

BAB I PENDAHULUAN

1.1 Latar Belakang

Pada era digital dan perkembangan teknologi yang sangat pesat ini media *social* bukanlah hal baru bagi masyarakat, hampir seluruh kalangan usia menggunakannya. Media *social* merupakan platform atau media digital yang digunakan untuk melakukan aktivitas *social* secara *online* seperti bertukar informasi, mengirimkan pesan, dan berbagi media. Menurut *website* datereportal, jumlah pengguna media *social* aktif di dunia pada Januari 2023 adalah 4,80 miliar, sekitar 59,9% dari total penduduk dunia [1].

Di Indonesia sendiri, jumlah pengguna media *social* aktif pada Januari 2023 adalah 167 juta, sekitar 60,4% dari total penduduk Indonesia. Salah satu media *social* yang digunakan masyarakat Indonesia adalah twitter, jumlah penggunanya sebanyak 24 juta [2]. Twitter adalah layanan untuk teman, keluarga, dan rekan kerja untuk berkomunikasi dan tetap terhubung melalui pertukaran pesan yang cepat dan sering [3]. Melalui twitter pengguna dapat mengetahui kabar, *thread*, atau topik yang sedang viral dan melakukan diskusi dengan pengguna lain terkait kabar tersebut.

Pada 25 September 2022 media *social* twitter diramaikan dengan kasus pengguna twitter dengan *username* @gandhoyy yang mendapatkan surat somasi dari PT Esteh Indonesia Makmur. PT Esteh Indonesia Makmur merupakan perusahaan *food & beverages* yang didirikan oleh Haidhar Hibatullah Wurjanto pada tahun 2018. Perusahaan ini menjual minuman berbahan dasar teh yang dikombinasikan dengan susu, coklat, taro, red velvet, dan varian lainnya [4].

Gandhoyy diberi somasi oleh PT Esteh Indonesia Makmur karena tweetnya yang mengkritik salah satu minuman dari perusahaan tersebut. Dalam tweetnya, Gandhoyy mengkritik bahwa minuman chizu red velvet memiliki rasa yang terlalu manis, tidak lupa ia menambahkan kata-kata kasar pada tweetnya. *Tweet* tersebut mengundang banyak reaksi dari warganet, baik masyarakat biasa, *public figure*,

tenaga kesehatan, maupun kompetitor. Namun karena banyaknya reaksi dari warganet, sulit untuk mendeteksi kecenderungan sentimen dari kasus tersebut. Oleh karena itu, dibutuhkan analisis sentimen untuk mengetahui sentimen dari pengguna twitter. Analisis sentimen merupakan salah satu bidang penelitian yang terdapat dalam NLP, komputasi linguistik, dan *text mining*. Analisis sentimen bertugas untuk mengelompokkan polaritas yang terdapat pada suatu teks, baik dokumen, kalimat, maupun fitur tingkat aspek apakah pendapat yang dikemukakan bersifat positif, negatif, atau netral [5].

Terdapat beberapa algoritma klasifikasi yang dapat digunakan untuk analisis sentimen, seperti *Naïve Bayes*, *Support Vector Machine*, *Decision Tree*, *Random Forest*, *Logistic Regression*, dll. Pada tahun 2019, V. A. Fitri *et.al* [6] melakukan penelitian tentang analisis sentimen terhadap kasus kampanye LGBT menggunakan algoritma *Naïve Bayes*, *Decision Tree*, dan *Random Forest*. Dari penelitian tersebut algoritma *Naïve Bayes* memiliki akurasi paling tinggi, yaitu 86,43%. Kemudian, pada tahun 2022, F. Albasithu dan A. Wibowo [7] membandingkan algoritma *Naïve Bayes* dan C4.5 untuk menganalisis sentimen presiden 3 periode. Hasilnya, algoritma *Naïve Bayes* mendapatkan akurasi yang lebih besar dibandingkan C4.5, dengan akurasi 85%.

Penelitian lain yang membandingkan algoritma SVM dengan algoritma klasifikasi juga dilakukan oleh K. A. Rokhman *et.al* [8] pada tahun 2021. Penelitian tersebut membandingkan algoritma SVM dengan *Decision Tree* untuk analisis sentimen *review* komentar pada aplikasi transportasi online. Algoritma SVM memiliki akurasi sebesar 90,20%, lebih tinggi daripada *Decision Tree* yang akurasinya hanya 89,90%. Masih di tahun yang sama, N. I. Wibowo *et.al* [9] melakukan perbandingan algoritma *Random Forest*, SVM, dan *Logistic Regression* untuk klasifikasi sentimen terhadap insiden kebocoran data Tokopedia. Pada penelitian tersebut *f1-score* digunakan untuk membandingkan performa dari ketiga algoritma, hasilnya SVM menghasilkan nilai *f1-score* paling tinggi, yaitu 0,503583.

Berdasarkan hasil penelitian terdahulu, algoritma *Naïve Bayes* dan *Support Vector Machine* sudah sering digunakan untuk analisis sentimen dan terbukti

memiliki performa yang lebih baik. Oleh karena itu, pada penelitian ini peneliti tertarik untuk membandingkan performa dan *confusion matrix* dari kedua algoritma tersebut.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan, permasalahan yang dapat dirumuskan dalam penelitian ini yaitu, bagaimana perbandingan performa algoritma *Naïve Bayes* dan *Support Vector Machine* dalam menganalisa sentimen pengguna twitter terhadap PT Esteh Indonesia Makmur.

1.3 Batasan Masalah

Berdasarkan rumusan masalah diatas, terdapat beberapa pembatasan masalah dalam penelitian ini, yaitu:

1. Permasalahan yang dibahas dalam penelitian ini adalah kasus somasi yang dilayangkan PT Esteh Indonesia Makmur kepada salah satu pengguna twitter dengan *username @gandhoyy*.
2. Bahasa pemrograman yang digunakan adalah python.
3. Algoritma *Naïve Bayes* yang digunakan adalah *Multinomial Naïve Bayes*.
4. Proses *scraping* data menggunakan *snsrape*.
5. *Tweet* diklasifikasikan menjadi dua kelas, yaitu positif dan negatif.
6. Data yang digunakan merupakan *tweet* Bahasa Indonesia dalam rentang waktu 24-30 September 2022.
7. *Keyword* yang digunakan untuk proses *scraping* adalah somasi dan es teh Indonesia.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk mengetahui algoritma yang memiliki performa paling baik diantara *Naïve Bayes* dan *Support Vector Machine* dalam melakukan analisis sentimen pengguna twitter terhadap PT Esteh Indonesia Makmur.

1.5 Manfaat Penelitian

1.5.1 Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan pengetahuan terkait performa algoritma *Naïve Bayes* dan *Support Vector Machine* dalam melakukan analisis sentimen.

1.5.2 Manfaat Praktis

1. Bagi Peneliti

Penelitian ini diharapkan dapat menambah pengetahuan dan pengalaman peneliti mengenai cara melakukan analisis sentimen dengan algoritma *Naïve Bayes* dan *Support Vector Machine* menggunakan bahasa pemrograman python.

2. Bagi Akademisi

Penelitian ini diharapkan dapat menjadi bahan referensi bagi penelitian selanjutnya, terutama penelitian sejenis yang membandingkan dua algoritma atau lebih.

3. Bagi Pembaca

Penelitian ini diharapkan dapat memberikan gambaran kepada pembaca terkait sentimen warganet terhadap PT Esteh Indonesia Makmur.

1.6 Sistematika Penulisan

Sistematika penulisan ditujukan kepada pembaca agar lebih mudah mengetahui garis besar isi skripsi untuk tiap-tiap bab. Berikut sistematika penulisan skripsi ini:

BAB I PENDAHULUAN

Bab ini terdiri dari 6 sub-bab, yaitu latar belakang yang berisi masalah dan penyebab penelitian ini dilakukan, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini terdiri dari 2 sub-bab, yaitu studi literatur yang berisi enam

penelitian terdahulu dan dasar teori yang berisi teori tentang analisis sentimen, algoritma *Naïve Bayes*, *Support Vector Machine*, dll.

BAB III METODE PENELITIAN

Bab ini menjelaskan tentang sistematika metode yang akan digunakan dalam proses perancangan, seperti objek penelitian, alur penelitian, dan alat dan bahan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi penyelesaian masalah dan penjelasan tentang program yang sudah dibuat berdasarkan topik penelitian.

BAB V PENUTUP

Bab ini berisi kesimpulan dan saran tentang penelitian yang telah dilakukan peneliti.

