

BAB I PENDAHULUAN

1.1 Latar Belakang

Perkembangan ilmu pada bidang teknologi dan informasi semakin hari kian semakin pesat. Teknologi menjadi alat yang dapat membantu kebutuhan manusia, dengan teknologi, apapun dapat dilakukan dengan lebih mudah. Begitu pentingnya peran teknologi inilah yang mulai membawa peradaban memasuki ke era digital. Seiring dengan pesatnya perkembangan teknologi tersebut kesadaran masyarakat terkait penggunaan teknologi pun semakin meningkat, hal itu dapat dilihat dengan semakin berkembangnya aplikasi-aplikasi sosial media, dimana kini hampir setiap orang mempunyai akun sosial media. Dengan adanya sosial media tersebut maka kita bisa dengan mudah berkomunikasi dengan seseorang yang jarak nya jauh bahkan letak geografisnya pun berbeda. Selain itu masih banyak sekali kegiatan-kegiatan yang dapat dilakukan dengan memanfaatkan sosial media. Maka dari itu, sosial media semakin digemari oleh semua kalangan umur karena bersifat fleksibel dan bisa diakses kapan pun dan dimana pun.

Bersosial media selain memudahkan untuk berkomunikasi dengan seseorang yang letaknya berjauhan, kita juga dapat dengan mudah menerima atau mencari suatu informasi. Informasi yang didapat pun bahkan bisa bersifat *realtime* atau langsung kita terima lewat seseorang yang mengunggah suatu peristiwa yang terjadi pada suatu *platform* sosial media. Para pengguna juga dapat mengetahui hal apa yang sedang banyak dibicarakan atau sedang *trend* di Indonesia maupun mengetahui *trend* di dunia.

Banyak sekali *platform* sosial media yang dapat kita gunakan, namun sosial media yang kini sedang banyak di gandrungi oleh masyarakat di Indonesia salah satunya adalah twitter. Twitter merupakan sebuah platform sosial media dimana penggunanya dapat mencurahkan isi pikirannya dengan membuat sebuah tweet. Selain itu, pengguna juga dapat mengunggah gambar ataupun video. Twitter juga dikenal dengan kecepatannya dalam menyebarkan informasi, kita dapat dengan

mudah mendapat informasi terbaru dan mengetahui hal apa yang sedang ramai dibicarakan oleh pengguna twitter dengan memanfaatkan fitur *trending topic*. *Trending topic* itu sendiri merupakan fitur untuk melihat frekuensi orang-orang dalam menggunakan kata-kata pada unggahan tweetnya. Semakin sering kata tersebut digunakan oleh pengguna maka kata atau kalimat tersebut akan menjadi *trending topic*. Dalam penerimaan informasi tidak selalu sebuah hal positif yang akan diterima, banyak sekali informasi-informasi *hoax* yang tersebar di sosial media dimana apabila kita tidak memilahnya dengan bijak dapat menyebabkan kesalahan dalam penerimaan informasi.

Topik perbincangan yang sedang ramai dibicarakan oleh para pengguna twitter belakangan ini adalah dengan adanya kehadiran ChatGPT. ChatGPT adalah *chatbot* berbasis teknologi kecerdasan buatan yang dikembangkan oleh OpenAI. ChatGPT ini memungkinkan untuk dapat mengobrol dengan manusia menggunakan kata yang sangat runtut, kontekstual, bahkan bisa diperintah untuk menuliskan baris kode pemrograman, sehingga disebut-sebut dapat menggantikan hasil dari mesin pencarian Google. OpenAI itu sendiri merupakan sebuah perusahaan yang bergerak dibidang kecerdasan buatan (*artificial intelligence*). Produk yang paling menjadi sorotan dari OpenAI adalah DALL-E dan GPT. DALL-E yaitu AI yang bisa menghasilkan gambar atau lukisan dari sepotong kalimat, sementara GPT yaitu AI pembuat teks. GPT memungkinkan AI membuatkan teks, mulai dari puisi, artikel, tutorial, hingga caption media sosial hingga pengetahuan umum dari sepotong pernyataan atau prompt. ChatGPT merupakan pengembangan dari GPT 3.0 yang diperkenalkan dua tahun lalu.

Kemunculan ChatGPT tersebut tentunya mendapatkan respon serta tanggapan yang berbeda-beda dari setiap pengguna sosial media twitter, ada pengguna yang mendukung kehadiran ChatGPT ini dan ada juga pengguna yang memandang ChatGPT sebagai sebuah ancaman. Variasi respon yang beragam itu menjadikan suatu hal yang menarik untuk dapat diteliti terkait

bagaimana sentimen para pengguna sosial media twitter terhadap kemunculan ChatGPT dengan menerapkan ilmu *text mining*.

Menurut Hofmann & Chisholm (2018), *text mining* adalah proses mengekstraksi informasi yang berguna dari data teks dengan mengenali pola, mengidentifikasi tren dan temuan yang bermanfaat, serta mengungkapkan pengetahuan dari data teks. Sedangkan, menurut Chakraborty et al. (2021), *text mining* adalah proses mengubah data teks menjadi informasi yang berguna melalui identifikasi pola, pengelompokan, klasifikasi, analisis sentimen, dan pencarian informasi, dengan tujuan meningkatkan pemahaman tentang topik tertentu dan mendukung pengambilan keputusan yang lebih baik. *Text mining* melibatkan penggunaan algoritma untuk mengekstraksi informasi yang berguna dari data teks dalam identifikasi pola serta mengidentifikasi tren dan temuan yang bermanfaat dengan penggunaan teknik seperti pengelompokan, klasifikasi, dan analisis sentimen untuk membantu pengambilan keputusan yang lebih baik dengan memahami pandangan dan opini pengguna, serta mencari tahu tren dan kecenderungan perilaku pengguna pada sosial media tertentu. Dalam keseluruhan, *text mining* telah menjadi topik yang semakin penting dalam analisis data modern dan dapat membantu organisasi dalam mengambil keputusan yang lebih baik dan efektif.

Penggunaan algoritma atau metode yang membantu dalam proses analisis data pada *text mining* sangat beragam, diantaranya yaitu metode klasifikasi. Metode klasifikasi adalah teknik-teknik atau algoritma-algoritma yang digunakan untuk memprediksi kelas atau label target dari suatu data berdasarkan fitur-fitur yang ada. Tujuan dari metode klasifikasi adalah untuk mengidentifikasi pola-pola atau aturan-aturan dari data pelatihan sehingga dapat digunakan untuk memprediksi label atau kelas dari data uji.

Dalam *text mining*, *Support Vector Machine* atau yang selanjutnya akan disebut juga SVM dapat digunakan untuk melakukan klasifikasi teks ke dalam beberapa kategori berdasarkan fitur-fitur atau atribut-atribut yang terkait

dengan mencari hyperplane terbaik untuk memisahkan kelas yang berbeda dalam ruang fitur. Sementara itu, *Naive Bayes Classifier* atau yang selanjutnya akan disebut NBC adalah algoritma klasifikasi berdasarkan teorema Bayes dengan menghitung probabilitas dari suatu data berada pada kelas tertentu berdasarkan probabilitas kelas dan probabilitas fitur dalam kelas tersebut. Selanjutnya *Logistic Regression* atau selanjutnya akan disebut LR adalah algoritma klasifikasi yang digunakan untuk masalah klasifikasi biner (dua kelas) atau masalah klasifikasi multikelas, algoritma ini menggunakan fungsi sigmoid untuk memodelkan probabilitas bahwa suatu data termasuk dalam kelas tertentu. Selanjutnya untuk *Random Forest* atau selanjutnya akan disebut RF adalah metode klasifikasi yang termasuk dalam kategori *ensemble learning*. Ini berarti model *random forest* menggunakan beberapa *decision trees* untuk meningkatkan performa dan mengurangi *overfitting*. Dan berikutnya yaitu Decision Tree atau selanjutnya akan disebut DT merupakan algoritma klasifikasi yang berbentuk seperti pohon keputusan dengan simpul-simpul yang mewakili keputusan dan daun-daun yang mewakili label kelas. Pada setiap simpul, algoritma melakukan pemilihan fitur yang paling informatif untuk memisahkan data berdasarkan label kelasnya.

Dalam ke-lima metode klasifikasi tersebut, fitur atau atribut yang digunakan biasanya berupa kata-kata atau frasa-frasa tertentu yang terkait dengan topik yang dibahas dalam teks. Oleh karena itu, SVM, NBC, LR, RF dan DT dapat membantu dalam mengidentifikasi pola dan relasi dalam teks yang dapat digunakan untuk mengambil informasi atau pengetahuan dari teks tersebut.

Berdasarkan dari uraian latar belakang di atas, maka penulis tertarik untuk melakukan penelitian dengan judul : **“Analisis Perbandingan Sentimen Pengguna Twitter Terhadap ChatGPT Menggunakan *Text Mining* Dengan Metode Klasifikasi.”**

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan diatas, penulis dapat menentukan rumusan masalah yang dijadikan dasar penelitian ini diantaranya yaitu :

1. Bagaimana analisis sentimen pengguna twitter terhadap kehadiran teknologi ChatGPT berdasarkan metode *Support Vector Machine*?
2. Bagaimana analisis sentimen pengguna twitter terhadap kehadiran teknologi ChatGPT berdasarkan metode *Naïve Bayes Classifier*?
3. Bagaimana analisis sentimen pengguna twitter terhadap kehadiran teknologi ChatGPT berdasarkan metode *Logistic Regression*?
4. Bagaimana analisis sentimen pengguna twitter terhadap kehadiran teknologi ChatGPT berdasarkan metode *Random Forest*?
5. Bagaimana analisis sentimen pengguna twitter terhadap kehadiran teknologi ChatGPT berdasarkan metode *Decision Tree*?
6. Bagaimana perbandingan kinerja hasil analisis sentimen pengguna Twitter terhadap ChatGPT berdasarkan metode *Support Vector Machine, Naïve Bayes Classifier, Logistic Regression, Random Forest* dan *Decision Tree*?

1.3 Batasan Masalah

Adapun batasan masalah berdasarkan judul yang peneliti pilih, yaitu :

1. Pengumpulan data set diperoleh melalui tweet pengguna twitter yang mengandung kata kunci “ChatGPT”.
2. Proses menganalisis data set menggunakan bahasa pemrograman *python*.
3. Metode klasifikasi yang digunakan yaitu SVM, NBC, LR, RF dan DT
4. Data yang akan diuji hanya data yang menggunakan bahasa Indonesia untuk kata kunci “ChatGPT”.
5. Data yang diambil merupakan data dari bulan Juni 2023.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dengan dilakukannya penelitian ini adalah sebagai berikut :

1. Mengetahui bagaimana sentimen para pengguna twitter terhadap kemunculan ChatGPT dengan menggunakan metode *Support Vector Machine*.
2. Mengetahui bagaimana sentimen para pengguna twitter terhadap kemunculan ChatGPT dengan menggunakan metode *Naïve Bayes Classifier*.
3. Mengetahui bagaimana sentimen para pengguna twitter terhadap kemunculan ChatGPT dengan menggunakan metode *Logistic Regression*.
4. Mengetahui bagaimana sentimen para pengguna twitter terhadap kemunculan ChatGPT dengan menggunakan metode *Random Forest*.
5. Mengetahui bagaimana sentimen para pengguna twitter terhadap kemunculan ChatGPT dengan menggunakan metode *Decision Tree*.
6. Mengetahui bagaimana perbedaan dari hasil perbandingan performa antara metode *Support Vector Machine (SVM)*, *Naïve Bayes Classifier*, *Logistic Regression*, *Random Forest*, dan *Decision Tree* pada pengklasifikasian sentiment pengguna sosial media twitter terhadap kemunculan ChatGPT.

1.5 Sistematika Penulisan

Sistematika penulisan membantu untuk memberikan gambaran menyeluruh terhadap masalah yang dibahas, sistematika penulisan dibagi ke dalam beberapa bab, yaitu:

Bab I Pendahuluan

Bab ini menyajikan mengenai latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, dan sistematika penulisan.

Bab II Kajian Pustaka

Bab ini menyajikan tentang kajian pustaka terkait dokumen, penelitian sebelumnya yang telah dilakukan yang mendasari penelitian dan dasar teori yang terkait dengan penelitian.

Bab III Metode Penelitian

Bab ini menyajikan metode dan langkah-langkah yang dilakukan oleh penulis dalam melakukan penelitian ini. Bab ini menguraikan terkait objek juga metode penelitian yang digunakan serta metode dari pengumpulan data seperti latar belakang, rumusan masalah, Batasan penelitian, tujuan, pengambilan data twitter yang kemudian dilakukan *preprocessing* sebelum diimplementasikan terhadap metode *support vector machine*, *naïve bayes classifier*, *logistic regression*, *random forest* dan *decision tree*.

Bab IV Hasil Penelitian dan Pembahasan

Bab ini menyajikan mengenai pembahasan hasil penelitian yang telah dilakukan dan analisis hasil penelitian yang diperoleh.

Bab V Kesimpulan dan Saran

Bab ini menyajikan kesimpulan dan saran yang didapatkan dari pembahasan hasil penelitian. Pada bab ini terdapat kesimpulan dari uraian yang telah dibahas pada bab sebelumnya, juga saran bagi peneliti selanjutnya dari peneliti.