

BAB I

PENDAHULUAN

Bab pendahuluan ini membahas mengenai latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan pada skripsi yang berjudul analisis sentimen pada ulasan pengguna aplikasi Character AI menggunakan seleksi fitur *Chi-Square* dan metode *Support Vector Machine* (SVM).

1.1 Latar Belakang

Perkembangan *Artificial Intelligence* (AI) telah mengalami kemajuan pesat yang membawa berbagai inovasi signifikan dalam kehidupan manusia. Salah satu kemajuan AI yang terkini yaitu *Generative AI*, yakni sistem kecerdasan buatan yang memiliki kemampuan untuk meniru ciptaan manusia, seperti gambar, musik, atau bentuk media lainnya melalui input teks ("*prompts*") (Sengar dkk., 2024). *Generative AI* berkembang dengan cepat berkat kemunculan *Large Language Models* (LLMs), yaitu model yang dilatih dengan data berukuran besar dan mampu memahami bahasa alami manusia. LLMs memungkinkan *Generative AI* menghasilkan respons yang kontekstual dan kompleks. Salah satu bentuk implementasi *Generative AI* dengan LLMs adalah *AI Companion* (Brown dkk., 2020).

AI Companion adalah aplikasi yang memanfaatkan *Artificial Intelligence* (AI) untuk menawarkan *user* terlibat dalam sebuah interaksi sosial dan emosional yang mencakup berbagai interaksi antara manusia dan komputer, baik dalam bentuk *chatbot*, avatar, maupun robot fisik (De Freitas dkk., 2024). Salah satu *AI Companion* yang mendapatkan banyak perhatian dari pengguna, yaitu aplikasi Character AI (bisa disebut sebagai c.ai atau Character.ai). Aplikasi ini memungkinkan pengguna untuk menciptakan dan berinteraksi dengan karakter virtual yang dapat disesuaikan, seperti tokoh publik, karakter fiksi, hingga selebriti. Aplikasi ini tidak hanya menawarkan percakapan berbasis teks yang kompleks, tetapi juga mencerminkan bagaimana pengguna berinteraksi dengan karakter yang memiliki kepribadian sesuai preferensi masing-masing pengguna (Character.ai, 2022). Karena keunikannya tersebut, hingga Agustus 2024, Character AI telah diunduh lebih dari 10 juta kali dan menerima lebih dari 1,11 juta ulasan di Google Play Store.

Ulasan pengguna di *platform* seperti Google Play Store dapat menjadi sumber informasi berharga yang mencerminkan pengalaman dan sentimen pengguna terhadap suatu teknologi (Putri dkk., 2024). Melihat tingginya antusiasme terhadap aplikasi Character AI, penting untuk mengetahui bagaimana persepsi pengguna terhadap aplikasi tersebut, karena banyaknya ulasan pengguna tidak selalu mencerminkan kepuasan secara menyeluruh. Ulasan yang tampak baik dari segi kuantitas belum tentu mengandung sentimen positif secara isi. Selain itu, aplikasi yang melibatkan interaksi emosional seperti Character AI sangat rentan terhadap kritik atau ekspektasi tinggi dari pengguna. Jadi memahami opini secara tekstual akan membantu pihak pengembang aplikasi dalam mengidentifikasi kelebihan, kekurangan, serta potensi perbaikan yang dibutuhkan untuk pengembangan aplikasi di masa depan (Luthfiana dkk., 2020). Oleh karena itu, penelitian ini akan melakukan analisis sentimen untuk memahami dan menganalisis persepsi pengguna terhadap aplikasi Character AI.

Analisis sentimen adalah bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi seseorang terhadap suatu entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya (Liu, 2012). Analisis sentimen merupakan salah satu metode yang efektif untuk mengukur dan memahami pandangan publik mengenai suatu topik. Melalui analisis sentimen, persepsi pengguna terhadap aplikasi Character AI dapat diketahui, apakah lebih cenderung optimis atau pesimis.

Dalam melakukan analisis sentimen, diperlukan algoritma *classifier* yang mendukung agar didapatkan hasil akurasi yang maksimal. Salah satu algoritma klasifikasi yang sering digunakan untuk melakukan analisis sentimen adalah algoritma *Support Vector Machine* (SVM). Prinsip dasar SVM secara sederhana dapat diartikan sebagai upaya menemukan *hyperplane* terbaik untuk memisahkan dua kelas, yaitu kelas positif dan kelas negatif (Nurkholis dkk., 2022). Algoritma SVM dapat bekerja dengan sangat baik untuk data yang bersifat linear, namun kenyataannya banyak *dataset* yang bersifat non-linear dan kompleks. Untuk mengatasi masalah tersebut, SVM memiliki beberapa fungsi kernel yang bisa digunakan untuk mengatasi masalah non-linear, dalam kata lain, kelas yang tidak bisa dipisahkan hanya dengan garis lurus. Kernel linear, kernel *polynomial*, dan kernel *Radial Basis Function* (RBF) adalah beberapa fungsi kernel yang umum untuk digunakan pada SVM. Penggunaan jenis kernel yang berbeda juga akan menghasilkan perbedaan performa dalam proses analisis sentimen yang dilakukan (Rochim dkk., 2021).

Beberapa penelitian terdahulu yang membahas mengenai penggunaan fungsi kernel SVM yaitu penelitian yang dilakukan oleh Nurkholis dkk. (2022), yang membandingkan dua fungsi kernel, yaitu kernel RBF dan kernel *polynomial* untuk menganalisis *multi-class* sentimen pengguna Twitter terhadap PPKM di Indonesia, mendapatkan hasil bahwa kernel *polynomial* lebih unggul dengan nilai akurasi sebesar 98,9%. Penelitian Rochim dkk. (2021) membandingkan performa dari tiga fungsi kernel yaitu kernel linear, kernel RBF, dan kernel *polynomial* untuk mengklasifikasikan sentimen pengguna Twitter terhadap COVID-19. Hasil yang diperoleh yaitu kernel *polynomial* lebih unggul dari nilai akurasi yaitu sebesar 91%. Husni dkk. (2023) juga melakukan penelitian untuk klasifikasi sentimen publik terhadap pembelajaran daring menggunakan tiga fungsi kernel SVM yaitu kernel linear, kernel RBF, dan kernel *polynomial*. Hasil dari penelitian ini adalah kernel linear lebih unggul dari dua fungsi kernel lainnya dengan memperoleh akurasi sebesar 84%.

Support Vector Machine (SVM) dengan fungsi kernel telah banyak digunakan dalam penelitian analisis sentimen karena menghasilkan akurasi yang baik seperti yang sudah dijelaskan pada penelitian-penelitian sebelumnya. Namun, dalam klasifikasi teks, *dataset* yang diperoleh seringkali memiliki dimensi yang besar dan fitur yang banyak. Oleh karena itu, seleksi fitur dapat dilakukan untuk mengurangi dimensi data atau fitur guna meningkatkan performa algoritma klasifikasi (Mustamu & Sibaroni, 2023). Hokijulandy dkk. (2023) melakukan penelitian uji kernel linear dengan seleksi fitur *Chi-Square* dalam analisis sentimen terhadap aplikasi *mobile* JKN, dan memperoleh *F1-score* sebesar 95,09% menjadi 96,43%. Penggunaan seleksi fitur *Chi-Square* untuk reduksi dimensi fitur terbukti dapat meningkatkan performa model klasifikasi (Hokijulandy dkk., 2023; Winata dkk., 2022). Prastyo dkk. (2020) melakukan penelitian menggunakan empat fungsi kernel SVM yaitu kernel linear, RBF, *polynomial*, dan *sigmoid* dengan TF-IDF untuk analisis sentimen pengguna Twitter terhadap kebijakan *Omnibus Law* di Indonesia. TF-IDF pada penelitian ini digunakan untuk ekstraksi fitur dan seleksi fitur. Hasil yang didapat yaitu fungsi kernel RBF memiliki nilai akurasi tertinggi yaitu 96,61% dengan 2000 fitur. Penelitian serupa dilakukan oleh Putri dkk. (2024) yang membandingkan empat fungsi kernel SVM dalam melakukan analisis sentimen terhadap *Chatbot* AI pada ulasan Google Play Store dengan hasil kernel linear memiliki nilai akurasi tertinggi sebesar 92,68%.

Setelah melakukan analisis terhadap permasalahan yang ada menggunakan studi pustaka terhadap penelitian sebelumnya, penulis memutuskan untuk menerapkan beberapa hal dari penelitian sebelumnya untuk diterapkan dalam penelitian penulis. Beberapa hal yang diterapkan yaitu penggunaan algoritma klasifikasi *Support Vector Machine* (SVM) dengan tiga fungsi kernel, yaitu kernel linear, kernel *Radial Basis Function* (RBF), dan kernel *polynomial*. Selain itu, penulis juga menggunakan TF-IDF untuk ekstraksi fitur dan *Chi-Square* untuk seleksi fitur.

Berdasarkan permasalahan yang diuraikan pada latar belakang di atas, penulis bermaksud untuk melakukan penelitian mengenai analisis sentimen pada ulasan pengguna aplikasi Character AI menggunakan seleksi fitur *Chi-Square* dan metode *Support Vector Machine* (SVM).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya, maka dapat dibuat rumusan masalah yaitu bagaimana menerapkan seleksi fitur *Chi-Square* dan *Support Vector Machine* (SVM) untuk analisis sentimen terhadap ulasan pengguna aplikasi Character AI.

1.3 Tujuan dan Manfaat

Tujuan dari penelitian ini adalah untuk menerapkan metode seleksi fitur *Chi-Square* dan algoritma *Support Vector Machine* (SVM) dalam melakukan analisis sentimen terhadap ulasan pengguna aplikasi Character AI, guna mengklasifikasikan sentimen ke dalam kelas positif dan negatif.

Penelitian ini diharapkan dapat memberikan manfaat, meliputi berbagai aspek yang mendukung pengembangan pengetahuan dalam melakukan analisis sentimen. Adapun manfaat spesifik dari penelitian ini adalah sebagai berikut:

1. Bagi penulis, untuk eksplorasi teori-teori yang dipelajari, serta menambah wawasan, ilmu pengetahuan, dan pengalaman dalam melakukan analisis sentimen.
2. Bagi pembaca, untuk memberikan informasi mengenai analisis sentimen terhadap ulasan pengguna aplikasi Character AI yang bisa digunakan sebagai referensi penelitian selanjutnya, terutama dalam penelitian mengenai seleksi fungsi *Chi-Square* dan fungsi kernel *Support Vector Machine* (SVM).

1.4 Ruang Lingkup

Penelitian ini diperlukan adanya ruang lingkup agar pembahasan terfokus dan lebih terarah berdasarkan perumusan masalah di atas dan sesuai dengan batasan kemampuan penulis untuk dapat menyelesaikan permasalahan, maka ruang lingkup dan batasan masalah dalam penelitian ini yaitu sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah data dari ulasan pengguna aplikasi Character AI yang dikumpulkan dengan cara *scraping* dari Google Play Store menggunakan *library google-play-scraper*. Data yang diambil hanya ulasan yang menggunakan bahasa Indonesia. Emoji, angka, dan idiom dalam ulasan akan diabaikan atau tidak diolah.
2. Data teks yang telah diperoleh akan melalui tahap *preprocessing*, dimana data teks akan melalui proses *case folding*, *cleaning*, normalisasi, tokenisasi, *filtering*, dan *stemming*.
3. Data teks yang diambil akan diberi label dengan klasifikasi biner (*dual-class*), yaitu sentimen positif dan sentimen negatif menggunakan *lexicon dictionary-based approach SentiStrength* Indonesia.
4. Ekstraksi fitur menggunakan TF-IDF, dan seleksi fitur menggunakan *Chi-Square* dengan taraf signifikansi (α) sebesar 0,3.
5. Pengolahan data analisis sentimen menggunakan bahasa pemrograman Python, perangkat lunak Google Colaborator, dan metode *Support Vector Machine* (SVM) dengan 3 fungsi kernel yaitu linear, *Radial Basis Function* (RBF), dan *polynomial*.
6. Proses *hyperparameter tuning* dilakukan dengan *Grid Search Cross Validation* pada 5 *subset* data ($K = 5$) untuk mencari kombinasi parameter terbaik, yaitu kernel linear dengan parameter C , kernel RBF dengan parameter C dan *gamma* (γ), serta kernel *polynomial* dengan parameter C dan *degree* (d).
7. Evaluasi model menggunakan *confusion matrix*, dengan nilai yang diperoleh akan digunakan untuk menghitung metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*, serta *misclassification rate*.

1.5 Sistematika Penulisan

Penulisan laporan ini terdiri dari lima bab untuk memberikan gambaran mengenai penyusunan penelitian tentang analisis sentimen pada ulasan pengguna aplikasi Character AI menggunakan seleksi fitur *Chi-Square* dan metode *Support Vector Machine* (SVM).

BAB I PENDAHULUAN

Bab ini berisikan latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika penulisan mengenai analisis sentimen pada ulasan pengguna aplikasi Character AI menggunakan seleksi fitur *Chi-Square* dan metode *Support Vector Machine* (SVM).

BAB II LANDASAN TEORI

Bab ini memuat tinjauan pustaka dan dasar teori yang digunakan untuk mendukung penelitian mengenai analisis sentimen pada ulasan pengguna aplikasi Character AI menggunakan seleksi fitur *Chi-Square* dan metode *Support Vector Machine* (SVM).

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metode yang digunakan dalam penelitian analisis sentimen pada ulasan pengguna aplikasi Character AI menggunakan seleksi fitur *Chi-Square* dan metode *Support Vector Machine* (SVM). Bab ini membahas tentang alur penelitian, seperti pengumpulan data, prapemrosesan data, pelabelan data, ekstraksi dan seleksi fitur, serta klasifikasi sentimen.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menggambarkan pembahasan hasil penelitian analisis sentimen pada ulasan pengguna aplikasi Character AI menggunakan seleksi fitur *Chi-Square* dan metode *Support Vector Machine* (SVM). Bab ini membahas tentang hasil implementasi dari pengumpulan data, prapemrosesan data, pelabelan data, ekstraksi dan seleksi fitur, dan klasifikasi sentimen dengan skenario pengujian menggunakan 3 fungsi kernel SVM, yaitu kernel linear, kernel *Radial Basis Function* (RBF), dan kernel *polynomial*.

BAB V PENUTUP

Bab ini berisikan kesimpulan dan saran dari penelitian mengenai analisis sentimen pada ulasan pengguna aplikasi Character AI menggunakan seleksi fitur *Chi-Square* dan metode *Support Vector Machine* (SVM)