

BAB I PENDAHULUAN

1.1 Latar Belakang

Di masa kini yang bergerak cepat di bidang ekonomi dan keuangan digital yaitu aplikasi dompet elektronik atau e-wallet sudah menjadi bagian penting dalam kehidupan sehari-hari masyarakat Indonesia dalam bertransaksi. Nilai transaksi melalui dompet digital di Indonesia mengalami peningkatan sangat pesat dalam beberapa tahun terakhir, dengan pertumbuhan hingga dua digit pertahun. Pengguna aplikasi seperti DANA, OVO, GoPay, dan lainnya terus bertambah seiring dengan semakin luasnya akses internet dan semakin banyaknya orang yang mulai menggunakan layanan keuangan berbasis smartphone. Dalam situasi seperti ini, pandangan dan persepsi pengguna menjadi sangat penting. Komentar dan opini pengguna yang tersampaikan melalui platform seperti Google Play Store atau media sosial bisa menunjukkan sejauh mana kepuasan, keluhan, atau harapan mereka terhadap layanan tersebut. Maka, analisis sentimen menjadi cara yang efektif untuk mengekstraksi dan memantau opini pengguna secara teratur dan sistematis.

Fenomena ini terjadi ketika ulasan pengguna terhadap aplikasi e-wallet tidak selalu sesuai dengan skor atau rating yang ditampilkan. Misalnya, dalam penelitian Analisis Sentimen Pengguna Aplikasi Dana pada kolom Google Play store ditemukan sekitar 65% dari ulasan pengguna menunjukkan sentimen negatif meskipun aplikasi memiliki rating yang baik di toko aplikasi [1]. Selain itu juga sebagian besar pengguna (84,7%) memiliki pandangan negatif terhadap aplikasi DANA, sementara hanya 8,8% yang memberikan penilaian positif. Perbedaan ini menunjukkan adanya ketidaksesuaian antara rating secara keseluruhan dan isi ulasan yang diberikan [2]. Di sisi lain, aplikasi GoPay menggunakan metode kernel RBF pada SVM dengan dataset sebanyak 29.554 ulasan melaporkan akurasi sekitar 65%, presisi 71,43%, recall 50%, dan f1-score 58,82%. Hasil-hasil ini menunjukkan bahwa meskipun metode klasifikasi teks sudah banyak digunakan,

kinerja model masih belum maksimal dalam konteks ulasan aplikasi keuangan digital [3].

Masalah utama di bidang ini adalah bagaimana membuat model klasifikasi sentimen yang tepat dan bisa diandalkan untuk menganalisis ulasan pengguna aplikasi DANA. Model ini harus mampu mengenali berbagai ciri khas bahasa Indonesia seperti istilah santai, emotikon, serta penggunaan kata-kata khas. Selain itu, juga harus mengatasi ketidakseimbangan antara ulasan positif dan negatif, serta kesulitan dalam menyetel parameter model (*hyperparameter*) secara optimal. Tantangan tambahannya adalah memilih jenis *kernel SVM* yang tepat, biasanya menggunakan *kernel RBF*, serta melakukan pengaturan parameter seperti *C* dan *gamma* agar model tidak terlalu sederhana (*underfitting*) atau terlalu kompleks (*overfitting*). Dalam situasi tersebut, diperlukan pendekatan yang mampu meningkatkan kinerja *SVM* dengan *kernel RBF* serta memperbaiki efektivitas metode *Naïve Bayes* dengan menyesuaikan parameter secara tepat. Tanpa ada proses pengoptimalan parameter model, performa algoritma tidak akan berkembang secara signifikan, meskipun volume data yang digunakan sudah cukup besar.

Misalnya, penelitian Analisis Sentimen Aplikasi DANA menggunakan *SVM* dengan representasi *Word2Vec* melaporkan tingkat akurasi sekitar 88% [4]. Selain itu juga dengan membandingkan *SVM* dan *Naïve Bayes* pada dataset ulasan aplikasi atau produk, seperti marketplace, menunjukkan bahwa *SVM* rata-rata lebih unggul dibanding *Naïve Bayes* (misalnya 86,14% vs 85,53%) dalam satu studi perbandingan produk kecantikan. Namun, sebagian besar penelitian tersebut belum melakukan proses tuning *hyperparameter* secara sistematis atau mengeksplorasi secara mendalam *kernel RBF*, terutama dalam konteks aplikasi dompet digital di Indonesia [5]. Adapun yang mengenai optimasi *SVM* atau pemilihan fitur dalam konteks bahasa Indonesia sudah dilakukan, seperti *A Combination of Query Expansion Ranking and GA-SVM* yang menggabungkan algoritma genetika dengan *SVM* untuk pemilihan fitur dalam analisis sentimen Indonesia. Namun, penelitian tersebut tidak secara khusus ditujukan untuk domain aplikasi keuangan

digital seperti DANA dan juga tidak membandingkan dengan model Naïve Bayes yang sudah dioptimasi [6].

Penelitian ini bertujuan mengatasi kekurangan sebelumnya dengan merancang eksperimen untuk membandingkan dua metode, yaitu *SVM* dengan *kernel RBF* dan *Naïve Bayes*, masing-masing dilengkapi dengan teknik optimasi parameter atau *tuning hyperparameter* menggunakan metode *Grid Search*. Data yang digunakan berasal dari dataset ulasan aplikasi DANA yang diambil dari platform Google Play Store. Dengan proses ini, diharapkan dapat dibuat model klasifikasi sentimen yang lebih akurat dan efisien, serta mampu menangani karakteristik bahasa Indonesia yang rumit, seperti penggunaan kata-kata informal, emotikon, dan variasi ekspresi yang sering muncul dalam ulasan pengguna. Selain itu, hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan model analisis sentimen di bidang aplikasi keuangan digital, serta menjadi referensi dalam penerapan *tuning hyperparameter* untuk meningkatkan kinerja algoritma pembelajaran mesin dalam konteks bahasa Indonesia.

1.2 Rumusan Masalah

Berdasarkan konteks tersebut, penelitian ini berfokus pada penerapan dan perbandingan dua algoritma populer, yaitu *Support Vector Machine (SVM)* dengan *kernel RBF* dan *Naïve Bayes* dengan *hyperparameter tuning*, untuk mengukur efektivitasnya dalam mengklasifikasikan sentimen pengguna aplikasi DANA.

Berdasarkan uraian tersebut, rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana performa algoritma *Support Vector Machine* (dengan *kernel RBF*) dan *Naïve Bayes* (dengan *hyperparameter tuning*) dalam menganalisis dan mengklasifikasikan sentimen ulasan pengguna aplikasi DANA?
2. Bagaimana optimasi parameter (*hyperparameter tuning*) pada *Support Vector Machine (Kernel RBF)* dan *Naïve Bayes* untuk meningkatkan *accuracy*, *precision*, *recall* dan *F-1 score* pada analisis sentimen pengguna DANA?

3. Model mana yang memberikan hasil terbaik antara *Support Vector Machine (kernel RBF)* dan *Naïve Bayes dengan (Hyperparameter Tuning)* dalam mengklasifikasikan sentimen positif dan negatif pada pengguna aplikasi DANA?

1.3 Batasan Masalah

Pada skripsi ini, untuk menghindari ruang lingkup yang terlalu jauh, adapun batasan masalah yang digunakan adalah sebagai berikut :

1. Menganalisis sentimen ulasan pengguna aplikasi DANA dengan membandingkan dua algoritma machine learning *SVM* dengan *kernel RBF* dan *Naïve Bayes* yang telah dilakukan tuning hyperparameter.
2. Data diambil dari Scraping Data sebanyak 5.200 ulasan pengguna aplikasi DANA dari Google Play Store, seluruhnya dalam bahasa Indonesia dan sudah melalui proses kurasi awal.
3. Analisis hanya dilakukan pada teks ulasan tanpa mempertimbangkan elemen non-teks (bintang, emoji, data pengguna), serta dibatasi pada bahasa Indonesia saja.
4. Penelitian ini membandingkan hasil kerja metode *SVM* dengan *kernel RBF* dan *Naïve Bayes* setelah mengatur parameter *hyperparameter*, untuk mengetahui peningkatan tingkat keakuratan.
5. Penelitian hanya dilakukan dalam lingkungan perangkat lunak saja, tanpa menguji bagaimana sistem tersebut diimplementasikan dalam bentuk aplikasi atau perangkat keras

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja dua algoritma machine learning, yaitu *Support Vector Machine (SVM)* dengan *kernel Radial Basis Function (RBF)* dan *Naïve Bayes*, dalam mengklasifikasikan sentimen ulasan pengguna aplikasi DANA berbahasa Indonesia ke dalam 2 kategori, yakni positif dan negatif.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini diharapkan :

1. Meningkatkan pemahaman dalam bidang analisis sentimen dan pembelajaran mesin, terutama dalam penerapan algoritma *Support Vector Machine (SVM)* dengan kernel *RBF* dan *Naïve Bayes*, yang melibatkan penyesuaian parameter hyper pada data teks berbahasa Indonesia.
2. Diharapkan memberikan referensi dan pemahaman mengenai cara meningkatkan kualitas pelayanan dan mengidentifikasi keluhan pengguna.
3. Memberikan kontribusi terhadap pengembangan penelitian baik secara teori maupun praktik, khususnya di bidang analisis sentimen dan *Natural Language Processing* pada teks berbahasa Indonesia

1.6 Sistematika Penulisan

Untuk memperoleh gambaran yang lebih mudah dipahami mengenai isi penulisan skripsi ini , sistematika penulisan ditulis dengan menguraikan bab demi bab, meliputi :

BAB I PENDAHULUAN

Berisikan tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA Berisi tentang studi literatur, dasar teori yang digunakan dalam skripsi yang berasal dari jurnal, buku, dan internet yang menjadi dasar dari penelitian ini.

BAB III METODE PENELITIAN

Bab ini akan membahas lebih detail tahapan analisis *Support Vector Machine (RBF)* dan *Naïve Bayes (Hyperparameter tuning)* yang dihasilkan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini akan membahas mengenai proses dan hasil dari analisis *Support Vector Machine* dan *Naïve Bayes* yang telah dilakukan , mengamati jumlah sentimen

analisis serta menyimpulkan hasil yang didapat dalam penelitian ini.

BAB V PENUTUP

Berisi tentang kesimpulan yang diambil dari proses penelitian dan saran dari penulis terkait pengembangan dan progres lanjutan untuk mendapatkan data yang lebih spesifik

