

BAB I

PENDAHULUAN

Bab ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang akan digunakan dalam dokumen penelitian yang berjudul Analisis Sentimen Ulasan Pengguna Aplikasi DANA di Google Play Store Menggunakan *Support Vector Machine* dan *word embedding FastText*.

1.1 Latar Belakang

DANA merupakan salah satu aplikasi *digital wallet* yang populer di Indonesia untuk memfasilitasi pembayaran, *transfer* uang, dan transaksi keuangan lainnya melalui *smartphone* (Nurcahyo dkk., 2023). Seiring meningkatnya penggunaan DANA, pemahaman terhadap kebutuhan dan ekspektasi pengguna menjadi semakin penting bagi pengembang untuk mempertahankan kualitas layanan yang kompetitif di tengah persaingan yang semakin ketat (Purnama dkk., 2021).

Platform Google Play Store menyediakan fitur ulasan yang memungkinkan pengguna aplikasi untuk memberikan ulasan berupa komentar dan *rating* bintang. Fitur ini menjadi sumber informasi yang penting bagi pengembang aplikasi, meskipun data yang tersedia bersifat tidak terstruktur (Zuo dkk., 2019). Ulasan pengguna dapat memberikan gambaran tentang kepuasan pengguna, masalah yang sering terjadi, serta fitur yang diinginkan. Namun, data yang dihasilkan oleh ulasan pengguna ini sering kali berjumlah besar, tidak terstruktur, dan sulit untuk dianalisis secara manual. Oleh karena itu, dibutuhkan metode yang efektif dan efisien untuk mengidentifikasi dan mengklasifikasikan sentimen pengguna berdasarkan ulasan yang diberikan.

Analisis sentimen telah menjadi alat penting dalam memahami persepsi pengguna terkait produk atau layanan tertentu, termasuk aplikasi keuangan seperti *digital wallet*. Dengan memanfaatkan teknologi *Natural Language Processing* (NLP) dan *text mining*, analisis sentimen dapat mengidentifikasi informasi subjektif dari ulasan pengguna, membantu pengembang dalam membuat keputusan yang lebih baik berdasarkan opini pengguna. Dalam konteks *digital wallet* seperti DANA, analisis ulasan yang diunggah di platform seperti Google Play Store dapat mengungkapkan perasaan positif, negatif, atau

netral pengguna terhadap berbagai fitur aplikasi, pengalaman pengguna, dan masalah yang dihadapi.

Meskipun demikian, tantangan utama dalam analisis sentimen melibatkan ketidakpastian bahasa alami, seperti ironi dan sarkasme, yang dapat mengubah interpretasi sentimen suatu teks. Selain itu, perbedaan gaya bahasa dan penggunaan bahasa informal memperumit proses pengklasifikasian sentimen secara akurat (Wankhade dkk., 2022). Oleh karena itu, metode yang efektif dan inovatif, seperti *machine learning* menjadi penting untuk mengatasi tantangan tersebut, memungkinkan analisis sentimen yang lebih mendalam dan akurat terhadap data ulasan yang tidak terstruktur.

Sudah terdapat beberapa penelitian sebelumnya mengenai analisis sentimen pada ulasan aplikasi DANA, yang pertama penelitian yang dilakukan oleh Muhammad dkk. (2022) menggunakan metode *Support Vector Machine* (SVM) terbukti efektif, di mana SVM dengan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) menghasilkan akurasi hingga 87,58% dalam mengklasifikasi ulasan positif dan negatif pengguna DANA. Hal ini menunjukkan bahwa SVM merupakan algoritma yang kuat untuk klasifikasi data yang besar, terutama ketika dilengkapi dengan seleksi fitur untuk menyederhanakan dimensi data.

Selanjutnya, terdapat juga penelitian yang dilakukan oleh Nurian & Sari (2023) menggunakan metode *Naïve Bayes* dalam analisis sentimen pada ulasan aplikasi DANA yang menghasilkan akurasi sebesar 85% dalam mengklasifikasikan ulasan positif, negatif dan netral. Hasil ini lebih rendah dari metode SVM, sehingga menegaskan bahwa SVM mampu menghasilkan akurasi lebih baik dalam mengklasifikasikan ulasan pengguna.

Selain SVM dan *Naïve Bayes*, metode *Random Forest* juga telah diaplikasikan pada analisis sentimen ulasan DANA. Penelitian yang dilakukan oleh Agustina dkk. (2024) menunjukkan bahwa *Random Forest* mencapai akurasi 84%, ketika pohon keputusan dan kedalamannya dioptimalkan. *Random Forest* memperlihatkan performa yang baik dalam menangani data dengan berbagai karakteristik ulasan pengguna, meskipun akurasinya sedikit di bawah hasil SVM pada klasifikasi terkait ulasan aplikasi DANA.

Penelitian serupa pada konteks yang berbeda juga telah dilakukan di tingkat internasional untuk menilai sentimen respons masyarakat terhadap vaksinasi. Pada

penelitian yang dilakukan oleh Agustina dkk. (2024) menunjukkan bahwa SVM yang dikombinasikan dengan *Word2Vec* dapat meningkatkan pemahaman konteks kata-kata dalam analisis sentimen. Namun, penelitian yang menggunakan kombinasi SVM dengan *Word2Vec* menunjukkan kinerja yang lebih rendah dibandingkan dengan *Term Frequency-Inverse Document Frequency* (TF-IDF), dengan hasil *precision* 83%, *recall* 82%, dan *f1-score* 76% untuk *Word2Vec*, sedangkan TF-IDF menghasilkan *precision* 85%, *recall* 86%, dan *f1-score* 83%.

Performa *Word2Vec* yang lebih rendah disebabkan oleh fokusnya pada makna semantik kata-kata, yang kurang optimal dalam pendekatan SVM yang lebih linear dan lebih efektif dengan fitur eksplisit seperti yang dihasilkan oleh TF-IDF. Kondisi ini mengakibatkan banyak prediksi yang tidak sesuai dengan label sebenarnya, terutama pada data yang tidak seimbang.

Melihat kelemahan performa *Word2Vec* dalam kombinasi dengan SVM, penelitian ini bertujuan untuk mengombinasikan SVM dengan *FastText* sebagai metode *feature extraction* alternatif. *FastText* menawarkan beberapa keunggulan signifikan yang diharapkan dapat meningkatkan akurasi analisis sentimen. Pada penelitian yang dilakukan oleh Khasanah (2021) menunjukkan bahwa penggunaan *embedding FastText* dapat meningkatkan kinerja model klasifikasi sentimen secara signifikan. Dalam studi ini, digunakan dua dataset utama yaitu *Stanford Sentiment Treebank 2* (SST2) yang berisi ulasan singkat dengan label sentimen, serta dataset *Movie Reviews* (MR) yang berisi ulasan film berlabel positif dan negatif. Hasil menunjukkan bahwa model *FastText* + CNN mencapai akurasi tertinggi dengan 84% pada dataset SST2 dan 80% pada dataset MR.

Keunggulan ini diperoleh karena *FastText* memiliki kemampuan menangani kata-kata tidak dikenal (*out-of-vocabulary*) dan kata yang jarang muncul dengan menggunakan informasi sub-kata, sehingga mampu menangkap variasi morfologis dan kontekstual yang lebih kompleks. Selain itu, penelitian ini juga menunjukkan bahwa model *FastText* + *Bidirectional Gated Recurrent Unit* (BiGRU) dan *FastText* + *Bidirectional Long Short-Term Memory* (BiLSTM) memberikan hasil yang kompetitif dibandingkan dengan baseline model yang menggunakan *embedding GloVe*. Penelitian ini menegaskan bahwa *embedding FastText* dapat meningkatkan performa model *deep learning* sederhana dengan arsitektur yang efisien.

Dalam konteks analisis sentimen yang melibatkan berbagai ekspresi emosional, keakuratan representasi kata dari *FastText* diharapkan dapat mendukung SVM dalam menghasilkan klasifikasi yang lebih akurat dibandingkan *Word2Vec*. Kombinasi SVM dan *FastText* diharapkan tidak hanya mempertahankan pemahaman kontekstual yang kuat tetapi juga meningkatkan kemampuan klasifikasi. Pendekatan ini dinilai kompetitif dalam analisis sentimen berbasis teks sekaligus efisien dalam hal komputasi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, dirumuskan permasalahan yang dibahas pada skripsi ini yaitu bagaimana melakukan analisis sentimen ulasan aplikasi DANA pada Google Play Store dengan metode *Support Vector Machine* dan *word embedding FastText*. Dari permasalahan tersebut diturunkan ke dalam permasalahan khusus yaitu:

1. Bagaimana cara melakukan analisis sentimen ulasan aplikasi DANA pada Google Play Store menggunakan *Support Vector Machine* yang dikombinasikan dengan *word embedding FastText*?
2. Bagaimana hasil pengujian dari pemilihan kernel terbaik metode *Support Vector Machine* dalam menghasilkan kinerja terbaik untuk analisis sentimen ulasan aplikasi DANA pada Google Play Store?

1.3 Tujuan dan Manfaat

Tujuan penelitian ini adalah untuk menganalisis sentimen terhadap ulasan aplikasi DANA pada Google Play Store. Secara khusus, penelitian ini bertujuan mengembangkan model klasifikasi sentimen yang dapat mengelompokkan ulasan pengguna ke dalam kategori positif dan negatif secara otomatis, serta mencapai tingkat akurasi terbaik melalui kombinasi metode *Support Vector Machine* (SVM) dengan *word embedding FastText*.

Manfaat dari penelitian ini diharapkan dapat menjadi bahan evaluasi bagi perusahaan yang mengelola aplikasi DANA untuk meningkatkan kualitas layanan dan pengalaman pengguna. Dengan memahami sentimen pengguna, perusahaan dapat lebih mudah mengidentifikasi aspek layanan yang perlu diperbaiki atau dikembangkan. Selain itu, penelitian ini juga diharapkan dapat menjadi referensi bagi penelitian-penelitian selanjutnya

yang berfokus pada analisis sentimen dalam industri *fintech* atau yang menggunakan metode SVM dan *word embedding FastText* untuk analisis sentimen.

1.4 Ruang Lingkup

Penelitian ini diperlukan adanya ruang lingkup agar penelitian yang dilakukan lebih terarah dan mencapai sasaran yang diharapkan. Ruang lingkup dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan sebagai dasar dalam penelitian ini adalah data ulasan aplikasi DANA yang diambil dari platform Google Play Store. Proses pengambilan data dilakukan dengan metode *scraping* menggunakan *library google-play-scraper*. Data yang diambil berjumlah 10.000 ulasan terbaru dengan proses *scraping* dilakukan pada tanggal 23 September 2024.
2. Analisis sentimen pada data dilakukan dengan klasifikasi sentimen ke dalam 2 kelas yaitu positif dan negatif.
3. Data dilabelisasi menggunakan model berbasis *Robustly Optimized BERT Pre-training Approach* (RoBERTa) khusus untuk Bahasa Indonesia dari Hugging Face.
4. Proses implementasi dilakukan menggunakan bahasa pemrograman Python dengan perangkat lunak Google Colaborator.

1.5 Sistematika Penulisan

Sistematika penulisan memberikan gambaran laporan dari skripsi ini secara urut dan jelas. Berikut adalah beberapa pokok bahasan dalam skripsi ini, yaitu:

BAB I PENDAHULUAN

Bab ini menyajikan latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan skripsi analisis sentimen ulasan pengguna aplikasi DANA pada Google Play Store menggunakan metode *Support Vector Machine* dan *word embedding FastText*.

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan penelitian-penelitian terkait dan dasar teori yang digunakan dalam skripsi analisis sentimen ulasan pengguna aplikasi DANA pada Google Play Store menggunakan metode *Support Vector Machine* dan *word embedding FastText*.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metodologi yang digunakan dalam skripsi analisis sentimen ulasan pengguna aplikasi DANA pada Google Play Store menggunakan metode *Support Vector Machine* dan *word embedding FastText*. Bab ini membahas alur penelitian, pengumpulan data melalui teknik *web scraping*, proses labelisasi data secara otomatis menggunakan model khusus Bahasa Indonesia, *preprocessing* data, pembagian data serta proses klasifikasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan mengenai analisis dan hasil penelitian analisis sentimen ulasan pengguna aplikasi DANA pada Google Play Store menggunakan metode *Support Vector Machine* dan *word embedding FastText*. Bab ini membahas tentang implementasi pelatihan dan skenario pengujian menggunakan 3 jenis kernel yaitu kernel linear, kernel RBF, dan kernel *Polynomial*.

BAB V PENUTUP

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan, serta saran untuk pengembangan dan penelitian lebih lanjut.