

BAB I

PENDAHULUAN

Pada bab ini akan dibahas mengenai pengenalan tentang penelitian yang dilakukan. Bab ini berisikan latar belakang, rumusan masalah, tujuan, manfaat dan batasan masalah dari penelitian ini. Latar belakang menjelaskan mengapa dan bagaimana penelitian akan dilakukan. Rumusan masalah memberikan pertanyaan yang akan dijawab dari penelitian ini. Tujuan penelitian menyatakan apa yang diharapkan dapat dicapai dalam pada penelitian ini. Manfaat penelitian berisikan kontribusi yang diharapkan bagi beberapa pihak. Batasan masalah memberikan batasan mengenai fokus penelitian yang dilakukan.

1.1. Latar Belakang

Teknologi di bidang informasi berkembang dengan pesat dan telah digunakan di berbagai bidang kehidupan, tak terkecuali di bidang kesehatan. Berbagai teknologi informasi kesehatan dikembangkan untuk mempermudah akses data penanganan kesehatan. Beberapa aplikasi juga diciptakan untuk menyediakan informasi dan layanan untuk pengguna mengenai cara hidup yang sehat (Ben-Zeev, et al., 2015). Salah satunya merupakan aplikasi SATUSEHAT yang dikeluarkan oleh kementerian kesehatan RI, aplikasi SATUSEHAT. Aplikasi SATUSEHAT ini merupakan aplikasi yang diperbaharui dari aplikasi sebelumnya, Peduli Lindungi yang sebelumnya digunakan untuk melacak kontak dan vaksinasi *Covid-19* telah diubah menjadi aplikasi kesehatan masyarakat secara umum (Hulisnaini & Juliangrace, 2023). Beberapa penambahan fitur diberikan untuk membuat aplikasi ini dapat bermanfaat di luar masa pandemi *Covid-19* dan dapat digunakan dalam jangka Panjang.

Masyarakat Indonesia memberikan respon yang beragam dalam pengalihfungsian aplikasi Peduli Lindungi menjadi SATUSEHAT. Oleh karena itu, dilakukan analisis sentimen untuk memetakan respon masyarakat terhadap aplikasi ini untuk perkembangan aplikasi kedepannya. Dengan melakukan analisis sentimen terhadap aplikasi SATUSEHAT,

dapat mempercepat proses klasifikasi ulasan dari para penggunanya ke depan sehingga dapat berfokus pada pengembangan berkelanjutan sebagai respon dari sentimen di masa depan. Analisis sentimen adalah teknik pemrosesan teks yang digunakan untuk mengumpulkan data dari berbagai sumber, termasuk media sosial, situs *web e-commerce*, dan laman unduhan aplikasi, dalam bentuk opini, baik positif maupun negatif. (Wahyudi & Kusumawardhana, 2021). Analisis Sentimen dapat dilakukan dengan mengumpulkan terlebih dahulu ulasan yang diberikan oleh para pengguna aplikasi SATUSEHAT. Ulasan-ulasan ini dapat ditemukan di aplikasi *Google Play Store*. Rating yang diberikan di ulasan *Google Play Store* terkadang tidak sesuai dengan sentimen yang diberikan oleh pengguna dalam bentuk komentar. Oleh karena itu diperlukan cara lain mengolah ulasan pada *Google Play Store*.

Dalam mengolah ulasan-ulasan aplikasi SATUSEHAT di *Google Play Store*, diperlukan proses *text mining*. Dari pencarian yang dilakukan, *text mining* bisa disebut sebagai sebuah proses saat pengguna melakukan prosedur untuk mengekstraksi informasi berharga dari sumber data dengan menemukan dan memeriksa pola-pola tertentu dalam kumpulan dokumen menggunakan alat analisis. (Ashari, Arifianto, & Faruq). *Text mining* yang dilakukan dalam penelitian ini bertujuan untuk mengklasifikasikan teks untuk kebutuhan analisis sentimen. Terdapat banyak pilihan algoritma maupun metode yang dapat digunakan untuk melakukan kegiatan pengklasifikasian teks ini. Dalam melakukan analisis sentimen, pemilihan algoritma merupakan kunci untuk mendapatkan akurasi tertinggi. Beberapa algoritma klasifikasi akan digunakan dalam melakukan penelitian kali ini. Algoritma klasifikasi pertama yang akan digunakan yaitu *Multinomial Naïve Bayes*. *Multinomial Naïve Bayes* merupakan model pengembangan dari algoritma *Naïve Bayes* yang dalam penggunaannya cocok untuk melakukan pengklasifikasian pada data teks ataupun dokumen. Dalam formula yang digunakan *Multinomial Naïve Bayes*, kelas pada sebuah dokumen tidak hanya ditentukan dengan kata yang muncul tetapi juga jumlah kemunculannya (Witten, Frank, & Hall, 2011). Algoritma ini dipilih

karena memberikan kinerja yang dalam penggunaannya bekerja cukup baik pada data teks dan klasifikasi sentimen. Algoritma ini efektif dalam menangani representasi teks yang dihasilkan oleh ekstraksi fitur *TF-IDF*, dan dapat menangkap hubungan probabilistik antara kata-kata dan sentimen. Lalu Metode *Random Forest*, algoritma kedua yang digunakan dalam penelitian ini, terbentuk dari penggabungan *Decision Tree* dalam melakukan klasifikasi. *Random Forest* merupakan algoritma *ensemble* yang mana dalam proses mendapatkan keputusan akhir dilakukan *voting majority* dari seluruh model *Decision Tree* yang terbentuk (Rohman, Purwanto, & Santoso, 2018). Algoritma ini dipilih karena dikenal sebagai model yang kuat dan fleksibel. Ia dapat menangani berbagai jenis data dan kompleksitas relasi antar fitur. Dalam konteks analisis sentimen, *Random Forest* dapat menangani pola yang kompleks dan variabel yang tidak terduga. Setelah itu, *Logistic Regression*, algoritma ketiga, merupakan algoritma yang dapat digunakan untuk menggambarkan hubungan antara satu atau lebih variabel respons dengan satu atau lebih faktor prediktor. Variabel respon dari *Logistic Regression* hanya bernilai biner, yaitu antara 0 dan 1. Maka algoritma ini akan menghasilkan kelas sentimen positif dan negatif (Bimantara & Dina, 2018). Algoritma ini dipilih karena memberikan interpretasi yang relatif mudah terhadap koefisien dan dampaknya terhadap *output*. Ini memudahkan pemahaman faktor-faktor apa yang mempengaruhi keputusan klasifikasi. Algoritma klasifikasi *Multinomial Naïve Bayes*, *Random Forest*, dan *Logistic Regression* merupakan metode yang memiliki akurasi cukup tinggi dalam berbagai penelitian klasifikasi teks. Oleh karena itu, algoritma klasifikasi *Multinomial Naïve Bayes*, *Random Forest*, dan *Logistic Regression* akan digunakan pada penelitian kali ini.

Dalam mengoptimasi algoritma klasifikasi saat melakukan analisis sentimen, terdapat banyak cara yang dapat dilakukan. Dalam penelitian kali ini, akan dicoba menggabungkan hasil klasifikasi dari masing-masing algoritma menggunakan metode *Majority Vote*. Setelah itu, akurasi hasil yang didapat akan dibandingkan dengan hasil dari masing-masing

algoritma secara individu untuk melihat apakah metode ini dapat mengoptimasi hasil dari masing-masing algoritma. Metode *Majority Vote* yaitu dengan memberikan nilai pada setiap fitur yang terdapat di sebuah *dataset* (Saputra, Sujatmika, & Arifin, 2011). Hasil dari masing-masing algoritma yang digunakan akan disimpan, lalu akan dibandingkan masing-masing hasil penyeleksian menggunakan metode *Majority Vote*. Dari penelitian yang dilakukan oleh (Abbas, Salih, Hussein, Hussein, & Abdulwahhab, 2020), didapatkan hasil bahwa metode *ensemble majority vote* menghasilkan tingkat akurasi yang lebih tinggi dari beberapa algoritma klasifikasi individu yang digunakan. Dalam penelitian ini, akan dilakukan penelitian guna menghasilkan hasil analisis sentimen dengan akurasi tinggi untuk aplikasi SATUSEHAT. Algoritma-algoritma digunakan yang digunakan memiliki pendekatan yang berbeda sehingga diharapkan dapat memberikan diversitas yang dapat meningkatkan kinerja *ensemble*. Dengan menggunakan ensemble dari beberapa model, terdapat potensi untuk mengurangi risiko *overfitting* pada data pelatihan. 3 algoritma juga dipilih untuk menghindari hasil voting netral sehingga tidak memberikan kecenderungan untuk memilih satu kelas sentimen di luar hasil voting metode *Majority Vote*.

1.2. Rumusan Masalah

Dari latar belakang yang telah dijelaskan, dilakukan perumusan masalah yang dapat dilihat sebagai berikut.

1. Apakah metode penggabungan *Majority Vote* dapat meningkatkan nilai akurasi pada algoritma klasifikasi *Multinomial Naïve Bayes*, *Random Forest*, dan *Logistic Regression* dalam kasus analisis sentimen?
2. Bagaimana pengaruh metode *Majority Vote* pada akurasi yang dihasilkan pada algoritma klasifikasi *Multinomial Naïve Bayes*, *Random Forest*, dan *Logistic Regression* untuk kasus analisis sentimen?
3. Bagaimana akurasi hasil analisis sentimen aplikasi SATUSEHAT yang didapat dari ulasan pada aplikasi *Google Play Store*?

1.3. Tujuan

Tujuan yang ingin dicapai di dalam penelitian ini untuk menjawab rumusan masalah yang ada adalah sebagai berikut.

1. Mengimplementasikan metode penggabungan *Majority Vote* untuk meningkatkan nilai akurasi pada algoritma klasifikasi *Multinomial Naïve Bayes*, *Random Forest*, dan *Logistic Regression* dalam kasus analisis sentimen.
2. Mengetahui pengaruh metode *Majority Vote* pada akurasi algoritma klasifikasi *Multinomial Naïve Bayes* dan *Random Forest*, dan *Logistic Regression* untuk analisis sentimen.
3. Menghasilkan hasil analisis sentimen dengan akurasi tinggi aplikasi SATUSEHAT dari ulasan aplikasi *Google Play Store*.

1.4. Manfaat Penelitian

Manfaat yang diharapkan dapat diberikan dari melakukan penelitian ini terhadap beberapa pihak sebagai berikut.

1. Bagi peneliti, manfaat yang didapatkan adalah tambahan ilmu baru mengenai algoritma klasifikasi dan metode pengotimasiannya.
2. Bagi Kementerian Kesehatan RI, manfaat yang didapatkan oleh Kementerian Kesehatan adalah dapat menghasilkan model algoritma yang dapat dikembangkan dan digunakan untuk mengetahui penilaian masyarakat dan dapat mengembangkan aplikasi SATUSEHAT.

1.5. Batasan Masalah

Batasan masalah yang digunakan di dalam penelitian ini sebagai berikut.

1. Analisis Sentimen dilakukan dengan ulasan aplikasi SATUSEHAT sebanyak 1592 data yang terdapat pada aplikasi *Google Play Store*.
2. Hasil penelitian adalah tingkat akurasi pengujian algoritma klasifikasi *Multinomial Naïve Bayes*, *Random Forest*, *Logistic Regression* dan hasil voting ketiganya menggunakan metode *Majority Vote*.
3. Ekstraksi Fitur dilakukan dengan metode *TF-IDF*.

Halaman ini sengaja dikosongkan