

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dengan semakin berkembangnya teknologi dan popularitas media sosial, banyak orang sekarang berbagi ulasan dan pendapat tentang produk atau layanan yang digunakan, termasuk tentang jamu madura. Jamu madura adalah salah satu jenis obat tradisional yang berasal dari Pulau Madura, Indonesia. Karena kepopulerannya sebagai produk kesehatan alami, jamu madura seringkali menjadi objek dari ulasan dan komentar oleh para konsumen. Untuk mengetahui respon dan pandangan konsumen dibutuhkan sistem dengan metode yang dapat menganalisis sentimen terhadap ulasan tersebut. Analisis sentimen adalah bidang studi yang menganalisis pendapat, perasaan, evaluasi, penilaian, sikap, dan perasaan seseorang terhadap entitas tertentu seperti layanan, organisasi, individu, masalah, peristiwa, topik dan atribut lainnya[1]. Analisis sentimen dilakukan sebagai acuan dalam meningkatkan suatu pelayanan, ataupun meningkatkan kualitas produk[2]. Analisis sentimen menggunakan algoritma *machine learning* untuk proses klasifikasinya. Algoritma *machine learning* yang sering kali digunakan antara lain SVM.

SVM merupakan metode klasifikasi yang memiliki performansi lebih baik dibanding metode yang lainnya[3]. SVM mempunyai beberapa kernel yang umum digunakan diantaranya kernel Radial Basis Function (RBF) juga dapat disebut dengan kernel Gaussian. Kernel RBF memiliki dua parameter yaitu Gamma yang digunakan untuk menentukan jauh atau tidaknya pengaruh dari satu sampel *training* dataset pada garis pemisahannya. Jika bernilai rendah maka menunjukkan bahwa pengaruhnya jauh, sedangkan jika bernilai tinggi maka menunjukkan bahwa pengaruhnya dekat. Nilai dari gamma ini kemudian dipertimbangkan dalam perhitungan. Parameter yang kedua yaitu C atau yang disebut dengan Cost. Cost merupakan parameter yang berguna untuk mengoptimalkan SVM agar terhidar dari misklasifikasi pada setiap sampel dalam *training* dataset[4]. Akan tetapi, *Support Vector Machine* memiliki kekurangan dalam pemilihan fitur yang sesuai. Pemilihan fitur dapat mempengaruhi hasil akurasi karena tidak semua fitur sama pentingnya. Seleksi fitur *Chi-Square* dapat menghilangkan fitur tanpa

mengurangi tingkat akurasi bertujuan melihat ketergantungan fitur terhadap label kelas untuk mendapatkan fitur yang optimal dan relevan[5]. Berdasarkan hal tersebut dalam penelitian ini peneliti melakukan analisis sentimen dengan algoritma *Support Vector Machine* (SVM) sebagai metode klasifikasi untuk analisis sentimen dan *Chi-Square* sebagai metode seleksi fitur untuk meningkatkan performa analisis sentimen. Dan data ulasan jamu madura yang digunakan dalam penelitian ini diambil dari Tokopedia sebanyak 556 data ulasan. Berikut penelitian - penelitian sebelumnya mengenai SVM dan *Chi-Square* yang digunakan penulis sebagai referensi antara lain: “Perbandingan Naïve Bayes, SVM, dan K-NN untuk Analisis Sentimen *Gadget* Berbasis Aspek” dalam penelitian ini terbukti bahwa model klasifikasi SVM menunjukkan hasil terbaik dengan rata-rata accuracy SVM sebesar 96.43% ditinjau dari empat aspek, yaitu aspek desain sebesar 94.40%, aspek harga sebesar 97.44%, aspek spesifikasi sebesar 96.22%, dan aspek citra merk sebesar 97.63%[6]. Selanjutnya pada penelitian lain yaitu “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma *Support Vector Machine* (SVM)” memperoleh hasil SVM lebih baik dari metode sebelumnya (BM25 + KNN dan Naive Bayes). Dengan akurasi =96,68%, precision=95.82%, recall=94.04% dan AUC = 0,979[7]. Selanjutnya penelitian “Analisis Sentimen Untuk Deteksi Depresi Pada Forum Reddit Dengan Fitur Seleksi *Chi-Square* menggunakan *Support Vector Machine*” diperoleh akurasi 98.94%[8]. Selanjutnya penelitian “Klasifikasi Sentimen Ulasan Pengguna Aplikasi PeduliLindungi di Google Play Menggunakan Algoritma *Support Vector Machine* dengan Seleksi Fitur *Chi-Square*” diperoleh akurasi sebesar 93%, *recall* sebesar 86%, *precision* sebesar 98%, *specificity* sebesar 98% dan *f1-score* sebesar 92%[9]. Dari penelitian tersebut metode SVM dibantu dengan seleksi fitur *Chi-Square* dapat menghasilkan performansi yang bagus dibandingkan dengan metode klasifikasi lainnya dalam pengklasifikasian data teks.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Jamu madura sering menjadi ulasan para konsumen karena kepopulerannya, untuk mengetahui respon dan pandangan konsumen dibutuhkan

sistem dengan metode yang dapat menganalisis sentimen terhadap ulasan tersebut. Metode SVM memiliki tingkat akurasi yang cukup baik dalam pengklasifikasian data teks. Akan tetapi, *Support Vector Machine* memiliki kekurangan dalam pemilihan fitur yang sesuai. Pemilihan fitur dapat mempengaruhi hasil akurasi karena tidak semua fitur sama pentingnya.

1.2.2 Metode Penelitian

Berdasarkan hal tersebut dalam penelitian ini melakukan analisis sentimen dengan SVM dan seleksi fitur *Chi-Square* untuk mengurangi dimensi data dan fitur-fitur yang tidak relevan sehingga dapat meningkatkan hasil akurasi pada algoritma *Support Vector Machine*

1.2.3 Pertanyaan Penelitian

Adapun pertanyaan penelitian yang didapat dari rumusan masalah diatas yaitu;

1. Bagaimana melakukan analisis sentimen pada ulasan jamu madura?
2. Bagaimana performa yang dihasilkan metode *Support Vector Machine* dan seleksi fitur *Chi-Square* untuk klasifikasi sentimen pada ulasan jamu madura?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Mengetahui proses analisis sentimen pada ulasan jamu madura.
2. Mengetahui performa yang dihasilkan metode *Support Vector Machine* dan seleksi fitur *Chi-Square* untuk klasifikasi sentimen pada ulasan jamu madura.
3. Menerapkan metode *Support Vector Machine* dengan seleksi fitur *Chi-Square* untuk analisis sentimen jamu madura berdasarkan data ulasan di tokopedia.

1.3.2 Manfaat Penelitian

Adapun manfaat dilakukannya penelitian ini adalah:

1. Memahami teknologi informasi dalam bidang *text mining* dan analisis sentimen.
2. Mengetahui kinerja metode *Support Vector Machine* dan seleksi fitur *Chi-Square* untuk klasifikasi sentimen.
3. Dapat menjadi rujukan dan referensi tambahan tentang analisis sentimen dengan *Support Vector Machine*

4. Memberikan informasi analisis sentimen yang dapat membantu mengevaluasi produk jamu Madura.

1.4 Batasan Masalah

Berikut adalah batasan-batasan penelitian yang dilakukan dalam penelitian ini:

1. Data dalam penelitian ini adalah data ulasan jamu Madura di Tokopedia dengan jumlah 556 ulasan yang diambil pada tahun 2022 sampai 2023
2. Pengambilan data menggunakan library python *Selenium* dan *BeautifulSoup*
3. pelabelan berdasarkan hasil diskusi dengan ahli bahasa / oleh Guru Bahasa Indonesia.
4. Metode yang digunakan pada penelitian ini yaitu metode *Support Vector Machine* dengan seleksi fitur *Chi-Square* dan tidak membandingkan dengan algoritma lain.
5. Pengklasifikasian sentimen dijadikan ke dalam dua kategori yaitu sentimen positif dan sentimen negative
6. Bahasa pemrograman menggunakan bahasa pemrograman *Python*.

1.5 Sistematika Penulisan

Berikut merupakan beberapa aturan dalam penulisan yang digunakan dalam penyusunan laporan. Adapun sistematika proposal yang digunakan terdiri dari beberapa bagian, yaitu :

BAB I PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat, batasan masalah dan sistematika proposal.

BAB II LANDASAN TEORI

Bab ini membahas mengenai beberapa teori untuk menunjang penelitian, menguraikan tinjauan pustaka, penelitian terdahulu, serta teori penunjang lainnya sebagai bahan referensi.

BAB III ANALISA DAN PERANCANGAN SISTEM

Bab ini membahas tentang penerapan konsep dasar yang sebelumnya telah dijelaskan pada Bab II untuk menganalisis dan merancang sistem sesuai tahap-tahap untuk menyelesaikan masalah tersebut dengan menggunakan metode *Support Vector Machine* dan seleksi fitur *Chi-Square*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas mengenai implementasi sistem yang dikerjakan sesuai dengan rancangan yang dibuat pada bab sebelumnya. Dimana terdapat hasil dari tahap *crawling* data pada *e-commerce* tokopedia, hasil dari tahap pelabelan manual, hasil dari tahap *preprocessing* data (remove duplicate, *case folding*, *cleaning*, *tokenizing*, *stopword removal*, *stemming*), hasil dari tahap pembobotan dengan TF-IDF, hasil dari seleksi fitur *CHI-SQUARE*, hasil dari tahap klasifikasi dengan metode *Support Vector Machine*, hasil dari tahap evaluasi model menggunakan *confusion matrix*, terakhir terdapat hasil dari uji skenario.

BAB V PENUTUP

Bab ini membahas tentang kesimpulan dari hasil penelitian yang meliputi hasil *accuracy* serta hasil nilai *precision*, *recall* dan *f-1 score* yang didapatkan. Selain itu juga terdapat saran bagi penelitian berikutnya yang akan mengangkat topik yang sama yaitu analisis sentimen.