

BAB I PENDAHULUAN

1.1 Latar Belakang

Teknologi saat ini berkembang karena konsekuensi dari pertumbuhan internet yang meluas di seluruh dunia. Selaku konsekuensi dari meningkatnya kebutuhan warga dengan bermacam tipe informasi yang nantinya akan digunakan untuk kebutuhan yang berbeda - beda. Dengan timbulnya media sosial di Indonesia, menjadi menarik untuk mengklasifikasikan opini publik tentang perokok yang senantiasa jadi bahan perdebatan. Penggunaan metodologi serta perlengkapan analisis yang cocok untuk membuat analisis media lebih mudah serta lebih efektif. Utamanya media sosial mempunyai akibat yang signifikan pada cara bangsa beroperasi. Setiap orang mempunyai hak untuk membaca serta mengatakan pendapatnya secara online, baik positif ataupun negatif.

Nilai data yang dimiliki pengguna platform media sosial ini sangat berarti. Orang - orang dari segala dunia dengan leluasa mengatakan komentar mereka di web media sosial semacam YouTube, Twitter, Weibo, ataupun WeChat, dan lewat toko online ataupun web individu. Aspek politik serta ekonomi warga mereka menjadi bahan perdebatan di antara banyak pengguna. Web media sosial yang kerap digunakan untuk memperdebatkan isu - isu politik serta sosial adalah Twitter [1]. Twitter merupakan website microblogging yang terkenal di mana pengguna bisa mempublikasikan pembaruan status (diucap "tweet"). Banyak suka, tidak suka, serta kontribusinya pada bermacam topik tercantum dalam tweet ini [2].

Orang-orang di Indonesia dan banyak negara lain menggunakan Twitter sebagai bagian dari platform media sosial utama. Yang mengejutkan 73% hingga 87% pengguna internet yang membaca ulasan online mengklaim bahwa mereka terpengaruh ulasan dalam memutuskan apakah akan membeli produk atau menggunakan layanan. Dari 81% pengguna internet telah melakukan setidaknya 1 pencarian produk [3]. Konsekuensinya, Twitter bertindak sebagai tempat mengumpulkan informasi yang diungkapkan secara publik melalui tweet yang

diposting ke timeline [4]. Istilah "tweet" berkaitan dengan setiap tulisan atau teks yang dikirim ke layanan Twitter. Twitter sering digunakan untuk bercerita tentang peristiwa selain berbagi informasi. Orang sering mengungkapkan emosinya saat bercerita [5].

Trending Topics merupakan fitur di Twitter yang mempermudah pengguna dalam mencari data tentang topik tertentu yang sedang trending. Tetapi, sebagian pengguna bisa jadi kesusahan menguasai topik pendek yang sedang tren di Twitter [6]. Sebagian riset lebih dahulu menggunakan informasi yang dikumpulkan dari Twitter untuk melaksanakan riset tentang topik yang berkaitan dengan tim. World Health Organization menegaskan bila tren merokok dikala ini terus bersinambung, hendak menyebabkan korban 10 juta orang pada tahun 2030, 70% di antara lain terletak di negeri berkembang seperti Indonesia. Meningkatnya jumlah perokok berakibat negatif terhadap kesehatan secara signifikan [7].

Selain itu, sebagian besar perokok penduduk Indonesia, memberikan kontribusi signifikan terhadap tingkat merokok terbesar di dunia. Aliansi Pengendalian Tembakau Asia Tenggara memperkirakan Indonesia memiliki 65,19 juta perokok pada tahun 2019, menjadikannya negara dengan prevalensi rokok tertinggi di kawasan ini [8]. Menurut World Health Organization pada tahun 2017, prevalensi perokok muda di Indonesia sebesar 11,5%; pria memiliki persentase lebih tinggi daripada wanita (1,5% vs 21,4%) [9]. Prevalensi merokok pada penduduk usia 10 hingga 18 tahun meningkat dari 7,2% menjadi 9,1% sejak tahun 2016, menunjukkan kecenderungan peningkatan prevalensi merokok secara nasional lebih besar pada kelompok anak-anak dan remaja [10]. Di Provinsi Jawa Barat, jumlah penduduk berusia di atas 15 tahun yang merokok meningkat dari 33,19% pada tahun 2017 menjadi 35,78% pada tahun 2018 [11]. Merokok merupakan kebiasaan yang dapat membahayakan kesehatan dan menimbulkan berbagai penyakit yang dapat menimbulkan morbiditas dan mortalitas, yang mana salah satu parameter mortalitas adalah angka kematian kasar (CDR) [12]. Bahkan Badan Kesehatan Dunia (WHO) menyatakan bahwa sekitar 225.700 perokok meninggal di Indonesia setiap tahun akibat asap tembakau atau penyakit terkait

tembakau lainnya. Tingkat rata-rata perokok di Indonesia tetap menjadi salah satu yang tertinggi di dunia. Indonesia merupakan satu-satunya negara di kawasan Asia Tenggara yang belum meratifikasi FCTC WHO (World Health Organization Framework Convention on Tobacco Control). Namun, penggunaan tembakau terus menjadi penyebab signifikan kematian dan penyakit seumur hidup yang masih bisa dicegah [13].

Ada berbagai metode pengumpulan data melalui Twitter, dan masing-masing memiliki kelebihan dan kekurangan. Dalam penelitian ini, penulis mencoba melakukan analisis komparatif tentang kemampuan metode Twitter untuk pengumpulan data. Data dari hasil laporan tersebut kemudian dapat digunakan untuk komunitas atau lembaga yang terkait dengannya sebagai alat penilaian. Ini akan mengenali pendapatan atau pola yang akan digunakan untuk proses pengujian SVM, yang bertujuan untuk memberi label sentimen pada tweet, dengan melakukan pelatihan tentang klasifikasi SVM. Untuk mendapatkan informasi sentimen yang terkandung dalam kalimat opini, analisis sentimen merupakan teknik pemahaman, penggalian informasi subjektif dari sumber, dan secara otomatis mengolah data kontekstual. Asumsi kuat digunakan dalam algoritme pembelajaran Naive Bayes, yang akan didasarkan pada teori Bayesian.

Teori Bayes adalah teori yang menggunakan data yang dikumpulkan sebelumnya sehingga memutuskan apa yang memiliki probabilitas tertinggi [14]. Pohon Keputusan, sebaliknya, adalah algoritma prediksi yang menggunakan data yang dikumpulkan sebelumnya [15]. Analisis sentimen adalah proses penambangan data yang digunakan untuk menemukan dan mengekstraksi data dari teks dengan tujuan memahami sentimen sosial. Untuk lebih memahami tentang sikap, pendapat, dan pendekatan dalam teks informasi, analisis sentimen juga dilakukan [16]. Python adalah bahasa pemrograman yang digunakan oleh penulis penelitian ini untuk menganalisis dan menerapkan metode yang akan digunakan. Python lebih seperti bahasa manusia daripada bahasa mesin karena merupakan bahasa pemrograman tingkat tinggi. Python dirancang untuk memberikan kenyamanan programmer dalam hal efisiensi waktu, kemudahan pengembangan, dan kompatibilitas sistem

[17].

Dalam penelitian ini penulis akan melakukan analisis opini masyarakat pengguna media sosial twitter terhadap perokok di Indonesia. Penulis juga akan melakukan perbandingan tiga metode dalam penerapan di analisis sentimen pengguna twitter yaitu metode Naïve Bayes, Support Vector Machine, dan Decision Tree. Alasan penulis melakukan penelitian ini adalah untuk menemukan sentimen masyarakat terhadap perokok di Indonesia dan juga menemukan algoritma mana yang terbaik dalam melakukan pengolahan data analisis sentimen ini.

1.2 Rumusan Masalah

Berdasar latar belakang masalah yang telah disampaikan seperti di 1.1, pokok permasalahan yang dirumuskan adalah:

1. Bagaimana opini masyarakat pengguna media sosial twitter mengenai perokok?
2. Berapa tingkat akurasi yang dihasilkan dari metode naïve bayes, support vector machine, dan decision tree?

1.3 Batasan Masalah

Agar penyusunan penelitian lebih terperinci dan terarah sesuai dengan permasalahan yang telah dirumuskan, maka diperlukan batasan masalah antara lain:

1. Penelitian akan menekankan pada pengujian metode.
2. Data yang digunakan hanya dari tweet mengandung kata kunci "perokok" yang dituliskan pengguna twitter.
3. Data tweet yang digunakan hanya tweet berbahasa Indonesia.
4. Data tweet diambil dengan 2 rentang waktu 25 Oktober 2022 sampai 3 November 2022 dengan perolehan data bersih 2.772 tweet dan 1 Oktober 2022 sampai 31 Januari 2023 dengan perolehan data bersih 6.300 tweet.
5. Hasil klasifikasi tweet hanya berupa sentimen negatif dan sentimen positif.

1.4 Tujuan Penelitian

Adapun tujuan yang akan dicapai dalam penelitian ini adalah mengklasifikasikan nilai sentimen tweet pengguna twitter terhadap perokok

kemudian melakukan pengujian terhadap tiga metode sehingga didapatkan opini terbesar masyarakat mengenai perokok dan diketahui metode dengan performa terbaik dalam melakukan pengujian sentimen.

1.5 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah:

1. Dapat mengetahui opini pengguna twitter terhadap perokok di Indonesia.
2. Dapat mengetahui performa terbaik dari metode yang digunakan.
3. Dapat menambah wawasan dan informasi kepada pihak yang berkepentingan.
4. Diharapkan dengan adanya penelitian ini dapat dijadikan sebagai referensi untuk penelitian selanjutnya.

1.6 Sistematika Penulisan

Pada tahapan ini digunakan sebagai sarana untuk memudahkan serta mengetahui pembahasan yang ditulis oleh penulis pada skripsi ini secara menyeluruh sesuai dengan ketentuan yang berlaku di Universitas Amikom Yogyakarta. Adapun sistematika penulisan yang digunakan adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisikan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini membahas mengenai penelitian sebelumnya yang telah dilakukan oleh peneliti lain yang dijadikan sebagai referensi dalam penelitian ini, bab ini juga berisi dasar – dasar teori yang digunakan dalam penelitian.

BAB III METODE PENELITIAN

Bab ini terdapat uraian mengenai objek penelitian, alur penelitian yang dilakukan, alat dan bahan yang digunakan dalam penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini merupakan tahapan yang penulis lakukan dalam menerapkan alur metode yang telah dirancang untuk menghasilkan tujuan penelitian.

BAB V PENUTUP

Bab ini berisi mengenai kesimpulan yang diperoleh dari hasil penelitian yang telah dilakukan dan saran yang dapat peneliti sampaikan untuk penelitian selanjutnya.

