ANALISIS ULASAN PELANGGAN PADA SEWA KAMERA DI JEMBERKAMERA MENGGUNAKAN ALGORITME NAIVE BAYES

SKRIPSI

Oleh

Moh. Bachtiar Rifki Habibi

E41220474

PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI JEMBER
2026

ANALISIS ULASAN PELANGGAN PADA SEWA KAMERA DI JEMBERKAMERA MENGGUNAKAN ALGORITME NAIVE BAYES

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains Terapan Komputer (S. Tr. Kom)
di Program Studi Teknik Informatika
Jurusan Teknologi Informasi

Oleh

Moh. Bachtiar Rifki Habibi

E41220474

PROGRAM STUDI TEKNIK INFORMATIKA

JURUSAN TEKNOLOGI INFORMASI

POLITEKNIK NEGERI JEMBER

2026

KEMENTERIAN PENDIDIKAN TINGGI, SAINS DAN TEKNOLOGI
POLITEKNIK NEGERI JEMBER
JURUSAN TEKNOLOGI INFORMASI

ANALISIS ULASAN PELANGGAN PADA SEWA KAMERA DI JEMBERKAMERA MENGGUNAKAN ALGORITME NAIVE BAYES

Moh. Bachtiar Rifki Habibi (E41220474)
Telah diuji tanggal 24 Juni 2026
Dan dinyatakan memenuhi syarat

Ketua Penguji,

Ratih Ayuninghemi, S.ST, M.Kom
NIP. 19860802 201504 2 002

Sekretaris Penguji,

Choirul Huda, S.Kom., M.Kom
NIP. 199212272022031007

Anggota Penguji,

Reza Putra Pratama, S.Kom., M.T.
NIP. 199703102024061001

Dosen Pembimbing,

Choirul Huda, S.Kom., M.Kom
NIP. 199212272022031007

Mengesahkan,
Ketua Jurusan Teknologi Informasi

Hendra Yufit Riskiawan, S.Kom, M.Cs
NIP. 19830203 200604 1 003

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

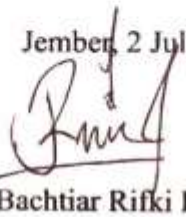
Nama: Moh. Bachtiar Rifki Habibi

NIM: E41220474

Menyatakan dengan sebenar-benarnya bahwa seluruh pernyataan dalam Laporan Akhir saya yang berjudul "Analisis Ulasan Pelanggan Pada Sewa Camera di Jemberkamera Menggunakan Algoritme Naive Bayes" merupakan gagasan dan hasil karya saya sendiri dengan arahan dosen pembimbing, serta belum pernah diajukan dalam bentuk apa pun pada perguruan tinggi mana pun.

Seluruh data dan informasi yang digunakan telah dinyatakan secara jelas dan dapat diperiksa kebenarannya. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan oleh penulis lain telah disebutkan dalam naskah dan dicantumkan dalam daftar pustaka di bagian akhir laporan skripsi ini.

Jember, 2 Juli 2026



Moh. Bachtiar Rifki Habibi

E41220474

PERNYATAAN
PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertandatangan di bawah ini,saya:

Nama	: Moh. Bachtiar Rifki Habibi
NIM	: E41220474
Program Studi	: Teknik Informatika
Jurusan	: Teknologi Informasi

Demi pengembangan Ilmu Pengetahuan, saya menyetujui untuk memberikan kepada UPT.Perpustakaan Politeknik Negeri Jember, Hak Bebas Royalti NonEksklusif (Non-Exclusive Royalty Free Right) atas **Karya Ilmiah berupa Laporan Skripsi saya yang berjudul:**

ANALISIS ULASAN PELANGGAN PADA SEWA KAMERA DI JEMBERKAMERA MENGGUNAAN ALGORITME NAIVE BAYES

Dengan Hak Bebas Royalti Non-Eksklusifini UPT. Perpustakaan Politeknik Negeri Jember berhak menyimpan, mengalih media atau format, mengelola dalam bentuk Pangkalan Data (Database), mendistribusikan karya dan menampilkan atau mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis atau pencipta.

Saya bersedia untuk menanggung secara pribadi tanpa melibatkan pihak Politeknik Negeri Jember, Segala bentuk tuntutan hukum yang timbul atas Pelanggaran Hak Cipta dalam Karya ilmiah ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jember
PadaTanggal : 2 Juli 2026
Yang menyatakan,

Nama : Moh. Bachtiar Rifki Habibi
NIM : E41220474

MOTTO

"Yang penting berusaha, hasilnya serahkan kepada Allah."

(KH. Muhammad Zuhri Zaini)

"Bukan bisa atau tidak, tapi mau atau tidak."

(Deddy Corbuzier)

HALAMAN PERSEMBAHAN

Puji syukur penulis panjatkan ke hadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul **"Analisis Ulasan Pelanggan Pada Sewa Kamera di Jemberkamera Menggunakan Algoritme Naive Bayes"**. Dengan penuh rasa syukur, penulis mempersembahkan karya ini kepada:

1. Kedua orang tua tercinta yang selalu memberikan doa, dukungan, dan semangat tanpa henti sehingga penulis dapat menyelesaikan skripsi ini.
2. Dosen pembimbing Bapak Choirul Huda, S.Kom., M.Kom. yang telah memberikan arahan, bimbingan, serta motivasi selama proses penyusunan skripsi.
3. Seluruh dosen dan staf Program Studi Teknik Informatika Politeknik Negeri Jember yang telah memberikan ilmu dan pengalaman berharga.
4. Teman-teman Teknik Informatika angkatan 2022 khususnya, terima kasih kepada Sofyan Pria Achmadi, Setia Nanda Pangestu, dan Ichie Revans Ali Fikri Ramadhani yang selalu hadir sejak awal perkuliahan, saling peduli, saling menguatkan, saling mengingatkan, serta berbagi suka dan duka dalam setiap perjalanan hingga skripsi ini terselesaikan.
5. Terima kasih kepada grup musik Payung Teduh yang lagu-lagunya telah menjadi teman di tengah sunyi dengan alunan melodi gitar sepanjang proses penyusunan skripsi ini.
6. Diri sendiri, Moh. Bachtiar Rifki Habibi telah berjuang dan bertahan dalam menyelesaikan proses perkuliahan hingga penyusunan skripsi ini. Penulis berharap bermanfaat untuk sekitar dan alam semesta.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan, oleh karena itu diharapkan kritik dan saran yang membangun. Semoga skripsi ini dapat memberikan manfaat bagi pembaca.

Analisis Ulasan Pelanggan Pada Sewa Kamera di Jemberkamera Menggunakan Algoritme Naive Bayes *(Analysis of Customer Reviews on Camera Rentals at Jemberkamera Using Naive Bayes Algorithm)*
Choirul Huda, S.Kom., M.Kom. *as chief counselor*

Moh. Bachtiar Rifki Habibi
Study Program of Informatics Engineering
Majoring of Information Technology
Program Studi Teknik Informatika
Jurusan Teknologi Informasi

ABSTRACT

Jemberkamera utilizes Google Maps for customer interaction, but the large volume of informal and subjective reviews complicates manual service evaluations. This study develops a web-based system to classify review sentiments into positive, neutral, and negative categories using the Multinomial Naive Bayes algorithm. To handle the extreme data imbalance in the negative class, the Synthetic Minority Over-sampling Technique (SMOTE) is applied. A dataset of 2,492 reviews was extracted via the Apify platform, then processed through text preprocessing and Term Frequency-Inverse Document Frequency (TF-IDF) weighting. Testing results show that a 50:50 data split scenario combined with SMOTE achieved the most optimal performance, yielding the highest accuracy of 70.47%. The integration of SMOTE effectively improved the model's sensitivity in recognizing minority negative sentiments and reduced majority class bias. The application's visual dashboard helps Jemberkamera management to evaluate business services and objectively monitor customer satisfaction in order to improve quality based on digital opinion data.

Keywords*: Sentiment Analysis, Naive Bayes, SMOTE, Google Maps, Jemberkamera, TF-IDF*

Analisis Ulasan Pelanggan Pada Sewa Kamera di Jemberkamera Menggunakan Algoritme Naive Bayes
Choirul Huda, S.Kom., M.Kom. sebagai dosen pembimbing

Moh. Bachtiar Rifki Habibi
Program Studi Teknik Informatika
Jurusan Teknologi Informasi

ABSTRAK

Jemberkamera menggunakan *Google Maps* untuk berinteraksi dengan pelanggan, tetapi besarnya volume ulasan informal dan subjektif menyulitkan evaluasi pelayanan secara manual. Penelitian ini bertujuan membangun sistem berbasis *web* untuk mengklasifikasikan sentimen ulasan pelanggan ke dalam kategori positif, netral, dan negatif menggunakan algoritma Multinomial Naive Bayes. *Metode Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan untuk mengatasi masalah ketidakseimbangan data latih. Dataset sebanyak 2.492 ulasan murni diperoleh melalui teknik *web scraping* menggunakan platform Apify. Data teks tersebut diproses melalui tahapan *preprocessing* (pembersihan, penyeragaman huruf, normalisasi kata baku, pemotongan kata, penghapusan kata hubung, dan pencarian kata dasar) serta pembobotan kata *Term Frequency-Inverse Document Frequency* (TF-IDF). Pengujian performa model dilakukan menggunakan beberapa skenario pembagian data *training* dan *testing*. Hasil eksperimen menunjukkan bahwa skenario pembagian data 50% *training* dan 50% *testing* dengan penerapan SMOTE menghasilkan performa paling optimal dengan akurasi tertinggi sebesar 70,47%. Integrasi SMOTE terbukti efektif meningkatkan sensitivitas model dalam mengenali sentimen minoritas negatif serta mengurangi bias kelas mayoritas. Sehingga, pengelola Jemberkamera bisa mengetahui kepuasan pelanggan dan evaluasi secara objektif berbasis data opini digital.

Kata Kunci: Analisis Sentimen, Naive Bayes, SMOTE, *Google Maps*, Jemberkamera, TF-IDF

RINGKASAN

Analisis Ulasan Pelanggan Pada Sewa Kamera di Jemberkamera Menggunakan Algoritme Naive Bayes, Moh. Bachtiar Rifki Habibi, NIM E41220474, Tahun 2026, Teknologi Informasi, Politeknik Negeri Jember, Choirul Huda, S.Kom., M.Kom. (Dosen Pembimbing).

Penelitian ini dilatar belakangi oleh sulitnya mengevaluasi secara manual ribuan ulasan pelanggan Jemberkamera *di Google Maps* yang bersifat subjektif dan informal. Tujuannya adalah membangun aplikasi berbasis web menggunakan algoritma *Naive Bayes* untuk mengklasifikasikan sentimen ulasan menjadi positif, netral, dan negatif sebagai media evaluasi pelayanan.

Dataset dikumpulkan melalui teknik web scraping dengan platform Apify, menghasilkan 3.933 ulasan mentah yang kemudian disaring menjadi 2.492 ulasan bersih. Pada tahap *preprocessing*, data teks diproses melalui *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal* Sastrawi, dan *stemming*. Fitur kata kemudian diekstraksi menggunakan pembobotan TF-IDF. Untuk mengatasi ketidakseimbangan data latih pada kelas minoritas negatif hanya 58 data, metode SMOTE (*Synthetic Minority Over-sampling Technique*) diterapkan sebelum proses pemodelan.

Hasil Pengujian menggunakan *confusion matrix* dengan variasi rasio data latih dan uji 80:20, 70:30, 60:40, dan 50:50. Hasil eksperimen menunjukkan skenario 50:50 dengan metode SMOTE menghasilkan performa paling optimal dan stabil tanpa *overfitting*, mencatatkan akurasi tertinggi sebesar 70,47%, presisi 56,99%, *recall* 57,05%, dan *F1-score* 55,32%. Hasil analisis menunjukkan bahwa mayoritas ulasan *Google Maps* Jemberkamera memiliki sentimen positif. Selain itu, aplikasi berbasis *web* yang dikembangkan telah berhasil melalui pengujian *Black Box Testing*, sehingga seluruh fungsionalitas sistem menjadi alat bantu pemantauan kepuasan pelanggan Jemberkamera secara objektif.

PRAKATA

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penulisan skripsi berjudul "Analisis Ulasan Pelanggan Pada Sewa Kamera di Jemberkamera Menggunakan Algoritm Naive Bayes" dapat diselesaikan dengan baik.

Tulisan ini adalah laporan hasil penelitian yang dilaksanakan mulai bulan maret 2025 bertempat di Politeknik Negeri Jember, sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains Terapan Komputer (S.Tr.Kom) di Program Studi Teknik Informatika Jurusan Teknologi Informasi.

Penulis menyampaikan penghargaan dan ucapan terima kasih yang sebesar besarnya kepada:

1. Bapak Saiful Anwar, S.Tp, M.P sebagai Direktur Politeknik Negeri Jember.
2. Bapak Hendra Yufit Riskiawan, S.Kom,M.Cs sebagai Ketua Jurusan Teknologi Informasi.
3. Ibu Bety Etikasari, S.Pd., M.Pd. selaku Ketua Program Studi Teknik Informatika.
4. Bapak Choirul Huda, S.Kom., M.Kom. sebagai Dosen Pembimbing.
5. Ibu Ratih Ayuninghemi, S.ST, M.Kom. sebagai Ketua Penguji.
6. Bapak Reza Putra Pradana, S.Kom., M.T. sebagai Anggota Penguji.
7. Bapak Anggik Budi Prasetiyo, S.Pd., M.Li., Gr. Sebagai Pakar Bahasa.
8. Orang Tua penulis dan keluarga yang selalu memberikan dukungan dan doa.

Skripsi ini masih kurang sempurna, mengharapkan kritik dan saran yang sifatnya membangun guna perbaikan di masa mendatang. Semoga tulisan ini bermanfaat.

Jember, 2 Juli 2026

Moh. Bachtiar Rifki Habibi

E41220474

DAFTAR ISI

Halaman

HALAMAN JUDUL ..... **ii**
HALAMAN PENGESAHAN ..... **iii**
SURAT PERNYATAAN ..... **iv**
SURAT PERNYATAAN PUBLIKASI ..... **v**
MOTTO ..... **vi**
HALAMAN PERSEMBAHAN ..... **vii**
ABSTRACT ..... **viii**
ABSTRAK ..... **ix**
RINGKASAN ..... **x**
PRAKATA ..... **xi**
DAFTAR ISI ..... **xii**
DAFTAR GAMBAR ..... **xv**
DAFTAR TABEL ..... **xvii**
DAFTAR KODE PROGRAM ..... **xviii**
DAFTAR LAMPIRAN ..... **xix**

BAB 1. PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 7
1.3 Tujuan Penelitian 7
1.4 Manfaat Penelitian 7
1.5 Batasan Masalah 8

BAB 2 . TINJAUAN PUSTAKA **9**
2.1 Ulasan Pelanggan 9
2.2 Google Maps 9
2.3 Jemberkamera 10
2.4 Web Scraping (Apify Google Maps Reviews Scraper) 10
2.5 Preprocessing Kata 10

2.6 Pembobotan Kata 11
2.7 Sastrawi 12
2.8 Naive Bayes 13
2.9 Confusion Matrix 15
2.10 State of The Art 17

BAB 3 . METODOLOGI PENELITIAN 22
3.1 Tempat dan Waktu Penelitian 22
3.2 Alat dan Bahan 22
3.2.1 Alat 22
3.2.2 Bahan 23
3.3 Tahapan Penelitian 23
3.3.1 Identifikasi Masalah 23
3.3.2 Analisis Kebutuhan 24
3.3.3 Pengumpulan Data 26
3.3.4 Perancangan Sistem 26
3.3.5 Implementasi dan Pengujian 32
3.3.6 Hasil dan Pembahasan 33
3.3.7 Kesimpulan dan Saran 34
3.4 Jadwal Pelaksanaan 34

BAB 4 . HASIL DAN PEMBAHASAN 35
4.1 Identifikasi Masalah 35
4.1.1 Wawancara Penelitian Dengan Mitra 35
4.2 Analisis Kebutuhan 37
4.3 Pengumpulan Data 37
4.4 Prepocessing Data 39
4.4.1 *Cleaning data* (Pembersihan Data) 40
4.4.2 *Case Folding* 41
4.4.3 Normalisasi 42
4.4.4 *Tokenizing* 43
4.4.5 *Stopword Removal* 44

4.4.6 *Stemming*.....45
4.5 Pembobotan Kata (TF-IDF).....47
4.5.1 Perhitungan TF (*Term Frequency*).....47
4.5.2 Perhitungan *Inverse Document Frequency* (IDF).....50
4.5.3 Perhitungan *Term Frequency-Inverse Document Frequency* (TF-IDF).....52
4.6 Implementasi dan Pengujian.....53
4.6.1 Split Data.....53
4.6.2 Klasifikasi Naive Bayes.....54
4.6.3 Evaluasi Model.....57
4.6.4 Visualisasi Hasil.....63
4.6.5 Antarmuka Sistem.....74
4.6.6 *Black Box Testing*.....81
4.6.7 *User Testing*.....82
4.7 Analisis Hasil.....82

BAB 5 . KESIMPULAN DAN SARAN.....84
5.1 Kesimpulan.....84
5.2 Saran.....85

DAFTAR PUSTAKA.....86

LAMPIRAN.....89

DAFTAR GAMBAR

Halaman

Gambar 1. 1 Data Statisik Tujuan Mengakses Internet, 2024 1
Gambar 1. 2 Data Statistik Ulasan Google Maps Jemberkamera 4
Gambar 3. 1 Tahapan Penelitian 23
Gambar 3. 2 Alur Kerja Sistem 24
Gambar 3. 3 Use Case Diagram 27
Gambar 3. 4 Wireframe Halaman Login 28
Gambar 3. 5 Wireframe Halaman Dashboard 29
Gambar 3. 6 Wireframe Halaman Upload Data 29
Gambar 3. 7 Wireframe Halaman Prepocessing 30
Gambar 3. 8 Wireframe Halaman TF-IDF 30
Gambar 3. 9 Wireframe Halaman Klasifikasi 31
Gambar 3. 10 Wireframe Halaman Hasil dan Laporan 31
Gambar 4. 1 Surat Izin Observasi Mitra 36
Gambar 4. 2 Dokumentasi Observasi Bersama Mitra 37
Gambar 4. 3 Google Maps Reviews Scraper Platform Apify 38
Gambar 4. 4 Scraping Data Review Apify 39
Gambar 4. 5 Hasil Data Terkumpul Review Google Maps 39
Gambar 4. 6 Hasil Confusion Matrix Skenario 80:20 58
Gambar 4. 7 Hasil Confusion Matrix Skenario 70:30 59
Gambar 4. 8 Hasil Confusion Matrix Skenario 60:40 60
Gambar 4. 9 Hasil Confusion Matrix Skenario 50:50 61
Gambar 4. 10 Hasil Label Ulasan 64
Gambar 4. 11 Hasil Model Skenario 80:20 65
Gambar 4. 12 Hasil Model Skenario 70:30 65
Gambar 4. 13 Hasil Model Skenario 60:40 66
Gambar 4. 14 Hasil Model Skenario 50:50 67
Gambar 4. 15 Hasil Tabel Split Data Skenario 80:20 68

Gambar 4. 16 Hasil Tabel Split Data Skenario 70:30 68
Gambar 4. 17 Hasil Tabel Split Data Skenario 60:40 69
Gambar 4. 18 Hasil Tabel Split Data Skenario 50:50 69
Gambar 4. 19 Hasil Distribusi Data Training 80:20 70
Gambar 4. 20 Hasil Distribusi Data Training 70:30 71
Gambar 4. 21 Hasil Distribusi Data Training 60:40 71
Gambar 4. 22 Hasil Distribusi Data Training 50:50 72
Gambar 4. 23 Hasil Classification Report Skenario 80:20 72
Gambar 4. 24 Hasil Classification Report Skenario 70:30 73
Gambar 4. 25 Hasil Classification Report Skenario 60:40 73
Gambar 4. 26 Hasil Classification Report Skenario 50:50 74
Gambar 4. 27 Halaman Login 74
Gambar 4. 28 Halaman Dashboard 75
Gambar 4. 29 Hasil Halaman Upload Dataset Diperoleh dari Scraping 76
Gambar 4. 30 Hasil Halaman Prepocessing 77
Gambar 4. 31 Hasil Pembobotan TF-IDF 78
Gambar 4. 32 Hasil Proses Klasifikasi Naive Bayes 79
Gambar 4. *33* Hasil dan Visualisasi Evaluasi Model 80

DAFTAR TABEL

Halaman

Tabel 2. 1 State of The Art 19
Tabel 3. 1 Jadwal Pelaksanaan 34
Tabel 4. 1 Hasil Scraping Data Google Maps 40
Tabel 4. 2 Cleaning data 41
Tabel 4. 3 Case Folding 42
Tabel 4. 4 Normalisasi 43
Tabel 4. 5 Tokenizing 44
Tabel 4. 6 Stopword Removal 45
Tabel 4. 7 Stemming 46
Tabel 4. 8 Data Sampel Perhitungan TF-IDF 47
Tabel 4. 9 Hasil Frekuensi Kemunculan Setiap Kata Dalam Dokumen 48
Tabel 4. 10 Hasil Perhitungan TF 49
Tabel 4. 11 Hasil Perhitungan DF 50
Tabel 4. 12 Hasil Perhitungan IDF 51
Tabel 4. 13 Hasil Perhitungan TF-IDF 52
Tabel 4. 14 Hasil Perhitungan Likelihood 55
Tabel 4. 15 Data Uji 56
Tabel 4. 16 Hasil Rekapitulasi Pengujian 62
Tabel 4. 17 Hasil Pengujian Blackbox Testing Bersama pihak jemberkamera dan Teknisi Jurusan 81

DAFTAR KODE PROGRAM

Halaman

Kode Program 4. 1 Cleaning data 40
Kode Program 4. 2 Case Folding 41
Kode Program 4. 3 Normalisasi 42
Kode Program 4. 4 Tokenizing 44
Kode Program 4. 5 Stopword Removal 45
Kode Program 4. 6 Stemming 46
Kode Program 4. 7 Split Data 54

DAFTAR LAMPIRAN

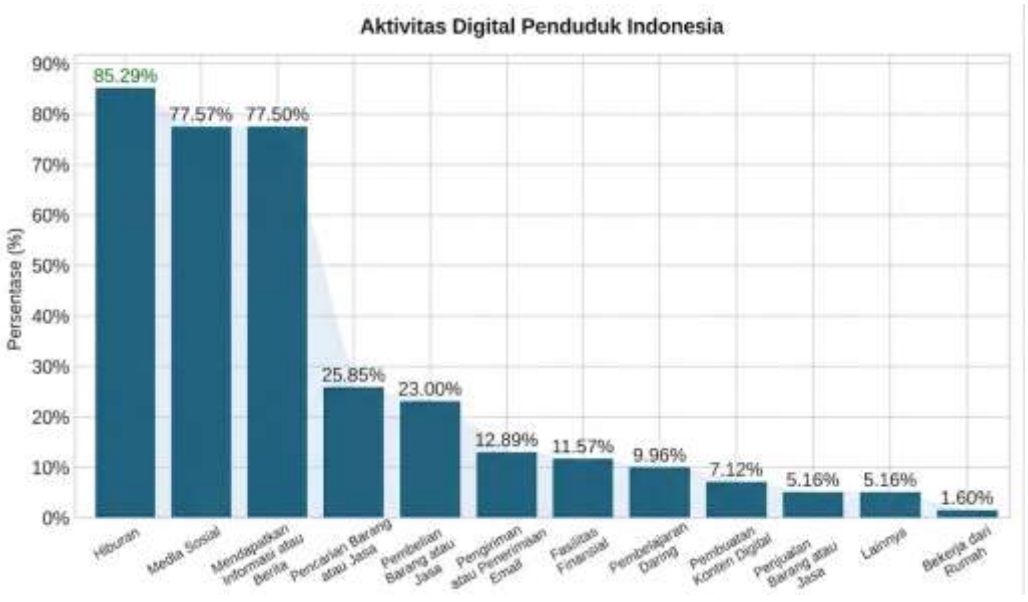
Halaman

Lampiran 1 Skenario Pengujian Website 89
Lampiran 2 Lembar Validasi Pakar Bahasa 95
Lampiran 3 Lembar Validasi 1 Pengujian Web App Oleh Mitra 96
Lampiran 4 Lembar Validasi 2 Pengujian Web App Oleh Teknisi Jurusan TIF.... 97
Lampiran 5 Detail Antarmuka Hasil Halaman Prepocessing..... 98
Lampiran 6 Detail Antarmuka Hasil Halaman TF 99
Lampiran 7 Detail Antarmuka Hasil Halaman DF-IDF 100
Lampiran 8 Detail Antarmuka Hasil Halaman TF-IDF 101
Lampiran 9 Dokumentasi Pengujian Web App Oleh Pihak Jemberkamera..... 102
Lampiran 10 Visualisasi Word Cloud Positif..... 102
Lampiran 11 Visualisasi Word Cloud Netral..... 102
Lampiran 12 Visualisasi Word Cloud Negatif 103

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Berdasarkan pengutipan data Badan Pusat Statistik (BPS) dalam laporan *Statistik Telekomunikasi Indonesia 2024*, aktivitas digital penduduk Indonesia masih didominasi oleh pemanfaatan internet. Diantaranya untuk hiburan 85,29%, media sosial 77,57%, mendapatkan informasi atau berita 77,50%, pencarian barang atau jasa 25,85% Pembelian barang atau jasa 23,00%, pengiriman atau penerimaan email 12,89%, Fasilitas Finansial 11,57%, pembelajaran daring 9,96%, pembuatan konten digital 7,12%, penjualan barang atau jasa 5,16%, lainnya 5,16% dan bekerja dari rumah 1,60%. Data tersebut menunjukkan bahwa selain dimanfaatkan sebagai sarana hiburan dan memperoleh informasi, internet juga mulai berperan dalam mendukung berbagai aktivitas ekonomi dan pendidikan, meskipun tingkat pemanfaatannya masih relatif rendah. Gambar 1.1 menampilkan data statistik penduduk yang mengakses internet berdasarkan tujuan penggunaannya pada tahun 2024.



Gambar 1. 1 Data Statisik Tujuan Mengakses Internet, 2024

Tingginya aktivitas pencarian informasi di internet khususnya terkait pencarian barang atau jasa, menunjukkan adanya perilaku pelanggan digital yang

semakin kritis dalam proses pengambilan keputusan. Pelanggan cenderung melakukan pencarian informasi, membandingkan berbagai alternatif, dan mempertimbangkan ulasan dari pengguna lain sebelum memutuskan untuk membeli atau menggunakan suatu produk maupun layanan. Oleh karena itu, respons seorang semakin selektif dalam memahami dinamika pemenuhan ekspektasi demi menjaga kepuasan.

Tingkat respons emosional individu seseorang, muncul sebuah persepsi mereka terhadap hasil kinerja yang mereka lakukan sesuai atau melebihi ekspektasi awal. Kepuasan merupakan tolok ukur yang searah luas antara hasil yang diperoleh dengan keinginan. Jika hasil diterima sesuai apa yang diharapkan ataupun lebih, ini akan menimbulkan kesenangan dan kepuasan pelanggan, namun sebaliknya, apabila yang diharapkan tidak sesuai, pelanggan akan merasa kecewa (Sasono dkk., 2023). Kondisi ini diperkuat oleh pengujian statistik lapangan yang menemukan bahwa kontribusi kualitas produk fisik bersama dengan kualitas pelayanan memiliki dampak yang sangat dominan, yakni sebesar 77,3% dalam menentukan dan menerangkan fluktuasi tingkat kepuasan yang dirasakan oleh pelanggan (Mayasari & Safina, 2021). Pentingnya kepuasan pelanggan sangat besar karena pelanggan yang merasa puas cenderung menjadi pelanggan setia mereka, lebih besar kemungkinannya untuk melakukan pembelian ulang, merekomendasikan produk, dan berkontribusi pada pertumbuhan bisnis jangka panjang. Tingkat kepuasan yang dirasakan pelanggan terhadap produk atau jasa yang mereka dapatkan dari sebuah bisnis atau institusi ini mencakup evaluasi pelanggan tentang seberapa baik suatu produk atau jasa yang bisa dipenuhi, atau bahkan melampaui keinginan, kebutuhan dan harapan mereka (Damanik dkk., 2024).

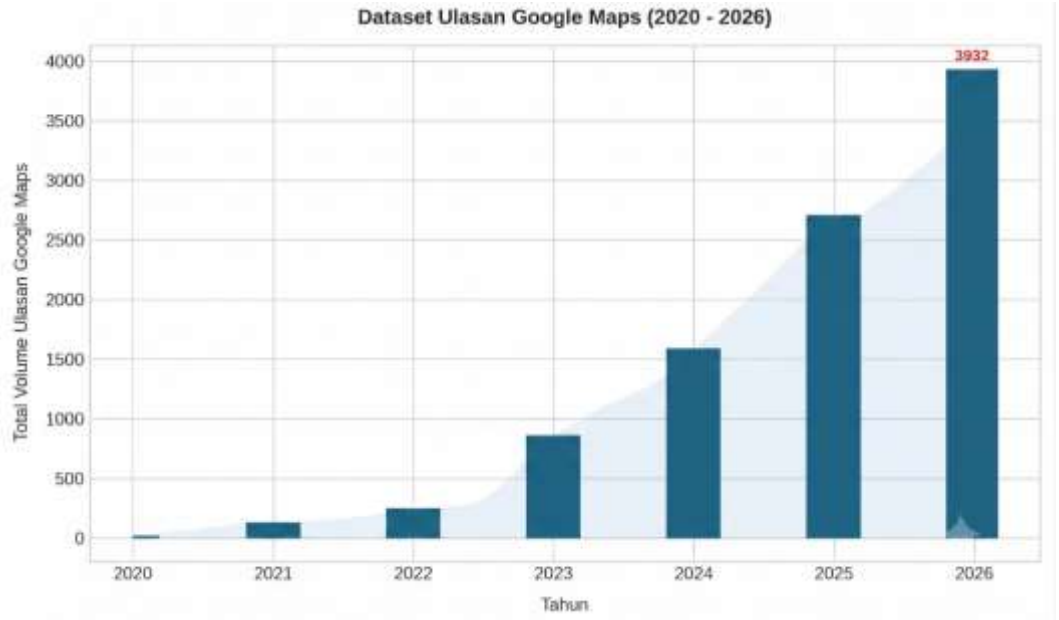
Kamera merupakan perangkat optik yang digunakan untuk menangkap cahaya dan mengubahnya menjadi gambar melalui proses tertentu. Teknologi kamera telah berkembang pesat dari waktu ke waktu, mulai dari kamera analog hingga digital. Pada era kontemporer, jenis kamera seperti DSLR (*Digital Single-Lens Reflex*) dan *Mirrorless* menjadi pilihan populer, karena mampu menghasilkan gambar dengan kualitas tinggi. Kedua jenis kamera ini banyak digunakan dalam berbagai kebutuhan, mulai dari dokumentasi pribadi hingga produksi profesional

seperti fotografi komersial dan videografi. Selain itu, keunggulan dari fitur manual yang dimiliki kamera ini memberikan kontrol lebih bagi pengguna dalam mengatur komposisi dan pencahayaan gambar (Fernanda dkk., 2023). Saat ini, kebutuhan untuk mengabadikan momen melalui fotografi telah menjadi bagian dari gaya hidup banyak orang. Tidak hanya untuk kepentingan pribadi, banyak pengguna yang secara aktif memotret momen-momen penting dalam kehidupan mereka untuk kemudian dibagikan melalui berbagai platform media sosial. Aktivitas ini tidak hanya berfungsi sebagai dokumentasi, tetapi juga sebagai sarana berbagi cerita, membangun citra diri, hingga memperluas jangkauan komunikasi visual. Oleh karena itu, kemampuan dasar dalam menggunakan kamera secara tepat menjadi keterampilan yang semakin relevan dan dibutuhkan oleh masyarakat umum (Irwanto dan Giantika, 2022).

Jemberkamera merupakan penyedia layanan rental kamera dan perlengkapan fotografi yang beroperasi di Jember. Dalam industri layanan seperti ini, ulasan dari pelanggan menjadi indikator utama dalam menilai kualitas produk dan layanan yang ditawarkan. Pelanggan cenderung membagikan pengalaman mereka melalui ulasan di berbagai platform salah satunya di *Google Maps*. *Google Maps*, sebagai platform pemetaan yang dikembangkan oleh *Google*, tidak hanya membantu pengguna dalam menemukan lokasi bisnis, tetapi juga menyediakan fitur ulasan pelanggan. Melalui *Google Maps*, pelanggan dapat memberikan penilaian berupa rating bintang dan komentar tertulis mengenai pengalaman mereka. Ulasan ini bersifat autentik karena berasal langsung dari pengguna layanan, sehingga memberikan gambaran yang jujur dan transparan tentang kualitas layanan yang diberikan (Zidny, 2024a). Saat ini Jemberkamera tidak memiliki *website* resmi sehingga penulis hanya bisa mengambil ulasan dari *Google Maps*.

Saat ini, Jemberkamera mengelola ulasan tersebut dilakukan secara manual. Hal ini menjadi tantangan besar bagi Jemberkamera karena jumlah ulasan mencapai lebih dari 4000. Sehingga, analisis secara manual menjadi tidak efisien, memakan waktu, dan rawan kesalahan interpretasi, serta menghambat respons cepat terhadap kebutuhan pelanggan. Tanpa analisis ulasan, Jemberkamera akan kehilangan wawasan berharga mengenai area perbaikan yang spesifik, tren kepuasan

pelanggan, atau bahkan potensi krisis reputasi yang bisa dihindari jika ulasan negatif terdeteksi lebih awal. Gambar 1.1 menampilkan grafik menyajikan data statistik pertumbuhan volume ulasan Jemberkamera tahun 2020 sampai 2026 yang dihasil *web scraping*.



Gambar 1. 2 Data Statistik Ulasan *Google Maps* Jemberkamera

Penelitian ini dilakukan untuk menganalisis dari ulasan pelanggan di *Google Maps*, bertujuan memahami bagaimana ulasan tersebut dapat mempengaruhi reputasi online, membangun kepercayaan dan kepuasan calon pelanggan, dan memperluas jangkauan pasar. Dengan memahami ulasan pelanggan, Jemberkamera dapat meningkatkan kualitas layanan dan strategi pemasaran mereka untuk memenuhi kebutuhan dan harapan pelanggan secara efektif. Hal ini menjadikan sebuah acuan untuk meningkatkan strategis perusahaan (Al Lutfani dkk., 2025).

Ulasan yang bersifat positif, netral atau negatif dapat menjadi masukan berharga untuk perbaikan layanan di masa mendatang. Oleh karena itu, mengelola pelanggan dengan baik menjadi strategi penting dalam mempertahankan dan meningkatkan kualitas layanan. Walaupun ulasan dari konsumen memiliki nilai yang tinggi, mengelola dan menganalisis ulasan yang jumlahnya melimpah dan bermacam-macam. Selain itu, ulasan yang diberikan konsumen bersifat subjektif

dan menggunakan bahasa informal. Sehingga, diperlukan metode yang dapat memproses dan menganalisis data ulasan dengan efektif.

Pada penelitian yang sama, beberapa metode yang digunakan termasuk dalam kategori klasifikasi dan terdiri dari empat pendekatan *Naive Bayes*, *Support Vector Machine* (SVM), *Logistic Regression*, dan *Deep Learning*. *Metode Naive Bayes* beroperasi dengan memanfaatkan *Teorema Bayes* untuk menghitung probabilitas kelas berdasarkan data historis, meskipun asumsi yang memudahkan model ini adalah bahwa setiap variabel dianggap saling independen. Prosesnya mencakup penentuan jumlah kelas, penghitungan probabilitas prior dan *likelihood* untuk setiap kelas, perkalian seluruh variabel dalam masing-masing kelas (posterior), dan perbandingan hasilnya untuk menentukan kelas paling mungkin (Heristian dkk., 2025). Sementara itu, *Support Vector Machine* (SVM) bekerja dengan mencari *hyperplane* optimal yang dapat memaksimalkan margin pemisah antara dua kelas positif dan negatif (Salsabillah dkk., 2024). *Logistic Regression* digunakan untuk memodelkan variabel dependen dikotomis, menghasilkan output biner (0 atau 1) berdasarkan satu atau lebih prediktor bebas (Budianto dkk., 2024). Sedangkan metode *Deep Learning* untuk mengekstraksi fitur dari data secara iteratif dan menyesuaikan model melalui proses pembelajaran berbasis error (Naquitasia dkk., 2022).

Oleh karena itu, dalam penelitian ini penulis memilih metode *Naive Bayes* karena beberapa alasan strategis. Metode ini menunjukkan performa klasifikasi yang efektif hanya dengan memanfaatkan probabilitas dasar tanpa membutuhkan asumsi distribusi fitur yang kompleks, sehingga sangat ideal untuk dataset dengan struktur sederhana. Meskipun mengandalkan asumsi independensi antar fitur, *Naive Bayes* tetap mampu memberikan prediksi yang andal melalui perhitungan probabilitas prior, *likelihood*, dan *posterior* secara sistematis

Salah satu cara yang sering dipakai dalam analisis ulasan ini disebabkan oleh keahliannya dalam mengategorikan teks dengan efektif. Algoritma *Naive Bayes* menentukan kemungkinan sebuah teks bersifat positif, netral atau negatif. menggunakan metode statistik yang mudah namun optimal (Gondowastradmojo dan Kusumastuti, 2024). Algoritma ini beroperasi berdasarkan prinsip probabilitas

dan asumsi independensi antara fitur, sehingga dapat mengklasifikasikan teks dengan cepat dan tepat. *Naive Bayes* telah banyak diterapkan dalam berbagai studi untuk mengklasifikasikan ulasan konsumen dengan hasil yang memuaskan, terutama pada dataset teks yang cukup sederhana dan tidak terlalu besar.

Ulasan tersebut bersifat informal dan bervariasi, sehingga metode *Naive Bayes* mampu menanganinya. Ulasan di *Google Maps* pada jemberkamera saat ini mencapai 4.000 lebih ulasan, umumnya terdiri dari kalimat pendek yang berisi pendapat langsung dari pengguna. Dengan memanfaatkan metode *Naive Bayes*, proses klasifikasi dapat dilaksanakan, sehingga memberikan gambaran menyeluruh mengenai pandangan pelanggan terhadap layanan Jemberkamera. Penelitian ini memanfaatkan data ulasan pengguna Jemberkamera yang diperoleh dari *Google Maps* sebagai sumber utama yang berupa teks komentar (Kamal dan Ratnasari, 2024). Informasi tersebut meliputi berbagai tanggapan pelanggan mengenai pengalaman mereka saat menggunakan layanan penyewaan kamera.

Teknik proses pada pengambilan data di *Google Maps* yaitu menggunakan platform web *scraping* otomatis bernama Apify dengan esktraksi data tertentu. Dalam topik ini, digunakan untuk mengumpulkan data ulasan pelanggan dari platform *Google Maps*, yang kemudian akan diproses untuk keperluan analisis sentimen atau pengelompokan informasi (Fikri dkk., 2022).

Selanjutnya, diproses dan dianalisis guna mengetahui sentimen yang terdapat di dalamnya. Dengan demikian, hasil analisis dapat memberikan gambaran yang objektif tentang kepuasan pelanggan dan elemen layanan yang harus ditingkatkan. Melalui klasifikasi sentimen, pengelola Jemberkamera dapat dengan cepat memantau dan menilai umpan balik pelanggan secara langsung. Ini akan mendukung pembuatan keputusan strategis untuk memperbaiki kualitas layanan, mengatasi kekurangan, serta meningkatkan kepuasan dan loyalitas pelanggan. Di samping itu, analisis sentimen juga dapat berfungsi sebagai alat ukur efektivitas promosi dan layanan yang telah dilaksanakan. Menyadari betapa pentingnya tinjauan konsumen dalam membangun reputasi perusahaan serta meningkatkan mutu layanan, penelitian ini bertujuan untuk mengimplementasikan metode *Naive Bayes* untuk mengklasifikasikan sentimen dari ulasan konsumen Jemberkamera.

Melalui hasil penelitian ini, diharapkan dapat memberikan sumbangan nyata untuk pengembangan sistem pengelolaan ulasan yang lebih efisien dan efektif serta mendukung Jemberkamera dalam menghadapi kompetisi di era digital. Hasil dari penelitian yang dilakukan pada analisis sentimen ini yaitu untuk mengetahui seberapa besar tingkat akurasi yang dihasilkan pada penggunaan metode algoritma dari *Naive Bayes*.

1.2 Rumusan Masalah

Bagaimana proses pengambilan data ulasan pelanggan pada *Google Maps* Jemberkamera?

a. Bagaimana implementasi *Naive Bayes* dalam klasifikasi ulasan pengguna pada *Google Maps* di Jemberkamera?
b. Bagaimana tingkat akurasi metode *Naive Bayes* dalam mengelompokkan ulasan pelanggan Jemberkamera?

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini yaitu:

a. Untuk menganalisis dan mendeskripsikan proses pengambilan data ulasan pelanggan Jemberkamera dari platform *Google Maps* menggunakan teknik web *scraping*.
b. Untuk mengelompokkan ulasan pelanggan Jemberkamera dengan pendekatan klasifikasi menggunakan metode *Naive Bayes*.
c. Untuk mengevaluasi tingkat akurasi metode *Naive Bayes* dalam mengelompokkan ulasan pelanggan Jemberkamera, sehingga dapat diketahui efektivitas metode tersebut.

1.4 Manfaat Penelitian

Adapun Manfaat penelitian ini yaitu:

a. Memberikan solusi praktis bagi Jemberkamera untuk menganalisis ulasan pengunjung menggunakan metode *Naive Bayes*, serta menjadi contoh penerapan analisis sentimen pada ulasan di *Google Maps*.

b. Menjadi referensi penerapan metode klasifikasi teks, khususnya algoritma *Naive Bayes*, dalam mengolah data ulasan pelanggan secara praktis dan aplikatif, terutama pada platform digital seperti *Google Maps*.
c. Membantu penulis memahami metode klasifikasi teks secara langsung melalui studi kasus nyata, sekaligus menambah wawasan dalam pengolahan data berbasis ulasan pelanggan.

1.5 Batasan Masalah

Untuk menjaga agar penelitian tetap fokus dan terarah sesuai dengan rumusan masalah serta tujuan yang telah ditetapkan, maka penulis membatasi ruang lingkup penelitian ini sebagai berikut:

a. Dataset ulasan (*review*) pelanggan yang dikumpulkan dibatasi pada periode tanggal 1 Januari 2020 hingga data terbaru yang berhasil diekstraksi pada saat proses pengambilan data di tahun 2026.
b. Proses klasifikasi teks dibatasi pada pengelompokan tiga kelas sentimen Positif, Netral, Negatif menggunakan algoritma Naive Bayes dan pembobotan TF-IDF.
c. Aplikasi berbasis *web* ini menggunakan sistem satu aktor (*single-actor*), hak akses dashboard terbatas hanya untuk Admin atau pengelola internal Jemberkamera tanpa antarmuka bagi pelanggan umum.

BAB 2. TINJAUAN PUSTAKA

2.1 Ulasan Pelanggan

Salah satu nilai penting bagi perusahaan akhir ini lebih mencari *feetback* ataupun informasi terkait produk melalui platfrom sosial media dan *google*. Disisi lain, *Google* juga meyediakan *Google Maps* agar *customer* bisa memberikan ulasan di *Google Maps* dan mereview positif, netral dan negatif mengenai produk mereka, hal ini bisa di akses ke seluruh orang (Riza Andrian Septian dan Sita Deliyana Firmialy, 2023). Ulasan pelanggan merupakan bentuk umpan balik yang diberikan oleh konsumen setelah mereka menggunakan suatu produk atau layanan. Ulasan ini biasanya berisi pendapat, pengalaman, maupun penilaian yang mencerminkan tingkat kepuasan atau ketidakpuasan terhadap kualitas produk, pelayanan, harga, maupun aspek lainnya yang mereka alami. Ulasan dapat disampaikan dalam bentuk teks, bintang rating, atau kombinasi keduanya, dan tersebar luas di berbagai platform digital seperti media sosial*, e-commerce*, *website* bisnis, maupun *Google Maps*. Bagi pelaku usaha, ulasan pelanggan bukan hanya sekadar komentar, tetapi merupakan sumber data penting yang merepresentasikan persepsi publik terhadap bisnis mereka (Shalihah dkk., 2024).

2.2 Google Maps

Google Maps merupakan layanan pemetaan digital gratis berbasis internet yang dikembangkan oleh *Google*. Aplikasi ini dapat diakses melalui *browser* maupun perangkat seluler, dan menyediakan berbagai banyak fitur seperti penunjuk arah, informasi lalu lintas, pencarian lokasi. Salah satu fitur unggulannya adalah kemampuan untuk menampilkan ulasan dan penilaian dari pengguna terhadap berbagai tempat seperti restoran, toko, tempat wisata, dan layanan publik lainnya. Melalui fitur ini, pengguna dapat saling berbagi pengalaman, memberikan saran, atau menyampaikan keluhan, sehingga membantu pengguna lain dalam mengambil keputusan sebelum mengunjungi suatu lokasi. (Sentimen dkk., 2022)

2.3 Jemberkamera

Jemberkamera adalah perusahaan yang menyediakan layanan rental peralatan fotografi dan videografi, termasuk kamera (DSLR, *mirrorless, camcorder*), iPhone, lensa, *lighting*, drone, dan aksesori pendukung seperti tripod. Berdiri sejak 2016, kini Jemberkamera telah memiliki lebih dari 9 tahun pengalaman dalam melayani kebutuhan visual di wilayah Jember. Untuk menjangkau wilayah yang lebih luas, Jemberkamera memiliki tiga galeri di Jember:

a. Galeri Kampus di Jl. Danau Toba No.15, Lingkungan Panji, Tegal Boto, Kec. Sumbersari.
b. Galeri Mangli di Jl. Kauman, Karang Miuwo, Mangli, Kec. Kaliwates.
c. Galeri Ambulu, berlokasi sekitar Alun-alun Ambulu.

Berdasarkan data *Google Maps pada galari kampus di* Jl. Danau Toba No.15, Lingkungan Panji, Tegal Boto, Kec. Sumbersari, Jemberkamera telah menerima ribuan ulasan pelanggan hingga saat ini, yang memberikan gambaran langsung tentang kualitas layanan mereka dan menjadi bahan utama untuk analisis ulasan pelanggan dalam penelitian ini.

2.4 Web Scraping (Apify Google Maps Reviews Scraper)

Tahap *scraping* didefinisikan sebagai proses pengumpulan data ulasan pengguna secara sistematis dari *Google Maps*, yang dilakukan secara otomatis menggunakan platform Apify dengan memanfaatkan *actor Google Maps Reviews Scraper*. *Tools* ini memungkinkan pengambilan data secara detail berdasarkan URL lokasi, meliputi teks ulasan, rating, tanggal publikasi, serta informasi tambahan seperti respons pemilik dan identitas pengguna. Melalui *actor* tersebut, proses *scraping* dapat dilakukan tanpa perlu membangun sistem dari awal, karena data dapat langsung diekstraksi dan diunduh dalam format terstruktur seperti CSV atau Excel. Hal ini sangat membantu dalam mempercepat proses pengumpulan data serta mempermudah tahap analisis selanjutnya (Ardiansyah dan Kurniawan, 2024).

2.5 Preprocessing Kata

Tahap *Preprocessing* data teks adalah serangkaian langkah yang dilakukan sebelum data dimasukkan ke dalam model *text mining*. Tujuan utamanya adalah mengubah teks mentah menjadi data berkualitas tinggi yang siap untuk analisis

lebih lanjut (Hakim, 2021). Proses ini melibatkan analisis sintaksis untuk memastikan struktur teks yang tepat dan analisis semantik untuk menjaga makna yang sesuai (Apriliansyah dkk., 2025). Umumnya dilakukan dalam beberapa tahap meliputi sebagai berikut:

a. *Cleaning*: Proses menghilangkan karakter atau informasi yang tidak relevan dari teks, seperti emotikon, angka, URL, dan tanda baca, agar data menjadi lebih bersih dan fokus pada isi utama.
b. *Normalisasi*: Mengubah kata tidak baku dan singkatan menjadi kata baku.
c. *Case Folding*: Mengubah semua huruf dalam teks menjadi huruf kecil untuk menyeragamkan data dan menghindari perbedaan penafsiran karena kapitalisasi yang berbeda.
d. *Stopword Removal*: Menghapus kata-kata umum yang sering muncul namun tidak memiliki makna penting dalam analisis sentimen, seperti kata penghubung (misalnya, "dan", "atau").
e. *Tokenisasi*: Memecah teks menjadi unit-unit yang lebih kecil, biasanya kata atau frasa, yang disebut token, untuk memudahkan analisis lebih lanjut.
f. *Stemming*: Mengubah kata-kata menjadi bentuk dasarnya (kata akar) untuk mengurangi variasi kata dan mempermudah analisis, seperti mengubah "menjari" menjadi "ajari".

2.6 Pembobotan Kata

Pembobotan kata menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*). Tujuan dari tahap ini adalah untuk mengetahui kata-kata apa saja yang paling penting dalam setiap ulasan. Metode TF-IDF menghitung seberapa sering sebuah kata muncul dalam satu ulasan dan membandingkannya dengan frekuensi kata tersebut di semua ulasan. Kata yang sering muncul di satu ulasan tapi jarang di ulasan lain akan diberi bobot lebih tinggi. Dengan begitu, kata-kata yang memiliki arti penting dalam menentukan sentimen bisa dikenali dan digunakan untuk proses klasifikasi oleh algoritma *Naive Bayes* (Ardiansyah dan Kurniawan, 2024). Perhitungan *Term Frequency* (TF) dilakukan untuk mengetahui tingkat kemunculan suatu kata dalam sebuah dokumen. Nilai ini diperoleh dengan

menghitung jumlah kemunculan kata tertentu dibandingkan dengan total seluruh kata dalam dokumen tersebut. Terdapat rumus TF sebagai berikut:

$$TF(t,d) = \frac{Jumlah\ kata\ t\ dalam\ dokumen\ d}{Total\ kata\ dalam\ dokumen\ d} \quad \ldots\ldots (2.1)$$

Keterangan:

t = kata muncul

d = dokumen

Tahapan berikutnya adalah menghitung *Inverse Document Frequency* (IDF), yaitu untuk mengetahui frekuensi kemunculan suatu term pada seluruh dokumen dengan menggunakan rumus sebagai berikut:

$$IDF(t) = log\left(\frac{N}{dF(t)}\right) \quad \ldots\ldots (2.2)$$

Keterangan:

N = Total jumlah dokumen

d = Jumlah dokumen yang mengandung kata

Selanjutnya, dilakukan perhitungan TF-IDF yang merupakan hasil perkalian antara nilai TF dan IDF untuk memperoleh bobot akhir setiap kata dengan menggunakan rumus sebagai berikut:

$$TF - IDF\ (t,d) = TF\ (t,d) \times IDF\ ) \quad \ldots\ldots (2.3)$$

2.7 Sastrawi

Sastrawi sebuah *library* di pemograman *python* yang berfungsi dalam diksi kata, sehingga untuk mengubah kata menjadi bentuk dasarnya (Pardede dan Darmawan, 2025). Sehingga proses tersebut mampu menyederhanakan data teks dengan mengurangi variasi kata yang memiliki makna sama menjadi satu bentuk dasar, sehingga jumlah fitur menjadi lebih terstruktur dan tidak redundan. Dengan berkurangnya kompleksitas data, model klasifikasi dapat bekerja lebih optimal karena fokus pada kata-kata inti yang relevan, bukan pada variasi imbuhan. Hal ini

berdampak pada peningkatan akurasi model, karena proses pembelajaran menjadi lebih efektif, mengurangi noise, serta meningkatkan konsistensi dalam pengenalan pola pada data teks (Nafi' dkk., 2022).

2.8 Naive Bayes

Naive Bayes merupakan salah satu metode klasifikasi dalam *Machine Learning* (ML) yang bekerja berdasarkan prinsip probabilitas dan statistik sederhana. Metode ini menggunakan Teorema Bayes, yang pertama kali dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes. Teorema tersebut digunakan untuk menghitung probabilitas suatu kejadian di masa depan berdasarkan informasi atau pengalaman yang telah terjadi di masa lalu. Penambahan kata "*naive*" mengacu pada asumsi bahwa setiap atribut atau fitur dalam data bersifat bebas (*independen*) satu sama lain, meskipun dalam kenyataan asumsi ini jarang sepenuhnya terpenuhi (Afriansyah dkk., 2024).

Sebagai metode klasifikasi statistik, *Naive Bayes* berguna untuk memprediksi kemungkinan suatu data termasuk ke dalam kelas tertentu berdasarkan fitur yang dimilikinya. Model ini menginterpretasikan probabilitas *posterior* berdasarkan hubungan antara probabilitas awal (prior) dan probabilitas kondisi (*likelihood*), sesuai dengan bentuk umum Teorema Bayes. Rumus *Naive Bayes* secara umum dapat diberikan sebagai berikut:

$$P(H|X) \frac{P(X|H).P(H)}{P(X)} = \ldots\ldots (2.4)$$

Keterangan:

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

P (H | X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob)

P (H) = Probabilitas hipotesis H (prior prob)

P (X |H) = Probabilitas X berdasarkan kondisi tersebut

$P\ (X)$ = Probabilitas dari X

Untuk menerapkan metode ini ke dalam klasifikasi teks (seperti analisis sentimen), terdapat tiga tahapan perhitungan probabilitas utama yang harus dilakukan, yaitu perhitungan *prior probability*, *likelihood*, dan *posterior probability*.

Prior probability merupakan probabilitas awal dari suatu kelas yang dihitung berdasarkan rasio jumlah dokumen dalam kelas tersebut terhadap total seluruh dokumen yang ada pada data latih. Rumus perhitungan *prior probability* dapat dinyatakan sebagai berikut:

$$P(\ Positif\ \mid\ Negatif\ )\ \frac{d\ (\ Positif\ \mid\ Negatif\ )}{|c|} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.5$$

Keterangan:

P (kelas) = Probabilitas awal (*prior probability*) dari suatu kelas (positif, netral, atau negatif).

d (kelas) = Jumlah dokumen dalam kelas tersebut.

|c| = Jumlah total dokumen dari semua kelas.

Likelihood atau probabilitas bersyarat digunakan untuk menghitung seberapa besar peluang kemunculan suatu kata (*term*) di dalam suatu kelas. Pada analisis sentimen yang menggunakan pembobotan *Term Frequency - Inverse Document Frequency* (TF-IDF), nilai frekuensi kemunculan kata digantikan oleh akumulasi bobot TF-IDF, dengan penambahan standardisasi *Laplace Smoothing* (+1) untuk menghindari nilai probabilitas nol. Rumus menghitung *likelihood* adalah sebagai berikut:

$$P(\ kata\ \mid\ kelas\ )\ \frac{Bobot\ TF-IDF\ term\ i\ pada\ kelas\ +\ 1}{Jumlah\ total\ TF-IDF\ dari\ semua\ term\ +\ |\mathcal{V}|} =\ldots\ldots\ldots 2.6$$

Keterangan:

|V| = Jumlah total kata unik dalam semua dokumen (*vocabulary*)

+1 = *Laplace* smoothing agar tidak nol jika kata tidak muncul dalam satu kelas

Posterior probability merupakan nilai akhir yang menunjukkan probabilitas suatu dokumen masuk ke dalam suatu kelas setelah memperhitungkan kata-kata yang terkandung di dalam dokumen tersebut. Nilai ini diperoleh dengan mengalikan *prior probability* dari suatu kelas dengan seluruh nilai *likelihood* (probabilitas bersyarat) dari setiap kata yang menyusun dokumen tersebut. Secara matematis, rumus umum untuk menghitung nilai *posterior probability* pada suatu dokumen dapat dinyatakan sebagai berikut:

$$P(\,c \mid d\,) \propto P(c) \times \prod_{i=1}^{n} P\,(W_i \mid c) \quad \ldots\ldots 2.7$$

Jika dijabarkan menjadi perkalian beruntun sebagai berikut:

$$P(\,c \mid d\,) \propto P(c)\ \times P(W_i \mid c)\ \times P(W_2 \mid c)\ \times P(W_3 \mid c) \ldots \times P(W_n \mid c)$$

Keterangan:

$P(\,c \mid d\,)$ = Probabilitas *posterior* kelas c setelah dokumen d diberikan.

$\propto$ = Simbol proporsional (*proportional to*), menunjukkan bahwa pembagi (P(d) diabaikan karena nilainya sama untuk semua kelas.

P(c) = Probabilitas *prior* dari kelas c.

$\prod_{i=1}^{n}$ = Notasi produk yang berarti perkalian beruntun dari kata pertama (i=1) hingga kata terakhir (n).

$P(W_i \mid c)$ = Probabilitas bersyarat (*likelihood*) kata ke-i (w_i) pada kelas c.

n = Total jumlah kata unik yang ada di dalam dokumen yang sedang diuji setelah melalui tahap *Preprocessing*.

2.9 Confusion Matrix

Evaluasi kinerja sistem berbasis data sangat bergantung pada proses klasifikasi, sebab ketepatan pengelompokan data menjadi tolok ukur kualitas sistem tersebut. Untuk mengukur performa ini, *confusion matrix* digunakan guna membandingkan kesesuaian antara hasil prediksi sistem dengan label *ground-truth*. Pada klasifikasi *multi* kelas yang melibatkan sentimen positif, netral, dan negatif, performa model dievaluasi berdasarkan distribusi prediksi pada setiap kelas

tersebut. Parameter utama yang diukur adalah *True Positive* (TP) untuk masing-masing kelas, yang merepresentasikan keberhasilan sistem dalam mengidentifikasi data aktual ke dalam kelas prediksi yang tepat secara akurat terletak pada garis diagonal matriks. (Fariz dkk., 2024). Adapun rumus untuk menghitung nilai akurasi disajikan sebagai berikut:

$$Akurasi\ \frac{Total\ Prediksi\ yang\ Benar\ (Diagonal\ Confusion\ Matrix)}{Total\ Seluruh\ Data\ Uji} \times 100\% \ldots\ldots 2.8$$

Keterangan:

TP = *True Po sitive*

TN = *True Negative*

FP = *False Positive*

FN = *False Negative*

Melalui pemanfaatan *confusion matrix*, evaluasi terhadap pengujian model tidak sekadar berfokus pada kuantitas prediksi yang tepat, melainkan juga mengidentifikasi karakteristik kesalahan klasifikasi yang dihasilkan. Analisis ini menjadi krusial karena tipe kesalahan tertentu berpotensi membawa dampak yang lebih fatal dibandingkan kesalahan lainnya.

Selain itu, *precision* dijelaskan sebagai metrik evaluasi yang berfungsi mengukur tingkat keakuratan model dalam memprediksi kelas positif dibandingkan dengan keseluruhan jumlah data yang diprediksi sebagai positif oleh sistem. rumus sebagai berikut:

$$Precision(kelas)\ \frac{TP_{os,eu,eg}}{TP_{osos,eu,eg} + FP_{os,eu,eg}} = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.9$$

Recall didefinisikan sebagai metrik evaluasi yang menggambarkan tingkat keberhasilan suatu model dalam mengidentifikasi kembali seluruh data yang termasuk dalam kelas tertentu secara benar. Nilai ini berfungsi untuk menjawab pertanyaan sejauh mana model mampu mengenali dan mengklasifikasikan data aktual di lapangan tanpa ada yang luput. Rumus sebagai berikut:

$$Recall\ (positif)\ \frac{TP_{osos,eu,eg}}{TP_{os,eu,eg} + FN_{os,eu,eg}} = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.10$$

F1-score merupakan metrik evaluasi yang digunakan untuk menunjukkan keseimbangan antara nilai precision dan recall melalui rata-rata harmonis dari kedua metrik tersebut. Nilai *F1-score* memberikan gambaran mengenai kemampuan model dalam mengklasifikasikan data secara akurat dengan mempertimbangkan ketepatan prediksi dan kemampuan mendeteksi seluruh data pada setiap kelas. Pada klasifikasi tiga kelas, perhitungan nilai *F1-score* dilakukan secara terpisah untuk masing-masing kelas, yaitu positif, netral, dan negatif, sebagaimana ditunjukkan pada rumus berikut:

$$f1 - class = 2 \times \frac{precision_{os,eu,eg} \times recall_{os,eu,eg}}{precision_{os,eu,eg} + recall_{os,eu,eg}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.11$$

Keterangan:

Recal = hasil dari recall

Precision = hasil dari Precision

2.10 State of The Art

Dalam penyusunan penelitian ini, telah ditemukan sejumlah penelitian terdahulu yang memiliki kemiripan dan pendekatan metodologis. Penelitian tersebut menjadikan refresnsi penting dan landasan sumber awal pada merumuskan arah serta focus penelitian.

Penelitian terdahulu yang dilakukan oleh Dianda Rifaldi dkk. (2025) dalam jurnal yang berjudul "Evaluasi Sentimen Pengguna ChatGPT Menggunakan Naive Bayes: Tinjauan dari Confusion Matrix dan Classification Report" mengevaluasi sentimen pengguna ChatGPT di media sosial Twitter menggunakan algoritma Naive Bayes dan pembobotan TF-IDF. Dengan rasio pembagian data 80:20, pengujian menghasilkan nilai akurasi yang relatif rendah yaitu sebesar 56%. Hasil evaluasi *Classification Report* dan confusion matrix pada penelitian tersebut menunjukkan adanya ketimpangan performa yang signifikan, di mana model mengalami bias pada kelas mayoritas dan kesulitan mengenali kelas sentimen

Netral. Untuk mengatasi kelemahan tersebut, peneliti menyarankan penerapan teknik penyeimbangan distribusi data seperti metode SMOTE pada penelitian selanjutnya.

Sementara itu, Fathir dkk. (2023) dalam penelitiannya yang berjudul "Analisis Sentimen Artikel Berita Pemilu Berbasis Metode Klasifikasi" melakukan analisis sentimen terhadap artikel berita pemilu menggunakan tiga algoritma komparasi, yaitu Naive Bayes, Random Forest, dan Support Vector Machine (SVM). Eksperimen dilakukan dalam dua kondisi, yaitu dengan menerapkan metode SMOTE dan tanpa SMOTE untuk menangani masalah ketidakseimbangan data latih. Hasil pengujian menunjukkan bahwa performa algoritma Random Forest meningkat hingga mencapai akurasi tertinggi sebesar 91,88% saat dikombinasikan dengan SMOTE. Sebaliknya, pada kondisi tanpa SMOTE, algoritma Support Vector Machine justru mencatatkan performa puncak dengan perolehan nilai akurasi mencapai 92,05%.

Pada penelitian "Analisis Sentimen Ulasan Mie Gacoan Solo Veteran Di *Google Maps* Menggunakan Algoritma *Naive Bayes*", peneliti bertujuan untuk membantu perusahaan Mie Gacoan Solo Veteran memahami tanggapan konsumen berdasarkan opini dan ulasan di *Google Maps*. Data dikumpulkan melalui crawling, diberi label, dan diproses (*Preprocessing*) sebelum diekstraksi menggunakan TF-IDF dan dimodelkan dengan *Naive Bayes*. Hasil awal pada data tidak seimbang menunjukkan akurasi 86%. Namun, setelah mengatasi ketidakseimbangan data dengan metode oversampling SMOTE (*Synthetic Minority Over-sampling Technique*), akurasi model meningkat menjadi 91%

Dalam penelitian berjudul "Analisis Sentimen Pengguna Aplikasi CapCut Pada Ulasan di Play Store Menggunakan Metode Naive Bayes", peneliti melakukan analisis sentimen untuk mengekstraksi dan memahami sentimen (positif atau negatif) dari ulasan pengguna aplikasi CapCut di *Play Store*. Penelitian ini bertujuan untuk mengetahui jumlah sentimen positif dan negatif dari ulasan pengguna, serta mengevaluasi akurasi, presisi, dan recall yang diperoleh dari confusion matrix, dengan harapan mencapai tingkat akurasi yang tinggi dalam klasifikasi menggunakan *Naive Bayes*. Data ulasan diambil dari *Play Store*

menggunakan teknik *web scraping* dan diproses melalui tahapan *Preprocessing* (pembersihan, *tokenisasi, transform cases, filter stopwords, dan filter tokens by length*). Kemudian, data diberi label secara manual menjadi sentimen positif atau negatif dan dibagi menjadi data latih dan data uji. Hasil implementasi *Naive Bayes* menunjukkan lebih banyak sentimen negatif dibandingkan sentimen positif, dan evaluasi menggunakan *confusion matrix* menghasilkan akurasi sebesar 84.09%, presisi 91.91%, dan recall 73.53%, menunjukkan bahwa metode *Naive Bayes* memiliki akurasi yang tinggi dalam klasifikasi data pada penelitian ini.

Tabel 2. 1 *State of The Art*

No	Nama Penulis	Judul Jurnal	Metode	Hasil
1	Dianda Rifaldi, dkk (2025)	Evaluasi Sentimen Pengguna ChatGPT Menggunakan Naive Bayes: Tinjauan dari Confusion Matrix dan Classification Report	Naive Bayes dan TF-IDF	Akurasi 56%, *F1-score* tertinggi pada kelas negatif (0.67) dan terendah pada kelas netral (0.38). Model mengalami bias akibat ketimpangan data ulasan teks.
2	Fathir, dkk (2023)	Analisis Sentimen Artikel Berita Pemilu Berbasis Metode Klasifikasi	Naive Bayes, Random Forest, SVM, dan SMOTE	Eksperimen SMOTE: Random Forest meraih akurasi tertinggi 91,88%. Eksperimen Tanpa SMOTE: SVM meraih akurasi tertinggi 92,05%.

3	Tariska Zidny Fatikhah, dkk (2024)	Analisis Sentimen Ulasan Mie Gacoan Solo Veteran Di *Google Maps* Menggunakan Algoritma *Naive Bayes*	Naive Bayes	Akurasi awal 86%, meningkat jadi 91% setelah SMOTE; model lebih akurat setelah mengatasi data imbalance
4	Meliyawati, dkk (2024)	Analisis Sentimen Pengguna Aplikasi CapCut Pada Ulasan di Play Store Menggunakan Metode Naive Bayes	Naive Bayes	Akurasi 84,09%, precision 91,91%, recall 73,53%; sentimen negatif lebih banyak dari positif

Berdasarkan pemetaan penelitian terdahulu pada Tabel 2.1, masing-masing peneliti memiliki kontribusi dan kelebihan spesifik dalam pengembangan bidang analisis sentimen. Penelitian oleh Dianda Rifaldi dkk. (2025) unggul dalam mengevaluasi metrik klasifikasi secara mendalam menggunakan kombinasi *confusion matrix* dan *classification report* pada data media sosial. Sementara itu, Fathir dkk. (2023) memiliki kelebihan dalam menyajikan studi komparasi yang komprehensif dengan membandingkan tiga algoritma sekaligus, yaitu Naive Bayes, Random Forest, dan SVM. Di sisi lain, Tariska Zidny dkk. (2024) berhasil membuktikan efektivitas metode oversampling SMOTE dalam mendongkrak akurasi model Naive Bayes secara signifikan dari 86% menjadi 91% pada ulasan Google Maps. Terakhir, penelitian oleh Meliyawati dkk. (2024) menunjukkan keunggulan pada penerapan tahapan *preprocessing* yang terstruktur rapi untuk mengekstraksi sentimen dari *platform Play Store*.

Meskipun penelitian terdahulu memiliki berbagai kelebihan, penelitian yang penulis lakukan saat ini memiliki keunggulan utama dalam hal otomatisasi sistem praktis. Kelebihan dari penelitian ini dijabarkan ke dalam tiga poin utama, yaitu:

1. Sistem yang dibangun ini menyatukan seluruh alur proses mulai dari pembacaan file mentah hasil *scraping*, pembersihan teks *preprocessing*, pembobotan TF-IDF,

penyeimbangan data latih lewat metode SMOTE, pelatihan model Naive Bayes, hingga pengujian performa hanya dalam satu kali eksekusi di dalam aplikasi.

2. Aplikasi berbasis *web* dengan hak akses terbatas yang diperuntukkan khusus bagi admin atau pengelola internal Jemberkamera saja, tanpa menyediakan antarmuka bagi pelanggan umum. Hal ini memastikan hasil evaluasi pelayanan aman sehingga mendukung pengambilan keputusan strategis manajemen.
3. Sistem ini dilengkapi dengan keunggulan visualisasi interaktif yang menampilkan grafik komparasi distribusi data sebelum dan sesudah SMOTE, split data, *confusion matrix*, *classification report*, serta *word cloud*.

BAB 3. METODOLOGI PENELITIAN

3.1 Tempat dan Waktu Penelitian

Penelitian akan dilakukan secara daring dan luring. Peneliti akan mengambil data ulasan pelanggan dari jemberkamera secara daring dari Politeknik Negeri Jember menggunakan platform *website* yaitu Apify. Peneliti juga akan melakukan verifikasi dengan persewaan jemberkamera yang beralamat di Jl. Danau Toba No.15, Lingkungan Panji, Tegal Boto, Kec. Sumbersari, Kabupaten Jember, Jawa Timur 68121. Adapun waktu pelaksanaan, penelitian ini akan dilaksanakan selama 8 bulan.

3.2 Alat dan Bahan

3.2.1 Alat

Alat pendukung yang digunakan untuk melakukan penelitian yang berjudul "Analisis Ulasan Pelanggan Pada Sewa Kamera di Jemberkamera Menggunakan Algoritme Naive Bayes" terdiri dari perangkat keras (*Hardware*) dan perangkat lunak (*Software*) sebagai berikut:

a. Perangkat Keras (*hardware*)

Perangkat Keras (*hardware*) yang digunakan dalam proses penelitian ini Laptop ASUS VivoBook 14_x441ubr dengan spesifikasi prosesor sebagai berikut:

1) Intel(R) Core (Tm) I3-7020u
2) RAM 8 GB (8192 MB)

b. Perangkat Lunak (*software*)

Perangkat lunak (*software*) yang digunakan untuk pelaksanaan penelitian ini sebagai berikut:

1) Sistem Operasi Windows 11
2) Microsoft Office 2016
3) *Google Chrome*
4) Excel 2016
5) Mendeley Reference Manager
6) Python 3.10.6

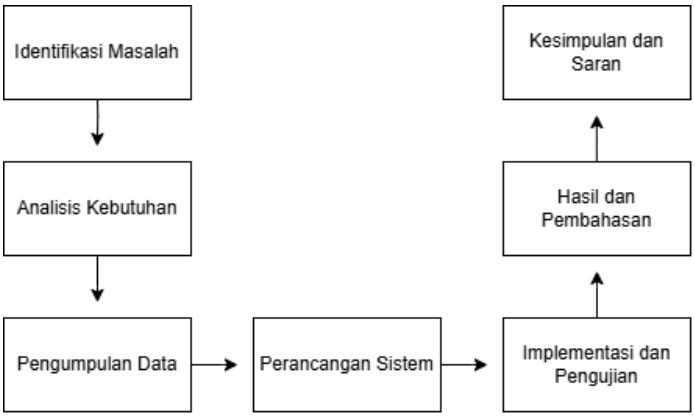
7) PHP 8.2.28
8) Laravel

3.2.2 Bahan

Bahan yang digunakan pada penelitian ini berupa data komentar pelanggan Jemberkamera yang diperoleh dari ulasan pada *Google Maps* pada link berikut (*https://www.google.com/maps/search/jemberkamera/*), yaitu kurang lebih kisaran 4000 data ulasan. Data tersebut diambil menggunakan bantuan dari platform *website* yaitu apify.

Lokasi Jember Kamera memiliki 3 cabang. Pada penelitian ini, data ulasan yang digunakan diambil dari lokasi di Jl. Danau Toba No.15, karena lokasi tersebut merupakan pusat Jemberkamera yang berada di wilayah kota Jember serta memiliki jumlah ulasan paling banyak di *Google Maps*, sehingga memudahkan peneliti dalam memperoleh dataset yang lebih lengkap dan relevan untuk mendukung proses penelitian.

3.3 Tahapan Penelitian



Gambar 3. 1 Tahapan Penelitian

Pada tahapan penelitian ini dilakukan beberapa tahapan pelaksanaan, yang akan disajikan dalam bentuk gambar 3.1.

3.3.1 Identifikasi Masalah

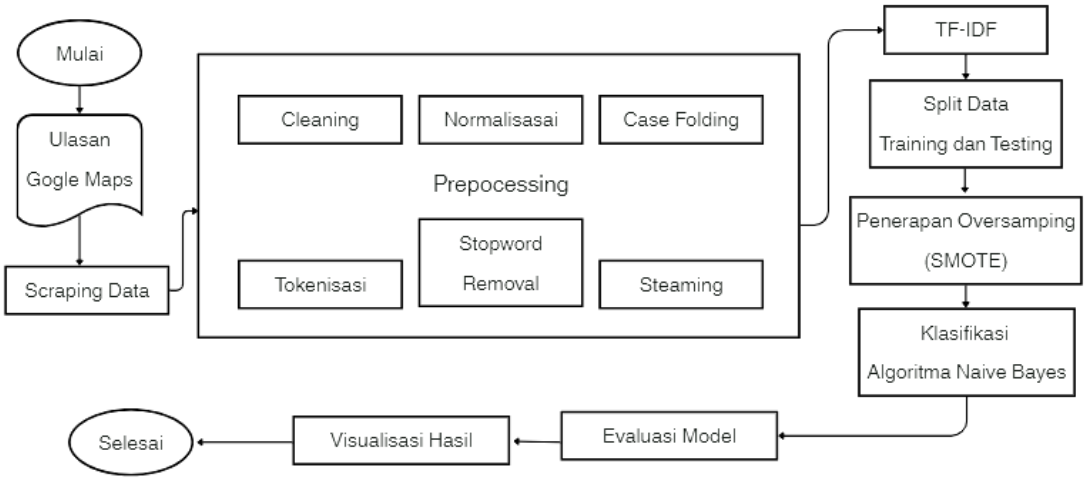
Proses identifikasi masalah dilakukan dengan mengkaji ulasan pelanggan yang pernah menggunakan layanan Jemberkamera di Kabupaten Jember. Langkah ini bertujuan untuk mengidentifikasi faktor-faktor penentu yang mendorong

pelanggan memberikan ulasan positif netral dan negatif. Hasil analisis ini digunakan sebagai bahan evaluasi agar Jemberkamera mampu mengoptimalkan pelayanan pada penyewaan kamera dan peralatan lainnya.

3.3.2 Analisis Kebutuhan

Analisis kebutuhan bertujuan untuk mengidentifikasi, merancang, dan mendokumentasikan kebutuhan teknis dan fungsional sistem klasifikasi sentimen berbasis *web*.

a. Alur Kerja Sistem



Gambar 3. 2 Alur Kerja Sistem

Alur kerja sistem klasifikasi sentimen ini dimulai dari tahap pengumpulan data hingga visualisasi hasil analisis. Proses diawali dengan pengambilan data ulasan dari *Google Maps*, khususnya ulasan pelanggan Jemberkamera, menggunakan teknik *web scraping* melalui platform Apify. Data ulasan yang telah dikumpulkan kemudian melalui tahapan *preprocessing* yang meliputi *cleaning* (penghapusan simbol, angka, dan tanda baca), *case folding* (penyeragaman huruf menjadi huruf kecil), normalisasi kata tidak baku, s*topword removal* (penghapusan kata umum yang tidak memiliki makna penting), tokenisasi (pemecahan kalimat menjadi kata), dan *stemming* (pengubahan kata menjadi bentuk dasar). Setelah proses *preprocessing* selesai, data teks yang telah bersih kemudian dikonversi menjadi

representasi numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) sebagai bobot kata yang digunakan dalam proses klasifikasi.

Dataset yang telah dibobot kemudian dibagi menjadi data training dan data testing untuk keperluan pelatihan dan pengujian model. Setelah pembagian data dilakukan, Penerapan *oversampling* (*Synthetic Minority Over-sampling Technique*) SMOTE yang diproses khusus pada data training untuk menyeimbangkan jumlah sampel kelas minoritas sebelum masuk ke fase pembelajaran. Pada tahap pelatihan, model klasifikasi sentimen dibangun menggunakan algoritma *Naive Bayes*. Model akan mempelajari pola kemunculan kata berdasarkan distribusi probabilitas pada masing-masing kelas sentimen, yaitu positif, netral, dan negatif. Probabilitas tersebut dihitung menggunakan *prior probability, likelihood dan posterior* untuk menentukan kecenderungan sentimen dari suatu ulasan.

Selanjutnya, model diuji menggunakan data testing untuk mengetahui performa klasifikasi. Evaluasi dilakukan menggunakan metrik *accuracy, precision, recall,* dan *F1-score* guna mengukur tingkat ketepatan model dalam melakukan klasifikasi sentimen. Hasil analisis kemudian divisualisasikan dalam bentuk distribusi sentimen positif, netral, dan negatif, evaluasi model, split data, distribusi data, *confusion matrix, classification report*, dan *word cloud* sehingga, bisa terlihat ditampilkan melalui antarmuka *website*.

Secara keseluruhan, sistem ini merupakan aplikasi klasifikasi sentimen berbasis web yang dirancang untuk proses analisis sentimen mulai dari pengambilan data ulasan *Google Maps*, *Preprocessing* teks, pembobotan TF-IDF, pemisahan data latih dan uji, penyeimbangan data menggunakan teknik SMOTE pelatihan dan pengujian model menggunakan algoritma *Naive Bayes*, hingga penyajian hasil evaluasi dan visualisasi sentimen secara komprehensif. Sistem ini bertujuan membantu pihak Jemberkamera dalam memantau opini pelanggan secara efisien sehingga dapat menjadi bahan pertimbangan dalam meningkatkan kualitas layanan dan produk.

Oleh sebab itu, pada penelitian ini fitur-fitur yang diperlukan agar mampu mengetahui ulasan pelanggan sebagai berikut:

1) Sistem harus mampu *Upload* data hasil dari *scraping* dari *Google Maps*.

2) Sistem mampu mengelola data *review* ulasan pelanggan, termasuk tahapan *Preprocessing* seperti *Cleaning, Case Folding*, normalisasi kata tidak baku, *tokenisasi*, *Stopword Removal*, dan *Stemming*, serta pembobotan TF-IDF.
3) Sistem mampu menampilkan hasil ulasan apakah tergolong positif, netral, atau negatif, setelah melalui proses pelatihan dan pengujian model *Naive Bayes*.
4) Sistem mampu menampilkan metrik evaluasi (akurasi, presisi, recall, *F1-score*) dan visualisasi hasil (seperti *word cloud* dan distribusi sentimen).

3.3.3 Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan menggunakan teknik web *scraping* dengan bantuan platform Apify melalui *actor Google Maps Reviews Scraper*. Proses ini digunakan untuk mengumpulkan data ulasan pelanggan pada *Google Maps* dalam format terstruktur seperti file .csv atau .xlsx. Data yang diperoleh berupa teks ulasan, rating, tanggal, serta informasi pengguna yang memberikan ulasan terhadap layanan Jemberkamera. Data tersebut kemudian digunakan sebagai bahan dalam proses analisis sentimen untuk memahami opini pelanggan, yang selanjutnya akan diklasifikasikan ke dalam kategori positif, netral dan negatif.

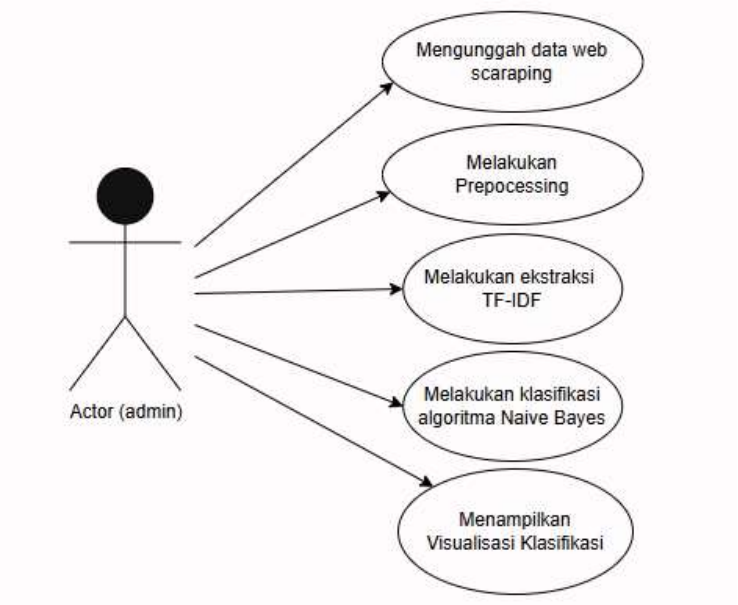
3.3.4 Perancangan Sistem

Perancangan sistem dalam penelitian ini mencakup pembuatan *use case* diagram untuk memetakan interaksi aktor terhadap fungsi-fungsi sistem, serta *wireframe* aplikasi sebagai rancangan visual antarmuka pengguna. Seluruh komponen tersebut dikembangkan secara terstruktur guna membentuk arsitektur sistem yang modular dan ramah pengguna *user-friendly*. Melalui perancangan ini, alur fungsional aplikasi analisis sentimen ulasan pelanggan Jemberkamera dapat dipastikan berjalan secara sistematis, mulai dari pengelolaan dataset, proses penyeimbangan data dengan SMOTE, hingga visualisasi hasil klasifikasi Naive Bayes.

a. *Use Case Diagram*

Sistem ini dirancang dengan satu aktor utama, yaitu Admin. Berdasarkan *use case diagram*, aktor memiliki akses untuk menjalankan beberapa proses utama dalam sistem analisis sentimen. Proses pertama adalah mengunggah data hasil *web scraping* yang digunakan sebagai sumber data ulasan. Setelah data berhasil diunggah, sistem akan melakukan *preprocessing* untuk membersihkan dan mempersiapkan data teks agar siap digunakan pada tahap analisis.

Tahap berikutnya adalah melakukan ekstraksi fitur menggunakan metode TF-IDF untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi. Setelah proses ekstraksi selesai, sistem akan menjalankan klasifikasi sentimen menggunakan algoritma Naive Bayes dan menampilkan hasil klasifikasi sentimen ke dalam kategori positif, netral, dan negatif. Dengan demikian, *use case diagram* menggambarkan alur utama sistem mulai dari pengunggahan data hingga menampilkan hasil klasifikasi sentimen secara terintegrasi.

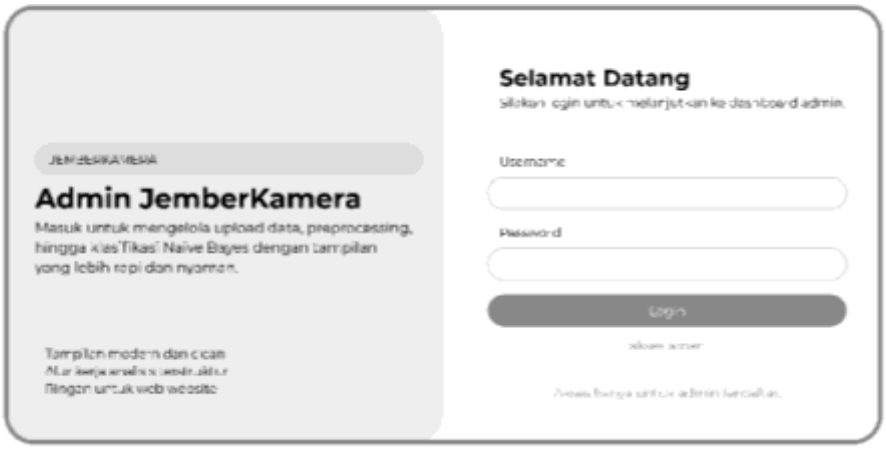


Gambar 3. 3 Use Case Diagram

b. *Wireframe Website*

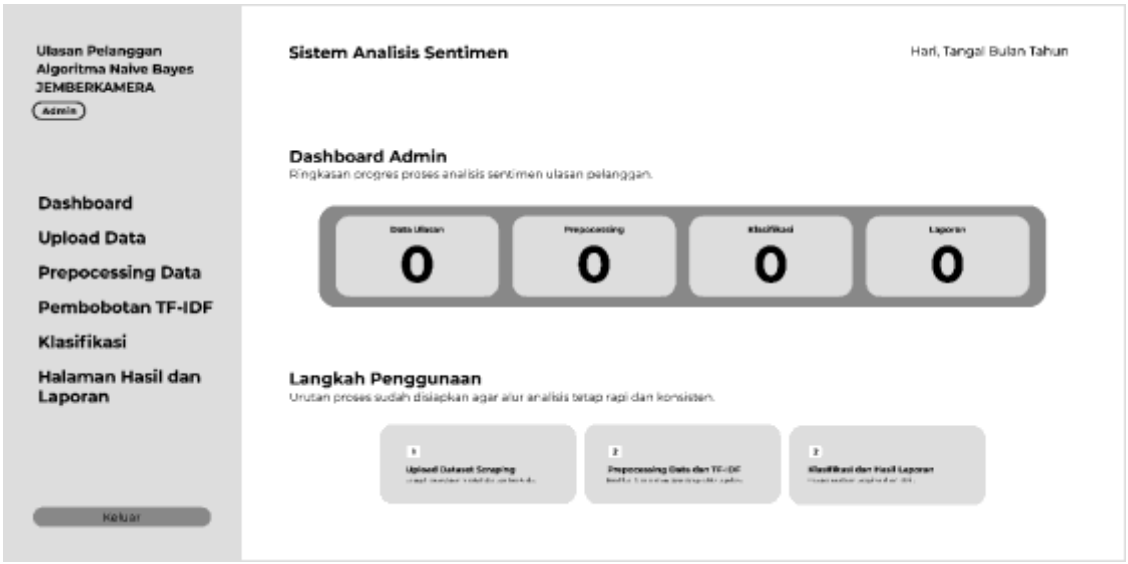
Wireframe aplikasi yang disajikan melalui serangkaian rancangan antarmuka bertujuan untuk memvisualisasikan struktur dan alur penggunaan sistem klasifikasi

sentimen berbasis *website* menggunakan metode *Naive Bayes*. Desain *wireframe* dibuat dengan pendekatan sederhana, terstruktur, dan mudah dipahami agar pengguna dapat mengikuti setiap tahapan proses analisis sentimen secara sistematis. Setiap halaman dirancang untuk mendukung proses pengelolaan data, *Preprocessing*, pembobotan kata, klasifikasi, hingga penyajian hasil analisis sentimen.



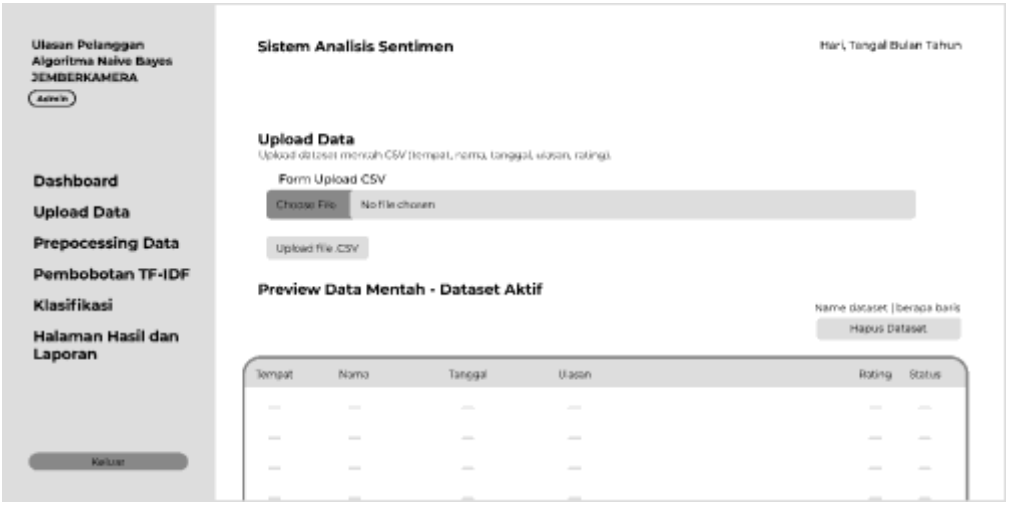
Gambar 3. 4 W*ireframe* Halaman Login

Gambar 3.4 halaman ini menampilkan antarmuka autentikasi pengguna (*login page*) yang dirancang dengan tata letak *split-screen modern* untuk mengamankan sekaligus membatasi hak akses ke dalam sistem klasifikasi sentimen Jemberkamera. Pada panel sisi kiri, disajikan informasi pengantar mengenai identitas aplikasi serta poin-poin keunggulan sistem seperti alur kerja analisis yang terstruktur dan performa website yang ringan. Sementara itu, panel sisi kanan memuat formulir interaktif berupa kolom masukan (*input field*) untuk *Username* dan *Password*, serta tombol eksekusi utama dan catatan kaki yang menegaskan bahwa otoritas masuk ke dalam dashboard admin hanya diperuntukkan bagi pengguna yang telah terdaftar secara resmi.



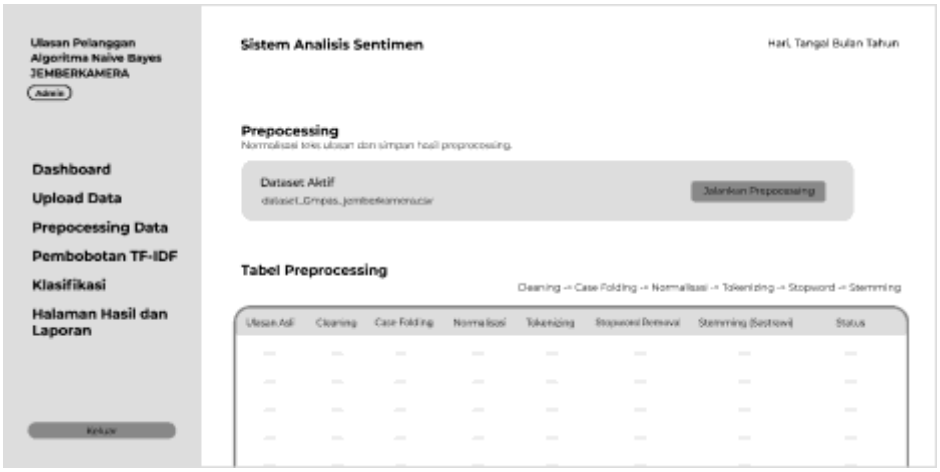
Gambar 3. 5 *Wireframe* Halaman Dashboard

Gambar 3.5 menampilkan ringkasan informasi utama sistem, seperti jumlah data ulasan, distribusi sentimen, serta menu navigasi menuju setiap tahapan proses analisis sentimen.



Gambar 3. 6 *Wireframe* Halaman Upload Data

Gambar 3.6 Halaman *Upload Data* digunakan untuk mengunggah dataset ulasan yang akan diproses pada sistem. Pada halaman ini pengguna dapat mengimpor data dalam format tertentu sehingga data dapat digunakan pada proses analisis berikutnya.



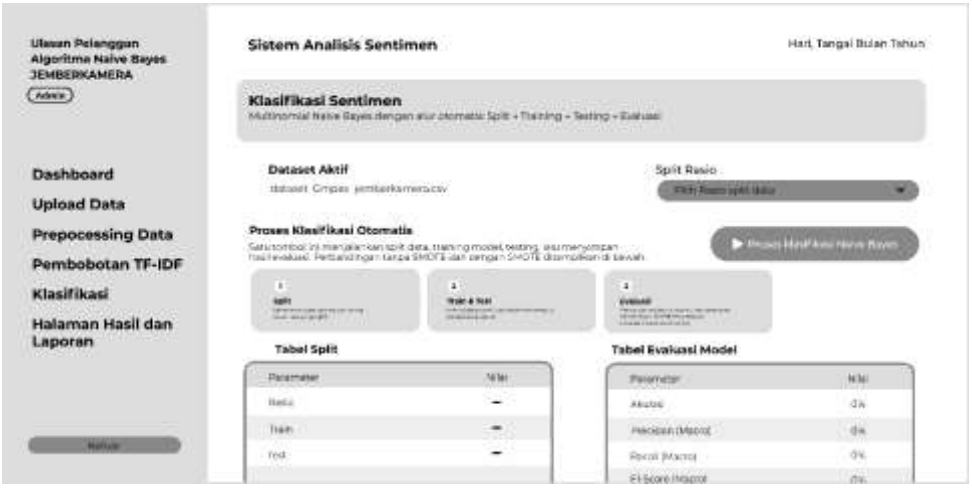
Gambar 3. 7 *Wireframe* Halaman *Prepocessing*

Gambar 3.7 halaman ini menampilkan struktur visual menu pengelolaan data teks murni hasil *upload*. Dilengkapi dengan tombol eksekusi utama "Jalankan Preprocessing", hasil dari proses tersebut disajikan dalam bentuk tabel tabular multi-kolom yang memetakan transformasi teks ulasan secara transparan, mulai dari teks asli pelanggan, hasil *cleaning*, *case folding*, normalisasi kata baku, *tokenizing*, *stopword removal*, hingga bentuk kata dasar pada kolom *stemming*.

Gambar 3. 8 *Wireframe* Halaman TF-IDF

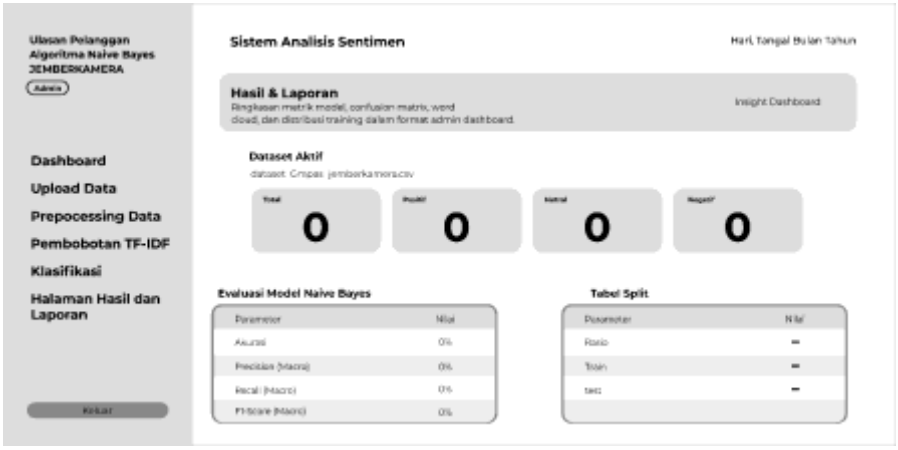
Gambar 3.8 halaman ini menampilkan menu utama pembobotan kata berbasis matriks numerik. Tata letak antarmuka dibagi menjadi dua komponen utama, sistem menyajikan kolom informasi dataset aktif yang dilengkapi dengan tombol eksekusi interaktif "Hitung Pembobotan", diikuti penjelasan rumus acuan kalkulasi di bagian

bawahnya. Selain itu, hasil ekstraksi fitur disajikan secara rapi menggunakan tiga baris komponen lipat (*accordion collapse*) yang memisahkan tabel data nilai *Term Frequency* (TF), *Document Frequency* (DF) - *Inverse Document Frequency* (IDF), serta hasil perkalian akhir TF-IDF secara terstruktur.



Gambar 3. 9 *Wireframe* Halaman Klasifikasi

Gambar 3.9 Halaman Klasifikasi Naive Bayes menampilkan proses klasifikasi sentimen menggunakan metode *Naive Bayes*. Pada tahap ini sistem melakukan perhitungan *prior probability*, *likelihood* atau *conditional probability*, serta *posterior probability* untuk menentukan kelas sentimen setiap ulasan ke dalam kategori positif, netral, atau negatif. Selain itu, halaman ini juga menampilkan hasil probabilitas terbesar yang digunakan sebagai dasar penentuan hasil klasifikasi akhir.



Gambar 3. 10 *Wireframe* Halaman Hasil dan Laporan

Gambar 3.10 Halaman Hasil dan Laporan menampilkan hasil akhir analisis sentimen yang telah dilakukan oleh sistem. Pada halaman ini pengguna dapat melihat distribusi jumlah sentimen positif, netral, dan negatif dalam bentuk tabel maupun grafik visualisasi. Selain itu, ditampilkan pula hasil evaluasi performa model menggunakan confusion matrix, *accuracy, precision, recall,* dan *F1-score* untuk mengetahui tingkat kinerja metode *Naive Bayes* dalam melakukan klasifikasi sentimen.

3.3.5 Implementasi dan Pengujian

Tahap ini merupakan realisasi dari seluruh rancangan sistem yang telah dibuat sebelumnya. Proses implementasi dimulai dengan pengembangan aplikasi berbasis *website* menggunakan *framework* Laravel. *Website* ini dirancang untuk memproses data ulasan pelanggan Jemberkamera, mulai dari pengumpulan data melalui *web scraping, Preprocessing teks*, hingga klasifikasi sentimen menggunakan algoritma *Naive Bayes*. Setelah proses implementasi selesai, dilakukan serangkaian pengujian untuk memastikan bahwa sistem berjalan sesuai fungsi yang diharapkan, serta mampu mengklasifikasikan sentimen ulasan secara akurat. Pengujian dilakukan dalam beberapa tahap berikut:

a. Pengujian Fungsional (*Black Box Testing*)

Pengujian fungsional dilakukan untuk memastikan bahwa seluruh fitur pada sistem analisis sentimen dapat berjalan sesuai dengan fungsi yang telah dirancang. Metode pengujian yang digunakan adalah *Black Box Testing*, yaitu pengujian yang berfokus pada kesesuaian input dan output tanpa memperhatikan proses internal sistem. Pengujian dilakukan pada setiap fitur utama, mulai dari proses unggah data hasil *web scraping*, *Preprocessing*, ekstraksi fitur TF-IDF, proses klasifikasi menggunakan algoritma Naive Bayes, hingga penampilan hasil klasifikasi sentimen.

Selain itu, sistem juga diuji pada proses pembagian data *split data* menggunakan beberapa skenario rasio, yaitu 80:20, 70:30, 60:40, dan 50:50 antara data *training* dan data *testing*. Penggunaan beberapa skenario ini bertujuan untuk memastikan seluruh fungsi sistem dapat berjalan dengan baik pada berbagai komposisi data latih dan data uji. Hasil pengujian menunjukkan bahwa seluruh fitur

sistem dapat berjalan sesuai dengan kebutuhan dan menghasilkan output yang sesuai dengan proses analisis sentimen yang dirancang.

b. Evaluasi Performa Model (*Confusion Matrix*)

Evaluasi performa model dilakukan untuk mengukur tingkat kinerja algoritma Naive Bayes dalam mengklasifikasikan sentimen ulasan pelanggan Jemberkamera. Proses evaluasi dilakukan dengan membandingkan hasil prediksi model terhadap label data sebenarnya menggunakan confusion matrix. Berdasarkan confusion matrix tersebut diperoleh nilai akurasi *accuracy*, presisi (*precision*), *recall*, dan *F1-score* sebagai indikator performa model klasifikasi.

Pengujian dilakukan menggunakan beberapa skenario pembagian data, yaitu 80:20, 70:30, 60:40, dan 50:50 antara data training dan data testing. Variasi rasio pembagian data tersebut digunakan sebagai parameter uji untuk mengetahui pengaruh komposisi jumlah data terhadap performa model klasifikasi Naive Bayes. Melalui pengujian multi-skenario ini, sistem dapat dianalisis secara objektif guna menentukan rasio pembagian data terbaik yang menghasilkan tingkat akurasi dan performa metrik paling optimal dalam mengelompokkan ulasan pelanggan ke dalam sentimen positif, netral, dan negatif.

c. Pengujian Pengguna (User Testing)

Pengujian lapangan dilakukan dengan menguji sistem secara langsung menggunakan data ulasan terbaru dari *Google Maps* yang belum pernah dilabeli sebelumnya. Ulasan ini diproses melalui *website* untuk mengevaluasi kecepatan klasifikasi, keakuratan prediksi, serta kemudahan penggunaan antarmuka sistem. Pengujian ini bertujuan untuk memastikan bahwa sistem layak dan siap digunakan oleh pihak pengelola Jemberkamera sebagai alat bantu dalam memantau opini serta kepuasan pelanggan.

3.3.6 Hasil dan Pembahasan

Hasil penelitian ini diperoleh melalui proses analisis dan pelabelan data ulasan pelanggan, yang diklasifikasikan menjadi tiga kategori, yaitu ulasan bernilai positif, netral dan negatif, menggunakan algoritma *Naive Bayes*. Setelah proses klasifikasi dilakukan, tingkat akurasi dari model dianalisis untuk mengetahui seberapa baik

algoritma tersebut dalam mengenali sentimen ulasan. Evaluasi dilakukan dengan menggunakan *Confusion Matrix*, yang berfungsi untuk membandingkan hasil klasifikasi dengan data asli.

3.3.7 Kesimpulan dan Saran

Tahap ini bertujuan untuk menyimpulkan hasil dari penelitian yang telah dilakukan. Melalui proses analisis yang diterapkan, dapat diketahui bagaimana sentimen pelanggan terhadap layanan Jemberkamera sebagai penyedia jasa penyewaan kamera di kabupaten jember. Temuan ini memberikan gambaran umum mengenai persepsi pelanggan, baik yang bersifat positif, netral maupun negatif. Penelitian ini juga membuka peluang untuk dilakukan studi lanjutan guna memperdalam pembahasan serta memberikan kontribusi yang bermanfaat bagi pembaca dan pihak-pihak terkait.

3.4 Jadwal Pelaksanaan

Perencanaan pelaksanaan kegiatan skripsi ini dengan judul "Analisis Ulasan Pelanggan Pada Sewa Kamera di Jemberkamera Menggunakan Algoritme Naive Bayes" dapat dilihat melalui Tabel 3.1 di bawah ini:

Tabel 3. 1 Jadwal Pelaksanaan

No	Keterangan	2025		2026					
		11	12	1	2	3	4	5	6
1	Indetifikasi Masalah	■							
2	Analisis Kebutuhan		■						
3	Pengumpulan Data			■					
4	Perancangan Sistem				■	■	■		
5	Implementasi dan Pengujian						■	■	
6	Hasil dan Pembahasan							■	■
7	Kesimpulan dan Saran								■

BAB 4. HASIL DAN PEMBAHASAN

4.1 Identifikasi Masalah

Penelitian mengidentifikasi masalah, yaitu beberapa hal yang dikesankan oleh pelanggan melewati ulasan di *Google Maps*. Terdapat ketidaksesuaian peralatan sewa alat yang dipromosikan di sosial media melainkan seperti, kamera, tripod, *handphone*, dan kualitas pelayanan mitra. Objek dalam penelitian ini adalah Jemberkamera sebagai penyedia layanan persewaan alat dan peralatan fotografi di Kabupaten Jember. Identifikasi permasalahan dilakukan dengan mengkaji berbagai kendala yang dihadapi pelanggan dalam memilih serta menggunakan layanan persewaan tersebut melalui analisis ulasan yang diberikan oleh pelanggan. Ulasan yang bersifat positif, netral maupun negatif digunakan sebagai sumber data untuk mengidentifikasi faktor-faktor yang memengaruhi tingkat kepuasan pelanggan, Melalui penelitian ini, dapat mempengaruhi pelanggan dalam penyewaan alat dan pihak mitra dapat melakukan pengembangan.

4.1.1 Wawancara Penelitian Dengan Mitra

Peneliti melakukan wawancara dengan pihak mitra sebagai langkah untuk memperoleh informasi awal yang relevan dengan penelitian yang dilakukan. Wawancara ini bertujuan untuk mengetahui gambaran umum mengenai kondisi mitra, khususnya terkait dengan ulasan pelanggan yang terdapat pada platform *Google Maps* serta bagaimana ulasan tersebut dapat dimanfaatkan sebagai bahan analisis dalam penelitian. Kegiatan wawancara dilakukan secara langsung dengan pihak mitra untuk mendapatkan persetujuan serta dukungan terhadap pelaksanaan penelitian. Dalam proses tersebut, peneliti juga menjelaskan tujuan penelitian, metode yang akan digunakan, serta manfaat yang diharapkan dari hasil penelitian bagi mitra.

Penulis juga menyertakan surat pernyataan berisi kesediaan mitra untuk dijadikan objek penelitian serta meminta izin kepada peneliti untuk mengambil dan menggunakan data ulasan pelanggan pada *Google Maps* sebagai bahan penelitian. Disisilain, penulis mendokumentasi ini bertujuan untuk memperkuat validasi

penelitian serta menunjukkan bahwa penelitian telah dilakukan dengan izin dan sepengetahuan pihak mitra pada gambar 4.1 dan 4.2.

KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
POLITEKNIK NEGERI JEMBER
Jalan Mastrip Kotak Pos 164 Jember 68121 Telp. (0331) 333532-34
Email : politeknik@polije.ac.id Website : https://www.polije.ac.id

Nomor : 5874 /DST.PL17/PP.02.10/2026 01 APR 2026
Perihal : Permohonan Ijin Survei dan Pengambilan Data

Kepada Yth.
Pimpinan Jember Kamera
Jl. Bangka VI No.25, Tegal Boto Lor, Sumbersari, Kec. Sumbersari, Kabupaten Jember, Jawa Timur 68121
Di
Tempat

Dalam rangka penyelenggaraan pendidikan Politeknik Negeri Jember yang berorientasi pada pendidikan profesional, mahasiswa wajib melaksanakan Tugas Akhir / Skripsi sebagai salah satu syarat kelulusan.

Sehubungan dengan hal tersebut mohon Bapak / Ibu berkenan mengijinkan mahasiswa kami dari Program Studi Sarjana Terapan - Teknik Informatika melakukan survei guna mendapatkan data dan informasi yang kompeten sesuai dengan bidang kajiannya di Instansi / perusahaan yang Bapak / Ibu pimpin.

Adapun mahasiswa yang dimaksud adalah :

Nama Mahasiswa	NIM	Judul Skripsi
Moh. Bachtiar Rifki Habibi	E41220474	Analisis Ulasan Pelanggan Pada Sewa Kamera di Jemberkamera Menggunakan Algoritme Naive Bayes

Konfirmasi kesediaan Bapak/Ibu untuk menerima ijin survey mahasiswa kami dapat disampaikan pada Sdri. Dia Bitari Mei Yuana, S.ST.,M.Tr.Kom dengan no Hp. 089699730305 selaku Koordinator Bidang Tugas Akhir/Skripsi Program Studi Sarjana Terapan - Teknik Informatika Politeknik Negeri Jember.

Demikian atas kebijakan dan kerjasama yang baik dari Bapak/Ibu dalam turut serta menunjang peningkatan keterampilan anak didik kami, diucapkan terima kasih.

an. Direktur
Wakil Direktur Bidang Akademik dan Perencanaan

M.Kom
NIP. 19[illegible]03 200312 1 001

Gambar 4. 1 Surat Izin Observasi Mitra

Gambar 4. 2 Dokumentasi Observasi Bersama Mitra

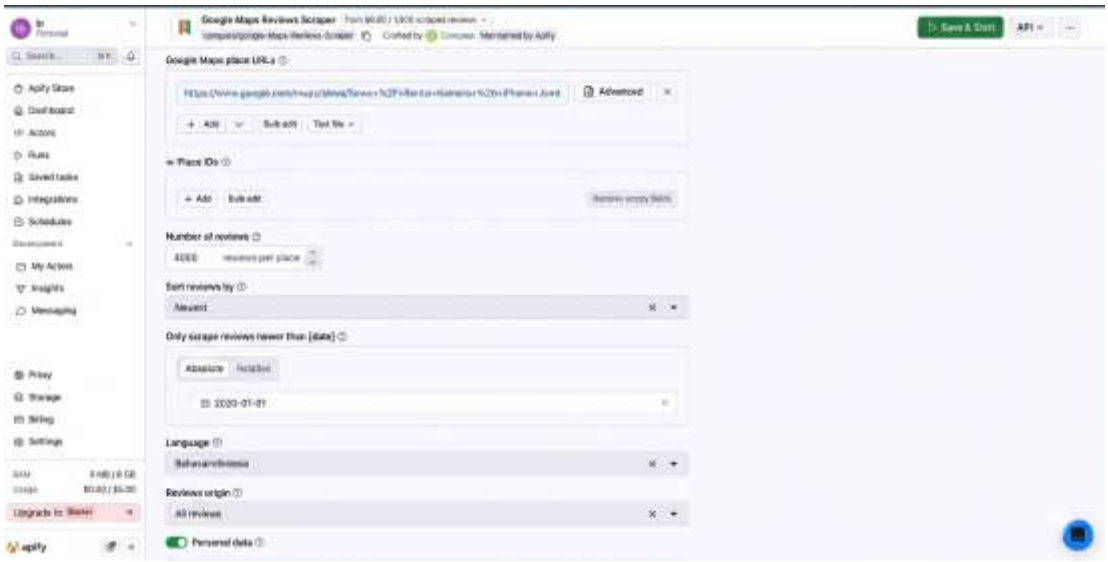
4.2 Analisis Kebutuhan

Analisis kebutuhan dilakukan untuk mengidentifikasi berbagai kebutuhan yang diperlukan dalam pengembangan sistem yang akan dibangun. Pada penelitian ini, penulis merancang sebuah *website* yang digunakan untuk melakukan proses klasifikasi teks terhadap ulasan pelanggan Jemberkamera yang terdapat pada *Google Maps*. Sistem yang dikembangkan diharapkan mampu membantu dalam mengelola data ulasan pelanggan mulai dari proses pengunggahan data, *Preprocessing*, hingga proses klasifikasi sentimen menggunakan metode yang telah ditentukan. Dengan adanya sistem ini, hasil analisis sentimen dapat ditampilkan secara lebih terstruktur sehingga dapat memberikan gambaran mengenai persepsi pelanggan terhadap layanan yang diberikan oleh Jemberkamera.

4.3 Pengumpulan Data

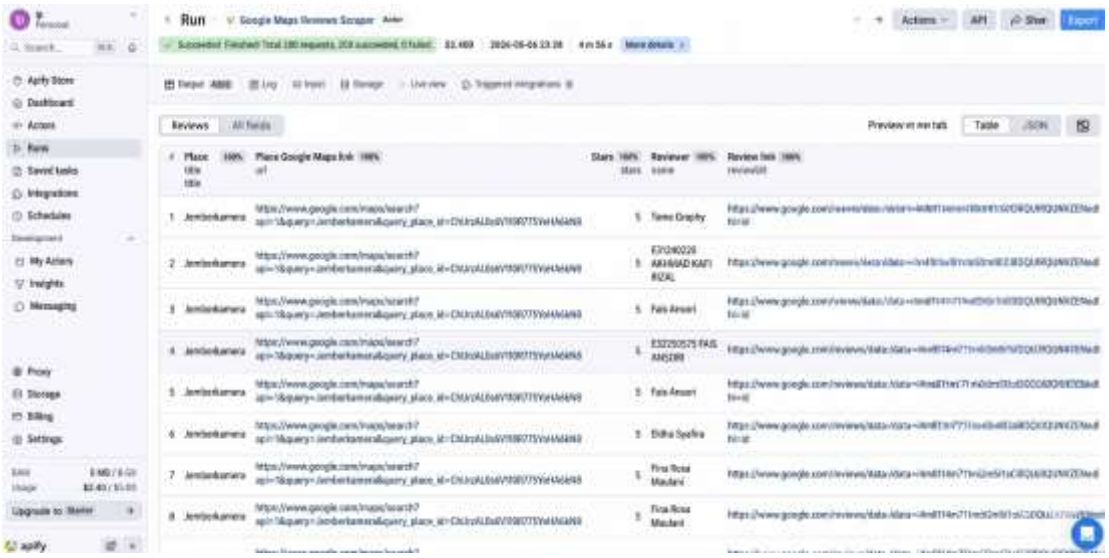
Proses pengumpulan data dilakukan dengan mengambil data ulasan pelanggan yang terdapat pada platform *Google Maps*. Pengambilan data dilakukan dengan memanfaatkan layanan web *scraping* melalui *website* Apify, dengan

menggunakan fitur *Google Maps Reviews Scraper* yang tersedia pada platform tersebut. Melalui *tools* tersebut, peneliti dapat mengumpulkan data ulasan pelanggan yang diberikan kepada mitra penelitian. Data yang diambil berupa teks ulasan, nama pengguna, tanggal ulasan, serta rating yang diberikan oleh pelanggan terhadap layanan yang tersedia. Pengambilan data ulasan dilakukan mulai dari tanggal 1 Januari 2020 hingga data terbaru pada saat penelitian dilakukan. Rentang waktu tersebut dipilih karena ulasan yang tersedia pada periode tersebut masih dianggap relevan dalam menggambarkan pengalaman pelanggan terhadap layanan terlihat pada gambar 4.3.



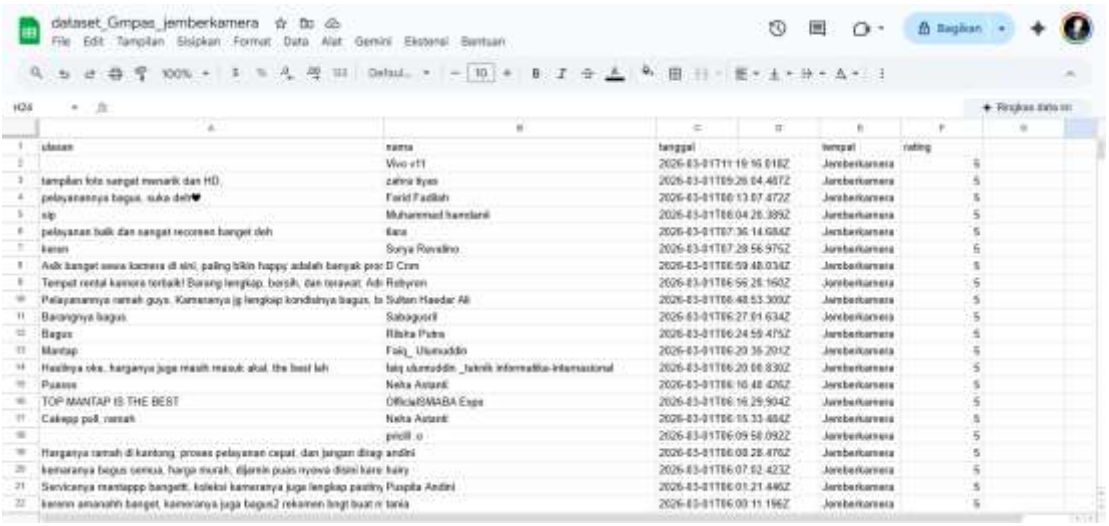
Gambar 4. 3 *Google Maps* Reviews Scraper Platform Apify

Berdasarkan proses *scraping* pada gambar 4.3 yang telah dilakukan, jumlah data ulasan yang berhasil dikumpulkan sebanyak 3.933 ulasan dalam rentang waktu kurang lebih lima tahun. Data tersebut selanjutnya digunakan sebagai dataset dalam proses analisis sentimen untuk mengetahui kecenderungan opini pelanggan terhadap layanan yang diberikan oleh mitra penelitian. Hasil *scraping* bisa di lihat pada gambar 4.3.



Gambar 4. 4 *Scraping* Data Review Apify

Kemudian selesai proses *scraping* lalu di simpan dalam format CSV *(Comma Separated Values)*. Hasil *scraping* yang telah diunduh di ekstrak ke dalam *spreadsheets* dapat dilihat pada gambar 4.5.



Gambar 4. 5 Hasil Data Terkumpul Review *Google Maps*

4.4 Prepocessing Data

Pada tahap penelitian ini, data yang diperoleh diproses *pre-processing* untuk meningkatkan kualitas data sebelum dilakukan klasifikasi. Proses ini bertujuan untuk membersihkan, menyusun, dan menyesuaikan data agar lebih terstruktur serta mudah dianalisis. Tahapan yang dilakukan meliputi *cleaning data* untuk menghapus

data yang tidak relevan atau data ulasan yang kosong, *case folding* untuk menyeragamkan huruf menjadi kecil, *tokenizing* untuk memecah teks menjadi kata-kata, *normalization* untuk mengubah kata tidak baku menjadi baku, *stopword removal* untuk menghilangkan kata umum yang tidak memiliki makna penting, serta *stemming* untuk mengubah kata menjadi bentuk dasar (Aziza dan Noor, 2026). Dengan adanya proses ini, data menjadi lebih optimal dan siap digunakan dalam proses klasifikasi. Hasil data ulasan pelenggan yang diperoleh dari hasil *scraping* melewati platform bisa dilihat pada tabel 4.1 Sebeleum diproses ke tahap *preprocessing*.

Tabel 4. 1 Hasil *Scraping* Data *Google Maps*

No	Nama	Ulasan	*Rating*
1.	rasifa aulia	pelayanan nya oke banget sukaaa	5
2.	Safina Azzahra	Recommended bangettt buat kalian yg pengen hunting foto dan murahh 👍👍👍	5
3.	Shelsie Malini Putri Y	pelayanannya bagus bgt, masnya mau jelaskan cara pake kamera	5

4.4.1 *Cleaning data* (Pembersihan Data)

Tahap pertama dalam proses *pre-processing* adalah *cleaning data*, yang bertujuan untuk membersihkan teks ulasan dari berbagai elemen yang tidak relevan. Proses ini dilakukan dengan menghapus simbol, tanda baca, angka, URL, serta karakter lain yang tidak memiliki kontribusi terhadap analisis sentimen. Dengan dilakukannya *cleaning data*, diharapkan teks ulasan menjadi lebih fokus pada informasi utama sehingga dapat meningkatkan kualitas analisis dan menghasilkan data yang lebih representatif untuk proses klasifikasi selanjutnya. Berikut merupakan kode program yang digunakan dalam proses *cleaning data*.

Kode Program 4. 1 *Cleaning data*

```
1) import pandas as pd
2) import re
3) def clean_text(text):
```

```
4) text = str(text)
5) text = text.lower()
6) text = re.sub(r'http\S+', '', text
7) text = re.sub(r'@\w+', '', text)
8) text = re.sub(r'#\w+', '', text)
9) text = re.sub(r'\d+', '', text)
10)  text = re.sub(r'[^\x00-\x7F]+', '', text)
11)  text = re.sub(r'[^\w\s]', '', text)
12)  text = re.sub(r'\s+', ' ', text ).strip()
13)  return text
14)  df['ulasan'] = df['ulasan'].apply(clean_text)
15)  print("Jumlah data setelah Cleaning:", len(df))
```

Setelah dilakukan *cleaning data* pada Kode Program 4.1, terlihat perbedaan antara data sebelum dan sesudah pembersihan. Data menjadi lebih bersih dari karakter yang tidak diperlukan sehingga siap untuk proses selanjutnya. Perbandingan tersebut ditampilkan pada Tabel 4.2.

Tabel 4. 2 *Cleaning data*

No	Sebelum *Cleaning data*	Sesudah *Cleaning data*
1.	pelayanan nya oke banget sukaaa	pelayanan nya oke banget sukaaa
2.	Recommended bangettt buat kalian yg pengen hunting foto dan murahh 👍👍👍	Recommended bangettt buat kalian yg pengen hunting foto dan murahh
3.	pelayanannya bagus bgt, masnya mau jelaskan cara pake kamera	pelayanannya bagus bgt masnya mau jelaskan cara pake kamera

4.4.2 *Case Folding*

Tahap selanjutnya yaitu *case folding*, merupakan proses mengubah seluruh huruf kapital menjadi huruf kecil (*lowercase*). Proses ini bertujuan untuk menyeragamkan teks sehingga meningkatkan konsistensi dalam analisis dan mengurangi kompleksitas pemrosesan data. Dengan demikian, hasil klasifikasi diharapkan menjadi lebih akurat. Berikut merupakan kode program c*ase folding*. Berikut merupakan kode program yang digunakan dalam proses *case folding*.

Kode Program 4. 2 *Case Folding*

```
1) import pandas as pd
2) df['ulasan'] = df['ulasan'].astype(str).str.lower()
3) df.to_csv(file_preprocess, index=False)
```

Pada kode program 4.2 merupakan proses untuk mengubah teks yang awalnya mengandung huruf kapital menjadi huruf kecil (*lowercase*). Tahapan ini dilakukan dengan baik sehingga teks menjadi lebih seragam. Hasil perubahan tersebut dapat dilihat pada Tabel 4.3.

Tabel 4. 3 *Case Folding*

No	Sebelum *Case Folding*	Sesudah *Case Folding*
1.	pelayanan nya oke banget sukaaa	pelayanan nya oke banget sukaaa
2.	Recommended bangettt buat kalian yg pengen hunting foto dan murahh	recommended bangettt buat kalian yg pengen hunting foto dan murahh
3.	pelayanannya bagus bgt masnya mau jelaskan cara pake kamera	pelayanannya bagus bgt masnya mau jelaskan cara pake kamera

4.4.3 Normalisasi

Tahap selanjutnya adalah normalisasi, yaitu proses mentransformasikan kata-kata tidak baku, penulisan gaul, maupun singkatan yang terdapat dalam ulasan menjadi bentuk kata yang seragam. Proses ini bertujuan untuk meningkatkan konsistensi data teks sehingga mempermudah proses pembobotan dan klasifikasi pada tahap berikutnya. Berikut merupakan kode program yang digunakan dalam proses normalisasi:

Kode Program 4. 3 Normalisasi

```
1) import pandas as pd
2) import re
3) kamus_normalisasi = {"bgt": "banget", "bgtt": "banget", "gk":
   "tidak", "ga": "tidak", "nggak": "tidak", "tdk": "tidak", "sm":
   "sama", "tp": "tapi", "dr": "dari", "yg": "yang", "dgn":
   "dengan", "udh": "sudah", "blm": "belum", "krn": "karena",
   "mantaaap": "mantap"}
4) kamus_normalisasi = {"reqomended": "recommended", "rekomended":
   "recommended", "recommendedd": "recommended", "banget":
   "banget", "kereen": "keren"}
5) def normalize_text(text):
6) text = str(text)
7) #Reducing repeated characters (SEMUA huruf dobel jadi 1)
```

```
8) text = re.sub(r'(.)\1+', r'\1', text)
9) #Normalisasi berdasarkan kamus
10)  kata_kata = text.split()
11)  hasil = []
12)  for kata in kata_kata:
13)  if kata in kamus_normalisasi:
14)  hasil.append(kamus_normalisasi[kata])
15)  else:
16)  hasil.append(kata)
17)  return " ".join(hasil)
18)  df['ulasan'] = df['ulasan'].apply(normalize_text)
```

Kode Program 4.3, proses ini berjalan dalam dua tahap. pertama, menggunakan Regular Expression re.sub(r'(.)\1+', r'\1', text) untuk mereduksi huruf berulang (seperti "kereen" menjadi "keren"); kedua, melakukan pemetaan manual menggunakan *dictionary Python* kamus_normalisasi untuk mengonversi singkatan seperti "bgt" menjadi "banget" dan "gk" menjadi "tidak" serta istilah serapan seperti "rekomended" menjadi "recommended". Melalui kombinasi reduksi huruf kembar dan pencocokan kamus buatan peneliti ini, teks ulasan menjadi lebih bersih serta konsisten, dengan detail hasil perubahan yang disajikan pada Tabel 4.4.

Tabel 4. 4 Normalisasi

No	Sebelum Normalisasi	Sesudah Normalisasi
1.	pelayanan nya oke banget sukaaa	pelayanan nya oke banget suka
2.	Recommended bangettt buat kalian yg pengen hunting foto dan murahh	rekomendasi banget buat kalian yg pengen hunting foto dan murah
3.	pelayanannya bagus bgt masnya mau jelaskan cara pake kamera	pelayanan bagus banget masnya mau jelaskan cara pake kamera

4.4.4 *Tokenizing*

Tahap Selanjutnya yaitu *Tokenizing* merupakan proses memecah teks menjadi bagian-bagian kecil berupa kata atau token. Tahap ini bertujuan untuk memudahkan analisis lebih lanjut karena setiap kata akan diproses secara terpisah dalam sistem. Berikut merupakan kode program yang digunakan dalam proses *Tokenizing*.

Kode Program 4. 4 *Tokenizing*

```
1) import pandas as pd
2) # Buat kolom baru 'tokens'
3) df['tokens'] = df['ulasan'].apply(lambda x: str(x).split())
4) # Cek contoh Tokenizing
5) print("\nContoh Tokenizing 5 data pertama:\n")
6) print(df[['ulasan', 'tokens']].head(5))
7) df.to_csv(file_preprocess, index=False)
```

Pada kode program 4.4 proses *Tokenizing* yang berfungsi untuk memecah kalimat menjadi unit-unit kecil (token) berupa kata. Metode ini dipilih karena sederhana dan efisien dalam memproses data teks ulasan, sehingga dapat membantu sistem dalam memahami struktur kata untuk proses analisis selanjutnya. Hasil perubahan tersebut dapat dilihat pada Tabel 4.5.

Tabel 4. 5 *Tokenizing*

No	Sebelum *Tokenizing*	Sesudah *Tokenizing*
1.	pelayanan nya oke banget suka	pelayanan | nya | oke | banget | suka
2.	rekomendasi banget buat kalian yg pengen hunting foto dan murah	rekomendasi | banget | buat | kalian | yg | pengen | hunting | foto | dan | murah
3.	pelayanan bagus banget masnya mau jelaskan cara pake kamera	pelayanan | bagus | banget | masnya | mau | jelaskan | cara | pake | kamera

4.4.5 *Stopword Removal*

Tahap *Stopword Removal* merupakan proses untuk menghapus kata-kata umum yang sering muncul namun tidak memiliki makna penting dalam analisis, seperti “dan”, “yang”, “di”, dan sejenisnya. Tahap ini bertujuan untuk mengurangi jumlah kata yang tidak relevan sehingga data menjadi lebih fokus dan dapat meningkatkan kualitas hasil analisis. Kode Program 4.5 menampilkan proses stop removal pada penelitian ini.

Kode Program 4. 5 *Stopword Removal*

```
1) !pip install Sastrawi
2) import pandas as pd
3) from Sastrawi.StopWordRemover.StopWordRemoverFactory import
   StopWordRemoverFactory
4) factory = StopWordRemoverFactory()
5) stopwords = set(factory.get_stop_words())
6) custom_stopwords =
   {"sangat","banget","sekali","deh","nih","dong","aja",
   "nya","lah","kok","pun","kan", "is","the","very","so","really"}
7) stopwords = stopwords.union(custom_stopwords)
8) def remove_stopwords(text):
9) words = str(text).split()
10)  filtered_words = [word for word in words if word not in
   stopwords]
11)  return " ".join(filtered_words)
12)  # Terapkan Stopword Removal
13)  df['ulasan'] = df['ulasan'].apply(remove_stopwords)
```

Pada kode program 4.5 proses *Stopword Removal* dilakukan menggunakan library Sastrawi untuk menghapus kata-kata umum yang tidak memiliki makna penting. Selain itu, ditambahkan stopword khusus yang sering muncul pada ulasan agar data menjadi lebih bersih dan fokus untuk proses analisis selanjutnya. Hasil perubahan tersebut dapat dilihat pada Tabel 4.6.

Tabel 4. 6 *Stopword Removal*

No	Sebelum *Stopword Removal*	Sesudah *Stopword Removal*
1.	pelayanan | nya | oke | banget | suka	pelayanan | oke | banget | suka
2.	rekomendasi | banget | buat | kalian | yg | pengen | hunting | foto | dan | murah	rekomendasi | banget | hunting | foto | murah
3.	pelayanan | bagus | banget | masnya | mau | jelaskan | cara | pake | kamera	pelayanan | bagus | banget | masnya | jelaskan | cara | pake | kamera

4.4.6 *Stemming*

Stemming merupakan proses mengubah kata menjadi bentuk dasar (kata akar) dengan menghilangkan imbuhan seperti awalan, akhiran, atau sisipan. Tahap ini bertujuan untuk menyederhanakan variasi kata sehingga mempermudah proses analisis dan meningkatkan konsistensi data. Berikut merupakan kode program yang digunakan *Stemming*.

Kode Program 4. 6 *Stemming*

```
1) import pandas as pd
2) from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
3) # Load file Preprocessing
4) file_preprocess = "data/dataset_Preprocessing.csv"
5) df = pd.read_csv(file_preprocess)
6) factory = StemmerFactory()
7) stemmer = factory.create_stemmer()
8) def Stemming_text(text):
9) return stemmer.stem(str(text))
10)  # Terapkan Stemming
11)  df['ulasan'] = df['ulasan'].apply(Stemming_text)
```

Pada kode program 4.6 proses *Stemming* dilakukan menggunakan *library* Sastrawi untuk mengubah kata menjadi bentuk dasarnya. Proses ini membantu menyederhanakan variasi kata sehingga data menjadi lebih konsisten dan mudah dianalisis. Hasil perubahan tersebut dapat dilihat pada Tabel 4.7.

Tabel 4. 7 *Stemming*

No	Sebelum *Stemming*	Sesudah *Stemming*
1.	pelayanan | oke | banget | suka	pelayanan oke banget suka
2.	rekomendasi | banget | hunting | foto | murah	rekomendasi banget hunting foto murah
3.	pelayanan | bagus | banget | masnya | jelaskan | cara | pake | kamera	pelayanan bagus banget mas jelas cara pake kamera

Berdasarkan data data asli ulasan pelanggan Jemberkamera hasil ekstraksi platform Apify yang berjumlah sebanyak 3.933 data, sistem secara otomatis melakukan penyaringan awal pada mengunggah data cvs dengan memotong 1.041 ulasan kosong yang hanya berisi rating bintang. Proses ini menyisakan sebanyak 2.892 data terunggah yang berhasil masuk ke dalam database sistem. Selanjutnya, ketika tahap preprocessing dieksekusi, pada proses cleaning ini kembali melakukan pembersihan intensif dengan menghapus sebanyak 348 data duplikat, 39 data ulasan yang terlalu pendek, serta 13 data ulasan dengan hasil stemming kosong. Hasil data yang diperloeh sebanyak 2.492 baris ulasan murni yang benar-benar valid, konsisten.

4.5 Pembobotan Kata (TF-IDF)

Metode TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan untuk memberikan bobot pada setiap kata. Dalam metode ini, pembobotan dilakukan melalui beberapa tahap, yaitu perhitungan TF (*Term Frequency*) untuk mengetahui tingkat kemunculan kata dalam suatu dokumen, DF (*Document Frequency*) untuk mengetahui jumlah dokumen yang mengandung kata tersebut, serta IDF (*Inverse Document Frequency*) untuk mengukur tingkat kepentingan kata dalam keseluruhan kumpulan dokumen. Dengan demikian, proses perhitungan TF-IDF dimulai dari tahapan-tahapan tersebut.

4.5.1 Perhitungan TF (*Term Frequency*)

Langkah pertama dalam proses pembobotan kata adalah menentukan nilai frekuensi kata (TF) *Term Frequency*. Perhitungan ini bertujuan untuk mengetahui seberapa sering suatu kata muncul dalam dokumen tertentu. Informasi tersebut digunakan untuk mengidentifikasi kata kunci yang dianggap penting dalam data. Nilai TF juga dimanfaatkan untuk mengetahui tingkat relevansi suatu kata terhadap isi dokumen. Tabel 4.8 menunjukkan contoh penggunaan perhitungan *Term Frequency* berdasarkan data sampel.

Tabel 4. 8 Data Sampel Perhitungan TF-IDF

Dokumen	Hasil prepocessing	Manual Label
Doc1	pelayanan oke banget suka	positif
Doc2	rekomendasi	positif
Doc3	rekomendasi banget hunting foto murah	positif
Doc4	terbaik banget	positif
Doc5	rekomendasi banget	positif
Doc6	sempet sewa iphone ternyata semua tanpa tempered glas harus lebih hati	netral
Doc7	murah sedikit lebih mahal	netral
Doc8	sewa kamera sony a6300 lensa 50m hasil lensa kit fix tidak tajam blur	negatif

Dokumen	Hasil prepocessing	Manual Label
Doc9	karyaw gak jujur harga layanan 150rb layanan trouble total 350rb harus	negatif
Doc10	agak kurang teliti kemarin sewa tripod tidak holder padahal rencana	negatif

Setelah proses labelling dihasilkan setiap ulasan menjadi dokumen dan seluruh tabel 4.8 diatas tahapan pre-processing diselesaikan dan kata-kata pada setiap dokumen berhasil diekstraksi, proses berikutnya adalah pemberian bobot pada dokumen dengan menggunakan *Term Frequency* (TF). Hal ini digunakan untuk menilai tingkat kepentingan suatu kata dalam sebuah dokumen dengan mempertimbangkan kemunculannya pada keseluruhan kumpulan dokumen.

Proses pembentukan term dilakukan menggunakan metode *TfidfVectorizer*, di mana seluruh kata hasil *preprocessing* dikumpulkan menjadi *vocabulary*. Jumlah term yang ditampilkan pada sistem merupakan hasil dari data sampel, sedangkan pada keseluruhan dataset jumlah term yang dihasilkan jauh lebih banyak. penulis hanya ditampilkan 10 term dengan nilai TF-IDF tertinggi pada setiap dokumen sebagai bentuk ringkasan hasil. Dapat ditampilkan dalam bentuk tabel 4.9 sebagai hasil pembobotan.

Tabel 4. 9 Hasil Frekuensi Kemunculan Setiap Kata Dalam Dokumen

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
rekomendasi	0	1	1	0	1	0	0	0	0	0
terbaik	0	0	0	1	0	0	0	0	0	0
pelayanan	1	0	0	0	0	0	0	0	0	0
oke	1	0	0	0	0	0	0	0	0	0
suka	1	0	0	0	0	0	0	0	0	0
sedikit	0	0	0	0	0	0	1	0	0	0
mahal	0	0	0	0	0	0	1	0	0	0

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
hunting	0	0	1	0	0	0	0	0	0	0
foto	0	0	1	0	0	0	0	0	0	0
murah	0	0	1	0	0	0	1	0	0	0

Nilai frekuensi terma pada setiap term dalam masing-masing dokumen diperoleh melalui proses perhitungan menggunakan rumus frekuensi terma sesuai pesamaan (2.1). Perhitungan tersebut digunakan untuk mengetahui jumlah kemunculan setiap kata, yang kemudian hasilnya disajikan dalam bentuk tabel 4.10.

$$TF(rekomendasi, doc1) = \frac{1}{4} = 0{,}25$$

Rumus yang digunakan pada persamaan (2.1) menunjukkan hasil perhitungan *Term Frequency* (TF) untuk setiap kata pada dokumen yang dianalisis. Berdasarkan hasil tersebut, tidak semua kata muncul pada setiap dokumen. Sebagai contoh, kata "rekomendasi" muncul di beberapa dokumen seperti Doc2, Doc3, dan Doc5, sehingga memiliki nilai frekuensi kemunculan yang lebih tinggi dibandingkan kata lainnya. Sebaliknya, terdapat kata-kata seperti "pelayanan", "oke", dan "suka" yang hanya muncul pada satu dokumen tertentu (Doc1), sehingga nilai TF-nya bernilai nol pada dokumen lainnya. Hal ini menunjukkan bahwa distribusi kemunculan kata tidak merata di setiap dokumen.

Selain itu, perbedaan jumlah kata dalam setiap dokumen juga memengaruhi nilai TF yang dihasilkan. Sebagai ilustrasi, kata "rekomendasi" muncul pada Doc2 dengan nilai TF sebesar 1, namun pada Doc5 nilainya menjadi 0,5 meskipun frekuensi kemunculannya sama-sama satu kali, hal ini dikarenakan adanya perbedaan total kata dalam dokumen tersebut. Hal yang sama juga terjadi pada kata-kata lain seperti "murah" yang memiliki nilai TF sebesar 0,2 pada Doc3 dan 0,25 pada Doc7, serta kata "terbaik" yang tidak selalu muncul di seluruh dokumen.

Tabel 4. 10 Hasil Perhitungan TF

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
rekomendasi	0	1	0,2	0	0,5	0	0	0	0	0

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
terbaik	0	0	0	0,5	0	0	0	0	0	0
pelayanan	0,25	0	0	0	0	0	0	0	0	0
oke	0,25	0	0	0	0	0	0	0	0	0
suka	0,25	0	0	0	0	0	0	0	0	0
sedikit	0	0	0	0	0	0	0,25	0	0	0
mahal	0	0	0	0	0	0	0,25	0	0	0
hunting	0	0	0,2	0	0	0	0	0	0	0
foto	0	0	0,2	0	0	0	0	0	0	0
murah	0	0	0,2	0	0	0	0,25	0	0	0

4.5.2 Perhitungan *Inverse Document Frequency* (IDF)

Dokumen *Frequency* (DF) digunakan untuk mengetahui tingkat penyebaran suatu kata dalam seluruh kumpulan dokumen. Nilai DF diperoleh dengan menghitung jumlah dokumen yang mengandung kata tertentu. Semakin banyak dokumen yang memuat kata tersebut, maka kontribusinya terhadap bobot TF-IDF akan semakin kecil karena kata tersebut dianggap kurang spesifik. Sebaliknya, kata yang hanya muncul pada sedikit dokumen akan memiliki nilai DF yang rendah dan cenderung memiliki bobot yang lebih besar dalam proses pembobotan TF-IDF. Dengan demikian, DF berperan dalam mengidentifikasi kata-kata yang umum maupun yang lebih spesifik dalam dataset. Pada penelitian ini, nilai DF dihitung berdasarkan seluruh dataset yang digunakan, sedangkan hasil yang ditampilkan pada tabel 4.11 hanya merupakan sebagian data sebagai sampel untuk mempermudah penyajian dan analisis.

Tabel 4. 11 Hasil Perhitungan DF

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	DF
rekomendasi	0	1	1	0	1	0	0	0	0	0	3
terbaik	0	0	0	1	0	0	0	0	0	0	1
pelayanan	1	0	0	0	0	0	0	0	0	0	1

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	DF
oke	1	0	0	0	0	0	0	0	0	0	1
suka	1	0	0	0	0	0	0	0	0	0	1
sedikit	0	0	0	0	0	0	1	0	0	0	1
mahal	0	0	0	0	0	0	1	0	0	0	1
hunting	0	0	1	0	0	0	0	0	0	0	1
foto	0	0	1	0	0	0	0	0	0	0	1
murah	0	0	1	0	0	0	1	0	0	0	2

Setelah nilai *Document Frequency* (DF) diperoleh, tahap selanjutnya adalah menghitung *Inverse Document Frequency* (IDF) untuk mengetahui tingkat kepentingan suatu kata dalam keseluruhan dokumen. Kata yang jarang muncul akan memiliki nilai IDF lebih tinggi karena dianggap lebih spesifik, sedangkan kata yang sering muncul akan memiliki nilai IDF lebih rendah. Perhitungan ini menggunakan sesuai persamaan (2.2) dengan bertujuan untuk mengurangi pengaruh kata umum dan menonjolkan kata yang lebih relevan dalam membedakan isi dokumen.

$$IDF(rekomendasi) = log\left(\frac{10}{3}\right) = 0,523$$

Untuk term kata "rekomendasi" menggunakan rumus logaritma menghasilkan nilai bobot sebesar 0,523. Secara keseluruhan, ringkasan hasil perhitungan bobot nilai IDF untuk setiap token kata lainnya di dalam sistem disajikan lengkap pada Tabel 4.12.

Tabel 4. 12 Hasil Perhitungan IDF

Term	**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**	**DF**	**IDF**
rekomendasi	0	1	1	0	1	0	0	0	0	0	3	0,523
terbaik	0	0	0	1	0	0	0	0	0	0	1	1,000
pelayanan	1	0	0	0	0	0	0	0	0	0	1	1,000
oke	1	0	0	0	0	0	0	0	0	0	1	1,000
suka	1	0	0	0	0	0	0	0	0	0	1	1,000

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	DF	IDF
sedikit	0	0	0	0	0	0	1	0	0	0	1	1,000
mahal	0	0	0	0	0	0	1	0	0	0	1	1,000
hunting	0	0	1	0	0	0	0	0	0	0	1	1,000
foto	0	0	1	0	0	0	0	0	0	0	1	1,000
murah	0	0	1	0	0	0	1	0	0	0	2	0,699

4.5.3 Perhitungan *Term Frequency-Inverse Document Frequency* (TF-IDF)

Perhitungan TF-IDF digunakan untuk memberikan bobot lebih besar pada kata yang sering muncul dalam suatu dokumen namun jarang muncul pada keseluruhan dokumen. Sebagai contoh, pada kata “rekomendasi” di dokumen Doc5, nilai TF sebesar 0,5 dikalikan dengan nilai IDF sebesar 0,523, sehingga menghasilkan nilai TF-IDF sebesar 0,262. Contoh lain dapat dilihat pada kata “terbaik” di Doc4, di mana nilai TF sebesar 0,5 dikalikan dengan nilai IDF sebesar 1,000 menghasilkan bobot akhir sebesar 0,5. Dengan demikian, nilai TF-IDF untuk setiap kata dapat dihitung berdasarkan hasil perkalian antara nilai frekuensi kata *Term Frequency* dan nilai kebalikan frekuensi dokumen *Inverse Document Frequency* tersebut mengikuti persamaan (2.3).

$$TF - IDF(rekomendasi, doc1) = TF(0) \times (0{,}523) = 0$$

Berikut hasl perhitungan TF-IDF terlihat pada tabel 4.13.

Tabel 4. 13 Hasil Perhitungan TF-IDF

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
rekomendasi	0	0,523	0,105	0	0,262	0	0	0	0	0
terbaik	0	0	0	0,5	0	0	0	0	0	0
pelayanan	0,25	0	0	0	0	0	0	0	0	0
oke	0,25	0	0	0	0	0	0	0	0	0
suka	0,25	0	0	0	0	0	0	0	0	0
sedikit	0	0	0	0	0	0	0,25	0	0	0
mahal	0	0	0	0	0	0	0,25	0	0	0

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
hunting	0	0	0,2	0	0	0	0	0	0	0
foto	0	0	0,2	0	0	0	0	0	0	0
murah	0	0	0,14	0	0	0	0,175	0	0	0

Hasil perhitungan TF-IDF digunakan untuk mengetahui bobot kata pada setiap kelas sentimen, yaitu Positif, Netral, dan Negatif. Nilai TF-IDF pada masing-masing kelas dijumlahkan sehingga diperoleh total TF-IDF kelas Positif sebesar 2,68, kelas Netral sebesar 0,675, dan kelas Negatif sebesar 0. Seluruh nilai tersebut kemudian dijumlahkan kembali untuk mendapatkan total keseluruhan TF-IDF sebesar 3,355.

4.6 Implementasi dan Pengujian

4.6.1 Split Data

Pada penelitian ini digunakan beberapa skenario pembagian data *split ratio*, yaitu 80:20, 70:30, 60:40, dan 50:50 antara data *training* dan *testing*. Penentuan jumlah data *training* dan *testing* pada masing-masing kelas sentimen disesuaikan dengan rasio pembagian data yang digunakan. variasi pembagian data ini bertujuan untuk membandingkan performa model klasifikasi pada setiap komposisi data latih dan data uji, sehingga dapat diketahui rasio yang memberikan hasil terbaik terhadap performa model Naive Bayes (Prasetyo dkk., 2024).

Dataset yang digunakan terdiri dari 2.492 data dengan distribusi label, 1.405 data sentimen Positif, 1.029 data sentimen Netral, dan 58 data sentimen negatif. Dataset tersebut dibagi ke dalam dua kelompok utama, yaitu data pelatihan *training* dan data pengujian *testing*. Pada skenario 80:20 diperoleh 1.994 data *training* awal dan 498 data *testing*, di mana setelah penerapan SMOTE data *training* meningkat menjadi 3.411 data. Pada skenario 70:30 diperoleh 1.744 data *training* awal dan 748 data *testing*, dan setelah SMOTE data *training* menjadi 2.976 data. Pada skenario 60:40 diperoleh 1.495 data *training* awal dan 997 data *testing*, dan setelah SMOTE data *training* menjadi 2.550 data. Sedangkan pada skenario 50:50 diperoleh 1.246 data *training* awal dan 1.246 data *testing*, dan setelah SMOTE data

training menjadi 2.109 data. Penggunaan beberapa rasio pembagian data dilakukan sebagai bentuk eksperimen untuk mengetahui pengaruh jumlah data *training* terhadap hasil klasifikasi model, sekaligus untuk mengevaluasi efektivitas penerapan SMOTE dalam menangani ketidakseimbangan kelas pada dataset.

Kode Program 4. 7 Split Data

```
# Mapping split ratio dari frontend Laravel
@if($selectedDataset)
<div class="col-12 col-lg-6"><label class="form-label fw-bold mb-1">Split Ratio</label>
<select class="form-select" id="testRatioSelect" name="test_size">
<option value="">Pilih rasio split data</option>

<option value="0.2"> 80:20 (80% training, 20% testing) </option>
<option value="0.3"> 70:30 (70% training, 30% testing) </option>
<option value="0.4"> 60:40 (60% training, 40% testing) </option>
<option value="0.5"> 50:50 (50% training, 50% testing) </option>

</select>
</div>
@endif
```

Proses pembagian data dilakukan menggunakan fungsi train_test_split dari library Scikit-learn. Parameter test_size digunakan untuk menentukan proporsi data testing sesuai rasio yang dipilih, sedangkan sisa data secara otomatis digunakan sebagai data training. Parameter random_state=42 digunakan agar hasil pembagian data tetap konsisten setiap kali sistem dijalankan. Selain itu, parameter stratify=labels diterapkan untuk menjaga distribusi kelas sentimen pada data training dan testing tetap seimbang. Data training digunakan dalam proses pelatihan model Naive Bayes, sedangkan data testing digunakan untuk mengevaluasi performa model secara objektif berdasarkan nilai *accuracy, precision, recall, dan F1-score.*

4.6.2 Klasifikasi Naive Bayes

Pada tahap klasifikasi menggunakan metode Naive Bayes, dilakukan perhitungan probabilitas prior untuk setiap kelas sentimen, yaitu positif, netral, dan negatif, guna mengetahui kemungkinan awal suatu dokumen termasuk ke dalam kelas tertentu. Selanjutnya, metode ini menghitung probabilitas kemunculan setiap

kata hasil pembobotan TF-IDF terhadap masing-masing kelas sentimen. Kata-kata yang lebih sering muncul pada suatu kelas akan memiliki peluang lebih besar dalam menentukan hasil klasifikasi. Proses klasifikasi dilakukan dengan memilih nilai probabilitas terbesar dari setiap kelas sehingga diperoleh hasil sentimen akhir pada setiap ulasan. Klasifikasi sampel sesuai dengan persamaan (2.5) perhitungan dilihat pada berikut:

P (Positif) = $\frac{5}{10}$ = 0,5

P (Netral) = $\frac{2}{10}$ = 0,2

P (Negatif) = $\frac{3}{10}$ = 0,3

Perhitungan *prior probability* telah dilakukan, tahap selanjutnya dilanjutkan dengan menghitung *likelihood* pada setiap term untuk masing-masing kelas, baik positif maupun negatif, sesuai dengan persamaan (2.6) yang digunakan.

Untuk kata yang tidak ditemukan dalam kamus data latih (term), maka nilai total bobot TF-IDF term tersebut pada setiap kelas dianggap 0.0000. Namun, dengan adanya *Laplace Smoothing* (+1), kata tersebut tetap memiliki probabilitas minimum sehingga proses perkalian pada klasifikasi tetap dapat dilakukan. Berikut merupakan contoh implementasi *likelihood*.

$$P(rekomendasi \mid positif)\ \frac{0{,}523+0{,}105+0{,}262+1}{2{,}68+10} = \frac{1{,}89}{12{,}68} \approx 0{,}149$$

$$P(rekomendasi \mid netral)\ \frac{0+1}{0{,}675+10} = \frac{1}{10{,}675} \approx 0{,}094$$

$$P(rekomendasi \mid negatif)\ \frac{0+1}{0+10} = \frac{1}{10} \approx 0{,}10$$

Hasil keseluruhan perhitungan *likelihood* ditampilkan pada tabel berikut:

Tabel 4. 14 Hasil Perhitungan *Likelihood*

Term	**Positif**	**Netral**	**Negatif**
rekomendasi	0,149	0,094	0,10

terbaik	0,118	0,094	0,10
pelayanan	0,099	0,094	0,10
oke	0,099	0,094	0,10
suka	0,099	0,094	0,10
sedikit	0,079	0,117	0,10
mahal	0,079	0,117	0,10
hunting	0,095	0,094	0,10
foto	0,095	0,094	0,10
murah	0,090	0,110	0,10

Setelah nilai *likelihood* untuk setiap term diperoleh, tahapan berikutnya dilakukan proses klasifikasi menggunakan data uji. Data testing yang telah melalui tahap *Preprocessing* kemudian digunakan untuk menghitung probabilitas pada masing-masing kelas sentimen. Proses pengujian tersebut disajikan pada Tabel 4.15.

Tabel 4. 15 Data Uji

Dokumen	**Hasil Ulasan Prepocessing**	**Label**
Doc11	terbaik sewa kamera	?

Tabel 4.15 menampilkan data uji yang telah selesai melalui tahapan sebelumnya, sehingga dapat digunakan dalam proses pengujian sistem. Setelah itu, dilakukan proses perhitungan klasifikasi untuk menentukan kecenderungan sentimen pada data, apakah termasuk ke dalam kelas positif, netral, atau negatif. Rumus yang dipakai persamaan pada (2.7) yaitu *posterior probability* pada *Naive Bayes* seperti berikut:

$$P(Positif \mid D11 \propto 0{,}118 \times 0{,}079 \times 0{,}079 \times 0{,}5$$
$$P(Positif \mid D11 \propto 0{,}000368219 \times 10^7 = 3682{,}19$$
$$P(Netral \mid D11 \propto 0{,}094 \times 0{,}094 \times 0{,}094 \times 0{,}2$$
$$P(Netral \mid D11 \propto 0{,}0001661168 \times 10^7 = 1661{,}168$$

$$P(Negatif \mid D11 \propto 0{,}10 \times 0{,}10 \times 0{,}10 \times 0{,}3$$
$$P(Negatif \mid D11 \propto 0{,}0003 \times 10^7 = 3000$$

Berdasarkan hasil perhitungan probabilitas menggunakan metode *Naive Bayes*, diperoleh nilai probabilitas untuk kelas positif sebesar 3682,19, kelas netral sebesar 1661,168, dan kelas negatif sebesar 3000. Dari hasil tersebut dapat diketahui bahwa kelas positif memiliki nilai probabilitas paling tinggi dibandingkan kelas netral dan negatif. Oleh karena itu, data ulasan D11 diklasifikasikan ke dalam kelas positif karena memiliki probabilitas terbesar sebagai hasil akhir klasifikasi.

4.6.3 Evaluasi Model

Proses evaluasi model dilakukan menggunakan data *testing* untuk mengetahui kemampuan metode *Naive Bayes* dalam mengklasifikasikan sentimen positif, netral, dan negatif. Dataset yang digunakan pada penelitian ini berjumlah 2.492 data, terdiri dari 1.405 data sentimen Positif, 1.029 data sentimen Netral, dan 58 data sentimen Negatif. Ketidakseimbangan distribusi kelas, khususnya pada kelas Negatif yang hanya berjumlah 58 data, menjadi tantangan utama dalam proses klasifikasi sehingga diterapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menyeimbangkan data *training* sebelum model dilatih. Hasil klasifikasi kemudian dianalisis menggunakan *confusion matrix* untuk membandingkan hasil prediksi model dengan label sebenarnya, sehingga dapat diketahui tingkat performa model secara objektif.

Pada tahap pengujian digunakan beberapa skenario pembagian data *split ratio*, yaitu 80:20, 70:30, 60:40, dan 50:50 antara data *training* dan data *testing*. Pada rasio 80:20 diperoleh 1.994 data *training* awal (3.411 setelah SMOTE) dan 498 data *testing*. Pada rasio 70:30 diperoleh 1.744 data *training* awal 2.976 setelah SMOTE dan 748 data *testing*. Pada rasio 60:40 diperoleh 1.495 data *training* awal 2.550 setelah SMOTE dan 997 data *testing*. Sedangkan pada rasio 50:50 diperoleh 1.246 data *training* awal 2.109 setelah SMOTE dan 1.246 data *testing*. Penggunaan beberapa skenario pembagian data tersebut bertujuan untuk mengetahui pengaruh proporsi data latih dan data uji terhadap kinerja model berdasarkan nilai *accuracy,*

precision, recall, dan *F1-score*, serta untuk menentukan rasio pembagian data yang menghasilkan performa klasifikasi terbaik.

a. Akurasi Skenario data 80:20

Pada skenario pembagian data 80:20, dataset yang digunakan berjumlah 2.492 data, dengan 1.994 data digunakan sebagai data *training* dan 498 data digunakan sebagai data *testing*. Pengujian dilakukan untuk mengetahui kemampuan model Naive Bayes dalam mengklasifikasikan sentimen positif, netral, dan negatif berdasarkan data uji yang telah diproses sebelumnya.

Confusion Matrix Data test

Visualisasi akurasi prediksi per kelas pada data test.

Aktual \ Pred	Positif	Netral	Negatif
Positif	227	33	8
Netral	103	97	13
Negatif	9	1	7

Gambar 4. 6 Hasil Confusion Matrix Skenario 80:20

Berdasarkan hasil pengujian pada Gambar 4.6 menggunakan *confusion matrix*, diperoleh hasil klasifikasi sebagai berikut: dari 268 data aktual Positif, sebanyak 227 diprediksi benar (TP), 33 salah diprediksi sebagai Netral, dan 8 diprediksi sebagai Negatif. Dari 213 data aktual Netral, sebanyak 97 diprediksi benar (TP), 103 diprediksi sebagai Positif, dan 13 diprediksi sebagai Negatif. Dari 17 data aktual Negatif, sebanyak 7 diprediksi benar (TP), 9 diprediksi sebagai Positif, dan 1 diprediksi sebagai Netral.

Perhitungan akurasi menggunakan persamaan (2.8) dilakukan untuk mengetahui tingkat ketepatan model dalam mengklasifikasikan seluruh data *testing*. Sampel perhitungan akurasi dengan skenario 80:20

$$akurasi\ \frac{(227+97+7)}{498} \times\ 100\%$$

$$akurasi\ \frac{331}{498} \times\ 100\%\ =\ 66{,}47\%$$

Hasil perhitungan menunjukkan bahwa model Naive Bayes pada skenario pembagian data 80:20 dengan penerapan metode SMOTE memperoleh nilai akurasi sebesar 66,47%, sebagaimana ditunjukkan pada Gambar 4.11.

b. Akurasi Skenario data 70:30

Pada skenario pembagian data 70:30, dataset yang digunakan berjumlah 2.492 data, dengan 1.744 data digunakan sebagai data training dan 748 data digunakan sebagai data testing. Pengujian dilakukan untuk mengetahui kemampuan model Naive Bayes dalam mengklasifikasikan sentimen positif, netral, dan negatif berdasarkan data uji yang telah diproses sebelumnya.

Confusion Matrix Data test

Visualisasi akurasi prediksi per kelas pada data test.

Aktual \ Pred	Positif	Netral	Negatif
Positif	352	52	9
Netral	153	145	15
Negatif	13	2	7

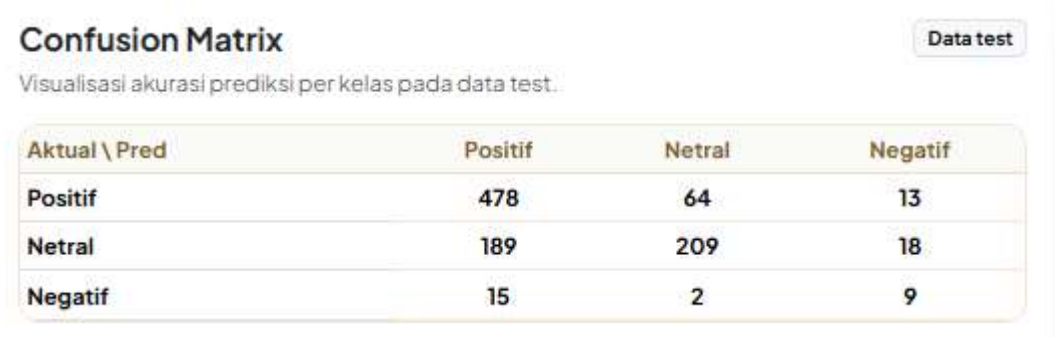
Gambar 4. 7 Hasil Confusion Matrix Skenario 70:30

Berdasarkan hasil pengujian pada Gambar 4.7 menggunakan confusion matrix, diperoleh hasil klasifikasi sebagai berikut: dari 413 data aktual Positif, sebanyak 352 diprediksi benar (TP), 52 salah diprediksi sebagai Netral, dan 9 diprediksi sebagai Negatif. Dari 313 data aktual Netral, sebanyak 145 diprediksi benar (TP), 153 diprediksi sebagai Positif, dan 15 diprediksi sebagai Negatif. Dari 22 data aktual Negatif, sebanyak 7 diprediksi benar (TP), 13 diprediksi sebagai Positif, dan 2 diprediksi sebagai Netral. Hasil perhitungan skenario pembagian data 70:30 dengan penerapan metode SMOTE memperoleh nilai akurasi sebesar 67.38% sebagaimana ditunjukkan pada Gambar 4.12.

c. Akurasi Skenario Data 60:40

Pada skenario pembagian data 60:40, dataset yang digunakan berjumlah 2.492 data, dengan 1.495 data digunakan sebagai data training dan 997 data digunakan sebagai data testing. Pengujian dilakukan untuk mengetahui kemampuan model

Naive Bayes dalam mengklasifikasikan sentimen positif, netral, dan negatif berdasarkan data uji yang telah diproses sebelumnya.



Confusion Matrix Data test

Visualisasi akurasi prediksi per kelas pada data test.

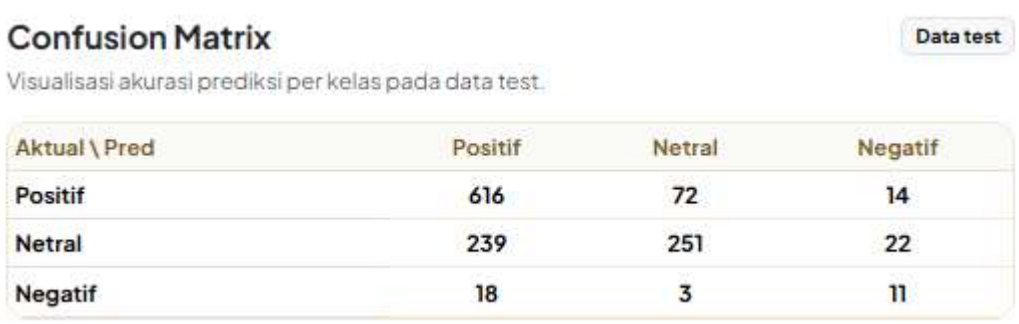
Aktual \ Pred	Positif	Netral	Negatif
Positif	478	64	13
Netral	189	209	18
Negatif	15	2	9

Gambar 4. 8 Hasil Confusion Matrix Skenario 60:40

Berdasarkan hasil pengujian pada Gambar 4.8 menggunakan confusion matrix, diperoleh hasil klasifikasi sebagai berikut: dari 555 data aktual Positif, sebanyak 478 diprediksi benar (TP), 64 salah diprediksi sebagai Netral, dan 13 diprediksi sebagai Negatif. Dari 416 data aktual Netral, sebanyak 209 diprediksi benar (TP), 189 diprediksi sebagai Positif, dan 18 diprediksi sebagai Negatif. Dari 26 data aktual Negatif, sebanyak 9 diprediksi benar (TP), 15 diprediksi sebagai Positif, dan 2 diprediksi sebagai Netral. Hasil perhitungan skenario pembagian data 60:40 dengan penerapan metode SMOTE memperoleh nilai akurasi sebesar 69,81% sebagaimana ditunjukkan pada Gambar 4.13.

d. Akurasi Skenario Data 50:50

Pada skenario pembagian data 50:50, dataset yang digunakan berjumlah 2.492 data, dengan 1.246 data digunakan sebagai data training dan 1.246 data digunakan sebagai data testing. Pengujian dilakukan untuk mengetahui kemampuan model Naive Bayes dalam mengklasifikasikan sentimen positif, netral, dan negatif berdasarkan data uji yang telah diproses sebelumnya.



Confusion Matrix Data test

Visualisasi akurasi prediksi per kelas pada data test.

Aktual \ Pred	Positif	Netral	Negatif
Positif	616	72	14
Netral	239	251	22
Negatif	18	3	11

Gambar 4. 9 Hasil Confusion Matrix Skenario 50:50

Berdasarkan hasil pengujian pada Gambar 4.9 menggunakan confusion matrix, diperoleh hasil klasifikasi sebagai berikut: dari 702 data aktual Positif, sebanyak 616 diprediksi benar (TP), 72 salah diprediksi sebagai Netral, dan 14 diprediksi sebagai Negatif. Dari 512 data aktual Netral, sebanyak 251 diprediksi benar (TP), 239 salah diprediksi sebagai Positif, dan 22 diprediksi sebagai Negatif. Dari 32 data aktual Negatif, sebanyak 11 diprediksi benar (TP), 18 diprediksi sebagai Positif, dan 3 diprediksi sebagai Netral. Hasil perhitungan skenario pembagian data 50:50 dengan penerapan metode SMOTE memperoleh nilai akurasi sebesar 70,47% sebagaimana ditunjukkan pada Gambar 4.14.

e. Precision, Recall dan *F1-score*

Selain akurasi, evaluasi model juga dilakukan menggunakan precision, recall, dan *F1-score* untuk mengetahui performa model secara lebih detail pada setiap kelas sentimen. Precision digunakan untuk mengukur ketepatan model dalam memberikan prediksi terhadap suatu kelas, recall digunakan untuk mengetahui kemampuan model dalam mengenali seluruh data pada kelas tertentu, sedangkan *F1-score* digunakan untuk melihat keseimbangan antara precision dan recall. Contoh perhitungan sampel evaluasi penelitian ini ditunjukkan pada skenario data 80:20 dengan penerapan metode SMOTE.

Perhitungan precision menggunakan persamaan (2.9).

$$Precision(netral)\ \frac{227}{227+112} = \frac{227}{339} = 66{,}96\%$$

$$Precision(netral)\ \frac{97}{97+34} = \frac{97}{131} = 74{,}05\%$$

$$Precision(negatif)\ \frac{7}{7+21} = \frac{7}{?28} = 25{,}00\%$$

Perhitungan recall menggunakan persamaan (2.10)

$$Recall\ (positif)\ \frac{227}{227+41} = \frac{227}{268} = 84{,}70\%$$

$$Recall\ (netral)\ \frac{97}{97+116} = \frac{97}{213} = 45{,}54\%$$

$$Recall\ (negatif)\ \frac{7}{7+10} = \frac{7}{17} = 41{,}18\%$$

Sedangkan Perhitungan *F1-score* menggunakan persamaan (2.11)

$$f1-positif = 2 \times \frac{66{,}96\% \times 84{,}70\%}{66{,}96\% + 84{,}70\%} = 2 \times \frac{5671{,}51}{151{,}66} = 74{,}79\%$$

$$f1-netral = 2 \times \frac{74{,}05\% \times 45{,}54\%}{74{,}05\% + 45{,}54\%} = 2 \times \frac{3372{,}24}{119{,}59} = 56{,}40\%$$

$$f1-negatif = 2 \times \frac{25{,}00\% \times 41{,}18\%}{25{,}00\% + 41{,}18\%} = 2 \times \frac{1029{,}50}{66{,}18} = 31{,}11\%$$

Tabel 4. 16 Hasil Rekapitulasi Pengujian

SMOTE	Precision	Recall	*F1-score*	Akurasi	Skenario
Positif	66,96%	84,70%	74,79%		
Netral	74,05%	45,54%	56,40%	66,47%	80:20
Negatif	25,00%	41,18%	31,11%		
Positif	67,95%	85,23%	75,62%		
Netral	72,86%	46,33%	56,64%	67,38%	70:30
Negatif	22,58%	31,82%	26,42%		
Positif	70,09%	86,13%	77,28%		
Netral	76,00%	50,24%	60,49%	69,81%	60:40
Negatif	22,50%	34,62%	27,27%		
Positif	70,56%	87,75%	78,22%		

SMOTE	Precision	Recall	*F1-score*	Akurasi	Skenario
Netral	76,99%	49,02%	59,90%	70,47%	50:50
Negatif	23,40%	34,38%	27,85%		

Berdasarkan Tabel 4.16 Hasil Rekapitulasi Pengujian, performa model Naive bayes menunjukkan tren peningkatan yang konsisten seiring bertambahnya porsi data uji testing. Akurasi model meningkat bertahap dari 66,47% pada skenario 80:20, menjadi 67,38% pada skenario 70:30, dan mencapai 69,81% pada skenario 60:40. Secara detail per kelas, kelas Positif konsisten mencatatkan *F1-score* tertinggi 74,79%-78,22% karena dominasi jumlah data aktual. Sebaliknya, kelas negatif cenderung tertahan dengan presisi rendah di kisaran 22%–25% akibat keterbatasan variasi kosakata asli pada data awal, meskipun data latihnya sudah dibantu dan diseimbangkan menggunakan metode SMOTE.

Dari seluruh konfigurasi yang diuji, skenario pembagian data 50:50 dengan penerapan SMOTE menjadi skenario terbaik. Konfigurasi ini berhasil mencapai performa puncak dengan nilai Akurasi tertinggi sebesar 70,47%. Keunggulan ini didorong oleh optimalnya model dalam mengenali kelas mayoritas, di mana kelas positif menyentuh recall 87,75% dan kelas netral mencatatkan precision tertinggi sebesar 76,99%. Keberhasilan model mempertahankan akurasi tertinggi pada porsi data uji terbesar 1.246 data membuktikan bahwa kombinasi skenario 50:50 dan SMOTE menghasilkan model klasifikasi yang paling stabil dan objektif.overfitting.

4.6.4 Visualisasi Hasil

Pada tahap visualisasi hasil, sistem menampilkan beberapa bentuk visual dan tabel untuk mempermudah analisis hasil klasifikasi sentimen menggunakan algoritma *Naive bayes.* Visualisasi ini bertujuan untuk memberikan gambaran distribusi data, performa model, serta hasil klasifikasi sentimen secara lebih informatif dan mudah dipahami diantaranya:

a. Hasil Label Ulasan (Positif, Netral, Negatif)

Hasil label ulasan menyajikan ringkasan total data serta persebaran distribusi label sentimen yang berhasil diidentifikasi oleh sistem setelah melalui tahap pelabelan. Berdasarkan data keseluruhan yang diolah, sistem mencatat total ulasan sebanyak 2.492 data. Distribusi sentimen pada dataset ini didominasi oleh ulasan berkategori Positif, yaitu sebanyak 1.405 data. Selanjutnya, ulasan dengan kategori Netral menempati posisi kedua dengan jumlah 1.029 data. Sementara itu, ulasan berkategori Negatif merupakan kelompok minoritas dengan jumlah yang sangat terbatas, yaitu hanya sebesar 58 data. Ringkasan distribusi visual dari hasil pelabelan sentimen ini secara lengkap dapat dilihat pada Gambar 4.10.

Gambar 4. 10 Hasil Label Ulasan

b. Evaluasi Model Naive Bayes

Pada skenario pembagian data 80:20, pengujian tanpa SMOTE menghasilkan nilai akurasi sebesar 67,27% dan presisi macro sebesar 70,01%. Namun, nilai recall macro cenderung rendah yaitu 48,07%. Setelah diterapkan metode SMOTE, meskipun nilai akurasi sedikit terkoreksi menjadi 66,47% dan presisi macro menjadi 55,34%, model mengalami peningkatan signifikan pada nilai recall macro menjadi 57,14% dan *F1-score* macro menjadi 54,10%. Hal ini menunjukkan bahwa SMOTE berhasil membantu model dalam mengenali kelas minoritas negatif dengan lebih baik. Ringkasan perbandingan parameter evaluasi untuk skenario 80:20 ini disajikan pada Gambar 4.11.

Evaluasi Model Multinomial Naive Bayes

Perbandingan hasil tanpa SMOTE dan dengan SMOTE.

Parameter	Hasil Tanpa SMOTE	Hasil Dengan SMOTE
Akurasi	67.27%	66.47%
Presisi (Macro)	70.01%	55.34%
Recall (Macro)	48.07%	57.14%
F1-Score (Macro)	49.58%	54.10%

Gambar 4. 11 Hasil Model Skenario 80:20

Selanjutnya, pada skenario pembagian data 70:30, model tanpa SMOTE memperoleh nilai akurasi sebesar 68,18% dengan *F1-score* macro sebesar 46,05%. Ketika metode SMOTE diimplementasikan pada data latih, nilai akurasi tercatat sebesar 67,38%. Penerapan SMOTE pada skenario ini secara nyata mendongkrak *F1-score* macro naik sebesar 6,84% menjadi 52,89% serta recall macro naik menjadi 54,46%. Peningkatan parameter makro ini membuktikan berkurangnya sifat bias model terhadap kelas mayoritas setelah dataset diseimbangkan. Visualisasi metrik evaluasi untuk skenario 70:30 dapat dilihat secara jelas pada Gambar 4.12.

Evaluasi Model Multinomial Naive Bayes

Perbandingan hasil tanpa SMOTE dan dengan SMOTE.

Parameter	Hasil Tanpa SMOTE	Hasil Dengan SMOTE
Akurasi	68.18%	67.38%
Presisi (Macro)	64.81%	54.47%
Recall (Macro)	45.82%	54.46%
F1-Score (Macro)	46.05%	52.89%

Gambar 4. 12 Hasil Model Skenario 70:30

Pada skenario pembagian data 60:40, performa model menunjukkan tren yang semakin positif. Pengujian tanpa SMOTE menghasilkan nilai akurasi 70,01%, presisi 66,15%, recall 46,72%, dan *F1-score* 47,07%. Sementara itu, pengujian dengan menerapkan metode SMOTE menghasilkan nilai akurasi yang sangat bersaing yaitu sebesar 69,81%. Keunggulan utama pada kondisi dengan SMOTE di

skenario ini terlihat dari melonjaknya nilai recall macro menjadi 56,99% dan *F1-score* macro menjadi 55,02%, menandakan model semakin stabil seiring bertambahnya porsi data pengujian. Detail visualisasi rekapitulasi performa skenario 60:40 ditunjukkan pada Gambar 4.13.

Evaluasi Model Multinomial Naive Bayes

Perbandingan hasil tanpa SMOTE dan dengan SMOTE.

Parameter	Hasil Tanpa SMOTE	Hasil Dengan SMOTE
Akurasi	**70.01%**	**69.81%**
Presisi (Macro)	**66.15%**	**56.20%**
Recall (Macro)	**46.72%**	**56.99%**
F1-Score (Macro)	**47.07%**	**55.02%**

Gambar 4. 13 Hasil Model Skenario 60:40

Skenario terakhir yaitu pembagian data dengan rasio 50:50. Pada pengujian tanpa SMOTE, diperoleh nilai akurasi sebesar 69,10%. Menariknya, ketika metode SMOTE diterapkan, skenario 50:50 ini sukses mencatatkan performa terbaik di antara seluruh eksperimen dengan nilai akurasi tertinggi mencapai 70,47%. Tidak hanya akurasi, metrik evaluasi lainnya dengan SMOTE juga menyentuh angka optimal, meliputi presisi sebesar 56,99%, recall sebesar 57,05%, dan *F1-score* sebesar 55,32%. Hal ini membuktikan bahwa pembagian data berimbang (50:50) yang dikombinasikan dengan teknik penyeimbangan data SMOTE menghasilkan performa klasifikasi sentimen yang paling kuat, objektif, dan stabil. Ringkasan hasil evaluasi untuk skenario puncak ini dapat dilihat pada Gambar 4.14.

Classification Report Per kelas

Metrik per kelas beserta macro dan weighted average.

Kelas	Precision	Recall	F1	Support
Positif	70.56%	87.75%	**78.22%**	702
Netral	76.99%	49.02%	**59.90%**	512
Negatif	23.40%	34.38%	**27.85%**	32
Macro avg	56.99%	57.05%	**55.32%**	1246
Weighted avg	71.99%	70.47%	**69.40%**	1246

Gambar 4. 14 Hasil Model Skenario 50:50

c. Confusion Matrix

Confusion Matrix digunakan untuk memvisualisasikan ketepatan prediksi model secara mendetail pada masing-masing kelas aktual positif, netral, dan negatif. Melalui visualisasi ini, dapat diketahui secara spesifik pada kelas mana model memberikan prediksi yang akurat terbaca pada diagonal utama matriks serta ke kelas mana model paling sering mengalami salah klasifikasi.

Detail visualisasi pencapaian confusion matrix untuk masing-masing eksperimen telah disajikan secara berurutan pada penjelasan sebelumnya, yaitu Gambar 4.6 untuk skenario pembagian data 80:20, Gambar 4.7 untuk skenario 70:30, Gambar 4.8 untuk skenario 60:40, dan Gambar 4.9 untuk skenario pengujian terakhir 50:50. Berdasarkan rujukan matriks-matriks tersebut, terlihat jelas bahwa model di semua skenario memiliki sensitivitas yang sangat tinggi dalam mengenali sentimen positif, namun masih memiliki tantangan dalam membedakan sentimen netral dan negatif akibat keterbatasan variasi data asli pada kelas minoritas.

d. Tabel Split Data

Visualisasi tabel split data digunakan untuk memberikan ringkasan informasi mengenai pembagian jumlah baris data dari total dataset awal sebanyak 2.492 data. Tabel ini memetakan secara transparan proporsi rasio data training awal, jumlah pembengkakan data training setelah dilakukan penyeimbangan menggunakan metode SMOTE, serta porsi data testing yang digunakan sebagai bahan pengujian model. Ringkasan visualisasi data ini dijabarkan ke dalam empat skenario pembagian rasio sebagai berikut:

Pada skenario pertama dengan rasio pembagian 80:20, sistem mengalokasikan 80% data untuk proses pelatihan dan 20% sisanya untuk pengujian. Dari pembagian tersebut, diperoleh jumlah data training awal sebanyak 1.994 data dan data testing sebanyak 498 data. Setelah data training awal diproses menggunakan metode SMOTE untuk menyeimbangkan kelas minoritas, jumlah data latih meningkat signifikan menjadi 3.411 data. Tampilan visual rekapitulasi data untuk skenario ini dapat dilihat pada Gambar 4.15.

Tabel Split Data

Ringkasan rasio split data, training, testing, dan total dataset.

Tahap	Jumlah
Rasio	80:20
Data Training Awal	1994
Data Training dengan SMOTE	3411
Testing	498
Total dataset	2492

Gambar 4. 15 Hasil Tabel Split Data Skenario 80:20

Pada skenario pembagian data dengan rasio 70:30, porsi data dialokasikan sebesar 70% sebagai data latih dan 30% sebagai data uji. Berdasarkan rasio ini, diperoleh jumlah data training awal sebanyak 1.744 data dan data testing sebanyak 748 data. Proses augmentasi data menggunakan metode SMOTE pada data latih menghasilkan peningkatan jumlah data menjadi 2.976 data training baru yang seimbang sebelum dimasukkan ke dalam tahap pemodelan. Detail tabel pembagian data ini ditunjukkan pada Gambar 4.16.

Tabel Split Data

Ringkasan rasio split data, training, testing, dan total dataset.

Tahap	Jumlah
Rasio	70:30
Data Training Awal	1744
Data Training dengan SMOTE	2976
Testing	748
Total dataset	2492

Gambar 4. 16 Hasil Tabel Split Data Skenario 70:30

Selanjutnya, untuk skenario pembagian data dengan rasio 60:40, sistem membagi dataset dengan proporsi 60% data pelatihan dan 40% data pengujian. Skenario ini menghasilkan total data training awal sebanyak 1.495 data dan data testing yang jauh lebih besar dari skenario sebelumnya, yaitu sebanyak 997 data. Setelah diterapkan teknik SMOTE pada data training awal, jumlah data latih bertambah secara sintetis menjadi 2.550 data agar ketiga kelas sentimen memiliki

bobot yang seimbang. Ringkasan informasi pembagian data ini disajikan pada Gambar 4.17.

Tabel Split Data

Ringkasan rasio split data, training, testing, dan total dataset.

Tahap	Jumlah
Rasio	60:40
Data Training Awal	1495
Data Training dengan SMOTE	2550
Testing	997
Total dataset	2492

Gambar 4. 17 Hasil Tabel Split Data Skenario 60:40

Skenario terakhir menggunakan rasio pembagian seimbang 50:50, di mana dataset dibagi sama rata menjadi 50% data pelatihan dan 50% data pengujian. Melalui pembagian ini, diperoleh jumlah data training awal sebanyak 1.246 data dan data testing sebanyak 1.246 data. Penerapan metode SMOTE pada porsi data latih di skenario ini menghasilkan total 2.109 data training akhir yang seimbang. Kombinasi pembagian porsi data uji yang luas ini terbukti menghasilkan metrik evaluasi terbaik pada sistem. Visualisasi ringkasan split data skenario 50:50 ini dapat dilihat secara lengkap pada Gambar 4.18.

Tabel Split Data

Ringkasan rasio split data, training, testing, dan total dataset.

Tahap	Jumlah
Rasio	50:50
Data Training Awal	1246
Data Training dengan SMOTE	2109
Testing	1246
Total dataset	2492

Gambar 4. 18 Hasil Tabel Split Data Skenario 50:50

e. Distribusi Data Training

Visualisasi ini menampilkan perbandingan grafik batang antara data training awal imbalance dataset dan Data Training dengan SMOTE balanced dataset. Perbandingan ini menunjukkan bagaimana metode SMOTE bekerja secara sintetis untuk menyamakan jumlah sampel kelas minoritas netral dan negatif agar setara dengan kelas mayoritas positif sebelum tahap pelatihan model. Penjelasan grafik distribusi untuk masing-masing skenario dipaparkan sebagai berikut::

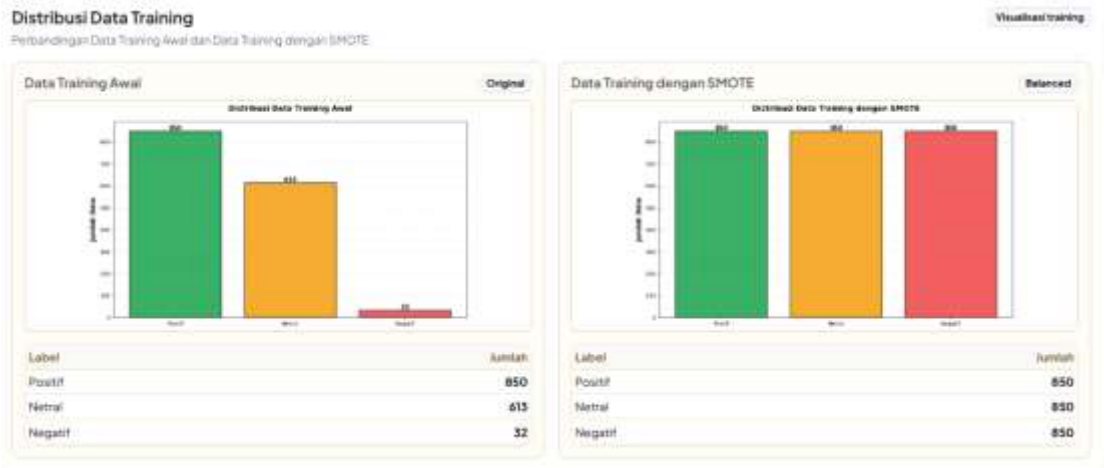
Pada skenario 80:20, grafik data awal menunjukkan dominasi kelas Positif (1.137 data), diikuti Netral (816 data), dan Negatif (41 data). Setelah diterapkan SMOTE, jumlah sampel kelas Netral dan Negatif diduplikasi secara sintetis hingga menyamai kelas mayoritas di angka 1.137 data per kelas. Perbandingan grafik skenario ini disajikan pada Gambar 4.19.

Gambar 4. 19 Hasil Distribusi Data Training 80:20

Untuk skenario 70:30, grafik original mencatat kelas Positif sebanyak 992 data, Netral 716 data, dan Negatif 36 data. Melalui algoritma SMOTE, kelas Netral dan Negatif ditingkatkan secara otomatis hingga mencapai jumlah seimbang, yaitu masing-masing 992 data sampel. Hasil penyeimbangan ini ditunjukkan pada Gambar 4.20.

Gambar 4. 20 Hasil Distribusi Data Training 70:30

Pada skenario 60:40, distribusi data awal terdiri atas 850 sampel Positif, 613 Netral, dan 32 Negatif. Penerapan teknik SMOTE berhasil menyeimbangkan porsi data latih dengan mendongkrak jumlah sampel kelas Netral dan Negatif hingga menyamai batas kelas mayoritas di angka 850 data. Grafik komparasi skenario ini dapat dilihat pada Gambar 4.21.



Gambar 4. 21 Hasil Distribusi Data Training 60:40

Pada skenario 50:50, grafik awal mencatat kelas Positif sebesar 703 data, Netral 517 data, dan Negatif 26 data. Setelah diolah menggunakan SMOTE, kuantitas data kelas Netral dan Negatif meningkat secara proporsional menjadi 703 data per kelas. Komposisi seimbang ini terbukti menghasilkan performa model yang paling optimal. Detail visualisasi grafik skenario ini tertera pada Gambar 4.22.

Gambar 4. 22 Hasil Distribusi Data Training 50:50

f. *Classification Report*

Visualisasi *Classification Report* menyajikan rincian metrik evaluasi model secara mendetail untuk setiap kelas sentimen. Pada skenario 80:20 dengan metode SMOTE, kelas Positif mencatatkan recall tertinggi sebesar 84,70% dan kelas Netral unggul pada precision sebesar 74,05%, sementara kelas Negatif memperoleh performa terendah dengan precision 25,00% dan *F1-score* 31,11%. Rincian nilai metrik evaluasi otomatis dari pengujian multi skenario ini dipaparkan pada Gambar 4.23.

Classification Report Per kelas

Metrik per kelas beserta macro dan weighted average.

Kelas	Precision	Recall	F1	Support
Positif	66.96%	84.70%	**74.79%**	268
Netral	74.05%	45.54%	**56.40%**	213
Negatif	25.00%	41.18%	**31.11%**	17
Macro avg	55.34%	57.14%	**54.10%**	498
Weighted avg	68.56%	66.47%	**65.43%**	498

Gambar 4. 23 Hasil *Classification Report* Skenario 80:20

Skenario 70:30, laporan hasil menunjukkan nilai *F1-score* untuk kelas Positif sebesar 75,62% dan kelas Netral sebesar 56,64%. Sementara itu, kelas Negatif

menghasilkan nilai recall sebesar 31,82% dan precision sebesar 22,58%. Visualisasi metrik evaluasi menyeluruh pada skenario ini ditunjukkan pada Gambar 4.24.

Classification Report Per kelas

Metrik per kelas beserta macro dan weighted average.

Kelas	Precision	Recall	F1	Support
Positif	67.95%	85.23%	**75.62%**	413
Netral	72.86%	46.33%	**56.64%**	313
Negatif	22.58%	31.82%	**26.42%**	22
Macro avg	54.47%	54.46%	**52.89%**	748
Weighted avg	68.67%	67.38%	**66.23%**	748

Gambar 4. 24 Hasil *Classification Report* Skenario 70:30

Pada skenario 60:40, terjadi peningkatan performa di mana kelas Positif mencatatkan nilai precision 70,09% dan recall 86,13%. Kelas Netral juga menunjukkan konsistensi dengan precision mencapai 76,00%. Laporan performa lengkap beserta nilai rata-rata makro untuk skenario ini disajikan pada Gambar 4.25.

Classification Report Per kelas

Metrik per kelas beserta macro dan weighted average.

Kelas	Precision	Recall	F1	Support
Positif	70.09%	86.13%	**77.28%**	555
Netral	76.00%	50.24%	**60.49%**	416
Negatif	22.50%	34.62%	**27.27%**	26
Macro avg	56.20%	56.99%	**55.02%**	997
Weighted avg	71.31%	69.81%	**68.97%**	997

Gambar 4. 25 Hasil *Classification Report* Skenario 60:40

Pada skenario 50:50, laporan pengujian mencatatkan hasil paling optimal di antara semua skenario. Kelas Positif menyentuh angka recall tertinggi sebesar 87,75% dengan *F1-score* mencapai 78,22%, sedangkan kelas Netral meraih precision sebesar 76,99%. Hasil laporan performa terbaik ini tertera secara lengkap pada Gambar 4.26.

Classification Report Per kelas

Metrik per kelas beserta macro dan weighted average.

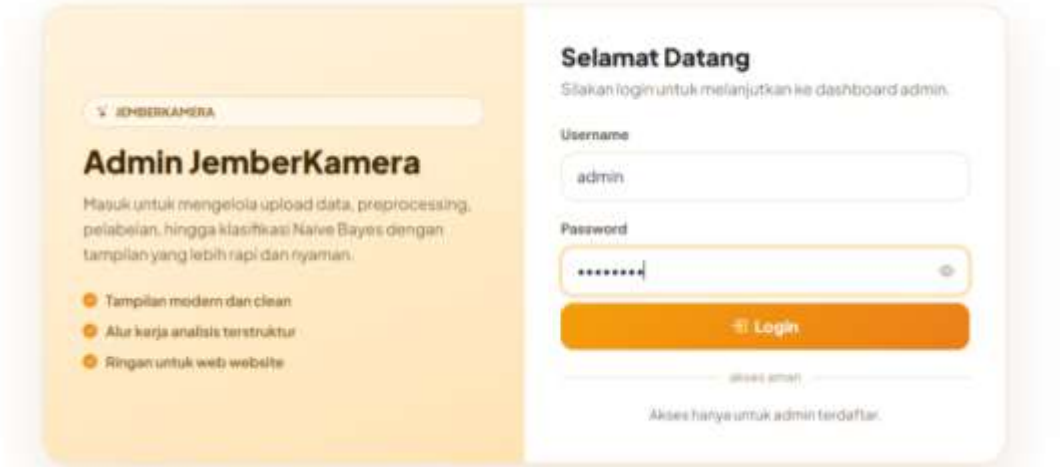
Kelas	Precision	Recall	F1	Support
Positif	70.56%	87.75%	**78.22%**	702
Netral	76.99%	49.02%	**59.90%**	512
Negatif	23.40%	34.38%	**27.85%**	32
Macro avg	56.99%	57.05%	**55.32%**	1246
Weighted avg	71.99%	70.47%	**69.40%**	1246

Gambar 4. 26 Hasil *Classification Report* Skenario 50:50

4.6.5 Antarmuka Sistem

a. Halaman Login

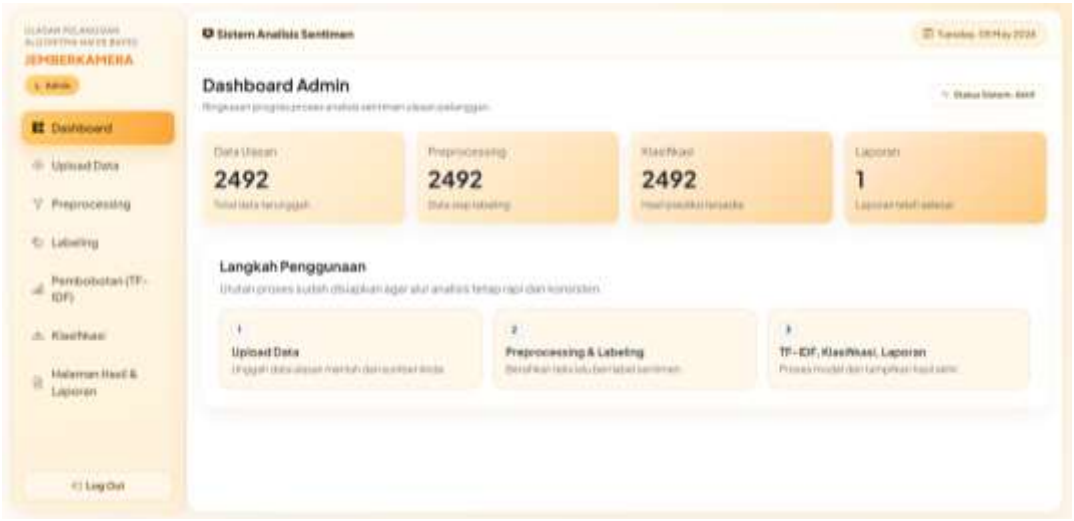
Halaman ini merupakan pintu masuk utama bagi administrator untuk mengakses sistem. Desain menggunakan tata letak dua sisi yang modern, di mana sisi kiri memberikan informasi singkat mengenai fitur utama sistem seperti tampilan yang bersih, alur kerja terstruktur, dan performa yang ringan. Di sisi kanan, terdapat formulir login yang terdiri dari input username dan password memudahkan akses menuju dashboard admin. terkhususkan oleh mitra jemberkamera.



Gambar 4. 27 Halaman Login

b. Halaman Dashboard Admin

Halaman *Dashboard* berfungsi sebagai pusat kendali dan ringkasan progres analisis. Pada bagian atas, terdapat empat kartu informasi utama yang menampilkan statistik data, meliputi total data ulasan, jumlah data hasil *Preprocessing*, jumlah data yang telah diklasifikasi, dan laporan yang telah selesai. Selain itu, terdapat bagian "Langkah Penggunaan" yang memberikan panduan alur kerja sistem, mulai dari unggah data, pembersihan teks, hingga proses klasifikasi akhir, guna memastikan pengguna mengikuti prosedur yang konsisten.

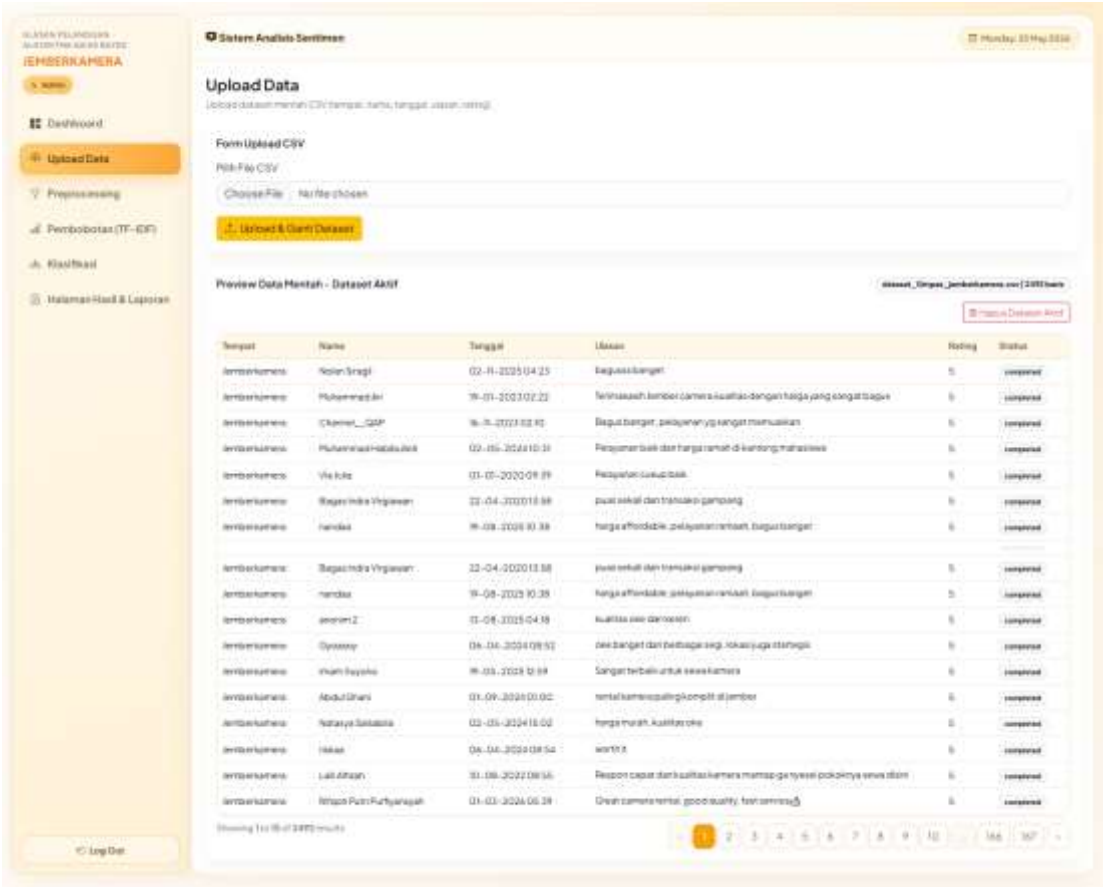


Gambar 4. 28 Halaman Dashboard

c. Halaman *Upload* Data

Halaman ini menampilkan menu interaktif pengelolaan dokumen yang berfungsi sebagai media memasukkan data ulasan konsumen Jemberkamera ke dalam basis data sistem. Komponen Form Upload CSV yang menyediakan tombol pemilihan berkas (*Choose File*) dan tombol aksi berwarna kuning mencolok bertuliskan "Upload & Ganti Dataset". Tepat di bawah formulir tersebut, sistem menyajikan panel visualisasi data bernama *Preview* Data Mentah - Dataset Aktif yang terintegrasi dengan fungsi penanda nama file berkas yang sedang aktif, jumlah total ulasan sebanyak 2492 baris, serta tombol merah "Hapus Dataset Aktif" untuk mengosongkan database secara instan. Hasil ekstraksi berkas pada Gambar 4.29 ditampilkan secara mendetail melalui tabel terstruktur yang memetakan enam

atribut utama ulasan, meliputi kolom Tempat, Nama pelanggan, Tanggal publikasi, isi komentar teks Ulasan, bobot nilai Rating, hingga label Status pemrosesan berkas. Selain itu, dilengkapi dengan komponen navigasi halaman (*pagination*) interaktif di sudut kanan bawah yang berfungsi untuk membatasi tampilan visual data agar rapi dan memudahkan administrator dalam meninjau seluruh baris ulasan murni secara sistematis.

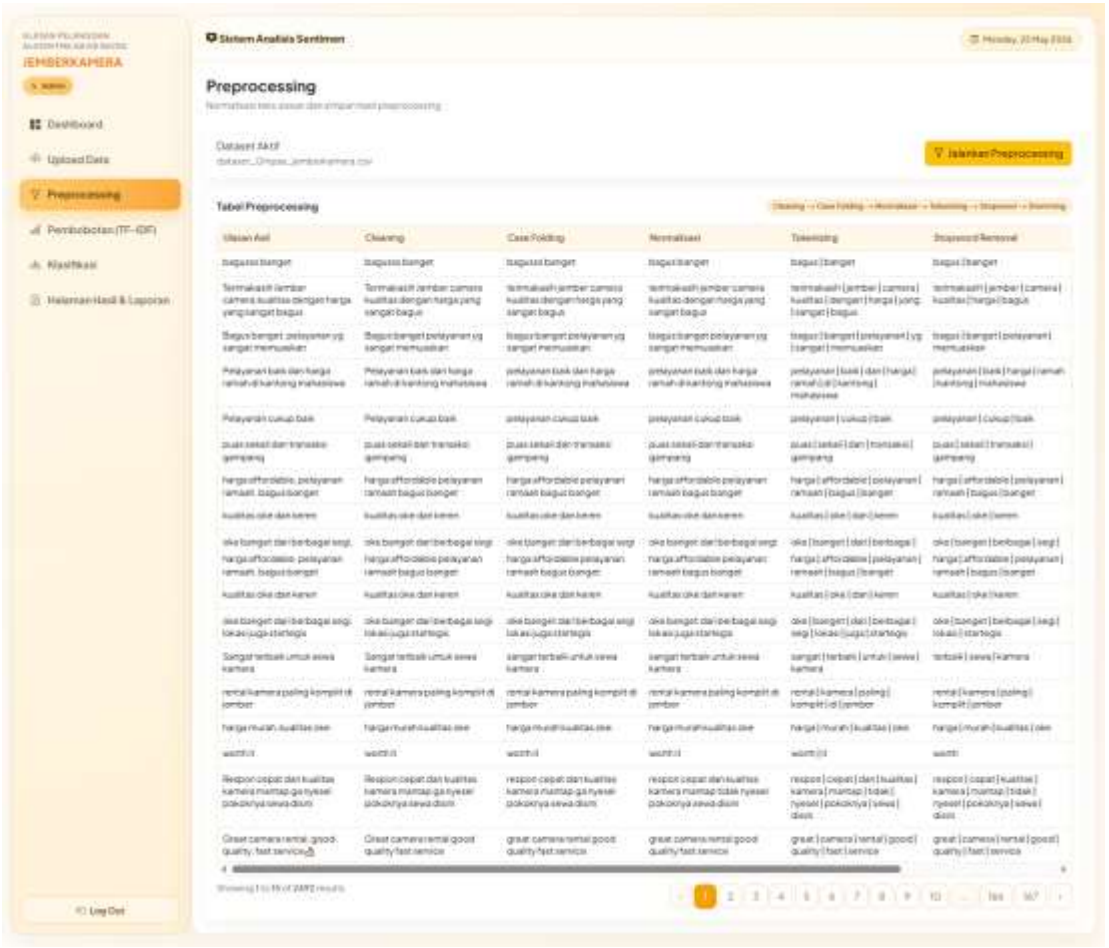


Gambar 4. 29 Hasil Halaman *Upload* Dataset Diperoleh dari *Scraping*

d. Halaman *Preprocessing*

Halaman *Preprocessing* menampilkan proses normalisasi teks ulasan secara transparan. Setelah pengguna menekan tombol "Jalankan *Preprocessing*", sistem akan melakukan serangkaian tahapan yang meliputi *Cleaning*, *Case Folding*, Normalisasi, *Tokenizing*, *Stopword Removal*, dan *Stemming*. Hasil dari setiap tahapan ditampilkan dalam bentuk tabel kolom-per-kolom, sehingga pengguna dapat melihat perubahan teks dari bentuk aslinya hingga menjadi data bersih yang

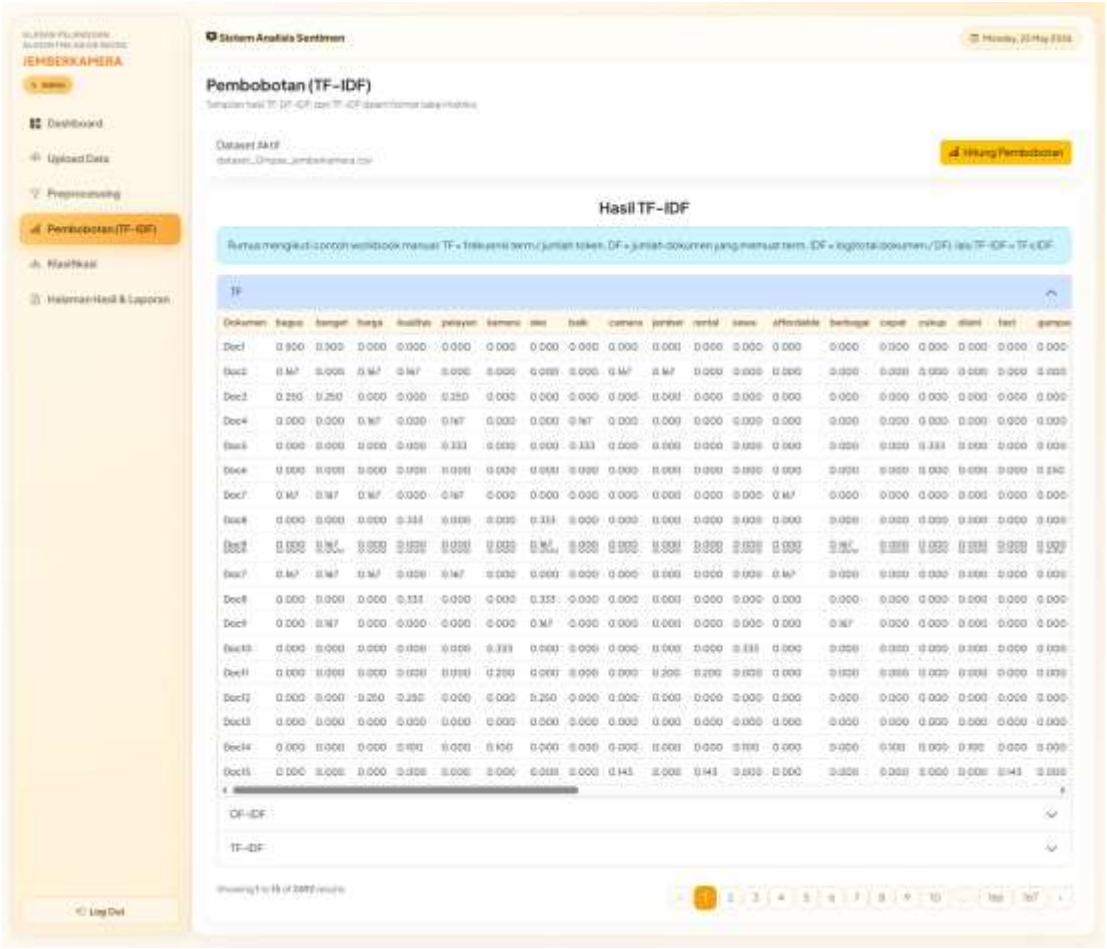
siap untuk diolah oleh algoritma. Tampilan secara lebih detail dapat dilihat pada Lampiran 3.



Gambar 4. 30 Hasil Halaman Prepocessing

e. Halaman Pembobotan (TF-IDF)

Halaman ini menyajikan hasil perhitungan bobot kata menggunakan metode TF-IDF. Terdapat penjelasan rumus yang digunakan sebagai referensi perhitungan manual. Hasil perhitungan ditampilkan dalam format tabel matriks yang menunjukkan nilai bobot setiap kata kunci (seperti "pelayanan", "rekomendasi", "harga") terhadap setiap dokumen ulasan. Proses ini merupakan langkah krusial untuk menentukan fitur kata mana yang paling berpengaruh dalam menentukan sentimen pada tahap klasifikasi Naive Bayes. Tampilan secara lebih detail untuk TF dilihat pada Lampiran 4, DF-IDF Lampiran 5 dan TF-IDF di Lampiran 6.

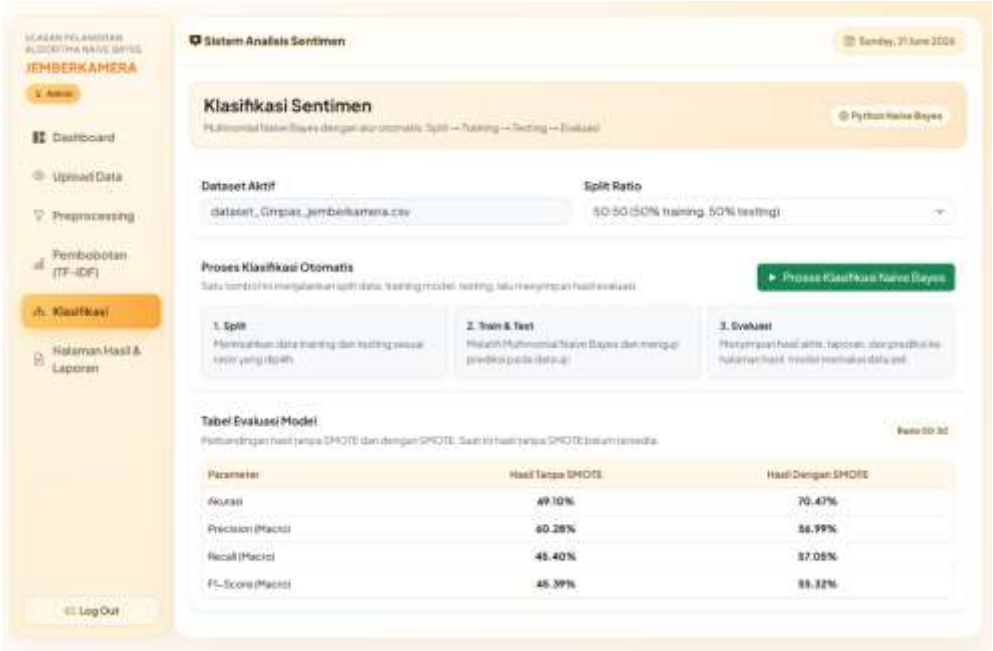


Gambar 4. 31 Hasil Pembobotan TF-IDF

f. Halaman Klasifikasi Naive Byes

Halaman klasifikasi Naive Bayes digunakan untuk menjalankan proses klasifikasi sentimen secara otomatis, mulai dari pembagian data split ratio, pelatihan model, pengujian, hingga evaluasi hasil. Pengguna dapat memilih dataset aktif dan menentukan rasio data, seperti 50:50 pada gambar. Seluruh proses dijalankan melalui tombol Proses Klasifikasi Naive Bayes, kemudian hasilnya ditampilkan pada Tabel Evaluasi Model berupa perbandingan nilai akurasi,

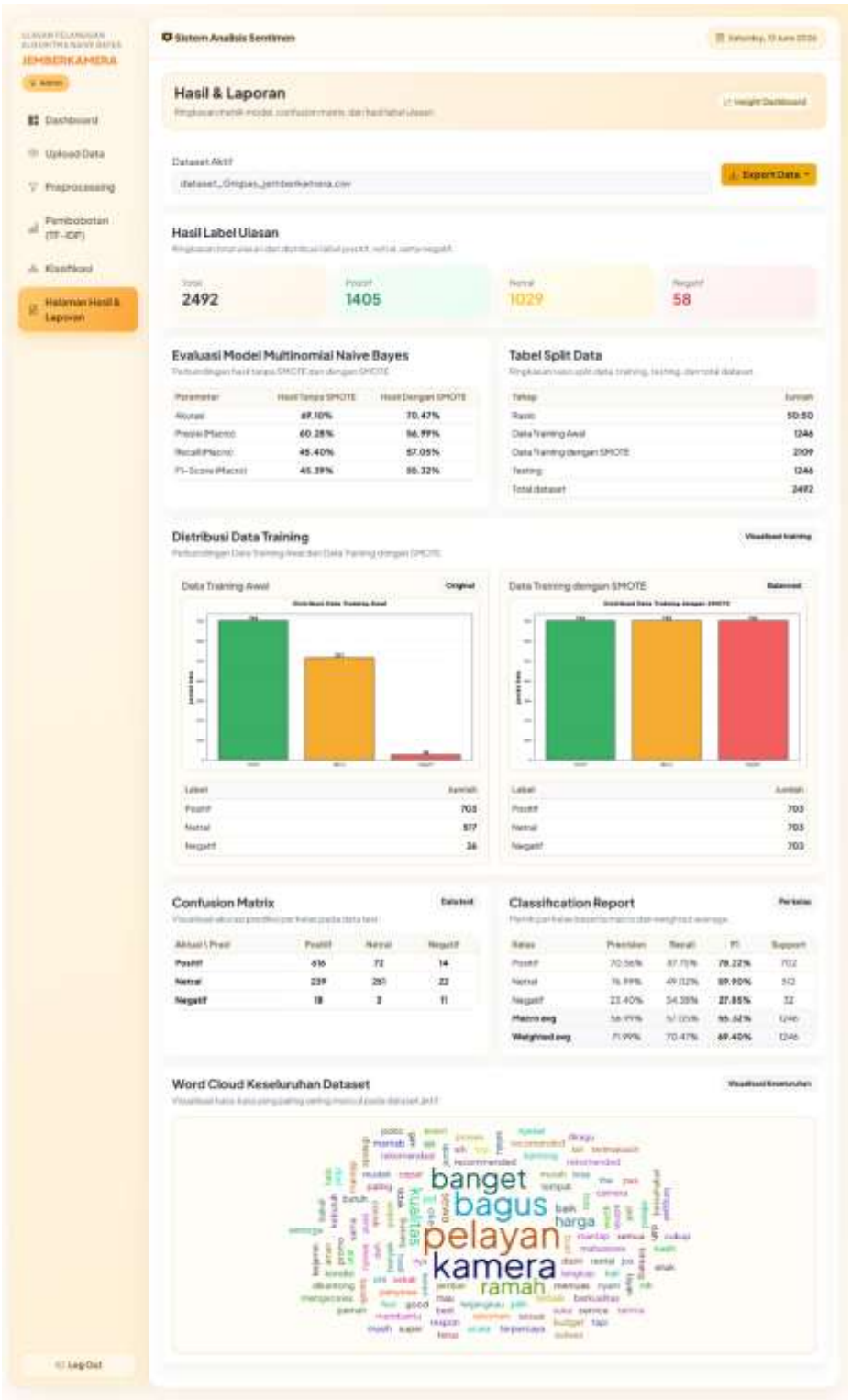
precision, recall, dan F1-score antara metode tanpa SMOTE dan dengan SMOTE untuk membantu menilai kinerja model secara objektif.



Gambar 4. 32 Hasil Proses Klasifikasi Naive Bayes

g. Halaman Hasil dan Laporan

Hasil menyajikan visualisasi data komprehensif yang diawali dengan fitur export data untuk mengunduh seluruh dataset ulasan hasil prediksi multinomial Naive bayes, diikuti kartu metrik hasil label ulasan yang memetakan total 2.492 data 1.405 positif, 1.029 netral, dan 58 negatif. Bagian tengah menampilkan panel evaluasi model dan tabel split data rasio 50:50, serta komponen distribusi data training yang memuat dua grafik batang komparatif; grafik kiri menunjukkan ketimpangan data asli, sedangkan grafik kanan mengonfirmasi keberhasilan penyeimbangan data latih lewat metode smote yang sama rata di angka 703 sampel untuk semua kelas. pada bagian bawah, antarmuka menyajikan tabel tabular berupa confusion matrix dan classification report untuk mengukur ketepatan prediksi data uji, serta ditutup dengan visualisasi *word cloud* yang menyediakan pilihan *dropdown* berupa kategori keseluruhan, positif dilihat pada lampiran 8 , netral pada lampiran 9, dan negatif pada lampiran 10, sehingga kata-kata yang paling dominan pada masing-masing kategori ulasan pelanggan Jemberkamera dapat ditampilkan secara ringkas.



Gambar 4. 33 Hasil dan Visualisasi Evaluasi Model

4.6.6 *Black Box Testing*

Pada penelitian ini, pengujian sistem dilakukan menggunakan metode *Black Box Testing* untuk menguji fungsi pada *website* analisis sentimen berbasis Naive Bayes. Pengujian dilakukan dengan melihat kesesuaian antara input dan output yang dihasilkan tanpa memperhatikan kode program yang digunakan. Fitur yang diuji meliputi halaman login, *Upload* data, *Preprocessing*, pembobotan TF-IDF, klasifikasi Naive Bayes, serta halaman hasil dan laporan.

Pengujian dilakukan untuk memastikan setiap proses pada sistem dapat berjalan sesuai alur yang telah dirancang. Sistem diuji mulai dari proses memasukkan data ulasan, melakukan *Preprocessing*, menghitung bobot TF-IDF, hingga proses klasifikasi sentimen ke dalam kelas positif, netral, dan negatif menggunakan metode *Naive Bayes*. Selain itu, sistem juga diuji dalam menampilkan hasil evaluasi model seperti *accuracy, precision, recall,* dan *F1-score*.

Berdasarkan hasil pengujian yang telah dilakukan, seluruh fitur pada *website* dapat berjalan dengan baik sesuai fungsinya. Sistem berhasil memproses data ulasan dan menampilkan hasil klasifikasi serta laporan evaluasi tanpa ditemukan kesalahan fungsional yang signifikan. Seluruh rangkaian pengujian fungsional ini juga telah didemonstrasikan dan diverifikasi secara langsung bersama pihak pengelola Jemberkamera dan teknisi jurusan guna memastikan kesesuaian sistem dengan kebutuhan operasional mitra, dengan bukti dokumentasi serta lembar validasi yang terlampir pada Lampiran 2 dan Lampiran 7. Hasil pengujian sistem selanjutnya ditampilkan pada tabel 4.17 pengujian *Black-box Testing*.

Tabel 4. 17 Hasil Pengujian *Blackbox Testing* Bersama pihak jemberkamera dan Teknisi Jurusan

NO	Pengujian Fitur	Hasil
1	Halaman Login	Berhasil
2	Halaman Dashboard	Berhasil
3	Halaman *Upload* Data	Berhasil
4	Halaman Prepocessing	Berhasil
5	Halaman TF-IDF	Berhasil

NO	Pengujian Fitur	Hasil
6	Halaman Klasifikasi Naive Bayes	Berhasil
7	Halaman Hasil dan Laporan	Berhasil

4.6.7 *User Testing*

Pengujian pengguna *user testing* atau pengujian lapangan dilakukan untuk mengevaluasi fungsionalitas, kegunaan, dan kesesuaian sistem analisis sentimen secara langsung dari perspektif pengguna di lingkungan kerja riil. Proses pengujian ini melibatkan pihak internal mitra secara langsung, yaitu *staf Front Officer* dari Jemberkamera yang bertindak sebagai operator utama sistem. Implementasi pengujian berfokus pada simulasi alur kerja analisis ulasan, dimulai dari pengunggahan dataset mentah hasil scraping, peninjauan transparansi data pada tabel preprocessing dan pembobotan TF-IDF, eksekusi pemodelan menggunakan opsi multi skenario rasio data, hingga interpretasi grafik visualisasi pada halaman hasil dan laporan.

Sebagai bukti bahwa aplikasi telah diuji, diverifikasi, dan diterima dengan baik untuk memenuhi kebutuhan operasional mitra, lembar pengesahan hasil uji coba yang ditandatangani secara resmi oleh perwakilan pihak validator Jemberkamera dapat dilihat secara mendetail pada Lampiran 2.

4.7 Analisis Hasil

Berdasarkan seluruh rangkaian eksperimen pada sistem analisis sentimen Jemberkamera, diperoleh kesimpulan bahwa skenario pembagian data 50:50 dengan penerapan metode SMOTE merupakan konfigurasi paling optimal. Skenario puncak ini berhasil mencatatkan nilai akurasi tertinggi sebesar 70,47% dan terbukti stabil tanpa mengalami *overfitting* meskipun diuji menggunakan porsi data pengujian terbesar 1.246 data. Penggunaan teknik *oversampling* melalui SMOTE secara konsisten sukses mendongkrak metrik rata-rata makro *macro avg* pada semua skenario pembagian data, yang membuktikan bahwa penyeimbangan data latih mampu memangkas sifat bias model terhadap kelas mayoritas sehingga sistem dapat mengklasifikasikan data secara lebih adil dan objektif.

Di sisi lain, hasil analisis juga menunjukkan adanya keterbatasan performa model pada kelas minoritas, di mana nilai presisi untuk sentimen negatif cenderung tertahan di kisaran rendah yaitu 22%–25% pada seluruh skenario pengujian. Hal ini menjadi catatan penting dalam penelitian bahwa meskipun kuantitas data latih telah direkayasa dan ditingkatkan secara sintetis oleh algoritma SMOTE, keterbatasan variasi kosakata asli pada data awal ulasan *Google Maps* Jemberkamera yang secara alami hanya berjumlah 58 ulasan negatif tetap membatasi kemampuan generalisasi model Naive Bayes saat menghadapi data uji riil.

Persamaan ketergantungan performa model terhadap karakteristik data awal, bahwa rendahnya performa model dan bias terjadi akibat ketimpangan distribusi data ulasan teks *data imbalance*, sehingga keterbatasan tersebut menjadi landasan ilmiah yang kuat mengapa akurasi algoritma berbasis probabilitas dapat tertahan (Rifaldi dkk., 2025). Namun, keputusan untuk mengintegrasikan teknik oversampling dalam riset ini terbukti menjadi langkah korektif yang tepat. Pada kondisi awal dataset yang tidak seimbang *imbalance data*, model Naive Bayes menghasilkan performa yang kurang akurat dalam mengenali kelas minoritas. Setelah diterapkan metode oversampling SMOTE, akurasi model mereka terbukti meningkat (Zidny, 2024). Dengan demikian, visualisasi peningkatan metrik makro pada website Jemberkamera menegaskan bahwa manipulasi data latih lewat SMOTE mampu memberikan hasil evaluasi yang jauh lebih objektif dibandingkan membiarkan dataset dalam kondisi timpang.

BAB 5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian, implementasi, dan pengujian yang telah dilakukan pada sistem analisis sentimen ulasan *Google Maps* Jemberkamera menggunakan algoritma Naive Bayes, maka dapat ditarik beberapa kesimpulan sebagai jawaban atas rumusan masalah yang diajukan:

1. Proses pengambilan data ulasan pelanggan pada *Google Maps* Jemberkamera dilakukan secara otomatis menggunakan teknik web *scraping* berbantuan alat Instant Data Scraper dan platform Apify. Data mentah yang berhasil diekstraksi berjumlah 2.492 data ulasan berbentuk file CSV, yang kemudian melewati serangkaian tahap *Preprocessing* teks meliputi cleansing, case folding, nomalisasi, *Tokenizing*, *Stopword Removal*, dan *Stemming*, hingga siap digunakan ke tahap pemodelan.
2. Sistem klasifikasi sentimen berbasis web ini menggunakan data ulasan pelanggan Jemberkamera yang diperoleh melalui web scraping. Data diproses melalui tahapan preprocessing, yaitu cleaning, normalisasi, case folding, tokenisasi, stopword removal, dan stemming. Selanjutnya, teks diberi bobot menggunakan TF-IDF. Data hasil pengolahan diberi label sentimen terdahulu pendekatan *lexicon-based*, kemudian digunakan untuk pelatihan dan pengujian model Multinomial Naive Bayes.
3. Tingkat akurasi terbaik metode Naive Bayes dalam mengelompokkan ulasan pelanggan diperoleh pada skenario pembagian data 50:50 dengan penerapan SMOTE. Dengan cara tersebut, sistem menghasilkan nilai Akurasi tertinggi sebesar 70,47%, Presisi 56,99%, Recall 57,05%, dan *F1-score* 55,32%. Pengujian membuktikan bahwa integrasi SMOTE efektif memangkas sifat bias model pada kelas mayoritas, sehingga menghasilkan performa klasifikasi sistem yang paling stabil dan objektif.
4. Berdasarkan Hasil analisis dan pelabelan akhir terhadap keseluruhan dataset ulasan *Google Maps* Jemberkamera menunjukkan dominasi pada kategori

sentimen positif. Dari total 2.492 ulasan bersih yang diolah oleh sistem, distribusi sentimen mencatat sebanyak 1.405 ulasan berkategori positif, 1.029 ulasan berkategori netral, dan 58 ulasan berkategori negatif. Dominasi ulasan positif ini membuktikan secara objektif bahwa mayoritas pelanggan merasa sangat puas terhadap kualitas pelayanan, performa, serta kelengkapan unit peralatan fotografi yang disediakan oleh pihak Jemberkamera.

5.2 Saran

Berdasarkan pengalaman selama pelaksanaan penelitian dan kendala riil yang dihadapi di lapangan, terdapat beberapa saran yang dapat diajukan untuk pengembangan penelitian selanjutnya, antara lain:

1. Imbalance Dataset: Penelitian ini adalah adanya ketimpangan data asli pada sentimen Negatif (hanya 58 data dari total 2.492 ulasan). Meskipun data latih sudah menerapkan SMOTE, variasi kosakata asli pada data uji tetap terbatas. Peneliti selanjutnya disarankan untuk memperluas jangkauan sumber data ulasan misalnya menggabungkan ulasan dari media sosial atau platform digital lain untuk memperkaya variasi kosakata asli.
2. Penelitian selanjutnya disarankan untuk menerapkan metode preprocessing yang lebih optimal karena pada penelitian ini terdapat beberapa kata yang terhapus sehingga berpotensi mengubah makna ulasan. Dengan demikian, informasi penting dalam ulasan dapat lebih terjaga tanpa mengurangi kualitas proses analisis sentimen.
3. Pengembangan Algoritma Klasifikasi: Mengingat karakteristik ulasan teks pendek memiliki kompleksitas makna yang tinggi, peneliti selanjutnya disarankan untuk melakukan komparasi atau mengombinasikan algoritma Naive Bayes dengan metode klasifikasi lain seperti *Support Vector Machine* (SVM) atau pendekatan berbasis *Deep Learning* seperti BERT untuk meningkatkan akurasi spesifik pada kelas-kelas sentimen yang sulit dibedakan.

DAFTAR PUSTAKA

Afriansyah, M., Saputra, J., Yoga Pudya Ardhana, V., Sa, Y., & Qamarul Huda Badaruddin, U. (2024). Algoritma Naive Bayes Yang Efisien Untuk Klasifikasi Buah Pisang Raja Berdasarkan Fitur Warna. *Hal. 236 Journal of Information Systems Management and Digital Business (JISMDB)*, *1*(2).

Al Lutfani, T. K., Astuti, R., Prihartono, W., & Hamonangan, R. (2025). Penerapan Naive Bayes Untuk Analisis Sentimen Pada Ulasan Pelanggan Di Lazada: Studi Kasus Toko Mawar Collection. *Jurnal Informatika dan Teknik Elektro Terapan*, *13*(2). https://doi.org/10.23960/jitet.v13i2.6391

Apriliansyah, R. D. R., Astuti, R., Prihartono, W., & Hamonangan, R. (2025). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pengunjung Di Pantai Kejawanan. *Jurnal Informatika dan Teknik Elektro Terapan*, *13*(1). https://doi.org/10.23960/jitet.v13i1.5774

Ardiansyah, & Kurniawan. (2024). Optimasi Metode Naïve Bayes Classifier Menggunakan Pendekatan Term Frequency-Inverse Document Frequency (TF-IDF) Pada Analisis Sentimen. *JSAI: Journal Scientific and Applied Informatics*, *7*(3). https://doi.org/10.36085

Aziza, A. F., & Noor, H. (2026). Klasifikasi Sentimen Ulasan Produk Tokopedia Menggunakan SVM dan Random Forest dengan Strategi Eliminasi Label Ambigu Berbasis Binary Classification. *Jurnal Sains Sistem Informasi*, *4*(2), 73. https://doi.org/10.31602/jssi.v4i2.23714

Badan Pusat Statistik. (2025). *Statistik Telekomunikasi Indonesia 2024* (Vol. 13). https://www.bps.go.id

Budianto, A. G., Rusilawati, R., Suryo, A. T. E., Cahyono, G. R., Zulkarnain, A. F., & Martunus, M. (2024). Perbandingan Performa Algoritma Support Vector Machine (SVM) dan Logistic Regression untuk Analisis Sentimen Pengguna Aplikasi Retail di Android. *Jurnal Sains dan Informatika*, *10*(2). https://doi.org/10.34128/jsi.v10i2.911

Damanik, K., Sinaga, M., Sihombing, S., Hidajat, M., Prakoso, O. S., & Penulis, K. (2024). *Pengaruh Kualitas Layanan, Kebijakan Publik dan Kepuasan Pelanggan terhadap Loyalitas Pelanggan*. https://doi.org/10.38035/jmpis.v5i2

Fikri, M. R., Handayanto, R. T., & Irwan, D. (2022). Web Scraping Situs Berita Menggunakan Bahasa Pemograman Python. *Journal of Students' Research in Computer Science*, *3*(1), 123–136. https://doi.org/10.31599/jsrcs.v3i1.1514

Gondowastradmojo, A., & Kusumastuti, R. (2024). *Seminar Nasional Amikom Surakarta (Semnasa) 2024 Analisa Sentimen Ulasan Di Tokopedia Dengan Metode Naive Bayes Prodi S1 Informatika, STMIK Amikom Surakarta Sukoharjo Indonesia*.

Hakim, B. (2021). Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning. *JBASE - Journal of Business and Audit Information Systems*, *4*(2). https://doi.org/10.30813/jbase.v4i2.3000

Heristian, S., Napiah, M., Erawati, W., & Author, C. (2025). Analisis Sentimen Ulasan Pelanggan Menggunakan Algoritma Naive Bayes pada Aplikasi Gojek. Dalam *Computer Science (CO-SCIENCE* (Vol. 5, Nomor 1). http://jurnal.bsi.ac.id/index.php/co-science

Irwanto, I., & Giantika, G. G. (2022). Penyuluhan Penggunaan Kamera HP untuk Kebutuhan Kalangan Wanita RW 13 Babelan Bekasi. *Jurnal Abdi Masyarakat Indonesia*, *2*(3), 833–842. https://doi.org/10.54082/jamsi.334

Kamal, W. W., & Ratnasari, C. I. (2024). *Analisis Sentimen Ulasan Produk: Kajian Pustaka*.

Mayasari, S., & Safina, W. D. (2021). *S., & Safina, W. D. (2021). Pengaruh Kualitas Produk Dan Pelayanan Terhadap Kepuasan Konsumen Pada Restoran Ayam Goreng Kalasan Cabang Iskandar Muda Medan (Vol. 1, Nomor 2).*

Nafi', A., Tri, A., Harjanta, J., Herlambang, B. A., Fahmi, S., Universitas, I., Semarang, P., Timur, J. S., Semarang, K., Tengah, J., & Korespondensi, E. (2022). *J-INTECH (Journal of Information and Technology) Analisis Sentimen Review Pelanggan Lazada dengan Sastrawi Stemmer dan SVM-PSO untuk Memahami Respon Pengguna*.

Naquitasia, R., Hatta Fudholi, D., & Iswari, L. (2022). *Analisis Sentimen Berbasis Aspek Pada Wisata Halal Dengan Metode Deep Learning* (Vol. 16, Nomor 2). https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index

Pardede, J., & Darmawan, D. (2025). Perbandingan Algoritma Stemming Porter, Sastrawi, Idris, Dan Arifin & Setiono Pada Dokumen Teks Bahasa Indonesia. *12*(1), 69–76. https://doi.org/10.25126/jtiik.2025128860

Perancangan Implementasi Dry Cabinet Untuk Menyimpan Kamera DSLR Atau Mirrorless Dengan Sistem Pendeteksi Jumlah Kamera Berbasis Microntroller. (t.t.).

Rifaldi, D., Stiyo Famuji, T., Okta Sari Miranda, B., Purma Ramadhan, F., Putri Mulyadi, I., Saputra, V., & Pramuja Inngam Fanani, G. (2025). Evaluasi Sentimen Pengguna ChatGPT Menggunakan Naive Bayes: Tinjauan dari Confusion Matrix dan Classification Report Evaluation ChatGPT User

Sentiment using Naive Bayes: A Review of Confusion Matrix and Classification Report. *Jurnal Riset Sistem dan Teknologi Informasi (RESTIA*, *3*(2), 81–89.

Riza Andrian Septian, & Sita Deliyana Firmialy. (2023). Pengaruh Citra Merek Dan Ulasan Pelanggan Terhadap Keputusan Pembelian Skintific. *Jurnal Ekuilnomi*, *5*(2), 425–432. https://doi.org/10.36985/ekuilnomi.v5i2.759

Salsabillah, D. F., Ratnawati, D. E., & Setiawan, N. Y. (2024). Analisis Sentimen Ulasan Rumah Makan Menggunakan Perbandingan Algoritma *Support Vector Machine* dengan *Naive bayes* (Studi Kasus: Ayam Goreng Nelongso Cabang Singosari, Malang). *Jurnal Teknologi Informasi dan Ilmu Komputer*, *11*(1), 107–116. https://doi.org/10.25126/jtiik.20241117584

Sasono, E., Adi Prasetyo, R., & Tinggi Ilmu Ekonomi Semarang, S. (2023). *Membangun Kepuasan Pelanggan Melalui Kualitas Pelayanan, Persepsi Harga dan Lokasi Pada Konsumen pt. Alya tour semarang*. *15*. https://doi.org/10.33747

Sentimen, A., Wisata, O., Di, B., Maps, G., Siti Utami, D., Utami, D. S., & Erfina, A. (2022). Analisis Sentimen Objek Wisata Bali Di Google Maps Menggunakan Algoritma Naive Bayes. Dalam *Jurnal Sains Komputer & Informatika (J-SAKTI* (Vol. 6, Nomor 1).

Shalihah, G., Kurniawan, R., & Suprapti, T. (2024). Analisis Sentimen Ulasan Pelanggan Mie Gacoan Menggunakan Metode Naïve Bayes Classifier. Dalam *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Nomor 1). https://dosbing.id/wp-content/uploads/2019/11/Annotation-

Zidny, T. (2024a). Analisis Sentimen Ulasan Mie Gacoan Solo Veteran Di Google Maps Menggunakan Algoritma Naive Bayes. Dalam *Analisis Sentimen Ulasan Mie Gacoan Solo Veteran…. ZONAsi: Jurnal Sistem Informasi* (Vol. 6, Nomor 3).

Zidny, T. (2024b). Analisis Sentimen Ulasan Mie Gacoan Solo Veteran Di Google Maps Menggunakan Algoritma Naive Bayes. Dalam *Analisis Sentimen Ulasan Mie Gacoan Solo Veteran…. ZONAsi: Jurnal Sistem Informasi* (Vol. 6, Nomor 3).

LAMPIRAN

Lampiran 1 Skenario Pengujian Website

1. Pengujian Halaman Login

NO	Skenario Pengujian	Input/Aksi	Output yang diharapkan	Hasil
1	Mengosongkan semua field input lalu menekan tombol Login	Mengosongkan field Username dan Password, klik tombol Login	"Please fill out this field."	Berhasil
2	Memasukkan kombinasi Username atau Password salah	Menginput data salah (misal: admin321 / salah password), klik tombol Login	Username atau password salah.	Berhasil
3	Memasukkan kombinasi Username dan Password yang benar	Menginput data akun admin yang terdaftar, klik tombol Login	Sistem berhasil memverifikasi akun dan mengalihkan redirect ke halaman Dashboard Admin.	Berhasil

2. Pengujian Halaman Dashboard

NO	Skenario Pengujian	Input/Aksi	Output yang diharapkan	Hasil
1	Memeriksa validitas tampilan ringkasan data statistik utama.	Membuka halaman Dashboard Admin sesaat	Menampilkan 4 kartu informasi (Total Data, *Preprocessing*, Klasifikasi, Laporan)	Berhasil

		setelah berhasil masuk ke sistem.	dengan angka kalkulasi yang sesuai.	
2	Menguji navigasi menu sidebar	Klik menu *Upload* Data, *Preprocessing* Data, dll. pada sidebar	Mengalihkan ke halaman sesuai dengan item menu yang dipilih.	Berhasil
3	Keluar dari sistem (Logout)	Klik tombol Keluar di bagian bawah sidebar	Menghapus sesi login admin dan mengalihkan kembali ke halaman Login.	Berhasil

3. Pengujian Halaman *Upload* Data

NO	Skenario Pengujian	Input/Aksi	Output yang diharapkan	Hasil
1	Mengunggah dataset tanpa memilih berkas terlebih dahulu.	Langsung mengklik tombol "*Upload* dan Ganti Dataset" saat kondisi form masih kosong (No file chosen).	“Please fill out this field.”	Berhasil
2	Mengunggah dataset ulasan pelanggan Jemberkamera berformat valid.	Memilih file .csv dari penyimpanan lokal, lalu klik "*Upload* dan Ganti Dataset".	Memperbarui status nama berkas aktif, dan memunculkan notifikasi sukses.	Berhasil
3	Memeriksa fungsi	Meninjau tabel "Preview Data Mentah" dan	Tabel sukses merender baris	Berhasil

	pratinjau data mentah dan tombol hapus dataset.	mencoba mengklik tombol "Hapus Dataset".	ulasan lengkap dengan pagination, dan tombol hapus mampu mengosongkan dataset aktif.	

4. Pengujian Halaman Prepocessing

NO	Skenario Pengujian	Input/Aksi	Output yang diharapkan	Hasil
1	Menjalankan proses pipeline pembersihan teks ulasan	Klik tombol Jalankan *Preprocessing* pada dataset aktif	Sistem mengeksekusi fungsi backend dan secara berurutan mengisi tabel tahapan teks: *CleaningCase Folding* Normalisasi *Tokenizing Stopword Removal Stemming* (Sastrawi) beserta indikator status sukses.	Berhasil

5. Pengujian Halaman TF-IDF

NO	Skenario Pengujian	Input/Aksi	Output yang diharapkan	Hasil
1	Menghitung bobot kata pada teks hasil *Preprocessing*	Klik tombol Hitung Pembobotan	Sistem memproses frekuensi kata dan nilai inversi dokumen, kemudian menampilkan hasilnya secara terstruktur di dalam komponen *accordion collapse matrix* (Term Frequency, DF-IDF, dan TF-IDF).	Berhasil
2	Membuka detail isi matriks pembobotan	Klik ikon *dropdown (arrow down)* pada baris *Term Frequency* (TF) / DF-IDF / TF-IDF	Komponen meluas (*expand*) ke bawah dan menampilkan detail baris token kata beserta nilai bobot numeriknya secara presisi.	Berhasil

6. Pengujian Halaman Klasifikasi

NO	Skenario Pengujian	Input/Aksi	Output yang diharapkan	Hasil
1	Menekan tombol proses klasifikasi tanpa memilih rasio split data	Membiarkan dropdown Split Rasio pada pilihan default ("Pilih Rasio	Sistem menampilkan notifikasi pilih split ratio terlebih dahulu sebelum proses klasifikasi.	Berhasil

		split data"), lalu klik tombol Proses klasifikasi Naive Bayes		
2	Memilih nilai proporsi pembagian data pada dropdown	Klik komponen dropdown Split Rasio dan memilih salah satu nilai (misal: 80:20 atau 70:30)	Sistem berhasil menangkap input parameter nilai rasio pembagian data secara presisi, lalu menyiapkan backend untuk memproses dual skema pengujian, yaitu pembagian data awal (Tanpa SMOTE) serta rekayasa penyeimbangan data latih (Dengan SMOTE) secara simultan.	Berhasil

7. Pengujian Halaman Hasil dan Laporan

NO	Skenario Pengujian	Input/Aksi	Output yang diharapkan	Hasil
1	Memeriksa keakuratan ringkasan statistik pembagian label ulasan.	Membuka halaman Hasil dan Laporan sesaat setelah sistem selesai melakukan pemodelan data.	Menampilkan total data, Positif, Netral dan Negatif.	Berhasil

2	Meninjau diagram komparasi grafik batang distribusi data training.	Memeriksa area visualisasi diagram pada blok menu "Distribusi Data Training".	Sistem sukses merender grafik batang Original (terlihat timpang) dan grafik Balanced (tinggi batang sama rata).	Berhasil
3	Mengevaluasi visualisasi matriks kesalahan (Confusion Matrix).	Memeriksa tabel silang sebaran prediksi pada blok menu "Confusion Matrix".	Sistem menampilkan tabel koordinat aktual vs prediksi untuk menghitung kebenaran klasifikasi data uji secara terstruktur.	
4	Sistem menampilkan tabel koordinat aktual vs prediksi untuk menghitung kebenaran klasifikasi data uji secara terstruktur.	Memeriksa baris data evaluasi pada blok menu komponen "Classification Report".	Sistem memuat rincian performa metrik precision, recall, dan *F1-score* spesifik untuk setiap baris kelas sentimen secara presisi.	Berhasil

Lampiran 2 Lembar Validasi Pakar Bahasa

KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
POLITEKNIK NEGERI JEMBER
JURUSAN TEKNOLOGI INFORMASI
PROGRAM STUDI TEKNIK INFORMATIKA
Jalan Mastrip Kotak Pos 164 Jember 68101 Telp. (0331) 333552-34, Faks. (0331) 333531 e-mail: politeknik@polije.ac.id ; website: http://www.polije.ac.id

SURAT PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama	: Anggik Budi Prasetiyo, S.Pd., M.Li., Gr.
Jabatan	: Guru Bahasa Indonesia
Instansi	: SMA Negeri 3 Jember

Menyatakan dengan sebenar-benarnya bahwa mahasiswa ini:

Nama	: Moh. Bachtiar Rifki Habibi
NIM	: E41220474
Judul Skripsi	:Analisis Ulasan Pelanggan Pada Sewa Kamera Di Jemberkamera Menggunakan Algoritme Naive Bayes
Jurusan	: Teknologi Informasi
Perguruan Tinggi	: Politeknik Negeri Jember

Sehubung dengan skripsi yang berjudul "Analisis Ulasan Pelanggan Pada Sewa Kamera Di Jemberkamera Menggunakan Algoritme Naive Bayes
Menyatakan bahwa:

Data yang meliputi ulasan pengguna aplikasi Google Maps yang telah dilabeling benar-benar valid tanpa ada perubahan yan mempengaruhi makna asli dari data yang di dapatkan.

Demikian surat pernyataan ini dibuat dengan sebenar-benarnya.

Jember, 6 Juli 2026

Anggik Budi Prasetiyo, S.Pd., M.Li., Gr.
NIP 199604302024211018

Lampiran 3 Lembar Validasi 1 Pengujian Web App Oleh Mitra

KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
POLITEKNIK NEGERI JEMBER
JURUSAN TEKNOLOGI INFORMASI
PROGRAM STUDI TEKNIK INFORMATIKA
Jalan Mastrip Kotak Pos 164 Jember 68101 Telp (0331) 333532-34, Faks (0331) 333531 e-mail: politeknik@polije.ac.id ; website: http://www.polije.ac.id

SURAT KETERANGAN VALIDASI PENGUJIAN APLIKASI

Yang bertanda tangan di bawah ini:

Nama	:	Dewi Milfin
Profesi / Jabatan	:	Front Officer
Instansi / Mitra	:	Jember Kamera

Menyatakan dengan sebenarnya bahwa saya telah melakukan proses peninjauan, validasi, dan pengujian aplikasi secara independen menggunakan metode *Black Box Testing* terhadap sistem perangkat lunak yang dikembangkan untuk keperluan penelitian skripsi / tugas akhir oleh:

Nama Mahasiswa	:	Moh. Bachtiar Rifki Habibi
NIM	:	E41220474
Program Studi	:	Teknik Informatika
Judul Skripsi	:	Analisis Ulasan Pelanggan Pada Sewa Kamera di Jemberkamera Menggunakan Algoritma Naive Bayes

Demikian surat keterangan validasi pengujian ini dibuat dengan sadar dan sebenar-benarnya secara profesional untuk dapat dipergunakan sebagaimana mestinya.

Penulis,

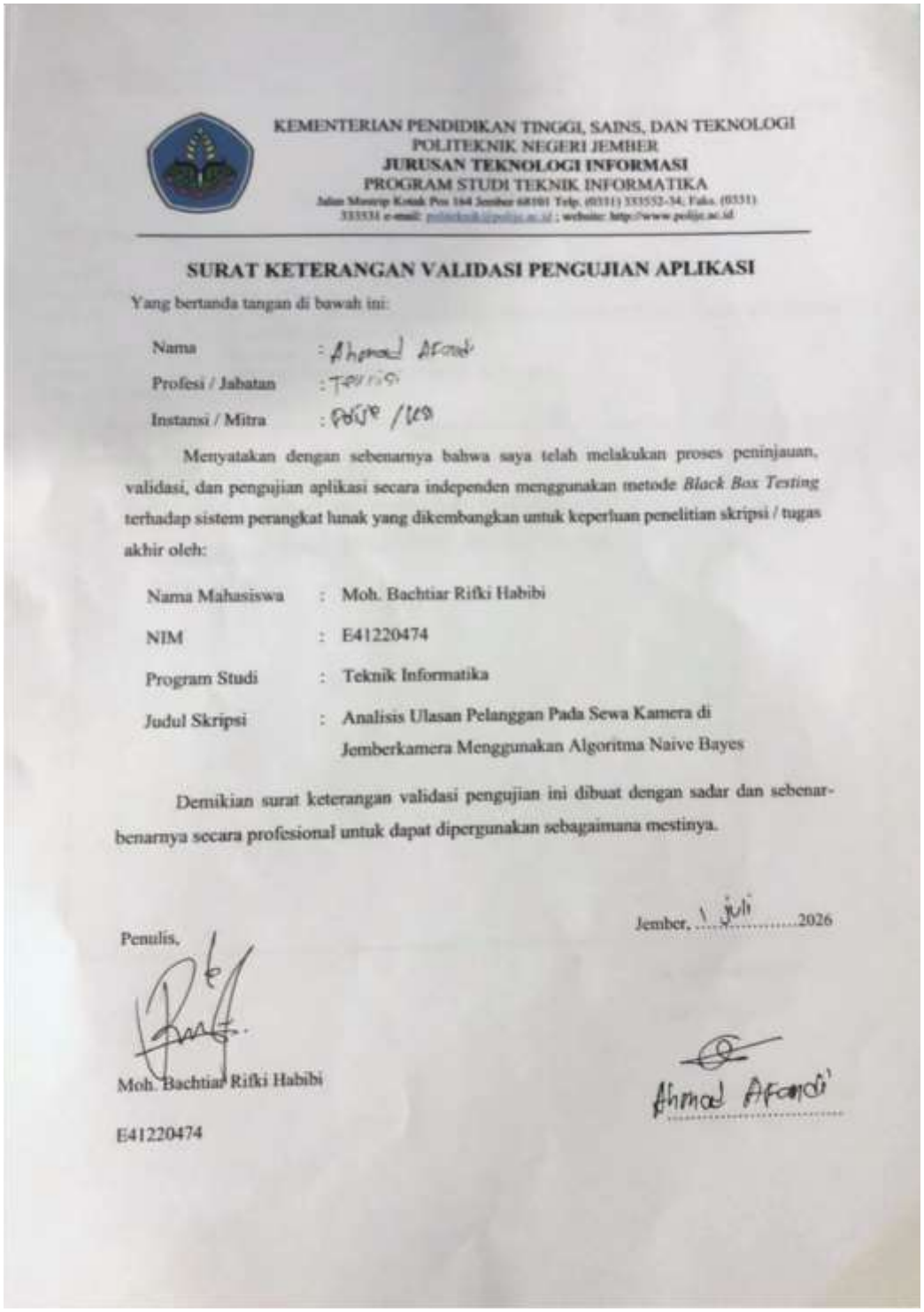
Moh. Bachtiar Rifki Habibi
E41220474

Jember, 19 Juni 2026
Pihak Mitra / Validator,

JEMBERKAMERA
rental • jasa foto video • service

Dewi Milfin

Lampiran 4 Lembar Validasi 2 Pengujian Web App Oleh Teknisi Jurusan TIF



KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
POLITEKNIK NEGERI JEMBER
JURUSAN TEKNOLOGI INFORMASI
PROGRAM STUDI TEKNIK INFORMATIKA
Jalan Mastrip Kotak Pos 164 Jember 68101 Telp. (0331) 333532-34, Faks. (0331) 333531 e-mail: politeknik@polije.ac.id ; website: http://www.polije.ac.id

SURAT KETERANGAN VALIDASI PENGUJIAN APLIKASI

Yang bertanda tangan di bawah ini:

Nama : Ahmad Afandi
Profesi / Jabatan : Teknisi
Instansi / Mitra : Polije / LCR

Menyatakan dengan sebenarnya bahwa saya telah melakukan proses peninjauan, validasi, dan pengujian aplikasi secara independen menggunakan metode *Black Box Testing* terhadap sistem perangkat lunak yang dikembangkan untuk keperluan penelitian skripsi / tugas akhir oleh:

Nama Mahasiswa	:	Moh. Bachtiar Rifki Habibi
NIM	:	E41220474
Program Studi	:	Teknik Informatika
Judul Skripsi	:	Analisis Ulasan Pelanggan Pada Sewa Kamera di Jemberkamera Menggunakan Algoritma Naive Bayes

Demikian surat keterangan validasi pengujian ini dibuat dengan sadar dan sebenar-benarnya secara profesional untuk dapat dipergunakan sebagaimana mestinya.

Jember, 1 Juli 2026

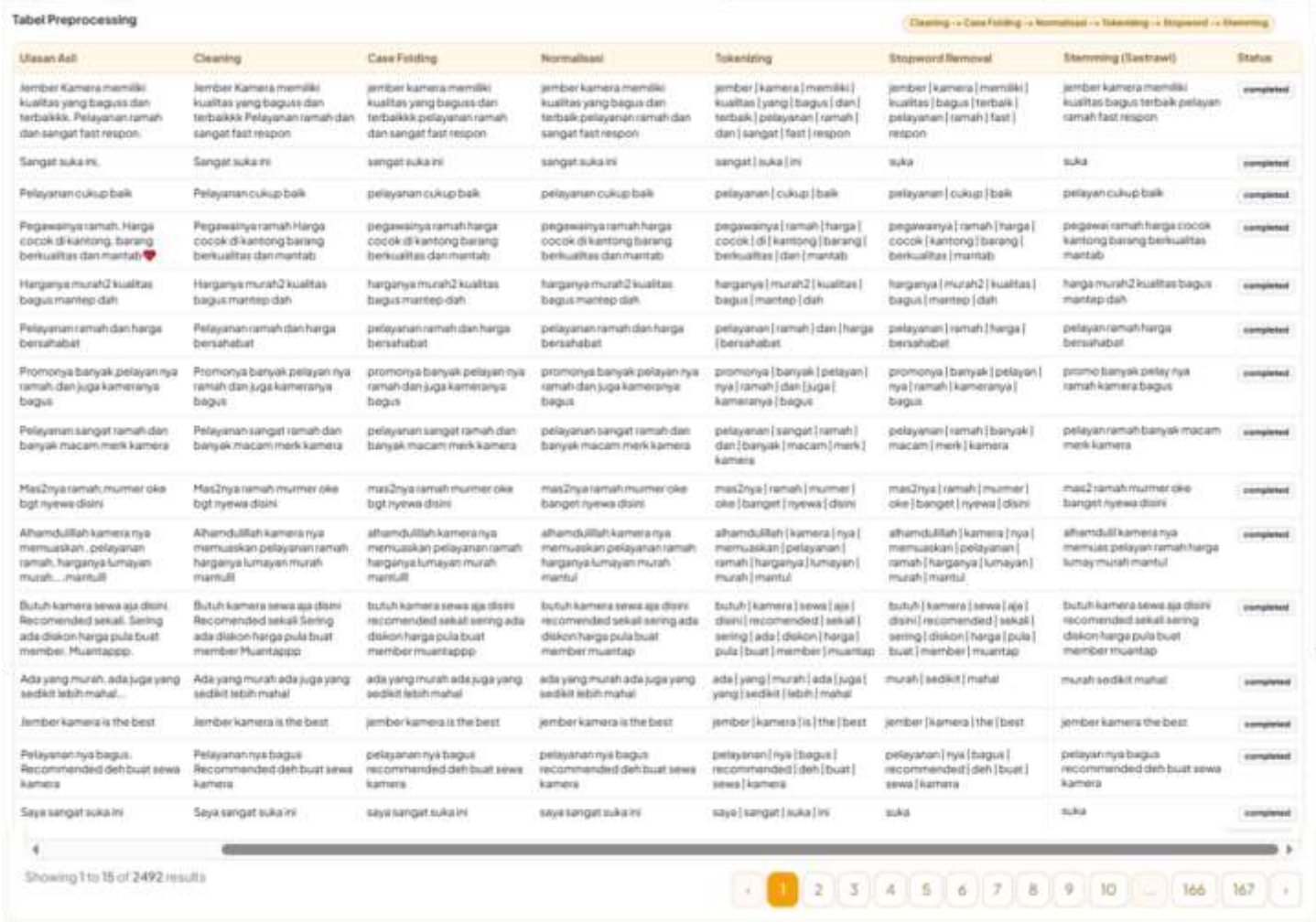
Penulis,

Moh. Bachtiar Rifki Habibi

E41220474

Ahmad Afandi

Lampiran 5 Detail Antarmuka Hasil Halaman Prepocessing



Tabel Preprocessing

Cleaning → Case Folding → Normalisasi → Tokenizing → Stopword → Stemming

Ulasan Asli	Cleaning	Case Folding	Normalisasi	Tokenizing	Stopword Removal	Stemming (Sastrawi)	Status
Jember Kamera memiliki kualitas yang baguss dan terbaikkk. Pelayanan ramah dan sangat fast respon.	Jember Kamera memiliki kualitas yang baguss dan terbaikkk Pelayanan ramah dan sangat fast respon	jember kamera memiliki kualitas yang baguss dan terbaikkk pelayanan ramah dan sangat fast respon	jember kamera memiliki kualitas yang bagus dan terbaik pelayanan ramah dan sangat fast respon	jember | kamera | memiliki | kualitas | yang | bagus | dan | terbaik | pelayanan | ramah | dan | sangat | fast | respon	jember | kamera | memiliki | kualitas | bagus | terbaik | pelayanan | ramah | fast | respon	jember kamera memiliki kualitas bagus terbaik pelayan ramah fast respon	completed
Sangat suka ini.	Sangat suka ini	sangat suka ini	sangat suka ini	sangat | suka | ini	suka	suka	completed
Pelayanan cukup baik	Pelayanan cukup baik	pelayanan cukup baik	pelayanan cukup baik	pelayanan | cukup | baik	pelayanan | cukup | baik	pelayan cukup baik	completed
Pegawainya ramah. Harga cocok di kantong, barang berkualitas dan mantab❤️	Pegawainya ramah Harga cocok di kantong barang berkualitas dan mantab	pegawainya ramah harga cocok di kantong barang berkualitas dan mantab	pegawainya ramah harga cocok di kantong barang berkualitas dan mantab	pegawainya | ramah | harga | cocok | di | kantong | barang | berkualitas | dan | mantab	pegawainya | ramah | harga | cocok | kantong | barang | berkualitas | mantab	pegawai ramah harga cocok kantong barang berkualitas mantab	completed
Harganya murah2 kualitas bagus mantep dah	Harganya murah2 kualitas bagus mantep dah	harganya murah2 kualitas bagus mantep dah	harganya murah2 kualitas bagus mantep dah	harganya | murah2 | kualitas | bagus | mantep | dah	harganya | murah2 | kualitas | bagus | mantep | dah	harga murah2 kualitas bagus mantep dah	completed
Pelayanan ramah dan harga bersahabat	Pelayanan ramah dan harga bersahabat	pelayanan ramah dan harga bersahabat	pelayanan ramah dan harga bersahabat	pelayanan | ramah | dan | harga | bersahabat	pelayanan | ramah | harga | bersahabat	pelayan ramah harga bersahabat	completed
Promonya banyak,pelayan nya ramah dan juga kameranya bagus	Promonya banyak pelayan nya ramah dan juga kameranya bagus	promonya banyak pelayan nya ramah dan juga kameranya bagus	promonya banyak pelayan nya ramah dan juga kameranya bagus	promonya | banyak | pelayan | nya | ramah | dan | juga | kameranya | bagus	promonya | banyak | pelayan | nya | ramah | kameranya | bagus	promo banyak pelay nya ramah kamera bagus	completed
Pelayanan sangat ramah dan banyak macam merk kamera	Pelayanan sangat ramah dan banyak macam merk kamera	pelayanan sangat ramah dan banyak macam merk kamera	pelayanan sangat ramah dan banyak macam merk kamera	pelayanan | sangat | ramah | dan | banyak | macam | merk | kamera	pelayanan | ramah | banyak | macam | merk | kamera	pelayan ramah banyak macam merk kamera	completed
Mas2nya ramah,murmer oke bgt nyewa disini	Mas2nya ramah murmer oke bgt nyewa disini	mas2nya ramah murmer oke bgt nyewa disini	mas2nya ramah murmer oke banget nyewa disini	mas2nya | ramah | murmer | oke | banget | nyewa | disini	mas2nya | ramah | murmer | oke | banget | nyewa | disini	mas2 ramah murmer oke banget nyewa disini	completed
Alhamdulillah kamera nya memuaskan , pelayanan ramah, harganya lumayan murah.....mantull	Alhamdulillah kamera nya memuaskan pelayanan ramah harganya lumayan murah mantull	alhamdulillah kamera nya memuaskan pelayanan ramah harganya lumayan murah mantull	alhamdulillah kamera nya memuaskan pelayanan ramah harganya lumayan murah mantul	alhamdulillah | kamera | nya | memuaskan | pelayanan | ramah | harganya | lumayan | murah | mantul	alhamdulillah | kamera | nya | memuaskan | pelayanan | ramah | harganya | lumayan | murah | mantul	alhamdulil kamera nya memuas pelayan ramah harga lumay murah mantul	completed
Butuh kamera sewa aja disini. Recomended sekali. Sering ada diskon harga pula buat member. Muantapppp.	Butuh kamera sewa aja disini Recomended sekali Sering ada diskon harga pula buat member Muantapppp	butuh kamera sewa aja disini recomended sekali sering ada diskon harga pula buat member muantapppp	butuh kamera sewa aja disini recomended sekali sering ada diskon harga pula buat member muantap	butuh | kamera | sewa | aja | disini | recomended | sekali | sering | ada | diskon | harga | pula | buat | member | muantap	butuh | kamera | sewa | aja | disini | recomended | sekali | sering | diskon | harga | pula | buat | member | muantap	butuh kamera sewa aja disini recomended sekali sering diskon harga pula buat member muantap	completed
Ada yang murah, ada juga yang sedikit lebih mahal...	Ada yang murah ada juga yang sedikit lebih mahal	ada yang murah ada juga yang sedikit lebih mahal	ada yang murah ada juga yang sedikit lebih mahal	ada | yang | murah | ada | juga | yang | sedikit | lebih | mahal	murah | sedikit | mahal	murah sedikit mahal	completed
Jember kamera is the best	Jember kamera is the best	jember kamera is the best	jember kamera is the best	jember | kamera | is | the | best	jember | kamera | the | best	jember kamera the best	completed
Pelayanan nya bagus. Recommended deh buat sewa kamera	Pelayanan nya bagus Recommended deh buat sewa kamera	pelayanan nya bagus recommended deh buat sewa kamera	pelayanan nya bagus recommended deh buat sewa kamera	pelayanan | nya | bagus | recommended | deh | buat | sewa | kamera	pelayanan | nya | bagus | recommended | deh | buat | sewa | kamera	pelayan nya bagus recommended deh buat sewa kamera	completed
Saya sangat suka ini	Saya sangat suka ini	saya sangat suka ini	saya sangat suka ini	saya | sangat | suka | ini	suka	suka	completed

Showing 1 to 15 of 2492 results

‹ 1 2 3 4 5 6 7 8 9 10 ... 166 167 ›

Lampiran 6 Detail Antarmuka Hasil Halaman TF

Rumus mengikuti contoh workbook manual: TF = frekuensi term / jumlah token, DF = jumlah dokumen yang memuat term, IDF = log(total dokumen / DF), lalu TF-IDF = TF x IDF.

TF

Dokumen	kamera	ramah	pelayan	harga	bagus	nya	banyak	buat	disini	jember	kualitas	murah	sewa	suka	aja	alhamdulil	baik	banget	barang	berkualitas
Doc1	0.100	0.100	0.100	0.000	0.100	0.000	0.000	0.000	0.000	0.100	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc3	0.000	0.000	0.333	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.333	0.000	0.000	0.000
Doc4	0.000	0.125	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.125
Doc5	0.000	0.000	0.000	0.167	0.167	0.000	0.000	0.000	0.000	0.000	0.167	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc6	0.000	0.250	0.250	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc7	0.143	0.143	0.000	0.000	0.143	0.143	0.143	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc8	0.167	0.167	0.167	0.000	0.000	0.000	0.167	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc9	0.000	0.143	0.000	0.000	0.000	0.000	0.000	0.000	0.143	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.143	0.000	0.000
Doc10	0.100	0.100	0.100	0.100	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000
Doc11	0.071	0.000	0.000	0.071	0.000	0.000	0.000	0.071	0.071	0.000	0.000	0.000	0.071	0.000	0.071	0.000	0.000	0.000	0.000	0.000
Doc12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.333	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc13	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc14	0.125	0.000	0.125	0.000	0.125	0.125	0.000	0.125	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000

Lampiran 7 Detail Antarmuka Hasil Halaman DF-IDF

DF-IDF

Term	DF	IDF
kamera	7	1.762
ramah	7	1.762
pelayan	6	1.916
harga	5	2.099
bagus	4	2.322
nya	3	2.609
banyak	2	3.015
buat	2	3.015
disini	2	3.015
jember	2	3.015
kualitas	2	3.015
murah	2	3.015
sewa	2	3.015
suka	2	3.015

Lampiran 8 Detail Antarmuka Hasil Halaman TF-IDF

TF-IDF

Dokumen	kamera	ramah	pelayan	harga	bagus	nya	banyak	buat	disini	jember	kualitas	murah	sewa	suka	aja	alhamdulil	baik	banget	barang	berkualitas
Doc1	0.176	0.176	0.192	0.000	0.232	0.000	0.000	0.000	0.000	0.301	0.301	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3.015	0.000	0.000	0.000	0.000	0.000	0.000
Doc3	0.000	0.000	0.639	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.236	0.000	0.000	0.000
Doc4	0.000	0.220	0.000	0.262	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.464	0.464
Doc5	0.000	0.000	0.000	0.350	0.387	0.000	0.000	0.000	0.000	0.000	0.502	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc6	0.000	0.441	0.479	0.525	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc7	0.252	0.252	0.000	0.000	0.332	0.373	0.431	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc8	0.294	0.294	0.319	0.000	0.000	0.000	0.502	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc9	0.000	0.252	0.000	0.000	0.000	0.000	0.000	0.000	0.431	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.530	0.000	0.000
Doc10	0.176	0.176	0.192	0.210	0.000	0.261	0.000	0.000	0.000	0.000	0.000	0.301	0.000	0.000	0.000	0.371	0.000	0.000	0.000	0.000
Doc11	0.126	0.000	0.000	0.150	0.000	0.000	0.000	0.215	0.215	0.000	0.000	0.000	0.215	0.000	0.265	0.000	0.000	0.000	0.000	0.000
Doc12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc13	0.441	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.754	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc14	0.220	0.000	0.240	0.000	0.290	0.326	0.000	0.377	0.000	0.000	0.000	0.000	0.377	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Doc15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3.015	0.000	0.000	0.000	0.000	0.000	0.000

Showing 1 to 15 of 2492 results

‹ 1 2 3 4 5 6 7 8 9 10 ... 166 167 ›

Lampiran 9 Dokumentasi Pengujian Web App Oleh Pihak Jemberkamera

Lampiran 10 Visualisasi Word Cloud Positif

Lampiran 11 Visualisasi Word Cloud Netral

Lampiran 12 Visualisasi Word Cloud Negatif