

Analisis Sentimen pada Media Sosial Menggunakan Word2Vec dan Gated Recurrent Unit (GRU) dengan Optimasi Genetic Algorithm

Syafa Fahreza¹, Erwin Budi Setiawan²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹syafafahreza@students.telkomuniversity.ac.id, ²erwinbudisetiawan@telkomuniversity.ac.id

1. Pendahuluan

Latar Belakang

Kemajuan teknologi informasi telah mengubah peran media sosial dari sekedar tempat penyimpanan informasi menjadi sebuah platform untuk menyampaikan pendapat dan aspirasi [1]. Salah satu platform media sosial yang banyak digunakan oleh masyarakat adalah Twitter [2]. Dengan jumlah pengguna sebanyak 19,5 juta, Indonesia menempati urutan kelima di dunia untuk pengguna Twitter, informasi ini bersumber dari Kominfo [3]. Pada Twitter, pengguna dapat dengan bebas memberikan opini sesuai dengan hati nurani mereka. Namun, tidak semua opini yang disampaikan bersifat positif [1], sehingga diperlukan analisis sentimen untuk mengklasifikasikan sebuah opini tersebut sebagai positif atau negatif. Analisis sentimen ini menjadi semakin penting dalam konteks pemilihan Presiden, yang merupakan salah satu fase krusial dalam sistem demokrasi suatu negara.

Analisis sentimen adalah proses mengidentifikasi, memahami, dan mengevaluasi pendapat, sikap, atau perasaan yang terkandung dalam teks, seperti ulasan, tweet, atau unggahan media sosial. Tujuan utama dari analisis sentimen adalah untuk menentukan apakah sebuah teks atau opini bersifat positif atau negatif [4]. Dalam penelitian ini, sentiment analisis akan dilakukan pada topik pemilihan presiden di Indonesia tahun 2024. Sentimen terkait dengan calon presiden, kebijakan, dan isu-isu politik yang relevan mencerminkan pandangan, harapan, kekhawatiran, dan preferensi pemilih. Melalui analisis sentimen, kita dapat memahami persepsi dan respons masyarakat terhadap kandidat serta isu-isu yang muncul selama kampanye pemilihan presiden. Informasi ini memberikan wawasan berharga bagi para pemimpin politik, kandidat, dan pemilih dalam memahami dinamika politik yang berkembang dan menginformasikan strategi kampanye serta kebijakan yang lebih responsif terhadap kebutuhan masyarakat.

Sudah berkembang banyak metode deep learning yang dapat digunakan untuk melakukan analisis sentimen, salah satunya adalah Gated Recurrent Unit (GRU). GRU merupakan salah satu jenis arsitektur jaringan syaraf tiruan yang digunakan dalam NLP untuk memodelkan urutan data, termasuk teks atau data bahasa manusia [6-8]. GRU memiliki kemiripan dengan *Long Short-Term Memory* (LSTM), namun dengan parameter yang lebih sedikit dari LSTM [9]. Pada GRU, terdapat mekanisme *gating* yang memungkinkan aliran informasi yang lebih baik antar sel pengulangan, sehingga GRU dapat memanggil kembali informasi yang relevan dari konteks sebelumnya dengan cara yang lebih efisien [8]. Hal ini membuat GRU menjadi pilihan yang baik dalam menganalisis teks yang panjang dengan urutan kata yang kompleks, seperti unggahan media sosial.

Penelitian ini menggunakan Word2Vec sebagai metode ekspansi fitur dan *Genetic Algorithm* sebagai optimasi fitur. Word2Vec sendiri digunakan untuk mengubah data menjadi vektor dalam bentuk angka, sedangkan *Genetic Algorithm* digunakan untuk menyelesaikan masalah optimasi [11]. Menurut Mikolov, Word2Vec memiliki beberapa kelebihan, diantaranya Word2Vec dapat bekerja dengan cepat dan efisien untuk data yang besar, mampu melihat dan menangkap hubungan semantik antar kata, dan mampu melakukan generalisasi ke kata-kata lain [12]. Di sisi lain, *Genetic Algorithm* memiliki keunggulan dalam mengatasi tantangan komputasi yang luas secara efektif.

Pengabungan *Genetic Algorithm* (GA) dan Word2Vec digunakan untuk mengoptimalkan *Gated Recurrent Unit* (GRU), prosesnya terjadi dalam beberapa langkah yang saling melengkapi. Pertama, Word2Vec digunakan untuk mengubah kata-kata dalam teks menjadi representasi vektor yang merefleksikan konteksnya. Vektor ini menyediakan pemahaman yang lebih baik tentang hubungan antara kata-kata dalam kalimat. Kemudian, GA digunakan untuk mengoptimalkan parameter GRU dengan mempertimbangkan representasi vektor tersebut. GA mencoba berbagai kombinasi parameter GRU dan mengevaluasi kinerjanya dalam tugas analisis sentimen menggunakan metrik yang ditetapkan. GA menggunakan mekanisme seleksi alamiah dan operasi genetik seperti *crossover* dan mutasi untuk menemukan parameter-parameter yang lebih baik yang meningkatkan kinerja GRU dalam analisis sentimen. Dengan cara ini, gabungan antara GA dan Word2Vec membantu meningkatkan representasi kata-kata dan mengoptimalkan struktur model GRU untuk meningkatkan akurasi dan kinerja dalam analisis sentimen.

Kontribusi utama dari penelitian ini adalah mengoptimalkan model *Gated Recurrent Unit* (GRU) menggunakan *Genetic Algorithm* dan mengombinasikan ekspansi fitur dengan Word2Vec untuk sentiment analisis pada topik pemilihan presiden 2024 di Indonesia. Sepengetahuan penulis, belum ada penelitian sebelumnya yang mengeksplorasi kombinasi ini, dan penulis melihat adanya potensi peningkatan akurasi melalui integrasi ini. Untuk menghasilkan model dengan akurasi yang tinggi, penulis menggunakan berbagai metode,

termasuk GRU, TF-IDF, Word2Vec, dan *Genetic Algorithm*. Penelitian ini melibatkan berbagai skenario, seperti pemilihan rasio optimal pada *baseline*, penentuan fitur maksimum dalam ekstraksi fitur, pemilihan korpus untuk ekspansi fitur, dan penggunaan optimasi untuk meningkatkan akurasi model.

Topik dan Batasannya

Batasan masalah dalam penelitian ini adalah data yang digunakan diambil dari media sosial Twitter. Tweet yang digunakan hanya tweet yang bahasa Indonesia dengan topik pemilihan presiden di Indonesia pada tahun 2024. Data akan dikelompokkan menjadi dua kelas yaitu positif dan negatif.

Tujuan

Penelitian ini bertujuan untuk melakukan analisis sentimen pada data Twitter dengan menerapkan teknik Word2Vec dan Gated Recurrent Unit (GRU), yang dioptimalkan menggunakan algoritma genetika. Tujuan kedua penelitian adalah untuk mengevaluasi tingkat performansi yang dihasilkan dari penggabungan teknik Word2Vec dan GRU dengan optimasi menggunakan algoritma genetika, dalam konteks analisis sentimen data Twitter. Diharapkan penelitian ini dapat memberikan wawasan mengenai efektivitas gabungan teknik tersebut dalam meningkatkan akurasi dan kemampuan model dalam memahami dan mengklasifikasikan sentimen dalam teks yang kompleks dan bervariasi seperti yang ditemukan di platform media sosial.

Organisasi Tulisan

Struktur penulisan dalam penelitian ini diatur sebagai berikut. Bagian awal berfokus pada pendahuluan, diikuti oleh bagian kedua yang membahas studi terkait yang relevan dengan penelitian ini. Bagian ketiga menjelaskan metode yang digunakan serta penerapan sistem yang telah dibangun. Bagian keempat memuat hasil pengujian dan analisisnya. Terakhir, bagian kelima berisi kesimpulan dari penelitian ini.