

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi digital dalam satu dekade terakhir telah mendorong transformasi besar dalam berbagai sektor, termasuk sektor telekomunikasi. Perusahaan penyedia layanan telekomunikasi berlomba-lomba menghadirkan inovasi berbasis aplikasi mobile untuk meningkatkan kualitas layanan dan memberikan kemudahan akses bagi pelanggan. Aplikasi mobile tidak hanya berfungsi sebagai media transaksi, tetapi juga sebagai sarana komunikasi, pelayanan mandiri, serta pengelolaan akun pelanggan secara real time. Transformasi ini menyebabkan interaksi antara perusahaan dan pelanggan semakin intensif melalui *platform* digital. Salah satu bentuk interaksi tersebut tercermin dalam fitur ulasan pengguna pada *platform* distribusi aplikasi. Ulasan pengguna merupakan sumber informasi yang sangat berharga karena memuat opini, pengalaman, serta persepsi pelanggan terhadap kualitas layanan yang diberikan. Berbeda dengan rating berbentuk angka, ulasan berbentuk teks mampu memberikan gambaran yang lebih mendalam mengenai kelebihan, kekurangan, serta permasalahan yang dialami pengguna. Oleh karena itu, analisis terhadap ulasan pengguna menjadi penting sebagai dasar evaluasi dan perbaikan layanan.

Namun demikian, jumlah ulasan yang sangat besar serta variasi bahasa yang digunakan pengguna menjadikan proses analisis secara manual tidak lagi efektif dan efisien. Ulasan yang masuk setiap hari dapat mencapai ratusan hingga ribuan, sehingga memerlukan pendekatan komputasional untuk mengolah dan mengklasifikasikan opini tersebut secara otomatis. Dalam konteks ini, analisis sentimen menjadi solusi yang relevan untuk mengidentifikasi polaritas opini pengguna ke dalam kategori tertentu, seperti sentimen positif dan negatif. Analisis sentimen merupakan bagian dari *text mining* yang bertujuan untuk mengekstraksi informasi subjektif dari data teks tidak terstruktur. Proses ini melibatkan tahapan praproses teks, representasi fitur, hingga klasifikasi menggunakan algoritma pembelajaran mesin. Representasi fitur yang umum digunakan dalam analisis sentimen adalah Term Frequency–Inverse Document Frequency (TF-IDF), yang mampu mengukur tingkat kepentingan suatu kata dalam dokumen relatif terhadap

keseluruhan korpus. Meskipun efektif, representasi TF-IDF menghasilkan vektor fitur berdimensi tinggi dan bersifat *sparse*, sehingga memerlukan metode klasifikasi yang mampu menangani karakteristik tersebut secara optimal.

Berbagai penelitian terdahulu menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memiliki performa yang baik dalam klasifikasi teks, termasuk analisis sentimen. SVM dikenal mampu menangani data berdimensi tinggi dan menghasilkan model dengan kemampuan generalisasi yang baik. Beberapa studi sebelumnya menerapkan SVM dalam analisis sentimen ulasan aplikasi dan menunjukkan tingkat akurasi yang relatif tinggi. Badriyah et al. (2025) menerapkan SVM dengan teknik Synthetic Minority Over-sampling Technique (SMOTE) untuk mengatasi ketidakseimbangan kelas dan memperoleh peningkatan performa klasifikasi. Audiansyah et al. (2022) juga menggunakan SVM untuk menganalisis sentimen ulasan aplikasi dan menunjukkan hasil yang efektif dalam mengklasifikasikan opini pengguna. Penelitian lain yang menerapkan SVM dalam analisis sentimen pada berbagai domain aplikasi digital juga menunjukkan performa yang kompetitif dibandingkan metode lain (Fitriyana et al., 2023; Mustasaruddin et al., 2023; Khoirunnisaa et al., 2024). Meskipun demikian, sebagian besar penelitian terdahulu cenderung berfokus pada perbandingan performa algoritma atau peningkatan akurasi melalui teknik optimasi tambahan. Kajian yang secara khusus menekankan kesesuaian metodologis antara karakteristik data teks hasil pembobotan TF-IDF dengan pemilihan fungsi *kernel* pada SVM masih relatif terbatas. Padahal, pemilihan *kernel* merupakan aspek penting dalam implementasi SVM karena berpengaruh langsung terhadap kemampuan model dalam membentuk *hyperplane* pemisah antar kelas.

SVM menyediakan beberapa jenis fungsi *kernel*, seperti *linear*, *polynomial*, *radial basis function* (RBF), dan *sigmoid*. Pada data teks yang direpresentasikan menggunakan TF-IDF, fitur yang dihasilkan memiliki dimensi yang sangat tinggi dan sebagian besar bernilai nol (*sparse*). Dalam kondisi seperti ini, *kernel linear* sering kali lebih sesuai karena bekerja langsung pada ruang fitur asli tanpa melakukan transformasi non-*linear* ke dimensi yang lebih tinggi. Selain itu, *kernel linear* memiliki kompleksitas komputasi yang lebih rendah dibandingkan *kernel non-linear*, sehingga lebih efisien dan stabil dalam implementasi praktis.

Penggunaan *kernel non-linear* memang berpotensi meningkatkan fleksibilitas model dalam memisahkan kelas yang tidak terpisah secara *linear*. Namun, pada data teks berdimensi tinggi, transformasi *non-linear* dapat meningkatkan risiko overfitting serta menambah beban komputasi yang signifikan. Oleh karena itu, diperlukan kajian yang tidak hanya berorientasi pada pencapaian akurasi tinggi, tetapi juga mempertimbangkan kesesuaian struktural antara data dan model yang digunakan. Sejumlah penelitian terdahulu menunjukkan bahwa Support Vector Machine (SVM) memiliki performa yang baik dalam klasifikasi teks karena kemampuannya menangani data berdimensi tinggi. Namun, mayoritas penelitian lebih berfokus pada perbandingan algoritma, peningkatan akurasi, atau penerapan teknik optimasi tambahan seperti penyeimbangan kelas dan tuning parameter. Pendekatan tersebut cenderung menempatkan akurasi sebagai tujuan utama tanpa membahas secara mendalam kesesuaian antara karakteristik data dan model yang digunakan.

Gap penelitian terletak pada masih terbatasnya kajian yang secara eksplisit menganalisis keselarasan metodologis antara data teks hasil pembobotan TF-IDF yang *high dimensional* dan *sparse* dengan pemilihan fungsi *kernel* pada SVM. Padahal, fungsi *kernel* merupakan komponen krusial yang menentukan bagaimana data dipetakan dan dipisahkan dalam ruang fitur, serta berpengaruh langsung terhadap stabilitas dan kemampuan generalisasi model. Minimnya pembahasan mengenai justifikasi teoretis dalam pemilihan *kernel* menyebabkan penggunaan *kernel* sering kali didasarkan pada pendekatan empiris semata, bukan pada pertimbangan struktural karakteristik data. Berdasarkan kesenjangan tersebut, penelitian ini mengambil posisi dengan mengimplementasikan Support Vector Machine menggunakan *kernel linear* dalam analisis sentimen ulasan pengguna aplikasi MyXL. Pemilihan *kernel linear* didasarkan pada pertimbangan metodologis bahwa data teks hasil pembobotan TF-IDF bersifat *high dimensional* dan *sparse*, sehingga lebih sesuai ditangani oleh model *linear* yang efisien dan memiliki kemampuan generalisasi yang baik.

Dengan pendekatan tersebut, penelitian ini tidak hanya berorientasi pada pencapaian akurasi, tetapi juga memberikan kontribusi metodologis dengan menegaskan bahwa pemilihan model klasifikasi harus didasarkan pada kesesuaian

struktural antara data dan metode. Hasil penelitian diharapkan memberikan manfaat praktis bagi pengembang aplikasi serta memperkaya literatur mengenai pentingnya keselarasan antara karakteristik data teks dan pemilihan *kernel* dalam implementasi SVM.

1.2. Perumusan Masalah

1.2.1 Permasalahan

Berdasarkan latar belakang di atas didapatkan beberapa permasalahan yaitu:

1. Tingginya jumlah ulasan pengguna aplikasi *MyXL* di Google Play Store menyulitkan proses analisis sentimen secara manual.
2. Data teks hasil pembobotan TF-IDF bersifat *high dimensional* dan *sparse* sehingga memerlukan metode klasifikasi yang sesuai secara metodologis.
3. Penelitian sebelumnya belum secara mendalam mengkaji kesesuaian pemilihan *kernel* pada SVM terhadap karakteristik data teks tersebut.
4. Diperlukan metode klasifikasi yang stabil, efisien secara komputasi, serta memiliki kemampuan generalisasi yang baik dalam mengklasifikasikan sentimen ulasan aplikasi

1.2.2 Metode Usulan

Mengimplementasikan *Support Vector Machine* (SVM) dengan *kernel linear* untuk klasifikasi sentimen ulasan aplikasi *MyXL*, melalui tahapan: *Scraping* data, *Preprocessing* teks, Pelabelan *Lexicon Based*, Klasifikasi *Support Vector Machine*, dan Evaluasi model. *Kernel linear* dipilih karena sesuai dengan karakteristik data teks yang berdimensi tinggi dan *sparse* serta memiliki efisiensi komputasi yang lebih baik dibandingkan *kernel non-linear*.

1.2.3 Pertanyaan Penelitian

Berdasarkan perumusan masalah di atas, maka pertanyaan penelitian yang diajukan adalah sebagai berikut:

1. Bagaimana implementasi SVM *kernel linear* dalam klasifikasi sentimen ulasan *MyXL*?
2. Bagaimana kinerja model SVM *kernel linear* berdasarkan *accuracy*, *precision*, *recall*, dan *F1 score*?
3. Apakah *kernel linear* sesuai secara metodologis untuk data teks TF-IDF pada analisis sentimen ulasan *MyXL*?

1.3. Tujuan dan Manfaat Penelitian

1.3.1. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Mengimplementasikan metode *Support Vector Machine* (SVM) dengan *kernel linear* dalam analisis sentimen ulasan aplikasi *MyXL* pada Google Play Store.
2. Menganalisis kinerja model SVM *kernel linear* dalam mengklasifikasikan sentimen positif dan negatif berdasarkan metrik evaluasi.
3. Membuktikan kesesuaian metodologis antara karakteristik data teks TF-IDF yang berdimensi tinggi dan *sparse* dengan penggunaan *kernel linear* pada SVM.

1.3.2. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis maupun praktis sebagai berikut:

1. Manfaat Teoritis
 - a. Memberikan kontribusi akademis dalam pengembangan metode analisis sentimen berbasis *Machine learning* khususnya penerapan *Support Vector Machine* untuk analisis sentimen.
 - b. Memperkuat kajian mengenai kesesuaian penggunaan *kernel linear* pada *Support Vector Machine* untuk klasifikasi teks berdimensi tinggi.
 - c. Menjadi referensi bagi penelitian selanjutnya dalam bidang *text mining* dan analisis sentimen berbasis pembobotan TF-IDF.
2. Manfaat Praktis
 - a. Memberikan informasi kepada pihak pengembang aplikasi *MyXL* mengenai persepsi dan kepuasan pengguna berdasarkan ulasan di Google Play Store.
 - b. Membantu dalam mengidentifikasi keluhan dominan pengguna sebagai dasar perbaikan layanan.
 - c. Menyediakan sistem analisis sentimen yang efisien dan dapat digunakan sebagai alat pendukung pengambilan keputusan berbasis data.
3. Manfaat Metodologis
 - a. Menunjukkan bahwa pemilihan *kernel linear* pada SVM bukan hanya alternatif sederhana, tetapi merupakan pendekatan yang sesuai secara struktural untuk data teks berbasis TF-IDF.

- b. Memberikan gambaran implementasi model klasifikasi yang stabil dan efisien secara komputasi.

1.4. Batasan Masalah

Agar penelitian lebih terarah dan tidak menyimpang dari tujuan, maka ditetapkan batasan masalah sebagai berikut:

1. Data yang digunakan berupa ulasan pengguna aplikasi *MyXL* yang diambil dari Google Play Store.
2. Jumlah dataset yang digunakan sebanyak 2.000 ulasan.
3. Klasifikasi sentimen dibatasi menjadi dua kelas, yaitu positif dan negatif.
4. Representasi fitur teks menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF).
5. Metode klasifikasi yang digunakan adalah *Support Vector Machine* (SVM) dengan *kernel linear* tanpa perbandingan dengan *kernel* lain atau algoritma lain.
6. Evaluasi model menggunakan *Confusion Matrix*, *accuracy*, *precision*, *recall*, dan *F1 score*.

1.5. Sistematika Penulisan

Secara umum, skripsi ini disusun ke dalam beberapa bab. Penyusunan tersebut bertujuan untuk menyajikan pembahasan secara sistematis dan terstruktur, sehingga memudahkan pembaca dalam memahami keseluruhan isi skripsi. Adapun susunan sistematika penulisan skripsi ini adalah sebagai berikut.

BAB I PENDAHULUAN

Dalam bab ini akan diuraikan mengenai latar belakang permasalahan, permasalahan, metode usulan, pertanyaan penelitian, tujuan penelitian, manfaat penelitian, batasan yang digunakan serta sistematika penulisan penelitian.

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan dasar teori dan metode yang digunakan sebagai landasan dan alat untuk menyelesaikan masalah. Kami akan menjelaskan *Support Vector Machine* (SVM) yang digunakan dalam analisis sentiment ulasan *MyXL* untuk mencapai tujuan penelitian yang ditetapkan.

BAB III METODOLOGI PENELITIAN

Bab ini merincikan runtutan tahapan yang dilakukan dalam penelitian ini meliputi studi literatur, pengambilan data, pengembangan model, dan evaluasi. Masing-masing tahapan memiliki subbab yang menjelaskan apa yang dilakukan pada tahap tersebut.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan mengenai hasil pelaksanaan penelitian yaitu program yang dibuat serta hasil skenario pengujian. Program dibahas secara rinci mengenai cara kerjanya, serta disebutkan perangkat-perangkat yang digunakan. Skenario pengujian dibahas sebagai bentuk evaluasi serta untuk menentukan skenario terbaik.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil pelaksanaan penelitian dan menjawab pertanyaan penelitian. Bab ini juga berisi saran yang dapat digunakan untuk penelitian selanjutnya.

DAFTAR PUSTAKA

Daftar pustaka berisi daftar referensi yang digunakan dalam penyusunan skripsi ini.