

BAB III

ANALISIS DAN PERANCANGAN SISTEM

3.1 Analisis Sistem

Universitas Muhammadiyah Gresik (UMG) sebagai salah satu institusi pendidikan tinggi menerima berbagai ulasan dari pengguna Google Maps yang mencerminkan persepsi publik terhadap layanan, fasilitas kampus, kualitas pendidikan, serta interaksi dengan mahasiswa. Ulasan-ulasan ini mencakup pengalaman positif maupun kritik yang dapat digunakan sebagai acuan untuk evaluasi dan perbaikan. Saat ini, analisis terhadap ulasan tersebut masih dilakukan secara manual, dengan pengumpulan data yang tidak sistematis dan penilaian subjektif, sehingga mengakibatkan ketidakakuratan dalam mengetahui kecenderungan opini publik.

Proses manual yang dilakukan oleh pihak universitas mengandalkan pengumpulan data lisan atau observasi sederhana, tanpa menggunakan metode yang terstruktur. Hal ini menyebabkan sulitnya mengidentifikasi pola sentimen secara objektif dan tepat waktu. Akibatnya, evaluasi terhadap kualitas layanan, fasilitas, maupun citra universitas menjadi kurang optimal. Selain itu, jumlah ulasan yang terus bertambah setiap tahun membuat proses manual semakin tidak efektif dan rentan terhadap kesalahan manusia.

Namun, terdapat beberapa keunggulan dari data ulasan Google Maps yang dapat dimanfaatkan untuk menyelesaikan permasalahan ini. Data yang diperoleh melalui metode *web scraping* dapat memberikan informasi yang lebih terperinci, seperti teks ulasan, rating bintang, tanggal posting, dan identitas pengguna (jika tersedia). Data ini memungkinkan analisis tren sentimen dari waktu ke waktu, sehingga dapat melihat bagaimana persepsi publik berubah pada periode tertentu, misalnya bulanan atau tahunan. Dengan informasi ini, pihak universitas dapat memahami lebih dalam pola opini publik terkait aspek-aspek tertentu, seperti fasilitas kampus, kualitas pendidikan, dan interaksi dengan mahasiswa.

Dalam konteks penelitian ini, analisis data dilakukan dengan tujuan utama untuk mengetahui pola sentimen publik terhadap Universitas Muhammadiyah Gresik. Data yang dikumpulkan akan dianalisis untuk mengidentifikasi ulasan yang

bersifat positif, negatif, atau netral, sehingga dapat menjadi dasar pengambilan keputusan bagi pihak universitas. Analisis awal ini menjadi fondasi penting sebelum diterapkan algoritma *Multinomial Naïve Bayes* (MNB) untuk klasifikasi sentimen secara otomatis.

Selain itu, analisis data juga mempertimbangkan distribusi kata kunci yang sering muncul dalam ulasan. Kata-kata tersebut mencerminkan topik yang paling banyak dibahas oleh pengguna, seperti kenyamanan fasilitas, kualitas pengajaran, keramahan staf, dan pengalaman akademik secara keseluruhan. Informasi ini dapat menjadi indikator bagi universitas untuk menentukan prioritas perbaikan atau pengembangan layanan.

Dengan adanya analisis data yang sistematis, pihak UMG dapat mempersiapkan strategi evaluasi dan komunikasi yang lebih efektif. Selain itu, pola sentimen yang diperoleh dari data historis memungkinkan universitas untuk melakukan perencanaan layanan yang berbasis bukti, meningkatkan kepuasan pengguna, dan memperkuat citra institusi. Analisis ini juga menjadi dasar bagi tahap selanjutnya, yaitu hasil analisis sistem, yang akan membahas secara rinci tahapan pengolahan data, penerapan algoritma MNB, dan evaluasi model untuk menghasilkan prediksi sentimen yang akurat.

3.2 Hasil Analisis Sistem

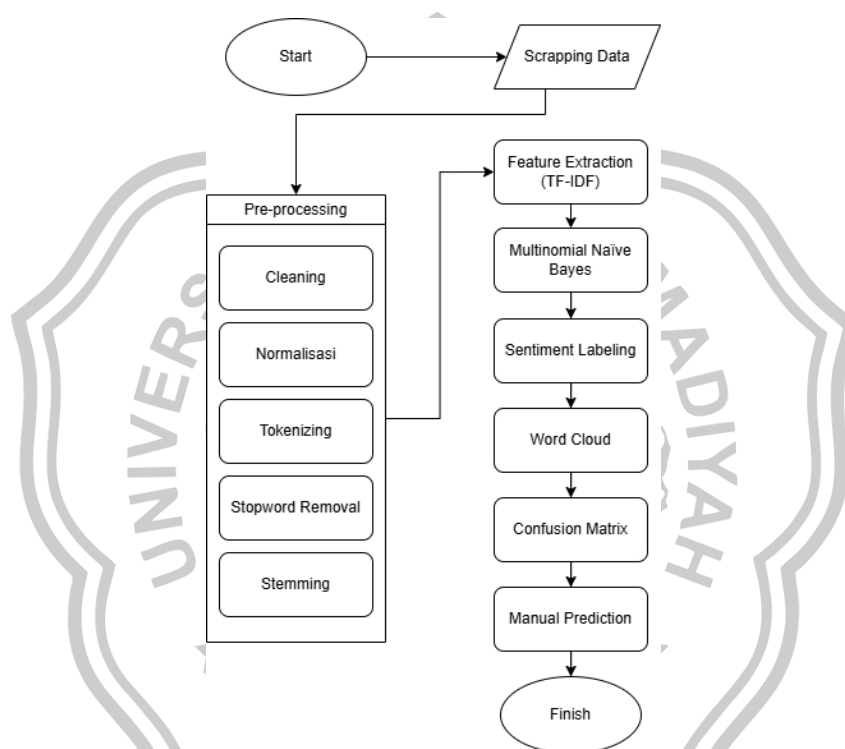
Pada tahap ini dilakukan proses analisis terhadap data ulasan yang diperoleh dari Google Maps Universitas Muhammadiyah Gresik. Data yang digunakan merupakan hasil *scraping* ulasan pengguna terhadap universitas tersebut. Dari proses *scraping* diperoleh sebanyak 487 data ulasan yang terdiri dari nama pengguna, rating, dan isi ulasan. Setelah melalui proses *preprocessing*, jumlah data yang dapat digunakan untuk analisis berkurang menjadi 282 data, dikarenakan sebagian besar ulasan hanya berisi rating bintang tanpa teks ulasan, sehingga tidak dapat dianalisis secara sentimen.

Data yang telah melalui tahap *preprocessing* kemudian digunakan untuk melakukan analisis sentimen menggunakan algoritma *Multinomial Naïve Bayes*. Proses analisis ini bertujuan untuk mengklasifikasikan ulasan menjadi tiga kategori

sentimen, yaitu positif, negatif, dan netral, berdasarkan isi teks ulasan yang telah dibersihkan.

Dari hasil pengujian, sistem menghasilkan tingkat akurasi sebesar 77,19%, yang menunjukkan bahwa model *Multinomial Naïve Bayes* mampu mengenali pola sentimen dengan tingkat ketepatan yang cukup baik terhadap data ulasan yang tersedia.

3.2.1 Flowchart Algoritma *Multinomial Naïve Bayes*



Gambar 3.1. Diagram Alur Penelitian

Diagram alir di atas menggambarkan tahapan proses analisis sentimen ulasan Google Maps Universitas Muhammadiyah Gresik menggunakan metode *Multinomial Naïve Bayes*. Proses dimulai dengan pengambilan data (*scrapping*) sebanyak 487 ulasan, yang kemudian diseleksi dan tersisa 282 data setelah melalui tahap *preprocessing*. Tahapan *preprocessing* meliputi *cleaning* untuk menghapus karakter yang tidak relevan, *normalisasi* untuk menyeragamkan kata tidak baku, *tokenizing* untuk memecah kalimat menjadi kata, *stopword removal* untuk menghapus kata umum yang tidak memiliki

makna penting, serta *stemming* untuk mengubah kata menjadi bentuk dasarnya. Data hasil pembersihan kemudian melalui tahap *feature extraction* menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*) agar teks dapat diubah menjadi representasi numerik yang dapat diolah oleh algoritma.

Tahap berikutnya adalah proses klasifikasi menggunakan algoritma *Multinomial Naïve Bayes* untuk mengelompokkan ulasan menjadi tiga kategori, yaitu positif, negatif, dan netral. Hasil analisis kemudian diberi label sentimen dan divisualisasikan menggunakan *Word Cloud* untuk menampilkan kata-kata yang paling sering muncul pada ulasan. Evaluasi dilakukan dengan *Confusion Matrix*, yang menunjukkan bahwa model menghasilkan akurasi sebesar 77,19%, menandakan tingkat ketepatan yang cukup baik dalam pengklasifikasian sentimen. Tahap terakhir adalah *manual prediction*, yang digunakan untuk membandingkan hasil prediksi model dengan penilaian manusia guna mengetahui tingkat konsistensi dan keandalan model terhadap data sebenarnya.

3.3 Representasi Data

Pada penelitian ini, data yang digunakan berupa ulasan pengguna Google Maps tentang Universitas Muhammadiyah Gresik (UMG) yang diperoleh melalui proses *web scraping*. Data hasil *scraping* terdiri dari 487 entri ulasan, yang mencakup tiga atribut utama, yaitu nama pengguna, rating bintang, dan teks ulasan. Namun, setelah dilakukan proses *preprocessing*, hanya 282 data yang dapat digunakan karena sebagian besar ulasan hanya berisi rating tanpa teks.

Data teks yang telah diperoleh kemudian direpresentasikan dalam bentuk fitur numerik agar dapat diproses oleh algoritma klasifikasi. Proses ini dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF), yang berfungsi untuk mengukur seberapa penting suatu kata dalam sebuah ulasan dibandingkan dengan keseluruhan kumpulan ulasan.

Selanjutnya, data ulasan diberi label berdasarkan kategori sentimen, yaitu:

1. **Positif**, jika ulasan menunjukkan penilaian baik atau apresiasi terhadap kampus.
2. **Negatif**, jika ulasan mengandung kritik atau keluhan.

3. **Netral**, jika ulasan bersifat deskriptif tanpa ekspresi emosi tertentu.

Data yang telah direpresentasikan dan dilabeli ini menjadi dataset utama untuk proses pelatihan model *Multinomial Naïve Bayes* (MNB). Hasil klasifikasi digunakan untuk mengetahui distribusi sentimen publik terhadap Universitas Muhammadiyah Gresik. Selain itu, visualisasi kata menggunakan *Word Cloud* juga dilakukan untuk menampilkan kata-kata yang paling sering muncul dalam ulasan, seperti “kampus”, “mahasiswa”, “baik”, dan “fasilitas”.

Dengan representasi data yang terstruktur dan terlabel dengan baik, model dapat dilatih secara optimal sehingga menghasilkan akurasi sebesar 77,19%. Nilai ini menunjukkan bahwa model mampu mengklasifikasikan sentimen dengan tingkat ketepatan yang cukup baik untuk data ulasan publik di platform digital.

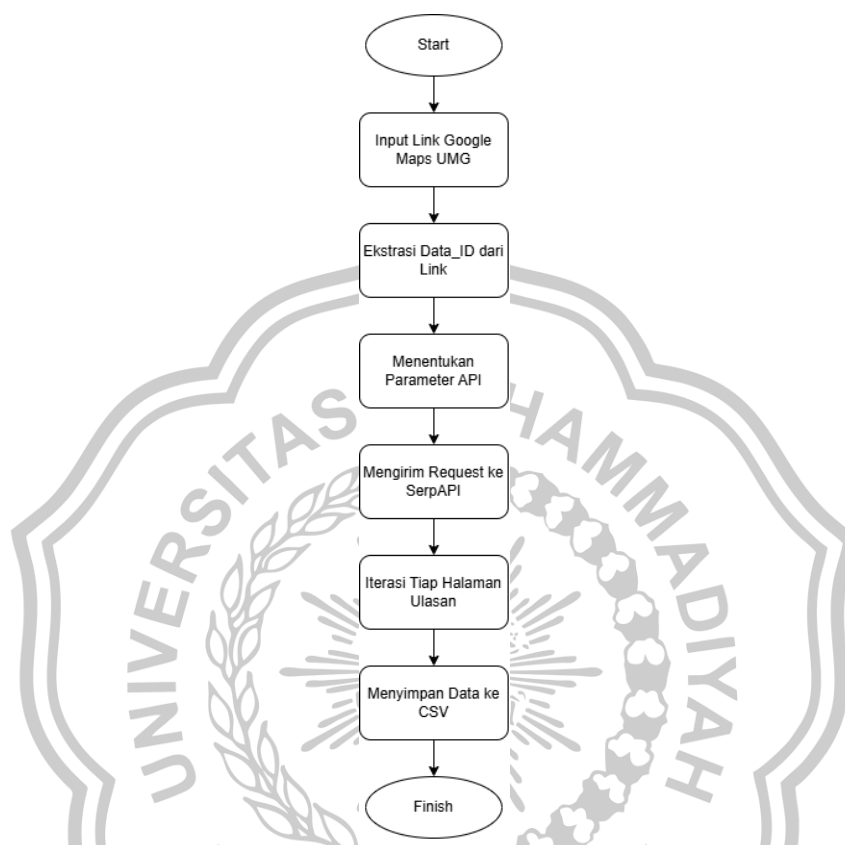
3.3.1 Tahapan Scraping Data

Pada penelitian ini, proses pengambilan data dilakukan menggunakan *Application Programming Interface* (API) melalui layanan SerpAPI. Metode ini digunakan untuk mengekstraksi data ulasan pengguna Google Maps terhadap Universitas Muhammadiyah Gresik (UMG) secara otomatis dengan format hasil yang terstruktur.

Proses *scraping* dimulai dengan memasukkan tautan Google Maps Universitas Muhammadiyah Gresik untuk mendapatkan *data_id*, yaitu identitas unik dari lokasi tersebut. Nilai *data_id* ini kemudian digunakan sebagai parameter penting dalam konfigurasi API. Selanjutnya, sistem mengirimkan permintaan (*request*) ke SerpAPI menggunakan parameter yang telah ditentukan seperti API key, *engine* (*google_maps_reviews*), bahasa halaman (*hl=id*), dan *data_id*.

Setelah itu, proses dilanjutkan dengan iterasi pada setiap halaman ulasan guna memastikan seluruh komentar pengguna berhasil diambil. Data hasil ekstraksi mencakup nama pengguna, rating bintang, tanggal ulasan, serta teks ulasan. Seluruh data yang diperoleh kemudian disimpan dalam format CSV agar dapat diolah lebih lanjut pada tahap *preprocessing*.

Hasil dari proses *scraping* ini menghasilkan 487 entri ulasan mentah, namun setelah proses pembersihan dan seleksi, hanya 282 data ulasan yang memiliki teks dan dapat digunakan sebagai dataset utama dalam analisis sentimen.



Gambar 3.2. Alur Scraping Data

Berikut adalah contoh sebagian hasil data ulasan yang diperoleh dari proses *scraping*:

Tabel 3.1. Ulasan Data Google Maps

NAME	RATING	ULASAN
SAGADES CV. MADANGKARA	1	ADA OKNUM SATPAM/ PENJAGA DI DEPAN GERBANG (Sudah Bapak2 45 tahunan) TIDAK KOMUNIKATIF, TERKESAN SEPERTI MALAS, DAN MOHON MAAF, SOK BGT! SAYA BERTAMU

		SEBAGAI TENAGA PENDIDIK, NAMUN BEDA UNIV, BARU KALI INI DIPERLAKUKAN TIDAK MENYENANGKAN. SAYA SUDAH SERING BERTAMU KE UNIV manapun. Semoga dapat perhatian khusus terlebih di ruang lingkup pendidikan. Tegas boleh tapi sopan santun dikedepankan. Jatuhnya bukan tegas tapi sok dan setahu saya SOP security, keamanan dan lain sebagainya tidak seperti itu.
Muhammad Iqbal Rahman	5	Kampusnya sangat megah dan mewah. Berada dilokasi strategis dan terdapat fasilitas yang mewah bagi mahasiswa. Sukses terus untuk UMG
Bagaskara Satria	4	Mohon, Info Lowongan Pekerjaan Lulusan Pendidikan D3/S1 diantaranya yaitu Pendidikan S1 *KEWIRAUSAHAAN* Bekerja sebagai Pegawai PNS/BUMN disuruh Orang Tua, menurut Perkataan Orang Tua bekerja dengan Pekerjaan tersebut dianggap berwibawa daripada jualan. #umggresik

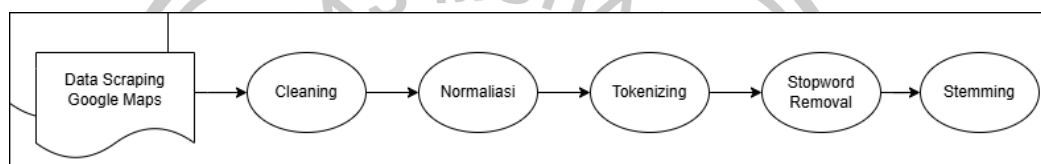
3.3.2 Tahapan Preprocessing Data

Setelah data ulasan berhasil diperoleh dari Google Maps melalui proses *scraping*, langkah berikutnya adalah melakukan *preprocessing* untuk menyiapkan data sebelum masuk ke tahap analisis dan klasifikasi. Pada tahap

ini, dilakukan serangkaian proses pembersihan dan transformasi data agar teks ulasan menjadi lebih terstruktur, relevan, dan siap digunakan oleh algoritma *Multinomial Naïve Bayes* .

Secara umum, *preprocessing* bertujuan untuk mengurangi *noise* pada data, menormalkan kata-kata, serta menghilangkan unsur yang tidak memiliki makna penting dalam analisis sentimen. Proses ini juga membantu meningkatkan akurasi model dengan memastikan setiap data memiliki bentuk representasi yang konsisten.

Alur tahapan *preprocessing* yang diterapkan dalam penelitian ini ditunjukkan pada Gambar dibawah ini, yang menggambarkan urutan proses pengolahan data mulai dari pembersihan hingga tahap stemming.



Gambar 3.3. Alur Processing Data

Adapun tahapan-tahapan *preprocessing* yang dilakukan adalah sebagai berikut:

1. *Cleanning* (Pembersihan Data)

Tahap ini bertujuan untuk menghapus elemen-elemen yang tidak relevan seperti angka, tanda baca, emotikon, URL, dan karakter khusus lainnya. Pembersihan dilakukan menggunakan *library Regular Expression (re)*, modul *string*, dan dukungan *pandas* untuk membersihkan kolom teks. Dengan demikian, teks yang dihasilkan hanya berisi kata-kata yang relevan dan bermakna.

2. Normalisasi

Proses normalisasi dilakukan untuk menyeragamkan bentuk kata tidak baku atau singkatan menjadi kata baku. Misalnya, kata “gk”, “ga”, dan “nggak” dinormalisasi menjadi “tidak”, sedangkan “bgt” atau “bngt” menjadi “banget”. Normalisasi menggunakan *kamus normalisasi* yang

dibuat secara manual berdasarkan variasi kata tidak baku pada ulasan, kemudian diterapkan melalui fungsi pemrosesan teks menggunakan *pandas*.

3. *Tokenizing*

Pada tahap ini, setiap kalimat dipecah menjadi potongan kata atau token. Sebagai contoh, kalimat “Pelayanannya bagus banget tempatnya bersih” akan diubah menjadi daftar kata [“pelayanannya”, “bagus”, “banget”, “tempatnya”, “bersih”]. Proses tokenisasi dilakukan menggunakan *Sastrawi Tokenizer* serta didukung oleh *library nltk* untuk memastikan pemisahan kata berjalan akurat.

4. *Stopword Removal* (Penyaringan Kata Umum)

Tahap ini bertujuan menghapus kata-kata umum yang sering muncul tetapi tidak memiliki makna penting, seperti “yang”, “dan”, “di”, atau “ke”. Penyaringan *stopword* menggunakan *Sastrawi StopWordRemoverFactory* serta custom *stopwords* tambahan apabila diperlukan, sehingga analisis dapat berfokus pada kata-kata yang membawa makna utama.

5. *Stemming*

Stemming dilakukan untuk mengembalikan setiap kata ke bentuk dasarnya. Contohnya, kata “berjalan”, “menjalankan”, dan “dijalankan” semuanya akan dikembalikan menjadi “jalan”. Pada penelitian ini, proses *stemming* menggunakan *Sastrawi StemmerFactory* yang secara khusus mendukung pemrosesan Bahasa Indonesia.

6. *Labeling* (Pelabelan Data Sentimen)

Setelah seluruh tahapan preprocessing selesai, data ulasan selanjutnya diberi label sentimen. Proses pelabelan pada sistem ini dilakukan secara otomatis (*auto labeling*) berdasarkan nilai atau informasi sentimen yang tersedia pada data ulasan Google Maps. Setiap ulasan diklasifikasikan ke dalam tiga kelas sentimen, yaitu positif, netral, dan negatif.

Pelabelan otomatis ini bertujuan untuk mempercepat proses pengolahan data serta mengurangi subjektivitas yang dapat muncul apabila pelabelan dilakukan secara manual. Hasil label yang diperoleh kemudian

digunakan sebagai data target (label kelas) dalam proses pelatihan dan pengujian model *Multinomial Naive Bayes*.

Alur tahapan *preprocessing* tersebut dapat dilihat pada tabel dibawah ini, yang menunjukkan urutan proses dari data mentah hingga data siap untuk tahap ekstraksi fitur menggunakan metode TF-IDF.

Tabel 3.2. Hasil *Preprocessing* Data

TAHAPAN	ULASAN
Data Asli	Kampusnya sangat megah dan mewah. Berada dilokasi strategis dan terdapat fasilitas yang mewah bagi mahasiswa. Sukses terus untuk UMG
<i>Cleaning</i>	kampusnya sangat megah dan mewah berada dilokasi strategis dan terdapat fasilitas yang mewah bagi mahasiswa sukses terus untuk umg
Normalisasi	kampusnya sangat megah dan mewah berada dilokasi strategis dan terdapat fasilitas yang mewah bagi mahasiswa sukses terus untuk umg
<i>Tokenizing</i>	[kampusnya, sangat, megah, dan, mewah, berada, dilokasi, strategis, dan, terdapat, fasilitas, yang, mewah, bagi, mahasiswa, sukses, terus, untuk, umg]
<i>Stopword Removal</i>	[kampusnya, sangat, megah, mewah, berada, dilokasi, strategis, terdapat, fasilitas, mewah, mahasiswa, sukses, terus, umg]
<i>Stemming</i>	[kampus, sangat, megah, mewah, berada, lokasi, strategis, dapat, fasilitas, mewah, mahasiswa, sukses, terus, umg]

Dari tabel di atas dapat disimpulkan bahwa teks ulasan yang semula panjang dan tidak terstruktur berhasil diubah menjadi kumpulan kata dasar yang bersih dan bermakna. Hasil inilah yang kemudian digunakan dalam proses ekstraksi fitur menggunakan metode TF-IDF pada tahap berikutnya.

3.3.3 Tahapan Feature Extraction TF-IDF

Setelah melewati seluruh proses *preprocessing*, tahap berikutnya adalah mengubah setiap kata pada dokumen ulasan menjadi representasi numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF–

IDF). Metode ini berfungsi memberikan bobot pada setiap kata berdasarkan tingkat kepentingannya dalam dokumen dan pada seluruh korpus.

Adapun komponen utama dalam TF-IDF adalah sebagai berikut:

1. Term Frequency (TF)

Menyatakan jumlah kemunculan kata t pada dokumen d .

2. Document Frequency (DF)

Mengambarkan jumlah dokumen yang mengandung kata t .

3. Inverse Document Frequency (IDF)

Mengukur tingkat kelangkaan kata dalam seluruh korpus. Semakin jarang muncul, semakin besar bobotnya.

Rumus perhitungan TF-IDF dinyatakan sebagai berikut:

$$TF - IDF(t, d) = TF(t, d) \times \log \left(\frac{N}{DF(t)} \right) \quad (3.1)$$

Keterangan

1. $TF(t, d)$ = frekuensi kata t pada dokumen d
2. $DF(t)$ = jumlah dokumen yang mengandung kata t
3. N = total dokumen pada korpus (pada penelitian ini menggunakan 10 dokumen latih)

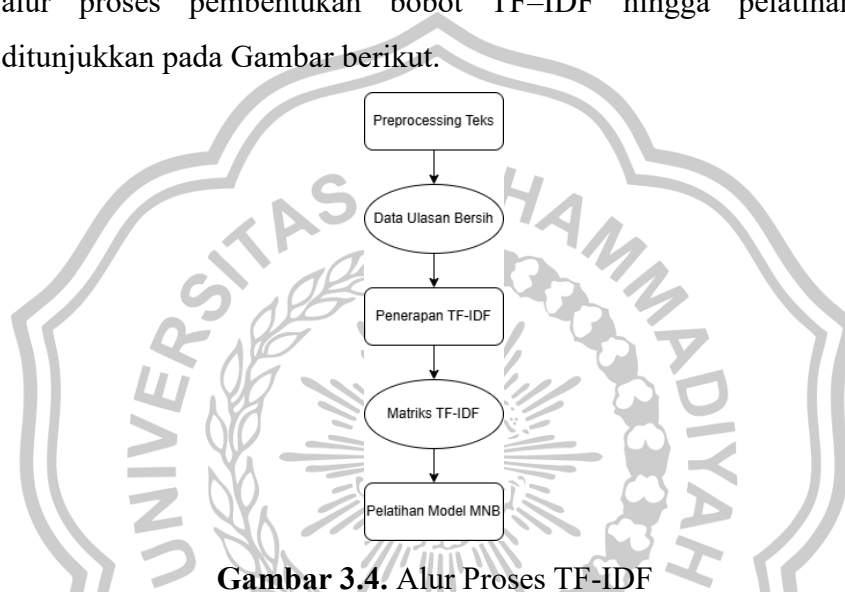
Setelah proses transformasi, setiap dokumen diubah menjadi vektor numerik. Berikut merupakan contoh hasil matriks TF-IDF yang dihasilkan pada 10 data ulasan yang digunakan sebagai dataset awal.

Tabel 3.3. Contoh Hasil Bobot TF-IDF pada 10 Dokumen Ulasan

Kata	TF	DF	IDF	TF-IDF
kampusnya	1	1	2.303	2.303
sangat	1	2	1.609	1.609
megah	1	1	2.303	2.303
dan	2	3	1.204	2.408
mewah	2	1	2.303	4.605
berada	1	1	2.303	2.303
dilokasi	1	1	2.303	2.303
strategis	1	2	1.609	1.609
terdapat	1	1	2.303	2.303
fasilitas	1	2	1.609	1.609
yang	1	3	1.204	1.204

bagi	1	1	2.303	2.303
mahasiswa	1	1	2.303	2.303
sukses	1	1	2.303	2.303
terus	1	1	2.303	2.303
untuk	1	3	1.204	1.204
umg	1	2	1.609	1.609

Matriks TF-IDF yang dihasilkan kemudian digunakan sebagai *feature representation* yang menjadi input pada tahap klasifikasi menggunakan algoritma *Multinomial Naïve Bayes* (MNB). Secara visual, alur proses pembentukan bobot TF-IDF hingga pelatihan model ditunjukkan pada Gambar berikut.



Gambar 3.4. Alur Proses TF-IDF

3.3.4 Tahapan Klasifikasi *Multinomial Naïve Bayes* (MNB)

Pada tahap ini dilakukan proses klasifikasi untuk menentukan polaritas sentimen dari setiap ulasan. Algoritma yang digunakan adalah *Multinomial Naïve Bayes* (MNB), yaitu salah satu varian dari algoritma *Naïve Bayes* yang umum digunakan untuk klasifikasi teks. Algoritma ini bekerja dengan menghitung peluang kemunculan setiap kata (fitur) terhadap kelas tertentu (positif, netral, atau negatif) menggunakan prinsip Teorema *Bayes*.

Algoritma MNB mengasumsikan bahwa setiap fitur bersifat saling bebas (*conditional independence assumption*), dan sangat cocok digunakan untuk data teks yang direpresentasikan melalui nilai frekuensi atau bobot kata seperti hasil *Term Frequency–Inverse Document Frequency* (TF-IDF).

Secara matematis, Teorema *Bayes* dituliskan sebagai berikut:

$$P(C|X) = \frac{P(C) \prod_{i=1}^n P(x_i|C)}{P(X)} \quad (3.2)$$

Dengan keterangan:

1. $P(C | X)$: probabilitas dokumen X termasuk kelas C
2. $P(C)$: probabilitas awal kelas C
3. $P(x_i | C)$: probabilitas kemunculan kata ke- i pada kelas C
4. $P(X)$: probabilitas total dokumen (konstanta)

Untuk mendapatkan hasil klasifikasi, algoritma MNB terlebih dahulu menghitung probabilitas kemunculan setiap kata pada masing-masing kelas. Oleh karena itu, diperlukan perhitungan manual likelihood sebagai contoh proses bagaimana model menentukan kelas sentimen suatu ulasan.

1. Perhitungan Manual Probabilitas Kata (*Likelihood*)

Perhitungan dilakukan menggunakan 10 dokumen latih. Probabilitas setiap kata dihitung berdasarkan frekuensinya pada masing-masing kelas menggunakan rumus Laplace smoothing:

$$P(x_i|C) = \frac{\text{Count}(x_i \text{ pada kelas } C) + 1}{\text{Total kata pada kelas } C + |V|} \quad (3.3)$$

Di mana $|V|$ adalah ukuran vocabulary.

Tabel 3.4. Perhitungan Probabilitas Kata pada *Multinomial Naïve Bayes*

Kata	TF	C_Pos	C_Neg	C_Net	P_Pos	P_Neg	P_Net	lnP_Pos	lnP_Neg	lnP_Net
kampus	1	1	0	0	0.0065	0.0024	0.0034	-5.03	-6.03	-5.70
sangat	1	2	1	0	0.0098	0.0048	0.0034	-4.63	-5.34	-5.70
megah	1	1	0	0	0.0065	0.0024	0.0034	-5.03	-6.03	-5.70
Mewah	2	1	0	0	0.0065	0.0024	0.0034	-5.03	-6.03	-5.70
berada	1	0	0	0	0.0033	0.0024	0.0034	-5.73	-6.03	-5.70
lokasi	1	0	0	0	0.0033	0.0024	0.0034	-5.73	-6.03	-5.70
strategis	1	1	0	0	0.0065	0.0024	0.0034	-5.03	-6.03	-5.70
dapat	1	0	0	0	0.0033	0.0024	0.0034	-5.73	-6.03	-5.70
fasilitas	1	1	0	0	0.0065	0.0024	0.0034	-5.03	-6.03	-5.70
mahasiswa	1	1	0	0	0.0065	0.0024	0.0034	-5.03	-6.03	-5.70
sukses	1	0	0	0	0.0033	0.0024	0.0034	-5.73	-6.03	-5.70
terus	1	0	0	0	0.0033	0.0024	0.0034	-5.73	-6.03	-5.70
umg	1	1	1	1	0.0065	0.0048	0.0067	-5.03	-5.34	-5.00

2. Perhitungan Skor Kelas (Log-Posterior)

Skor kelas dihitung menggunakan:

$$Score(C) = \ln P(C) + \sum_{i=1}^n \ln P(x_i|C) \quad (3.2)$$

Tabel 3.5. Perhitungan Skor Sentimen

Kata	lnP_ Pos	lnP_ Neg	lnP_ Net	Contrib_ Pos	Contrib_ Neg	Contrib_ Net
kampus	-5.03	-6.03	-5.70	-5.034	-6.031	-5.697
sangat	-4.63	-5.34	-5.70	-4.628	-5.338	-5.697
megah	-5.03	-6.03	-5.70	-5.034	-6.031	-5.697
mewah	-5.03	-6.03	-5.70	-10.067	-12.061	-11.394
berada	-5.73	-6.03	-5.70	-5.727	-6.031	-5.697
lokasi	-5.73	-6.03	-5.70	-5.727	-6.031	-5.697
strategis	-5.03	-6.03	-5.70	-5.034	-6.031	-5.697
dapat	-5.73	-6.03	-5.70	-5.727	-6.031	-5.697
fasilitas	-5.03	-6.03	-5.70	-5.034	-6.031	-5.697
mahasiswa	-5.03	-6.03	-5.70	-5.034	-6.031	-5.697
sukses	-5.73	-6.03	-5.70	-5.727	-6.031	-5.697
terus	-5.73	-6.03	-5.70	-5.727	-6.031	-5.697
umg	-5.03	-5.34	-5.00	-5.034	-5.338	-5.004

Tabel 3.6. Total Skor Akhir

Kelas	Score
Positif	-97.3788
Negatif	-109.2609
Netral	-107.0816

Kelas dengan skor terbesar (paling mendekati 0) = Positif, ulasan diklasifikasikan sebagai sentimen positif.

3. Implementasi Model Menggunakan *Pipeline*

Tabel 3.7. Hasil Implementasi *Pipeline Multinomial Naïve Bayes*

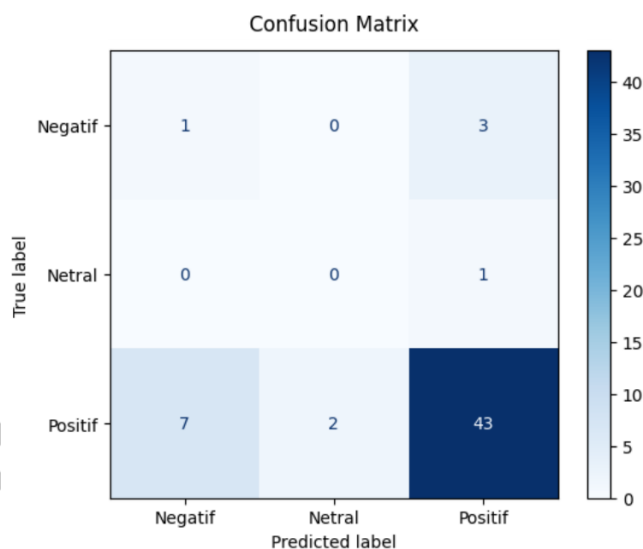
Komponen	Konfigurasi
TfidfVectorizer	max_df=0.95, max_features=8000, min_df=2, ngram_range=(1,2), sublinear_tf=True
MultinomialNB	alpha=0.5

Pipeline ini memastikan proses ekstraksi fitur dan klasifikasi berlangsung otomatis serta terstruktur.

4. Hasil *Sentiment Labeling*

Tabel 3.8. Hasil *Labeling*

Konsep confusion matrix memberikan gambaran lebih detail mengenai performa prediksi setiap kelas. Berikut adalah nilai confusion matrix berdasarkan hasil prediksi model:



Gambar 3.6. Confusion Matrix Model *Multinomial Naïve Bayes*

7. Prediksi Manual

Tabel 3.10. Hasil Evaluasi Model

Kalimat	Prediksi
kampus sangat megah mewah	Positif
fasilitas lumayan lengkap	Netral
atm nya rusak	Negatif

Model menghasilkan prediksi sesuai makna kalimat.

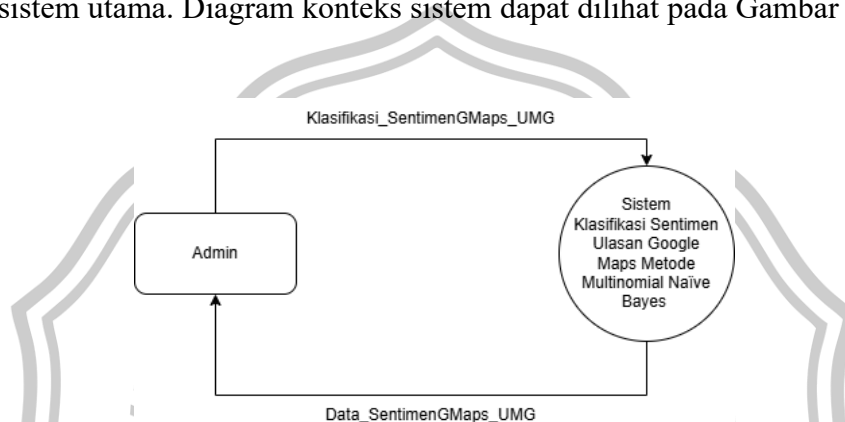
Secara keseluruhan, tahapan perhitungan manual pada algoritma *Multinomial Naïve Bayes* menunjukkan bagaimana setiap kata pada ulasan diberi bobot melalui TF-IDF, kemudian dihitung probabilitas kemunculannya dalam masing-masing kelas sentimen berdasarkan data latih. Nilai-nilai probabilitas tersebut digunakan untuk menentukan tingkat kemungkinan sebuah ulasan termasuk ke kelas positif, negatif, atau netral. Melalui proses ini, setiap kata turut berkontribusi dalam pembentukan model dengan memperkuat kecenderungan kelas berdasarkan pola kemunculannya. Tahapan perhitungan manual ini sekaligus memastikan bahwa model klasifikasi dibangun secara transparan, terukur, dan sesuai prinsip kerja dasar *Multinomial Naïve Bayes* yang mengasumsikan

independensi antar fitur. Prosedur ini juga memberikan gambaran nyata bahwa hasil klasifikasi sentimen diperoleh melalui proses matematis yang sistematis dan dapat dipertanggungjawabkan.

3.4 Perancangan Sistem

3.4.1 Diagram Konteks Sistem

Dalam diagram konteks ini terlihat entitas atau kesatuan luar yang terlibat dalam sistem, yang menggambarkan aliran data antara entitas eksternal dan sistem utama. Diagram konteks sistem dapat dilihat pada Gambar di bawah ini.



Gambar 3.7. Diagram Konteks Sistem

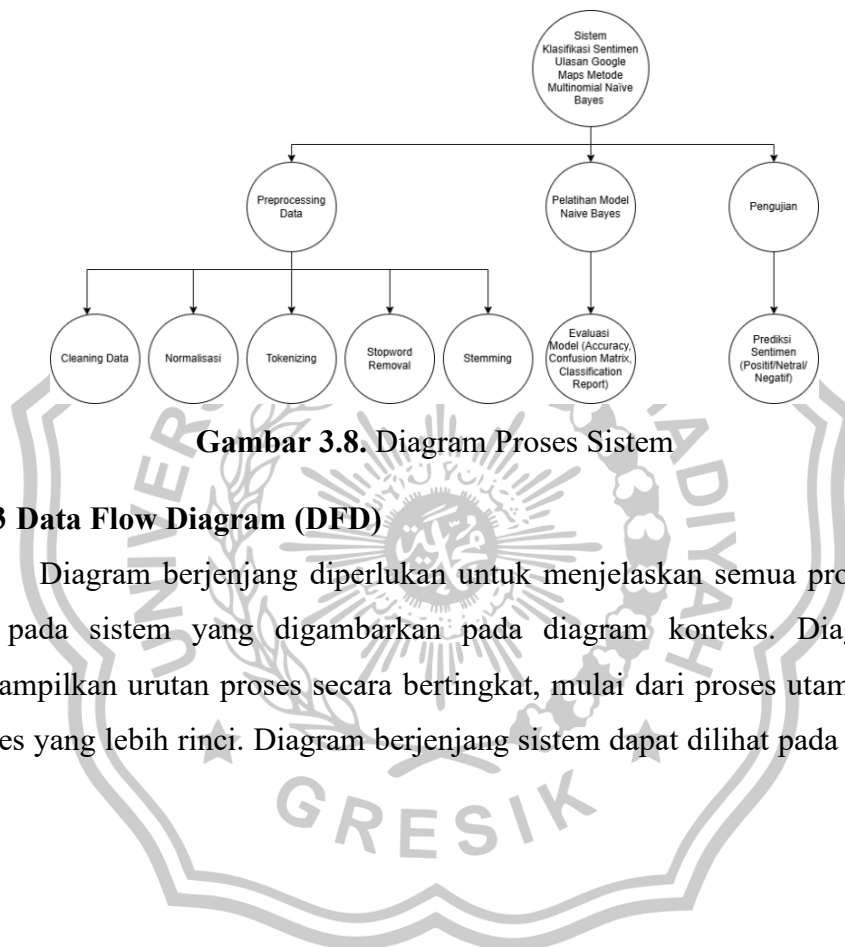
Diagram konteks pada Gambar tersebut menggambarkan proses yang terjadi pada Sistem Klasifikasi Sentimen Ulasan Google Maps Universitas Muhammadiyah Gresik menggunakan metode *Multinomial Naive Bayes*.

Dalam sistem ini, admin berperan sebagai entitas utama yang melakukan input data berupa data ulasan Google Maps UMG ke dalam sistem. Data tersebut kemudian diproses melalui tahapan *preprocessing*, ekstraksi fitur (TF-IDF), dan klasifikasi sentimen menggunakan metode *Multinomial Naive Bayes* untuk menghasilkan keluaran berupa hasil klasifikasi sentimen (positif, negatif, dan netral).

Hasil analisis sentimen yang diperoleh kemudian dikirim kembali kepada admin untuk dilakukan evaluasi, pelaporan, atau analisis lanjutan terhadap persepsi publik terhadap Universitas Muhammadiyah Gresik. Dengan demikian, sistem ini menghubungkan input data dari admin dengan proses klasifikasi yang memberikan hasil analisis secara objektif dan terstruktur.

3.4.2 Hirarki Proses Sistem

Diagram berjenjang diperlukan untuk menjelaskan semua proses yang ada pada sistem yang digambarkan pada diagram konteks. Diagram ini menampilkan urutan proses secara bertingkat, mulai dari proses utama hingga proses yang lebih rinci. Diagram berjenjang sistem dapat dilihat pada di bawah ini.

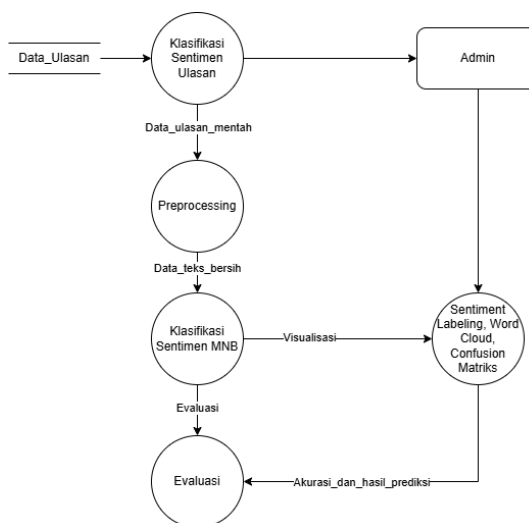


Gambar 3.8. Diagram Proses Sistem

3.4.3 Data Flow Diagram (DFD)

Diagram berjenjang diperlukan untuk menjelaskan semua proses yang ada pada sistem yang digambarkan pada diagram konteks. Diagram ini menampilkan urutan proses secara bertingkat, mulai dari proses utama hingga proses yang lebih rinci. Diagram berjenjang sistem dapat dilihat pada di bawah ini.

3.4.3.1 DFD Level 1



Gambar 3.9. Diagram DFD

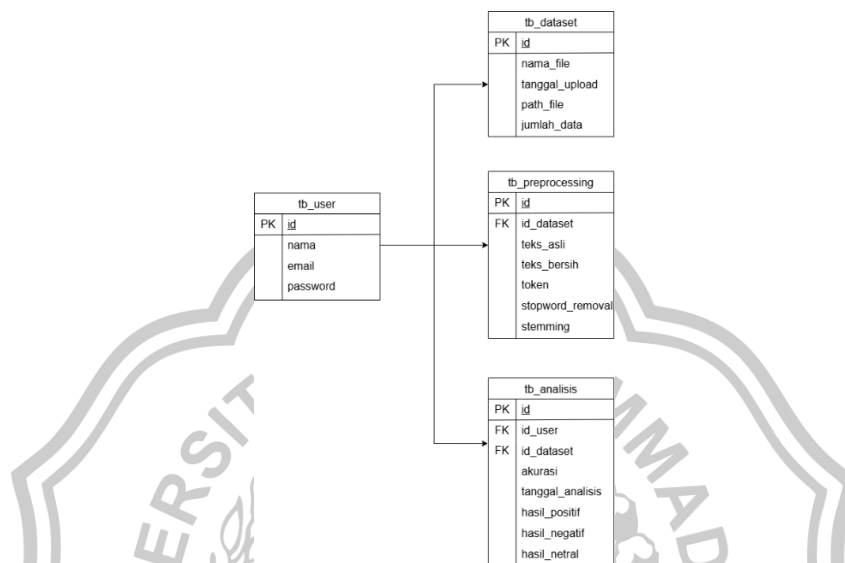
Diagram ini menggambarkan alur proses klasifikasi sentimen ulasan yang dimulai dari input berupa data ulasan yang diberikan oleh admin. Data ulasan mentah terlebih dahulu melalui tahap *preprocessing* untuk dibersihkan dari karakter yang tidak diperlukan, seperti tanda baca, huruf besar, dan kata yang tidak relevan, sehingga menghasilkan data teks bersih. Data tersebut kemudian diproses pada tahap klasifikasi sentimen menggunakan metode *Multinomial Naïve Bayes* (MNB) dengan bantuan ekstraksi fitur TF-IDF untuk mengelompokkan ulasan ke dalam kategori positif, negatif, atau netral. Hasil klasifikasi ini selanjutnya divisualisasikan dalam bentuk *sentiment labeling*, *word cloud*, dan *confusion matrix* guna menampilkan persebaran sentimen serta performa model. Tahap akhir berupa evaluasi dilakukan untuk mengukur akurasi dan hasil prediksi dari sistem, yang kemudian disajikan kembali kepada admin sebagai laporan analisis sentimen secara menyeluruh dan terstruktur.

3.5 Perancangan Basis Data

Perancangan basis data merupakan proses untuk menyusun struktur penyimpanan data agar dapat dikelola dan diakses secara efisien. Tujuan utama perancangan ini adalah untuk memastikan integritas data, menghindari redundansi,

dan mendukung kebutuhan sistem dalam analisis sentimen terhadap ulasan Google Maps menggunakan metode *Naïve Bayes Multinomial*.

Dalam sistem ini, metode yang digunakan untuk menggambarkan hubungan antar entitas adalah *Entity Relationship Diagram (ERD)*, yang ditunjukkan pada gambar berikut.



Gambar 3.10. Diagram ERD

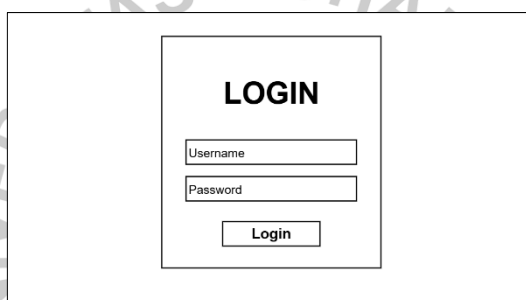
Pada gambar tersebut menunjukkan sistem basis data analisis sentimen ulasan Google Maps Universitas Muhammadiyah Gresik yang terdiri dari empat tabel utama dengan relasi yang terstruktur, di mana "tb_user" menyimpan data pengguna berupa *id*, *nama*, *email*, dan *password*, yang dapat melakukan banyak aktivitas analisis yang dicatat dalam "tb_dataset" dengan atribut *id*, *nama_file*, *tanggal_upload*, *jumlah_data*, dan *id_user* sebagai *foreign key* yang menghubungkannya ke "tb_user". Setiap dataset ulasan yang diunggah akan melalui tahap *preprocessing* yang hasilnya disimpan dalam "tb_preprocessing" dengan atribut *id*, *id_dataset*, *teks_asli*, *teks_bersih*, *token*, *stopword_removal*, dan *stemming*. Selanjutnya, dataset yang telah diproses dianalisis menggunakan metode *Naïve Bayes Multinomial*, dan hasil perhitungannya disimpan dalam "tb_analisis" dengan atribut *id*, *id_user*, *id_dataset*, *tanggal_analisis*, *akurasi*, *hasil_positif*, *hasil_negatif*, dan *hasil_netral*, di mana *id_user* dan *id_dataset* berperan sebagai *foreign key* yang menghubungkan hasil analisis dengan pengguna serta dataset terkait. Dengan struktur relasional antar tabel yang terintegrasi ini, sistem mampu

mendukung proses analisis sentimen secara otomatis mulai dari pengunggahan data, pembersihan teks, hingga menghasilkan klasifikasi sentimen dengan tingkat akurasi yang terukur dan efisien.

3.6 Perancangan Antar Muka Sistem

3.6.1 Halaman Login

Halaman login merupakan tampilan awal bagi admin untuk mengakses sistem analisis sentimen ulasan Google Maps Universitas Muhammadiyah Gresik. Tampilan dibuat sederhana dan terpusat dengan form berisi kolom *username*, *password*, serta tombol “Login”. Setelah data login diverifikasi, admin diarahkan menuju halaman beranda untuk mengelola proses *preprocessing*, pelatihan model *Naïve Bayes*, dan prediksi sentimen.



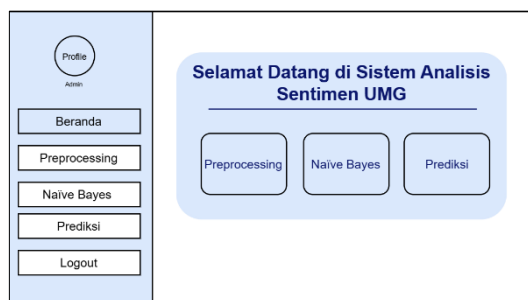
The image shows a login form titled "LOGIN". It contains three input fields: "Username", "Password", and a "Login" button. The form is centered on a white background with a faint watermark of the Universitas Muhammadiyah Gresik logo in the background.

Gambar 3.11. Halaman Login

3.6.2 Halaman Beranda

Halaman beranda merupakan tampilan utama setelah admin berhasil login ke dalam sistem analisis sentimen ulasan Google Maps Universitas Muhammadiyah Gresik. Pada bagian kiri terdapat *sidebar* yang berisi foto profil berbentuk lingkaran serta menu navigasi berupa tombol Beranda, *Preprocessing*, *Naïve Bayes*, *Prediksi*, dan *Logout*.

Di bagian tengah halaman ditampilkan ucapan “Selamat Datang di Sistem Analisis Sentimen UMG” sebagai pengantar utama. Selain itu, tersedia pula tombol navigasi cepat menuju halaman Beranda, *Preprocessing*, *Naïve Bayes*, dan *Prediksi*, untuk memudahkan admin dalam berpindah antar fitur sistem.

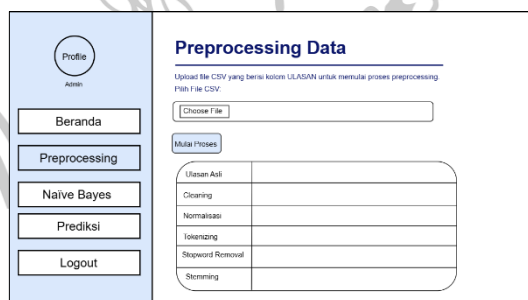


Gambar 3.12. Halaman Beranda

3.6.3 Halaman Preprocessing

Halaman *preprocessing* berfungsi untuk memproses data ulasan hasil *scraping* sebelum dilakukan analisis sentimen. Pada halaman ini, admin dapat mengunggah file berformat CSV yang berisi data ulasan Google Maps Universitas Muhammadiyah Gresik. Setelah file diunggah, admin dapat menekan tombol “Mulai Proses” untuk menjalankan tahapan pengolahan data secara otomatis.

Hasil proses akan ditampilkan dalam bentuk tabel yang berisi urutan tahapan pengolahan teks, dimulai dari ulasan asli, kemudian melalui proses *cleaning*, normalisasi, *tokenizing*, *stopword removal*, hingga *stemming*. Tahapan ini memastikan teks ulasan telah bersih dan siap digunakan untuk tahap analisis selanjutnya pada metode *Naïve Bayes*



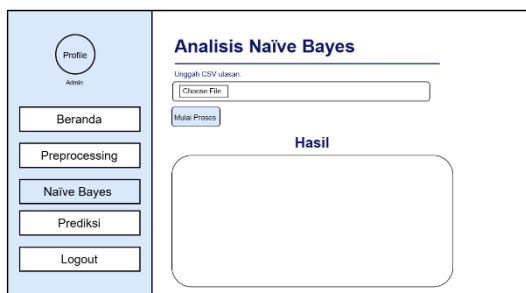
Gambar 3.13. Halaman Preprocessing

3.6.4 Halaman Naïve Bayes

Halaman *Naïve Bayes* berfungsi untuk melakukan proses klasifikasi sentimen menggunakan metode *Multinomial Naïve Bayes* (MNB). Pada halaman ini, admin perlu mengunggah file berformat CSV yang berisi data ulasan bersih hasil dari tahap *preprocessing*. Setelah file berhasil diunggah,

sistem akan menjalankan proses klasifikasi untuk menentukan kategori sentimen, yaitu positif, negatif, atau netral.

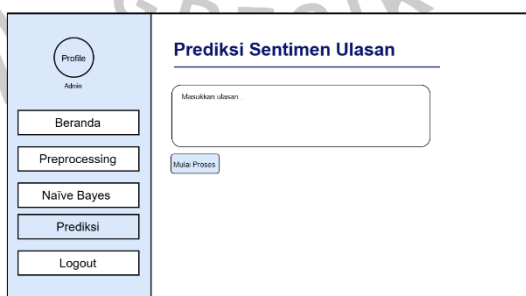
Hasil klasifikasi ditampilkan dalam bentuk tabel dan grafik evaluasi yang mencakup akurasi, presisi, *recall*, dan *f1-score*, sehingga admin dapat menilai sejauh mana performa model dalam mengklasifikasikan sentimen ulasan.



Gambar 3.14. Halaman *Naïve Bayes*

3.6.5 Halaman Prediksi

Halaman Prediksi berfungsi untuk memprediksi sentimen dari teks ulasan baru menggunakan model yang telah dilatih dengan metode *Multinomial Naïve Bayes*. Pada halaman ini, admin dapat memasukkan teks ulasan secara langsung ke dalam kolom input yang telah disediakan, kemudian menekan tombol “Mulai Proses” untuk menjalankan prediksi. Setelah proses selesai, sistem akan menampilkan hasil klasifikasi sentimen berupa kategori positif, negatif, atau netral, sesuai dengan analisis model terhadap isi teks yang dimasukkan.



Gambar 3.15. Halaman Prediksi

3.7 Perencanaan Pengujian Sistem

Berikut adalah langkah-langkah yang direncanakan dalam pengujian sistem untuk analisis sentimen ulasan Google Maps:

1. Pengujian dilakukan menggunakan dataset ulasan Google Maps Universitas Muhammadiyah Gresik yang telah dikumpulkan melalui proses *scraping* dan melewati tahapan *preprocessing* (*cleaning*, normalisasi, *tokenizing*, *stopword removal*, dan *stemming*).
2. Data yang telah bersih kemudian dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk membangun model klasifikasi sentimen dengan metode *Multinomial Naïve Bayes*, sedangkan data uji digunakan untuk mengukur performa model.
3. Setelah model dilatih, dilakukan evaluasi terhadap hasil klasifikasi sentimen menggunakan metrik pengujian seperti akurasi, precision, recall, dan f1-score, serta divisualisasikan melalui *confusion matrix* untuk melihat tingkat ketepatan prediksi terhadap masing-masing kategori sentimen (positif, negatif, dan netral).
4. Pengujian juga mencakup prediksi manual, di mana pengguna dapat memasukkan teks ulasan baru untuk melihat hasil klasifikasi sentimen secara langsung. Hal ini digunakan untuk memastikan bahwa sistem dapat berjalan dengan baik dalam kondisi nyata (*real-time input*).
5. Hasil pengujian kemudian dianalisis untuk menentukan tingkat keakuratan model dan keefektifan sistem dalam mengklasifikasikan sentimen ulasan pengguna terhadap Universitas Muhammadiyah Gresik. Nilai akurasi yang lebih tinggi menunjukkan bahwa model bekerja dengan baik dalam mengenali pola sentimen pada data teks.