

BAB I

PENDAHULUAN

1.1. Latar Belakang

Digitalisasi merupakan proses mengubah elemen fisik atau nondigital ke dalam format digital yang membuka peluang baru bagi pelaku industri untuk memperluas pasar. Melalui pemanfaatan teknologi digital, produk dapat dipasarkan secara lebih luas dengan menjadikan platform digital sebagai saluran utama dalam menjangkau konsumen [1]. Pemanfaatan teknologi digital dalam aktivitas pemasaran ini disebut sebagai *digital marketing* [2]. Dibutuhkan kemampuan dalam membuat konten berkualitas melalui pemanfaatan analisis data secara mendalam agar strategi *digital marketing* ini dapat berjalan efektif. Dengan pendekatan *data science* tersebut, perusahaan dapat memperoleh wawasan mendalam untuk memaksimalkan efektivitas kampanye digital [3]. *Digital marketing* memanfaatkan beragam saluran digital untuk mempromosikan produk atau layanan, salah satunya melalui kampanye media sosial (*social media campaign*) [2].

Media sosial berpotensi menjadi sumber data besar (*big data*) yang dapat dimanfaatkan dalam pendekatan *digital marketing*. Selain itu, media sosial telah menjadi bagian tak terpisahkan dari kehidupan masyarakat modern, sehingga menjadikannya sarana strategis dalam *digital marketing* [4]. Salah satu platform media sosial yang dapat dimanfaatkan adalah X (sebelumnya dikenal sebagai Twitter), yang unggul dalam penyebaran informasi secara cepat dan *real-time* melalui fitur *retweet* dan *trending topic*, sehingga efektif meningkatkan visibilitas konten [5] [6]. Berdasarkan data *The Global Statistic*, X digunakan oleh 58,30% pengguna internet di Indonesia, menjadikannya platform keempat paling populer setelah Instagram, Facebook, dan TikTok [7]. Tingginya tingkat interaksi melalui fitur seperti *tweet*, *retweet*, dan *hashtag* menjadikan platform ini sebagai sarana potensial bagi perusahaan untuk membangun keterlibatan langsung dengan audiens [8].

Platform X terbukti efektif sebagai sarana *social media marketing* atau *social media campaign* dengan kemampuannya dalam menjangkau audiens secara luas dengan biaya relatif rendah [9]. Penelitian [10] menunjukkan bahwa percakapan dan interaksi di X, seperti pada kampanye #TokopediaWIB dan #ShopeeGajian, dapat menjadi indikator keberhasilan pemasaran karena berkontribusi dalam membangun *brand awareness* sekaligus meningkatkan keterlibatan konsumen. Efektivitas ini juga diperkuat dengan fitur-fitur khas X seperti *retweet*, *mention*, *hashtag*, dan *trending topic* yang mampu meningkatkan peluang konten untuk menjadi viral, sehingga pesan kampanye dapat menjangkau lebih banyak pengguna [10]. Dengan demikian, X menjadi salah satu platform strategis dalam mendukung efektivitas pemasaran digital berbasis data.

Namun, keberhasilan *social media campaign* di X tidak selalu mudah dicapai. Salah satu tantangan utamanya terletak pada penentuan isi konten *tweet* yang tepat, karena tidak semua konten yang dianggap menarik atau relevan mampu menjangkau audiens secara luas maupun menghasilkan interaksi tinggi. Pemilihan topik, penggunaan *hashtag*, serta pemilihan kata yang digunakan dalam *tweet* menjadi faktor penting yang memengaruhi ketertarikan audiens sejak awal [11]. Meskipun demikian, tingkat interaksi konten tetap sulit diprediksi, karena beberapa *tweet* dapat memperoleh ribuan interaksi dalam waktu singkat, sementara *tweet* lain dengan karakteristik serupa justru tidak mendapatkan respons yang signifikan [12]. Metode analisis data tradisional yang hanya mengandalkan statistik *engagement* dasar masih belum mampu memahami hubungan kompleks antara isi konten dan pola penyebarannya [13]. Oleh karena itu, dibutuhkan pendekatan berbasis data yang lebih canggih untuk mengidentifikasi keterkaitan antara karakteristik isi *tweet* dan pola penyebarannya, sehingga strategi *social media campaign* dapat dioptimalkan secara lebih efektif.

Pendekatan berbasis *Machine Learning* (ML) dan *Deep Learning* (DL) menawarkan solusi analitis yang mampu mengidentifikasi keterkaitan antara karakteristik isi *tweet* dan pola penyebarannya. ML merupakan kumpulan algoritma yang dirancang untuk mengidentifikasi pola dalam data dan membuat prediksi berdasarkan informasi yang dipelajari dari data historis [14]. Sementara itu, DL

merupakan cabang dari ML yang terinspirasi dari cara kerja sistem saraf manusia yang telah banyak diterapkan dalam berbagai bidang, salah satunya analisis tren media sosial [15]. Dengan memanfaatkan model ML dan DL, karakteristik isi tweet dan tingkat interaksinya dapat dianalisis lebih baik, sehingga membantu mengatasi keterbatasan metode statistik konvensional dalam penyusunan strategi konten.

Beberapa penelitian telah dilakukan dengan memanfaatkan pendekatan ML dan DL untuk menganalisis tren pada berbagai jenis data teks. Salah satunya adalah penelitian yang berjudul “*Forecasting topic trends of blockchain utilizing topic modeling and deep learning-based time-series prediction on different document types*” [16], berfokus pada peramalan tren topik di bidang *blockchain* dengan mengintegrasikan *topic modeling* (LDA, DMR), *clustering*, serta model *deep learning* berbasis deret waktu, baik *non-graph* (LSTM, GRU) maupun *graph-based* (AGCRN, A3T-GCN). Data yang digunakan mencakup 56.300 artikel akademik, 13.524 paten, dan 36.991 artikel berita terkait *blockchain* yang dipublikasikan 2017–2023. Hasil penelitian menunjukkan bahwa LSTM lebih baik untuk pola periodik, sementara AGCRN dan A3T-GCN unggul pada pola non-periodik. Penelitian ini menawarkan kerangka kerja yang *robust* untuk analisis tren *blockchain*, tetapi masih terbatas pada data artikel akademik, paten, dan berita, sehingga belum menyentuh data media sosial yang lebih dinamis.

Berdasarkan keterbatasan pada penelitian [16] tersebut, penelitian ini mengembangkan konsep *forecasting* tren topik dengan menerapkannya pada data media sosial X yang memiliki karakter lebih cepat berubah, lebih dinamis, dan sangat dipengaruhi oleh interaksi pengguna. Tidak hanya berfokus pada topik, penelitian ini juga memperluas objek *forecasting* pada *hashtag* dan kata yang sering muncul dalam percakapan, sehingga hasil prediksi yang diperoleh dapat dimanfaatkan sebagai dasar dalam menyusun konten *tweet* yang lebih relevan, sesuai dengan tren, dan berpotensi menjangkau audiens lebih luas. Dengan demikian, penelitian ini tidak sekadar memprediksi kecenderungan kemunculan topik percakapan, tetapi juga berupaya mengidentifikasi *hashtag* dan pilihan kata dalam konten yang berpotensi efektif digunakan untuk mendukung strategi pemasaran di media sosial.

Perluasan *forecasting* pada elemen *hashtag* tersebut juga didukung oleh penelitian [17] yang mengusulkan metode *hashTag Recommendation for Enhancing Social popularITy* (TRANSIT) untuk merekomendasikan *hashtag* yang relevan dan berpotensi meningkatkan popularitas *tweet*. Metode ini menggabungkan popularitas pengguna, *tweet*, dan kata kunci teks melalui pembentukan matriks AUFWP yang diproses menggunakan algoritma *Markov Chain*, dan hasil eksperimennya menunjukkan performa yang lebih baik dibandingkan metode *state-of-the-art*. Meskipun demikian, penelitian tersebut masih terbatas pada fokus penggunaan *hashtag* sebagai faktor utama popularitas tanpa mempertimbangkan keterkaitannya dengan topik percakapan dan kata dominan yang berkembang. Oleh karena itu, pada penelitian ini tidak hanya memanfaatkan *hashtag* sebagai elemen rekomendasi tunggal, tetapi menempatkannya sebagai bagian dari hasil *forecasting* bersama topik dan kata agar rancangan konten yang dihasilkan menjadi lebih lengkap dan kontekstual.

Selanjutnya, agar hasil *forecasting* topik, *hashtag*, dan kata yang diperoleh tidak hanya berhenti pada identifikasi tren, diperlukan pula analisis yang mampu menilai potensi performa *tweet* ketika elemen-elemen tersebut diterapkan dalam sebuah konten. Analisis ini berkaitan dengan penelitian berjudul “*Virality classification from Twitter data using pre-trained language model and multi-layer perceptron*” [18], yang berhasil mengklasifikasikan tingkat popularitas *tweet* berbahasa Indonesia menggunakan kombinasi analisis sentimen, deteksi emosi, *text embedding* BERT, metadata *tweet*, dan model *Multi-Layer Perceptron* (MLP). Penelitian tersebut menunjukkan bahwa pendekatan MLP efektif dalam memprediksi popularitas *tweet*, namun belum mencakup dimensi tren topik, *hashtag*, dan kata yang berkembang secara temporal. Berdasarkan hal tersebut, penelitian ini melengkapi hasil *forecasting* dengan analisis pendukung berupa klasifikasi tingkat interaksi *tweet* untuk memprediksi kelas viral dan tidak viral serta estimasi *engagement like*, *retweet*, *quote*, dan *reply*. Dengan kombinasi dua pendekatan ini, strategi *social media marketing campaign* yang disusun tidak hanya berbasis pada tren yang diperkirakan akan muncul, tetapi juga pada potensi respons interaksi yang mungkin dihasilkan oleh konten tersebut.

Berdasarkan uraian penelitian terdahulu tersebut, meskipun berbagai studi sebelumnya telah menerapkan pendekatan ML dan DL untuk menganalisis tren pada data teks, sejumlah keterbatasan masih ditemukan. Penelitian *forecasting* [16] memang berhasil memprediksi tren topik *blockchain*, namun tidak menangani konten berbahasa Indonesia dan belum mengintegrasikan karakteristik khusus media sosial. Penelitian rekomendasi *hashtag* TRANSIT [17] juga masih bergantung pada analisis popularitas berbasis *hashtag* semata, sehingga pemanfaatan aspek lain belum menjadi fokus utama dalam model untuk meningkatkan popularitas konten. Sementara itu, Penelitian klasifikasi viralitas [18] hanya menggunakan data *tweet* tahun 2021 dan belum mencakup dimensi tren topik, *hashtag*, dan kata yang berkembang secara temporal.

Berdasarkan keterbatasan tersebut, penelitian ini mengusulkan pendekatan *forecasting* frekuensi topik, *hashtag*, dan kata untuk memprediksi potensi peningkatan kemunculannya di masa depan. Model yang digunakan berbasis *transformer* dan ML, yaitu BERTopic (*transformer*) untuk ekstraksi topik [19] dan LightGBM sebagai model ML [20] untuk memprediksi tren deret waktu pada topik, *hashtag*, dan kata yang telah diproses. Hasil *forecasting* kemudian digunakan sebagai dasar dalam menyusun konten yang lebih tepat sasaran dengan mempertimbangkan topik, *hashtag*, dan kata yang berpotensi naik di masa depan, sehingga mampu mendukung optimalisasi strategi *social media marketing campaign* di platform X.

Sementara itu, sebagai analisis pendukung untuk mengoptimalkan strategi *social media marketing campaign*, penelitian ini juga memanfaatkan pendekatan klasifikasi tingkat interaksi *tweet* untuk memahami hubungan antara tren yang diprediksi dengan potensi performa konten. Pendekatan ini menggunakan model berbasis *transformer*, yaitu IndoBERTweet dan RoBERTa, untuk menghasilkan representasi teks [21] [22], yang kemudian digabungkan dengan fitur metadata *tweet* dan diproses menggunakan model *deep learning* [23] [24]. Hasil klasifikasi tersebut digunakan sebagai informasi tambahan dalam mengevaluasi potensi keberhasilan konten, sehingga dapat memperkuat strategi *social media marketing campaign* pada platform X.

Dengan demikian, penelitian ini memberikan alur yang terstruktur, mulai dari pengumpulan dan pengolahan data, pembangunan model, analisis hasil, hingga penyediaan antarmuka pengguna (*Graphical User Interface/GUI*), sehingga dapat mendukung pengambilan keputusan yang lebih tepat dalam strategi pemasaran digital. Kinerja dari model yang dibangun juga akan dianalisis berdasarkan metrik evaluasi yang sesuai untuk menilai tingkat akurasi dan keandalan model dalam melakukan prediksi. Implementasi model ini diharapkan mampu memberikan wawasan strategis mengenai topik, *hashtag*, dan kata kunci relevan untuk potensi perluasan jangkauan suatu kampanye secara signifikan. Informasi tersebut memungkinkan perusahaan merancang strategi pemasaran digital yang lebih efektif, adaptif, dan berbasis data, sehingga mampu meningkatkan visibilitas merek melalui pemanfaatan konten yang lebih tepat sasaran.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, terdapat beberapa permasalahan yang menjadi fokus dalam penelitian ini, yaitu:

1. Bagaimana proses pengumpulan dan pengolahan data dilakukan agar dapat digunakan sebagai *input* pada model *forecasting* frekuensi topik, *hashtag*, dan kata di platform X?
2. Bagaimana melakukan *forecasting* frekuensi topik, *hashtag*, dan kata di platform X dengan memanfaatkan BERTopic untuk *topic modeling* serta LightGBM untuk peramalan deret waktu?
3. Bagaimana membangun model klasifikasi tingkat interaksi *tweet* sebagai analisis pendukung untuk optimalisasi *social media marketing campaign* di platform X?
4. Bagaimana strategi *social media marketing campaign* di platform X dapat dioptimalkan melalui integrasi hasil *forecasting* frekuensi topik, *hashtag*, dan kata yang didukung oleh analisis tingkat interaksi *tweet*?
5. Bagaimana tingkat akurasi dan kinerja model yang diusulkan dalam menghasilkan prediksi untuk optimalisasi strategi kampanye pemasaran digital berbasis data?

1.3. Batasan Masalah

Agar penelitian lebih terfokus, penelitian ini memiliki beberapa batasan sebagai berikut:

1. Data yang digunakan dalam penelitian ini berupa *tweet* berbahasa Indonesia yang dikumpulkan menggunakan metode *scraping* dan pengumpulan manual pada platform X, dengan rentang waktu mulai 27 Mei 2024 hingga 27 November 2025. Pengumpulan data difokuskan pada *keyword* yang berkaitan dengan promosi, *sale*, diskon, dan *flash sale*.
2. *Forecasting* frekuensi topik, *hashtag*, dan kata pada penelitian ini terbatas pada frekuensi kemunculan, bukan prediksi isi konten atau makna semantik.
3. Pada analisis pendukung klasifikasi tingkat interaksi *tweet*, pelabelan tingkat interaksi *tweet* didasarkan pada jumlah *retweet*, *like*, *reply*, *quote*, *followers*, *followings* dan *verified account*. Klasifikasi dilakukan dengan beberapa metrik seperti teks *tweet*, jumlah *followers*, *following*, *hashtags*, *mentions*, status *verified account*, panjang teks *tweet*, dan sentimen teks *tweet*.
4. Penelitian ini berfokus pada konteks kampanye *digital marketing* di platform X.
5. Penelitian ini berasumsi bahwa tingkat interaksi yang didapatkan tidak disebabkan oleh aktivitas *buzzer* atau akun yang dimobilisasi secara artifisial.

1.4. Tujuan Penelitian

Tujuan utama dari penelitian ini adalah untuk mengoptimalkan *social media marketing campaign* di platform X melalui pendekatan *data science*. Untuk mendukung pencapaian tujuan utama tersebut, penelitian ini memiliki beberapa tujuan pendukung, yaitu:

1. Merancang dan menganalisis proses pengumpulan serta pengolahan data agar dapat digunakan sebagai *input* dalam model *forecasting* frekuensi topik, *hashtag*, dan kata di platform X.
2. Mengembangkan model *forecasting* frekuensi topik, *hashtag*, dan kata di platform X menggunakan BERTopic untuk *topic modeling* serta LightGBM untuk peramalan deret waktu.

3. Membangun model klasifikasi tingkat interaksi *tweet* berbahasa Indonesia dengan mengintegrasikan representasi teks, analisis sentimen, dan metadata *tweet* sebagai analisis pendukung optimalisasi *social media marketing campaign* di platform X.
4. Mengintegrasikan hasil *forecasting* frekuensi topik, *hashtag*, dan kata guna memberikan *insight* berbasis data yang didukung dengan analisis tingkat interaksi *tweet* sehingga mengoptimalkan strategi *social media campaign* dalam *digital marketing*.
5. Mengevaluasi performa model yang diusulkan dalam menghasilkan prediksi untuk optimalisasi strategi kampanye pemasaran digital berbasis data.

1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat dalam berbagai aspek, baik akademik maupun non-akademik. Adapun beberapa manfaat dari penelitian ini, di antaranya:

1. Menambah literatur mengenai *forecasting* frekuensi topik, *hashtag*, dan kata pada media sosial.
2. Menjadi referensi bagi penelitian selanjutnya yang ingin mengembangkan model serupa pada bahasa Indonesia maupun platform media sosial lainnya.
3. Memberikan *insight* bagi perusahaan dan pelaku bisnis dalam merancang strategi *social media marketing campaign* yang lebih efektif dan tepat sasaran.
4. Membantu meningkatkan *engagement* audiens, memperkuat citra merek, serta mengevaluasi efektivitas kampanye pemasaran digital secara lebih akurat dan berbasis data.
5. Menjadi acuan dalam penyusunan strategi konten yang lebih adaptif terhadap tren yang berkembang di media sosial.