebabkan oleh perbedaan pendapat. Berbagai berita disuguhkan dalam bentuk video pada media sosial YouTube. Di YouTube, pengguna memiliki hak untuk berkomentar dan mengungkapkan pandangannya pada kolom komentar dalam bentuk tulisan/teks. Untuk dapat mengetahui respon dari pengguna, maka dapat dilakukan analisis dengan mengimplementasikan Convolutional Neural Network mengenai opini pengguna terkait dengan pemindahan ibu kota negara Indonesia sehingga nantinya akan diperoleh informasi berupa hasil klasifikasi dari opini tersebut. Opini pengguna dikelompokkan berdasarkan kelasnya, yaitu opini yang bersifat positif, negatif, maupun netral.

Penelitian terkait klasifikasi sentimen dengan topik kebijakan publik di Indonesia yang datanya diambil dari API Twitter dan menggunakan metode GloVe dalam melakukan ekspansi fiturnya. Tahapan metode yang dilakukan terdiri dari data crawling, preprocessing data, representasi data, ekspansi fitur, serta pemodelan analisis sentimen. Terdapat dua algoritma klasifikasi yang digunakan, yakni Support Vector Machine (SVM) dan Naïve Bayes. Hasil akurasi dari kedua algoritma dibandingkan dan didapatkan kesimpulan bahwa akurasi untuk Naïve Bayes sebesar 0.7786 dan Support Vector Machine (SVM) sebesar 0.8323 (Sreya & Setiawan, 2022).

Analisis sentimen data berupa teks mengenai program vaksinasi Covid-19 menggunakan algoritma Support Vector Machine (SVM) dan metode word embedding FastText dikumpulkan dari media sosial Twitter. Data melalui proses balancing dataset, seleksi fitur, dan tuning parameter. Dari penelitian, diperoleh hasil bahwa tingkat akurasi yang didapatkan adalah sebesar 68% (Kusairi & Agustian, 2022).

Setelah melalui beberapa riset, banyak penelitian terdahulu yang mengangkat permasalahan tentang analisis sentimen dengan algoritma Convolutional Neural Network. Akan tetapi, masih sedikit yang melakukan komparasi hasil dalam penggunaan ekstraksi fitur menggunakan metode GloVe dengan FastText dalam suatu penelitian utuh. Maka dari itu, peneliti mengusung topik tersebut dalam penelitian ini dengan data teks yang bersumber dari komentar pengguna YouTube mengenai pemindahan ibu kota negara Indonesia.

1.2 Rumusan Masalah

Dari latar belakang yang telah dijelaskan pada bagian sebelumnya, penulis merumuskan masalah penelitian ini, yaitu “Bagaimana hasil akurasi antara penggunaan ekstraksi fitur metode GloVe dengan metode FastText pada algoritma Convolutional Neural Network dalam analisis sentimen komentar pengguna YouTube mengenai pemindahan ibu kota negara Indonesia?”. Penjelasan lebih terperinci dari permasalahan yang dirumuskan adalah sebagai berikut.

1. Apakah penentuan metode ekstraksi fitur yang digunakan dalam Convolutional Neural Network memiliki dampak yang signifikan terhadap hasil akurasi dalam analisis sentimen komentar YouTube mengenai pemindahan ibu kota negara Indonesia?
2. Apa metode ekstraksi fitur yang memiliki akurasi lebih tinggi pada analisis sentimen komentar YouTube mengenai pemindahan ibu kota negara Indonesia menggunakan Convolutional Neural Network?

1.3 Batasan Masalah

Penulis memberikan batasan permasalahan dalam penelitian ini karena banyaknya jenis algoritma yang dapat diterapkan dalam analisis sentimen agar mengurangi terjadinya ambiguitas dalam proses berlangsungnya percobaan. Berikut merupakan batasan masalahnya.

1. Data yang digunakan memiliki bahasan pokok, yakni terkait dengan opini publik mengenai pemindahan ibu kota negara Indonesia.
2. Data diambil dari platform media sosial YouTube dan berfokus pada keyword “Pemindahan Ibu Kota Negara Indonesia”. Video YouTube yang digunakan dalam penelitian merupakan lima video teratas yang muncul dengan keyword

tersebut. Banyaknya data yang digunakan dalam penelitian ini adalah 44957 data.

3. Data komentar YouTube yang digunakan adalah komentar dari tanggal diunggahnya masing-masing video hingga tanggal dilakukannya crawling data.
4. Ekstraksi fitur yang diimplementasikan hanya berfokus pada metode GloVe dan FastText.

1.4 Tujuan Penelitian

Selaras dengan latar belakang dan rumusan masalah yang telah dijelaskan di atas, tujuan umum dari dilakukannya penelitian ini adalah untuk mengetahui perbandingan tingkat akurasi ketika menggunakan ekstraksi fitur metode GloVe dan metode FastText dengan algoritma Convolutional Neural Network untuk suatu analisis sentimen. Detail tujuan dari penelitian ini adalah sebagai berikut.

1. Mengidentifikasi dampak dari penentuan metode ekstraksi fitur yang digunakan terhadap tingkat akurasi dari analisis sentimen komentar YouTube tentang pemindahan ibu kota negara Indonesia menggunakan Convolutional Neural Network.
2. Mengidentifikasi alur penggunaan dari setiap metode ekstraksi fitur yang digunakan.
3. Mengidentifikasi metode ekstraksi fitur yang memiliki hasil akurasi lebih tinggi pada saat diimplementasikan menggunakan Convolutional Neural Network pada analisis sentimen data berupa teks dari komentar YouTube mengenai pemindahan ibu kota negara Indonesia.

1.5 Manfaat Penelitian

Dari penelitian yang dilakukan, manfaat yang bisa diambil adalah dapat mengetahui metode ekstraksi fitur dengan tingkat akurasi lebih tinggi antara metode GloVe dan metode FastText menggunakan algoritma Convolutional Neural Network pada analisis sentimen dari komentar YouTube bertopik bahasan tentang pemindahan ibu kota negara Indonesia.