

BAB II

LANDASAN TEORI

2.1 Sentimen Masyarakat di Media Sosial

Sentimen masyarakat di media sosial merupakan ekspresi opini, perasaan, atau persepsi individu terhadap suatu topik, layanan, atau institusi yang disampaikan melalui platform digital seperti Twitter, Instagram, dan Facebook. Dalam konteks akademik, sentimen terhadap institusi pendidikan tinggi dapat mencerminkan tingkat kepuasan, kritik, maupun rekomendasi yang berasal dari mahasiswa aktif, alumni, maupun masyarakat umum. Media sosial telah menjadi platform penting untuk berbagi konten dan interaksi, dengan jutaan gambar diposting setiap hari yang mencerminkan berbagai pengalaman dan emosi pengguna (Darwis *et al.*, 2024).

Menurut (Keahlian *et al.*, 2021) dijelaskan bahwa analisis sentimen merupakan proses untuk mengenali, mengekstrak, dan mengklasifikasikan opini masyarakat dalam bentuk teks menjadi kategori sentimen positif, negatif, dan netral. Sentimen positif biasanya menunjukkan dukungan atau apresiasi terhadap suatu objek, sedangkan sentimen negatif memuat opini berupa keluhan, kritik, atau ketidakpuasan.

Sementara itu, sentimen netral bersifat informatif tanpa memuat ekspresi emosi tertentu. Dengan mengetahui proporsi dari ketiga jenis sentimen ini, suatu institusi dapat memanfaatkan informasi tersebut untuk merancang kebijakan atau strategi pelayanan yang lebih responsif dan sesuai dengan kebutuhan masyarakat. dijelaskan bahwa analisis sentimen merupakan proses untuk

mengenali, mengekstrak, dan mengklasifikasikan opini masyarakat dalam bentuk teks menjadi kategori sentimen positif, negatif, dan netral. Sentimen positif biasanya menunjukkan dukungan atau apresiasi terhadap suatu objek, sedangkan sentimen negatif memuat opini berupa keluhan, kritik, atau ketidakpuasan.

2.2 Analisis Sentimen

Analisis sentimen merupakan salah satu teknik dalam text mining yang digunakan untuk mengenali, mengekstrak, dan mengelompokkan opini masyarakat yang terkandung dalam sebuah teks ke dalam kategori positif, negatif, atau netral. Proses ini bertujuan untuk mengetahui kecenderungan sikap, perasaan, maupun pandangan publik terhadap suatu isu atau layanan tertentu. Dalam ranah pendidikan, penerapan metode ini sangat relevan, mengingat banyaknya opini mahasiswa dan masyarakat yang disampaikan secara terbuka melalui media sosial. Dengan analisis sentimen, institusi pendidikan dapat memetakan persepsi publik terhadap pelayanan kampus secara sistematis dan akurat, sehingga hasilnya dapat dimanfaatkan untuk evaluasi layanan maupun perbaikan kebijakan di lingkungan universitas (Fridom Mailo *et al.*, 2021).

Seiring dengan pesatnya perkembangan teknologi kecerdasan buatan, metode analisis sentimen turut mengalami kemajuan signifikan, khususnya dengan hadirnya algoritma berbasis deep learning. Salah satu metode paling efektif saat ini adalah model berbasis transformer seperti BERT (Bidirectional Encoder Representations from Transformers), yang terbukti mampu memahami konteks kalimat secara lebih akurat dibandingkan dengan metode konvensional berbasis kamus maupun machine learning klasik. penggunaan BERT dalam analisis sentimen media sosial memungkinkan ekstraksi informasi sentimental dan

kontekstual dari teks secara lebih optimal, sehingga sangat potensial diterapkan dalam menganalisis opini masyarakat terhadap layanan institusi pendidikan seperti Universitas Dharma Persada (Tabinda Kokab *et al.*, 2022).

2.2.1 Transformer dalam Analisis Sentimen

Seiring berkembangnya teknologi kecerdasan buatan, khususnya dalam bidang pemrosesan bahasa alami (Natural Language Processing/NLP), penerapan algoritma berbasis deep learning telah membawa kemajuan signifikan dalam analisis sentimen. Salah satu model berbasis transformer yang paling banyak digunakan saat ini adalah BERT (Bidirectional Encoder Representations from Transformers), yang terbukti lebih efektif dalam memahami konteks teks dibandingkan metode tradisional seperti word embedding konvensional. Menurut Kokab ST *et al.*, (2022) penggunaan model transformer mampu menangkap informasi sintaksis, semantik, dan konteks secara menyeluruh, sehingga hasil klasifikasi sentimen lebih akurat pada data media sosial, termasuk data berbahasa Indonesia seperti opini terhadap layanan di institusi Pendidikan

Dalam penelitian analisis sentimen, model berbasis Transformer seperti BERT telah terbukti unggul dibandingkan pendekatan konvensional. Hal ini disebabkan oleh kemampuannya dalam menangkap informasi semantik dan kontekstual secara mendalam, serta mengatasi kendala seperti kehilangan makna akibat kata di luar kosakata (OOV) yang umum terjadi di media sosial. Selain itu, BERT memungkinkan pemrosesan data dengan konteks yang lebih akurat berkat mekanisme perhatian (attention mechanism) yang kompleks dan representasi dinamis antar kata. Oleh karena itu, penggunaan pendekatan Transformer dalam studi ini diharapkan dapat memberikan analisis sentimen yang lebih akurat dan

representatif terhadap opini masyarakat mengenai Universitas Dharma Persada (Tabinda Kokab *et al.*, 2022)

2.2.2 Penerapan Analisis Sentimen pada Pendidikan Tinggi

Analisis sentimen memainkan peran penting dalam menilai pandangan alumni terhadap institusi pendidikan tinggi. Dengan mengelompokkan opini menjadi kategori positif, netral, atau negatif, institusi dapat memperoleh gambaran yang lebih jelas mengenai pengalaman belajar mahasiswa selama masa studi. Penelitian oleh Komang Dharmendra *et al.*, (2022) menunjukkan bahwa pemanfaatan analisis sentimen terhadap opini alumni mampu memberikan informasi yang relevan untuk mendukung pengambilan keputusan dan peningkatan kualitas layanan akademik .

2.3 Text mining

Pengertian Text Mining Text mining merupakan teknik pengolahan data teks secara otomatis untuk memahami, mengekstrak, dan menganalisis informasi opini yang terkandung dalam kalimat atau dokumen. Tujuan dari proses ini adalah untuk memperoleh pola sentimen—baik positif maupun negatif—dari data teks yang tersedia, seperti media sosial. Dalam penelitian Lestari (2022), text mining diterapkan untuk mengelompokkan opini masyarakat di instgram terhadap suatu topik, dengan metode klasifikasi berbasis SVM

Dalam konteks penelitian ini, text mining digunakan sebagai fondasi awal untuk mengevaluasi opini masyarakat terhadap Universitas Dharma Persada, khususnya persepsi yang muncul di media sosial terkait layanan akademik, reputasi fakultas, serta pengalaman mahasiswa dan alumni. Proses ini diawali dengan tahap

pra-pemrosesan teks (*text preprocessing*), yang mencakup pembersihan data dari elemen tidak relevan (seperti simbol dan tanda baca), normalisasi huruf (case folding), pemisahan kata (tokenisasi), penghapusan kata-kata umum yang tidak bermakna (stopwords), serta pengembalian kata ke bentuk dasar (stemming). Tahap ini sangat krusial untuk memastikan data yang dianalisis memiliki kualitas tinggi dan representatif.

Setelah data melalui tahap pra-pemrosesan, langkah selanjutnya adalah penerapan model klasifikasi berbasis Transformer, yang dalam hal ini menggantikan metode klasik seperti SVM. Model Transformer, seperti IndoBERT, memiliki kemampuan memahami struktur kalimat dan konteks bahasa Indonesia dengan lebih mendalam. Penerapannya dalam penelitian ini ditujukan untuk mengatasi keterbatasan metode konvensional yang kurang mampu menangkap makna kontekstual, terutama dalam kalimat yang sarat dengan makna implisit, bahasa gaul, atau ironi yang sering dijumpai dalam komentar media sosial.

Dengan memanfaatkan teknik text mining dan mengintegrasikannya dengan model Transformer, penelitian ini bertujuan untuk membangun sistem analisis sentimen yang tidak hanya akurat secara teknis, tetapi juga mampu memberikan gambaran menyeluruh tentang persepsi publik terhadap Universitas Dharma Persada. Hasil dari analisis ini dapat dijadikan sebagai bahan pertimbangan dalam pengambilan keputusan strategis oleh pihak universitas, baik dalam hal peningkatan mutu layanan, pengembangan kebijakan akademik, maupun dalam membangun citra institusi yang lebih positif di mata masyarakat.

2.4 Natural Language Processing (NLP)

Natural Language Processing (NLP) merupakan cabang multidisipliner dari kecerdasan buatan (AI) yang berfokus pada kemampuan mesin untuk memahami, menafsirkan, dan menghasilkan bahasa manusia secara kontekstual. NLP mengintegrasikan prinsip-prinsip linguistik, ilmu komputer, dan psikologi kognitif untuk menjembatani komunikasi antara manusia dan mesin. Seperti yang dijelaskan oleh Rahyan Abu *et al.*, (2023) NLP memungkinkan komputer mengolah data teks yang tidak terstruktur dan mengubahnya menjadi informasi yang bermakna untuk berbagai aplikasi, seperti analisis sentimen, penerjemahan bahasa, hingga sistem asisten virtual.

Dalam penelitian ini, NLP berperan penting dalam mengolah komentar dan opini publik mengenai Universitas Dharma Persada yang tersebar di media sosial. Komponen NLP seperti analisis sintaksis, pengenalan entitas, dan klasifikasi sentimen digunakan untuk menilai persepsi masyarakat. Salah satu pendekatan utama dalam NLP adalah tokenisasi, yaitu proses memecah teks menjadi unit-unit kecil (token) seperti kata atau frasa, yang menjadi dasar bagi analisis lanjutan seperti pengenalan jenis kata dan parsing struktur kalimat.

Proses NLP ini juga mencakup penandaan bagian kalimat (part-of-speech tagging), yaitu pelabelan setiap token berdasarkan fungsi katanya, serta parsing sintaksis, yang menganalisis hubungan antar kata dalam sebuah kalimat untuk membentuk struktur kalimat yang logis. Dengan teknik ini, sistem dapat mengenali konteks kalimat seperti “Dharma Persada sangat maju dalam bidang teknik” sebagai opini positif.

Lebih lanjut, NLP digunakan untuk analisis sentimen, yaitu teknik untuk mengidentifikasi emosi atau sikap dalam sebuah teks—apakah bersifat positif, negatif, atau netral. Hal ini sangat relevan dalam penelitian ini, di mana persepsi publik terhadap universitas perlu dianalisis untuk memahami bagaimana institusi dipandang oleh masyarakat. Dengan mengintegrasikan NLP dan model Transformer yang telah dilatih dalam konteks Bahasa Indonesia, seperti IndoBERT, proses klasifikasi sentimen menjadi lebih kontekstual dan akurat.

2.5 Peran NLP di dalam Klasifikasi Sentimen

Natural Language Processing (NLP) memiliki peran yang sangat penting dalam analisis dan klasifikasi sentimen, khususnya dalam mengolah ulasan atau komentar pengguna yang berbentuk teks tidak terstruktur. NLP digunakan untuk mengelompokkan sentimen pengguna terhadap aplikasi Spotify ke dalam tiga kategori utama: positif, netral, dan negatif. Proses ini melibatkan serangkaian tahapan pra-pemrosesan teks seperti expanding contractions, case folding, stopword removal, dan lemmatisasi, yang bertujuan meningkatkan kualitas data sebelum dilakukan klasifikasi sentimen menggunakan algoritma pembelajaran mesin (Prasetya Aureli *et al.*, 2024).

Penerapan NLP memungkinkan sistem untuk memahami makna dan struktur bahasa yang digunakan oleh pengguna dalam ulasan mereka, bahkan dalam bentuk bahasa informal yang umum ditemui di media sosial. Setelah tahap pra-pemrosesan, teks kemudian diubah menjadi representasi numerik melalui vektorisasi seperti metode TF-IDF, yang selanjutnya digunakan untuk melatih model klasifikasi seperti Naïve Bayes atau Logistic Regression. Studi ini menunjukkan bahwa pendekatan NLP yang tepat dapat meningkatkan akurasi

model dalam mengklasifikasikan sentimen, membuktikan relevansi NLP dalam mendukung pengambilan keputusan berbasis opini publik.

Dalam konteks penelitian ini, yang berfokus pada analisis sentimen terhadap Universitas Dharma Persada menggunakan metode Transformer, NLP digunakan sebagai fondasi penting dalam memahami opini masyarakat dari media sosial atau platform ulasan. Dengan dukungan model bahasa modern berbasis Transformer seperti IndoBERT, klasifikasi sentimen menjadi lebih akurat karena mampu menangkap makna kontekstual dari bahasa Indonesia secara lebih mendalam.

2.6 Transformer & IndoBERT

2.6.1 Transformer

Model Transformer merupakan arsitektur jaringan saraf dalam bidang kecerdasan buatan yang dirancang khusus untuk menangani data berurutan seperti teks. Tidak seperti pendekatan tradisional seperti RNN atau CNN yang memproses data secara bertahap, Transformer memanfaatkan mekanisme *self-attention* yang memungkinkan pemrosesan paralel terhadap seluruh bagian input secara simultan. Mekanisme ini memungkinkan model untuk mempertimbangkan hubungan antar kata di seluruh bagian teks, tanpa harus mengikuti urutan secara berurutan. Hal ini menjadikan Transformer unggul dalam memahami konteks panjang dan kompleks, serta meningkatkan efisiensi dan akurasi dalam berbagai tugas pemrosesan bahasa alami, termasuk analisis sentimen (Dwi Purnomo & Sutopo, 2024).

Arsitektur Transformer secara umum terdiri dari dua bagian utama, yaitu encoder dan decoder. Namun, dalam tugas seperti analisis sentimen, hanya bagian

encoder yang digunakan, seperti pada model BERT. Encoder berfungsi untuk mengubah teks yang dimasukkan menjadi bentuk angka atau representasi digital yang bisa dimengerti oleh komputer. Representasi ini tidak hanya memperhatikan urutan kata dari kiri ke kanan, tetapi juga dari kanan ke kiri secara bersamaan, sehingga konteks kalimat bisa dipahami secara lebih menyeluruh (*bidirectional*).

Dalam konteks penelitian yang berfokus pada analisis sentimen terhadap Universitas Dharma Persada, model Transformer seperti IndoBERT digunakan untuk memahami opini masyarakat yang terekam melalui komentar-komentar di media sosial atau platform daring. Penelitian oleh Purnomo & Sutopo (2024) menunjukkan bahwa model berbasis Transformer yang telah dilatih dengan data bahasa Indonesia, seperti IndoBERT mampu memberikan performa sangat tinggi dalam mengklasifikasikan sentimen teks. Bahkan, IndoBERT varian besar (IndoNLU) mencatat skor F1 tertinggi sebesar 0.9353, yang menunjukkan efektivitas tinggi dalam memahami konteks lokal dan nuansa bahasa pengguna.

2.6.2 BERT dan IndoBERT (Pre-trained Model Bahasa Indonesia)

BERT (Bidirectional Encoder Representations from Transformers) adalah sebuah model bahasa berbasis arsitektur Transformer yang dikembangkan oleh Google pada tahun 2018. Berbeda dengan model sebelumnya yang hanya membaca teks dari satu arah, BERT memproses kalimat dari dua arah sekaligus—dari kiri ke kanan dan dari kanan ke kiri—sehingga mampu memahami konteks kata secara lebih menyeluruh. Pendekatan ini menjadikan BERT sangat unggul dalam berbagai tugas NLP (Natural Language Processing), termasuk dalam mendeteksi makna atau emosi dalam suatu kalimat.

IndoBERT adalah versi BERT yang dilatih khusus menggunakan data teks dalam Bahasa Indonesia dalam jumlah besar. Model ini digunakan untuk tugas klasifikasi sentimen karena mampu menangkap makna dan konteks bahasa lokal dengan lebih baik. Dalam penelitian), IndoBERT digunakan untuk mengklasifikasikan opini masyarakat Indonesia di Twitter mengenai Pemilu 2024. Hasilnya menunjukkan bahwa IndoBERT, khususnya versi large-p1, mampu mengklasifikasikan sentimen secara jauh lebih akurat dibandingkan model klasik seperti Naïve Bayes, SVM, Transformer maupun model berbasis kamus seperti TextBlob (Geni *et al.*, 2023).

Model IndoBERT bekerja dengan cara mempelajari pola-pola kata dan struktur kalimat dari data pelatihan, lalu menyempurnakannya (*fine-tuning*) agar mampu mengenali apakah suatu tweet bernada positif, negatif, atau netral. Akurasi tertinggi yang dicapai oleh IndoBERT large-p1 mencapai 83,5%, dengan peningkatan signifikan dibandingkan metode tradisional—membuktikan bahwa model ini sangat efektif untuk memahami opini publik di media sosial.

2.7 Data Mining

Data mining adalah proses untuk menemukan informasi yang penting dan bermanfaat dari kumpulan data yang sangat besar. Teknik ini menggunakan ilmu statistik, matematika, dan kecerdasan buatan untuk menggali pola atau hubungan yang tersembunyi di dalam data. Dalam konteks pendidikan, proses ini sering disebut Educational Data Mining (EDM), yaitu penerapan data mining untuk memahami perilaku dan kinerja mahasiswa berdasarkan data akademik yang tersedia.

Misalnya, di perguruan tinggi, terdapat banyak data tentang mahasiswa seperti nilai, latar belakang sekolah, jenis kelamin, asal kota, pekerjaan orang tua, dan sebagainya. Semua data ini bisa diolah untuk menemukan pola yang menunjukkan siapa saja mahasiswa yang kemungkinan besar akan lulus tepat waktu atau justru terlambat. Dengan cara ini, kampus dapat mengambil tindakan lebih awal untuk membantu mahasiswa agar tidak tertinggal dalam studinya (Galih, 2020).

Penerapan Data Mining dan NLP dalam Analisis Sentimen Media Sosial :

a. Pemodelan

Menggunakan teknik penambangan data seperti klasifikasi, pengelompokan, atau peramalan. Sebagai contoh, memakai model Transformer IndoBERT untuk klasifikasi sentimen.

b. Evaluasi

Menilai model menggunakan metrik tertentu seperti akurasi, precision, recall, dan F1-score untuk memastikan hasil yang diperoleh memenuhi tujuan bisnis.

c. Penerapan

Mengimplementasikan model dalam lingkungan nyata, misalnya membuat dashboard analisis sentimen Instagram secara langsung.

d. Penerapan Data Mining dalam Analisis Media Sosial

Platform media sosial menghasilkan data dengan volume yang besar dan beragam (teks, gambar, video). Penambangan data di sektor ini digunakan untuk memahami pandangan, perilaku, dan tren masyarakat.

Sebagai contoh, di Instagram, pengguna sering kali meninggalkan komentar yang bisa dikategorikan sebagai positif, netral, atau negatif terhadap produk atau layanan tertentu. Penambangan data memberikan kesempatan bagi perusahaan untuk:

- a. Mengotomatiskan klasifikasi komentar
- b. Mendeteksi isu yang sedang hangat
- c. Mengenal perubahan persepsi masyarakat terhadap merek
- d. Peran Data Mining dalam Klasifikasi Sentimen di Instagram

Klasifikasi sentimen merupakan salah satu fungsi penambangan data yang bertujuan untuk mengidentifikasi opini individu dalam teks sebagai positif, negatif, atau netral. Di Instagram, penambangan data memanfaatkan teknik Text Mining dan Natural Language Processing (NLP) untuk:

- a. Melakukan pembersihan teks komentar (praproses).
- b. Menerapkan model klasifikasi seperti Support Vector Machine (SVM), Random Forest, atau model berbasis Transformer (IndoBERT).
- c. Memprediksi polaritas sentimen dari komentar yang telah dianalisis.

Contoh Penerapan:

Memanfaatkan model IndoBERT untuk mengklasifikasi sentimen komentar di Instagram mengenai produk kosmetik lokal. Hasil penelitian menunjukkan bahwa akurasi model mencapai lebih dari 88%, melebihi metode tradisional seperti Tranformer

2.8 Instagram sebagai Sumber Data

Instagram adalah salah satu platform media sosial yang paling populer dan banyak digunakan, termasuk oleh para pendidik. Awalnya dibuat untuk berbagi foto dan video, Instagram kini berkembang menjadi tempat interaksi sosial yang sangat aktif, tidak hanya untuk keperluan pribadi, tetapi juga untuk kepentingan profesional.

Dalam konteks pendidikan dan penelitian, Instagram menjadi sumber data yang sangat kaya, karena pengguna secara rutin membagikan pengalaman, opini, dan refleksi mereka melalui unggahan, caption, dan komentar. Berdasarkan studi yang dilakukan oleh para Carpenter *et al.*, (2020) guru dan pendidik menggunakan Instagram bukan hanya untuk berbagi materi ajar atau dekorasi kelas, tetapi juga untuk membangun komunitas profesional, mendapatkan dukungan emosional, serta berbagi pengalaman dan ide pembelajaran.

Dengan fitur-fitur seperti hashtag, Instagram Stories, dan kolom komentar, platform ini memudahkan pengumpulan data berupa opini publik yang sangat kontekstual dan relevan. Banyak pengguna juga mencampurkan konten pribadi dan profesional dalam satu akun, yang memberikan pandangan lebih lengkap tentang karakter dan sikap mereka terhadap isu-isu tertentu—misalnya terhadap institusi pendidikan seperti Universitas Dharma Persada. Instagram sangat cocok digunakan sebagai sumber data dalam analisis sentimen, terutama karena:

- a) Komentar dan caption mencerminkan opini nyata pengguna terhadap suatu topik, layanan, atau institusi.
- b) Banyaknya data teks yang dapat diolah dengan metode text mining dan NLP (Natural Language Processing).

- c) Struktur visual dan emosional dari konten memungkinkan analisis konteks secara lebih mendalam.

Dalam studi ini, komentar-komentar pengguna Instagram terhadap Universitas Dharma Persada dapat dikumpulkan dan dianalisis untuk mengetahui apakah persepsi mereka bersifat positif, negatif, atau netral. Ini berguna untuk mengevaluasi citra universitas dan merancang strategi peningkatan kualitas layanan akademik.

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

CRISP-DM adalah singkatan dari *Cross-Industry Standard Process for Data Mining*, yaitu kerangka kerja atau panduan standar yang digunakan dalam proses data mining. Metode ini dikembangkan agar proses pengolahan data menjadi lebih terstruktur, terarah, dan bisa diterapkan di berbagai bidang industri.

CRISP-DM membagi proses data mining menjadi enam tahap utama, yang masing-masing memiliki tujuan spesifik dan saling berkaitan. Tahapan-tahapan ini membantu peneliti atau analis data untuk memahami masalah, menyiapkan data, membangun model, dan mengimplementasikan hasilnya secara nyata (Purnama *et al.*, 2020).

Tahapan sederhana CRISP-DM

1. Business Understanding (Pemahaman Bisnis). Tahap ini bertujuan memahami tujuan utama dari proyek data mining. Misalnya, ingin mengetahui pola konsumsi, memprediksi penjualan, atau dalam konteks skripsi Anda memahami opini publik

terhadap Universitas Dharma Persada.

Di tahap ini, kita tentukan dulu: “masalah apa yang ingin diselesaikan?”

2. Data Understanding (Pemahaman Data). Setelah tahu tujuannya, tahap berikutnya adalah mengumpulkan dan memahami data yang tersedia. Kita periksa dari mana data berasal, apakah kualitasnya bagus, dan apakah perlu dibersihkan atau diperbaiki.
3. Data Preparation (Persiapan Data). Ini adalah tahap "bersih-bersih" data. Kita memilih variabel yang dibutuhkan, membuang data tidak relevan, dan mengubah format data agar bisa digunakan oleh algoritma. Dalam skripsi Anda, ini bisa mencakup pembersihan komentar Instagram sebelum dianalisis.
4. Modeling (Pemodelan). Di sini, kita memilih dan menerapkan algoritma tertentu—misalnya *IndoBERT* untuk klasifikasi sentimen. Model dilatih agar bisa mengenali pola dalam data dan memprediksi sentimen (positif, negatif, netral).
5. Evaluation (Evaluasi). Setelah model selesai dibangun, kita uji akurasi: Apakah hasilnya sesuai harapan? Jika belum, mungkin perlu kembali ke tahap sebelumnya untuk diperbaiki.
6. Deployment (Penerapan). Tahap terakhir adalah mengimplementasikan model ke dalam sistem nyata. Misalnya, membuat dashboard analisis sentimen untuk melihat opini publik terhadap Universitas Dharma Persada secara otomatis.

2.8.1 Permodelan Uml

Unified Modeling Language (UML) merupakan bahasa visual standar yang digunakan untuk menggambarkan rancangan sistem perangkat lunak dalam bentuk diagram. UML membantu para pengembang perangkat lunak untuk menyusun dan

menjelaskan struktur sistem, alur kerja, serta hubungan antar bagian dalam sistem secara jelas dan terstruktur.

Menurut Sumiati *et al.*, (2021), UML memiliki kelebihan berupa fleksibilitas dan kemampuan untuk menjelaskan sistem secara lebih rinci dan terperinci. UML juga berperan sebagai cetak biru (blueprint) dalam pengembangan perangkat lunak, yang digunakan oleh semua pihak yang terlibat, seperti developer, user, hingga pemilik sistem (Sumiati *et al.*, 2021).



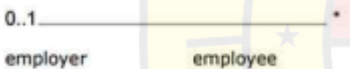

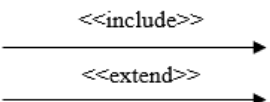
Pemodelan UML sangat penting dalam pengembangan sistem karena dapat:

- a) Menyampaikan ide atau desain sistem kepada stakeholder secara visual,
- b) Membantu merancang struktur sistem sebelum memasuki tahap implementasi,
- c) Menjadi dokumentasi yang berguna dalam proses evaluasi dan pengembangan lanjutan.

2.9 Usecase Diagram

Use Case Diagram adalah salah satu jenis diagram dalam pemodelan sistem yang digunakan untuk menjelaskan interaksi antara pengguna (aktor) dan sistem. Diagram ini menunjukkan fitur atau layanan apa saja yang tersedia dalam sistem, serta siapa saja yang bisa menggunakannya.

Use case diagram digunakan untuk menggambarkan kebutuhan sistem secara umum, dengan menunjukkan hubungan antara pengguna dan sistem, serta layanan yang disediakan oleh sistem tersebut. Dengan kata lain, diagram ini membantu memetakan siapa yang melakukan apa terhadap sistem, dan bagaimana sistem merespons setiap tindakan tersebut (Puturuhu Victor, 2022).

Simbol	Nama Komponen	Fungsi
	Aktor (<i>Actor</i>)	untuk memberikan input ke sistem, menerima informasi dari sistem, atau melakukan keduanya—mengirim dan menerima informasi.
	Dependency	Untuk menghubungkan antara dua elemen, di mana perubahan pada satu elemen memengaruhi elemen lainnya
	Association	hubungan antara elemen-elemen struktural yang saling terhubung dalam sebuah sistem.
	Generalization	Untuk menggambarkan relasi khusus antara objek anak (<i>child</i>) dan induk (<i>parent</i>).
	Include dan extends	include menggambarkan perilaku yang wajib dipenuhi agar suatu <i>event</i> dapat terjadi

		<p>extends menunjukkan perilaku tambahan yang hanya terjadi di bawah kondisi tertentu</p>
--	--	---



Tabel 2. 1 Usecase Diagram


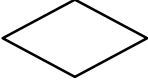

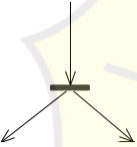
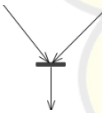
2.10 Diagram Activity

Activity Diagram atau diagram aktivitas adalah salah satu jenis diagram dalam pemodelan UML yang digunakan untuk menggambarkan alur kerja atau proses kegiatan di dalam suatu sistem. Diagram ini membantu menjelaskan langkah-langkah yang dilakukan pengguna ketika menjalankan fitur tertentu, misalnya saat melakukan pemesanan, login, atau pembayaran.

Diagram aktivitas menunjukkan urutan proses yang terjadi dalam sistem dari awal hingga akhir secara logis dan berurutan. Diagram ini menggambarkan bagaimana sistem akan merespons tindakan pengguna secara otomatis, dan sangat berguna untuk memahami alur logika sistem sebelum dikembangkan ke dalam bentuk kode program (Tabrani & Priyandaru, 2021).

Tabel 2. 2 Diagram Activity

Simbol	Nama Komponen	Fungsi
 	<ul style="list-style-type: none"> • <i>Start State</i> • <i>End State</i> 	<p>- Sebagai tanda awal proses dari <i>activity diagram</i>.</p> <p>- Menggambarkan akhir atau terminal dari pada sebuah <i>activity diagram</i></p>

	<i>Activity</i>	mengambarkan sebuah pekerjaan atau tugas dalam workflow
	<i>Decision</i>	mengindikasikan suatu kondisi kemungkinan perbedaan transisi
	<i>State Transition</i>	menunjukkan kegiatan apa berikutnya setelah suatu kegiatan sebelumnya
	<i>Fork/percabangan</i>	menunjukkan kegiatan yang dilakukan secara paralel atau untuk menggabungkan dua kegiatan paralel menjadi satu
	<i>Join (penggabungan)</i>	menunjukkan adanya dekomposisi

2.11 Sequence Diagram

Sequence Diagram adalah salah satu jenis diagram dalam UML (Unified Modeling Language) yang digunakan untuk menggambarkan urutan komunikasi antar objek dalam sistem berdasarkan waktu. Diagram ini memperlihatkan bagaimana objek saling mengirim pesan dan bereaksi terhadap suatu proses, seperti ketika pengguna melakukan login, melakukan transaksi, atau mengakses informasi.

Sequence diagram sangat berguna untuk memodelkan perilaku dinamis sistem karena menyajikan alur komunikasi antar entitas secara kronologis, dari awal hingga akhir suatu aktivitas. Diagram ini membantu menjelaskan "siapa melakukan apa" dan dalam urutan seperti apa, sehingga bisa mempermudah pengembang dan tim analisis memahami sistem yang akan dibangun (Al-Fedaghi, 2021).

2.12 Database dan DBMS

Database adalah tempat penyimpanan data digital yang dirancang agar informasi di dalamnya mudah diakses dan dikelola. Semua data yang dibutuhkan oleh perusahaan—seperti data pelanggan, transaksi, produk, dan laporan—disusun secara teratur di dalam database agar tidak berantakan dan bisa digunakan kapan saja.

Untuk mengelola database ini secara efisien, perusahaan menggunakan Database Management System (DBMS). DBMS adalah perangkat lunak yang berfungsi untuk menyimpan, mengatur, memperbarui, dan mengamankan data yang ada di dalam database. Sistem ini memungkinkan berbagai proses data seperti pencarian, pengolahan, hingga pelaporan dilakukan dengan cepat dan akurat (Aulia Isra & Nasution Padli, 2025).

Pencarian dan akses informasi.

Contoh DBMS yang banyak digunakan saat ini antara lain: MySQL, SQL Server, Microsoft Access, Oracle, dan PostgreSQL.

2.12.1 MySQL

MySQL adalah salah satu perangkat lunak sistem manajemen basis data (DBMS) yang paling banyak digunakan di dunia. MySQL bersifat open source, artinya siapa saja dapat menggunakan, memodifikasi, dan mendistribusikannya secara gratis. Hal ini membuat MySQL menjadi pilihan utama dalam berbagai pengembangan aplikasi, baik dalam skala kecil seperti website pribadi, maupun dalam skala besar seperti sistem informasi perusahaan, layanan e-commerce, hingga aplikasi akademik berbasis cloud.

MySQL merupakan sistem manajemen basis data yang dibangun menggunakan bahasa pemrograman C dan C++. MySQL bekerja dengan mengandalkan bahasa SQL (Structured Query Language) sebagai alat komunikasi antara pengguna dan sistem database. SQL memungkinkan pengguna untuk membuat, mengubah, menampilkan, atau menghapus data dari database melalui perintah-perintah sederhana namun sangat kuat (Putri Permata & Pratama Adhi, 2023).

2.12.2 Software dan Pemrograman Terkait

1. Python

Python adalah salah satu bahasa pemrograman yang saat ini paling banyak digunakan, baik di dunia industri maupun dalam dunia pendidikan. Python dikenal karena kesederhanaan sintaksisnya, kemudahan belajar bagi pemula, dan fleksibilitasnya dalam membangun berbagai macam aplikasi, seperti aplikasi desktop, website, hingga mobile.

Dalam jurnal karya Romzi & Kurniawan (2020) Python dijelaskan sebagai bahasa pemrograman yang cocok untuk pembelajaran logika dan algoritma, karena mendukung struktur pemrograman dasar seperti runtunan (sequence), percabangan (selection), dan perulangan (looping). Python pertama kali dikembangkan oleh Guido van Rossum pada tahun 1990 di Belanda, dan namanya terinspirasi dari acara televisi Monty Python's Flying Circus. Sejak itu, Python berkembang menjadi bahasa pemrograman global yang digunakan oleh banyak pengembang dan perusahaan besar.

2. Editor Jupyter

Jupyter Notebook merupakan platform open-source berbasis web yang memungkinkan integrasi kode, visualisasi, dan dokumentasi dalam satu dokumen interaktif. Fitur ini menjadikannya alat yang sangat efektif dalam bidang data science, pembelajaran mesin, dan penelitian ilmiah.

Menurut Rule *et al.*, (2019), Jupyter Notebook telah menjadi standar *de facto* dalam penelitian data science karena kemampuannya mengintegrasikan kode, output, dan narasi secara dinamis. Penelitian ini menggunakan Jupyter untuk pengembangan model Decision Tree Regression dan eksplorasi data.

3. Bahasa Pemrograman Python dan Library Terkait

Python merupakan bahasa pemrograman tingkat tinggi yang bersifat open-source dan mendukung berbagai paradigma pemrograman, termasuk prosedural, berorientasi objek, dan fungsional. Sintaksisnya yang sederhana dan kemampuannya dalam mengintegrasikan berbagai pustaka (*library*)

menjadikan Python sebagai pilihan utama dalam pengembangan aplikasi data science dan machine learning. Menurut (Joshi & Tiwari, 2023), Python menyediakan berbagai pustaka yang dapat dikategorikan ke dalam tiga kelompok utama: pengumpulan data, analisis dan pemrosesan data, serta visualisasi data.

Menurut Sundaram *et al.*, (2023) ,Dalam konteks penelitian ini, beberapa pustaka Python yang digunakan meliputi:

- a. Pandas: Digunakan untuk manipulasi dan analisis data tabular. Pandas menyediakan struktur data *DataFrame* yang memungkinkan operasi seperti pembersihan data, transformasi, dan agregasi dilakukan dengan sintaks yang ringkas namun ekspresif. Kemampuan Pandas dalam menangani data yang memiliki nilai hilang (*missing values*), melakukan penggabungan (*merge*) dan penyatuan (*join*) dataset, serta operasi *time series* sangat mendukung tahap pra-pemrosesan data dalam penelitian ini.
- b. NumPy: Digunakan untuk komputasi numerik, menyediakan array multidimensi yang efisien dan berbagai fungsi matematika tingkat tinggi. NumPy menjadi fondasi bagi banyak pustaka ilmiah Python lainnya dan memungkinkan komputasi numerik berjalan dengan performa tinggi, yang sangat krusial dalam proses pelatihan model machine learning.
- c. Matplotlib dan Seaborn: Digunakan untuk visualisasi data. Matplotlib adalah pustaka standar yang digunakan untuk membuat grafik dan plot dalam Python, mendukung berbagai jenis visualisasi seperti diagram batang, histogram, dan scatter plot. Seaborn dibangun di atas Matplotlib dan menyediakan antarmuka tingkat tinggi yang lebih intuitif dalam membuat

visualisasi statistik yang menarik secara visual. Kombinasi kedua pustaka ini digunakan dalam penelitian untuk mengeksplorasi distribusi data, hubungan antar variabel, serta mempresentasikan hasil prediksi model.

- d. Scikit-learn: Merupakan pustaka Python yang populer untuk implementasi algoritma machine learning, termasuk *Decision Tree Regression*. Scikit-learn menyediakan antarmuka yang konsisten dan dokumentasi yang lengkap, sehingga memudahkan pengguna dalam membangun dan mengevaluasi model prediktif.

4. Library Scikit-learn

- a) Scikit-learn merupakan pustaka open-source berbasis Python yang secara luas digunakan dalam pengembangan model machine learning. Pustaka ini menyediakan berbagai algoritma pembelajaran mesin, baik yang bersifat supervised maupun unsupervised. Salah satu keunggulan utamanya adalah kemudahan penggunaan, efisiensi dalam kinerja, dokumentasi yang lengkap, serta konsistensi pada antarmuka pemrogramannya (API). Dengan ketergantungan yang minimal dan lisensi BSD yang terbuka, Scikit-learn banyak digunakan dalam lingkungan akademis maupun industry (Pedregosa *et al.*, 2012).
- b) Pustaka ini mendukung integrasi dengan pustaka lain seperti NumPy dan SciPy, serta menyediakan fitur-fitur penting seperti pemilihan model, preprocessing, dan validasi silang (*cross-validation*), yang membuatnya sangat fleksibel dan kuat untuk berbagai kebutuhan analisis data. Antarmuka yang konsisten juga memudahkan pengguna dalam menerapkan dan membandingkan berbagai metode algoritma secara efisien.

c) Dalam penelitian ini, Scikit-learn dimanfaatkan untuk membangun model Decision Tree Regression. Penggunaannya mempermudah proses pelatihan model dan evaluasi performa secara sistematis, sehingga mendukung analisis prediktif yang dibutuhkan dalam studi ini.

d) Bahasa Pemrograman PHP

PHP adalah bahasa pemrograman yang paling banyak digunakan untuk membangun website dinamis. PHP bekerja di sisi server (*server-side scripting*), artinya seluruh proses atau perintah yang ditulis dalam PHP akan dijalankan di server, lalu hasilnya dikirim ke browser dalam bentuk HTML. Karena sifatnya yang open-source dan fleksibel, PHP dapat digunakan secara bebas, dimodifikasi, dan dikembangkan oleh siapa saja sesuai kebutuhan.

Menurut Endra *et al.*, (2021) PHP terbagi menjadi dua pendekatan utama, yaitu PHP Native dan PHP Framework seperti Laravel.

- a) PHP Native adalah penulisan kode PHP secara langsung tanpa bantuan kerangka kerja apa pun. Pendekatan ini memberi kebebasan penuh kepada pengembang untuk menentukan struktur dan alur logika program, namun juga memiliki kelemahan karena pengelolaan kodenya bisa menjadi rumit dan rawan error jika tidak didokumentasikan dengan baik.
- b) Sementara itu, framework PHP seperti Laravel menawarkan struktur kerja yang lebih rapi dan efisien. Framework ini sudah menyediakan banyak pustaka (*library*) dan fungsi siap pakai, sehingga proses pengembangan lebih cepat, aman, dan mudah dikelola, terutama dalam proyek skala besar atau kerja tim.

2.13 Tinjauan Penelitian Terdahulu

Beberapa penelitian yang relevan dengan topik analisis sentimen menggunakan metode Transformer dan data mining antara lain:

2.13.1 Paper 1

Akhmad (2023) dalam penelitiannya yang berjudul "*Analisis Sentimen Ulasan Aplikasi DLU Ferry*" menggunakan model IndoBERT untuk mengklasifikasikan opini pelanggan ke dalam tiga kategori, yaitu positif, netral, dan negatif. Hasil penelitian ini menunjukkan bahwa model IndoBERT mampu mencapai akurasi sebesar 86%. Penelitian ini membuktikan bahwa metode Transformer sangat cocok untuk menganalisis opini terhadap layanan publik berbasis digital.

2.13.2 Masalah Penelitian

Bagaimana menganalisis sentimen ulasan pelanggan terhadap aplikasi DLU Ferry secara otomatis agar dapat diketahui aspek mana saja yang perlu diperbaiki.

2.13.3 Penyelesaian yang Dilakukan

Penelitian ini melakukan analisis sentimen terhadap ulasan pelanggan aplikasi DLU Ferry di Google Play Store menggunakan metode Bidirectional Encoder Representations from Transformers (BERT) dengan model bahasa Indonesia IndoBERT-base.

2.13.4 Metodologi yang Digunakan

Metode penelitian yang dipakai adalah:

- a) Pengumpulan Data: Scraping 1575 ulasan aplikasi dari Google Play Store.
- b) Labelisasi Data: Berdasarkan rating bintang.

- a. Rating 1-2 = Negatif
 - b. Rating 3 = Netral
 - c. Rating 4-5 = Positif
- c) Preprocessing:
- a. Case folding
 - b. Data cleaning
 - c. Tokenisasi (NLTK)
 - d. Normalisasi kata tidak baku
- d) Model: IndoBERT-base (pre-trained model Bahasa Indonesia)
- e) Teknik: Fine-tuning
- f) Hyperparameter:
- a. Batch size: 32
 - b. Learning rate: $3e-6$
 - c. Epoch: 5
- g) Evaluasi:
- a. Accuracy
 - b. Confusion Matrix
 - c. Precision
 - d. Recall
 - e. F1-Score

2.13.5 Hasil Penelitian

Hasil pengujian pada penelitian memperoleh akurasi sebesar 86% dengan pemilihan *hyperparameter*, yaitu *batch size* 32, *learning rate* $3e-6$, dan *epoch* 5. Penelitian ini menggunakan data sebanyak 1575 ulasan yang terdiri dari sentimen positif, netral, dan negatif. Penelitian ini akan mengukur performa metode *Bidirectional*

Encoder Representations from Transformers (BERT) dalam melakukan klasifikasi sentimen menggunakan *Pretrained* model *IndoBERT-base* dengan teknik *fine-tuning*.

2.14 Paper 2

Menurut Putu *et al.*, (2023) dalam penelitiannya yang berjudul “*Analisis Sentimen terhadap Perundungan Siber pada Twitter Menggunakan BERT*” menggunakan model BERT untuk mengklasifikasikan komentar pengguna Twitter ke dalam tiga kategori sentimen, yaitu netral, ujar kebencian, dan kasar. Penelitian ini bertujuan mendeteksi perundungan siber secara otomatis dengan pendekatan deep learning berbasis transformer. Hasil penelitian menunjukkan bahwa model BERT berhasil mencapai akurasi sebesar 81%, dengan f1-score yang stabil pada ketiga kelas sentimen. Hal ini menunjukkan bahwa BERT cukup efektif digunakan untuk mendeteksi konten negatif di media sosial secara otomatis.

2.14.1 Masalah penelitian

Masalah yang diangkat dalam jurnal ini adalah maraknya perundungan siber di media sosial, terutama Twitter, yang sering kali tidak terdeteksi secara cepat. Banyak ujaran kebencian dan kata kasar menyebar luas tanpa adanya filter yang akurat, sehingga dibutuhkan sistem cerdas untuk mendeteksi dan mengklasifikasikan konten semacam itu secara otomatis

2.14.2 Penyelesaian yang Dilakukan

Penelitian ini menggunakan model BERT (Bidirectional Encoder Representations from Transformers) untuk mendeteksi dan mengklasifikasi komentar di Twitter yang termasuk kategori netral, ujaran kebencian, dan bahasa

kasar. Data diambil dari dataset publik yang telah dilabeli dan divalidasi sebelumnya

2.14.3 Metodologi yang Digunakan

- a) Pengumpulan data dari Twitter.
- b) Preprocessing: tokenisasi, filtering, dan stemming.
- c) Model: fine-tuning BERT untuk klasifikasi teks.
- d) Evaluasi: menggunakan confusion matrix dan metrik akurasi, precision, recall, dan f1-score pada data validasi dan data uji

2.14.4 Hasil penelitian

Model BERT mampu mengklasifikasikan komentar dengan akurasi sebesar 81%, baik pada data validasi maupun data uji. Metrik evaluasi lainnya juga menunjukkan performa yang baik, dengan f1-score rata-rata mendekati 0.81, menunjukkan bahwa sistem dapat mengidentifikasi perundungan siber secara cukup akurat.

2.15 Paper 3

Menurut penelitian Ihsan Septian & Lucia Kharisma (2023) dalam penelitiannya yang berjudul *“Implementasi BERT untuk Analisis Sentimen Komentar Pengguna Aplikasi Dana di Instagram”* mengembangkan sistem klasifikasi komentar Instagram menggunakan IndoBERT, yang merupakan varian BERT untuk Bahasa Indonesia. Penelitian ini berfokus pada klasifikasi sentimen positif dan negatif terhadap layanan aplikasi Dana. Dengan melakukan pelatihan model dan pengujian secara menyeluruh, sistem berhasil mencapai

akurasi sebesar 91% pada data uji, dan f1-score di atas 0.90. Selain itu, model ini juga diintegrasikan ke dalam aplikasi web berbasis Streamlit untuk penggunaan secara langsung. Hasil penelitian membuktikan bahwa IndoBERT efektif untuk menganalisis opini publik di media sosial berbahasa Indonesia.

2.15.1 Masalah Penelitian

Penelitian ini berangkat dari banyaknya komentar negatif di Instagram resmi aplikasi Dana, yang menunjukkan keluhan atau ketidakpuasan pengguna. Namun, perusahaan tidak bisa memeriksa semua komentar secara manual karena jumlahnya besar. Maka dibutuhkan sistem otomatis yang mampu mengklasifikasikan komentar berdasarkan sentimen positif dan negatif

2.15.2 Penyelesaian yang Dilakukan

Solusi yang diambil adalah mengimplementasikan model IndoBERT (versi BERT khusus Bahasa Indonesia) untuk menganalisis komentar pengguna di Instagram. Model dilatih menggunakan dataset hasil scraping dan dilabeli otomatis dengan bantuan NLP transformer. Setelah training, model ini digunakan dalam antarmuka berbasis web (Streamlit) untuk melakukan klasifikasi secara langsung.

2.15.3 Metodologi yang Digunakan

- a) Scraping komentar dari akun Instagram Dana.
- b) Label otomatis menggunakan NLP dan validasi manual.
- c) Penyeimbangan data (oversampling) karena ketidakseimbangan antara sentimen negatif dan positif.

- d) Preprocessing (lowercase, stemming, tokenisasi).
- e) Pelatihan model IndoBERT dengan parameter: batch size 32, epoch 10.
- f) Evaluasi menggunakan confusion matrix, akurasi, precision, recall, dan f1-score.
- g) Deployment sistem menggunakan Streamlit agar dapat digunakan secara langsung oleh pengguna melalui web

2.15.4 Hasil Penelitian

Model IndoBERT mampu menghasilkan:

- a. Akurasi data latih sebesar 98%,
- b. Akurasi validasi sebesar 93%,
- c. Akurasi pengujian akhir sebesar 91%, dengan f1-score pada sentimen positif dan negatif di atas 0.90. Ini menunjukkan bahwa metode ini sangat efektif untuk mengklasifikasikan komentar pengguna secara otomatis dan akurat.