

BAB II

LANDASAN TEORI

2.1 Tinjauan Pustaka

2.1.1 Analisis Sentimen

Analisis sentimen adalah suatu pendekatan dalam pemrosesan bahasa alami yang bertujuan mengidentifikasi dan mengukur sikap, emosi, atau pendapat orang terhadap suatu entitas atau topik melalui teks yang mereka tulis. Secara lebih spesifik, analisis ini mencoba mengklasifikasikan teks ke dalam kategori-kategori seperti positif, negatif, dan netral, dengan harapan memahami bagaimana opini publik terbentuk dan tersebar dalam ruang digital. Menurut (Mohamed et al., 2022) analisis sentimen memiliki peran penting dalam membantu organisasi memahami opini publik secara langsung dan mengambil keputusan strategis berdasarkan isu yang sedang berkembang. Dalam konteks penelitian ini, analisis sentimen dimanfaatkan untuk menilai pandangan masyarakat terhadap pembatalan pembelian BBM dari Pertamina oleh SPBU swasta, yang tercermin melalui berbagai komentar dan unggahan di platform media sosial Instagram dan X.

2.1.1.1 Pentingnya Analisis Sentimen di Media Sosial dan Ulasan Daring

Media sosial dan platform ulasan daring telah menjadi sumber informasi yang sangat berpengaruh dalam membentuk opini publik terhadap suatu isu, produk, maupun kebijakan. Analisis sentimen pada media sosial memungkinkan peneliti dan organisasi untuk memahami pola emosi serta persepsi masyarakat secara lebih cepat dan mendalam. Melalui analisis ini, data dalam bentuk teks

seperti komentar, cuitan, atau ulasan dapat dikonversi menjadi informasi bernilai yang menggambarkan kecenderungan opini publik, baik positif, negatif, maupun netral. (Matulessy et al., 2023) analisis sentimen terhadap opini pengguna di media sosial memiliki peran penting dalam memahami pandangan masyarakat terhadap suatu kebijakan atau lembaga. Dalam konteks penelitian ini, analisis sentimen diterapkan untuk menilai reaksi publik terhadap pembatalan pembelian BBM dari Pertamina oleh SPBU swasta yang diungkapkan melalui komentar dan diskusi di platform Instagram dan X. Melalui pemahaman terhadap opini tersebut, pihak terkait dapat mengidentifikasi respons positif maupun negatif dari masyarakat serta menyusun strategi komunikasi yang lebih efektif untuk menjaga kepercayaan publik dan memperbaiki citra lembaga terkait. (Matulessy et al., 2023)

2.1.1.2 Teknik dalam Analisis Sentimen

Teknik analisis sentimen mencakup serangkaian tahapan yang dimulai dari pra-pemrosesan teks, ekstraksi fitur, hingga klasifikasi sentimen. Tahap pra-pemrosesan dilakukan untuk membersihkan data dari elemen yang tidak relevan seperti tanda baca, URL, dan emoji, serta melakukan tokenisasi dan normalisasi teks agar siap dianalisis. Selanjutnya, tahap ekstraksi fitur bertujuan untuk mengubah teks menjadi representasi numerik menggunakan pendekatan seperti *Term Frequency–Inverse Document Frequency (TF-IDF)* atau representasi vektor berbasis *word embedding*. Menurut (Saragih et al., 2024) penerapan BERT pada analisis sentimen media sosial memberikan akurasi yang lebih tinggi dibandingkan model pembelajaran mesin konvensional.

2.1.2 Machine Learning dalam Analisis Sentimen

Machine Learning (ML) merupakan cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada pengembangan algoritma yang mampu belajar dari data dan membuat prediksi atau keputusan tanpa harus diprogram secara eksplisit. Dalam konteks analisis sentimen, metode ML digunakan untuk mengklasifikasikan opini publik ke dalam kategori seperti positif, negatif, atau netral berdasarkan pola yang ditemukan dalam teks. Menurut (Nikhil Sanjay Suryawanshi, 2024) algoritma *machine learning* seperti Naive Bayes, Support Vector Machines (SVM), dan Deep Learning banyak digunakan dalam analisis sentimen karena kemampuannya dalam mengklasifikasikan teks dalam jumlah besar secara efisien. Naive Bayes dikenal unggul karena kesederhanaannya serta efektivitasnya dalam mengolah data teks berskala besar. Dalam konteks penelitian ini, penerapan *machine learning* digunakan untuk memproses dan mengklasifikasikan berbagai komentar dan unggahan di Instagram maupun X (Twitter) terkait isu pembatalan pembelian BBM dari Pertamina oleh SPBU swasta secara otomatis, sehingga dapat memberikan gambaran umum mengenai sentimen publik terhadap kebijakan tersebut.

2.1.3 Data Mining dalam Analisis Sentimen

Data mining merupakan proses untuk menemukan pola, hubungan, dan informasi yang bernilai dari kumpulan data dalam jumlah besar. Dalam konteks analisis sentimen, data mining berfungsi untuk mengekstraksi dan mengidentifikasi opini publik dari data teks yang tidak terstruktur, seperti unggahan dan komentar di media sosial. Menurut (Norlaila et al., 2024) data mining memiliki peran penting

dalam menggali pola dan kecenderungan dari data tidak terstruktur yang bersumber dari media sosial, karena teknik seperti *clustering* dan *classification* mampu mengelompokkan opini menjadi kategori positif, negatif, dan netral secara efisien.

Dalam penelitian ini, data mining diterapkan untuk menganalisis sentimen publik terhadap pembatalan pembelian BBM dari Pertamina oleh SPBU swasta, dengan memanfaatkan data komentar dan percakapan dari platform Instagram dan X (Twitter). Melalui penerapan teknik data mining, pola opini masyarakat dapat diidentifikasi secara sistematis, sehingga memberikan pemahaman yang lebih mendalam tentang bagaimana persepsi publik terbentuk terhadap isu tersebut.

2.1.4 Prediksi dalam Analisis Sentimen

Prediksi dalam analisis sentimen merupakan tahap lanjutan yang bertujuan untuk memperkirakan kecenderungan atau arah opini publik berdasarkan data teks yang telah dianalisis sebelumnya. Proses ini melibatkan penggunaan model pembelajaran mesin atau pembelajaran mendalam yang telah dilatih untuk mengenali pola sentimen dalam data. prediksi sentimen memungkinkan peneliti untuk mengidentifikasi bagaimana persepsi publik dapat berubah terhadap suatu peristiwa, kebijakan, atau entitas tertentu berdasarkan data historis dan tren yang berkembang di media sosial. Dalam konteks analisis sentimen berbasis media sosial, model prediksi tidak hanya berfungsi untuk menentukan polaritas sentimen (positif, negatif, atau netral), tetapi juga untuk memprediksi arah perubahan opini publik dari waktu ke waktu. Model seperti BERT (Bidirectional Encoder Representations from Transformers)

2.1.5 Metode BERT dalam Analisis Sentimen

Model BERT (Bidirectional Encoder Representations from Transformers) merupakan salah satu terobosan besar dalam bidang Natural Language Processing (NLP) yang di kembangkan pada tahun 2018 dan dipublikasikan pada tahun 2019 oleh Google AI.(Devlin et al., 2019) Secara teoritis, BERT dibangun berdasarkan arsitektur Transformer yang menekankan mekanisme self-attention untuk memahami hubungan antar kata dalam sebuah kalimat secara efektif. Keunggulan utama BERT terletak pada kemampuannya memahami konteks secara dua arah (*bidirectional*), sehingga model dapat menangkap makna kata berdasarkan konteks sebelum dan sesudahnya secara simultan.

Secara teknis, cara kerja BERT melibatkan dua tahap utama: (1) Pre-training dan (2) Fine-tuning.

Pada tahap pre-training, BERT dilatih menggunakan dua mekanisme penting, yaitu Masked Language Modeling (MLM) dan Next Sentence Prediction (NSP). MLM bekerja dengan menyembunyikan sebagian kata dalam kalimat dan meminta model memprediksi kata yang hilang berdasarkan konteks di sekitarnya. Sementara itu, NSP melatih model untuk memahami hubungan antar dua kalimat, apakah kalimat kedua merupakan kelanjutan dari kalimat pertama. Tahap ini membuat BERT mampu memahami struktur semantik, relasi antarkalimat, serta konteks linguistik secara mendalam.

Setelah proses pre-training, model BERT dapat di-fine-tune pada berbagai tugas spesifik, termasuk klasifikasi teks dan analisis sentimen. Pada tugas analisis sentimen, BERT memproses teks masukan melalui lapisan *encoder* untuk

menghasilkan representasi vektor kontekstual, yang kemudian diklasifikasikan menjadi kategori sentimen seperti positif, negatif, atau netral. Berbagai studi menunjukkan bahwa BERT mampu memberikan akurasi yang lebih tinggi dibandingkan model NLP generasi sebelumnya.(Devlin et al., 2019) membuktikan bahwa BERT melampaui model-model terdahulu pada sejumlah benchmark NLP seperti GLUE, SQuAD, dan analisis sentimen, sehingga menjadikannya salah satu fondasi utama dalam pengembangan model bahasa modern.

2.1.6 CRISP-DM (Cross Industry Standard Process for Data Mining)

Metodologi CRISP-DM (Cross-Industry Standard Process for Data Mining) merupakan pendekatan standar yang digunakan untuk mengelola dan mengembangkan proyek *data mining* secara sistematis. Menurut (Laturiuw et al., 2023) , CRISP-DM terdiri dari enam tahapan utama, yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan. Dalam konteks penelitian “Analisis Sentimen Publik di Instagram dan X terhadap Pembatalan Pembelian BBM dari Pertamina oleh SPBU Swasta menggunakan Metode BERT”, metodologi ini berfungsi sebagai panduan untuk mengatur alur penelitian mulai dari proses pengumpulan, pembersihan, hingga analisis dan pemodelan data teks. Tahap pemahaman bisnis berfokus pada identifikasi tujuan utama penelitian, yaitu menganalisis persepsi publik terhadap isu pembatalan pembelian BBM. Tahap pemahaman data dilakukan dengan mengeksplorasi komentar dan unggahan publik dari Instagram serta X untuk menemukan pola opini masyarakat. Selanjutnya, pada tahap persiapan data, dilakukan proses *cleaning*, *tokenization*, dan *labeling* sebelum masuk ke tahap pemodelan yang menggunakan

algoritma BERT (Bidirectional Encoder Representations from Transformers) untuk melakukan klasifikasi sentimen positif, negatif, dan netral. Setelah model dibangun, tahap evaluasi digunakan untuk menilai performa model melalui metrik seperti akurasi, presisi, dan F1-score, sedangkan tahap penerapan digunakan untuk mengimplementasikan hasil model sebagai dasar dalam memahami opini publik secara lebih mendalam dan berbasis data.

2.1.7 Pemodelan UML

Unified Modeling Language (UML) merupakan bahasa pemodelan standar yang digunakan untuk mempermudah proses pengembangan perangkat lunak, terutama pada tahap analisis dan perancangan sistem. UML membantu pengembang menggambarkan sistem secara visual melalui berbagai jenis diagram yang menampilkan struktur serta perilaku dari sistem yang akan dibangun.

Menurut (John, 2025) , diagram Unified Modeling Language (UML) memiliki peran penting dalam proses pengembangan sistem berbasis data dan kecerdasan buatan, termasuk pada penelitian analisis sentimen. Dalam konteks penelitian “Analisis Sentimen Publik di Instagram dan X terhadap Pembatalan Pembelian BBM dari Pertamina oleh SPBU Swasta Menggunakan Metode BERT,” UML digunakan untuk memvisualisasikan alur kerja sistem, mulai dari proses pengumpulan data, praproses teks, pelatihan model BERT, hingga analisis hasil sentimen. Diagram seperti use case, activity, class, dan sequence membantu menggambarkan interaksi antar komponen sistem, mempermudah pemahaman alur analisis, serta mendukung pengembangan sistem yang terstruktur dan efisien.

Menurut (Deshmukh et al., n.d.) , penerapan Unified Modeling Language (UML) juga berperan penting dalam menjaga konsistensi serta efisiensi pengembangan sistem analisis sentimen. Dokumentasi yang jelas melalui diagram UML, seperti class diagram dan sequence diagram, dapat membantu pengembang memahami struktur dan alur sistem dengan lebih baik, sehingga mengurangi potensi kesalahan dalam implementasi serta mempercepat proses pemeliharaan. Dengan adanya model UML yang terorganisir, sistem analisis sentimen berbasis BERT dapat dikembangkan secara lebih efisien, mudah dipelihara, dan memiliki tingkat keandalan yang tinggi dalam menganalisis opini publik di media sosial.

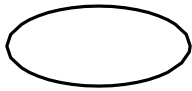
2.1.7.1 Use Case

Use Case Diagram berfungsi untuk memvisualisasikan hubungan antara aktor, baik itu pengguna maupun sistem eksternal, dengan sistem yang sedang dibangun. Diagram ini menggambarkan berbagai skenario interaksi yang dilakukan pengguna untuk mencapai tujuan atau fungsi tertentu dalam sistem secara lebih terstruktur dan mudah dipahami.

Menurut (Refianti et al., 2023) , use case diagram berperan dalam memperlihatkan fungsi utama dari suatu sistem beserta interaksi yang terjadi antara sistem dan aktor yang terlibat. Diagram ini membantu pengembang dalam memahami serta mengidentifikasi kebutuhan sistem berdasarkan sudut pandang pengguna, sehingga rancangan yang dihasilkan dapat lebih sesuai dengan tujuan dan alur penggunaan sebenarnya.

Dalam penelitian ini, use case diagram digunakan untuk memodelkan interaksi antara pengguna sistem, seperti admin dan publik, dengan berbagai fitur utama yang dikembangkan. Diagram ini menggambarkan alur aktivitas mulai dari proses pengambilan data komentar dari Instagram dan X (Twitter), pra-pemrosesan teks, klasifikasi sentimen menggunakan metode BERT, hingga penyajian hasil analisis sentimen terkait isu pembatalan pembelian BBM dari Pertamina oleh SPBU swasta. Use case diagram ini membantu memvisualisasikan fungsi sistem secara menyeluruh dan memperjelas hubungan antara pengguna dengan setiap proses utama yang ada di dalam sistem. Berikut adalah komponen yang terdapat pada Use Case Diagram:

Table 2. 1 Komponen Usecase Diagram (Refianti et al., 2023)

Nama Komponen	Kegunaan / Fungsi	Notasi / Simbol
<i>Actor</i>	<i>Actor</i> berfungsi sebagai titik awal dalam mendefinisikan bagaimana dan kapan suatu fungsi atau proses harus dijalankan.	
<i>Use Case</i>	Berfungsi untuk mendefinisikan dan menggambarkan interaksi antara pengguna (<i>actor</i>) dan sistem yang akan dirancang, serta bagaimana sistem tersebut akan memenuhi kebutuhan pengguna.	

<i>Association</i>	Berfungsi untuk menggambarkan hubungan antara <i>actor</i> dan <i>use case</i> , menunjukkan interaksi yang terjadi antara pengguna atau sistem eksternal dengan sistem yang sedang dikembangkan.	
<i>Generalization atau inheritance (pewarisan)</i>	Berfungsi untuk menggambarkan hubungan hierarkis antara elemen yang lebih umum dan elemen yang lebih spesifik.	
<i>Dependency</i>	Berfungsi untuk menggambarkan hubungan antara dua elemen di mana satu elemen (misalnya, <i>use case</i> atau <i>actor</i>) bergantung atau membutuhkan elemen lain untuk berfungsi dengan baik.	
<i>Include</i>	Berfungsi untuk menunjukkan bahwa satu <i>use case</i> selalu melibatkan atau menyertakan eksekusi <i>use case</i> lainnya sebagai bagian dari alurnya.	<<include>>
<i>Extend</i>	Berfungsi untuk menunjukkan bahwa sebuah <i>use case</i> dapat diperluas atau ditambah dengan fungsionalitas tambahan dalam kondisi tertentu.	<<extend>>
<i>Communicates</i>	Berfungsi untuk menunjukkan interaksi atau pertukaran informasi antara dua elemen, seperti antara dua <i>actor</i> atau antara <i>actor</i> dan sistem.	<<communicates>>

2.1.7.2 Activity

Activity Diagram berfungsi untuk memvisualisasikan urutan aktivitas atau proses bisnis yang terjadi dalam suatu sistem, dengan menampilkan langkah-langkah logis dari awal hingga akhir proses. Diagram ini membantu menggambarkan bagaimana alur kerja berjalan, termasuk tahapan pengambilan keputusan maupun aktivitas yang berlangsung secara bersamaan. Activity Diagram merupakan bentuk representasi grafis yang menunjukkan aliran aktivitas secara

berurutan, sehingga mempermudah pemahaman terhadap keseluruhan proses dan logika sistem yang dikembangkan.(Umesh Awate et al., 2024)

Dalam konteks penelitian ini, Activity Diagram digunakan untuk menggambarkan alur proses analisis sentimen publik terkait pembatalan pembelian BBM dari Pertamina oleh SPBU swasta, mulai dari pengambilan data komentar di Instagram dan X (Twitter), tahap pra-pemrosesan teks, pemodelan dan klasifikasi sentimen menggunakan metode BERT, hingga penyimpanan serta visualisasi hasil analisis. Diagram ini membantu memperjelas setiap langkah proses yang dilakukan sistem secara berurutan, sehingga memudahkan pemahaman terhadap alur kerja keseluruhan dan keterkaitan antaraktivitas dalam sistem analisis sentimen yang dikembangkan. Berikut komponen yang terdapat dalam Activity Diagram:

Table 2. 2 Komponen Activity Diagram(Umesh Awate et al., 2024)

Nama Komponen	Kegunaan / Fungsi	Notasi / Simbol
<i>Start State</i>	Berfungsi untuk menunjukkan keadaan awal dalam sebuah diagram status (<i>state diagram</i>) yang menggambarkan titik awal dari proses atau alur dalam sistem.	
<i>State</i>	Berfungsi untuk menggambarkan kondisi atau situasi tertentu dalam siklus hidup suatu objek atau entitas dalam sistem, yang dapat berubah berdasarkan interaksi atau peristiwa tertentu.	
<i>Activity</i>	Berfungsi untuk menggambarkan kondisi atau situasi tertentu dalam siklus hidup suatu objek atau	

	entitas dalam sistem, yang dapat berubah berdasarkan interaksi atau peristiwa tertentu.	
<i>State Transition</i>	Berfungsi untuk menggambarkan perubahan dari satu <i>state</i> (keadaan) ke <i>state</i> lainnya dalam sistem, berdasarkan peristiwa atau kondisi tertentu.	
<i>Transition to Self</i>	Berfungsi untuk menggambarkan perubahan atau transisi dari satu <i>state</i> ke <i>state</i> yang sama dalam diagram status (<i>state diagram</i>).	
<i>Horizontal Synchronization</i>	Biasanya merujuk pada koordinasi atau penyelarasan elemen atau proses dalam satu baris atau jalur yang sejajar secara horizontal.	
<i>Vertical Synchronization</i>	Berfungsi untuk menggambarkan sinkronisasi alur aktivitas yang terjadi secara bersamaan dalam urutan vertikal.	
<i>Decision</i>	Berfungsi untuk menggambarkan titik di mana suatu alur aktivitas bercabang berdasarkan kondisi tertentu.	
<i>End State</i>	Berfungsi untuk menunjukkan titik akhir dari suatu alur aktivitas atau proses dalam sistem.	

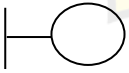
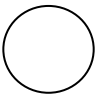
2.1.7.3 Sequence Diagram

Sequence diagram merupakan salah satu jenis diagram dalam UML yang digunakan untuk menunjukkan urutan interaksi antar objek atau komponen di

dalam sistem berdasarkan aliran waktu. Diagram ini menjelaskan bagaimana objek saling berkomunikasi dan bertukar pesan untuk menjalankan suatu proses atau fungsi tertentu. Menurut (Al-Fedaghi, n.d.) , sequence diagram termasuk salah satu diagram UML yang paling sering digunakan karena mampu memperlihatkan hubungan dinamis antar objek serta urutan pesan yang terjadi selama sistem beroperasi.

Secara umum, Sequence Diagram digunakan untuk memodelkan urutan interaksi antara aktor dengan sistem, maupun antar komponen di dalam sistem yang memiliki struktur lebih kompleks. Diagram ini berfungsi untuk mempermudah visualisasi proses yang melibatkan banyak entitas dan menampilkan urutan komunikasi di antara mereka secara terperinci dan sistematis.

Table 2. 3 Komponen Sequence Diagram (Al-Fedaghi, n.d.)

SIMBOL	NAMA	KETERANGAN
	Aktor	Menggambarkan seseorang atau sesuatu yang berinteraksi dengan sistem
	<i>Interface/ Boundary Object</i>	Menangani komunikasi antar lingkungan sistem
	<i>Entity Object</i>	Gambaran sistem sebagai landasan dalam menyusun basis data
	<i>Control Object</i>	Komponen yang mengatur dan menjadwalkan aktivitas dan elemen lainnya

	<i>Object Message</i>	Menggambarkan pesan atau hubungan antar objek yang menunjukkan urutan kejadian yang terjadi.
	<i>Message to Self</i>	Menggambarkan pesan atau hubungan objek itu sendiri.
	<i>Lifeline</i>	Garis titik-titik yang terhubung dengan objek, di sepanjang lifeline terdapat aktivasi.
	<i>Activation</i>	Mewakili pelaksanaan operasi dari objek, panjang kotak ini sebanding dengan durasi aktivitas suatu operasi.

2.1.8 Database dan DBMS

Database merupakan sekumpulan data yang saling berkaitan dan diatur secara sistematis agar dapat disimpan, diakses, serta dikelola dengan efisien. Menurut (Purnaningsih et al., 2022) basis data adalah himpunan data yang tersusun dengan rapi dan terorganisir, yang berfungsi untuk memenuhi kebutuhan informasi dalam suatu organisasi secara efektif.

Database Management System (DBMS) merupakan perangkat lunak yang berfungsi untuk mengelola, mengatur, dan memelihara basis data agar data dapat diakses dengan aman, efisien, serta terstruktur. Menurut (Kalsum Siregar et al., 2024) DBMS adalah sistem perangkat lunak yang dirancang untuk mengatur dan mengelola data dalam suatu database, sehingga memudahkan pengguna dalam proses penyimpanan, pengambilan, dan pengolahan informasi.

2.1.9 Software Dan Pemrograman Terkait

Pengembangan sistem analisis sentimen modern sangat bergantung pada kombinasi perangkat lunak (software) dan teknik pemrograman (programming) yang tepat. Pada tahap pengolahan data teks dari media sosial seperti Instagram dan X, software-analitik seperti pustaka NLP (misalnya NLTK, spaCy) serta kerangka deep-learning (misalnya PyTorch atau TensorFlow) digunakan bersama pemrograman Python untuk memfasilitasi tahapan pembersihan data, tokenisasi, dan pelabelan. Selanjutnya, model canggih seperti BERT (Bidirectional Encoder Representations from Transformers) di-fine-tune menggunakan dataset lokal agar sesuai dengan karakter bahasa pengguna Indonesia dan konteks isu yang dikaji. (Saragih et al., 2024)

2.1.9.1 Bahasa Pemrograman Python dan Library Terkait

Bahasa pemrograman Python telah menjadi pilihan utama dalam pengembangan sistem analisis sentimen karena fleksibilitasnya serta dukungan pustaka (*library*) yang kaya untuk pemrosesan bahasa alami dan pembelajaran mesin. Beberapa pustaka seperti TextBlob, NLTK, dan kerangka kerja seperti TensorFlow atau PyTorch memungkinkan proses *pre-processing* teks (cleaning, tokenisasi, normalisasi) serta penerapan model klasifikasi sentimen secara cepat dan efisien. Menurut (Sadam et al., 2024) secara khusus menunjukkan bahwa Python digunakan untuk pendidikan dan praktik analisis teks dalam konteks bahasa Indonesia, yang menegaskan bahwa pemilihan Python dan library-nya sangat relevan dalam penelitian-penelitian analisis sentimen lokal.

2.1.9.2 Library Scikit-learn

Library Scikit-learn merupakan salah satu pustaka Python yang banyak digunakan dalam pengembangan model machine learning, termasuk dalam analisis sentimen. Scikit-learn menyediakan berbagai algoritma pembelajaran mesin seperti klasifikasi, regresi, dan clustering yang mudah diimplementasikan. Dalam konteks analisis sentimen, library ini berperan penting dalam proses pelatihan dan evaluasi model, seperti Support Vector Machine (SVM), Naive Bayes, dan Logistic Regression, serta mendukung pengolahan data teks dengan teknik *feature extraction* seperti TF-IDF dan CountVectorizer. Menurut (Aminudin et al., 2024) Scikit-learn digunakan untuk membangun model analisis sentimen pada aplikasi New Sakpole dengan mengombinasikan pendekatan Natural Language Processing (NLP) guna mengukur tingkat kepuasan pengguna terhadap sistem, sehingga membantu dalam pengambilan keputusan berbasis data.

2.1.9.3 Bert

Model BERT (Bidirectional Encoder Representations from Transformers) merupakan salah satu inovasi penting dalam bidang *Natural Language Processing (NLP)* yang dikembangkan oleh Google pada tahun 2018 dan dipublikasikan secara resmi pada tahun 2019. BERT dirancang untuk memahami konteks bahasa secara dua arah (*bidirectional*), di mana model membaca teks dari kiri ke kanan dan dari kanan ke kiri secara bersamaan. Pendekatan ini membuat BERT mampu memahami makna kata berdasarkan keseluruhan kalimat, bukan hanya dari urutan kata tertentu. Secara umum, BERT dilatih menggunakan dua mekanisme utama, yaitu Masked Language Modeling (MLM) yang bertujuan memprediksi kata yang disembunyikan

dalam sebuah kalimat, serta Next Sentence Prediction (NSP) yang berfungsi mempelajari hubungan antar kalimat. Kemampuan ini menjadikan BERT unggul dalam tugas-tugas NLP seperti klasifikasi teks, ekstraksi informasi, analisis sentimen, dan pemahaman bahasa.

2.2.1 Paper 1 : Analisis Sentimen terhadap Game Genshin Impact

Menggunakan BERT

Penelitian yang dilakukan oleh Ryo Kusnadi, Yusuf, Andriantony, Richard Ardian Yaputra, dan Melna Caintan (2021) dengan judul *“Analisis Sentimen terhadap Game Genshin Impact Menggunakan BERT”* yang dipublikasikan dalam RABIT: Jurnal Teknologi dan Sistem Informasi Universitas Internasional Batam, membahas penerapan metode Bidirectional Encoder Representations from Transformers (BERT) dalam mengklasifikasikan sentimen pengguna terhadap permainan Genshin Impact berdasarkan ulasan yang diperoleh dari Google Play Store.

Latar belakang penelitian ini berangkat dari meningkatnya aktivitas pengguna internet dan media sosial yang menghasilkan data dalam jumlah besar, termasuk ulasan terhadap aplikasi dan game. Kondisi tersebut mendorong perlunya analisis sentimen otomatis agar perusahaan pengembang dapat memahami opini dan kepuasan pengguna secara efisien. Tujuan penelitian ini adalah untuk mengidentifikasi kategori sentimen positif, negatif, atau netral pada ulasan pengguna guna membantu pengembang dalam menentukan arah pengembangan game selanjutnya.

Metode penelitian menggunakan pendekatan CRISP-DM (Cross Industry Standard Process for Data Mining) yang mencakup tahapan pemahaman bisnis, pengumpulan data melalui teknik *web scraping*, pra-pemrosesan data, pemodelan menggunakan BERT, evaluasi model, hingga implementasi (*deployment*). Model BERT dipilih karena kemampuannya memahami konteks dua arah dari teks, yang membuatnya unggul dibandingkan metode konvensional seperti Naïve Bayes. Proses pelatihan dilakukan menggunakan parameter *batch size* 32, *learning rate* $2e-5$, dan 10 *epochs*, dengan hasil evaluasi menunjukkan nilai akurasi yang tinggi — *precision* sebesar 0.79 untuk sentimen negatif, 0.77 untuk positif, dan 0.70 untuk netral, serta nilai ROC AUC sebesar 0.88.

Penelitian ini relevan dengan skripsi ini karena sama-sama menerapkan teknik analisis sentimen berbasis BERT untuk memahami persepsi publik dari data teks. Perbedaannya terletak pada konteks dan sumber data, di mana penelitian Kusnadi dkk. berfokus pada ulasan aplikasi di Google Play Store, sedangkan skripsi ini menggunakan data opini publik dari Instagram dan X (Twitter) terkait pembatalan pembelian BBM dari Pertamina oleh SPBU swasta. Namun demikian, penelitian tersebut menjadi rujukan penting karena membuktikan efektivitas BERT dalam menangani data teks berbahasa alami dan menghasilkan klasifikasi sentimen yang akurat dalam mendukung pengambilan keputusan berbasis opini publik.

2.2.1.1 Tujuan Penelitian

Tujuan utama dari penelitian “Analisis Sentimen terhadap Game Genshin Impact Menggunakan BERT” yang dilakukan oleh Ryo Kusnadi dkk. (2021) adalah untuk mengklasifikasikan sentimen pengguna terhadap permainan Genshin Impact

berdasarkan ulasan di Google Play Store menggunakan metode Bidirectional Encoder Representations from Transformers (BERT) guna mengetahui persepsi pengguna dalam tiga kategori, yaitu positif, negatif, dan netral. Dalam konteks skripsi ini yang berjudul “Analisis Sentimen Publik di Instagram dan X terhadap Pembatalan Pembelian BBM dari Pertamina oleh SPBU Swasta Menggunakan Metode BERT,” tujuan penelitian tersebut menjadi acuan dalam menunjukkan efektivitas BERT dalam menganalisis sentimen teks berbahasa alami secara otomatis serta membuktikan kemampuannya dalam mengklasifikasikan opini publik secara akurat berdasarkan konteks kalimat. Selain itu, penelitian ini menjadi referensi dalam penerapan tahapan CRISP-DM mulai dari pengumpulan, pra-pemrosesan, pemodelan, hingga evaluasi yang dapat diadaptasi untuk memahami persepsi masyarakat terhadap kebijakan ekonomi dan energi. Dengan demikian, penelitian Ryo Kusnadi dkk. memberikan dasar yang kuat bagi pengembangan model analisis sentimen berbasis BERT untuk menilai opini publik secara objektif dan komprehensif, terutama dalam isu strategis seperti pembatalan pembelian BBM oleh SPBU swasta.

2.2.1.2 Metodologi Yang Digunakan

Penelitian yang dilakukan oleh Ryo Kusnadi dkk. (2021) menggunakan metode CRISP-DM (Cross Industry Standard Process for Data Mining) dalam melakukan analisis sentimen terhadap game *Genshin Impact* menggunakan BERT (Bidirectional Encoder Representations from Transformers). Metode ini mencakup beberapa tahapan, mulai dari memahami tujuan penelitian, mengumpulkan data ulasan pengguna dari Google Play Store melalui proses *scraping*, hingga

menyiapkan data dengan cara membersihkan teks, menghapus duplikasi, dan mengelompokkan sentimen menjadi positif, netral, dan negatif. Data yang telah diproses kemudian dimasukkan ke dalam model BERT untuk dilatih menggunakan *AdamW optimizer* dengan pengaturan *batch size* 32, *learning rate* $2e-5$, dan 10 *epochs*. Model BERT digunakan karena mampu memahami konteks bahasa dari dua arah sehingga hasil klasifikasi sentimen menjadi lebih akurat. Evaluasi dilakukan dengan melihat nilai akurasi, *precision*, dan *recall* untuk menilai seberapa baik model mengenali setiap jenis sentimen. Dalam konteks skripsi ini, metodologi tersebut menjadi acuan penting karena memberikan gambaran yang jelas tentang langkah-langkah penerapan BERT untuk menganalisis opini publik di media sosial seperti Instagram dan X, sehingga dapat membantu memahami persepsi masyarakat terhadap isu pembatalan pembelian BBM oleh SPBU swasta secara lebih sistematis dan akurat.

2.2.1.3 Temuan Utama

Penelitian yang dilakukan oleh Ryo Kusnadi dkk. (2021) menghasilkan temuan bahwa metode BERT (Bidirectional Encoder Representations from Transformers) mampu memberikan hasil analisis sentimen yang sangat akurat terhadap ulasan pengguna game *Genshin Impact* di Google Play Store. Model yang dikembangkan berhasil mengklasifikasikan sentimen pengguna ke dalam tiga kategori, yaitu positif, netral, dan negatif, dengan tingkat *precision* sebesar 0.79 untuk sentimen negatif, 0.77 untuk positif, dan 0.70 untuk netral, serta nilai ROC AUC mencapai 0.88, yang menunjukkan kinerja model yang baik dalam membedakan tiap jenis sentimen. Hasil tersebut juga memperlihatkan bahwa

sentimen netral merupakan kategori yang paling sulit diprediksi karena berada di antara dua kutub emosi yang berlawanan. Selain itu, penelitian ini menunjukkan bahwa penggunaan BERT lebih unggul dibandingkan metode tradisional seperti Naïve Bayes karena kemampuannya memahami konteks kalimat dari dua arah sekaligus. Dalam konteks skripsi ini, temuan tersebut menjadi dasar penting bahwa BERT efektif digunakan untuk menganalisis opini publik pada platform media sosial seperti Instagram dan X, terutama dalam memahami persepsi masyarakat terhadap isu pembatalan pembelian BBM oleh SPBU swasta secara lebih mendalam dan akurat.

2.2.1.4 Kesimpulan Penelitian

Penelitian yang dilakukan oleh Ryo Kusnadi dkk. (2021) menyimpulkan bahwa metode BERT (Bidirectional Encoder Representations from Transformers) sangat efektif dalam melakukan analisis sentimen berbasis teks karena mampu memahami konteks bahasa secara dua arah sehingga menghasilkan klasifikasi sentimen yang akurat. Melalui pengujian terhadap ribuan ulasan pengguna game *Genshin Impact* di Google Play Store, model BERT berhasil membedakan sentimen positif, negatif, dan netral dengan tingkat akurasi yang tinggi. Penelitian ini juga membuktikan bahwa model BERT lebih unggul dibandingkan algoritma pembelajaran mesin tradisional dalam mengolah data teks yang kompleks dan tidak terstruktur. Selain itu, penelitian ini menunjukkan pentingnya proses *data preprocessing* dan penanganan *class imbalance* agar model dapat memberikan hasil yang lebih optimal. Dalam konteks skripsi ini, kesimpulan tersebut menjadi landasan bahwa BERT dapat diterapkan secara efektif untuk menganalisis sentimen

publik di media sosial seperti Instagram dan X, sehingga mampu memberikan gambaran yang lebih objektif dan mendalam mengenai persepsi masyarakat terhadap kebijakan pembatalan pembelian BBM oleh SPBU swasta.

2.2.2 Paper 2 : Analisis Sentimen Hate Speech Mengenai Calon Wakil

Presiden Indonesia Menggunakan Algoritma BERT

Penelitian berjudul “*Analisis Sentimen Hate Speech Mengenai Calon Wakil Presiden Indonesia Menggunakan Algoritma BERT*” membahas penerapan algoritma Bidirectional Encoder Representations from Transformers (BERT) dalam menganalisis dan mengklasifikasikan sentimen yang mengandung ujaran kebencian (*hate speech*) terhadap calon wakil presiden Indonesia di media sosial. Latar belakang penelitian ini didasarkan pada fenomena meningkatnya aktivitas pengguna media sosial dalam menanggapi isu politik menjelang pemilihan umum, di mana opini publik sering kali disertai ujaran kebencian yang dapat memicu polarisasi sosial. Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasikan opini masyarakat yang bersifat positif, negatif, maupun netral, serta mendeteksi ujaran kebencian yang beredar di dunia maya dengan menggunakan pendekatan berbasis pemrosesan bahasa alami (*Natural Language Processing*).

Metode penelitian dilakukan dengan mengumpulkan data komentar dari media sosial seperti Twitter dan Instagram, yang kemudian melalui proses pra-pemrosesan teks seperti *case folding*, *cleansing*, *tokenizing*, *stopword removal*, dan *stemming* untuk menyiapkan data agar dapat dianalisis oleh model BERT. Setelah tahap pra-pemrosesan, dilakukan pelatihan model menggunakan arsitektur BERT-

Base Multilingual Cased, karena model ini mampu memahami konteks bahasa Indonesia dengan baik. Hasil penelitian menunjukkan bahwa BERT mampu mendeteksi sentimen dan ujaran kebencian dengan tingkat akurasi yang tinggi, di mana model menunjukkan kinerja yang lebih baik dibandingkan metode konvensional seperti Naïve Bayes dan SVM. Model BERT dapat menangkap makna kontekstual kata dalam kalimat, sehingga mampu membedakan opini yang bersifat kritik dengan ujaran kebencian secara lebih akurat.

Temuan utama penelitian ini menunjukkan bahwa BERT efektif dalam melakukan klasifikasi sentimen dan deteksi hate speech pada teks berbahasa Indonesia, terutama di media sosial yang memiliki bahasa tidak baku dan penuh variasi. Dalam konteks skripsi ini, penelitian tersebut menjadi rujukan penting karena memiliki kesamaan dalam penggunaan metode BERT serta sumber data yang berasal dari media sosial seperti Instagram dan X (Twitter). Perbedaannya terletak pada objek kajian, di mana penelitian sebelumnya berfokus pada ujaran kebencian dalam konteks politik, sedangkan skripsi ini menitikberatkan pada analisis sentimen publik terhadap isu ekonomi dan energi, yakni pembatalan pembelian BBM oleh SPBU swasta dari Pertamina. Dengan demikian, penelitian ini menjadi dasar yang kuat dalam menunjukkan kemampuan BERT sebagai model deep learning yang efektif untuk memahami konteks bahasa alami dan menganalisis opini publik secara lebih mendalam dan akurat.

2.2.2.1 Tujuan Penelitian

Tujuan utama dari penelitian *“Analisis Sentimen Hate Speech Mengenai Calon Wakil Presiden Indonesia Menggunakan Algoritma BERT”* adalah untuk

menganalisis dan mengklasifikasikan sentimen masyarakat di media sosial yang berkaitan dengan isu politik, khususnya terhadap calon wakil presiden Indonesia, sekaligus mendeteksi ujaran kebencian (*hate speech*) yang muncul dalam percakapan daring. Penelitian ini bertujuan untuk menunjukkan efektivitas algoritma Bidirectional Encoder Representations from Transformers (BERT) dalam memahami konteks bahasa alami serta membedakan antara opini, kritik, dan ujaran kebencian dengan tingkat akurasi yang tinggi. Melalui proses pengumpulan data, pra-pemrosesan teks, pelatihan model, dan evaluasi performa, penelitian ini menghasilkan model yang mampu mengenali pola ujaran kebencian dalam bahasa Indonesia secara akurat. Dalam konteks skripsi ini yang berjudul “*Analisis Sentimen Publik di Instagram dan X terhadap Pembatalan Pembelian BBM dari Pertamina oleh SPBU Swasta Menggunakan Metode BERT*,” tujuan penelitian tersebut menjadi acuan penting dalam pemanfaatan BERT untuk menganalisis opini publik secara otomatis dan mendalam, di mana fokus penelitian diarahkan pada persepsi masyarakat terhadap isu ekonomi dan energi, khususnya kebijakan pembatalan pembelian BBM oleh SPBU swasta.

2.2.2.2 Metodologi yang Digunakan

Penelitian “*Analisis Sentimen Hate Speech Mengenai Calon Wakil Presiden Indonesia Menggunakan Algoritma BERT*” menggunakan pendekatan berbasis Bidirectional Encoder Representations from Transformers (BERT) untuk melakukan analisis sentimen dan deteksi ujaran kebencian pada data media sosial. Metodologi yang diterapkan dimulai dari pengumpulan data komentar dan unggahan dari platform media sosial seperti Twitter dan Instagram, yang kemudian

diproses melalui tahapan pra-pemrosesan teks meliputi *case folding*, *cleansing*, *tokenizing*, *stopword removal*, dan *stemming* agar data siap digunakan untuk pelatihan model. Setelah data bersih diperoleh, dilakukan proses pelabelan sentimen dan hate speech menjadi tiga kategori utama, yaitu positif, negatif, dan netral. Model BERT yang digunakan dalam penelitian ini adalah yang telah dilatih pada berbagai bahasa termasuk bahasa Indonesia, sehingga mampu memahami konteks kalimat secara dua arah. Proses pelatihan model dilakukan dengan parameter tertentu seperti *batch size*, *learning rate*, dan *epoch* yang diatur untuk mendapatkan performa terbaik. Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk menilai tingkat keakuratan klasifikasi. Hasil evaluasi menunjukkan bahwa model BERT mampu mengidentifikasi ujaran kebencian dan sentimen publik dengan akurasi tinggi. Dalam konteks skripsi ini, metodologi penelitian tersebut menjadi acuan dalam penerapan tahapan serupa untuk menganalisis sentimen publik di Instagram dan X, dengan menyesuaikan konteks data dan isu penelitian mengenai pembatalan pembelian BBM oleh SPBU swasta dari Pertamina agar model dapat memberikan hasil analisis yang relevan dan akurat.

2.2.2.3 Temuan Utama

Penelitian “*Analisis Sentimen Hate Speech Mengenai Calon Wakil Presiden Indonesia Menggunakan Algoritma BERT*” menunjukkan bahwa algoritma BERT mampu mengklasifikasikan sentimen publik dan mendeteksi ujaran kebencian dengan tingkat akurasi yang tinggi pada data media sosial berbahasa Indonesia. Model BERT Base Multilingual Cased yang digunakan terbukti efektif dalam

memahami konteks bahasa secara dua arah, sehingga dapat membedakan opini, kritik, dan ujaran kebencian dengan lebih akurat dibandingkan metode konvensional seperti Naïve Bayes atau SVM. Hasil penelitian memperlihatkan bahwa BERT dapat menangkap makna tersembunyi di balik gaya bahasa informal yang umum digunakan di media sosial, serta mampu mengenali ekspresi negatif atau provokatif yang mengandung indikasi kebencian. Selain itu, ditemukan bahwa sebagian besar komentar publik di media sosial menunjukkan kecenderungan sentimen negatif terhadap tokoh politik tertentu menjelang masa pemilihan umum. Temuan ini membuktikan bahwa BERT sangat efektif digunakan dalam analisis sentimen dan deteksi hate speech pada teks berbahasa Indonesia. Dalam konteks skripsi ini, hasil penelitian tersebut menjadi landasan penting bahwa BERT juga dapat diterapkan secara optimal untuk menganalisis opini publik di Instagram dan X, khususnya dalam memahami persepsi masyarakat terhadap kebijakan ekonomi dan energi seperti pembatalan pembelian BBM oleh SPBU swasta, karena kedua penelitian sama-sama memanfaatkan data teks tidak terstruktur dari media sosial dengan gaya bahasa yang beragam.

2.2.2.4 Kesimpulan Penelitian

Penelitian “*Analisis Sentimen Hate Speech Mengenai Calon Wakil Presiden Indonesia Menggunakan Algoritma BERT*” menyimpulkan bahwa penggunaan algoritma BERT (Bidirectional Encoder Representations from Transformers) sangat efektif dalam melakukan analisis sentimen dan deteksi ujaran kebencian pada teks berbahasa Indonesia, terutama yang berasal dari media sosial. BERT mampu memahami konteks bahasa secara dua arah sehingga dapat membedakan

antara opini netral, kritik, dan ujaran kebencian dengan tingkat akurasi yang tinggi. Hasil penelitian juga menunjukkan bahwa model BERT memiliki performa yang lebih unggul dibandingkan metode tradisional seperti Naïve Bayes dan SVM dalam menangani data teks tidak terstruktur yang bersifat kompleks dan penuh variasi. Selain itu, penelitian ini menegaskan pentingnya tahap pra-pemrosesan data dalam meningkatkan kinerja model, karena data media sosial sering kali mengandung banyak noise dan bahasa informal. Dalam konteks skripsi ini yang berjudul *“Analisis Sentimen Publik di Instagram dan X terhadap Pembatalan Pembelian BBM dari Pertamina oleh SPBU Swasta Menggunakan Metode BERT,”* kesimpulan tersebut menjadi dasar kuat bahwa BERT dapat digunakan secara efektif untuk memahami opini publik secara mendalam, sekaligus memberikan analisis sentimen yang akurat terhadap isu-isu sosial dan ekonomi berdasarkan data nyata dari media sosial.

2.2.3 Paper 3 : Analisis Sentimen terhadap Pelayanan Kesehatan Berdasarkan Ulasan Google Maps Menggunakan BERT

Penelitian yang dilakukan oleh Ardiansyah, Adika Sri Widagdo, Krisna Nuresa Qodri, Fachruddin Edi Nugroho Saputro, dan Nisrina Akbar Rizky P. (2023) berjudul *“Analisis Sentimen terhadap Pelayanan Kesehatan Berdasarkan Ulasan Google Maps Menggunakan BERT”* membahas penerapan algoritma Bidirectional Encoder Representations from Transformers (BERT) dalam mengklasifikasikan opini masyarakat terhadap pelayanan kesehatan melalui ulasan yang diambil dari platform Google Maps. Fokus penelitian ini adalah untuk mengetahui kecenderungan sentimen publik—positif, negatif, dan netral—terhadap

layanan rumah sakit, puskesmas, dan klinik kesehatan, serta mengidentifikasi topik yang paling sering dibahas oleh masyarakat terkait kualitas layanan tersebut. Latar belakang penelitian ini berangkat dari pentingnya pengukuran kepuasan pengguna layanan kesehatan, karena ulasan masyarakat di platform digital sering kali tidak sejalan antara teks ulasan dan rating bintang yang diberikan. Oleh karena itu, analisis sentimen berbasis NLP (Natural Language Processing) digunakan sebagai pendekatan ilmiah untuk menilai opini publik secara lebih objektif dan terukur.

Penelitian ini menggunakan metode BERT IndoBenchmark/IndoBERT-base-p1 dengan dataset berjumlah 4.748 ulasan yang setelah proses pra-pemrosesan menjadi 4.228 data. Tahapan penelitian meliputi pengumpulan data dengan Selenium dan BeautifulSoup, pembersihan teks (*data cleaning*), *case folding*, *tokenizing*, dan penghapusan *stopwords*, kemudian dilakukan labeling otomatis menggunakan model RoBERTa bahasa Indonesia. Setelah itu, data dibagi menjadi tiga bagian dengan rasio 70% data latih, 30% validasi, dan 30% data uji. Model BERT dilatih menggunakan GPU Tesla T4 dengan parameter *learning rate* $3e-6$, *batch size* 32, dan 5 *epochs*. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa model mencapai akurasi 85% pada data validasi dan 86% pada data uji, dengan rata-rata *macro average* 74% dan *weighted average* 85%. Selain itu, analisis menunjukkan bahwa kata “pelayanan” menjadi topik paling dominan, diikuti oleh kata “ramah”, “cepat”, dan “dokter,” menandakan bahwa aspek pelayanan menjadi perhatian utama masyarakat terhadap fasilitas kesehatan.

Dalam konteks skripsi ini, penelitian tersebut sangat relevan karena sama-sama menerapkan metode BERT untuk menganalisis opini publik berbasis teks dalam bahasa Indonesia. Meskipun objek kajian berbeda—penelitian Ardiansyah dkk. berfokus pada layanan kesehatan, sedangkan skripsi ini pada kebijakan ekonomi dan energi—keduanya memiliki kesamaan dalam pendekatan metodologis dan penggunaan sumber data publik berbasis platform digital. Penelitian ini menjadi referensi penting dalam menunjukkan efektivitas BERT dalam memahami konteks bahasa alami dan mengklasifikasikan sentimen publik secara akurat, sehingga dapat diterapkan pula untuk menganalisis persepsi masyarakat terhadap pembatalan pembelian BBM oleh SPBU swasta dari Pertamina di media sosial seperti Instagram dan X.

2.2.3.1 Tujuan Penelitian

Tujuan utama dari penelitian *“Analisis Sentimen terhadap Pelayanan Kesehatan Berdasarkan Ulasan Google Maps Menggunakan BERT”* adalah untuk mengidentifikasi dan menganalisis sentimen masyarakat terhadap layanan kesehatan di Indonesia berdasarkan ulasan pengguna di platform Google Maps. Penelitian ini berupaya mengklasifikasikan opini publik ke dalam tiga kategori utama, yaitu positif, negatif, dan netral, guna memperoleh gambaran objektif mengenai persepsi masyarakat terhadap kualitas pelayanan rumah sakit, puskesmas, dan klinik. Dengan menggunakan algoritma Bidirectional Encoder Representations from Transformers (BERT), penelitian ini bertujuan meningkatkan akurasi analisis sentimen berbahasa Indonesia, terutama dalam konteks teks yang bersifat tidak terstruktur dan informal. Hasil analisis diharapkan dapat menjadi

dasar evaluasi bagi instansi kesehatan untuk memahami kepuasan masyarakat secara lebih mendalam. Dalam konteks skripsi ini yang berjudul *“Analisis Sentimen Publik di Instagram dan X terhadap Pembatalan Pembelian BBM dari Pertamina oleh SPBU Swasta Menggunakan Metode BERT,”* tujuan penelitian tersebut menjadi acuan penting dalam menunjukkan kemampuan BERT dalam mengolah data teks publik secara kontekstual dan akurat, sehingga dapat diterapkan pula untuk memahami persepsi masyarakat terhadap kebijakan publik di sektor ekonomi dan energi.

2.2.3.2 Metodologi yang Digunakan

Penelitian *“Analisis Sentimen terhadap Pelayanan Kesehatan Berdasarkan Ulasan Google Maps Menggunakan BERT”* menggunakan metodologi berbasis Bidirectional Encoder Representations from Transformers (BERT) untuk menganalisis opini masyarakat terhadap layanan kesehatan melalui ulasan pengguna di platform Google Maps. Tahapan penelitian dimulai dengan pengumpulan data menggunakan Selenium dan BeautifulSoup, di mana peneliti berhasil memperoleh sebanyak 4.748 ulasan dari berbagai fasilitas kesehatan seperti rumah sakit, puskesmas, dan klinik. Data tersebut kemudian melalui tahap pra-pemrosesan teks yang meliputi *data cleaning*, *case folding*, *tokenizing*, dan *stopword removal* untuk memastikan data siap diolah oleh model. Proses pelabelan data dilakukan secara otomatis menggunakan model *RoBERTa bahasa Indonesia*, dan data dibagi menjadi tiga bagian: 70% untuk data latih, 30% untuk validasi, dan 30% untuk pengujian. Model yang digunakan adalah IndoBERT-base-p1 dari IndoBenchmark, yang dilatih menggunakan GPU Tesla T4 dengan parameter

learning rate 3e-6, *batch size* 32, dan 5 *epochs*. Evaluasi performa dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk menilai hasil klasifikasi. Hasil penelitian menunjukkan bahwa model BERT mampu memberikan kinerja yang baik dalam mengidentifikasi sentimen masyarakat terhadap layanan kesehatan. Dalam konteks skripsi ini, metodologi tersebut menjadi acuan penting dalam penerapan BERT untuk analisis sentimen publik di Instagram dan X, dengan menyesuaikan konteks data dan topik penelitian terkait pembatalan pembelian BBM oleh SPBU swasta dari Pertamina agar hasil analisis lebih relevan dan akurat.

2.2.3.3 Temuan Utama

Penelitian “*Analisis Sentimen terhadap Pelayanan Kesehatan Berdasarkan Ulasan Google Maps Menggunakan BERT*” menghasilkan temuan bahwa model BERT (Bidirectional Encoder Representations from Transformers) sangat efektif dalam mengklasifikasikan sentimen masyarakat terhadap layanan kesehatan di Indonesia berdasarkan ulasan di Google Maps. Hasil pengujian menunjukkan bahwa model IndoBERT-base-p1 mencapai tingkat akurasi sebesar 85% pada data validasi dan 86% pada data uji, dengan rata-rata *macro average* 74% dan *weighted average* 85%. Temuan ini membuktikan bahwa BERT mampu memahami konteks bahasa Indonesia secara mendalam, termasuk dalam teks yang bersifat tidak baku dan informal. Analisis juga menunjukkan bahwa sebagian besar sentimen pengguna bersifat positif, dengan topik utama yang paling sering muncul adalah “pelayanan”, “ramah”, “cepat”, dan “dokter”, yang menandakan bahwa kualitas pelayanan menjadi faktor paling berpengaruh dalam penilaian publik terhadap fasilitas kesehatan. Selain itu, model BERT terbukti lebih unggul dibandingkan metode

konvensional seperti Naïve Bayes dan LSTM dalam menangani data teks yang kompleks dan bervariasi. Dalam konteks skripsi ini, temuan tersebut memperkuat bukti bahwa BERT dapat digunakan secara efektif untuk menganalisis opini publik di media sosial seperti Instagram dan X, karena memiliki kemampuan memahami konteks bahasa alami dan mengenali sentimen masyarakat terhadap kebijakan atau isu sosial, seperti pembatalan pembelian BBM oleh SPBU swasta dari Pertamina, secara lebih akurat dan komprehensif.

2.2.3.4 Kesimpulan Penelitian

Penelitian “*Analisis Sentimen terhadap Pelayanan Kesehatan Berdasarkan Ulasan Google Maps Menggunakan BERT*” menyimpulkan bahwa penggunaan algoritma BERT (Bidirectional Encoder Representations from Transformers) sangat efektif untuk melakukan analisis sentimen berbasis teks berbahasa Indonesia, khususnya dalam konteks ulasan publik terhadap layanan kesehatan. Model IndoBERT-base-p1 terbukti mampu memahami konteks kalimat secara dua arah, sehingga dapat mengidentifikasi perbedaan antara sentimen positif, negatif, dan netral dengan tingkat akurasi tinggi. Hasil penelitian menunjukkan bahwa mayoritas ulasan masyarakat terhadap layanan kesehatan bersifat positif, terutama pada aspek keramahan dan kecepatan pelayanan, sementara sentimen negatif umumnya berkaitan dengan waktu tunggu dan prosedur administrasi. Penelitian ini juga menegaskan bahwa BERT memiliki kemampuan lebih unggul dibandingkan metode tradisional seperti Naïve Bayes dan LSTM dalam menangani teks tidak terstruktur dan bahasa informal. Dalam konteks skripsi ini yang berjudul “*Analisis Sentimen Publik di Instagram dan X terhadap Pembatalan Pembelian BBM dari*

Pertamina oleh SPBU Swasta Menggunakan Metode BERT,” kesimpulan tersebut menjadi landasan kuat bahwa BERT dapat diterapkan secara efektif untuk memahami opini publik di media sosial, serta menghasilkan analisis sentimen yang akurat terhadap isu-isu kebijakan publik di bidang ekonomi dan energi.

