

SENTIMENT ANALYSIS OF CAMPUS SERVICES
CAPSTONE PROJECT



DI SUSUN OLEH :

M. ZUHRI AL KAURI	G1A023029
GHAZI AL-GHIFARI	G1A023053
I NYOMAN DIMAS KRESNA A.	G1A023077

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS BENGKULU
2025

HALAMAN PENGESAHAN
PROPOSAL CAPSTONE PROJECT
SENTIMENT ANALYSIS OF CAMPUS SERVICES
KOMPUTASI CERDAS DAN DATA SCIENCE

DI SUSUN OLEH :

M. ZUHRI AL KAURI	G1A023029
GHAZI AL-GHIFARI	G1A023053
I NYOMAN DIMAS KRESNA A.	G1A023077

Telah disetujui oleh :

Dosen Pembimbing Capstone

Ir. Arie Vatesia, S.T, M.TI, Ph.D.

NIDN: 0004028501

DAFTAR ISI

HALAMAN PENGESAHAN.....	1
DAFTAR ISI	2
DAFTAR TABEL	3
DAFTAR GAMBAR	4
DAFTAR ISTILAH	4
BAB I. PENDAHULUAN	4
1.1 Latar Belakang	4
1.2 Rumusan Masalah	6
1.3 Batasan Masalah	6
1.4 Tujuan Capstone Project	7
1.5 Manfaat Capstone Project	7
BAB II. ANALISIS KEBUTUHAN DAN PERANCANGAN	9
2.1 Analisis Kebutuhan	9
2.2 Perancangan	11
BAB III. RENCANA CAPSTONE PROJECT	6
3.1 Tahapan Capstone Project	6
3.2 Timeline Capstone Project	7
3.3 Alat dan Bahan	8
DAFTAR PUSTAKA	8
LAMPIRAN	9

DAFTAR TABEL

Table 1 Timeline Capstone Project	22
Table 2 Hardware	24
Table 3 Platform Pengembangan.....	25
Table 4 Library Python untuk Analis Sentimen	26
Table 5 Library untuk Visualisasi dan Analisis	27
Table 6 Tools Pengumpulan dan Manajemen Data.....	28
Table 7 Dataset dan Bahan Penelitian	29
Table 8 Platform Kolaborasi Tim.....	29
Table 9 Anggaran	31

DAFTAR GAMBAR

Gambar 1 Flow Chart Sentiment Analysis	13
Gambar 2 Use Case Diagram	Error! Bookmark not defined.
Gambar 3 Activity Diagram	Error! Bookmark not defined.
Gambar 4 Sequence Diagram	Error! Bookmark not defined.
Gambar 5 Interface Upload Dataset	15
Gambar 6 Interface Panduan Penggunaan	16
Gambar 7 Interface Pratinjau Data	17
Gambar 8 Interface Konfigurasi Analisis	17
Gambar 9 Interface Hasil Training dan Evaluasi	17
Gambar 10 Interface Distribusi Sentiment	18

DAFTAR ISTILAH

C <i>Campus Services</i>	Bagian <i>gameplay</i> yang menjelaskan tantangan atau halangan apa saja yang terdapat didalam <i>game</i> yang dimainkan
D <i>Data-driven</i>	Pendekatan dalam pengambilan keputusan yang berdasarkan pada analisis dan interpretasi data, bukan intuisi atau pengalaman semata.
F <i>Feature Extraction</i>	Tahap dalam <i>machine learning</i> di mana data teks yang telah dibersihkan diubah menjadi representasi numerik (fitur) agar dapat diproses oleh algoritma.
M <i>Machine Learning</i>	Cabang dari kecerdasan buatan yang memungkinkan sistem untuk belajar dari data, mengidentifikasi pola, dan membuat keputusan dengan intervensi manusia yang minimal.
N <i>Natural Language Processing</i>	Bidang ilmu komputer dan kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa manusia, khususnya dalam hal pemrosesan dan analisis data teks.
P <i>Preprocessing Data</i>	Proses pembersihan dan persiapan data teks mentah sebelum dianalisis, meliputi tahapan seperti <i>case folding</i> , <i>tokenizing</i> , <i>stopword removal</i> , dan <i>stemming</i> .
S <i>Scraping</i>	PTeknik pengumpulan data dari situs web atau media sosial secara otomatis untuk mengekstrak informasi spesifik, seperti komentar atau ulasan pengguna.
<i>Stopword Removal</i>	Salah satu tahap dalam <i>preprocessing</i> data teks untuk menghapus kata-kata umum yang tidak memiliki makna signifikan (contoh: dan, yang, di).
<i>Stemming</i>	Proses mengubah kata menjadi bentuk dasarnya dengan menghilangkan imbuhan untuk menormalkan teks.
<i>Tokenizing</i>	Proses memecah kalimat atau teks panjang menjadi unit-unit yang lebih kecil seperti kata atau frasa, yang disebut token.

BAB I.

PENDAHULUAN

1.1 Latar Belakang

Peningkatan kualitas layanan pendidikan tinggi merupakan aspek penting dalam upaya membangun sistem akademik yang efektif dan berdaya saing. Perguruan tinggi dituntut tidak hanya menyediakan fasilitas akademik, tetapi juga layanan administrasi, sistem informasi, serta dukungan non-akademik yang mampu menunjang pengalaman belajar mahasiswa. Di Indonesia, survei kepuasan mahasiswa menunjukkan bahwa masih terdapat gap antara harapan dan realita pelayanan kampus, baik dari sisi akademik maupun non-akademik (Utami, 2020). Ketidakpuasan tersebut dapat menurunkan kepercayaan mahasiswa terhadap institusi pendidikan dan berdampak pada citra universitas di tingkat nasional maupun internasional (Napitupulu et al., 2020).

Salah satu penyebab utama permasalahan layanan kampus adalah keterbatasan mekanisme umpan balik yang tersedia. Sistem evaluasi yang masih berbasis survei manual sering kali tidak efektif dalam menangkap opini mahasiswa secara luas, cepat, dan berkelanjutan (Telkom University, 2022). Selain itu, banyak institusi belum memanfaatkan potensi data mahasiswa yang tersebar di media sosial, forum daring, maupun platform internal sebagai sumber informasi berharga untuk peningkatan layanan (Baihaqi & Fitriani, 2023). Kurangnya sistem analisis kepuasan berbasis teknologi juga membuat pihak kampus sulit memahami tren sentimen mahasiswa secara real time (Shaik et al., 2020).

Permasalahan ini penting untuk segera diselesaikan, mengingat kepuasan mahasiswa menjadi salah satu indikator kunci keberhasilan perguruan tinggi dalam menyelenggarakan pendidikan berkualitas. Hasil penelitian menunjukkan bahwa kualitas layanan akademik, administrasi, dan fasilitas memiliki pengaruh signifikan

terhadap motivasi belajar dan loyalitas mahasiswa (Dake & Gyimah, 2022). Lebih lanjut, analisis sentimen dari ulasan mahasiswa dapat menjadi sumber data strategis untuk pengambilan keputusan manajemen pendidikan tinggi (Dalipi et al., 2018). Tanpa adanya mekanisme berbasis data, kampus berisiko kehilangan peluang untuk meningkatkan kualitas layanan dan daya saing di era digital.

Kesesuaian masalah ini dengan tema teknologi informasi untuk peningkatan kualitas pendidikan sangat jelas. Analisis sentimen sebagai bagian dari natural language processing (NLP) telah terbukti efektif dalam menggali opini pengguna terhadap layanan publik, termasuk dalam konteks pendidikan (Kastrati et al., 2020). Dengan memanfaatkan pendekatan berbasis data science, perguruan tinggi dapat mengembangkan sistem yang mampu mengekstraksi insight dari data teks mahasiswa untuk mendukung pengambilan keputusan strategis (Alzaid et al., 2022). Hal ini sejalan dengan tren global yang menekankan pemanfaatan kecerdasan buatan untuk mendukung tata kelola institusi pendidikan (IJIET, 2024).

Beberapa solusi dapat ditawarkan untuk mengatasi permasalahan ini. Pertama, survei manual berbasis kuesioner dapat digunakan untuk mengumpulkan data kepuasan layanan. Namun, metode ini cenderung memakan waktu, membutuhkan biaya besar, serta terbatas pada responden tertentu. Kedua, analisis sentimen berbasis machine learning terhadap ulasan mahasiswa, baik di media sosial maupun platform internal kampus, dapat digunakan untuk memperoleh informasi secara lebih cepat, akurat, dan komprehensif. Metode ini dinilai lebih efektif karena mampu mengolah data dalam jumlah besar sekaligus mengidentifikasi pola sentimen mahasiswa terhadap berbagai aspek layanan kampus (Baihaqi & Fitriani, 2023).

Berdasarkan studi terdahulu, solusi berbasis analisis sentimen menggunakan NLP dan machine learning terbukti unggul dibanding survei manual. Misalnya, penelitian oleh Kastrati et al. (2020) menunjukkan efektivitas deep learning dalam mengidentifikasi persepsi mahasiswa terkait pembelajaran daring. Begitu pula

penelitian Dalipi et al. (2018) yang menemukan bahwa pemanfaatan data media sosial untuk analisis kepuasan mahasiswa mampu memberikan insight lebih mendalam dibanding metode survei konvensional. Oleh karena itu, capstone project ini memilih analisis sentimen berbasis machine learning sebagai solusi terbaik untuk mendukung peningkatan kualitas layanan kampus melalui pendekatan data-driven.

1.2 Rumusan Masalah

Dari latar belakang yang telah dipaparkan, rumusan masalah dalam Capstone Project ini adalah:

1. Bagaimana cara mengidentifikasi dan mengklasifikasikan sentimen mahasiswa terhadap *campus services* (positif, negatif, atau netral) secara otomatis menggunakan pendekatan *Natural Language Processing* (NLP) dan *Machine Learning*?
2. Bagaimana perbandingan performa antara model *machine learning* klasik (Naive Bayes, Logistic Regression) dengan model *deep learning* pra-latih (IndoBERT) dalam mengklasifikasikan sentimen pada konteks layanan kampus di Indonesia?
3. Bagaimana merancang sebuah sistem analisis sentimen yang mampu menyajikan hasil klasifikasi sentimen mahasiswa dalam bentuk visualisasi interaktif sehingga dapat membantu pihak kampus dalam mengambil keputusan berbasis data?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan dapat diselesaikan dengan baik, maka ditetapkan batasan masalah sebagai berikut:

1. Data yang digunakan berasal dari data primer hasil yaitu komentar mahasiswa yang diperoleh dari kanal internal maupun media sosial, dengan jumlah terbatas sesuai kebutuhan penelitian.

2. Analisis sentimen hanya dikategorikan ke dalam tiga kelas, yaitu positif, negatif, dan netral, tanpa melakukan analisis lebih mendalam terhadap emosi spesifik (misalnya marah, senang, kecewa, dan sebagainya)
3. Batasan algoritma pada proyek ini berfokus pada implementasi dan perbandingan dua pendekatan analisis sentimen yang sudah teruji dalam sebuah aplikasi web interaktif. Pendekatan pertama adalah *machine learning* klasik yang mencakup model Naive Bayes dan Logistic Regression, di mana model-model ini dilatih dari nol (*training from scratch*) menggunakan *dataset* berlabel yang disediakan oleh pengguna. Sementara itu, pendekatan kedua adalah *deep learning* dengan memanfaatkan model *Transformer* pra-latih (*pre-trained*) yaitu IndoBERT/Roberta untuk melakukan prediksi langsung (*inference*) tanpa memerlukan proses *training* pada data pengguna. Dengan demikian, proyek ini tidak mencakup perancangan atau pengembangan arsitektur model baru dari awal, melainkan hanya berfokus pada penerapan dan evaluasi model-model tersebut. Sistem yang dibangun hanya berupa prototipe sederhana yang menampilkan hasil analisis sentimen dalam bentuk visualisasi (grafik/tabel/dashboard), bukan sistem produksi berskala besar.
4. Lingkup penelitian ini berfokus pada penerapan *sentiment analysis* dalam pendidikan tinggi untuk mengidentifikasi kepuasan mahasiswa. Hasil tinjauan sistematis menunjukkan bahwa penerapan terbanyak ada pada aspek akademik (77,8%), diikuti kesejahteraan mahasiswa (18,5%), dan citra institusi (3,7%). Kontribusi utamanya meliputi identifikasi pengalaman mahasiswa, perbaikan aktivitas pengajaran (48,15%), serta dukungan pengambilan keputusan manajerial. Namun, keterbatasan juga muncul, seperti rendahnya partisipasi mahasiswa, ketidakseimbangan data, keterbatasan populasi, hingga ketiadaan leksikon khusus (Chamorro-Atalaya et al., 2025). Dengan demikian, penelitian ini berada dalam lingkup

tersebut untuk memperkuat pemanfaatan *sentiment analysis* dalam evaluasi layanan kampus.

1.4 Tujuan Capstone Project

Tujuan dari Capstone Project ini adalah membangun sebuah prototipe sistem analisis sentimen berbasis *Natural Language Processing* (NLP) dan *Machine Learning* yang mampu mengolah opini mahasiswa terhadap *campus services* yang bersifat tekstual menjadi data terstruktur, mengklasifikasikannya ke dalam kategori positif, negatif, atau netral, serta menyajikan hasil analisis dalam bentuk visualisasi interaktif untuk mendukung evaluasi dan pengambilan keputusan pihak kampus

1.5 Manfaat Capstone Project

Capstone Project ini diharapkan memberikan manfaat nyata baik bagi mahasiswa maupun pihak kampus. Bagi kampus, sistem analisis sentimen dapat membantu memperoleh gambaran objektif mengenai opini mahasiswa terhadap *campus services*, sehingga perbaikan layanan dapat dilakukan lebih tepat sasaran. Bagi mahasiswa, proyek ini memberi saluran terstruktur untuk menyampaikan pendapat mereka agar lebih diperhatikan oleh pihak universitas. Dalam jangka panjang, penerapan sistem ini mendukung terciptanya budaya pengambilan keputusan berbasis data di lingkungan perguruan tinggi, meningkatkan kualitas layanan secara berkelanjutan, memperkuat citra positif universitas, sekaligus memberikan kontribusi akademis dalam pengembangan bidang *Data Science* dan *Komputasi Cerdas* melalui penerapan NLP dan Machine Learning di sektor pendidikan

BAB II.

ANALISIS KEBUTUHAN DAN PERANCANGAN

2.1 Analisis Kebutuhan

Analisis kebutuhan merupakan langkah penting untuk merumuskan solusi yang sesuai dengan permasalahan yang diangkat pada Capstone Project ini. Proses analisis kebutuhan dilakukan melalui tiga tahapan utama, yaitu requirement gathering (pengambilan data), requirement analysis (analisis kebutuhan), dan perumusan daftar kebutuhan sistem.

2.1.1 Proses Pengambilan Data

Pengumpulan data dilakukan untuk mendapatkan informasi yang akurat mengenai opini mahasiswa terhadap layanan kampus. Sesuai batasan masalah yang telah ditentukan pada Bab I, pengambilan data dilakukan melalui:

1. Data Primer

Observasi dilakukan dengan mengumpulkan data opini melalui *crawling* pada media sosial. Proses *crawling* dilakukan terhadap komentar, ulasan, dan diskusi publik yang relevan dengan topik penelitian. Data yang diperoleh berasal dari berbagai platform media sosial yang banyak digunakan oleh masyarakat, sehingga dapat mencerminkan pandangan dan pengalaman pengguna secara lebih luas.

Data primer ini dapat menjadi gambaran yang cukup representatif mengenai opini publik, karena bersumber dari interaksi nyata pengguna di dunia digital. Meskipun demikian, data tetap perlu diproses dan dibersihkan terlebih dahulu agar dapat diolah secara sistematis dalam analisis sentimen.

2.1.2 Proses Analisis Kebutuhan

Data yang diperoleh kemudian dianalisis menggunakan beberapa pendekatan:

1. Persona

Persona yang digunakan adalah mahasiswa kampus sebagai pengguna utama layanan. Mahasiswa ini memiliki kebutuhan untuk menyampaikan keluhan/opini dengan cara yang mudah, serta berharap pihak kampus merespon dan memperbaiki layanan.

2. *User Journey Map*

Alur pengalaman mahasiswa:

1. Pemanfaatan layanan kampus, meliputi administrasi, jaringan Wi-Fi, dan klinik kampus
 2. Merasakan pengalaman yang tidak memuaskan.
 3. Mengungkapkan keluhan di media sosial, grup chat, atau forum mahasiswa.
 4. Masukan tersebut tidak tersampaikan ke pihak kampus secara resmi.
- Dari alur diatas maka didapatkan titik permasalahan yaitu masukan mahasiswa tidak tersampaikan ke pihak kampus sehingga masukan tidak terkelola dengan baik

3. *Thematic Analysis*

Dari hasil survei dan observasi, keluhan mahasiswa dapat dikelompokkan menjadi empat tema utama :

1. Administrasi
2. Fasilitas
3. Teknologi Informasi
4. Layanan Kesehatan

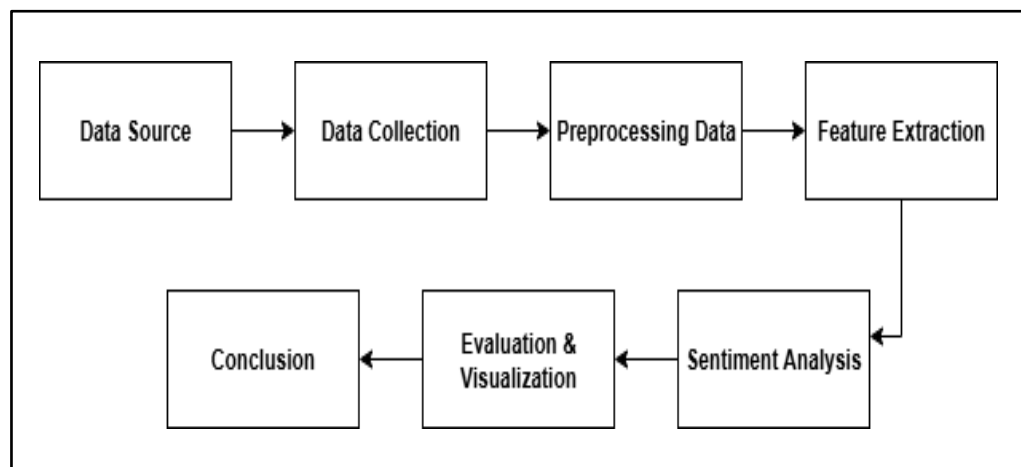
Dari analisis tersebut terlihat bahwa sistem yang dirancang harus mampu mengolah opini mahasiswa secara otomatis, menyajikan hasil dalam

kategori sentimen, serta memvisualisasikan hasilnya agar mudah dipahami oleh pihak kampus.

2.2 Perancangan

2.2.1 Alur Kerja *Sentiment Analysis*

Dalam perancangan ini digunakan alur kerja *Sentiment Analysis* yang disusun secara sistematis untuk mengubah data opini menjadi informasi yang terukur. Pendekatan ini sesuai dengan tema penelitian karena berfokus pada pengolahan teks ulasan atau komentar dari berbagai sumber menjadi data yang dapat dianalisis dengan bantuan *Natural Language Processing (NLP)* dan *Machine Learning*.



Gambar 1 *Flow Chart Sentiment Analysis*

Diatas merupakan bentuk flow chart akan digunakan dalam perancangan Capstone Project ini, terdapat beberapa tahapan pada metode tersebut yang dapat dijelaskan seperti berikut :

1. Data Source

Pada tahap ini ditentukan sumber data yang akan digunakan untuk analisis. Data dapat berasal dari berbagai sumber seperti API publik (Twitter API, YouTube API), web scraping dari situs e-commerce atau forum, dataset

terbuka dari Kaggle, maupun hasil survei manual. Pemilihan sumber data yang relevan akan memengaruhi kualitas hasil analisis. Data

2. Data Collection

Setelah sumber ditentukan, langkah berikutnya adalah mengumpulkan data berupa teks ulasan, komentar, atau opini. Data biasanya juga disertai dengan informasi tambahan seperti waktu, rating, atau identitas pengguna. Hasil pengumpulan data disimpan dalam format terstruktur (CSV, database) agar mudah diproses lebih lanjut.

3. Data Preparation

Tahap ini berfokus pada pembersihan dan normalisasi data teks. Proses umum yang dilakukan meliputi:

- a. *Cleaning* : menghapus tanda baca, angka, URL, dan karakter khusus.
 - b. *Case Folding* : mengubah semua huruf menjadi huruf kecil.
 - c. *Tokenizing* : memecah kalimat menjadi potongan kata.
 - d. *Stopword Removal* : menghapus kata-kata umum yang tidak memiliki makna penting (contoh: “yang”, “dan”).
 - e. *Stemming/Lemmatization*: mengubah kata ke bentuk dasarnya.
- Preprocessing sangat penting untuk meningkatkan akurasi analisis.

4. Feature Extraction

Pada tahap ini, data teks yang sudah dibersihkan diubah menjadi representasi numerik agar bisa diproses oleh algoritma komputer. Beberapa metode umum adalah Bag of Words, TF-IDF, dan Word Embedding (Word2Vec, GloVe, BERT). Representasi ini akan menjadi fitur input bagi model analisis sentimen.

5. Sentiment Analysis

Setelah data direpresentasikan, dilakukan klasifikasi sentimen. Terdapat dua pendekatan utama:

- a. **Lexicon-Based**: menggunakan kamus kata yang memiliki skor positif atau negatif.

- b. **Machine Learning/Deep Learning:** menggunakan algoritma klasifikasi seperti Naïve Bayes, SVM, Logistic Regression, LSTM, atau BERT.

Hasil analisis akan berupa label sentimen (positif, negatif, atau netral).

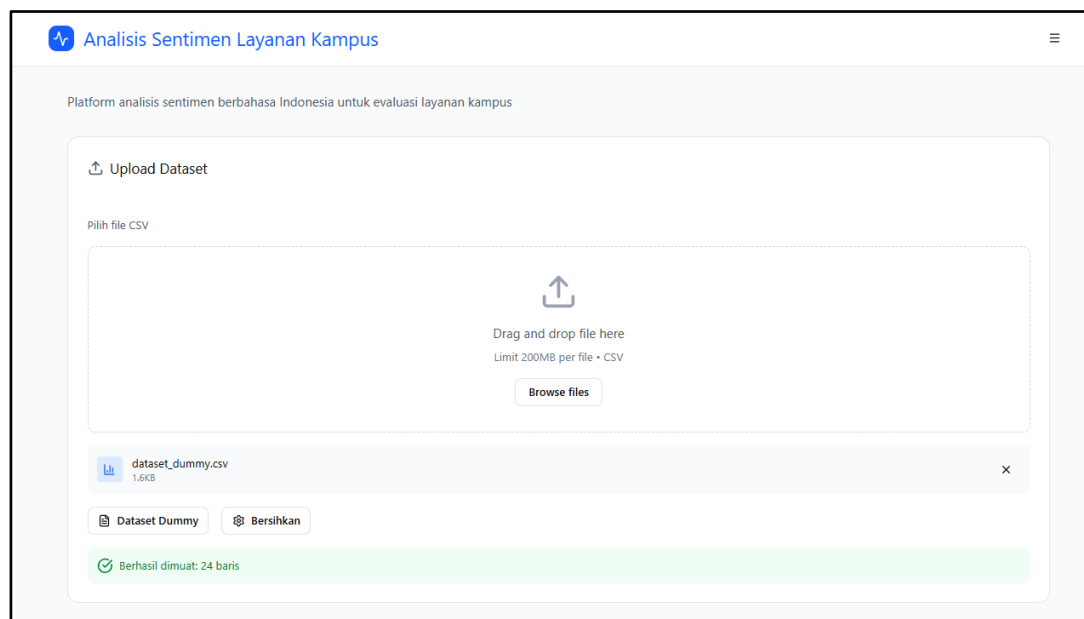
6. Evaluation & Visualization

Untuk memastikan kualitas hasil analisis, dilakukan evaluasi menggunakan metrik seperti akurasi, precision, recall, dan F1-score. Selain itu, hasil juga divisualisasikan dalam bentuk grafik, pie chart, histogram, atau wordcloud untuk memudahkan interpretasi pola sentimen.

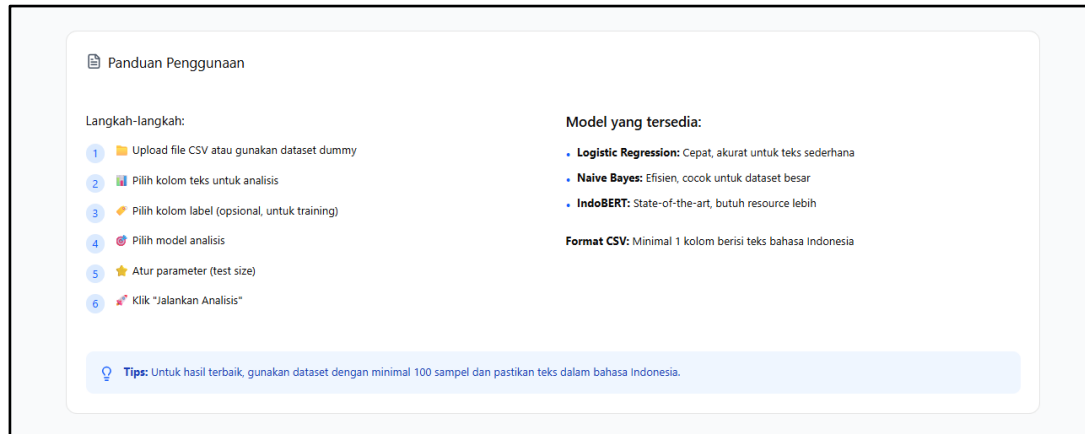
7. Conclusion

Tahap terakhir adalah menyimpulkan hasil analisis. Kesimpulan berisi gambaran umum mengenai kecenderungan opini (misalnya mayoritas positif atau negatif), pola yang ditemukan, serta rekomendasi berdasarkan hasil analisis. Tahap ini sangat penting untuk mendukung pengambilan keputusan yang berbasis data.

2.2.2 Perancangan Interface



Gambar 2 Interface Upload Dataset



Gambar 3 Interface Panduan Penggunaan

Pratinjau Data

Baris: 24 Kolom: 2 Missing: 0 Ukuran: 1.6KB

Info Kolom

#	content	label
0	Pelayanan akademik hari ini sangat cepat dan ramah, terima kasih!	positive
1	Sistem informasi kampus sering down, bikin kesal.	negative
2	Antrian administrasi cukup tertib, tapi masih bisa ditingkatkan.	neutral
3	Fasilitas perpustakaan lengkap dan nyaman untuk belajar.	positive
4	Akses WiFi kampus lambat sekali, tidak bisa upload tugas.	negative

Gambar 4 Interface Pratinjau Data

Konfigurasi Analisis

Kolom Teks: content Model: Logistic Regression

Kolom Label (optional): label Test Set: 0.2 (slider from 0.10 to 0.40)

Jalankan Analisis

Preprocessing selesai: 24/24 teks valid

Gambar 5 Interface Konfigurasi Analisis

Hasil Training & Evaluasi

Metrik Evaluasi

Akurasi: 60.00%

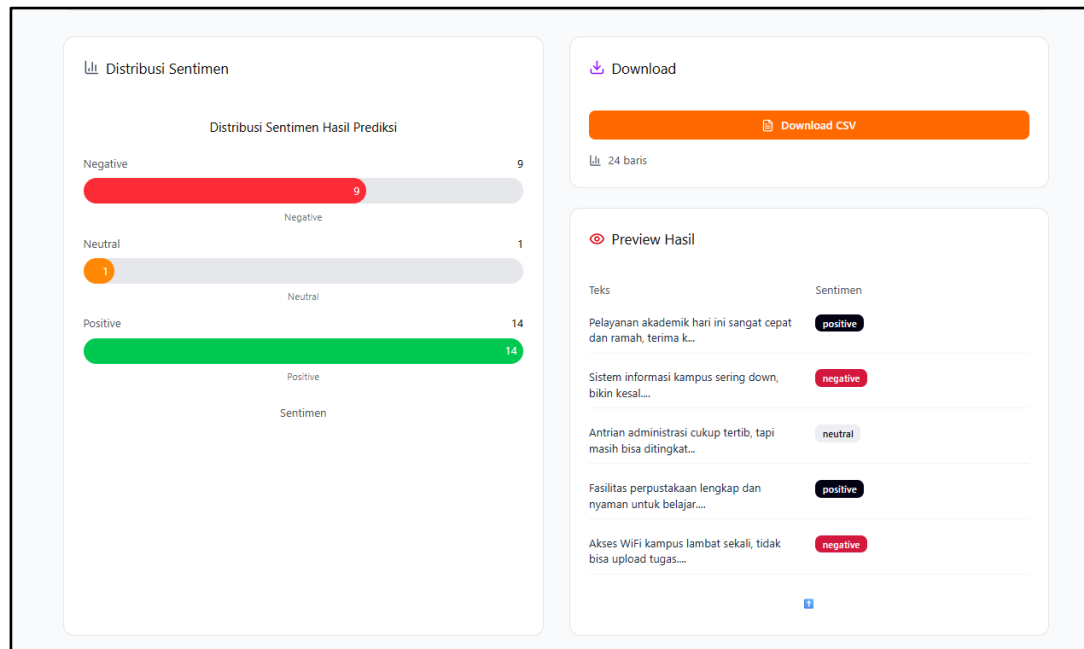
Sentimen	Precision	Recall	F1	Support
Negative	1.000	0.500	0.667	2
Neutral	0.000	0.000	0.000	1
Positive	0.500	1.000	0.667	2

Confusion Matrix

		Predicted		
		Neg	Neu	Pos
Actual	Neg	1	0	1
	Neu	0	0	1
	Pos	0	0	2

Legend: 0.0 (light blue), 1.0 (medium blue), 2.0 (dark blue)

Gambar 6 Interface Hasil Training dan Evaluasi



Gambar 7 Interface Distribusi Sentiment

Interface ini dirancang untuk memudahkan evaluasi layanan kampus melalui *Sentiment Analysis*. Dengan antarmuka yang sederhana dan terstruktur, pengguna dapat mengupload dataset berupa file CSV yang didalamnya merupakan ulasan dan komentar mahasiswa terkait dengan layanan yang dimiliki oleh kampus tersebut. Setelah data berhasil dimuat, sistem akan menampilkan pratinjau yang mencakup jumlah baris, kolom, ukuran file, serta isi teks dan label sentimennya.

Pada bagian konfigurasi analisis ini pengguna dapat memilih kolom teks yang akan dianalisis, kolom label untuk training, model analisi yang tersedia, serta mengatur proposi data uji. Setelah menjalankan analisis, sistem menampilkan distribusi sentimen yang terdeteksi, baik positif, negatif, maupun netral. Selanjutnya tersedia preview hasil berupa daftar komentar lengkap dengan prediksi sentimennya, serta opsi untuk mengunduh hasil analisis dalam format CSV. Kinerja model juga dievaluasi menggunakan metrik seperti

akurasi, precision, recall, dan F1-score, yang divisualisasikan melalui tabel evaluasi dan confusion matrix.

BAB III.

RENCANA CAPSTONE PROJECT

3.1 Tahapan Capstone Project

Kemukakan tahapan-tahapan yang akan dilalui untuk membangun solusi yang ditawarkan. Utamakan menjelaskan tahapan dengan bantuan gambar.

1. Studi Literatur & Identifikasi Masalah
 - 1) Mempelajari teori dasar **analisis sentimen**, *Natural Language Processing* (NLP), dan arsitektur model seperti *Transformer* (IndoBERT) serta model klasik (*Machine Learning*).
 - 2) Mengidentifikasi kebutuhan utama: sebuah *platform* yang memungkinkan pihak kampus menganalisis sentimen dari *dataset* ulasan secara mandiri dan interaktif.
 - 3) Merancang arsitektur aplikasi, termasuk alur kerja pengguna, model yang akan diimplementasikan (IndoBERT), dan fitur-fitur utama (unggah data, visualisasi, dan unduh hasil).
2. Pengumpulan dan Persiapan Data
 - 1) Mengumpulkan *dataset* ulasan mahasiswa dari berbagai sumber (media sosial, survei internal) sebagai data utama.
 - 2) Melakukan **pelabelan data secara manual** untuk sebagian *dataset* dengan tiga kategori: positif, negatif, dan netral. Data berlabel ini krusial untuk melatih model Logistic Regression dan Naive Bayes.
 - 3) Menyusun *dataset* akhir dalam format CSV, memastikan konsistensi kolom dan kebersihan data awal.
3. Implementasi Model & Pengembangan Aplikasi
 - 1) Implementasi *Backend*: Mengembangkan skrip Python (`sentiment_analysis.py`) yang berisi fungsi-fungsi inti:
 - 2) *Data Preprocessing* otomatis (*case folding*, *stopword removal*, dll.).

- 3) Logika *training* dan evaluasi untuk model Logistic Regression dan Naive Bayes menggunakan Scikit-learn.
 - 4) Logika *inference* (prediksi) menggunakan model *pre-trained* IndoBERT dari pustaka Hugging Face Transformers.
 - 5) Pengembangan *Frontend*: Membangun antarmuka pengguna (UI) yang interaktif menggunakan *framework* Streamlit (`streamlit_app.py`):
 - 6) Membuat fitur unggah *file* CSV.
 - 7) Membuat *widget* interaktif (*tombol*, *dropdown*, *slider*) untuk memilih kolom data, model, dan parameter.
 - 8) Mengintegrasikan fungsi *backend* untuk dieksekusi saat pengguna menekan tombol "Jalankan Analisis".
4. Evaluasi Model dan Visualisasi Interaktif
- 1) Mengimplementasikan fitur evaluasi model otomatis di dalam aplikasi, yang menampilkan metrik (*accuracy*, *precision*, *recall*, *F1-score*) dan *Confusion Matrix* secara *real-time* setelah proses *training* selesai.
 - 2) Membuat visualisasi data interaktif, seperti diagram batang yang menunjukkan distribusi sentimen dari hasil prediksi.
 - 3) Mengembangkan fitur unduh hasil analisis dalam format CSV, yang berisi data asli ditambah dengan kolom sentimen hasil prediksi
5. Penyusunan Laporan
- 1) Menyusun dokumentasi lengkap tentang tahapan analisis mulai dari studi literatur hingga evaluasi model.
 - 2) Menyajikan hasil analisis dalam bentuk visualisasi sederhana (tabel dan grafik).
 - 3) Menyusun laporan akhir capstone project sesuai format akademik yang berlaku.

3.2 Timeline Capstone Project

Berdasarkan sub-bab 3.1, kemukakan timeline Capstone Project dari awal hingga akhir beserta beserta penanggung jawab pada tahapan itu. Tahapan pada sub-

	Distribusi Sentimen	Nyoman Dimas											
	Pengujian Fungsional & <i>Usability Testing</i> Aplikasi secara Menyeluruh	Semuanya											
Penyusunan Laporan	Penyusunan Laporan Akhir Capstone Project (Bab 1-5)	Semuanya											

3.3 Alat dan Bahan

3.3.1 Perangkat Keras (Hardware)

Table 2 *Hardware*

No	Nama Alat	Spesifikasi	Fungsi
1.	Laptop/PC	<ul style="list-style-type: none"> • Prosesor: Intel Core i3/AMD Ryzen 3 atau lebih tinggi • RAM: Minimum 4GB (disarankan 8GB) • Storage: 5GB ruang kosong • Koneksi internet stabil 	Sebagai perangkat utama untuk pengembangan model, eksplorasi data, dan dokumentasi proyek dalam Jupyter Notebook

3.3.2 Perangkat Lunak (Software)

A. Platform Pengembangan

Table 3 Platform Pengembangan

No	Nama Software	Versi/Spesifikasi	Fungsi
1.	Python	Versi 3.7 - 3.11	Bahasa pemrograman utama untuk implementasi model analisis sentimen dan preprocessing data
2.	Streamlit	$\geq 1.10.0$	<i>Framework</i> utama untuk membangun dan menjalankan antarmuka aplikasi web yang interaktif.
3.	Google Colab	Platform cloud gratis	Platform untuk eksperimen awal, <i>prototyping</i> model, dan analisis data sebelum diimplementasikan ke dalam aplikasi.
4.	Terminal / Command Prompt	Standar Sistem Operasi	Lingkungan eksekusi untuk skrip <i>backend</i> (sentiment_analysis.py) melalui <i>Command-Line Interface</i> (CLI).

B. Library Python untuk Analisis Sentimen

Table 4 Library Python untuk Analisis Sentimen

No	Nama	Versi/Spesifikasi	Fungsi
1.	pandas	$\geq 1.3.0$	Manipulasi dan analisis data terstruktur, membaca file CSV/Excel
2.	numpy	$\geq 1.21.0$	Komputasi numerik dan operasi array untuk preprocessing data
3.	scikit-learn	$\geq 1.0.0$	Implementasi algoritma machine learning (SVM, Naive Bayes, Logistic Regression), evaluasi model
4.	Transformers & PyTorch	$\geq 4.0.0$	Memuat dan menjalankan model <i>deep learning</i> canggih (IndoBERT/Roberta) untuk analisis sentimen secara langsung (<i>inference</i>).
5.	argparse	Standar Python	Membuat antarmuka baris perintah (CLI) untuk skrip

			<i>backend</i> , memungkinkan eksekusi dengan argumen seperti --csv dan --model.
6.	Sastrawi	$\geq 1.0.1$	Melakukan Stemming (mengubah kata ke bentuk dasar) dan menyediakan daftar Stopwords khusus untuk Bahasa Indonesia.

C. Library untuk Visualisasi dan Analisis

Table 5 Library untuk Visualisasi dan Analisis

No	Nama	Versi/Spesifikasi	Fungsi
1.	matplotlib	$\geq 3.5.0$	Pustaka dasar untuk membuat visualisasi data statis seperti grafik batang dan <i>heatmap</i>
2.	Streamlit	$\geq 1.10.0$	Membangun seluruh komponen antarmuka pengguna (UI) dari aplikasi web, termasuk tombol, <i>dropdown</i> , dan

			area unggah <i>file</i>
3.	seaborn	$\geq 0.11.0$	Membangun visualisasi statistik yang lebih menarik di atas Matplotlib, digunakan untuk membuat <i>Confusion Matrix heatmap</i> .

D. Tools Pengumpulan dan Manajemen Data

Table 6 Tools Pengumpulan dan Manajemen Data

No	Nama	Versi/Spesifikasi	Fungsi
1.	Google Forms	Platform gratis	Pengumpulan data feedback mahasiswa melalui survei online
2.	Google Sheets	Platform gratis	Penyimpanan dan manajemen dataset sementara sebelum diproses
3.	CSV/Excel Files	Format .csv, .xslv	Format data untuk dataset training dan testing

3.3.3 Dataset dan Bahan Penelitian

Table 7 Dataset dan Bahan Penelitian

No	Jenis Data	Sumber	Fungsi
1.	Dataset simulasi	Data buatan tim sendiri	Data untuk testing awal model dan validasi pipeline
2.	Data Survei Mahasiswa	Google Forms	Data real feedback mahasiswa tentang pelayanan mahasiswa
3.	Dataset Publik Bahasa Indonesia	Kaggle, Github (gratis)	Dataset tambahan untuk training model analisis sentimen
4.	Stopwords Indonesia	NLTK, repository Github	Daftar kata yang akan dihilangkan dalam preprocessing

3.3.4 Platform Kolaborasi Tim

Table 8 Platform Kolaborasi Tim

No	Platform	Spesifikasi	Fungsi
1.	GitHub	Repository gratis	Version control dan kolaborasi kode tim
2.	Google Drive	Storage gratis 15GB	Sharing dataset, dokumentasi, dan backup

			notebook
3.	Whatsapp/Telegram	Aplikasi messaging	Komunikasi tim dan koordinasi harian

3.3.5 Justifikasi Pemilihan Alat dan Bahan

Kriteria Pemilihan :

1. Gratis dan Open Source: Semua tools yang dipilih tersedia gratis untuk meminimalkan biaya proyek
2. Mudah Dipelajari: Tools yang familiar dan memiliki dokumentasi yang baik
3. Community Support: Tools dengan komunitas besar sehingga mudah mendapat bantuan
4. Sesuai Kebutuhan: Memenuhi requirement untuk analisis sentimen sederhana berbahasa Indonesia

Alasan Khusus :

- a) Python Bahasa pemrograman ini dipilih karena merupakan standar industri untuk *data science* dan *machine learning*. Ekosistemnya yang kaya dengan *library* pendukung memungkinkan semua tahapan proyek, dari analisis data hingga pengembangan web, dapat dilakukan dalam satu bahasa yang sama.
- b) Pandas Digunakan sebagai *library* fundamental untuk manipulasi data. Kemampuannya dalam membaca *file* CSV, membersihkan, dan menstrukturkan data ke dalam format DataFrame sangat penting untuk persiapan data sebelum diolah oleh model.
- c) Scikit-learn Dipilih karena menyediakan implementasi model *machine learning* klasik (Logistic Regression & Naive Bayes) yang stabil, efisien, dan mudah digunakan. Selain itu, fitur TfidfVectorizer dan metrik evaluasi yang lengkap sangat esensial untuk alur kerja *training* model.
- d) Sastrawi *Library* ini dipilih secara khusus karena merupakan standar terbaik untuk pemrosesan teks (NLP) Bahasa Indonesia. Algoritma *stemming* dan daftar *stopwords*-nya yang dirancang untuk Bahasa Indonesia memberikan hasil normalisasi kata yang jauh lebih akurat dibandingkan *library* generik lainnya.

- e) Transformers (dari Hugging Face) & PyTorch Pustaka ini dipilih untuk mengakses model *deep learning* canggih dan *state-of-the-art* seperti IndoBERT/Roberta. Keunggulannya adalah menyediakan model pra-latih dengan pemahaman konteks bahasa yang mendalam, memungkinkan analisis sentimen yang sangat akurat tanpa perlu melatih model dari nol.
- f) Streamlit *Framework* ini dipilih untuk membangun antarmuka aplikasi web yang interaktif. Alasannya adalah kesederhanaannya yang luar biasa, di mana aplikasi web yang fungsional dapat dibuat hanya dengan menggunakan skrip Python, mempercepat proses pengembangan dari model ke produk yang dapat digunakan.
- g) Matplotlib & Seaborn Kombinasi *library* ini dipilih untuk membuat visualisasi data seperti *Confusion Matrix* dan diagram distribusi sentimen. Keduanya sangat fleksibel dan terintegrasi dengan baik di dalam aplikasi Streamlit untuk menyajikan hasil analisis secara visual.

3.1 Anggaran Capstone Project (Opsional)

Kemungkinan anggaran yang akan dikeluarkan selama pelaksanaan Capstone Project. Anggaran bersifat opsional

Table 9 Anggaran

PENGELUARAN					
No	Rincian	Jumlah	Satuan	Biaya Satuan	Total
1.	Internet (kuota)	1	Paket	Rp. 150.000	Rp. 150.000
2.	x	x	x	x	x
3.	x	x	x	x	x
4.	x	x	x	x	x
Total Pengeluaran					Rp.150.000

DAFTAR PUSTAKA

- Alzaid, H., Alturki, R., Meccawy, Z., Meccawy, M., & Alshehri, A. (2022). Sentiment analysis for evaluating student feedback using deep learning. *Education and Information Technologies*, 27(5), 6753–6775. <https://doi.org/10.1007/s10639-021-10867-w>
- Baihaqi, R. A., & Fitriani, F. (2023). Analisis sentimen terhadap ulasan mahasiswa menggunakan IndoBERT. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(1), 45–52. <https://doi.org/10.25126/jtiik.202310123>
- Chamorro-Atalaya, O., Aldana-Trejo, F., Alvarado-Bravo, N., Nieves-Barreto, C., Aguilar-Loyaga, S., Gamarra-Bustillos, C., Torres-Quiroz, A., Riveros-Cuellar, A., & Pérez-Samanamud, M., & Pérez-Guevara, L. (2025). Identification of the satisfaction of university students through sentiment analysis: A systematic review. *International Journal of Evaluation and Research in Education (IJERE)*, 14(1), 37–49. <https://doi.org/10.11591/ijere.v14i1.26640>
- Dake, A. A., & Gyimah, E. (2022). Service quality and student satisfaction in higher education. *International Journal of Educational Management*, 36(4), 567–583. <https://doi.org/10.1108/IJEM-10-2021-0378>
- Dalipi, F., Imran, A. S., & Kastrati, Z. (2018). Mooc dropout prediction using machine learning techniques: A review. *2018 IEEE Global Engineering Education Conference (EDUCON)*, 1109–1114. <https://doi.org/10.1109/EDUCON.2018.8363314>
- IJIET. (2024). Artificial intelligence in higher education: A systematic review. *International Journal of Instructional and Educational Technology*, 11(2), 55–72. <https://doi.org/10.3991/ijiet.v11i2.45329>

- Kastrati, Z., Dalipi, F., & Imran, A. S. (2020). Sentiment analysis of students' feedback with deep learning. *Applied Sciences*, 10(4), 1328. <https://doi.org/10.3390/app10041328>
- Napitupulu, D., Syafrullah, M., Rahim, R., Abdullah, D., & Setiawan, M. I. (2020). Analysis of student satisfaction toward quality of service facility. *Journal of Physics: Conference Series*, 1477(5), 052052. <https://doi.org/10.1088/1742-6596/1477/5/052052>
- Shaik, K. B., Ramesh, G., & Suresh, P. (2020). Sentiment analysis of student feedback data using machine learning techniques. *Procedia Computer Science*, 172, 599–606. <https://doi.org/10.1016/j.procs.2020.05.087>
- Telkom University. (2022). Student satisfaction survey report. *Telkom University Research Repository*. Retrieved from <https://repository.telkomuniversity.ac.id>
- Utami, C. W. (2020). Analisis kepuasan mahasiswa terhadap pelayanan akademik. *Jurnal Administrasi Pendidikan*, 27(1), 35–46. <https://doi.org/10.17509/jap.v27i1.2447>