

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Media sosial telah menjadi bagian dalam kehidupan masyarakat. Selain berfungsi sebagai alat komunikasi, media sosial juga menjadi tempat untuk menyuarakan pendapat, memberikan kritik, serta membagikan pengalaman (Umbara, 2021). Kemudahan akses dan jangkauan informasi yang luas menjadikan media sosial sebagai sarana untuk memperoleh dan menyebarluaskan informasi secara cepat dan efisien (Zaini dkk., 2024).

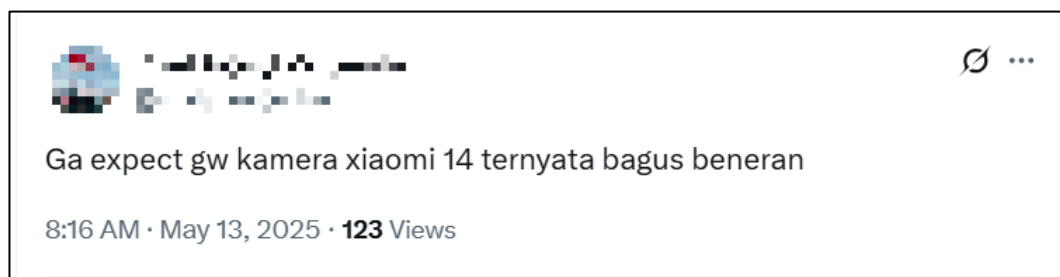
Berdasarkan laporan dari *Hootsuite (We Are Social)*, per awal tahun 2025, tercatat sebanyak 143 juta identitas pengguna media sosial di Indonesia, yang merepresentasikan sekitar 50,2% dari total populasi (Riyanto, 2025). Salah satu platform media sosial yang digunakan masyarakat Indonesia dan dikenal karena kemampuannya menyajikan informasi secara *real-time* adalah X (sebelumnya Twitter).

Pada mulanya, X dikenal sebagai platform *microblogging* yang berfokus pada penyampaian pesan dalam bentuk konten teks singkat. Melalui format yang mendukung penyebaran informasi secara terbuka, independen, serta *real-time*, platform *microblogging* seperti X kemudian berkembang menjadi salah satu wadah utama dalam komunikasi opini publik (Zhao dkk., 2022). X juga saat ini tercatat memiliki sekitar 25,2 juta pengguna di Indonesia (Riyanto, 2025). Jumlah pengguna tersebut menandakan X sebagai salah satu media sosial populer di Indonesia yang juga menghasilkan volume data diskusi dan opini publik yang besar.

Dengan karakteristiknya yang terbuka, ringkas, dan cepat, X sebagai media sosial juga menjadi tempat bagi penggunanya dalam menyampaikan beragam opini, pengalaman, termasuk ulasan terhadap suatu produk atau layanan. Penyebaran opini konsumen secara *online* ini dikenal sebagai *electronic word-of-mouth (e-WOM)*, yang didefinisikan sebagai segala bentuk pernyataan dari pelanggan mengenai suatu produk atau perusahaan, yang dapat diakses oleh banyak orang dan institusi melalui internet (Hennig-Thurau dkk., 2004). Seiring meningkatnya

penggunaan media sosial, konsumen juga semakin aktif menyuarakan opini secara terbuka, yang mana hal tersebut dapat memengaruhi citra dan reputasi perusahaan di mata publik (Kusumawati dkk., 2024).

Umpan balik berupa opini yang diungkapkan konsumen tidak hanya menggambarkan pandangan dan pengalaman yang dirasakan, tetapi juga menjadi sumber data yang berharga bagi perusahaan. Opini yang terbentuk tersebut dapat dianalisis lebih lanjut untuk mengidentifikasi tren konsumen dan menilai persepsi publik terhadap perusahaan (Giachanou & Crestani, 2016). Selain itu, opini dalam bentuk ulasan juga dapat digunakan untuk peningkatan kualitas produk atau layanan (Qi dkk., 2016). Oleh karena itu, perusahaan perlu menyadari dan memahami opini publik yang beredar di media sosial terkait produk, layanan, maupun merek mereka, sebagai langkah untuk memperoleh wawasan yang bermanfaat.



Gambar 1.1 Contoh Opini Konsumen pada *Post* di X

Dengan menganalisis teks opini konsumen di media sosial, perusahaan dapat memahami emosi yang diungkapkan serta isu atau aspek yang menjadi fokus perhatian konsumen. Misalnya, pada Gambar 1.1 dapat terlihat bahwa tanggapan pengguna bernada positif dan pembahasan mengarah pada keunggulan fitur kamera produk. Informasi ini berguna sebagai masukan bagi perusahaan untuk mengenali kekuatan yang diapresiasi konsumen maupun potensi kelemahan apabila muncul tanggapan berupa keluhan.

Namun, mengolah data opini di media sosial bukanlah hal yang mudah. Volume data dalam bentuk teks yang dihasilkan di media sosial, seperti X, sangat besar dan beragam, serta sering kali menggunakan bahasa yang tidak baku, seperti singkatan, akronim, afiks, bentuk yang tidak beraturan, dan campuran ragam bahasa (Ambiya & Anggriani, 2018). Pendekatan teknologi yang tepat untuk menangani tantangan

tersebut secara otomatis, cepat, dan akurat adalah Pemrosesan Bahasa Alami atau *Natural Language Processing* (NLP).

Natural Language Processing (NLP) adalah bagian dari kecerdasan buatan (*artificial intelligence/AI*) yang mempelajari bagaimana komputer mampu memahami serta mengolah bahasa alami yang digunakan oleh manusia, baik dalam bentuk tulisan maupun suara (Sulistyo dkk., 2021). Salah satu penerapan umum NLP adalah pada *task* klasifikasi untuk analisis sentimen, yaitu proses mengelompokkan opini dalam teks menjadi kategori seperti positif, negatif, atau netral (Çelikutğ, 2018). Analisis sentimen kini menjadi alat yang penting bagi perusahaan untuk mengumpulkan umpan balik pelanggan dalam upaya merumuskan strategi yang lebih efektif, misalnya dalam hal pemasaran (Gunasekaran, 2023). Melakukan analisis sentimen pada pandangan masyarakat di media sosial terhadap perusahaan juga dapat dimanfaatkan untuk memantau dan mengevaluasi reputasi merek (Adiguna, 2023).

Pemanfaatan NLP lain pada proses analisis teks yakni NLP dapat digunakan untuk mengeksplorasi dan mengelompokkan topik dari kumpulan data teks atau yang dikenal dengan istilah *topic modeling*. *Topic modeling* atau pemodelan topik merupakan teknik yang digunakan untuk menemukan topik-topik abstrak atau tersembunyi dalam kumpulan dokumen teks yang besar (Barde & Bainwad, 2017). Topik dalam *topic modeling* direpresentasikan sebagai distribusi probabilitas atas kumpulan kata dalam korpus, dengan kata-kata yang memiliki kecenderungan muncul bersama akan membentuk suatu topik tertentu (Jelodar dkk., 2018). Dengan memanfaatkan *topic modeling*, perusahaan dapat mengidentifikasi jenis-jenis topik pembicaraan yang disampaikan konsumen tanpa harus memberikan label atau kategori pada teks tersebut sebelumnya (Wayasti & Surjandari, 2018).

Kemajuan *deep learning* turut mendorong kemampuan model NLP dalam memahami bahasa secara lebih kompleks dan kontekstual. *Deep learning* merupakan bagian dari *machine learning* yang menggunakan jaringan saraf tiruan (*neural networks*) yang terinspirasi dari cara kerja otak manusia untuk memecahkan permasalahan kompleks, termasuk dalam mempelajari data yang tidak terstruktur dan tidak berlabel (Sharifani & Amini, 2023). Salah satu terobosan di bidang NLP yang memanfaatkan *deep learning* ini adalah hadirnya arsitektur Transformer.

Transformer memperkenalkan mekanisme *self-attention* yang memungkinkan model memahami hubungan antar kata dalam satu kalimat atau dokumen secara paralel dan kontekstual (Vaswani dkk., 2017). Dalam *task* klasifikasi sentimen misalnya, model BERT (*Bidirectional Encoder Representations from Transformers*) yang menggunakan arsitektur Transformer memiliki performa yang lebih baik dibandingkan algoritma *machine learning* tradisional (González-Carvajal & Garrido-Merchán, 2020). Pada pemodelan topik, model yang menggunakan arsitektur Transformer seperti BERTopic juga mampu menghasilkan *cluster* topik yang lebih koheren dan relevan dibandingkan metode tradisional (Grootendorst, 2022).

Penelitian sebelumnya telah memanfaatkan arsitektur Transformer untuk analisis sentimen dan pemodelan topik di media sosial. Misalnya, penelitian yang menganalisis 16.611 komentar di situs Reddit menggunakan BERTopic dan *fine-tuned* BERT yang berhasil menemukan bahwa sebanyak 57% komentar dari data tersebut bersifat positif tentang masa depan AI dengan 36 topik teridentifikasi (Ocal, 2024). Contoh lain yakni penggunaan IndoBERT dan BERTopic untuk menganalisis *tweet* berbahasa Indonesia tentang layanan transportasi *online*, dengan temuan akurasi model mencapai 97% dan 6 topik berhasil diidentifikasi (Mahfudiyah & Alamsyah, 2023).

Berdasarkan pemaparan urgensi yang telah diidentifikasi, skripsi ini memilih untuk melakukan *fine-tuning* model Transformer untuk *task* klasifikasi dalam analisis sentimen, serta menerapkan pendekatan berbasis Transformer untuk pemodelan topik. Keduanya diaplikasikan pada kumpulan *post* terkait sebuah merek teknologi di platform X. Skripsi ini bertujuan untuk mengembangkan model klasifikasi sentimen yang akurat dan mengidentifikasi topik-topik perbincangan utama sekaligus menganalisis bagaimana sentimen publik terdistribusi dalam setiap topik tersebut.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya, dapat dirumuskan beberapa pertanyaan penelitian ini sebagai berikut.

- a. Bagaimana cara menerapkan metode *fine-tuning* model Transformer untuk klasifikasi sentimen serta metode berbasis Transformer untuk pemodelan topik pada data *post* terkait merek teknologi di media sosial X?
- b. Bagaimana evaluasi kinerja model Transformer yang telah dilakukan *fine-tuning* dalam melakukan klasifikasi sentimen serta kualitas model topik yang dihasilkan dengan metode berbasis Transformer pada data *post* terkait merek teknologi di media sosial X?
- c. Bagaimana hasil temuan terkait sentimen publik dan topik-topik pembahasan dominan mengenai merek teknologi di media sosial X?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut.

- a. Menerapkan metode *fine-tuning* pada pengembangan model Transformer untuk *task* klasifikasi sentimen, serta mengaplikasikan metode berbasis Transformer dalam pemodelan topik terhadap data *post* mengenai merek teknologi di media sosial X.
- b. Mengevaluasi kinerja model Transformer yang telah melalui proses *fine-tuning* dalam melakukan klasifikasi sentimen serta menganalisis kualitas topik yang dihasilkan dengan metode berbasis Transformer dari data *post* terkait merek teknologi di media sosial X.
- c. Mengidentifikasi dan menganalisis sentimen publik beserta topik-topik pembahasan dominan mengenai merek teknologi di media sosial X, serta mengetahui sebaran sentimen untuk tiap topik.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut.

- a. Penelitian ini diharapkan dapat menambah referensi literatur di bidang *Natural Language Processing* (NLP), khususnya mengenai penerapan model Transformer seperti IndoBERT dan BERTopic untuk analisis data media sosial berbahasa Indonesia dalam kombinasi *task* klasifikasi sentimen dan pemodelan topik.

- b. Penelitian ini juga diharapkan dapat memberikan wawasan berbasis data yang dapat dimanfaatkan oleh pembaca untuk memahami persepsi publik terhadap merek Xiaomi di media sosial X.

1.5 Batasan Masalah

Dalam penelitian ini, batasan masalah yang ditetapkan adalah sebagai berikut.

- a. Data yang digunakan terbatas pada kumpulan *post* dalam bahasa Indonesia yang diperoleh secara *scraping* pada media sosial X.
- b. *Post* yang dikumpulkan hanya bersifat teks (tidak mencakup media seperti gambar dan video).
- c. Merek teknologi yang menjadi fokus penelitian adalah Xiaomi. Pemilihan merek tersebut dengan alasan bahwa merek Xiaomi merupakan merek populer di Indonesia dan dikenal memproduksi berbagai perangkat elektronik sehingga dengan harapan topik-topik yang diperbincangkan beragam.
- d. Eksperimen yang dilakukan yakni menerapkan *fine-tuning* dari *pre-trained* model IndoBERT untuk pengembangan model klasifikasi dalam *task* analisis sentimen dan pengaplikasian metode berbasis Transformer yaitu BERTopic untuk *task* pemodelan topik.
- e. Evaluasi proses *fine-tuning* model IndoBERT untuk klasifikasi sentimen dilakukan dengan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score* dan kurva ROC. Sementara untuk evaluasi hasil topik dari proses pemodelan topik menggunakan BERTopic digunakan metrik kuantitatif *topic coherence score* (C_v) dan pendekatan kualitatif yakni *human judgment* oleh penulis.
- f. Penelitian ini tidak mempertimbangkan faktor *real-time processing*, sehingga hanya menggunakan data yang telah dikumpulkan selama masa penelitian.

1.6 Sistematika Penulisan

Skripsi ini disusun dengan sistematika penulisan yang terdiri dari lima bab. Berikut ini adalah penjelasan untuk masing-masing bab tersebut.

BAB I PENDAHULUAN

Bab ini mengawali penelitian dengan memaparkan konteks dan urgensi masalah yang diangkat sebagai latar belakang. Selanjutnya, dirumuskan

pertanyaan-pertanyaan penelitian spesifik. Tujuan yang ingin dicapai melalui penelitian ini kemudian ditetapkan secara jelas, beserta potensi manfaat yang dapat dihasilkan. Terakhir, bab ini juga menjelaskan batasan ruang lingkup penelitian serta memberikan gambaran umum struktur penulisan.

BAB II KAJIAN PUSTAKA

Bab ini menyajikan landasan teoritis mengenai konsep-konsep, model, dan algoritma utama yang relevan dengan penelitian. Selain itu, dilakukan tinjauan terhadap penelitian-penelitian terdahulu yang terkait, merangkum temuan dan metode yang digunakan.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai rancangan dan pendekatan yang digunakan untuk menjawab pertanyaan penelitian. Langkah-langkah atau prosedur pelaksanaan penelitian diuraikan secara sistematis dan kronologis dari awal hingga akhir. Terakhir, dijelaskan pula alat dan bahan yang digunakan untuk mengolah temuan penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan temuan-temuan utama penelitian secara objektif dan terstruktur. Hasil tersebut kemudian diinterpretasikan secara mendalam untuk menjelaskan maknanya terkait dengan pertanyaan penelitian.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari keseluruhan penelitian yang telah dilakukan, serta saran dari penulis untuk kegiatan penelitian selanjutnya terkait dengan topik yang sedang dibahas.