

1. Pendahuluan

Latar Belakang

Di media sosial telah ramai mempercibangkan pemilihan presiden di masa yang akan datang. Pemilihan presiden telah menjadi bagian penting bagi masyarakat Indonesia. Tentunya dalam pemilihan ini sangat berkaitan dengan adanya keterlibatan politik dimana dunia politik selalu dikaitkan dengan isu hoaks. Saat masa pemilihan presiden Indonesia 2019, terjadi banyak penyebaran konten hoaks yang bertujuan untuk merusak citra kandidat lawan dan pendukung partai serta pendukung kandidat lawan oleh politisi [1]. Di Indonesia konten hoaks dalam berita politik sudah sering terjadi terutama di media sosial [1]. Penggunaan media sosial tidak pernah lepas dari kehidupan sosial sehari-hari. Wadah yang sering digunakan sebagai media komunikasi dan menyebarkan berita termasuk berita politik dengan membagikan visi dan misi politik sehingga mengikat masyarakat untuk bergerak memberikan dukungan kepada partai [2].

Disinformasi dalam penyebaran konten politik di Indonesia pun dilancarkan baik dalam skala nasional hingga internasional. Contoh bentuk penyebaran konten disinformasi selama pemilihan presiden terjadi saat disinformasi tersebar guna mendorong Donald Trump memenangkan pemilihan [3, 4]. Disinformasi merupakan bentuk isu negatif atau berita palsu yang menjadi perhatian masyarakat setiap tahunnya dengan menyebar secara luas dan cepat untuk tujuan menyesatkan atau menipu golongan dan masyarakat tertentu. Twitter memiliki potensial dalam melakukan penyebaran disinformasi dikarenakan banyaknya pengguna twitter saat ini [5]. Sehingga, data twitter dapat digunakan untuk analisis dengan proses teks mining yang hasilnya digunakan untuk analisis lebih lanjut dengan memanfaatkan analisis sentimen.

Analisis sentimen adalah bidang penelitian komputasional berdasarkan bidang text mining yang mengacu pada pendapat, sentimen dan emosi orang melalui kata – kata atau kalimat untuk menuju suatu entitas [6]. Berdasarkan penelitian [5], sentimen orang – orang bisa tergambarkan dengan baik melalui media sosial salah satunya pada twitter. Selama pemilihan presiden tentunya masyarakat dapat mengekspresikan pendapat masing – masing dengan adanya sentimen positif maupun sentimen negatif yang kebanyakan mengarah kepada kandidat lawan. Menurut penelitian [7], analisis sentimen bisa memberikan hasil prediksi yang akurat untuk melihat nilai tingkat akurasi berdasarkan sentimen positif dan negatif.

Menurut penelitian [6] metode Support Vector Machine memberikan hasil akurasi terbaik untuk analisis sentimen dengan tingkat keakuratan mencapai 77,4%. Sementara itu, penelitian [8] menunjukkan bahwa metode Random Forest dapat memberikan kinerja akurasi yang baik. Menurut penelitian [9], penggunaan algoritma yang salah satunya ialah Gradient Boosting menghasilkan nilai akurasi 81.82%. Pada ketiga algoritma tersebut akan diuji secara bersamaan dengan metode Ensemble. Menurut penelitian [10], metode Ensemble memberikan hasil akurat yang cukup tinggi hingga mencapai nilai akurasi 97.66% dengan pengujian tiga algoritma yang mana salah satu algoritma nya adalah algoritma SVM.

Penelitian ini bertujuan untuk menemukan algoritma terbaik, baik secara individu maupun melalui gabungan kombinasi dengan Ensemble, yang dapat menghasilkan akurasi yang tinggi dan efisien dalam menganalisis sentimen. Hal ini penting untuk mengidentifikasi bentuk-bentuk disinformasi yang tersebar di media sosial. Dalam hal ini dilihat bagaimana sentimen masyarakat di media sosial terhadap pemilihan presiden Indonesia 2024 dari setiap konten yang memiliki sentimen positif dan negatif. Menganalisis disinformasi merupakan tantangan bagi banyak orang karena data yang diperoleh bervariasi, seperti topik, bahasa, dan platform media. Dalam konteks berita palsu, sering terdapat perubahan gaya bahasa yang digunakan untuk menjelekkan atau mencemooh berita tersebut [11]. Ensemble Learning adalah metode pemodelan dengan menggabungkan beberapa algoritma bersamaan untuk membuat prediksi yang akurat daripada hanya dengan menggunakan satu algoritma saja paper [12]. Pada proses nya dilakukan pengujian dengan menggunakan algoritma Random Forest, SVM (Support Vector Machine), dan Gradient Boosting Classifier yang kemudian ketiga metode tersebut dijadikan base model dalam algoritma Stacked Ensemble untuk menganalisis sentimen yang berkaitan dengan pilpres 2024 berdasarkan sentimen positif dan negatif.

Topik dan Batasannya

Penyebaran konten disinformasi dalam pemilihan presiden melalui media sosial menjadi isu yang mendesak dan memerlukan pemahaman yang lebih mendalam. Dalam konteks pemilihan presiden di masa depan, media sosial telah menjadi wadah penting bagi masyarakat Indonesia dalam berinteraksi dan mendapatkan informasi politik. Namun, fenomena penyebaran konten disinformasi yang sering kali tersebar di media sosial menjadi tantangan yang harus diatasi. Terlebih lagi, media sosial, khususnya Twitter, telah menjadi sumber utama

penyebaran informasi dan pendapat masyarakat terkait pemilihan presiden. Oleh karena itu, penting untuk mengidentifikasi masalah-masalah yang terkait dengan penyebaran konten hoaks dan disinformasi politik melalui media sosial, serta pentingnya analisis sentimen dan identifikasi algoritma terbaik untuk memahami pendapat masyarakat secara akurat. Untuk mengatasi masalah ini, penelitian ini menggunakan pendekatan analisis sentimen dengan menerapkan beberapa metode klasifikasi seperti Support Vector Machine (SVM), Random Forest, dan Gradient Boosting Classifier. Ketiga algoritma tersebut digabungkan menjadi satu dalam bentuk model dasar (base model) untuk membentuk algoritma stacked ensemble. Dalam penelitian ini, fokus utama adalah mencari metode yang paling efektif dan dapat menghasilkan tingkat akurasi yang tinggi dalam menganalisis sentimen terkait pemilihan presiden.

Dalam penelitian ini, ada beberapa batasan masalah yang perlu diperhatikan. Pertama, keterbatasan terkait jumlah dataset nya dikarenakan Twitter menerapkan API berbayar sehingga sulit untuk kembali melakukan crawling data untuk menambah jumlah dataset. Kemudian batasan lainnya, terkait dalam proses pelabelan data yang dilakukan secara manual. Dikarenakan belum adanya pelabelan otomatis yang valid sehingga diharuskan menggunakan cara manual. Selain itu, sumber daya komputasi yang terbatas ketika menggunakan tools khusus untuk pemrograman berbahasa python.

Tujuan

Adapun beberapa tujuan dalam penelitian ini sebagai berikut: pertama, memanfaatkan sentimen dalam melakukan klasifikasi antara disinformasi dan non-disinformasi. Kemudian, membandingkan kinerja tiga metode klasifikasi yaitu SVM, Random Forest, dan Gradient Boosting Classifier. Selanjutnya, Mengembangkan model dasar (base model) berdasarkan gabungan ketiga algoritma tersebut untuk membentuk algoritma stacked ensemble yang dapat meningkatkan akurasi dalam analisis sentimen. Kemudian, Menentukan metode klasifikasi yang paling efektif dan menghasilkan tingkat akurasi yang tinggi untuk menganalisis sentimen terkait pemilihan presiden di media sosial.

Organisasi Tulisan

Setelah bagian pendahuluan pada penelitian ini, kemudian dilanjutkan dengan bagian studi terkait mengenai sentimen, disinformasi, SVM, Random Forest, Gradient Boosting Classifier, dan Stacked Ensemble. Bagian selanjutnya, sistem yang dibangun, dimana menjelaskan metodologi yang digunakan dalam penelitian ini. Bagian berikutnya, evaluasi, menjelaskan hasil kinerja dan performansi pada model yang sudah dibangun dan penjelasan terkait evaluasi. Bagian terakhir, kesimpulan, menjelaskan rangkuman dari keseluruhan penelitian yang dilakukan.