

1. Pendahuluan

Kebakaran hutan dan kabut asap yang diakibatkannya merupakan bencana ekologi yang menjadi perhatian nasional dan internasional. Penyebab utama kebakaran hutan di Indonesia adalah pembukaan lahan menggunakan api untuk perluasan kebun kelapa sawit di lingkungan lahan gambut [14]. Pada bulan Januari hingga September tahun 2019, terjadi kebakaran hutan di Indonesia yang membakar 942.484 hektar hutan [9]. Kejadian tersebut menimbulkan reaksi yang bervariasi di media sosial. Tidak hanya orang Indonesia, penduduk negara lain yang terdampak oleh kabut asap kebakaran hutan di Indonesia juga ikut membicarakan bencana tersebut pada media sosial, seperti Twitter.

Twitter merupakan layanan jejaring sosial dimana pengguna dapat membuat pesan singkat sepanjang 280 karakter bernama *tweet* atau cuitan. Twitter melaporkan 237,8 juta pengguna aktif harian pada laporan keuangan Twitter kuartal kedua tahun 2022 [3]. Indonesia termasuk sebagai salah satu negara dengan pengguna Twitter terbanyak di dunia, mencapai 24 juta pengguna aktif pada Januari 2023. [2] Twitter telah menjadi sarana utama masyarakat Indonesia untuk mengungkapkan opininya. Opini tersebut dapat dimanfaatkan sebagai data untuk menganalisis opini publik terhadap suatu topik menggunakan metode analisis sentimen.

Analisis sentimen atau *opinion mining* adalah sebuah teknik yang menganalisis dan menentukan opini, perasaan, penilaian, dan sikap seseorang dengan mengekstrak elemen subjektif dalam sebuah teks [6]. Terdapat dua pendekatan yang dapat diterapkan pada analisis sentimen, yaitu pendekatan dengan model berdasarkan *lexicon* dan berdasarkan *machine learning*. Pendekatan dengan model berdasarkan *lexicon* menggunakan kata-kata yang diberikan bobot, yang dapat dibuat secara manual atau secara otomatis, dan tidak memerlukan proses pelatihan model [13]. Sedangkan, pendekatan dengan model berdasarkan *machine learning*, atau pendekatan *supervised learning*, menggunakan teks yang diberi label untuk melatih model klasifikasi dan memiliki kelebihan lebih akurat dibandingkan model berdasarkan *lexicon* [10]. Berdasarkan akurasi yang lebih tinggi, pendekatan berdasarkan *machine learning* digunakan di penelitian ini.

Terdapat beberapa model klasifikasi analisis sentimen dengan pendekatan *machine learning* yang umum digunakan, salah satunya adalah *Support Vector Machine* (SVM). Berdasarkan riset yang dilakukan Sharma dan Dey (2012) [11] tentang pengaruh seleksi fitur terhadap performansi algoritma klasifikasi SVM, *Naïve Bayes*, *Maximum Entropy*, *Decision Tree*, *K-Nearest Neighbor*, *Winnow Classifier*, dan *Ada-boost Classifier*; SVM memperoleh tingkat performansi keseluruhan yang tertinggi dalam klasifikasi teks dan menghasilkan akurasi hingga 90,15%. Penelitian yang dilakukan oleh Utami (2017) [17] tentang analisis sentimen opini publik berita terhadap kebakaran hutan dengan membandingkan algoritma SVM dan *K-Nearest Neighbor* (K-NN) berbasis *Particle Swarm Optimization* (PSO) mendapatkan hasil perhitungan menggunakan metode SVM memiliki akurasi sebesar 80,83% dengan AUC sebesar 0,947. Performansi terbaik diperoleh metode SVM+PSO dengan akurasi sebesar 86,11% dan AUC sebesar 0,922.

Pada penelitian ini, penulis mempelajari bagaimana penerapan dan performansi metode SVM dalam menganalisis sentimen pengguna Twitter terhadap kebakaran hutan dan kabut asap dan bagaimana tingkat performansi metode SVM dalam permasalahan tersebut. Analisis sentimen dilakukan terhadap cuitan yang berkaitan dengan topik kebakaran hutan dan kabut asap berbahasa Indonesia dan dikirim dalam jangka waktu 1 September 2019 hingga 30 September 2019. Penelitian ini memiliki tujuan untuk mengevaluasi penerapan SVM dalam analisis sentimen Twitter terhadap kebakaran hutan dan kabut asap. Evaluasi dilakukan menggunakan *performance metrics* yang terdiri dari *precision*, *recall* dan *f1-score*. Kontribusi utama penelitian ini adalah dibangun sebuah model klasifikasi yang dapat memprediksi sentimen pengguna Twitter terhadap kebakaran hutan dan kabut asap dengan akurat.

Selanjutnya, paper disusun menjadi empat bagian. Studi terkait pada analisis sentimen didiskusikan pada Bagian 2. Penjelasan sistem yang dibangun terdapat pada Bagian 3. Evaluasi yang terdiri dari hasil pengujian dan analisis hasil pengujian dipaparkan pada Bagian 4. Kesimpulan dan saran tertulis pada Bagian 5.