

BAB I

PENDAHULUAN

1.1 Latar Belakang

Berdasarkan riset yang dilakukan oleh Hootsuite pada 2020 mengenai Hootsuite (We are Social): Indonesian Digital Report 2020 [1] jumlah pengguna internet di Indonesia mencapai 175,4 juta pada Januari 2020. Jumlah pengguna internet yang besar ini berimbas pada banyaknya informasi yang tersebar di internet. Informasi yang paling sering ditemukan berupa angka, teks, gambar, dan suara. Saat ini penyebaran informasi tidak terlepas dari perang dari sosial media. Beberapa sosial media menjadi lebih populer dengan kemajuan teknologi seperti Instagram, *Twitter*, Tiktok dan Facebook. Jumlah pengguna sosial media di Indonesia saat ini mencapai 160 juta pengguna. Selain sosial media aktivitas lain yang dilakukan oleh pengguna internet di Indonesia adalah bermain gim, berselancar dan melakukan transaksi *online*. Beberapa *e-commerce* yang cukup terkenal saat ini ada Tokopedia, Shopee, Lazada, Blibli dan Bhinneka.

Salah satu website dengan yang paling banyak dikunjungi adalah *twitter*, *Twitter* adalah sebuah layanan jejaring sosial dan mikroblog. *Twitter* memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks dengan 280 karakter. Berdasarkan data yang dilansir oleh Hootsuite pengguna *twitter* di Indonesia mencapai 10 juta pengguna aktif. *Twitter* menjadi salah satu jejaring sosial media untuk berbagi opini dalam bentuk cuitan.

Kemudahan bertransaksi di era saat ini juga mempermudah pembeli dalam mendapatkan sebuah barang, calon pembeli cukup menuju ke *marketplace* atau *e-commerce* yang disukai. Dilansir dari website iPrice insights salah satu *e-commerce* yang paling populer di Indonesia adalah Tokopedia. Tokopedia memiliki jumlah pengunjung bulanan mencapai 84 juta dalam kuartal ketiga 2020. Tokopedia melakukan promosi dengan memberikan banyak promo, diskon serta *cashback* dalam setiap transaksi

Setiap interaksi yang terdapat di internet dapat dimanfaatkan menjadi informasi berharga. Respon atau tanggapan yang ada dapat berupa sebuah percakapan, komentar, atau balasan teks. Setiap respon berupa teks ini dapat dimanfaatkan menjadi sebuah analisis khusus yang disebut *text mining*. *Text mining* memiliki definisi menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antara dokumen.

Analisis Sentimen merupakan sebuah perpaduan dari *text mining* dan *data mining*, atau sebuah metode yang digunakan untuk memahami, mengekstrak dan mengolah teks yang diberikan konsumen atau publik melalui banyak media mengenai sebuah produk, jasa atau sebuah instansi.

Dengan berkembangnya sosial media dan e-commerce yang ada penulis ingin memanfaatkan informasi yang ada dengan melakukan analisis sentimen untuk mengetahui bagaimana tingkat kepuasan publik terkait Tokopedia di Indonesia. Analisis sentimen yang digunakan pada penelitian kali menggunakan data *twitter* mengenai Tokopedia.

Berdasarkan uraian diatas maka penulis bertujuan mengimplementasikan algoritma *support vector machine* untuk mengetahui bagaimana analisis sentimen terhadap Tokopedia menggunakan data *twitter*.

Oleh karena itu, berdasarkan uraian diatas judul yang diambil dalam penelitian ini adalah "Implementasi Algoritma Support Vector Machine untuk Analisis Sentimen pada Data Twitter Tokopedia".

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan di atas, maka dalam penelitian ini permasalahan yang akan dirumuskan yaitu sebagai berikut :

1. Bagaimana sentimen yang diberikan pengguna *twitter* terhadap Tokopedia menggunakan Metode *Support Vector Machine*

2. Mengetahui nilai akurasi dari algoritma *Support Vector Machine* pada klasifikasi sentimen pada Tokopedia

1.3 Batasan Masalah

1. Pengambilan data *twitter* yang digunakan dalam penelitian ini menggunakan *Twitter Streaming API*.
2. Sumber data yang digunakan dalam penelitian ini adalah kalimat opini yang akan dianalisis melalui *tweet* berkaitan dengan sentimen pengguna *twitter* terhadap Tokopedia.
3. Data sentimen mengenai Tokopedia Indonesia diperoleh dari media sosial *twitter* dengan kata kunci "Tokopedia".
4. Jumlah data yang dipakai dalam penelitian ini berjumlah 5000 *tweet*, Dimana 80% digunakan sebagai data latih dan 20% digunakan sebagai data uji, yang diambil dari tanggal 1 Januari–31 Desember 2020.
5. Data *twitter* yang digunakan merupakan data *tweet* berbahasa Indonesia.
6. Data sentimen diklasifikasikan menjadi 2 kelas yaitu positif dan negatif.

1.4 Maksud dan Tujuan Penelitian

1. Mengetahui performa dari algoritma *Support Vector Machine* yang diterapkan pada proses analisis sentimen berbahasa Indonesia yang memiliki dua kelas klasifikasi yaitu positif dan negatif.
2. Penelitian ini diharapkan dapat memberikan gambaran mengenai bagaimana sentimen pengguna *twitter* terhadap Tokopedia.

1.5 Metode Penelitian

1.5.1 Metode Pengumpulan Data

Dataset yang digunakan merupakan teks berbahasa Indonesia yang diambil dari API Twitter dengan kata kunci "Tokopedia". Proses pengumpulan data/*scrapping* menggunakan *tool* khusus bernama *twint*. Jumlah data yang diambil sebanyak 5000 data.

Setelah data *twitter* terkumpul dilakukan pelabelan data apakah data tersebut tergolong sebagai cuitan bersentimen positif atau bersentimen negatif. Data yang ada kemudian dibagi menjadi dua bagian yaitu data latih dan data uji. Perbandingan dari kedua data ini adalah 80:20. Pada data latih berjumlah 4000 data dimana 2000 data diberi label positif dan 2000 data diberi label negatif. Sisanya 1000 data digunakan sebagai data uji.

1.5.2 Metode Analisis dan Perancangan

1. Studi Pustaka

Melakukan mengumpulkan literatur yang dibutuhkan seperti jurnal ilmiah, tesis dan buku untuk mendukung penelitian ini terkait algoritma *Support Vector Machine*.

2. Eksperimen

Melakukan pengujian terhadap kode program yang dibuat agar mudah untuk dikembangkan. Membuat tampilan yang mudah dipahami dan informatif.

3. Analisis dan Perancangan

Melakukan analisa terhadap tahapan-tahapan pembuatan model klasifikasi yang akan dilakukan. Hasil tersebut dibuat perancangan, tahapan yang akan dirancang adalah

- **Pengumpulan Data:** Pengumpulan data yang dilakukan pada penelitian ini berupa dataset tweet dari pengguna *twitter* berbahasa Indonesia mengenai sentimen terhadap Tokopedia. Seluruh data tweet tersebut akan disimpan dalam sebuah dataset dan diberikan label secara manual untuk memberikan atribut apakah suatu data tweet dianggap positif atau negatif. Jumlah dataset yang akan di *crawling* dari *twitter* yang berkenaan dengan subjek penelitian adalah sebanyak 5.000 data tweet. Sejumlah 80 persen dari data tersebut akan digunakan sebagai data latih dan 20 persen akan digunakan sebagai data uji untuk mengetahui seberapa besar akurasi dari proses training dalam memprediksi suatu data.

- *Preprocessing*: Tahapan *preprocessing* adalah dimana dilakukannya penyeleksian data yang akan diproses. Setiap kata akan dipecah menjadi beberapa bagian yang lebih kecil sehingga mempunyai arti yang lebih sempit. Ada beberapa tahapan dalam proses *preprocessing* yaitu *tokenization*, *lower casing*, *cleaning* dan *stemming*.
- Pembobotan Data: Pembobotan merupakan metode untuk mengubah input data menjadi suatu fitur vektor. Metode pembobotan yang umum digunakan adalah TF-IDF. Sebagai contoh terdapat sederetan fitur seperti pada vektor $\{f_1, f_2, \dots, f_n\}$ dimana yaitu sekumpulan fitur-fitur sebanyak n yang sudah ditentukan sebelumnya. Misalkan kata "puas" maka fitur vektor dari data adalah vektor.
- Pembuatan Model Klasifikasi: Tahapan pembuatan klasifikasi ini menggunakan Support Vector Machine untuk mendapatkan sebuah nilai atau pola yang disimpan menjadi model yang bertujuan untuk memprediksi pemberian label pada tahap pengujian.

1.5.3 Metode Testing

Melakukan perhitungan akurasi dari model klasifikasi yang dihasilkan menggunakan *Confusion Matrix*.

1.5.4 Metode Implementasi

Membuat sebuah tampilan/*interface* sederhana dari fitur yang akan dikembangkan serta menunjukan bagaimana langkah algoritma *Support Vector Machine* ini berjalan dengan menggunakan *backend flask* dan *frontend React.js*.

1.6 Sistematika Penulisan

Sistematika penulisan terbagi menjadi beberapa bab pokok pembahasan yang dijabarkan sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini menjelaskan mengenai latar belakang masalah, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Pada bab ini menjelaskan mengenai referensi – referensi dari analisis sentimen yang telah ada sebelumnya untuk dijadikan acuan dalam melakukan penelitian, menjelaskan mengenai perbandingan dari beberapa analisis yang telah ada dengan analisis yang penulis buat. Serta memberikan serta menyampaikan teori-teori yang dapat mendukung penelitian ini.

BAB III METODE PENELITIAN

Pada bab ini menjelaskan mengenai Bab ini memaparkan populasi dan sampel penelitian, jenis dan sumber data, variabel penelitian, metode pengumpulan data, metode analisis data, dan tahapan penelitian.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan mengenai analisis dan perancangan yang dilakukan terhadap hasil pengumpulan data dan penguraian hasil dari pengolahan dan analisa dimana hasil penelitian ini untuk mengetahui hasil keakuratan dari hasil pengklasifikasian dengan metode *support vector machine* dan hasil dari asosiasi berdasarkan kelas pengklasifikasian.

BAB V PENUTUP

Pada bab ini menjelaskan mengenai kesimpulan dan saran. Kesimpulan dapat mengemukakan kembali masalah penelitian, menjawab pertanyaan di rumusan masalah dan mampu membuktikan capaian tujuan penelitian, menyimpulkan bukti-bukti yang diperoleh dan akhirnya menarik kesimpulan apakah hasil yang didapat (dikerjakan), layak untuk digunakan (diimplementasikan)

DAFTAR PUSTAKA
LAMPIRAN

