

1. Pendahuluan

Latar Belakang

Perkembangan media sosial telah meningkat dengan cukup pesat, dengan menggunakan internet sebagai alat komunikasi tentunya masyarakat akan mudah untuk menggunakannya. Hal inilah yang melatarbelakangi adanya perubahan dalam bagaimana cara masyarakat berkomunikasi yang awalnya konvensional menjadi modern dan digital [1]. Media sosial itu sendiri adalah sebuah platform yang digunakan untuk berkomunikasi dengan seseorang ataupun khalayak ramai dapat melalui konten berupa teks, video maupun foto. Pada media sosial juga seseorang dapat mengeluarkan pendapatnya terhadap informasi yang diperoleh.

Media sosial yang memiliki fitur untuk mencari informasi sekaligus berkomunikasi akan mendapatkan banyak perhatian dari penggunanya. Berdasarkan hasil riset *Wearesocial Hootsuite* pada bulan Januari 2019 pengguna media sosial di Indonesia mencapai 150 juta atau sebesar 56% dari total populasi [2]. Media sosial yang ada di Indonesia sangatlah beragam, mulai dari fungsi hingga ciri khasnya masing-masing. Seperti Twitter dan Facebook, keduanya memiliki keunggulan dalam menyampaikan informasi atau opini, Twitter dengan fitur *trending topic* dan Facebook dengan *feeds* postingannya. Menurut laporan statistika, berjudul "*Pengguna Twitter Indonesia Masuk Daftar Terbanyak di Dunia, Urutan Berapa?*", pengguna Twitter di Indonesia telah mencapai 16,32 juta pengguna pada tahun 2022 awal [3]. Dengan angka pengguna yang cukup besar, hal ini dapat dimanfaatkan oleh sebuah organisasi atau berita untuk menilai bagaimana opini masyarakat terhadap suatu masalah, apakah hasilnya cenderung baik, buruk atau netral.

Kebijakan pemerintah merupakan salah satu opini publik yang sering menjadi *trending topic* di media sosial Twitter khususnya *tweet* mengenai politik, Pertamina, BPJS, pertalite, dan bansos. Akan tetapi *tweet* seringkali memiliki konteks yang rumit dan tidak jelas, dan menjadi sulit untuk menentukan emosi yang sebenarnya sehingga diperlukannya analisis sentimen. Analisis sentimen adalah sebuah proses yang berfungsi untuk menentukan opini, pendapat, dan juga emosi yang diterjemahkan melalui teks, biasanya sentimen ini nantinya diklasifikasikan menjadi opini positif dan negatif [4]. Analisis sentimen dianggap sebagai salah satu sub bagian terpenting dari Natural Research dalam pemrosesan bahasa (NLP) [5]. Proses analisis sentimen dapat berupa konten yang berisi ulasan teks, forum, *tweet* atau blog [6]. Dengan menggunakan analisis sentimen, informasi yang beredar di media sosial dapat diolah menjadi data yang lebih terstruktur [7].

Penelitian ini akan berkontribusi untuk menganalisis data sentimen publik pengguna Twitter menggunakan metode CNN untuk pemodelannya lalu menggabungkan fitur ekstraksi dari TF-IDF N-gram dengan fitur ekstraksi RoBERTa dan fitur ekspansi dari Glove. *Labeling* pada penelitian ini dibagi menjadi 2 *labeling* manual dan *labeling* system. Untuk *labeling* manual dibagi menjadi 3 kelas yaitu positif, negatif, dan netral. *Tweet* positif dapat berupa motivasi ataupun ilmu pengetahuan. Untuk *tweet* negatif dapat berupa ujaran kebencian, hinaan, maupun penistaan. Sedangkan *tweet* netral berupa kata-kata yang bersifat objektif, tidak mengandung unsur positif dan negatif. Untuk *labeling* system dilakukan dengan melakukan survey kata dengan range *labeling* -5 hingga 5 lalu di konversi menjadi 3 label. Penelitian ini merupakan yang pertama dalam analisis sentimen menggunakan 2 cara *labeling* dengan menggunakan model CNN dan RoBERTa dengan fitur ekspansi Glove.

Topik dan Batasannya

Topik pada penelitian ini akan menggunakan metode *Robustly Optimized BERT Pretrained Approach* (RoBERTa) untuk *word embedding* sebagai fitur ekstraksi dan dibantu dengan metode *deep learning* yaitu *Convolutional Neural Network* (CNN). Penelitian ini memiliki batasan yaitu: (1) dataset yang digunakan adalah cuitan atau *tweet* di Twitter dengan *keyword* yang berkaitan dengan kebijakan pemerintah Indonesia; (2) Metode yang digunakan dalam penelitian ini menggunakan *Convolutional Neural Network* (CNN) dan *Robustly Optimized BERT Pretraining Approach* (RoBERTa); (3) Tidak mendeteksi emoji yang terdapat pada *tweet*.

Tujuan

Tujuan dari penelitian ini adalah untuk mengimplementasikan sistem analisis sentimen menggunakan kombinasi *Convolutional Neural Network* (CNN) dan *Robustly Optimized BERT Pretrained Approach* (RoBERTa) dan mengukur sistem performansi yang dibangun.

Organisasi Tulisan

Penelitian disusun sebagai berikut. Bagian pertama berisi pendahuluan. Bagian kedua berisi mengenai penjelasan studi terkait dari penelitian ini. Bagian ketiga berisi penjelasan metode yang digunakan dan penerapannya. Bagian keempat berisi hasil pengujian dan analisisnya. Bagian kelima berisi kesimpulan dari penelitian ini.