

1. Pendahuluan

1.1 Latar Belakang

Kenaikan harga bahan bakar merupakan masalah yang sering terjadi dalam beberapa tahun terakhir, salah satunya disebabkan oleh kenaikan harga bahan bakar internasional yang tidak stabil [1] Pemerintah Indonesia mulai menetapkan kenaikan harga bahan bakar pada awal September [2]. Kenaikan harga BBM ini menjadi perbincangan hangat di masyarakat sehingga menimbulkan pro dan kontra. Berdasarkan laporan keuangan Twitter kuartal ketiga tahun 2021, pengguna aktif Twitter telah mencapai 211 juta, dan Indonesia menjadi salah satu negara dengan pengguna Twitter aktif [3]. Twitter merupakan platform yang banyak digunakan dalam penelitian analisis sentimen, karena Twitter memiliki fitur yang disebut dengan tweet. Dalam analisis sentimen, tweet merupakan objek yang cocok digunakan sebagai data penelitian karena terdapat opini publik yang umumnya berbentuk teks. Proses analisis sentimen secara otomatis mengekstrak informasi dari data tentang isu-isu tertentu [4]. Teknik yang digunakan untuk menganalisis informasi dari platform media sosial seperti Twitter dan Facebook [5].

Meskipun sudah banyak penelitian di bidang ini, beberapa tantangan masih perlu diatasi, seperti meningkatkan kinerja model, mengurangi waktu pemrosesan, mengadaptasi teknik proses ke jenis dan domain data tertentu [5].

Deep learning baru-baru ini menjadi semakin populer dalam kasus klasifikasi teks karena penggunaan kumpulan data yang besar dan kompleks. Pembelajaran mesin memiliki beberapa area untuk perbaikan, termasuk waktu pemrosesan pelatihan yang cenderung lama dan kurang akurat [6]. Selain single deep learning, penggabungan deep learning juga semakin banyak digunakan. Penelitian [7] membuktikan bahwa hybrid deep learning memiliki performa yang lebih baik daripada machine learning.

Pada klasifikasi teks, deep learning tidak dapat langsung memproses data dalam bentuk teks, sehingga data teks harus diubah ke dalam bentuk vektor [8]. Perkembangan machine learning menjadi deep learning diikuti dengan perkembangan ekstraksi fitur kata, yaitu penyisipan kata [9]. Word embedding merupakan salah satu representasi kata yang paling populer dalam NLP.

Word2vec adalah teknik penyematan kata yang populer yang biasa digunakan dalam analisis sentimen dan dikembangkan pada tahun 2013 oleh Mikolov dkk [10]. Menurut Mikolov, ada dua arsitektur untuk Word2vec yaitu Continuous Bag of Words (CBOW) dan Skip-gram. Meskipun kedua arsitektur ini beroperasi secara berbeda, keduanya merepresentasikan kata-kata dengan cara yang sama. Performa model Word2vec dapat dipengaruhi oleh berbagai parameter selama pelatihan, termasuk dimensi vektor dan ukuran window [11]. Perbedaan arsitektur, dimensi vektor, dan ukuran window dapat mempengaruhi performa Word2vec.

Pada penelitian yang dilakukan oleh Farhan dkk., 2020 [3] yang memanfaatkan SVM untuk membandingkan dimensi vektor CBOW dan Skip-gram Word2vec pada analisis sentimen Tweet berbahasa Indonesia dengan menggunakan corpus wikipedia mendapatkan hasil 64.4% precision, 58% recall, dan 61.1% f-score.

Berdasarkan penelitian terdahulu menyatakan bahwa penggunaan Word2vec pada fitur ekstraksi dapat meningkatkan performansi model. Namun penelitian sebelumnya masih menggunakan metode machine learning yang dimana metode tersebut dianggap sudah terlalu klasik dan telah banyak metode terbaru yang digunakan saat ini. Oleh karena itu pada penelitian ini penulis melakukan hal yang serupa namun dalam mencari arsitektur dan dimensi vektor terbaik word2vec menggunakan metode terbaru yaitu hybrid deep learning CNN-BiLSTM pada dataset bahasa Indonesia yang berfokus pada kebijakan publik kenaikan harga BBM

1.2 Topik dan Batasannya

Topik dari penelitian ini yaitu bagaimana penerapan word2vec yang tepat sebagai model ekstraksi fitur terhadap analisis sentimen Twitter berbahasa Indonesia dengan menggunakan metode hybrid deep learning CNN-BiLSTM. Metode diimplementasikan dengan menggunakan bahasa python. Data yang dipakai berupa cuitan(tweet) yang berfokus pada kebijakan publik kenaikan BBM di Indonesia.

1.3 Tujuan

Tujuan dari penelitian ini yaitu menganalisis faktor-faktor yang mempengaruhi hasil penerapan metode word2vec seperti jenis arsitektur model dan ukuran dimensi pada analisis sentimen Twitter berbahasa Indonesia. Hasil penerapan akan dilihat dari skor presisi, recall dan f-score.

Organisasi Tulisan

Penulisan tugas akhir ini ditulis sebagai berikut. Bagian 1 menjelaskan tentang pendahuluan. Bagian 2 menjelaskan studi terkait. Bagian 3 menjelaskan sistem yang dibangun. Bagian 4 menjelaskan evaluasi yang berisi hasil dan diskusi. Bagian 5 menjelaskan kesimpulan

