

BAB I

PENDAHULUAN

1.1. Latar Belakang

Seiring dengan kemajuan zaman, teknologi kecerdasan buatan telah menjadi bagian tak terpisahkan dari berbagai aktivitas. Dalam beberapa tahun terakhir, penggunaan kecerdasan buatan untuk membantu berbagai tugas dan pekerjaan semakin populer [1]. Kemunculan platform media sosial telah mengubahnya menjadi arena utama bagi publik untuk berdiskusi dan memberikan opini mengenai berbagai inovasi teknologi, termasuk kecerdasan buatan.

Salah satu model kecerdasan buatan yang menjadi topik hangat diskusi adalah DeepSeek AI, yang dikembangkan untuk memahami, menghasilkan, dan menganalisis teks, gambar, dan file dengan akurasi tinggi. Popularitasnya memicu lahirnya beragam komentar dan ulasan dari pengguna di platform seperti Instagram. Namun, volume opini yang besar ini masih bersifat tidak terstruktur, sehingga diperlukan analisis sistematis untuk memahami bagaimana persepsi publik sesungguhnya terhadap teknologi ini.

Untuk dapat mengitung sentimen atau opini masyarakat, digunakan metode analisis sentimen. Analisis sentimen merupakan salah satu cabang dari *text mining* yang melakukan sebuah proses dari klasifikasi pada dokumen teks. Metode ini digunakan untuk menentukan pendapat tentang masalah, apakah permasalahan tersebut mempunyai kecenderungan ke arah yang positif atau ke arah yang negatif [2]. Untuk penelitian ini, metode yang digunakan yaitu metode Decision Tree dengan algoritma C4.5, yang keunggulannya telah dibuktikan dalam berbagai riset sebelumnya.

Beberapa penelitian terdahulu secara konsisten menunjukkan keunggulan algoritma Decision Tree C4.5. Kajian oleh Devi Yunita dan Ines Heidiani Ikasari [3] memperlihatkan seberapa unggulnya algoritma C4.5 saat dibandingkan dengan Naïve Bayes. Dalam studi tersebut, metode C4.5 berhasil mencapai tingkat akurasi 94,17%, jauh melampaui Naïve Bayes yang hanya mencatatkan 85,83%. Temuan serupa juga dilaporkan oleh Sriani, Suhardi, dan La Saufa Yardha [4] dalam analisis sentimen untuk aplikasi Mobile JKN, di

mana algoritma C4.5 lebih efektif dengan perolehan akurasi 79,75% dibandingkan Naïve Bayes Classifier. Dukungan lebih lanjut datang dari riset oleh Rachmat Hidayat, Abdul Barir Hakim, dan Risman Nugraha [5] yang menguji kedua metode pada sentimen produk Es Teh Indonesia. Hasilnya sekali lagi membuktikan Decision Tree C4.5 sebagai metode yang lebih unggul, dengan tingkat akurasi mencapai 71,96% dibandingkan 66,11% yang didapat oleh Naïve Bayes.

Dari hasil beberapa penelitian sebelumnya didapatkan bahwa Decision tree metode C4.5 memiliki hasil yang lebih baik dan akurat. Oleh karena itu, algoritma C4.5 Decision Tree dipilih sebagai metode penelitian dengan target postingan instagram yang membahas mengenai DeepSeek AI dengan judul Analisis Sentimen Ulasan Pengguna Terhadap Postingan Deepseek Di Instagram Menggunakan Metode Decision Tree.

1.2. Rumusan Masalah

Berdasarkan pada uraian latar belakang maka dapat dipaparkan rumusan masalah sebagai berikut:

1.2.1. Permasalahan

Menganalisis *komentar* pada *Instagram* untuk mengetahui komentar masyarakat terhadap postingan yang membahas Deepseek AI di *Instagram* berupa sentimen *positif*, *negatif*, dan *netral*.

1.2.2. Solusi Permasalahan

Berdasarkan masalah tersebut, peneliti menggunakan metode klasifikasi algoritma *C4.5 Decision Tree* pada klasifikasi komentar dengan harapan mendapat nilai akurasi yang optimal terhadap klasifikasi analisis sentiment social media Instagram terhadap postingan yang membahas Deepseek AI.

1.2.3. Pertanyaan Penelitian

1. Berapa nilai akurasi yang diperoleh dengan melakukan percobaan parameter *max_depth* yang berbeda pada *Decision Tree C4.5*?
2. Bagaimana pengaruh jumlah pembagian data dengan ratio yang berbeda terhadap nilai akurasi yang didapat?

3. Bagaimana perbandingan kinerja model klasifikasi berdasarkan variasi metode pelabelan, nilai threshold (0.1 dan -0.1, 0.2 dan -0.2, 0.3 dan -0.3), serta penggunaan hyperparameter tuning dan tanpa tuning dalam menghasilkan akurasi tinggi dan meminimalkan overfitting?

1.3. Batasan Masalah

Dari pembahasan sebelumnya, penulis mendapatkan hasil untuk di gunakan ruang lingkup penelitian sehingga menghasilkan output yang diinginkan dan optimal dengan Batasan yang diterapkan sebagai berikut :

1. Data yang digunakan penelitian hanya berasal dari Instagram
2. Data yang diperoleh berasal dari komentar pada postingan yang mengandung unsur Deepseek AI di Instagram.
3. Komentar yang digunakan pada penelitian ini hanya menggunakan Bahasa Inggris.
4. Postingan yang digunakan pada penelitian ini di mulai dari bulan Januari 2025 sampai Juni 2025.
5. Tidak ada batasan berapa banyak komentar yang akan digunakan pada setiap postingan yang mengandung unsur Deepseek AI di Instagram.
6. Pelabelan sentiment dilakukan menggunakan metode *lexicon-based* yaitu *TextBlob*.
7. Label dibatasi menjadi 3 kelas yaitu *positif, negatif, netral*.
8. Model yang digunakan adalah *Decision Tree (C4.5 berbasis entropy)*.
9. Tidak membandingkan dengan model lain sebagai focus utama penelitian.
10. Sistem dirancang menggunakan Bahasa pemrograman python dengan bantuan *library WordNet* dan *Lemma*.
11. Data *Scraping komentar* menggunakan *tools IG Exporter & Scraper* dari *Chrome Web Store*
12. Klasifikasi pohon keputusan dibangun menggunakan *library* dari *skitch learn decisiontreeClassifier*.
13. Langkah *stopwords* menggunakan *library* dan *nlk* berbahasa Inggris, pada penelitian ini memodifikasi list *stopwords* pada *library bltk* sesuai dengan kasus yang dilakukan. Daftar stopword dimodifikasi dengan mengecualikan

beberapa kata tanya penting seperti "who", "what", "where", "when", "why", dan "how

1.4. Manfaat dan Tujuan Penelitian

1.4.1. Manfaat

Diharapkan memberikan manfaat yang berguna bagi masyarakat tentang gambaran sentiment terhadap komentar pengguna Instagram mengenai postingan Deepseek AI.

1.4.2. Tujuan

Berdasarkan tujuan masalah tersebut, tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui nilai akurasi yang diperoleh ketika melakukan scenario yang berbeda pada *max_depth C4.5 Decision Tree*
2. Untuk mengetahui nilai akurasi yang didapat dari pembagian data dengan ratio yang berbeda.
3. Untuk mengetahui scenario terbaik dari kinerja model klasifikasi berdasarkan variasi metode pelabelan, nilai threshold (0.1 dan -0.1, 0.2 dan -0.2, 0.3 dan -0.3), serta penggunaan hyperparameter tuning dan tanpa tuning dalam menghasilkan akurasi tinggi dan meminimalkan overfitting

1.5. Sistematika Proposal

Struktur penulisan proposal tugas akhir dirangkai seperti berikut:

BAB I PENDAHULUAN

Bab ini membahas latar belakang mengenai ruang lingkup masalah yang dibahas, solusi yang diusulkan untuk menyelesaikan permasalahan, tujuan serta manfaat dengan diterapkannya solusi, dan sistematika penulisan proposal.

BAB II KAJIAN PUSTAKA

Bab ini akan menjelaskan Mengenai materi dan metode yang dijadikan landasan teori dalam penelitian ini. Diantaranya materi yang dibahas adalah algoritma Decision Tree C4.5. dan penelitian sebelumnya terkait analisis sentimen.

BAB III METODE PENELITIAN

Pada bab ini dijelaskan tentang dataset yang akan digunakan , alur sistem, tahapan pengujian, scenario uji coba dan jadwal pengerjaan.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan tentang hasil dari penelitian terkait dengan menjelaskan penelitian berlangsung dan mendapatkan hasil pada penelitian

BAB V PENUTUP

Pada bab ini dijelaskan tentang kesimpulan pada penelitian yang menghasilkan beberapa kesimpulan pada penelitian ini dan saran yang dapat dilakukan dan dikembangkan untuk peneliti kali ini.