

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi telah mengubah cara masyarakat berinteraksi dan mengekspresikan opini mereka. Melalui media sosial, platform *e-commerce*, forum daring, dan aplikasi berbasis web, setiap harinya puluhan hingga ratusan ribu komentar, ulasan, dan posting baru muncul yang mengandung nilai opini atau sentimen publik (Suhendra, 2024). Data bentuk-teks tersebut menyimpan potensi besar untuk dijadikan sumber insight dalam pengambilan keputusan bisnis, pemasaran, hingga kebijakan publik. Namun, potensi ini belum sepenuhnya dimanfaatkan secara optimal.

Namun demikian, besarnya volume data opini yang tersedia belum sepenuhnya dimanfaatkan secara optimal. Salah satu penyebab utama adalah kompleksitas teknis dalam melakukan analisis terhadap data teks tersebut. Analisis sentimen memerlukan serangkaian tahapan, mulai dari pra-pemrosesan teks, ekstraksi fitur, hingga penerapan algoritma klasifikasi (Tan & Lee, 2023). Bagi sebagian besar pihak atau institusi non-teknis, keterbatasan pengetahuan dan akses terhadap sistem analisis menjadi hambatan sehingga data opini publik hanya tersimpan tanpa diolah lebih lanjut menjadi informasi yang bernilai. Untuk banyak pihak, walaupun data sudah tersedia, kemampuan teknis atau akses ke alat analisis masih menjadi penghalang utama sehingga data opini publik hanya “terkumpul” dan tidak “tergunakan” secara maksimal.

Penelitian yang dilakukan terhadap berbagai dataset berbahasa Indonesia menunjukkan bahwa pendekatan khusus dan metode yang sesuai konteks diperlukan agar performa klasifikasi menjadi optimal. Dalam konteks Bahasa Indonesia, analisis sentimen memiliki tantangan tersendiri karena karakteristik morfologi, variasi imbuhan, serta penggunaan bahasa informal, Bahasa gaul atau *slang* yang cukup tinggi dan juga mengingat Bahasa yang digunakan oleh *Internet Citizen* (Netizen) pada era moderen dalam memberikan opini atau pendapat terdapat banyak imbuhan kata maupun kalimat yang belum terdaftar dalam KBBI bahkan pelesetan sebuah kata atau kalimat yang terdengar lebih gaul. Penelitian terhadap berbagai dataset berbahasa Indonesia menunjukkan bahwa pemilihan metode klasifikasi yang tepat sangat berpengaruh terhadap performa sistem. Sebagai contoh, penelitian terhadap ulasan produk berbahasa Indonesia menemukan bahwa algoritma *Support Vector Machine* (SVM) mampu memberikan performa klasifikasi yang baik di antara metode-lainnya (Romadhony et al., 2024).

Dominasi penggunaan bahasa gaul atau slang di media sosial menjadi tantangan serius dalam analisis sentimen berbasis teks (Budiasa et al., 2021). mengungkapkan bahwa bentuk slang di media sosial Indonesia sangat produktif dan kreatif, dengan kemunculan kosakata baru seperti *Gelo*, *Gacor*, *Kelas King*, *bucin*, dan berbagai bentuk akronim maupun *clipping* yang tidak terdaftar dalam Kamus Besar Bahasa Indonesia (KBBI). Selain itu, banyak kata mengalami pergeseran makna sehingga berbeda dari arti leksikal aslinya. Kondisi ini berpotensi menimbulkan kesalahan interpretasi dalam sistem analisis sentimen otomatis, karena model klasifikasi berbasis machine learning umumnya mengandalkan representasi numerik dari kata yang terdeteksi secara literal. Jika kata slang tidak dikenali atau tidak dipetakan dengan tepat pada tahap preprocessing dan ekstraksi fitur, maka sentimen yang terkandung dalam teks dapat salah diklasifikasikan. Akibatnya, hasil analisis dapat menjadi bias atau kurang akurat, yang pada akhirnya berpengaruh terhadap validitas insight yang digunakan untuk pengambilan keputusan. Oleh karena itu, penelitian ini menjadi penting untuk memastikan bahwa sistem analisis sentimen yang dikembangkan mampu menangani variasi bahasa informal secara adaptif dan tetap menghasilkan performa klasifikasi yang optimal dalam konteks teks media sosial berbahasa Indonesia.

Salah satu algoritma yang paling sering digunakan dalam tugas klasifikasi sentimen adalah *Support Vector Machine* (SVM). Algoritma ini termasuk dalam kategori *supervised learning* yang bekerja dengan mencari *hyperplane* optimal untuk memisahkan data ke dalam kelas-kelas tertentu dengan margin maksimum. Keunggulan utama SVM terletak pada kemampuannya dalam menangani data berdimensi tinggi, seperti data teks hasil representasi TF-IDF yang dapat menghasilkan ribuan hingga puluhan ribu fitur dalam satu dokumen. Selain itu, SVM dikenal memiliki kemampuan generalisasi yang baik serta relatif tahan terhadap *overfitting*, terutama ketika dikombinasikan dengan pemilihan kernel yang tepat seperti *Radial Basis Function* (RBF). Studi oleh Dea Andini, Rani Laple, dan Taufik Hidayat (2025) yang berjudul “Implementasi Metode Support Vector Machine untuk Analisis Sentimen pada Ulasan Aplikasi Sayurbox di Google Play Store” menunjukkan bahwa metode SVM mampu mencapai tingkat akurasi hingga 91% dalam mengklasifikasikan sentimen pengguna terhadap aplikasi berbahasa Indonesia (Andriati et al., 2025).

Temuan tersebut menguatkan bahwa SVM tetap menjadi metode yang kompetitif dibandingkan algoritma lain seperti *Naïve Bayes* atau *K-Nearest Neighbor* (KNN), khususnya dalam pengolahan teks Bahasa Indonesia yang memiliki kompleksitas morfologi dan variasi

leksikal yang tinggi. Dibandingkan *Naïve Bayes* yang mengasumsikan independensi antar fitur, SVM mampu membentuk batas pemisah yang lebih optimal pada ruang fitur multidimensi. Sementara itu, dibandingkan KNN yang sensitif terhadap distribusi data dan jumlah fitur yang besar, SVM cenderung lebih stabil dalam menghadapi dataset teks berukuran besar. Oleh karena itu, penggunaan SVM dalam penelitian ini menjadi pilihan yang rasional dan relevan untuk menghasilkan model klasifikasi sentimen yang akurat dan memiliki kemampuan generalisasi yang baik terhadap data media sosial berbahasa Indonesia.

Berdasarkan permasalahan yang telah diuraikan, tujuan utama penelitian ini adalah untuk merancang dan mengimplementasikan sebuah sistem analisis sentimen berbasis algoritma *Support Vector Machine* (SVM) yang mampu mengolah data opini publik pada media sosial secara otomatis dan terstruktur. Penelitian ini bertujuan untuk membangun model klasifikasi sentimen yang akurat dalam mengidentifikasi kategori positif, negatif, dan netral dari teks berbahasa Indonesia, termasuk teks yang mengandung bahasa gaul atau slang yang tidak terstandarisasi dalam Kamus Besar Bahasa Indonesia (KBBI). Dengan demikian, diharapkan data opini publik tidak hanya tersimpan, tetapi juga dapat diolah dengan efektif untuk menghasilkan insight yang berguna.

Secara khusus, penelitian ini juga bertujuan untuk mengintegrasikan tahapan pra-pemrosesan teks, ekstraksi fitur menggunakan metode *TF-IDF*, serta optimasi hyperparameter melalui *GridSearch* dan validasi menggunakan *cross validation* guna menghasilkan model yang memiliki performa optimal dan kemampuan generalisasi yang baik. Selain itu, penelitian ini tidak hanya berhenti pada pengujian model secara eksperimental, tetapi juga mengembangkan platform berbasis sistem yang dapat memfasilitasi pengguna dalam melakukan analisis sentimen secara praktis tanpa memerlukan pemahaman teknis mendalam mengenai machine learning.

Dengan demikian, penelitian ini memiliki dua kontribusi utama, yaitu kontribusi metodologis berupa pengujian efektivitas algoritma SVM dalam konteks teks media sosial berbahasa Indonesia yang kaya akan variasi bahasa informal, serta kontribusi aplikatif berupa pengembangan platform analisis sentimen yang mampu mengubah data opini publik menjadi insight yang informatif dan mudah dipahami. Melalui pendekatan ini, diharapkan penelitian dapat menjembatani kesenjangan antara pengembangan model klasifikasi secara akademis dan implementasi sistem yang bermanfaat secara nyata dalam mendukung pengambilan keputusan berbasis data.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan, maka didapatkan rumusan masalah sebagai berikut:

- 1) Bagaimana merancang dan mengembangkan platform analisis sentimen berbasis web yang mampu mengintegrasikan proses preprocessing, ekstraksi fitur, klasifikasi, dan visualisasi hasil dalam satu sistem terstruktur?
- 2) Bagaimana mengimplementasikan algoritma Support Vector Machine (SVM) dengan kernel RBF sebagai metode klasifikasi sentimen dalam platform tersebut?
- 3) Bagaimana tingkat performa model SVM yang diintegrasikan dalam sistem berdasarkan hasil evaluasi menggunakan metrik accuracy, precision, recall, dan F1-score pada dataset ulasan pengguna?

## 1.3 Tujuan

Adapun tujuan yang ingin dicapai sebagai berikut:

- 1) Merancang dan mengembangkan platform analisis sentimen berbasis web yang mengintegrasikan proses preprocessing, ekstraksi fitur, klasifikasi, serta visualisasi hasil dalam satu sistem yang terstruktur.
- 2) Mengimplementasikan algoritma *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF) sebagai metode klasifikasi sentimen ke dalam platform yang dibangun.
- 3) Mengevaluasi dan menganalisis performa model SVM yang diintegrasikan dalam sistem berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score* pada dataset ulasan pengguna.

## 1.4 Manfaat

Adapun manfaat yang dapat diperoleh sebagai berikut:

- 1) **Bagi Penulis** Menerapkan pemahaman teoritis mengenai *Machine Learning* dan *Software Engineering* dalam menghasilkan produk teknologi tepat guna.

- 2) **Bagi Akademisi** Platform ini dapat digunakan sebagai alat bantu penelitian atau sarana pembelajaran praktis mengenai tahapan kerja *Text Mining* tanpa hambatan teknis pemrograman.
- 3) **Bagi Pengguna Umum/Industri** Menyediakan alternatif alat analisis sentimen yang praktis untuk memonitor opini publik atau umpan balik pelanggan secara cepat dan otomatis.

### 1.5 Ruang Lingkup

Berikut ruang lingkup yang digunakan peneliti agar lebih terarah dan sesuai dengan permasalahan, sebagai berikut:

- 1) Objek penelitian berupa data teks opini publik yang mengandung sentimen, seperti ulasan produk, komentar media sosial, dataset publik dari *Kaggle*, serta data yang diunggah oleh pengguna platform.
- 2) Metode klasifikasi utama yang digunakan adalah *Support Vector Machine* (SVM) dengan pendekatan *supervised learning*.
- 3) Kategori yang dihasilkan dibatasi pada dua atau tiga kelas, yaitu positif, negatif, dan/atau netral.

## 1.6 Kebaharuan

Kebaharuan (*Novelty*) dalam penelitian ini terletak pada pengembangan integrasi sistem analisis sentimen yang biasanya hanya dilakukan melalui eksekusi kode (*scripting*) menjadi sebuah **Platform Berbasis Web (GUI)** yang ramah pengguna.

Berdasarkan tinjauan terhadap penelitian terdahulu, sebagian besar penerapan algoritma *Support Vector Machine* (SVM) untuk analisis sentimen, seperti pada penelitian (TARIGAN et al., 2025), berfokus pada hasil akurasi analisis terhadap objek studi kasus semata (seperti Pilpres), namun tidak menghasilkan produk perangkat lunak yang dapat digunakan kembali (*reusable*) oleh pengguna awam.

Penelitian ini mengisi celah tersebut dengan merancang bangun platform yang memadukan alur kerja *Text Mining* lengkap (mulai dari *Preprocessing* data tidak baku/bahasa gaul hingga Klasifikasi SVM) ke dalam satu antarmuka terintegrasi. Hal ini memungkinkan proses analisis sentimen dilakukan secara otomatis tanpa memerlukan keahlian pemrograman, serta mendukung penggunaan data mentah dari berbagai sumber (seperti Excel) secara dinamis.