

BAB 1 PENDAHULUAN

1.1 Latar belakang

Analisis sentimen merupakan salah satu teknik dalam bidang text mining yang digunakan untuk mengidentifikasi serta mengklasifikasikan opini atau perasaan seseorang terhadap suatu objek ke dalam kategori sentimen tertentu, seperti positif dan negatif. Analisis sentimen banyak dimanfaatkan untuk memahami opini publik terhadap produk, layanan, maupun isu tertentu melalui media digital. Dalam konteks *e-commerce*, analisis sentimen berperan penting dalam mengetahui persepsi pelanggan terhadap suatu produk sehingga dapat mendukung pengambilan keputusan bisnis, peningkatan kualitas layanan, serta pengembangan strategi pemasaran dan sistem rekomendasi.

Dalam analisis sentimen berbasis teks, salah satu tantangan utama adalah representasi teks dan pemilihan algoritma klasifikasi yang tepat. Metode Term Frequency–Inverse Document Frequency (TF-IDF) merupakan salah satu teknik representasi teks yang paling banyak digunakan karena mampu merepresentasikan kata secara efektif dalam ruang vektor berdimensi tinggi. Beberapa penelitian menunjukkan bahwa kombinasi TF-IDF dan Support Vector Machine (SVM) mampu menghasilkan tingkat akurasi yang tinggi dan lebih unggul dibandingkan metode representasi berbasis word embedding seperti Word2Vec pada dataset berbahasa Indonesia, dengan akurasi mencapai lebih dari 85% [1]. Penelitian lain juga menunjukkan bahwa TF-IDF dengan SVM menghasilkan akurasi sebesar 76%, lebih tinggi dibandingkan algoritma KNN yang hanya mencapai 61% [2]. Hasil-hasil tersebut menunjukkan bahwa TF-IDF masih relevan dan efektif untuk analisis sentimen teks berbahasa Indonesia.

Dari sisi algoritma klasifikasi, Support Vector Machine (SVM) dikenal sebagai metode yang efektif untuk klasifikasi teks karena mampu membentuk hyperplane dengan margin pemisah optimal pada data berdimensi tinggi. Beberapa penelitian melaporkan bahwa SVM mampu mencapai tingkat akurasi yang sangat tinggi pada analisis sentimen, bahkan melebihi algoritma Naïve Bayes dan Perceptron pada berbagai dataset ulasan digital [3][4]. Temuan-temuan tersebut

memperlihatkan bahwa SVM memiliki performa yang konsisten dan kompetitif dalam menangani permasalahan klasifikasi sentimen.

Meskipun memiliki performa yang baik, Support Vector Machine (SVM) memiliki beberapa keterbatasan, yaitu sensitif terhadap outlier dan variasi data latih, serta performanya dapat menurun pada dataset dengan distribusi kelas yang tidak seimbang (imbalanced dataset) [6][11]. Selain itu, performa SVM sangat dipengaruhi oleh pemilihan fungsi kernel yang digunakan, dan tidak terdapat satu kernel SVM yang selalu optimal untuk semua jenis data [26]. Oleh karena itu, pengujian beberapa kernel SVM diperlukan untuk menemukan pemetaan fitur yang paling sesuai dengan karakteristik data ulasan yang dianalisis.

Permasalahan tersebut menjadi semakin kompleks ketika diterapkan pada data ulasan e-commerce, yang umumnya bersifat subjektif, bervariasi, serta memiliki distribusi sentimen yang tidak seimbang [8][28]. Kondisi ini juga ditemukan pada ulasan pelanggan produk Batik Madura, di mana sentimen positif cenderung mendominasi dibandingkan sentimen negatif dan netral [25]. Batik Madura merupakan salah satu warisan budaya lokal Indonesia yang memiliki ciri khas motif berani dan warna cerah. Seiring berkembangnya platform e-commerce, popularitas Batik Madura semakin meningkat, diikuti oleh bertambahnya jumlah ulasan pelanggan. Distribusi sentimen yang tidak seimbang tersebut berpotensi menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas dan menghasilkan performa yang kurang stabil [19][28].

Permasalahan utama dalam penelitian ini adalah bagaimana meningkatkan akurasi dan stabilitas klasifikasi sentimen ulasan Batik Madura pada dataset yang tidak seimbang serta sensitif terhadap variasi data latih. Untuk mengatasi keterbatasan SVM dan permasalahan tersebut, digunakan pendekatan ensemble learning dengan metode Bootstrap Aggregating (Bagging) yang bertujuan untuk mengurangi varians model dan meningkatkan stabilitas prediksi [6][7]. Bagging bekerja dengan melatih beberapa model SVM pada subset data hasil bootstrap sampling dan menggabungkan hasil prediksi menggunakan mekanisme majority voting. Beberapa penelitian menunjukkan bahwa kombinasi Bootstrap Aggregating (Bagging) dan Support Vector Machine (SVM) mampu menghasilkan peningkatan performa yang signifikan pada analisis sentimen. Salah satu penelitian melaporkan

bahwa penerapan Bagging pada SVM mampu mencapai akurasi sebesar 98,3% dengan nilai F1-score mendekati 99%, yang menunjukkan peningkatan kinerja klasifikasi yang sangat tinggi [5]. Selain itu, penelitian lain juga membuktikan bahwa metode Bagging mampu meningkatkan akurasi serta menghasilkan model klasifikasi yang lebih stabil dibandingkan dengan model tunggal [6][7][8].

Selain itu, untuk menangani ketidakseimbangan kelas pada dataset, penelitian ini menerapkan metode Synthetic Minority Over-sampling Technique (SMOTE). SMOTE menghasilkan sampel sintesis baru pada kelas minoritas melalui interpolasi antar tetangga terdekat, sehingga lebih efektif dibandingkan metode *oversampling* sederhana yang hanya menduplikasi data [19]. Penerapan SMOTE dipilih untuk mengurangi bias model terhadap kelas mayoritas dan meningkatkan kemampuan klasifikasi pada kelas minoritas [5][28].

Berdasarkan permasalahan tersebut, tujuan penelitian ini adalah untuk menentukan kernel SVM yang paling optimal serta menganalisis pengaruh penerapan Bagging terhadap peningkatan akurasi dan stabilitas SVM pada analisis sentimen ulasan Batik Madura. Oleh karena itu, penelitian ini berfokus pada analisis sentimen ulasan pelanggan Batik Madura di platform *e-commerce* Tokopedia dan Shopee dengan pendekatan Support Vector Machine yang dikombinasikan dengan *Bootstrap Aggregating (Bagging)* dan SMOTE. Penelitian ini diharapkan menghasilkan model klasifikasi sentimen yang lebih akurat dan stabil serta memberikan manfaat bagi pelaku UMKM dan pemasar digital dalam pengambilan keputusan berbasis data, sekaligus mendukung pelestarian produk budaya lokal melalui media digital.

1.2 Perumusan masalah

1.2.1 Permasalahan

Permasalahan utama dalam penelitian ini adalah bagaimana meningkatkan akurasi dan stabilitas klasifikasi sentimen ulasan Batik Madura pada data yang tidak seimbang dan bervariasi, mengingat algoritma Support Vector Machine (SVM) sebagai model tunggal sensitif terhadap variasi data latih dan noise, sehingga berpotensi menghasilkan model dengan varians tinggi dan bias terhadap kelas mayoritas.

1.2.2 Solusi Permasalahan

Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan pendekatan terpadu dengan tahapan sebagai berikut:

1. Representasi teks menggunakan TF-IDF untuk mengekstrak fitur yang relevan dari ulasan pelanggan.
2. Penerapan SVM dengan beberapa fungsi kernel (linear, polynomial, RBF, dan sigmoid) guna memperoleh model klasifikasi yang paling sesuai dengan karakteristik data.
3. Penggunaan metode ensemble Bagging-SVM untuk mengurangi varians model, meningkatkan stabilitas, dan memperbaiki akurasi klasifikasi.
4. Penerapan teknik SMOTE untuk menangani ketidakseimbangan distribusi kelas sentimen pada dataset.
5. Evaluasi kinerja model menggunakan metrik akurasi, presisi, recall, dan F1-score untuk menilai efektivitas pendekatan yang diusulkan.

1.2.3 Pertanyaan penelitian

1. Bagaimana konfigurasi parameter terbaik dari algoritma SVM dengan berbagai jenis kernel dalam melakukan klasifikasi sentimen?
2. Bagaimana penerapan Bagging-SVM dengan parameter estimator menggunakan SVM konfigurasi terbaik sebelumnya, mampu meningkatkan performa klasifikasi sentimen dibandingkan SVM ?

1.3 Tujuan dan manfaat

1.3.1 Tujuan penelitian

1. Menentukan konfigurasi parameter terbaik dari algoritma SVM dengan berbagai jenis kernel untuk memperoleh performa klasifikasi sentimen yang optimal.
2. Menganalisis pengaruh variasi jumlah *estimators* pada metode Bagging-SVM, dengan menggunakan base estimator SVM terbaik, dalam meningkatkan kualitas performa model dibandingkan SVM.

1.3.2 Manfaat Penelitian

1. Memberikan gambaran tentang efektivitas metode representasi teks TF-IDF dalam analisis sentimen ulasan pelanggan, serta perbandingan kinerja algoritma klasifikasi SVM dengan berbagai kernel dan Bagging-SVM dengan variasi jumlah *estimators* untuk memperoleh hasil yang optimal.
2. Membantu penjual Batik Madura dalam memahami opini pelanggan sehingga dapat meningkatkan strategi pemasaran berbasis data.
3. Menambah kontribusi akademik dalam penelitian analisis sentimen berbasis *Bagging-SVM*.

1.4 Batasan masalah

Agar penelitian lebih terarah, maka batasan masalah yang diterapkan dalam penelitian ini adalah:

1. Data yang digunakan dalam penelitian ini hanya berasal dari ulasan pelanggan Batik Madura pada *e-commerce* Tokopedia dan Shopee dengan jumlah total 1896 ulasan. Dimana untuk data tersebut didapat dari beberapa *seller* yaitu "ervabatik", "shoppingnaris", "alifahbatikmadura", dan "minarti1723".
2. Metode representasi teks yang digunakan adalah TF-IDF.
3. Pelabelan sentimen dilakukan secara manual dengan melibatkan 3 anotator yaitu mahasiswa dan 1 validator yaitu guru Bahasa Indonesia.
4. Model klasifikasi yang digunakan adalah SVM dengan berbagai kernel (linear, polynomial, RBF, sigmoid) dan *Bagging-SVM dengan parameter $n_estimators$* .
5. Untuk menangani ketidakseimbangan kelas dalam data, penelitian ini menggunakan metode SMOTE (*Synthetic Minority Over-sampling Technique*).
6. Evaluasi model dilakukan berdasarkan metrik akurasi, presisi, recall, dan F1-score.
7. Bahasa Pemrograman menggunakan bahasa pemrograman *Python*.

1.5 Sistematika Penulisan

Sistematika proposal ini disusun dalam tiga bab sebagai berikut:

BAB 1 PENDAHULUAN

Bagian ini memberikan tinjauan umum mengenai penelitian yang akan dilakukan. Isinya mencakup latar belakang, perumusan masalah, batasan-batasan yang ditetapkan, tujuan serta manfaat dari penelitian ini, serta sistematika dalam penyusunan laporan.

BAB 2 TINJAUAN PUSTAKA

Bagian ini menjelaskan dasar teori yang menjadi acuan dan pedoman dalam penelitian analisis sentimen pada UMKM Batik Madura dengan menggunakan Pendekatan *Bagging-SVM*, serta mencakup tinjauan literatur yang relevan dalam bidang penelitian ini.

BAB 3 METODE PENELITIAN

Bab ini membahas metode penelitian yang digunakan dalam penelitian ini. Pada bab ini dijelaskan tahapan-tahapan penelitian secara sistematis, mulai dari pengumpulan data ulasan pelanggan dari *e-commerce* Tokopedia dan Shopee, proses pelabelan data sentimen, serta tahapan preprocessing teks yang meliputi *case folding*, *tokenization*, *stopword removal*, dan *stemming*. Selain itu, bab ini juga menjelaskan metode pembentukan fitur menggunakan TF-IDF, pembagian data latih dan data uji, penerapan algoritma klasifikasi *Support Vector Machine* (SVM), serta penerapan metode Bagging. Tahapan evaluasi model dengan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* juga dijelaskan pada bab ini.

BAB 4 HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil eksperimen dari skenario pengujian yang telah dilakukan. Penjelasan meliputi performa model dalam berbagai kombinasi parameter, visualisasi hasil evaluasi AUC, serta implementasi sistem dalam bentuk antarmuka menggunakan Streamlit. Pembahasan dilakukan untuk menganalisis hasil serta membandingkannya dengan studi sebelumnya.

BAB 5 KESIMPULAN DAN SARAN

Bab terakhir berisi kesimpulan dari hasil penelitian berdasarkan tujuan yang telah ditetapkan, serta saran untuk pengembangan penelitian di masa yang akan datang, baik dalam hal metode, data, maupun aplikasi sistem.