

BAB III

ANALISIS DAN PERANCANGAN SISTEM

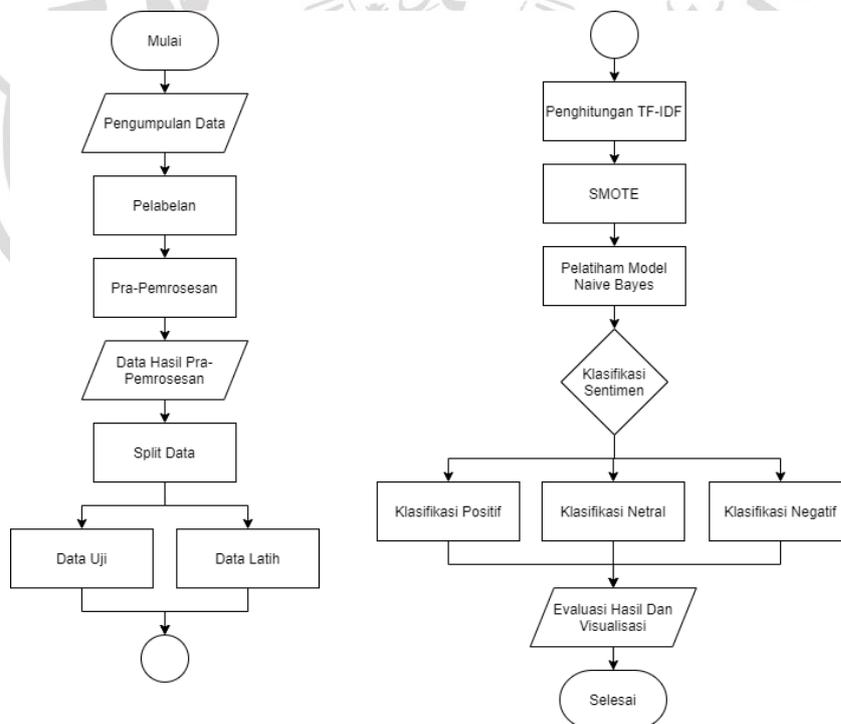
3.1 Analisis Sistem

Analisis sentimen adalah proses otomatis untuk memahami dan mengolah data teks guna menentukan sentimen dalam sebuah opini, baik positif, negatif, maupun netral. Ulasan pelanggan mencakup berbagai aspek layanan, seperti kualitas air, kecepatan respons pelayanan, dan kemudahan akses informasi. Sentimen dalam ulasan ini menunjukkan persepsi masyarakat, baik berupa opini positif, negatif, maupun netral, terhadap layanan yang diberikan oleh perusahaan. Dengan memahami pola sentimen dari ulasan pelanggan, perusahaan dapat memperoleh wawasan penting terkait persepsi publik dan menemukan aspek layanan yang perlu ditingkatkan. Penelitian ini dilakukan dengan tahapan-tahapan sistematis. Data ulasan pelanggan dikumpulkan dari platform *Google Maps*, difokuskan pada ulasan yang relevan dengan Kantor Pusat Perumda Giri Tirta Gresik. Data ini kemudian diproses menggunakan teknik pra-pemrosesan teks, seperti *casefolding*, *cleansing*, *tokenizing*, normalisasi, *stopword*, dan *stemming*, untuk menyederhanakan teks sehingga siap dianalisis. Metode *Naive Bayes* dipilih untuk klasifikasi sentimen karena kemampuannya yang efisien dalam menangani data teks yang tidak terstruktur. Algoritma ini digunakan untuk mengelompokkan ulasan menjadi sentimen positif, negatif, atau netral. Hasil dari analisis ini dievaluasi menggunakan metrik performa seperti *accuracy*, *precision*, *recall*, dan *F1-score* untuk memastikan keandalan model. Melalui pendekatan ini, penelitian diharapkan mampu memberikan gambaran mendalam tentang persepsi masyarakat terhadap layanan Perumda Giri Tirta Gresik, yang dapat digunakan sebagai bahan evaluasi dan peningkatan kualitas layanan.

3.2 Hasil analisis Sistem

Hasil analisis dari klasifikasi sentimen ulasan pelanggan terkait layanan Perumda Giri Tirta Gresik yang diunggah di *Google Maps* memberikan gambaran tentang opini publik terhadap kualitas layanan yang diberikan. Sistem klasifikasi sentimen yang diterapkan menggunakan metode *Naive Bayes*, di mana data ulasan yang telah memiliki label sentimen digunakan sebagai data latih. Proses ini

memungkinkan pengelompokan ulasan baru ke dalam tiga kategori utama: positif, netral, atau negatif. Sebanyak 214 ulasan pelanggan yang diambil dari *Google Maps* dianalisis dalam penelitian ini. Data yang dikumpulkan berasal dari ulasan pada lokasi Kantor Pusat Perumda Giri Tirta Gresik dan diproses menggunakan beberapa tahap, termasuk *casefolding*, *cleaning*, *tokenizing*, normalisasi, *stopword*, dan *stemming*, untuk memastikan data siap digunakan dalam klasifikasi sentimen. Setelah klasifikasi selesai dilakukan, setiap ulasan berhasil dikelompokkan berdasarkan sentimen yang relevan. Hasil pengelompokan ini menunjukkan pola persepsi masyarakat yang dapat diinterpretasikan sebagai masukan terkait kualitas layanan perusahaan. Variasi sentimen yang ditemukan mencerminkan beragam pengalaman pelanggan, mulai dari apresiasi, deskripsi netral, hingga kritik terhadap layanan yang diberikan. Dengan sistem ini, pengolahan ulasan yang sebelumnya tidak terstruktur menjadi lebih terorganisir dan informatif. Hasil klasifikasi ini juga memperlihatkan bagaimana opini pelanggan dapat dimanfaatkan untuk menganalisis aspek-aspek layanan yang memerlukan perhatian lebih. Alur sistem algoritma *naive bayes* untuk menentukan persepsi pelanggan dapat dilihat dari gambar 3.1.



Gambar 3. 1 Metodologi Penelitian

Flowchart ini menjelaskan langkah-langkah sistematis dalam mengklasifikasikan persepsi pelanggan menggunakan algoritma *Naive Bayes*. Setiap langkah memiliki fungsi sebagai berikut:

1. Pengumpulan Data

Data ulasan pelanggan dikumpulkan dari platform tertentu, seperti *Google Maps*. Data ini menjadi sumber utama untuk menentukan persepsi pelanggan.

2. Pelabelan

Data ulasan diberi label sentimen positif, netral, atau negatif secara manual atau otomatis. Proses ini memastikan bahwa setiap data memiliki kategori sentimen yang jelas.

3. Preprocessing

Data ulasan yang telah terkumpul diproses untuk menghilangkan elemen yang tidak relevan, seperti tanda baca, angka, atau kata-kata umum yang tidak memiliki kontribusi signifikan dalam analisis.

4. Data Hasil Pre-Processing

Data ulasan yang telah diproses disimpan sebagai dataset yang siap digunakan pada tahap selanjutnya.

5. Split Data

Data dibagi menjadi dua bagian:

- Data Latih: Digunakan untuk membangun model *Naive Bayes*.
- Data Uji: Digunakan untuk menguji akurasi model yang telah dilatih.

6. Penghitungan *TF-IDF*

Data ulasan yang sudah diberi label dianalisis menggunakan metode *TF-IDF* untuk menghitung bobot penting dari setiap kata dalam data ulasan.

7. *SMOTE*

Data ulasan yang telah diberi bobot *TF-IDF* diproses menggunakan metode *SMOTE* untuk menyeimbangkan jumlah data antar kelas sentimen. Teknik ini membuat data sintesis pada kelas minoritas berdasarkan *k*-neighbor terdekat agar model dapat belajar secara lebih seimbang.

8. Pelatihan Model *Naive Bayes*

Hasil klasifikasi diuji untuk menilai performa model menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Data hasil klasifikasi divisualisasikan dalam bentuk grafik atau tabel untuk memudahkan analisis.

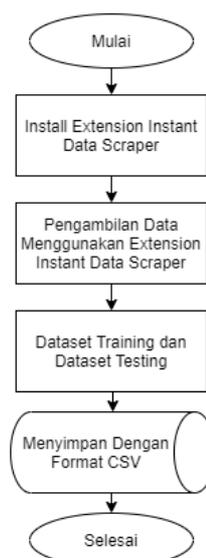
9. Klasifikasi Sentimen

Model dilatih menggunakan data latih berlabel untuk mengenali pola sentimen. Setelahnya, performa diuji dengan data uji.

10. Evaluasi Hasil dan Visualisasi

Hasil klasifikasi diuji untuk menilai performa model menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Data hasil klasifikasi divisualisasikan dalam bentuk grafik atau tabel untuk memudahkan analisis.

Pada tahap pengumpulan data ulasan pelanggan Perumda Giri Tirta Gresik dari *Google Maps* untuk menganalisis sentimen masyarakat terhadap layanan yang diberikan. Proses pengambilan data dilakukan menggunakan *extension Chrome "Instant Data Scraper"*. Dengan menggunakan tool ini, seluruh ulasan yang tersedia dapat diambil dengan lebih cepat dan akurat. Data yang diperoleh sebanyak 571 ulasan, yang mencakup berbagai opini pelanggan mengenai pengalaman mereka dalam menggunakan layanan Perumda Giri Tirta Gresik.



Gambar 3. 2 Alur *Scraping* Dataset

a. Mengakses Data dengan *Instant Data Scraper*

Menggunakan ekstensi *Instant Data Scraper*, dilakukan proses scraping untuk mengambil semua data yang dibutuhkan. Ekstensi ini digunakan untuk mengumpulkan informasi dari situs *website* tanpa perlu melakukan *coding* manual.

b. Proses Penyimpanan

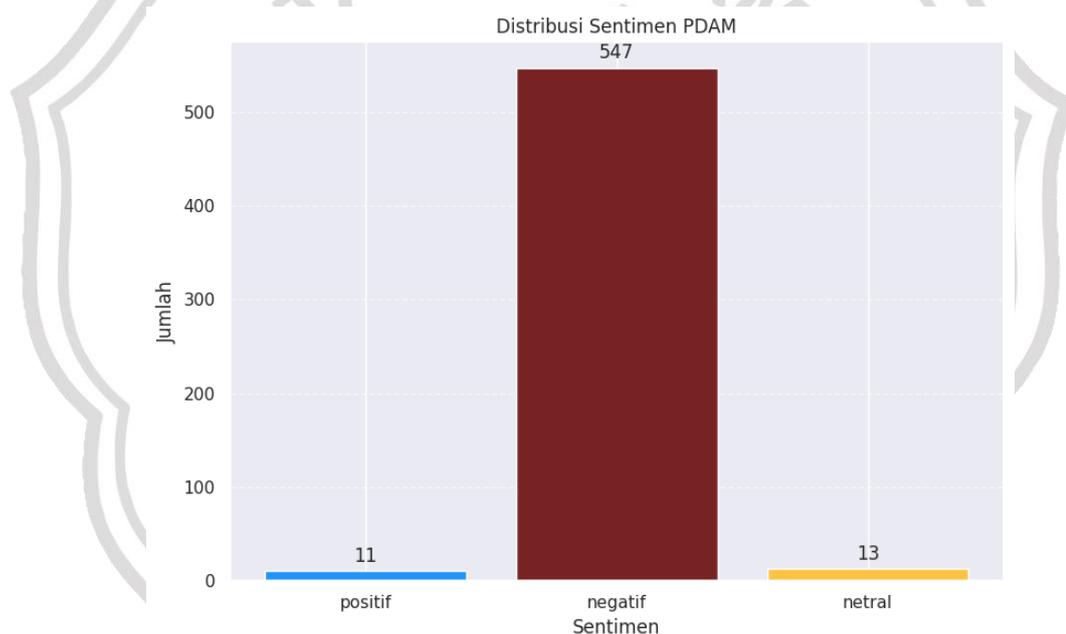
Semua data yang telah dikumpulkan disimpan dalam format *CSV* agar lebih mudah digunakan dalam tahap analisis dan pemrosesan lebih lanjut. Berikut adalah contoh hasil *scraping* dataset dengan menggunakan ekstensi *Instant Data Scraper*.

Tabel 3. 1 Contoh hasil *scraping* menggunakan *Instant Data Scraper*

username	ulasan
01_Abidatul Husnah	PDAM gk beres air gk pernah lancar satu bulan pasti ada matinya kami itu bayar. Trs bulan kemaren aja cuma bayar 45 ribu kok bulan iki jadi 190 ribu padahal makek air ya kayak biasanya gimana ini
Agus Prasetya Hariadi	Sdah brpa tahun kjadian sperti ni terulang trus, apa gk pernah mikir solusinya... Prcuma gnti atasan, rapat gk ada solusi Tempat kejadian itu"" trus
Ahmad Zhangkaru	Perusahaan ga becus, air mengalir 4 hr 1 bulan mati begitu terus, kapan berbenah beri layanan terbaik, kalau telat dendanya minta ampun, tapi air ga ngalir diam saja.
mahzanul hakim	Cara hitung PDAM sudah kayak KPU, stan meter awal 2091 akhir 2091 krn

username	ulasan
	memang air gak pernah keluar tapi pemakaian ditulis 18 kubik. Sudah bertahun-tahun air gak keluar dan sudah berkali-kali komplain tapi gak ada respon dan tagihan tiap bulan ditulis 58rb

Dalam tahap analisis data, setiap komentar dievaluasi untuk menentukan sentimennya. Sentimen tersebut kemudian dikelompokkan ke dalam tiga kategori utama yaitu positif, netral, dan negatif. Hasil klasifikasi sentimen ini kemudian divisualisasikan dalam Gambar 3.3.



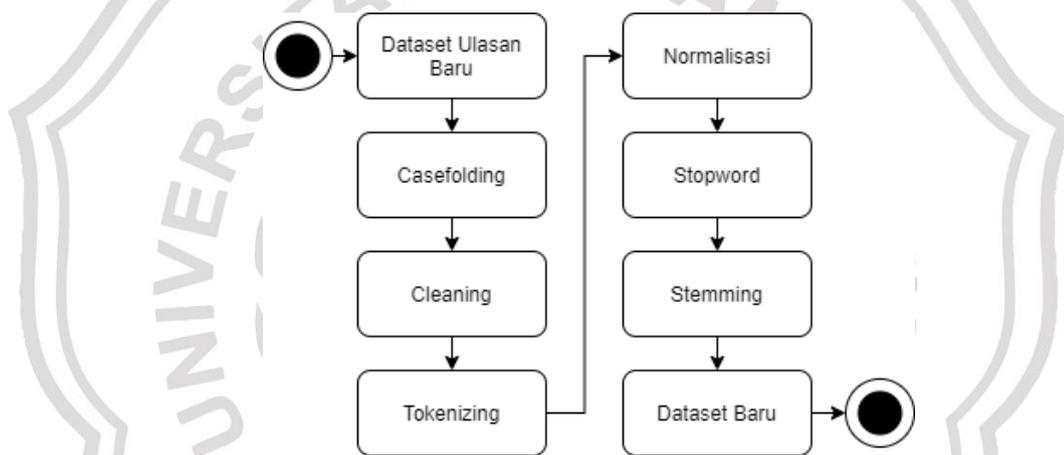
Gambar 3. 3 Distribusi Sentimen Terhadap PDAM Giri Tirta Gresik

Gambar di atas menampilkan distribusi jumlah data berdasarkan kategori sentimen terhadap PDAM, hasil dari proses pelabelan manual. Pelabelan dilakukan dengan membaca setiap komentar pelanggan dan menilai makna serta nada kalimat untuk menentukan kategorinya. Sentimen positif diberikan pada komentar yang menunjukkan kepuasan atau pujian, negatif untuk keluhan atau kekecewaan, dan netral untuk komentar informatif tanpa emosi. Proses manual

ini dipilih untuk memastikan akurasi dalam memahami konteks serta makna tersirat dalam setiap ulasan.

A. *Preprocessing*

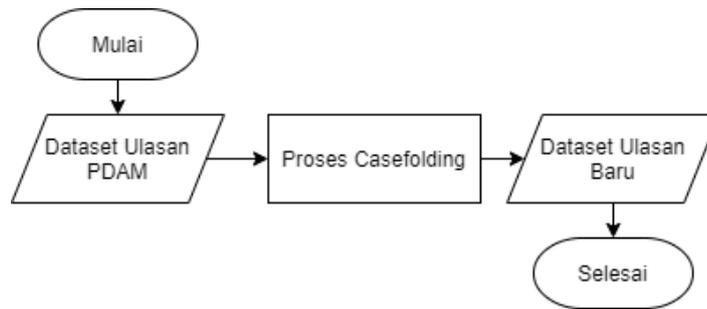
Tahap *preprocessing* memiliki peran penting dalam mendukung proses selanjutnya, terutama dalam menghilangkan atribut yang kurang relevan untuk keperluan klasifikasi. Pada tahap ini, data yang digunakan masih dalam bentuk mentah dan belum diproses. Tujuan utama *preprocessing* adalah meningkatkan kualitas data agar proses klasifikasi dapat berjalan lebih optimal. Langkah-langkah dalam *preprocessing* dijelaskan lebih lanjut pada Gambar 3.4.



Gambar 3. 4 Alur *Preprocessing*

a. *Casefolding*

Dalam proses *casefolding*, seluruh huruf kapital pada ulasan pelanggan Perumda Giri Tirta Gresik dikonversi menjadi huruf kecil. Langkah ini bertujuan untuk menyeragamkan teks agar perbedaan antara huruf besar dan kecil tidak memengaruhi analisis sentimen. Diagram alur proses *casefolding* dapat dilihat pada Gambar 3.5.



Gambar 3. 5 Alur Proses *Casefolding*

Sebagai representasi dari proses *Casefolding* dalam penelitian ini, hasilnya dapat dilihat pada Tabel 3.2.

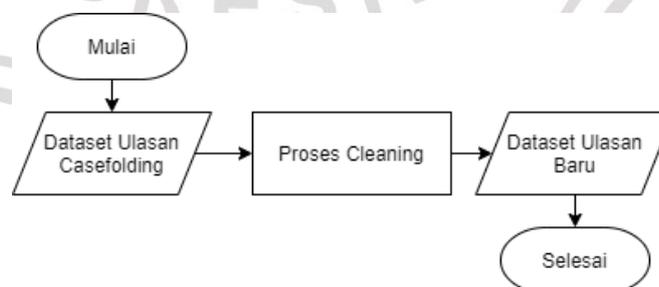
Tabel 3. 2 Contoh Hasil Proses *Casefolding*

Sebelum	Sesudah
PDAM gk beres air gk pernah lancar satu bulan pasti ada matinya kami itu bayar. Trs bulan kemaren aja cuma bayar 45 ribu kok bulan iki jadi 190 ribu padahal makek air ya kayak biasanya gimana ini	pdam gk beres air gk pernah lancar satu bulan pasti ada matinya kami itu bayar. trs bulan kemaren aja cuma bayar 45 ribu kok bulan iki jadi 190 ribu padahal makek air ya kayak biasanya gimana ini
Sdah brpa tahun kjadian sperti ni terulang trus, apa gk pernah mikir solusinya... Prcuma gnti atasan, rapat gk ada solusi Tempat kejadian itu" trus	sdah brpa tahun kjadian sperti ni terulang trus, apa gk pernah mikir solusinya... prcuma gnti atasan, rapat gk ada solusi tempat kejadian itu" trus
Perusahaan ga becus, air mengalir 4 hr 1 bulan mati begitu terus, kapan berbenah beri layanan terbaik, kalau telat dendanya minta ampun, tapi air ga ngalir diam saja.	perusahaan ga becus, air mengalir 4 hr 1 bulan mati begitu terus, kapan berbenah beri layanan terbaik, kalau telat dendanya minta ampun, tapi air ga ngalir diam saja.

Sebelum	Sesudah
Cara hitung PDAM sudah kayak KPU, stan meter awal 2091 akhir 2091 krn memang air gak pernah keluar tapi pemakaian ditulis 18 kubik. Sudah bertahun ² air gak kluar dan sudah berkali ² komplain tapi gak ada respon dan tagihan tiap bulan ditulis 58rb	cara hitung pdam sudah kayak kpu, stan meter awal 2091 akhir 2091 krn memang air gak pernah keluar tapi pemakaian ditulis 18 kubik. sudah bertahun ² air gak kluar dan sudah berkali ² komplain tapi gak ada respon dan tagihan tiap bulan ditulis 58rb

b. *Cleaning*

Proses *cleaning* bertujuan untuk membersihkan data teks dengan menghapus elemen yang tidak diperlukan atau kurang relevan, seperti tanda baca, angka, simbol, *URL*, dan karakter khusus lainnya. Beberapa contoh elemen yang dihilangkan mencakup tautan yang diawali dengan *http* atau *bit.ly*, serta simbol khusus seperti `!@#%&()_+{}[]<>?,:. .`. Langkah ini dilakukan untuk meningkatkan kualitas data sehingga analisis dapat dilakukan dengan lebih mudah serta meningkatkan akurasi pada tahap pemrosesan berikutnya, seperti klasifikasi atau analisis sentimen. Diagram alur yang menggambarkan proses *cleaning* dapat dilihat pada Gambar 3.6.



Gambar 3. 6 Alur Proses *Cleaning*

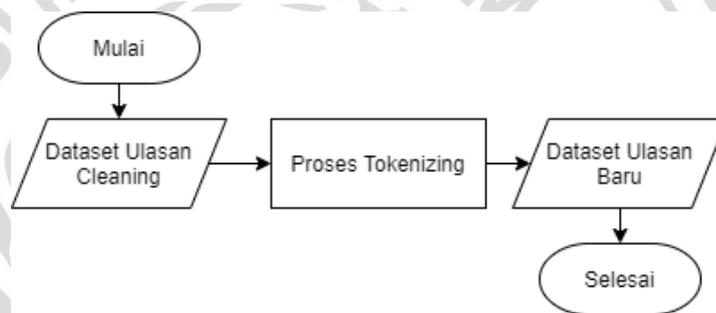
Sebagai representasi dari proses *Cleaning* dalam penelitian ini, hasilnya dapat dilihat pada Tabel 3.3.

Tabel 3. 3 Contoh Hasil Proses *Cleaning*

Sebelum	Sesudah
pdam gk beres air gk pernah lancar satu bulan pasti ada matinya kami itu bayar. trs bulan kemaren aja cuma bayar 45 ribu kok bulan iki jadi 190 ribu padahal makek air ya kayak biasanya gimana ini	pdam gk beres air gk pernah lancar satu bulan pasti ada matinya kami itu bayar trs bulan kemaren aja cuma bayar 45 ribu kok bulan iki jadi 190 ribu padahal makek air ya kayak biasanya gimana ini
sdah brpa tahun kjadian sperti ni terulang trus, apa gk pernah mikir solusinya... prcuma gnti atasan, rapat gk ada solusi tempat kejadian itu" trus	sdah brpa tahun kjadian sperti ni terulang trus apa gk pernah mikir solusinya prcuma gnti atasan rapat gk ada solusi tempat kejadian itu trus
perusahaan ga becus, air mengalir 4 hr 1 bulan mati begitu terus, kapan berbenah beri layanan terbaik, kalau telat dendanya minta ampun, tapi air ga ngalir diam saja.	perusahaan ga becus air mengalir hr bulan mati begitu terus kapan berbenah beri layanan terbaik kalau telat dendanya minta ampun tapi air ga ngalir diam saja
cara hitung pdam sudah kayak kpu, stan meter awal 2091 akhir 2091 krn memang air gak pernah keluar tapi pemakaian ditulis 18 kubik. sudah bertahun ² air gak kluar dan sudah berkali ² komplain tapi gak ada respon dan tagihan tiap bulan ditulis 58rb	cara hitung pdam sudah kayak kpu stan meter awal akhir krn memang air gak pernah keluar tapi pemakaian ditulis kubik sudah bertahun air gak kluar dan sudah berkali komplain tapi gak ada respon dan tagihan tiap bulan ditulis rb

c. *Tokenizing*

Proses *tokenizing* adalah memecah ulasan pelanggan PDAM Giri Tirta Gresik menjadi bagian-bagian kecil yang disebut token. Token ini dapat berupa kata, frasa, atau simbol, tergantung pada tujuan analisis. Dengan langkah ini, teks menjadi lebih terstruktur, memungkinkan setiap token dianalisis secara terpisah. Proses ini berperan penting dalam berbagai aplikasi pemrosesan bahasa alami *Natural Language Processing* atau *NLP*, seperti pengelompokan teks, analisis sentimen, dan ekstraksi data. Diagram alur yang menggambarkan proses *tokenizing* dapat ditemukan pada Gambar 3.7.



Gambar 3. 7 Alur Proses *Tokenizing*

Sebagai representasi dari proses *Tokenizing* dalam penelitian ini, hasilnya dapat dilihat pada Tabel 3.4.

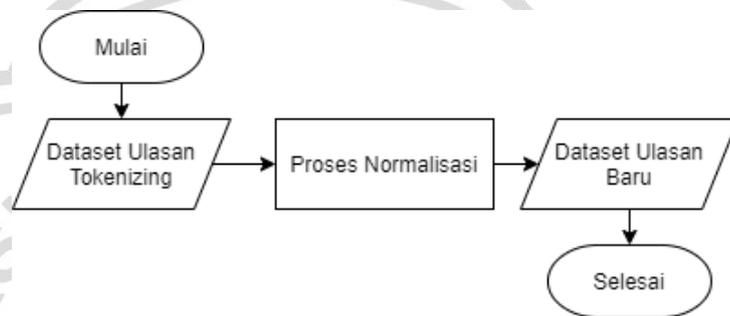
Tabel 3. 4 Contoh Hasil *Tokenizing*

Sebelum	Sesudah
pdam gk beres air gk pernah lancar satu bulan pasti ada matinya kami itu bayar trs bulan kemaren aja cuma bayar ribu kok bulan iki jadi ribu padahal makek air ya kayak biasanya gimana ini	['pdam', 'gk', 'beres', 'air', 'gk', 'pernah', 'lancar', 'satu', 'bulan', 'pasti', 'ada', 'matinya', 'kami', 'itu', 'bayar', 'trs', 'bulan', 'kemaren', 'aja', 'cuma', 'bayar', 'ribu', 'kok', 'bulan', 'iki', 'jadi', 'ribu', 'padahal', 'makek', 'air',

Sebelum	Sesudah
	'ya', 'kayak', 'biasanya', 'gimana', 'ini']
sdah brpa tahun kjadian sperti ni terulang trus apa gk pernah mikir solusinya prcuma gnti atasan rapat gk ada solusi tempat kejadian itu trus	['sdah', 'brpa', 'tahun', 'kjadian', 'sperti', 'ni', 'terulang', 'trus', 'apa', 'gk', 'pernah', 'mikir', 'solusinya', 'prcuma', 'gnti', 'atasan', 'rapat', 'gk', 'ada', 'solusi', 'tempat', 'kejadian', 'itu', 'trus']
perusahaan ga becus air mengalir hr bulan mati begitu terus kapan berbenah beri layanan terbaik kalau telat dendanya minta ampun tapi air ga ngalir diam saja	['perusahaan', 'ga', 'becus', 'air', 'mengalir', 'hr', 'bulan', 'mati', 'begitu', 'terus', 'kapan', 'berbenah', 'beri', 'layanan', 'terbaik', 'kalau', 'telat', 'dendanya', 'minta', 'ampun', 'tapi', 'air', 'ga', 'ngalir', 'diam', 'saja']
cara hitung pdam sudah kayak kpu stan meter awal akhir krn memang air gak pernah keluar tapi pemakaian ditulis kubik sudah bertahun air gak kluar dan sudah berkali komplain tapi gak ada respon dan tagihan tiap bulan ditulis rb	['cara', 'hitung', 'pdam', 'sudah', 'kayak', 'kpu', 'stan', 'meter', 'awal', 'akhir', 'krn', 'memang', 'air', 'gak', 'pernah', 'keluar', 'tapi', 'pemakaian', 'ditulis', 'kubik', 'sudah', 'bertahun', 'air', 'gak', 'kluar', 'dan', 'sudah', 'berkali', 'komplain', 'tapi', 'gak', 'ada', 'respon', 'dan', 'tagihan', 'tiap', 'bulan', 'ditulis', 'rb']

d. Normalisasi

Normalisasi teks adalah proses mengubah kata tidak baku, singkatan, atau slang menjadi bentuk standar sesuai KBBI, seperti "abis" menjadi "habis". Dalam analisis sentimen ulasan pelanggan PDAM Giri Tirta Gresik, normalisasi meningkatkan keseragaman data sehingga lebih mudah diproses oleh algoritma seperti *Naive Bayes*. Proses ini membantu meningkatkan akurasi klasifikasi, seperti ditunjukkan pada Gambar 3.8.



Gambar 3. 8 Alur Proses Normalisasi

Sebagai representasi dari proses Normalisasi dalam penelitian ini, hasilnya dapat dilihat pada Tabel 3.5.

Tabel 3. 5 Contoh Hasil Normalisasi

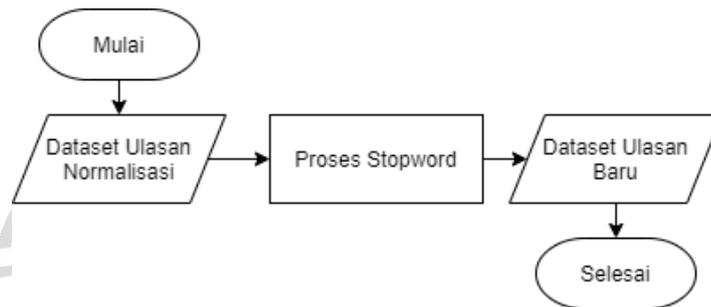
Sebelum	Sesudah
['pdam', 'gk', 'beres', 'air', 'gk', 'pernah', 'lancar', 'satu', 'bulan', 'pasti', 'ada', 'matinya', 'kami', 'itu', 'bayar', 'trs', 'bulan', 'kemaren', 'aja', 'cuma', 'bayar', 'ribu', 'kok', 'bulan', 'iki', 'jadi', 'ribu', 'padahal', 'makek', 'air', 'ya', 'kayak', 'biasanya', 'gimana', 'ini']	pdam tidak beres air tidak pernah lancar satu bulan pasti ada matinya kami itu bayar terus bulan kemarin aja cuma bayar ribu kok bulan ini jadi ribu padahal pakai air ya seperti biasanya gimana ini
['sdah', 'brpa', 'tahun', 'kjadian', 'sperti', 'ni', 'terulang', 'trus', 'apa', 'gk', 'pernah', 'mikir', 'solusinya', 'percuma']	sudah berapa tahun kejadian seperti ini terulang terus apa tidak pernah mikir solusinya percuma

Sebelum	Sesudah
'solusinya', 'prcuma', 'gnti', 'atasan', 'rapat', 'gk', 'ada', 'solusi', 'tempat', 'kejadian', 'itu', 'trus']	ganti atasan rapat tidak ada solusi tempat kejadian itu terus
['perusahaan', 'ga', 'becus', 'air', 'mengalir', 'hr', 'bulan', 'mati', 'begitu', 'terus', 'kapan', 'berbenah', 'beri', 'layanan', 'terbaik', 'kalau', 'telat', 'dendanya', 'minta', 'ampun', 'tapi', 'air', 'ga', 'ngalir', 'diam', 'saja']	perusahaan tidak becus air mengalir hari bulan mati begitu terus kapan berbenah beri layanan terbaik kalau telat dendanya minta ampun tapi air tidak ngalir diam saja
['cara', 'hitung', 'pdam', 'sudah', 'kayak', 'kpu', 'stan', 'meter', 'awal', 'akhir', 'krn', 'memang', 'air', 'gak', 'pernah', 'keluar', 'tapi', 'pemakaian', 'ditulis', 'kubik', 'sudah', 'bertahun', 'air', 'gak', 'kluar', 'dan', 'sudah', 'berkali', 'komplain', 'tapi', 'gak', 'ada', 'respon', 'dan', 'tagihan', 'tiap', 'bulan', 'ditulis', 'rb']	cara hitung pdam sudah seperti kpu stan meter awal akhir karena memang air tidak pernah keluar tapi pemakaian ditulis kubik sudah bertahun air tidak kluar dan sudah berkali komplain tapi tidak ada respon dan tagihan tiap bulan ditulis ribu

e. *Stopword*

Proses *Stopword* dimulai dengan mengidentifikasi dan menghapus kata-kata umum yang sering muncul tetapi tidak memiliki makna signifikan dalam analisis sentimen, seperti "dan," "yang," "adalah," dan "di." Kata-kata ini tidak memberikan kontribusi penting terhadap pemahaman konteks ulasan pelanggan

terhadap layanan Perumda Giri Tirta Gresik di *Google Maps*. Dengan menghapus stopwords, proses analisis menjadi lebih efisien karena algoritma *Naive Bayes* dapat lebih fokus pada kata-kata yang relevan dan informatif. Diagram alur proses *Stopword* dapat dilihat pada Gambar 3.9.



Gambar 3. 9 Alur Proses *Stopword*

Sebagai representasi dari proses *Stopword* dalam penelitian ini, hasilnya dapat dilihat pada Tabel 3.6.

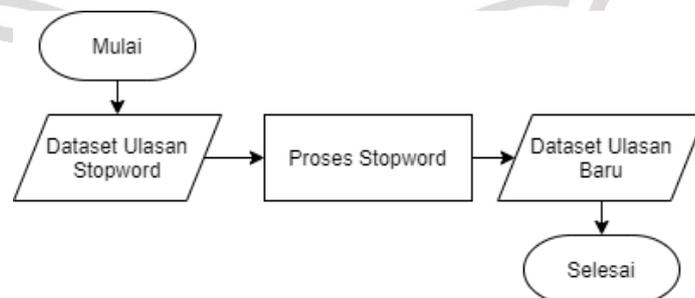
Tabel 3. 6 Contoh Hasil *Stopword*

Sebelum	Sesudah
pdam tidak beres air tidak pernah lancar satu bulan pasti ada matinya kami itu bayar terus bulan kemarin aja cuma bayar ribu kok bulan ini jadi ribu padahal pakai air ya seperti biasanya gimana ini	pdam tidak beres air tidak pernah lancar satu bulan matinya bayar terus bulan kemarin aja cuma bayar ribu kok bulan jadi ribu padahal pakai air ya biasanya gimana
sudah berapa tahun kejadian sperti ini terulang terus apa tidak pernah mikir solusinya percuma ganti atasan rapat tidak ada solusi tempat kejadian itu terus	berapa tahun kejadian sperti terulang terus apa tidak pernah mikir solusinya percuma ganti atasan rapat tidak solusi tempat kejadian terus
perusahaan tidak becus air mengalir hari bulan mati begitu	perusahaan tidak becus air mengalir hari bulan mati terus

Sebelum	Sesudah
terus kapan berbenah beri layanan terbaik kalau telat dendanya minta ampun tapi air tidak ngalir diam saja	kapan berbenah beri layanan terbaik kalau telat dendanya minta ampun air tidak ngalir diam
cara hitung pdam sudah seperti kpu stan meter awal akhir karena memang air tidak pernah keluar tapi pemakaian ditulis kubik sudah bertahun air tidak keluar dan sudah berkali komplain tapi tidak ada respon dan tagihan tiap bulan ditulis ribu	cara hitung pdam kpu stan meter awal akhir memang air tidak pernah keluar pemakaian ditulis kubik bertahun air tidak keluar berkali komplain tidak respon tagihan tiap bulan ditulis ribu

f. *Stemming*

Proses stemming bertujuan mengubah kata menjadi bentuk dasar dengan menghapus imbuhan atau akhiran, sehingga variasi kata seperti "berjalan," "berlarian," dan "berlari-lari" disederhanakan menjadi "lari." Dalam analisis sentimen ulasan pelanggan Perumda Giri Tirta Gresik di *Google Maps*, stemming membantu mengurangi keragaman kata, sehingga meningkatkan efisiensi dan akurasi algoritma Naive Bayes. Diagram alur proses stemming ditampilkan pada Gambar 3.10.



Gambar 3. 10 Alur Proses *Stemming*

Sebagai representasi dari proses *Stemming* dalam penelitian ini, hasilnya dapat dilihat pada Tabel 3.7.

Tabel 3. 7 Contoh Hasil *Stemming*

Sebelum	Sesudah
pdam tidak beres air tidak pernah lancar satu bulan matinya bayar terus bulan kemarin aja cuma bayar ribu kok bulan jadi ribu padahal pakai air ya biasanya gimana	pdam tidak beres air tidak pernah lancar satu bulan mati bayar terus bulan kemarin aja cuma bayar ribu kok bulan jadi ribu padahal pakai air ya biasa gimana
berapa tahun kejadian sperti terulang terus apa tidak pernah mikir solusinya percuma ganti atasan rapat tidak solusi tempat kejadian terus	berapa tahun jadi sperti ulang terus apa tidak pernah mikir solusi percuma ganti atas rapat tidak solusi tempat jadi terus
perusahaan tidak becus air mengalir hari bulan mati terus kapan berbenah beri layanan terbaik kalau telat dendanya minta ampun air tidak ngalir diam	usaha tidak becus air alir hari bulan mati terus kapan benah beri layan baik kalau telat denda minta ampun air tidak ngalir diam
cara hitung pdam kpu stan meter awal akhir memang air tidak pernah keluar pemakaian ditulis kubik bertahun air tidak keluar berkali komplain tidak respon tagihan tiap bulan ditulis ribu	cara hitung pdam kpu stan meter awal akhir memang air tidak pernah keluar pakai tulis kubik tahun air tidak kluar kali komplain tidak respon tagih tiap bulan tulis ribu

B. Ekstraksi Fitur

Dalam tahap ekstraksi fitur menggunakan *TF-IDF*, ulasan pelanggan Perumda Giri Tirta Gresik di *Google Maps* dikonversi ke dalam bentuk vektor numerik. Setiap kata dalam ulasan diberikan nilai bobot berdasarkan frekuensi kemunculannya dalam teks serta relevansinya terhadap keseluruhan kumpulan ulasan. Metode *TF-IDF* memungkinkan teks untuk direpresentasikan dengan lebih informatif, sehingga dapat meningkatkan efektivitas analisis sentimen menggunakan *Naive Bayes*. Nilai bobot yang diperoleh akan digunakan sebagai input dalam tahap klasifikasi untuk mengidentifikasi apakah ulasan memiliki sentimen positif, netral, atau negatif.

Tabel 3. 8 Contoh Ulasan Pelanggan Dari Data Uji

D1	layan bagus air bagus
D2	apakah pdam masih layan distribusi air ke rumah atau sudah alih ke bidang lain
D3	air mati minggu gak ada pemberitahuan lapor ke kantor tidak ada tindak lanjut sangat buruk

Tahap pembobotan kata diawali dengan mengambil kata-kata yang telah melewati proses *preprocessing* dari *database*. Kemudian, dilakukan penghitungan jumlah kemunculan setiap kata dalam dokumen untuk menentukan tingkat bobotnya. Proses ini hanya menggunakan kata-kata yang sudah melalui tahap *preprocessing*, sehingga menghasilkan nilai bobot yang mencerminkan tingkat kepentingan masing-masing kata. Hasil akhir dari proses pembobotan dapat dilihat pada Tabel 3.9.

Tabel 3. 9 Nilai Bobot Masing-Masing Kata

TERM	TF			DF	idf=log(D/DF)	TF-IDF		
	D1	D2	D3			D1	D2	D3
layan	1	1	0	2	0.1761	0.1761	0.1761	0.0000
bagus	2	0	0	2	0.1761	0.3522	0.0000	0.0000
air	1	1	1	3	0.0000	0.0000	0.0000	0.0000
apakah	0	1	0	1	0.4771	0.0000	0.4771	0.0000
pdam	0	1	0	1	0.4771	0.0000	0.4771	0.0000
masih	0	1	0	1	0.4771	0.0000	0.4771	0.0000

TERM	TF			DF	idf=log(D/DF)	TF-IDF		
	D1	D2	D3			D1	D2	D3
distribusi	0	1	0	1	0.4771	0.0000	0.4771	0.0000
ke	0	2	1	3	0.0000	0.0000	0.0000	0.0000
rumah	0	1	0	1	0.4771	0.0000	0.4771	0.0000
atau	0	1	0	1	0.4771	0.0000	0.4771	0.0000
sudah	0	1	0	1	0.4771	0.0000	0.4771	0.0000
alih	0	1	0	1	0.4771	0.0000	0.4771	0.0000
bidang	0	1	0	1	0.4771	0.0000	0.4771	0.0000
lain	0	1	0	1	0.4771	0.0000	0.4771	0.0000
mati	0	0	1	1	0.4771	0.0000	0.0000	0.4771
minggu	0	0	1	1	0.4771	0.0000	0.0000	0.4771
gak	0	0	1	1	0.4771	0.0000	0.0000	0.4771
ada	0	0	2	2	0.1761	0.0000	0.0000	0.3522
pemberitahuan	0	0	1	1	0.4771	0.0000	0.0000	0.4771
lapor	0	0	1	1	0.4771	0.0000	0.0000	0.4771
kantor	0	0	1	1	0.4771	0.0000	0.0000	0.4771
tidak	0	0	1	1	0.4771	0.0000	0.0000	0.4771
tindak	0	0	1	1	0.4771	0.0000	0.0000	0.4771
lanjut	0	0	1	1	0.4771	0.0000	0.0000	0.4771
sangat	0	0	1	1	0.4771	0.0000	0.0000	0.4771
buruk	0	0	1	1	0.4771	0.0000	0.0000	0.4771

Tabel yang ditampilkan merupakan hasil perhitungan *TF-IDF* (*Term Frequency - Inverse Document Frequency*) untuk berbagai kata yang terdapat dalam tiga dokumen berbeda D1, D2, dan D3. Perhitungan ini dilakukan untuk menentukan seberapa penting suatu kata dalam suatu dokumen berdasarkan frekuensi kemunculannya dan seberapa unik kata tersebut dalam keseluruhan dokumen.

Pada tabel, kolom *TF* (*Term Frequency*) menunjukkan jumlah kemunculan kata dalam masing-masing dokumen. Misalnya, kata "air" muncul sekali dalam setiap dokumen D1, D2, dan D3, sehingga nilai *TF* untuk kata ini di setiap dokumen adalah 1.

Pada kolom *DF* (*Document Frequency*) menunjukkan jumlah dokumen yang mengandung kata tertentu. Misalnya, kata "air" muncul di ketiga dokumen, sehingga nilai *DF*-nya adalah 3, sedangkan kata "pemberitahuan" hanya muncul dalam satu dokumen, sehingga *DF*-nya 1.

Perhitungan *IDF* (*Inverse Document Frequency*) dilakukan menggunakan rumus $IDF = \log\left(\frac{D}{DF}\right)$ dengan D adalah jumlah total

dokumen 3 dalam kasus ini dan DF adalah jumlah dokumen yang mengandung kata tersebut. Misalnya, untuk kata "air", yang muncul di semua dokumen $DF = 3$, perhitungannya adalah $IDF = \log\left(\frac{3}{3}\right) = \log(1) = 0$ Nilai ini menunjukkan bahwa kata "air" terlalu umum dan tidak memberikan informasi signifikan dalam analisis. Sebaliknya, kata "pemberitahuan" yang hanya muncul di satu dokumen memiliki nilai $IDF = \log\left(\frac{3}{1}\right) = \log(3) = 0.4771$.

Bobot akhir $TF-IDF$ diperoleh dengan mengalikan nilai TF dan IDF untuk setiap kata dalam setiap dokumen. Sebagai contoh, untuk kata "bagus" di dokumen D1, yang memiliki $TF = 2$ dan $IDF = 0.1761$, maka $TF - IDF = 2 \times 0.1761 = 0.3522$ Nilai ini menunjukkan bahwa kata "bagus" memiliki bobot 0.3522 dalam dokumen D1, yang berarti kata tersebut cukup penting dalam dokumen tersebut dibandingkan dengan dokumen lain dalam koleksi karena memiliki frekuensi yang relatif tinggi dalam dokumen tetapi tidak terlalu sering muncul di seluruh dataset.

C. Penerapan *SMOTE*

Dalam penelitian ini, teknik *SMOTE* (*Synthetic Minority Oversampling Technique*) digunakan untuk menangani masalah ketidakseimbangan kelas sentimen dalam dataset ulasan pelanggan Perumda Giri Tirta Gresik. Ketidakseimbangan ini terjadi karena sebagian besar ulasan mengandung sentimen negatif, sementara ulasan dengan sentimen netral dan positif jauh lebih sedikit. Ketidakseimbangan ini berpotensi menyebabkan model klasifikasi bias terhadap kelas mayoritas dan mengabaikan pola dari kelas minoritas.

SMOTE bekerja dengan cara membuat data sintetis untuk kelas minoritas melalui interpolasi antar data yang berdekatan *k-nearest neighbors* dalam ruang fitur. Tidak seperti metode *oversampling* konvensional yang hanya menggandakan data, *SMOTE* menciptakan sampel baru berdasarkan jarak antara titik data yang ada, sehingga

menghasilkan variasi data yang lebih alami dan memperkaya representasi kelas minoritas.

Terdapat beberapa variasi teknik *SMOTE*, yaitu *Regular SMOTE*, *Borderline-SMOTE*, *SVM-SMOTE*, *ADASYN*, dan *KMeans-SMOTE*. Penulis memilih *Regular SMOTE* karena teknik ini sederhana dan sesuai untuk data teks yang telah direpresentasikan dalam bentuk numerik seperti *TF-IDF*. Karakteristik data ulasan dalam penelitian ini tidak menunjukkan batasan kelas yang kompleks atau distribusi yang ekstrem, sehingga pendekatan interpolasi standar dinilai cukup untuk meningkatkan representasi kelas minoritas.

Penerapan *SMOTE* dilakukan setelah proses pembobotan *TF-IDF*, sehingga data yang diperoleh sudah dalam bentuk vektor numerik yang siap digunakan untuk proses *oversampling*. Dengan menyeimbangkan distribusi antar kelas, model *Naive Bayes* yang digunakan dalam penelitian ini dapat mempelajari pola dari setiap sentimen dengan lebih adil, dan hasil klasifikasi menjadi lebih representatif. Berikut hasil *oversampling SMOTE* pada tabel 3.10.

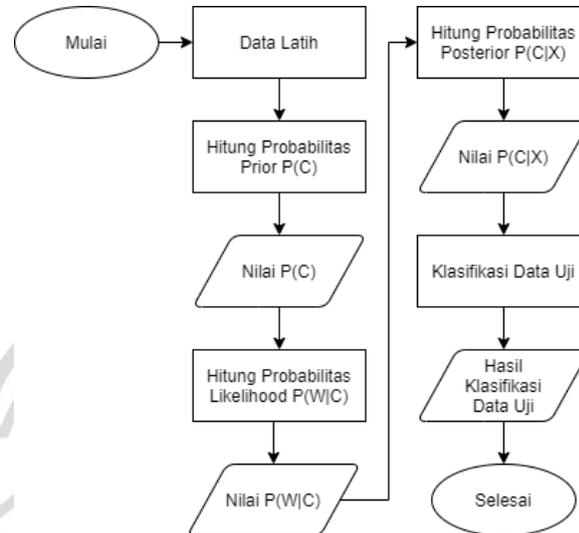
Tabel 3. 10 Hasil *Oversampling SMOTE*

Kelas Sentimen	Jumlah Data Sebelum <i>SMOTE</i>	Jumlah Data Setelah <i>SMOTE</i>
Negatif	437	437
Netral	10	437
Positif	9	437

D. *Naive Bayes*

Flowchart berikut menggambarkan proses algoritma *Naive Bayes* dalam pelatihan model dan pengujian data untuk klasifikasi sentimen. Proses dimulai dari pembagian dataset, perhitungan probabilitas, hingga penentuan kelas sentimen berdasarkan nilai probabilitas tertinggi. Selain itu, dilakukan evaluasi model secara sistematis untuk memastikan keakuratan

hasil klasifikasi serta meningkatkan performa model dalam mengelola data dan mengidentifikasi kategori sentimen dengan tepat.



Gambar 3. 11 Flowchart Algoritma Naive Bayes

Berikut adalah penjelasan tiap langkah dalam flowchart algoritma Naive Bayes:

1. Mulai, Proses dimulai untuk melakukan klasifikasi menggunakan algoritma *Naive Bayes*.
2. Data Latih, Dataset yang digunakan untuk membangun model. Data ini berisi fitur (variabel input) dan label kelas.
3. Hitung Probabilitas Prior $P(C)$, Probabilitas awal dari masing-masing kelas dihitung berdasarkan distribusi kelas dalam data latih.
4. Nilai $P(C)$, Menyimpan nilai probabilitas prior untuk masing-masing kelas.
5. Hitung Probabilitas Likelihood $P(W|C)$, Menghitung peluang tiap fitur (W) muncul pada masing-masing kelas (C).
6. Nilai $P(W|C)$, Menyimpan nilai probabilitas likelihood untuk digunakan dalam klasifikasi data uji.
7. Hitung Probabilitas Posterior $P(C|X)$, Menggunakan teorema Bayes untuk menghitung probabilitas sebuah data uji (X) termasuk ke kelas tertentu (C).

8. Nilai $P(C|X)$, Menyimpan hasil perhitungan probabilitas posterior untuk setiap kelas berdasarkan fitur data uji.
9. Klasifikasi Data Uji, Menentukan kelas dengan probabilitas posterior tertinggi sebagai hasil klasifikasi.
10. Hasil Klasifikasi Data Uji, Menampilkan kelas akhir dari data uji berdasarkan hasil perhitungan sebelumnya.
11. Selesai, Model dievaluasi dengan metrik seperti akurasi, lalu hasil sentimen divisualisasikan dalam grafik untuk analisis.

3.3 Representasi Model

Model *Naive Bayes* merepresentasikan probabilitas kemunculan kata dalam setiap kategori sentimen, memungkinkan sistem mengenali pola dominan dalam ulasan pelanggan. Analisis ini memberikan wawasan lebih dalam tentang persepsi masyarakat serta membantu perusahaan dalam meningkatkan kualitas layanan berdasarkan umpan balik pelanggan. Data ulasan dibagi terlebih dahulu menjadi data latih dan data uji menggunakan fungsi *train_test_split* dari pustaka *Python*, kemudian dikonversi ke dalam bentuk vektor menggunakan teknik *TF-IDF*. Model kemudian dilatih dengan data pelatihan untuk mempelajari pola sentimen, sementara data pengujian digunakan untuk mengukur efektivitas klasifikasi sentimen pelanggan terhadap layanan Perumda Giri Tirta Gresik.

1. Probabilitas *prior*

Probabilitas ini dihitung dengan membandingkan jumlah sampel dalam setiap kelas dengan total sampel pada data latih. Sebagai contoh, perhitungan dilakukan berdasarkan ulasan pada Tabel 3.8, di mana D1 termasuk dalam kategori positif, D2 dalam kategori netral, dan D3 dalam kategori negatif. Dalam contoh perhitungan ini, digunakan dokumen 1 (D1) untuk contoh perhitungan menggunakan rumus pada persamaan (2.6)

Maka probabilitas *prior* untuk setiap kategori dihitung sebagai berikut:

$$P(\text{Positif}) = \frac{1}{3} = 0.3333$$

$$P(\text{Netral}) = \frac{1}{3} = 0.3333$$

$$P(\text{Negatif}) = \frac{1}{3} = 0.3333$$

2. Probabilitas *likelihood*

Dalam *Naive Bayes*, setelah menghitung probabilitas *prior*, kita perlu menghitung probabilitas setiap kata muncul dalam masing-masing kategori sentimen. Ini disebut probabilitas *likelihood*, yang dihitung dengan rumus pada persamaan (2.7):

$$P(w|C) = \frac{n_w + 1}{N_c + V}$$

Di ketahui total jumlah kata dalam kelas C adalah $D1 = 4$, $D2 = 14$, $D3 = 15$ dengan total keseluruhan kata V adalah 26.

- Probabilitas “layan” dalam kategori positif:

$$P(\text{layan}|\text{Positif}) = \frac{1 + 1}{4 + 26} = \frac{2}{30} = 0.0667$$

- Probabilitas “layan” dalam kategori netral:

$$P(\text{layan}|\text{netral}) = \frac{1 + 1}{14 + 26} = \frac{2}{40} = 0.0500$$

- Probabilitas “layan” dalam kategori negatif:

$$P(\text{layan}|\text{negatif}) = \frac{0 + 1}{15 + 26} = \frac{1}{41} = 0.0244$$

Tabel 3. 11 Hasil Perhitungan Probabilitas Kata Dalam Setiap Kategori

Kata	Positif	Netral	Negatif
Layan	0.0667	0.0500	0.0244
Bagus	0.1000	0.0250	0.0244
Air	0.0667	0.0500	0.0488

3. Probabilitas *posterior*

Dalam *Naive Bayes*, kita menghitung probabilitas suatu ulasan termasuk dalam kategori positif, netral, atau negatif menggunakan *Teorema Bayes*.

Rumus dasar *Naive Bayes* Menggunakan pada persamaan (2.4):

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)}$$

Karena $P(X)$ konstan untuk semua kelas, kita hanya perlu fokus pada persamaan (2.5):

$$P(C|X) \propto P(X|C) \times P(C)$$

Berikut perhitungannya digunakan tabel 3.10 hasil dari perhitungan probabilitas kata setiap kategori:

- Kategori positif

$$\begin{aligned} P(\text{Positif}|X) &\propto P(\text{Positif}) \times P(\text{layan}|\text{Positif}) \times \dots \times P(\text{air}|\text{positif}) \\ &= 0.3333 \times 0.0667 \times 0.1000 \times 0.0667 \\ &= 0.3333 \times 0.00044444 \\ &= 0.00014813 \end{aligned}$$

- Kategori Netral

$$\begin{aligned} &= 0.3333 \times 0.0500 \times 0.0250 \times 0.0500 \\ &= 0.3333 \times 0.00006250 \\ &= 0.00002083 \end{aligned}$$

- Kategori Negatif

$$\begin{aligned} &= 0.3333 \times 0.0244 \times 0.0244 \times 0.0488 \\ &= 0.3333 \times 0.00002902 \\ &= 0.00000967 \end{aligned}$$

4. Klasifikasi data uji

Klasifikasi data uji adalah tahap pengujian model yang telah dilatih untuk menilai akurasi dalam mengenali sentimen ulasan positif, netral, atau negatif berdasarkan pola yang dipelajari. Berdasarkan hasil perhitungan probabilitas *posterior*, kalimat dalam dokumen D1 yang terdapat pada Tabel 3.8 "layan bagus air bagus" memiliki nilai probabilitas tertinggi dalam kategori positif, yaitu sebesar 0.00014813. Nilai ini menunjukkan bahwa kata-kata dalam dokumen D1 lebih sering muncul dalam kategori sentimen positif dibandingkan kategori lainnya. Oleh karena itu, dokumen D1 diklasifikasikan sebagai ulasan positif terhadap layanan. Hasil Perhitungan untuk ulasan D1, D2 dan D3 ditampilkan pada tabel 3.12.

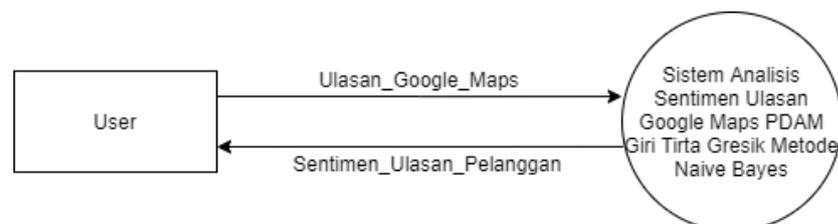
Tabel 3. 12 Hasil Data Uji

	Ulasan	Probabilitas Positif	Probabilitas Netral	Probabilitas Negatif	Sentimen
D1	layan bagus air bagus	0.00014813	0.00002083	0.00000967	Positif
D2	apakah pdam masih layan distribusi air ke rumah atau sudah alih ke bidang lain	0.00000000 0000000000 083622	0.00000000 0000000006 102905	0.00000000 0000000000 001441	Netral
D3	air mati minggu gak ada pemberitahua n lapor ke kantor tidak ada tindak lanjut sangat buruk	0.00000000 0000000000 0013937	0.00000000 0000000000 0000745	0.00000000 0000000000 2159596	Negatif

3.4 Perancangan system

3.4.1 Diagram Konteks

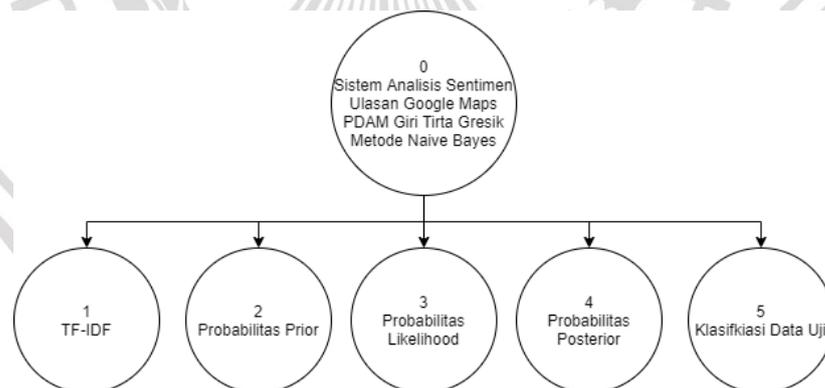
Diagram konteks dalam suatu sistem digunakan untuk merepresentasikan entitas eksternal atau pihak yang berhubungan dengan sistem tersebut.

**Gambar 3. 12** Diagram Konteks Sistem

Gambar 3.12 menggambarkan alur sistem analisis sentimen terhadap ulasan pelanggan PDAM Giri Tirta Gresik yang diperoleh dari *Google Maps* dengan metode *Naive Bayes*. Ulasan *Google Maps* merupakan komentar atau tanggapan pelanggan terkait pelayanan PDAM, seperti kualitas air, kecepatan pelayanan, dan respons petugas. Ulasan ini bersifat teks bebas dan menjadi sumber utama dalam analisis. Selanjutnya, sistem memproses ulasan menggunakan algoritma *Naive Bayes* untuk mengklasifikasikan sentimen menjadi positif, negatif, atau netral, berdasarkan kata-kata yang digunakan. Hasil klasifikasi disebut sebagai sentimen ulasan pelanggan, yang kemudian dikembalikan kepada pengguna sebagai bentuk evaluasi layanan. Informasi ini membantu PDAM memahami persepsi pelanggan dan meningkatkan kualitas pelayanan secara berkelanjutan.

3.4.2 Diagram Jenjang

Diagram jenjang adalah representasi visual yang menggambarkan hubungan serta tingkatan dalam suatu sistem berdasarkan urutan atau hierarki yang ada. Diagram ini membantu dalam memahami struktur sistem, proses klasifikasi sentimen, dan keterkaitan antara komponen di dalamnya.



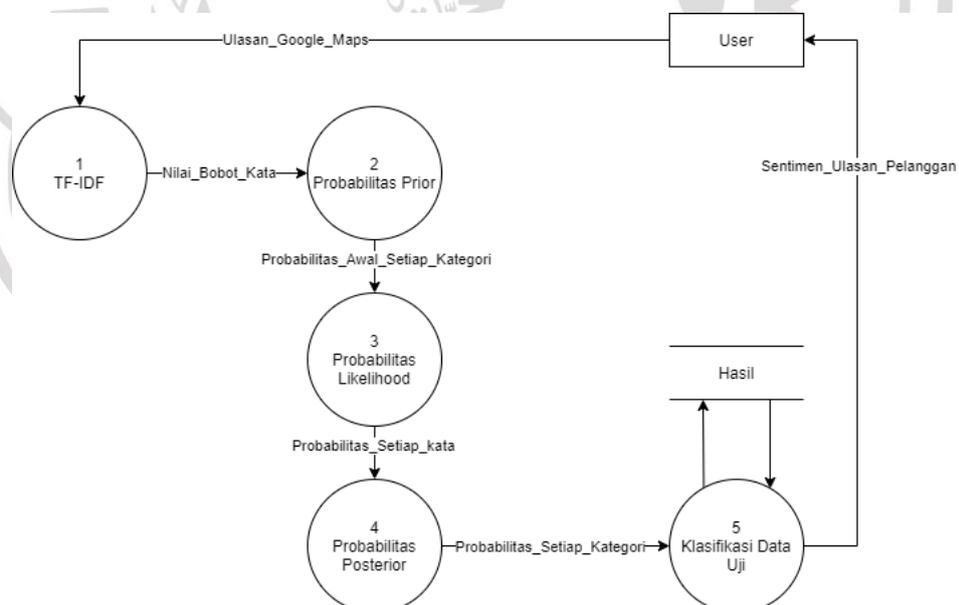
Gambar 3. 13 Diagram Jenjang Sistem

Pada gambar 3.13 menunjukkan diagram jenjang sistem yang menggambarkan tahapan proses terdiri dari:

1. Top level : Sistem ini menganalisis sentimen ulasan pelanggan PDAM Giri Tirta Gresik menggunakan *Naive Bayes* untuk mengklasifikasikan ulasan sebagai positif, netral, atau negatif.
2. Level 1 : Diagram ini menunjukkan tahapan utama dalam analisis sentimen ulasan *Google Maps* terhadap layanan PDAM Giri Tirta Gresik menggunakan *Naive Bayes*. Proses dimulai dengan perhitungan TF-IDF, lalu dilanjutkan dengan probabilitas prior, likelihood, dan posterior untuk menentukan sentimen ulasan. Hasil akhir berupa klasifikasi sentimen positif, negatif, atau netral yang ditampilkan kepada *user*.

3.4.3 Data Flow Diagram

Data Flow Diagram (DFD) menggambarkan alur proses dalam sistem analisis sentimen ulasan menggunakan metode *Naive Bayes*. Diagram ini memecah proses utama menjadi beberapa subproses untuk memberikan pemahaman lebih rinci kepada pengguna mengenai cara kerja sistem. Berikut adalah gambar dari DFD level 1 pada gambar 3.14:



Gambar 3. 14 Data Flow Diagram Level 1 Sistem

Berdasarkan pada gambar 3.14 dapat di jelaskan bahwa *DFD* Level 1 menjelaskan 5 proses utama yakni:

1. Proses 1 *TF-IDF*

Pada proses ini ulasan yang diperoleh dari *Google Maps* diproses menggunakan metode Term Frequency-Inverse Document Frequency (*TF-IDF*) untuk menentukan bobot setiap kata dalam teks. Bobot ini mencerminkan seberapa penting suatu kata dalam dokumen dibandingkan dengan keseluruhan *dataset*.

2. Proses 2 Perhitungan Probabilitas *Prior*

Pada proses ini hasil dari *TF-IDF* digunakan untuk menghitung probabilitas awal *prior* dari setiap kategori sentimen positif, netral, atau negatif berdasarkan distribusi data latih. Probabilitas ini merepresentasikan kemungkinan suatu ulasan termasuk dalam kategori tertentu sebelum mempertimbangkan kata-kata yang terkandung di dalamnya.

3. Proses 3 Perhitungan Probabilitas *Likelihood*

Pada proses ini sistem menganalisis probabilitas *likelihood*, yaitu probabilitas kemunculan setiap kata dalam masing-masing kategori sentimen. Proses ini menggunakan formula *Naive Bayes* untuk menghitung kemungkinan suatu kata muncul dalam sebuah kategori tertentu.

4. Proses 4 Perhitungan Probabilitas *Posterior*

Pada proses ini sistem mengombinasikan probabilitas *prior* dan *likelihood*, sistem menghitung probabilitas *posterior*, yang menentukan kategori sentimen yang paling mungkin untuk sebuah ulasan tertentu. Proses ini didasarkan pada prinsip *Teorema Bayes*, di mana sistem mempertimbangkan kemunculan kata-kata dalam ulasan dan distribusi kata dalam setiap kategori.

5. Proses 5 Klasifikasi Data Uji

Pada proses ini, sistem mengklasifikasikan ulasan ke dalam kategori sentimen positif, netral, atau negatif. Hasil klasifikasi ini kemudian dikembalikan kepada *user* dalam bentuk informasi sentimen dari ulasan pelanggan.

3.5 Perancangan basis data

Perancangan basis data adalah tahap penyusunan struktur penyimpanan data yang digunakan untuk mengelola dan menganalisis sentimen pelanggan dengan metode *Naive Bayes*. Sistem ini menggunakan 7 tabel utama dalam databasenya, yaitu tabel *user*, tabel *pdam*, tabel *pelabelan*, tabel *preprocessing*, tabel *latih*, tabel *uji* dan tabel *klasifikasi*. Setiap tabel memiliki fungsi krusial dalam mendukung analisis sentimen, mulai dari penyimpanan data ulasan hingga proses klasifikasi sentimen pelanggan.

3.5.1 Tabel *user*

Tabel *user* pada database sistem digunakan untuk menyimpan data *user*. Berikut merupakan tabel dari *user* dapat dilihat pada Tabel 3.13.

Tabel 3. 13 Tabel *user*

Nama	Type	Deskripsi
id_user	int	PK
username	varchar	-
password	varchar	-

3.5.2 Tabel *pdam*

Tabel *pdam* pada database sistem digunakan untuk menyimpan data ulasan *pdam* yang telah di import. Berikut merupakan tabel dari *pdam* dapat dilihat pada Tabel 3.14.

Tabel 3. 14 Tabel *pdam*

Nama	Type	Deskripsi
id_pdam	Int	PK
tgl_ulasan	varchar	-
ulasan	varchar	-
usernamee	varchar	-
sentimen	varchar	-

3.5.3 Tabel *preprocessing*

Tabel *preprocessing* pada database sistem digunakan untuk menyimpan data yang sudah melakukan *preprocessing*. Berikut merupakan tabel dari *preprocessing* dapat dilihat pada Tabel 3.15.

Tabel 3. 15 Tabel Preprocessing

Nama	Type	Deskripsi
id_preprocessing	int	PK
text	text	-
casefolding	text	-
cleaning	text	-
tokenizing	text	-
normalisasi	text	-
stopword	text	-
stemming	text	-
sentimen	text	-

3.5.4 Tabel Split Data

Tabel split data pada basis data sistem digunakan untuk menyimpan hasil pembagian data menjadi data latih dan data uji. Struktur dari tabel split data dapat dilihat pada Tabel 3.16.

Tabel 3. 16 Tabel Split Data

Nama	Type	Deskripsi
Id_split_data	int	PK
text	text	-
sentimen	text	-
tipe_data	text	-

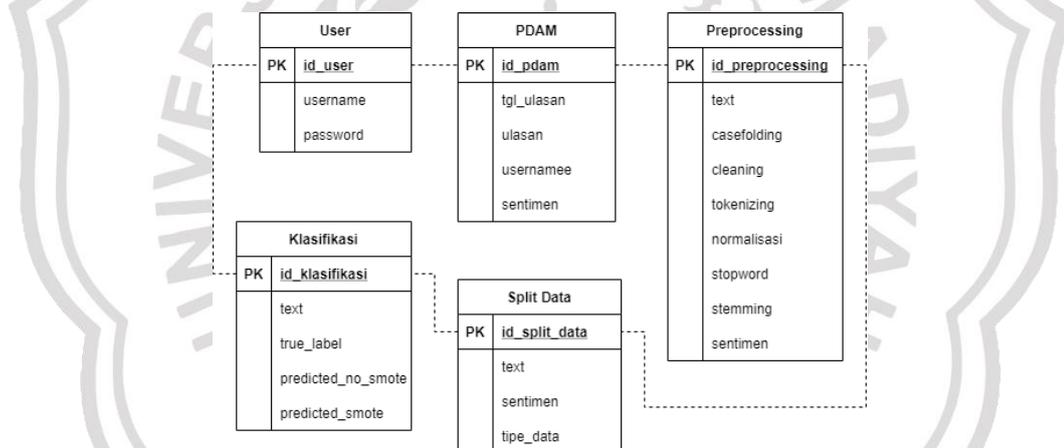
3.5.5 Tabel klasifikasi

Tabel klasifikasi pada database sistem digunakan untuk menyimpan data yang sudah melakukan proses klasifikasi sentimen *naive bayes*. Berikut merupakan tabel dari klasifikasi dapat dilihat pada Tabel 3.17.

Tabel 3. 17 Tabel Klasifikasi

Nama	Type	Deskripsi
Id_klasifikasi	int	PK
text	text	-
true_label	text	-
predicted_no_smote	text	-
predicted_smote	text	-

Perancangan Sistem ini memiliki database yang dirancang untuk menyimpan data yang diperlukan dalam analisis sentimen pelanggan. Struktur data ini disusun agar mendukung proses klasifikasi sentimen menggunakan metode *Naive Bayes*. Berikut adalah gambaran hubungan antar tabel yang divisualisasikan dalam *ERD* pada Gambar 3.15.

**Gambar 3. 15** *ERD (Entity Relationship Diagram)*

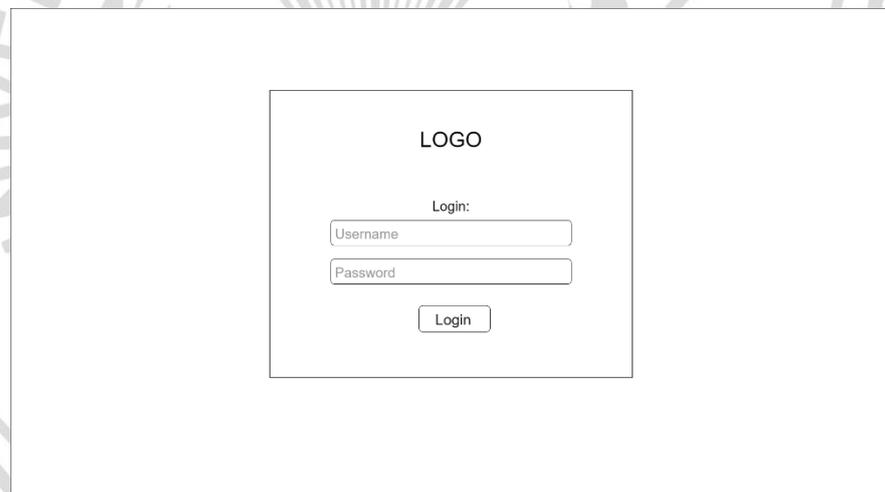
ERD di atas menunjukkan alur sistem analisis sentimen terhadap ulasan pelanggan PDAM. Data pengguna disimpan dalam tabel *User*, sedangkan ulasan dan label sentimen dicatat di tabel *PDAM*. Ulasan kemudian diproses di tabel *Preprocessing* melalui tahapan *casefolding*, *cleaning*, *tokenizing*, *normalisasi*, *stopword*, dan *stemming*. Hasilnya dibagi ke dalam tabel *Split Data* sebagai data latih dan uji. Proses klasifikasi menggunakan algoritma *Naive Bayes* dicatat di tabel *Klasifikasi*, yang menyimpan label asli dan hasil prediksi dengan maupun tanpa *SMOTE*.

3.6 Perancangan Antarmuka

Desain antarmuka berfungsi sebagai media interaksi antara sistem dan pengguna, menyajikan informasi dalam bentuk tampilan yang memuat data yang diperlukan. Pada penelitian ini, antarmuka dirancang untuk memfasilitasi input data ulasan pelanggan terkait layanan PDAM serta menampilkan hasil analisis sentimen dari ulasan tersebut.

a. Halaman Login

Halaman login adalah tampilan pertama yang harus diakses oleh user sebelum masuk ke dalam sistem. User diwajibkan memasukkan username dan password yang benar, kemudian menekan tombol "Login" untuk memperoleh akses ke sistem. Setelah berhasil login, user dapat menggunakan seluruh fitur yang tersedia, termasuk mengunggah data, melakukan analisis sentimen, melihat hasil visualisasi, serta mengevaluasi kinerja model secara menyeluruh.



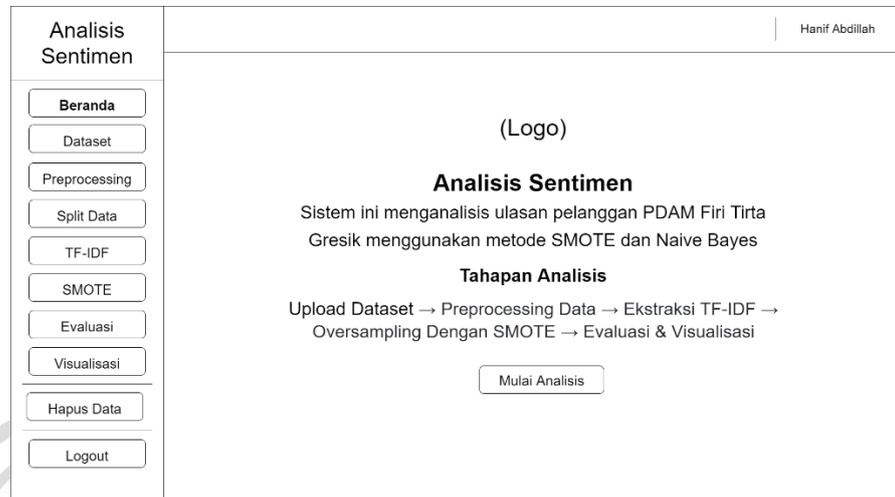
The image shows a wireframe of a login page. It consists of a large outer rectangle containing a smaller inner rectangle. At the top center of the inner rectangle is the word "LOGO". Below it is the text "Login:". Underneath "Login:" are two horizontal input fields. The first is labeled "Username" and the second is labeled "Password". At the bottom center of the inner rectangle is a button labeled "Login".

Gambar 3. 16 Perancangan Antarmuka Halaman Login

b. Halaman Beranda

Halaman beranda menampilkan informasi sistem analisis sentimen PDAM Giri Tirta Gresik dengan metode SMOTE dan Naive Bayes serta tahapan analisis. Halaman ini juga menyediakan tombol untuk memulai proses analisis. Selain itu, ditampilkan pula deskripsi singkat tujuan sistem,

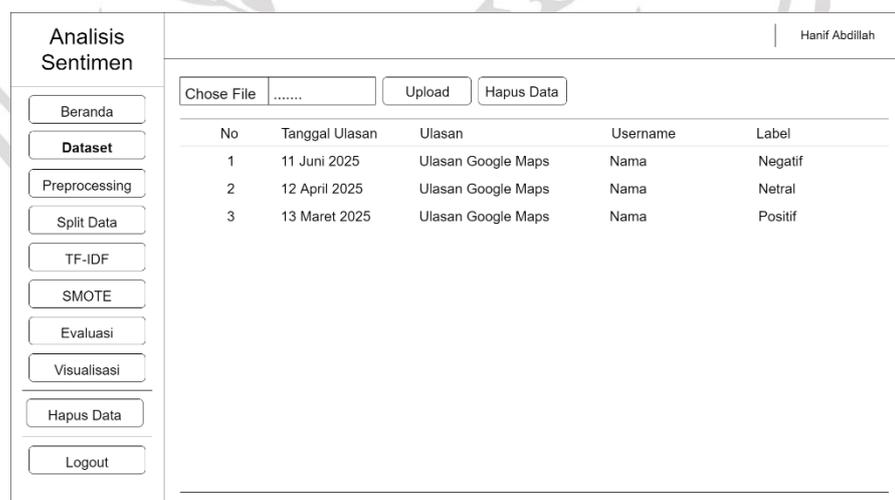
visualisasi alur kerja analisis, serta akses menuju fitur-fitur utama seperti unggah data, hasil klasifikasi, dan evaluasi model.



Gambar 3. 17 Prancangan Antarmuka Halaman Beranda

c. Halaman Dataset

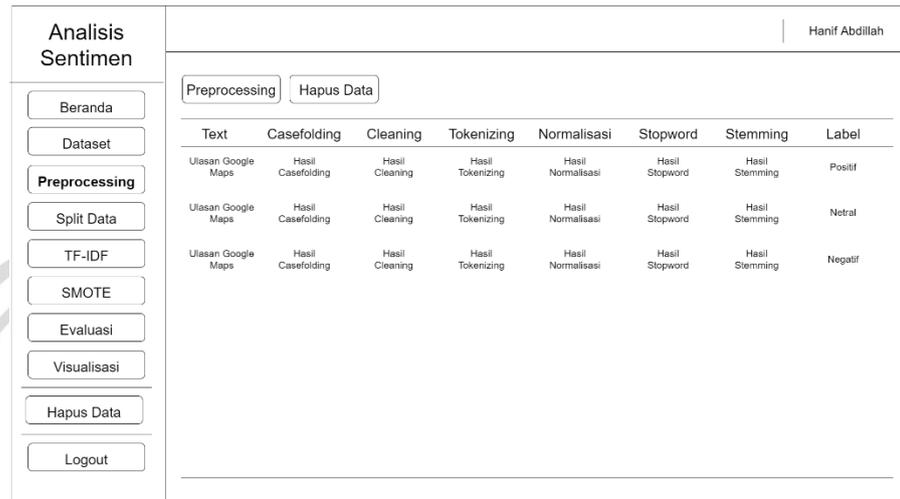
Halaman dataset digunakan untuk mengimpor dan mengelola data ulasan yang diambil dari Google Maps. Terdapat fitur untuk memilih file, mengunggah dataset, serta menghapus data yang sudah dimasukkan. Data yang ditampilkan mencakup nomor, tanggal ulasan, dan isi ulasan. Selain itu, pengguna dapat melihat pratinjau data secara langsung untuk memastikan kesesuaian sebelum analisis dilakukan.



Gambar 3. 18 Perancangan Antarmuka Halaman Dataset

d. Halaman *Preprocessing*

Halaman preprocessing berfungsi untuk membersihkan dan mempersiapkan data ulasan secara otomatis melalui proses casefolding, cleaning, tokenizing, normalisasi, stopwords removal, dan stemming agar data siap digunakan dalam analisis sentimen.

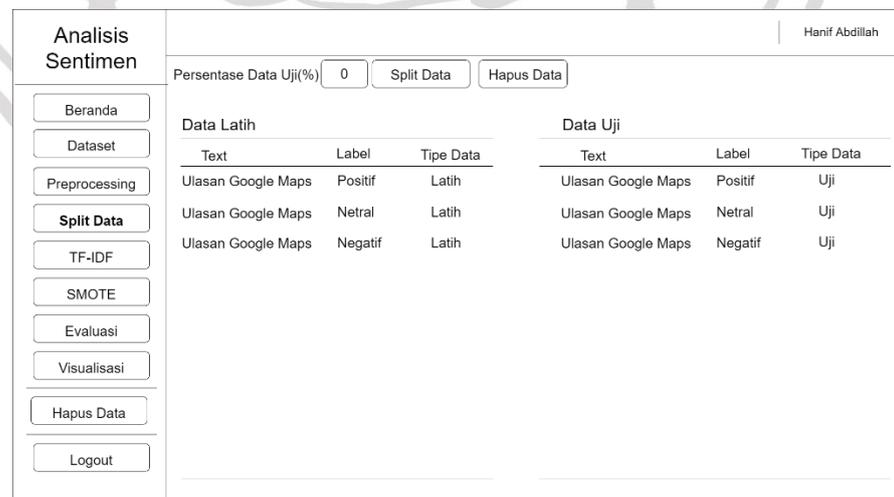


Analisis Sentimen		Hanif Abdillah						
Beranda Dataset Preprocessing Split Data TF-IDF SMOTE Evaluasi Visualisasi Hapus Data Logout		Preprocessing Hapus Data						
Text	Casefolding	Cleaning	Tokenizing	Normalisasi	Stopword	Stemming	Label	
Ulasan Google Maps	Hasil Casefolding	Hasil Cleaning	Hasil Tokenizing	Hasil Normalisasi	Hasil Stopword	Hasil Stemming	Positif	
Ulasan Google Maps	Hasil Casefolding	Hasil Cleaning	Hasil Tokenizing	Hasil Normalisasi	Hasil Stopword	Hasil Stemming	Netral	
Ulasan Google Maps	Hasil Casefolding	Hasil Cleaning	Hasil Tokenizing	Hasil Normalisasi	Hasil Stopword	Hasil Stemming	Negatif	

Gambar 3. 19 Perancangan Antarmuka Halaman *Preprocessing*

e. Halaman Split Data

Halaman Split Data digunakan untuk membagi data ulasan menjadi data latih dan data uji berdasarkan persentase tertentu. Data ditampilkan beserta label sentimen dan tipe data. Halaman ini juga menyediakan tombol untuk melakukan split dan menghapus data jika diperlukan.



Analisis Sentimen		Hanif Abdillah			
Beranda Dataset Preprocessing Split Data TF-IDF SMOTE Evaluasi Visualisasi Hapus Data Logout		Persentase Data Uji(%) <input type="text" value="0"/> Split Data Hapus Data			
Data Latih			Data Uji		
Text	Label	Tipe Data	Text	Label	Tipe Data
Ulasan Google Maps	Positif	Latih	Ulasan Google Maps	Positif	Uji
Ulasan Google Maps	Netral	Latih	Ulasan Google Maps	Netral	Uji
Ulasan Google Maps	Negatif	Latih	Ulasan Google Maps	Negatif	Uji

Gambar 3. 20 Perancangan Antarmuka Halaman Split Data

f. Halaman *TF-IDF*

Halaman TF-IDF menampilkan hasil transformasi teks ke dalam bentuk vektor menggunakan metode TF-IDF. Tabel menunjukkan bobot setiap kata (fitur) pada masing-masing dokumen, yang akan digunakan sebagai input untuk model klasifikasi. Pengguna dapat memproses atau menghapus data TF-IDF melalui tombol yang disediakan.

Analisis Sentimen		Hanif Abdillah																							
Beranda Dataset Preprocessing Split Data TF-IDF SMOTE Evaluasi Visualisasi Hapus Data Logout		Proses TF-IDF Hapus Data	<table border="1"> <thead> <tr> <th>No</th> <th>abcde</th> <th>abcd</th> <th>abc</th> <th>ab</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0.0</td> <td>0.0</td> <td>0.2</td> <td>0</td> </tr> <tr> <td>2</td> <td>0.0</td> <td>0.5</td> <td>0.3</td> <td>0.6</td> </tr> <tr> <td>3</td> <td>0.0</td> <td>0.7</td> <td>0.4</td> <td>0</td> </tr> </tbody> </table>			No	abcde	abcd	abc	ab	1	0.0	0.0	0.2	0	2	0.0	0.5	0.3	0.6	3	0.0	0.7	0.4	0
No	abcde	abcd	abc	ab																					
1	0.0	0.0	0.2	0																					
2	0.0	0.5	0.3	0.6																					
3	0.0	0.7	0.4	0																					

Gambar 3. 21 Perancangan Antarmuka Halaman *TF-IDF*

g. Halaman *SMOTE*

Halaman SMOTE menunjukkan distribusi data sentimen sebelum dan sesudah dilakukan oversampling menggunakan metode SMOTE untuk menyeimbangkan jumlah data pada tiap kelas secara lebih adil dan representatif.

Analisis Sentimen		Hanif Abdillah																					
Beranda Dataset Preprocessing Split Data TF-IDF SMOTE Evaluasi Visualisasi Hapus Data Logout		Proses Smote Hapus Data	<table border="1"> <thead> <tr> <th colspan="2">Distribusi Sebelum SMOTE</th> </tr> <tr> <th>Label</th> <th>Jumlah</th> </tr> </thead> <tbody> <tr> <td>Negatif</td> <td>100</td> </tr> <tr> <td>Netral</td> <td>30</td> </tr> <tr> <td>Positif</td> <td>40</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th colspan="2">Distribusi Setelah SMOTE</th> </tr> <tr> <th>Label</th> <th>Jumlah</th> </tr> </thead> <tbody> <tr> <td>Negatif</td> <td>100</td> </tr> <tr> <td>Netral</td> <td>100</td> </tr> <tr> <td>Positif</td> <td>100</td> </tr> </tbody> </table>	Distribusi Sebelum SMOTE		Label	Jumlah	Negatif	100	Netral	30	Positif	40	Distribusi Setelah SMOTE		Label	Jumlah	Negatif	100	Netral	100	Positif	100
Distribusi Sebelum SMOTE																							
Label	Jumlah																						
Negatif	100																						
Netral	30																						
Positif	40																						
Distribusi Setelah SMOTE																							
Label	Jumlah																						
Negatif	100																						
Netral	100																						
Positif	100																						

h. Halaman Evaluasi

Halaman evaluasi menampilkan perbandingan probabilitas, label asli, dan hasil prediksi antara model tanpa *SMOTE* dan dengan *SMOTE*. Evaluasi dilakukan menggunakan akurasi dan confusion matrix.

Text	Label Asli	Prediksi Tanpa Smote	Prediksi Dengan Smote
Ulasan Google Maps	Negatif	Netral	Netral
Ulasan Google Maps	Netral	Netral	Negatif
Ulasan Google Maps	Positif	Negatif	Positif

Gambar 3. 22 Perancangan Antarmuka Halaman Evaluasi

i. Halaman Visualisasi

Halaman visualisasi data menampilkan grafik perbandingan sentimen dan word cloud untuk memvisualisasikan hasil analisis sentimen. Visualisasi ini membantu pengguna memahami distribusi sentimen dan kata-kata yang paling sering muncul dalam ulasan pelanggan.

Gambar 3. 23 Peancangan Antarmuka Halaman Visualisasi

3.7 Perancangan pengujian system

Pengujian sistem dirancang untuk memastikan setiap komponen bekerja sebagaimana mestinya dan mendeteksi potensi kesalahan sebelum sistem diimplementasikan. Proses ini berfokus pada evaluasi input, pemrosesan, dan output dari setiap fitur utama sesuai dengan kebutuhan pengguna. Berikut untuk tabel pengujian sistem:

Tabel 3. 18 Perancangan Pengujian Sistem

No	Pengujian	Hasil yang Diharapkan	Status
1	Pengguna memasukkan username dan password	Jika <i>username</i> dan <i>password</i> benar, sistem mengarahkan ke halaman utama. Jika <i>username</i> dan <i>password</i> kosong tidak dapat login	
2	Pengguna mengunggah file dataset ulasan	Dataset berhasil ditampilkan dalam tabel.	
3	Pengguna menjalankan proses <i>preprocessing</i>	Data berhasil di proses dan ditampilkan di dalam tabel	
4	Pengguna menjalankan proses split data	Sistem berhasil membagi data sesuai input dan di tampilkan dalam tabel	
5	Pengguna menjalankan proses perhitungan <i>TF-IDF</i>	Sistem menampilkan hasil perhitungan <i>TF-IDF</i> untuk setiap dokumen	
6	Pengguna menjalankan proses <i>SMOTE</i>	Sistem menampilkan hasil data sebelum <i>SMOTE</i> dan sesudah <i>SMOTE</i>	
7	Pengguna menjalankan proses klasifikasi <i>naive bayes</i>	Sistem menampilkan hasil perhitungan probabilitas <i>Naive Bayes</i> sebelum dan sesudah penerapan <i>SMOTE</i> ,	

No	Pengujian	Hasil yang Diharapkan	Status
		serta menyajikan hasil klasifikasi, confusion matrix, dan perbandingan nilai akurasi antar model	
8	Menampilkan visualisasi grafik hasil klasifikasi	Sistem menampilkan grafik distribusi perbandingan sentimen sebelum dan sesudah <i>SMOTE</i>	
9	Pengguna melakukan <i>logout</i>	Pengguna berhasil melakukan <i>logout</i> pada sistem	

3.8 Spesifikasi Pengembangan system

Spesifikasi pengembangan sistem mencakup perangkat keras (*hardware*) dan perangkat lunak (*software*) yang digunakan dalam proses pembangunan serta pengujian sistem analisis sentimen pelanggan Perumda Giri Tirta Gresik.

a. Perangkat Keras

Pengembangan sistem ini dilakukan menggunakan Laptop dengan spesifikasi sebagai berikut:

- Prosesor: *AMD Ryzen 7 5700U with Radeon Graphics*
- Ram: 8 GB
- Penyimpanan: *SSD 512 GB*
- Layar: 15.6 inch *FHD (1920x1080)*
- Mouse: *Logitech M171*

b. Perangkat Lunak

Perangkat lunak yang digunakan dalam pengembangan sistem ini meliputi:

- Sistem Operasi: *Windows 11*
- Bahasa Pemrograman: *Python*
- Database: *MySQL*

3.9 Skenario Pengujian Dan Evaluasi Model

Dalam rangka mengevaluasi sistem analisis yang dikembangkan, dilakukan pengujian untuk menentukan tingkat akurasi perhitungan dengan metode *Naive Bayes Classifier*. Pengujian ini menggunakan skenario Confusion Matrix dalam analisis sentimen guna memperoleh nilai akurasi. Dalam pelaksanaannya, dataset yang berisi 571 ulasan dibagi dengan perbandingan 80:20, 70:30 dan 60:40. Rincian pembagian dataset tersebut disajikan pada Tabel 3.19.

Tabel 3. 19 Tabel Pembagian Dataset

Rasio Pembagian % Data Latih : Data Uji	Jumlah Data Latih	Jumlah Data Uji
80:20	456	115
70:30	399	172
60:40	342	229

Evaluasi model dilakukan untuk menilai kinerja algoritma *Naïve Bayes* dalam menganalisis sentimen ulasan pelanggan terhadap layanan Perumda Giri Tirta Gresik di *Google Maps*. Model dikembangkan menggunakan data latih yang telah diproses dan diseimbangkan dengan teknik *SMOTE* untuk mengatasi ketidakseimbangan kelas dalam data. Pengujian dilakukan menggunakan data uji yang telah dipisahkan sebelumnya, dan hasil prediksi model dibandingkan dengan label aktual untuk menghasilkan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Metrik-metrik ini digunakan untuk mengetahui sejauh mana model mampu mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral secara akurat. Hasil evaluasi disajikan dalam bentuk confusion matrix yang menggambarkan distribusi prediksi benar dan salah dari masing-masing kelas. Melalui evaluasi ini, efektivitas algoritma *Naive Bayes* dalam melakukan klasifikasi sentimen pada data yang telah diseimbangkan dengan *SMOTE* dapat dianalisis secara menyeluruh.