

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era perkembangan teknologi informasi yang pesat telah membawa perubahan besar dalam berbagai aspek kehidupan manusia, terutama dalam hal konsumsi media hiburan. Layanan streaming film seperti Netflix, Disney+, dan Amazon Prime Video telah menjadi pilihan utama masyarakat modern untuk menikmati film dan serial secara fleksibel. Dengan hanya menggunakan perangkat mobile dan koneksi internet, pengguna dapat menikmati berbagai tayangan dari seluruh dunia. Aplikasi streaming tersebut umumnya tersedia di Google Play Store, dan pengguna diberi kebebasan untuk memberikan ulasan (review) dan penilaian (rating) berdasarkan pengalaman mereka.

Ulasan-ulasan tersebut merupakan sumber data berharga yang dapat digunakan untuk mengetahui tingkat kepuasan pengguna serta kekurangan dari layanan yang disediakan. Namun, dengan volume ulasan yang sangat besar dan terus bertambah setiap hari, analisis secara manual menjadi tidak efisien dan rentan terhadap bias. Oleh karena itu, dibutuhkan pendekatan berbasis teknologi seperti analisis sentimen (sentiment analysis) untuk memahami opini publik secara otomatis. Analisis sentimen merupakan salah satu bidang dalam pemrosesan bahasa alami (Natural Language Processing/NLP) yang bertujuan untuk mengidentifikasi dan mengklasifikasikan opini dalam teks menjadi kategori seperti positif, negatif, atau netral.

Dalam proses klasifikasi sentimen, algoritma machine learning banyak digunakan karena kemampuannya dalam mengenali pola dari data teks yang kompleks. Di antara algoritma yang sering digunakan adalah Support Vector Machines (SVM) dan K-Nearest Neighbor (K-NN). Support Vector Machines dikenal mampu menghasilkan pemisah kelas (hyperplane) terbaik melalui margin maksimum, serta tahan terhadap overfitting pada data berdimensi tinggi (Cortes & Vapnik, 1995). Sementara itu, K-Nearest Neighbor adalah algoritma berbasis instance yang sederhana namun cukup efektif, yang bekerja dengan mencari kedekatan data uji terhadap sejumlah data latih terdekat (Cover & Hart, 1967).

Berbagai penelitian telah dilakukan untuk menilai efektivitas SVM dan K-NN dalam tugas klasifikasi sentimen. Wicaksono dan Mardiyanto (2021) menunjukkan bahwa SVM menghasilkan akurasi tinggi pada analisis ulasan produk e-commerce. Pratama dan Sari (2020) membandingkan SVM dan K-NN dalam klasifikasi review film dan menemukan bahwa K-NN

cukup kompetitif dalam performa. Santoso et al. (2019) membuktikan bahwa SVM mampu mencapai akurasi hingga 85% dalam klasifikasi sentimen aplikasi mobile. Penelitian Putri dan Haryanto (2018) menggunakan K-NN pada ulasan restoran dan menunjukkan bahwa metode ini cocok untuk dataset kecil dan mudah diinterpretasikan. Nugroho dan Lestari (2022) juga menyimpulkan bahwa SVM lebih stabil dalam menangani data Twitter yang bersifat pendek dan berisik dibandingkan K-NN.

Penelitian lainnya oleh Maulana dan Yusuf (2020) menggunakan SVM untuk analisis sentimen review hotel dan menunjukkan performa yang unggul dibanding metode lainnya. Sementara itu, Rahayu et al. (2019) membandingkan performa K-NN dan Decision Tree dalam klasifikasi sentimen produk fashion, dan menemukan bahwa K-NN lebih unggul dalam presisi. Septiani dan Wardhani (2021) mengimplementasikan SVM pada review pengguna aplikasi transportasi online dan memperoleh akurasi lebih dari 90%. Penelitian oleh Firmansyah dan Fauzi (2022) mengkaji perbandingan SVM dan K-NN pada analisis review aplikasi edukasi, dan menemukan bahwa SVM memberikan hasil klasifikasi lebih konsisten.

Selain itu, penelitian dari Aulia dan Ramdani (2020) menemukan bahwa integrasi preprocessing seperti TF-IDF dan stemming sangat memengaruhi performa K-NN dalam klasifikasi review makanan. Hasanah dan Kusuma (2022) menunjukkan bahwa penggunaan SVM dengan kernel linear sangat efektif dalam analisis sentimen e-learning. Penelitian oleh Yuliana dan Sembiring (2021) menunjukkan bahwa K-NN dengan pemilihan nilai k yang tepat dapat meningkatkan akurasi dalam klasifikasi sentimen sosial media. Ramadhan dan Akbar (2020) mengimplementasikan SVM dan K-NN untuk klasifikasi review aplikasi perbankan digital dan menyatakan bahwa SVM lebih cocok untuk dataset besar. Penelitian oleh Lestari dan Mahendra (2023) juga mengungkap bahwa kombinasi preprocessing yang optimal mampu meningkatkan performa K-NN hingga mendekati akurasi SVM. Terakhir, Sari dan Anwar (2021) membandingkan kinerja SVM, K-NN, dan Naive Bayes dalam ulasan produk elektronik dan menyimpulkan bahwa SVM memiliki tingkat akurasi tertinggi.

Berdasarkan paparan tersebut, dapat disimpulkan bahwa SVM dan K-NN sama-sama memiliki keunggulan dan keterbatasan masing-masing. Oleh karena itu, dalam penelitian ini dilakukan analisis dan perbandingan antara algoritma Support Vector Machines dan K-Nearest Neighbor dalam mengklasifikasikan sentimen review aplikasi streaming film yang diambil dari Google Play Store.

Dalam konteks keilmuan Sistem Informasi, penelitian ini tidak hanya berfokus pada kinerja algoritma klasifikasi, tetapi juga pada bagaimana hasil analisis sentimen dapat dimanfaatkan untuk mendukung proses pengambilan keputusan strategis. Informasi yang

diperoleh dari hasil klasifikasi ulasan pengguna dapat menjadi dasar evaluasi kinerja sistem, peningkatan kualitas layanan, perbaikan fitur, serta pengembangan strategi *user experience* yang lebih efektif. Dengan demikian, penelitian ini memberikan kontribusi terhadap pemanfaatan data tidak terstruktur sebagai bagian dari sistem pendukung keputusan (Decision Support System) dalam pengelolaan dan pengembangan layanan berbasis teknologi informasi.

1.2 Identifikasi Masalah

Berdasarkan uraian latar belakang di atas, dapat diidentifikasi beberapa permasalahan sebagai berikut :

1. Belum tersedia perbandingan langsung analisis sentiment pada ulasan pengguna aplikasi Netflix menggunakan algoritma *Support Vector Machine* dan *K – Nearest Neighbor* dengan pelabelan *Lexicon 3 class*.
2. Terjadinya kasus *Imbalance data* dalam proses pelabelan menggunakan vader sentimen

1.3 Rumusan Masalah

Berdasarkan identifikasi masalah yang telah dijelaskan, dapat dirumuskan masalah-masalah sebagai berikut :

1. Belum tersedia perbandingan analisis sentiment pada ulasan pengguna aplikasi Netflix menggunakan algoritma *Support Vector Machine* dan *K – Nearest Neighbor* dengan pelabelan *Lexicon 3 class*.
2. Bagaimana penanganan *imbalance data* yang tepat agar model klasifikasi sentimen dapat menghasilkan performa yang optimal?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, tujuan penelitian ini adalah sebagai berikut :

1. Menyediakan analisis perbandingan sentimen pada ulasan pengguna aplikasi Netflix menggunakan algoritma *Support Vector Machines* dan *K-Nearest Neighbor* dengan pelabelan *Lexicon 3 class*.

2. Menangani kasus *imbalance* data agar model klasifikasi sentimen dapat menghasilkan performa yang optimal.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Menerapkan dua algoritma klasifikasi yaitu, *Support Vector Machines* dan *K-Nearest Neighbor* pada model analisis sentimen.
2. Dataset berupa ulasan pengguna *Netflix* yang bersumber dari *Google Play Store* dikumpulkan selama periode bulan Maret – Juni tahun 2025, berjumlah 5000 ulasan berbahasa Indonesia.
3. Menggunakan *tools google colab* dan bahasa pemrograman *Python*.

1.6 Kontribusi

Penelitian ini diharapkan mampu memberikan kontribusi sebagai berikut :

1. Memberikan masukan atau *support* kepada pengembang *Netflix* mengenai aspek aplikasi yang perlu ditingkatkan atau perlu diperbaiki.
2. Menambah wawasan akademik mengenai performa algoritma *Support Vector Machines* dan *K-Nearest Neighbor* dalam konteks analisis sentimen.
3. Menyediakan literatur atau referensi bagi penelitian lanjutan yang ingin melakukan analisis sentimen terhadap ulasan aplikasi yang berbeda.
4. Memberikan kontribusi pada keilmuan Sistem Informasi khususnya dalam pemanfaatan text mining dan machine learning untuk mendukung proses evaluasi sistem berbasis ulasan pengguna.