

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi informasi yang semakin pesat telah membawa perubahan signifikan dalam cara manusia memperoleh informasi, berkomunikasi, dan menjalankan aktivitas sehari-hari. Kemajuan di bidang kecerdasan buatan (AI) menjadi salah satu inovasi paling berpengaruh, terutama karena kemampuannya dalam mengolah data dalam jumlah besar serta mendukung berbagai kegiatan seperti pendidikan, administrasi, dan layanan publik. Salah satu bentuk teknologi AI yang menarik perhatian global adalah model bahasa generatif yang dapat menanggapi permintaan pengguna secara interaktif. Kehadiran generatif AI sejak tahun 2023 mendorong percepatan adopsi AI secara masif, karena teknologi ini lebih mudah diimplementasikan tanpa memerlukan keahlian teknis mendalam dan mampu menekan biaya penggunaan. OECD mencatat bahwa pada tahun 2024 lebih dari 13% perusahaan di wilayah EU27 telah mengadopsi solusi AI, dengan tingkat pemanfaatan yang meningkat hampir dua kali lipat dibandingkan tahun sebelumnya, menunjukkan bahwa AI menjadi pendorong penting dalam peningkatan efisiensi, pengambilan keputusan, serta daya saing berbagai sektor industri (Kergroach & Héritier, 2025).

ChatGPT dirilis ke publik pada 30 November 2022 sebagai *research preview* dan sejak awal menunjukkan adopsi massal lintas konteks penggunaan (percakapan, penulisan, analitik). Tanggal rilis ini dikonfirmasi oleh pengumuman resmi pengembang. Secara sosial, kemunculan *ChatGPT* memicu adopsi cepat di pendidikan tinggi, ditopang oleh persepsi kegunaan (*performance expectancy*), kemudahan penggunaan (*effort expectancy*), dan pengaruh sosial. Sejumlah tinjauan dan studi empiris menemukan faktor-faktor tersebut berperan dalam niat dan penggunaan nyata, sekaligus menyoroti hambatan seperti kecemasan dan kekhawatiran integritas akademik (Albadarin et al., 2024).

Di dunia kerja dan produktivitas, bukti eksperimental menunjukkan efek yang substansial: uji terkontrol pada tugas penulisan profesional menemukan waktu penyelesaian turun ~40% dan kualitas output naik ~18% ketika peserta menggunakan ChatGPT. Temuan ini memperkuat narasi bahwa model generatif

mempercepat produksi konten dan menaikkan performa, terutama bagi pekerja dengan produktivitas awal lebih rendah. Ringkasan kebijakan mutakhir juga menyimpulkan bukti bahwa generatif AI menaikkan efisiensi dan mengubah proses bisnis (Noy & Zhang, 2023).

Secara makro, tinjauan sistematis lintas sektor (kesehatan, pendidikan, kerja, media sosial, dsb.) mendeskripsikan manfaat (efisiensi, akses informasi) dan tantangan (bias, transparansi, kesenjangan literasi) sebagai dampak sosial utama dari adopsi ChatGPT/LLM sejak 2022–2023. Ruang lingkup pengetahuan ini terus bertambah dan menekankan bahwa efek sosialnya bersifat berlapis: peningkatan produktivitas dan akses disertai kebutuhan tata kelola dan literasi yang lebih kuat (Hussein et al., 2025). Di ekosistem akademik, pembahasan tentang integritas akademik dan etika meningkat tajam. Kajian di *Teaching in Higher Education dan Computers & Education* menekankan peluang (umpan balik cepat, personalisasi) sekaligus risiko (plagiarisme/automasi tugas tanpa pemahaman), sehingga mendorong penyesuaian kebijakan penugasan, asesmen, dan literasi AI (Cotton et al., 2024).

Dalam kehidupan sosial, kita sering mengungkapkan sentimen / pendapat kepada orang lain telah menjadi kebiasaan yang kita lakukan setiap hari. Sentimen menurut KBBI adalah pendapat atau pandangan yang didasarkan pada perasaan yang berlebih-lebihan terhadap sesuatu (bertentangan dengan pertimbangan pikiran). Seiring dengan perkembangan teknologi dan internet, semakin mudah bagi masyarakat untuk mengungkapkan opini dan pendapatnya dalam berbagai bentuk seperti unggahan di media sosial, penilaian atau ulasan produk, artikel berita, maupun transkrip wawancara. Dari opini dan pendapat tersebut terkandung banyak informasi berharga tentang emosi dan pandangan manusia.

Twitter menjadi salah satu platform populer bagi masyarakat Indonesia untuk menyampaikan opini. Pada Januari 2022, jumlah pengguna *Twitter* di Indonesia tercatat mencapai 18,45 juta, menjadikannya peringkat kelima terbanyak di dunia (Harahap et al., 2024). Dengan jutaan pengguna aktif harian, *Twitter* menghasilkan volume data yang besar dan bersifat publik, sehingga dapat dimanfaatkan secara legal sebagai sumber data representatif dalam analisis sentimen maupun penelitian sosial (Cholid & Wahyu Mubarak, 2024). *Twitter*

memungkinkan penggunaanya berkirim dan membaca pesan yang lebih dari 280 karakter yang disebut tweet. Sebelumnya, pesan di Twitter hanya sampai 140 karakter tetapi pada tanggal 7 November 2017 ditambah menjadi 280 karakter) (Gligoric et al., 2022). Pada tanggal 31 Juli 2023 *Twitter* melakukan rebranding dengan mengubah nama menjadi *X* putih dengan latar belakang hitam. Ini merupakan bagian dari upaya rebranding yang lebih luas yang dipimpin oleh Elon Musk (Lee et al., 2023). Selasa (1/8/2023), istilah *tweet* untuk utas yang diunggah diganti dengan post. Selain itu, fitur yang sebelumnya dikenal sebagai retweet kini telah diubah namanya menjadi *repost* (Rostanti, 2023). *OpenAI* merilis model *chatbot*, *ChatGPT*, yang dapat berinteraksi dengan manusia melalui percakapan. Model *chatbot* dioptimalkan dengan Teknik pembelajaran pengawasan dan penguatan. *ChatGPT* menggunakan teknik kecerdasan buatan dengan pembelajaran mendalam untuk menghasilkan respon seperti manusia terhadap pertanyaan bahasa alami, menjadikannya model bahasa yang sangat besar (Roumeliotis & Tselikas, 2023).

Perkembangan *ChatGPT* ini memberikan dampak positif dan negatif. Salah satu dampak positifnya adalah bisa mengembangkan berbagai keterampilan. Dampak dari hal tersebut ialah adanya kemungkinan pekerjaan yang sifatnya rutin seperti programmer, penerjemah bahasa, dan pekerjaan yang berkaitan dengan administrasi. Perkembangan ini menuai banyak tanggapan di sosial seperti *Facebook*, *Instagram*, *Youtube*, *Twitter*.

Beberapa user di *Youtube* memberikan komen pada sebuah video *Youtube* berjudul 'Bagaimana *ChatGPT* Secara Perlahan Menghancurkan Otak Anda' oleh seorang Youtuber Justin Sung pada tanggal 3 Agustus 2025. Salah satunya dari user dengan nama @johnex2154 memberikan komentar (dengan terjemahan) 'Saya secara pribadi belajar hal ini dengan cara yang sulit ketika menggunakan *ChatGPT* untuk membantu saya dalam pemrograman. Saya juga bukan pengguna berat AI, tetapi dalam seminggu saya menyadari bahwa masalah kode yang saya atasi dalam proyek saya tidak mengajarkan apa pun kepada saya ketika dibantu oleh AI. Sebelumnya, setiap masalah yang saya selesaikan dalam pemrograman secara mandiri dan dengan sumber daya online lainnya, membuat prosesnya jauh lebih lama, tetapi saya sering belajar begitu banyak sehingga saya menemukan

solusi yang lebih baik daripada yang saya kira bisa saya lakukan karena saya memahami konsepnya dan menemukan cara untuk menyesuaikan solusi berdasarkan kebutuhan proyek spesifik saya. Saya melihat masa depan di mana AI dibangun dengan tujuan pembelajaran yang sejati, tetapi saat ini *ChatGPT* tidak melakukannya secara *default*'. Pada sebuah postingan *Facebook* dari user Nadhir Ashafiq terdapat komentar dari seorang user dengan nama Arif Zuhairi (dengan terjemahan) 'Di masa depan, manusia akan menjadi lebih bodoh, bergantung pada kecerdasan buatan (AI). Bukan sekarang, tapi jika AI menjadi lebih sadar di kemudian hari... kecuali mereka menghancurkannya sebelum AI menghancurkan manusia...'. Di sosial media lain seperti *Instagram* juga terjadi, seperti sebuah postingan dari akun *@technology*. Terdapat sebuah komentar dari akun dengan user *@kulture_* (dengan terjemahan) 'Karena memberikan saran yang lebih baik daripada "para profesional". Sambil menghemat uang. AI yang mundur.'. Di *Twitter* juga muncul berbagai sentimen salah satunya dari akun *@pulismen* dengan tweet 'sorry nih tapi kok *chatgpt* 4.o jelek banget ya?'. Selain beragamnya sentimen publik lintas platform, peneliti memutuskan untuk menggunakan *Twitter* sebagai sumber data utama dalam penelitian ini. *Twitter* merupakan platform yang banyak digunakan untuk mengungkapkan opini publik di Indonesia karena sifatnya yang cepat, terbuka, serta mudah diakses untuk kebutuhan analisis sentimen (Wicaksana et al., 2022). Hal ini didasari oleh karakteristik *Twitter* yang memungkinkan pengguna menyampaikan opini secara cepat, spontan, serta bersifat *real-time* sehingga cocok untuk menangkap dinamika pandangan masyarakat terhadap fenomena tertentu (Uliniansyah et al., 2024).

Penelitian yang dilakukan oleh (Pramono et al., 2024) menunjukkan bahwa analisis sentimen di *Twitter* masih sangat relevan terutama dalam konteks sosial seperti pemilihan Presiden pada tahun 2024. Hal ini menguatkan bahan media sosial *Twitter* masih bisa menjadi sumber data yang cukup penting untuk memahami persepsi masyarakat, sehingga pemanfaatannya untuk opini publik seperti *ChatGPT* juga menjadi sangat relevan. Dengan jumlah data yang besar dan aksesibilitas yang luas, *Twitter* dapat digunakan untuk memperoleh representasi opini masyarakat secara lebih komprehensif (Akbar et al., 2020).

Penelitian sebelumnya menunjukkan bahwa berbagai algoritma pembelajaran mesin seperti *Support Vector Machine* (SVM), *Logistic Regression*, *K-Nearest Neighbors* (KNN), dan *Naive Bayes*, serta model berbasis *deep learning*, telah banyak diterapkan untuk analisis sentimen. Hasil studi komparatif tersebut mengungkap bahwa SVM dan model *deep learning* umumnya memberikan akurasi lebih tinggi, namun membutuhkan sumber daya komputasi yang lebih besar dan waktu pelatihan yang lebih panjang. Sebaliknya, algoritma seperti KNN dan *Decision Tree* cenderung kurang efisien pada data berskala besar, sementara *Naive Bayes* menawarkan kinerja yang relatif stabil pada teks berukuran pendek (Seo et al., 2020).

Penelitian terbaru menunjukkan bahwa *Naive Bayes* mampu memberikan kinerja yang kompetitif dalam melakukan klasifikasi sentimen pada data berteks pendek di *Twitter*. Dalam studi yang membandingkan metode *Naive Bayes* dengan VADER dan *TextBlob* pada analisis sentimen layanan *ShopeeFood*, *Naive Bayes* terbukti efektif dalam menangani karakteristik teks yang ringkas dan dinamis serta menghasilkan performa yang baik dibandingkan metode lainnya (Huda et al., 2025).

Terdapat suatu kendala dalam menggunakan algoritma NBC yaitu probabilitas nol/*Zero Probability*. Hal ini terjadi jika label kelas individu hilang, maka estimasi probabilitas berbasis frekuensi akan menjadi nol dan kita akan mendapatkan nilai nol ketika semua probabilitas dikalikan. Hal ini dapat diatasi dengan teknik *Laplace Smoothing*. Jika kita memilih nilai $\alpha \neq 0$ (tidak sama dengan 0), probabilitas tidak akan lagi menjadi nol meskipun sebuah kata tidak ada di dalam dataset pelatihan (Rizky et al., 2024).

Dalam menganalisis data terkadang data yang diperoleh tidak seimbang (*Imbalance Data*). Data yang memiliki rasio yang tidak berimbang antara data satu dengan data lainnya dapat dikatakan sebagai *imbalanced*. *Data mining* mengartikan *imbalanced* dengan jumlah data kelas mayoritas lebih banyak dibandingkan dengan kelas minoritas maka perlu dilakukan *Synthetic Minority Oversampling Technique* (SMOTE). Teknik SMOTE dilakukan karena terbukti dapat mengatasi masalah *imbalanced data* dan mendapatkan hasil klasifikasi yang lebih baik seperti pada penelitian (Suchrady & Purwarianti, 2023; Obiedat et al.,

2022 ; Zhang et al., 2023). Salah satu teknik yang digunakan untuk mengatasi ketidakseimbangan data adalah *Random Undersampling*, yaitu teknik penyeimbangan data yang bekerja dengan cara mengurangi jumlah data pada kelas mayoritas secara acak hingga mencapai distribusi yang seimbang dengan kelas minoritas (Yousefimehr et al., 2026). Pada penelitian ini, *Random Undersampling* dikombinasikan dengan SMOTE sehingga menghasilkan teknik resampling yang lebih optimal dibandingkan menggunakan satu teknik saja (Arianto & Nurrahmasita, 2025).

Berdasarkan pemaparan serta hasil temuan penelitian sebelumnya, penulis merumuskan pendekatan metodologis yang akan diterapkan dalam penelitian ini. Pemilihan metode dilakukan dengan mempertimbangkan karakteristik data berupa teks pendek di platform *Twitter*, kondisi distribusi kelas yang tidak seimbang, serta kebutuhan penguatan performa model dalam proses klasifikasi sentimen. Adapun langkah metodologis yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Penelitian ini menggunakan algoritma *Naive Bayes Classifier* (NBC) dalam proses klasifikasi sentimen. Pemilihan NBC didasarkan pada temuan penelitian yang menyatakan bahwa algoritma ini efektif untuk menganalisis teks pendek, seperti data *Twitter*, serta mampu menghasilkan performa yang kompetitif dalam tugas klasifikasi sentimen (Huda et al., 2025).
2. Penelitian ini menerapkan pendekatan *Term Frequency* (TF) sebagai metode seleksi fitur utama. Pendekatan TF dipilih karena mampu menyoroti kata-kata yang paling sering muncul dalam dokumen sehingga memberikan kontribusi bermakna terhadap proses pemodelan. Secara teknis, metode ini dipandang sesuai diterapkan pada teks pendek karena sederhana, komputasional efisien, dan dapat mendukung peningkatan kinerja klasifikasi.
3. Penelitian ini menggunakan *Laplace Smoothing* dan *Synthetic Minority Oversampling Technique* (SMOTE) sebagai teknik pendukung pemodelan. *Laplace Smoothing* diterapkan untuk mengatasi permasalahan *zero probability* dalam NBC, yaitu kondisi ketika suatu kata tidak muncul pada kategori tertentu sehingga probabilitas kelas menjadi nol (Rizky et al., 2024). Sementara itu, SMOTE digunakan untuk menyeimbangkan distribusi kelas pada

data sentimen yang tidak seimbang (*imbalanced data*), mengingat teknik ini terbukti mampu meningkatkan kinerja klasifikasi dalam berbagai konteks analisis teks (Obiedat et al., 2022 ; Suchrady & Purwarianti, 2023) (Zhang et al., 2023)

Berdasarkan pada latar belakang yang telah penulis uraikan sebelumnya, maka penulis akan melakukan penelitian terhadap pengguna (user) *Twitter* terkait topik dengan *ChatGPT* dengan menggunakan Algoritma NBC serta mengukur tingkat akurasi *Laplace Estimator* dengan menjadikan topik tersebut sebagai bahan kajian yang tertuang dalam bentuk skripsi dengan judul “**Analisis Sentimen *ChatGPT* Pada *Twitter* Dengan Menggunakan Algoritma *Naive Bayes Classifier* (NBC) Dengan *Laplace Estimator*** “

1.2 Identifikasi Masalah

Di bawah ini adalah indikator masalah yang muncul:

1. Perbedaan opini masyarakat tentang adanya *ChatGPT* di media sosial *Twitter*.
2. Perlunya analisis sentimen untuk menggali lebih lanjut mengenai opini masyarakat mengenai *ChatGPT* menggunakan media sosial *Twitter*.
3. Belum adanya analisis menggunakan algoritma *Naive Bayes Classifier* (NBC) dengan *Laplace Estimator* pada pembahasan *ChatGPT* di media sosial *Twitter*.

1.3 Pembatasan Masalah

Luasnya bidang keahlian yang dikerjakan dalam merealisasikan penelitian ini, maka peneliti membatasi permasalahan pada:

1. Sentimen publik yang diambil dari data *twitter* menggunakan *query ChatGPT, GPT3, GPT4, Sora OpenAI, Open AI, OpenAI, Kecerdasan Buatan ChatGPT, Chatbot ChatGPT*.
2. Jumlah data *tweet* yang diambil berkisar 4.590 tweet dari tanggal 10 Maret 2025 sampai 23 Maret 2025 dengan data yang diambil dari tahun 2022-2024 dengan tingkat *error rate* $\pm 10\%$ (Fitriana & Sibaroni, 2020).

3. Dalam melakukan penelitian peneliti mengambil data berbahasa Indonesia dan akan mengabaikan bahasa asing, bahasa daerah, penggunaan emoji dan *retweet*.
4. Dalam melakukan penelitian peneliti akan menggunakan 3 perbandingan kelas yaitu *True/Positive*, *False/Negative* dan *Neutral*.

1.4 Perumusan Masalah

Pada penelitian ini dapat dirumuskan permasalahan yang dapat diangkat, yaitu “Bagaimana hasil tingkat akurasi analisis sentimen *ChatGPT* pada *twitter* dengan menggunakan metode algoritma *Naive Bayes Classifier* (NBC) dengan *Laplace Estimator*?”.

1.5 Tujuan Penelitian

Peneliti melakukan penelitian ini dengan tujuan untuk mengetahui hasil tingkat akurasi sentiment masyarakat terhadap *ChatGPT* dan *OpenAI* menggunakan Algoritma *Naive Bayes Classifier* (NBC) dengan *Laplace Estimator* di media sosial *Twitter*.

1.6 Manfaat Penelitian

Dengan dilakukannya penelitian ini, diharapkan hasil dari penelitian dapat memberikan beberapa manfaat di bawah ini:

1. Manfaat bagi penulis
 - a. Memberikan pengetahuan lebih kepada penulis dalam menganalisa sentimen masyarakat terhadap topik tentang *ChatGPT* dan *OpenAI* menggunakan Algoritma *Naive Bayes Classifier* (NBC) dengan metode penelitian dengan *Laplace Estimator* dan Teknik SMOTE.
 - b. Hasil penelitian ini dapat menjadi referensi penulis apabila hendak melanjutkan penelitian di masa yang akan datang.

2. Manfaat bagi universitas
 - a. Mengetahui kemampuan mahasiswa dalam menguasai dan menerapkan ilmu yang telah didapatkan di bangku perkuliahan dan juga sebagai bahan evaluasi.
 - b. Menjadi sebuah tolak ukur bagi universitas dalam menentukan keberhasilan dan kemampuan penulis dalam mengimplementasikan ilmu yang sudah didapatkan selama menempuh pendidikan perkuliahan di universitas.
 - c. Sebagai referensi untuk penelitian selanjutnya.
3. Manfaat bagi masyarakat
 - a. Memberikan informasi kepada masyarakat dalam menganalisa sentimen masyarakat terhadap topik tentang *ChatGPT* dan *OpenAI* yang diambil dari *twitter*.
 - b. Hasil dari penelitian ini dapat digunakan untuk penelitian tahap selanjutnya.

