



**ANALISIS SENTIMEN MASYARAKAT TERHADAP DESTINASI  
PRIORITAS PARIWISATA DI INDONESIA MENGGUNAKAN  
ALGORITMA NAÏVE BAYES CLASSIFIER**

**TUGAS AKHIR**



**Oleh:**  
**ALEXANDER STEVEN SUTANTO**  
**21410100033**

---

**FAKULTAS TEKNOLOGI DAN INFORMATIKA**  
**UNIVERSITAS DINAMIKA**  
**2025**

**ANALISIS SENTIMEN MASYARAKAT TERHADAP DESTINASI  
PRIORITAS PARIWISATA DI INDONESIA MENGGUNAKAN  
ALGORITMA NAÏVE BAYES CLASSIFIER**

**TUGAS AKHIR**

**Diajukan sebagai salah satu syarat untuk menyelesaikan  
Program Sarjana**



**UNIVERSITAS  
Dinamika**

**Oleh:**

**Nama : Alexander Steven Sutanto  
NIM : 21410100033  
Program Studi : S1 Sistem Informasi**

**FAKULTAS TEKNOLOGI DAN INFORMATIKA  
UNIVERSITAS DINAMIKA**

**2025**

## Tugas Akhir

# ANALISIS SENTIMEN MASYARAKAT TERHADAP DESTINASI PRIORITAS PARIWISATA DI INDONESIA MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER

Dipersiapkan dan disusun Oleh

**Alexander Steven Sutanto**

**NIM: 21410100033**

Telah diperiksa, dibahas dan disetujui oleh Dewan Pembahas

Pada: Rabu, 6 Agustus 2025

### Susunan Dewan Pembahas

#### Pembimbing

I. Julianto Lemantara, S.Kom., M.Eng.

NIDN. 0722108601

II. Prof. Dr. M.J. Dewiyani Sunarto

NIDN. 0725076301

#### Pembahas

I. Tutut Wuriyanto, M.Kom.

NIDN. 0703056702

Digitally signed  
by Julianto

Date: 2025.08.06  
17:58:25 +07'00'

Digitally signed by  
Dewiyani

Tugas Akhir ini telah diterima sebagai salah satu persyaratan

Untuk memperoleh gelar sarjana

Digitally signed by  
Julianto

Date: 2025.08.08  
16:52:17 +07'00'

**Julianto Lemantara, S.Kom., M.Eng**

NIDN. 0722108601

Dekan Fakultas Teknologi dan Informatika

UNIVERSITAS DINAMIKA



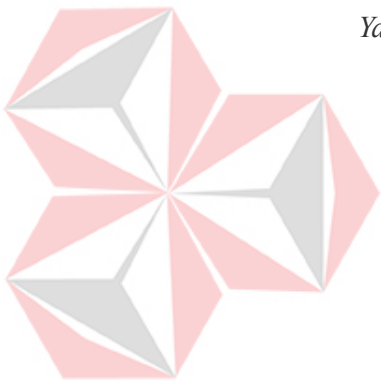
*Katakan pada dirimu*  
*“Besok, mungkin kita sampai”*  
*“Besok mungkin tercapai”*  
*- Hindia -*

UNIVERSITAS  
**Dinamika**

*Saya persembahkan Tugas Akhir ini untuk diri saya sendiri,*

*Alexander Steven Sutanto*

*Yang telah berhasil menuntaskan studi dengan seluruh usahanya*



UNIVERSITAS  
**Dinamika**

**PERNYATAAN**  
**PERSETUJUAN PUBLIKASI DAN KEASLIAN KARYA ILMIAH**

Sebagai mahasiswa Universitas Dinamika, Saya :

Nama : Alexander Steven Sutanto  
NIM : 21410100033  
Program Studi : S1 Sistem Informasi  
Fakultas : Fakultas Teknologi dan Informatika  
Jenis Karya : Tugas Akhir  
Judul Karya : ANALISIS SENTIMEN MASYARAKAT TERHADAP  
DESTINASI PRIORITAS PARIWISATA DI  
INDONESIA MENGGUNAKAN ALGORITMA  
NAÏVE BAYES CLASSIFIER

Menyatakan dengan sesungguhnya bahwa :

1. Demi pengembangan Ilmu Pengetahuan, Teknologi dan Seni, Saya menyetujui memberikan kepada Universitas Dinamika Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty Free Right*) atas seluruh isi/sebagian karya ilmiah Saya tersebut diatas untuk disimpan, dialihmediakan, dan dikelola dalam bentuk pangkalan data (*database*) untuk selanjutnya didistribusikan atau dipublikasikan demi kepentingan akademis dengan tetap mencantumkan nama Saya sebagai penulis atau pencipta dan sebagai pemilik Hak Cipta.
2. Karya tersebut diatas adalah hasil karya asli Saya, bukan plagiat baik sebagian maupun keseluruhan. Kutipan, karya, atau pendapat orang lain yang ada dalam karya ilmiah ini semata-mata hanya sebagai rujukan yang dicantumkan dalam Daftar Pustaka Saya.
3. Apabila dikemudian hari ditemukan dan terbukti terdapat tindakan plagiasi pada karya ilmiah ini, maka Saya bersedia untuk menerima pencabutan terhadap gelar kesarjanaan yang telah diberikan kepada Saya.

Demikian surat pernyataan ini Saya buat dengan sebenar-benarnya.

Surabaya, 30 Juli 2024



Alexander Steven Sutanto  
NIM : 21410100033

## ABSTRAK

Pertumbuhan sektor pariwisata memberikan kontribusi signifikan terhadap ekonomi nasional dalam dua dekade terakhir. Untuk memperkuat industri ini, Kementerian Pariwisata dan Ekonomi Kreatif menetapkan 10 destinasi prioritas, lima di antaranya menjadi fokus dalam penelitian ini, yaitu Pulau Morotai, Pantai Tanjung Kelayang, Danau Toba, Wakatobi, dan Pantai Tanjung Lesung yang mencakup wilayah pantai, laut, dan danau. Salah satu komponen penting dalam meningkatkan minat kunjungan wisatawan, khususnya di era digital, adalah ulasan yang mereka sampaikan secara daring. Oleh karena itu, penting untuk memahami persepsi mereka terhadap destinasi wisata. Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan wisatawan berbasis data Google Maps menggunakan algoritma *Naïve Bayes Classifier*, guna membantu pengelola memahami persepsi pengunjung serta mengidentifikasi aspek-aspek yang perlu diperbaiki. Ulasan diklasifikasikan ke dalam dua kategori, yaitu sentimen positif dan negatif, kemudian dilatih pada model *Naïve Bayes* yang dievaluasi menggunakan *confusion matrix* dan teknik *10-fold cross-validation*. Hasil klasifikasi menunjukkan akurasi yang tinggi di seluruh destinasi, dengan Danau Toba mencapai akurasi tertinggi sebesar 95.91%, Wakatobi sebesar 94.67%, Pantai Tanjung Kelayang sebesar 93.94%, Pulau Morotai sebesar 90.69%, dan Pantai Tanjung Lesung sebesar 90.13%. Tingkat akurasi sebesar 90–100% menunjukkan bahwa hasil klasifikasi sangat baik, mencerminkan ketepatan prediksi yang tinggi terhadap data aktual. Sentimen positif terbesar didapat oleh Wakatobi sebesar 97.48% dan sentimen negatif tertinggi didapat oleh Pantai Tanjung Lesung sebesar 27.85%. Selain itu, visualisasi kata menggunakan Wordcloud yang dibuat berdasarkan data ulasan yang telah dilabeli namun belum melalui proses klasifikasi dengan algoritma *Naïve Bayes* berhasil mengungkap daya tarik utama dari masing-masing destinasi pada ulasan positif. Sementara itu, pada ulasan negatif, visualisasi ini berhasil mengidentifikasi kata-kata dominan seperti 'sampah', 'parkir', dan 'mahal' yang menjadi indikator aspek yang perlu mendapat perhatian dari pengelola. Dengan demikian, hasil penelitian ini tidak hanya menggambarkan persepsi pengunjung secara akurat, tetapi juga memberikan masukan konkret untuk meningkatkan kualitas layanan dan pengalaman wisata.

**Kata Kunci:** *Analisis Sentimen, Destinasi Prioritas, Naïve Bayes*

## KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat, karunia, serta bimbingan-Nya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang berjudul “ANALISIS SENTIMEN MASYARAKAT TERHADAP DESTINASI PRIORITAS PARIWISATA DI INDONESIA MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER” ini dengan baik dan lancar. Laporan ini disusun sebagai salah satu syarat untuk menyelesaikan program studi di Universitas Dinamika. terselesaikannya laporan ini tentunya tidak terlepas dari dukungan, bimbingan, kritik, dan saran dari berbagai pihak yang telah membantu penulis selama proses penyusunan laporan. Oleh karena itu, pada kesempatan ini penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Keluarga tercinta yang senantiasa memberikan doa, semangat, dan dukungan sepenuhnya kepada penulis. Terima kasih atas kepercayaan dan keyakinannya bahwa penulis mampu menyelesaikan studi ini dengan baik hingga tuntas.
2. Bapak Julianto Lemantara, S.Kom., M.Eng. selaku Dosen Pembimbing I yang telah memberikan bimbingan, membantu memberikan masukan perbaikan, dan menyempurnakan isi laporan agar lebih baik.
3. Ibu Prof. Dr. M.J. Dewiyani Sunarto selaku Dosen Pembimbing II yang telah memberikan bimbingan dan membantu memberikan masukan perbaikan selama proses penyelesaian tugas akhir.
4. Bapak Tutut Wuriyanto, M.Kom., selaku Dosen Penguji yang telah berkenan menguji hasil tugas akhir ini.
5. Teman-teman S1 Sistem Informasi angkatan 2021 yang telah memberikan semangat dan menemani penulis sejak menjadi mahasiswa baru hingga mampu menyelesaikan laporan tugas akhir ini.
6. Raysita Syahnas Sharon, yang senantiasa menjadi pengingat di saat penulis mulai lelah dan ragu, serta tak pernah berhenti memberikan semangat dan dukungan kepada penulis. Terima kasih atas setiap waktu, perhatian, dan ketulusan yang telah diberikan dalam menemani penulis melalui setiap proses hingga tugas akhir ini dapat terselesaikan. Semoga kebahagiaan selalu menemani tiap langkahmu.



7. Kepada penulis yang berulang kali ragu terhadap keputusan yang diambil namun tetap meyakinkan dirinya dan menyelesaikan tugas akhir ini hingga tuntas.

Penulis menyadari bahwa laporan tugas akhir ini masih memiliki banyak kekurangan. Untuk itu, penulis sangat mengharapkan masukan dan saran yang bersifat membangun dari berbagai pihak demi penyempurnaan di masa mendatang. Sebagai penutup, penulis berharap laporan ini dapat memberikan manfaat bagi para pembaca serta menjadi sumbangsih positif dalam pengembangan ilmu pengetahuan.

Surabaya, 23 Juli 2025

Penulis



UNIVERSITAS  
**Dinamika**

## DAFTAR ISI

	Halaman
<b>ABSTRAK .....</b>	<b>vii</b>
<b>KATA PENGANTAR.....</b>	<b>viii</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>DAFTAR TABEL .....</b>	<b>xiv</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xv</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Masalah.....	4
1.4 Tujuan.....	4
1.5 Manfaat.....	4
<b>BAB II LANDASAN TEORI .....</b>	<b>5</b>
2.1 Penelitian Sebelumnya .....	5
2.2 Sektor Wisata Alam Pantai .....	6
2.3 CRISP-DM.....	6
2.4 Analisis Sentimen.....	8
2.5 <i>Web Scraping</i> .....	8
2.6 SerpAPI .....	9
2.7 <i>Data Cleaning</i> .....	9
2.8 <i>Text Mining</i> .....	9
2.9 Naïve Bayes.....	11
2.10 Confusion Matrix .....	12
2.11 <i>K-Fold Cross-Validation</i> .....	13
2.12 Augmentasi Data .....	13
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>14</b>
3.1 <i>Business Understanding</i> .....	14
3.1.1 Studi Literatur.....	14
3.1.2 Identifikasi Proses Bisnis .....	15

3.1.3	Identifikasi Masalah .....	15
3.2	<i>Data Understanding</i> .....	16
3.2.1	Analisis Kebutuhan Sistem dan Data .....	16
3.2.2	<i>Data Scraping</i> .....	17
3.3	<i>Data Preparation</i> .....	17
3.3.1	<i>Data Cleaning</i> .....	18
3.3.2	Pelabelan.....	19
3.3.3	Pembagian <i>Data Training</i> dan <i>Data Testing</i> .....	20
3.3.4	Augmentasi Data .....	20
3.3.5	<i>Data Preprocessing</i> .....	21
3.4	<i>Modelling</i> .....	23
3.5	Evaluasi ( <i>Confusion Matrix</i> ).....	28
3.6	<i>Deployment</i> (Visualisasi).....	28
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>		<b>29</b>
4.1	Analisis.....	29
4.2	Hasil <i>Web Scraping</i> .....	29
4.3	<i>Data Cleaning</i> .....	30
4.4	Pelabelan .....	31
4.5	Pembagian <i>Data Training</i> dan <i>Data Testing</i> .....	32
4.6	Augmentasi Data .....	32
4.7	<i>Data Preprocessing</i> .....	33
4.7.1	<i>Case Folding</i> .....	33
4.7.2	<i>Tokenizing</i> .....	34
4.7.3	<i>Stop Word Removal</i> .....	35
4.7.4	<i>Stemming</i> .....	35
4.7.5	Hasil <i>Preprocessing</i> .....	36
4.8	Implementasi Naïve Bayes.....	37
4.9	Hasil Evaluasi Naïve Bayes .....	43
4.9.1	Hasil <i>Classification Report</i> .....	44
4.9.2	Hasil <i>10-fold cross-validation</i> .....	44
4.10	Hasil Akhir Naïve Bayes .....	45
4.11	Visualisasi.....	46

4.11.1	Visualisasi Pulau Morotai .....	46
4.11.2	Visualisasi Pantai Tanjung Kelayang .....	47
4.11.3	Visualisasi Danau Toba .....	48
4.11.4	Visualisasi Wakatobi .....	48
4.11.5	Visualisasi Pantai Tanjung Lesung .....	49
4.11.6	Hasil Akhir Visualisasi .....	50
<b>BAB V</b>	<b>PENUTUP .....</b>	<b>51</b>
5.1	Kesimpulan .....	51
5.2	Saran .....	51
<b>DAFTAR PUSTAKA .....</b>		<b>52</b>
<b>LAMPIRAN .....</b>		<b>57</b>



UNIVERSITAS  
**Dinamika**

## DAFTAR GAMBAR

	Halaman
Gambar 1. 1. Jumlah Kunjungan Wisatawan Mancanegara Agustus 2020 - 2024 .	1
Gambar 2. 1. <i>Framework</i> CRISP-DM .....	6
Gambar 3. 1. Metode CRISP – DM .....	14
Gambar 3. 2. Tahap <i>Business Understanding</i> .....	14
Gambar 3. 3. Tahap <i>Data Understanding</i> .....	16
Gambar 3. 4. Tahap <i>Data Preparation</i> .....	18
Gambar 4. 1 Contoh Hasil Pelabelan .....	32
Gambar 4. 2 Contoh Hasil Augmentasi.....	33
Gambar 4. 3 Tahap <i>Case Folding</i> .....	33
Gambar 4. 4 Hasil <i>Case Folding</i> .....	34
Gambar 4. 5 Tahap <i>Tokenizing</i> .....	34
Gambar 4. 6 Hasil <i>Tokenizing</i> .....	35
Gambar 4. 7 Tahap <i>Stop Word Removal</i> .....	35
Gambar 4. 8 Hasil <i>Stop Word Removal</i> .....	35
Gambar 4. 9 Tahap <i>Stemming</i> .....	35
Gambar 4. 10 Hasil <i>Stemming</i> .....	36
Gambar 4. 11 Tahap Akhir <i>Preprocessing</i> .....	36
Gambar 4. 12 Hasil Akhir <i>Preprocessing</i> .....	37
Gambar 4. 13 Hasil Naive Bayes Python.....	40
Gambar 4. 14 Hasil <i>Confusion Matrix</i> .....	41
Gambar 4. 15 Hasil <i>Classification Report</i> .....	43
Gambar 4. 16 Hasil Evaluasi <i>10-fold cross-validation</i> .....	43
Gambar 4. 17 Hasil <i>Wordcloud</i> Pulau Morotai .....	46
Gambar 4. 18 Hasil <i>WordCloud</i> Pantai Tanjung Kelayang.....	47
Gambar 4. 19 Hasil <i>WordCloud</i> Danau Toba .....	48
Gambar 4. 20 Hasil <i>WordCloud</i> Wakatobi .....	48
Gambar 4. 21 Hasil <i>WordCloud</i> Pantai Tanjung Lesung .....	49

## DAFTAR TABEL

	Halaman
Tabel 2. 1. Referensi Penelitian Terdahulu .....	5
Tabel 2. 2. Confusion Matrix .....	12
Tabel 3. 1. Contoh Tahap Pelabelan .....	20
Tabel 3. 2. Hasil <i>Case Folding</i> .....	22
Tabel 3. 3. Hasil <i>Tokenizing</i> .....	22
Tabel 3. 4. Hasil <i>Stop-Word Removal</i> .....	23
Tabel 3. 5. Hasil <i>Stemming</i> .....	23
Tabel 3. 6. Hasil <i>Text Preprocessing</i> dan Jumlah Kata per Ulasan .....	25
Tabel 3. 7. Frekuensi Probabilitas Kata .....	25
Tabel 3. 8. Hasil Perhitungan Rumus <i>Laplace Correction</i> .....	26
Tabel 4. 1 Hasil <i>Web Scraping</i> dari tiap destinasi .....	30
Tabel 4. 2 Jumlah data kosong pada tiap destinasi .....	30
Tabel 4. 3 Jumlah ulasan kurang dari 3 kata pada tiap destinasi .....	31
Tabel 4. 4 Jumlah ulasan tidak relevan pada tiap destinasi .....	31
Tabel 4. 5 Jumlah <i>Review</i> Setiap Destinasi .....	31
Tabel 4. 6. Tabel pembagian data <i>training</i> dan data <i>testing</i> .....	32
Tabel 4. 7 Frekuensi Probabilitas Kata <i>Testing</i> .....	38
Tabel 4. 8 Hasil Testing dengan <i>Laplace Correction</i> .....	39
Tabel 4. 9 Hasil <i>Classification Report</i> kelima destinasi .....	44
Tabel 4. 10. Hasil <i>10-fold-cross-validation</i> .....	44
Tabel 4. 11. Hasil Akhir Naive Bayes .....	45

## DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Naïve Bayes Python.....	57
Lampiran 2. Kartu Bimbingan .....	59
Lampiran 3. Plagiasi.....	60
Lampiran 4. Biodata.....	61



UNIVERSITAS  
**Dinamika**

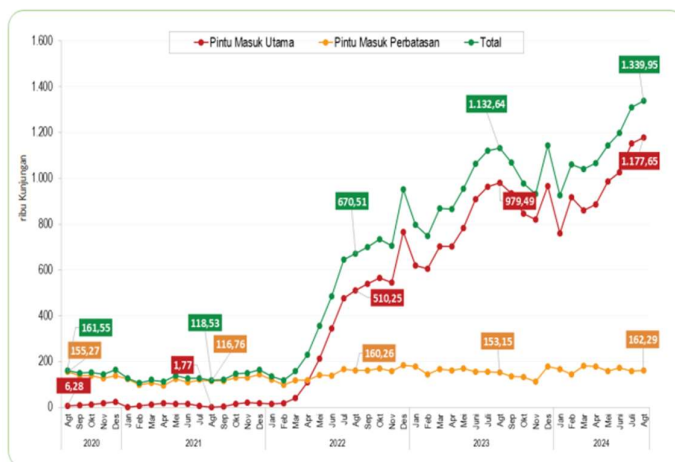
## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang

Pertumbuhan ekonomi merupakan proses peningkatan output per kapita di suatu negara dalam jangka panjang. Dalam konteks pariwisata, pertumbuhan ekonomi saling mempengaruhi dengan sektor pariwisata. Seiring perkembangan, pertumbuhan ekonomi suatu negara atau daerah diperoleh dari berbagai sektor ekonomi, dan salah satu sektor yang menjadi unggulan dalam dua dekade terakhir adalah sektor pariwisata (Mukaffi & Haryanto, 2022).

Badan Pusat Statistik (BPS) mengeluarkan Berita Resmi Statistik terkait Perkembangan Pariwisata di Indonesia per Agustus 2024. Pada Agustus 2024, jumlah kunjungan wisatawan mancanegara ke Indonesia mencapai 1,34 juta. Angka ini mengalami peningkatan sebesar 2,23 persen dibandingkan dengan Juli 2024 (*month-to-month*) dan naik 18,30 persen dibandingkan dengan periode yang sama tahun sebelumnya (*year-on-year*). Wisatawan mancanegara yang berkunjung ke Indonesia pada Agustus 2024 didominasi oleh wisatawan asal Malaysia (14,09 persen), Australia (11,47 persen), dan Tiongkok (9,41 persen) (Badan Pusat Statistik, 2024). Data tersebut dapat dilihat pada Gambar 1.1.



Gambar 1. 1. Jumlah Kunjungan Wisatawan Mancanegara Agustus 2020 - 2024



Kementerian Pariwisata dan Ekonomi Kreatif (Kemenparekraf) melaporkan bahwa devisa dari sektor pariwisata dan ekonomi kreatif telah mencapai US\$ 7,46 miliar atau setara dengan Rp 113,69 triliun (dengan kurs Rp 15.240) hingga Juli 2024. Jumlah ini telah memenuhi target devisa yang ditetapkan untuk tahun ini, di mana Kemenparekraf menetapkan target antara US\$ 7,38 miliar hingga US\$ 13,08 miliar. Perolehan devisa tersebut meningkat dibandingkan semester I 2023, yang hanya mencapai US\$ 6,09 miliar (Syaharani, 2024).

Menteri Pariwisata dan Ekonomi Kreatif, Sandiaga Uno, menyampaikan hasil survei Kementerian Pariwisata dan Ekonomi Kreatif (Kemenparekraf) yang melibatkan 761 responden terkait preferensi destinasi wisata. Berdasarkan survei tersebut, 64% responden memilih pantai, danau, dan laut sebagai destinasi utama. Pusat kuliner menjadi pilihan bagi 54%, pegunungan dipilih oleh 51,3%, taman rekreasi oleh lebih dari 36%, dan desa wisata hampir 30% dengan durasi wisata berkisar antara dua hingga empat hari. Pantai selalu menjadi daya tarik utama dalam wisata bahari, sebuah pandangan yang juga didukung oleh Nia Niscaya, Deputi Bidang Kebijakan Strategis (Simangunsong & Widyanti, 2023).

Meskipun sektor wisata pantai, danau, dan laut memiliki potensi besar, masih terdapat berbagai ulasan dari pengunjung yang menunjukkan ketidakpuasan terhadap kualitas layanan, aksesibilitas, dan pengelolaan lingkungan. Hal ini menegaskan perlunya perbaikan dan peningkatan layanan di sektor wisata alam di Indonesia. Ulasan pelanggan secara online memiliki pengaruh signifikan terhadap minat kunjungan wisatawan, terutama di era digital saat ini, di mana calon wisatawan sering mencari ulasan tentang destinasi sebelum memutuskan perjalanan mereka (Makarim & Pradana, 2022).

Misi Kemenparekraf yaitu menciptakan destinasi pariwisata berstandar internasional, menjalankan strategi pemasaran sesuai kebutuhan wisatawan, serta berkomitmen meningkatkan kelestarian lingkungan dan memperkuat kapasitas industri pariwisata Indonesia agar memiliki daya saing tinggi (Kemenparekraf/Baparekraf, 2023). Untuk mendukung pencapaian misi dan pengembangan wisata pantai, Kemenparekraf telah menobatkan sepuluh destinasi prioritas yaitu Mandalika, Pulau Morotai, Tanjung Kelayang, Danau Toba,

Wakatobi, Borobudur, Kepulauan Seribu, Tanjung Lesung, Bromo, dan Labuan Bajo (Kemenparekraf/Baparekraf RI, 2020).

Oleh karena itu, penelitian ini berfokus pada analisis sentimen ulasan wisatawan terhadap destinasi prioritas tersebut, dengan memanfaatkan ulasan yang diambil dari Google Maps. Penelitian ini bertujuan untuk mengidentifikasi sentimen positif dan negatif terhadap destinasi, memberikan rekomendasi strategis untuk meningkatkan kualitas layanan, dan mendukung pengembangan sektor wisata alam yang berdaya saing dan berkelanjutan. Analisis sentimen merupakan studi komputasi yang fokus pada opini, sentimen, serta emosi yang diungkapkan dalam teks. Melalui analisis ini, teks dalam kalimat atau dokumen akan dikelompokkan berdasarkan polaritasnya untuk mengetahui apakah pendapat yang disampaikan dalam teks tersebut bernada positif, negatif, atau netral (Amrullah, Anas, & Hidayat, 2020).

Dalam penelitian ini digunakan metode Naïve Bayes untuk mengklasifikasikan data teks karena strukturnya yang sederhana dan mudah diterapkan. Algoritma ini memiliki rumus yang sederhana dan biasanya memberikan akurasi yang lebih tinggi dibandingkan metode lainnya. Salah satu fitur utamanya adalah asumsi kuat tentang independensi antara setiap kondisi atau peristiwa dalam data (Nugroho, et al., 2024). Di dalam metode Naïve Bayes digunakan *laplace correction* untuk mengatasi masalah *zero probability*, yaitu ketika suatu kategori atau fitur tidak muncul dalam data pelatihan. *Laplace correction* menambahkan nilai kecil (biasanya 1) pada frekuensi kemunculan kategori, sehingga perhitungan probabilitas tetap valid terhadap data baru.

Ulasan yang didapat akan diklasifikasikan menjadi dua kategori, yaitu positif dan negatif, menggunakan algoritma *Naïve Bayes Classifier*. Dengan hasil ini, pengelola wisata lokal ataupun Kementerian Pariwisata dan Ekonomi Kreatif dapat terbantu dalam mendapatkan feedback dan melakukan evaluasi serta perbaikan pada kualitas layanan, sehingga pengunjung dapat menikmati pengalaman wisata yang lebih baik.

## 1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka rumusan masalah dari Tugas Akhir (TA) ini yaitu bagaimana melakukan analisis sentimen *review* pengunjung pada destinasi prioritas menggunakan algoritma naïve bayes *classifier* untuk memberikan bahan evaluasi kepada pengelola wisata dalam meningkatkan layanan.

## 1.3 Batasan Masalah

Adapun batasan masalah dalam melakukan penelitian ini sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah destinasi prioritas yang dinobatkan oleh Kementerian Pariwisata dan Ekonomi Kreatif di bidang perairan yang lebih spesifik.
2. Dari 10 Destinasi Prioritas diambil 5 destinasi yaitu Pulau Morotai, Tanjung Kelayang, Danau Toba, Wakatobi, dan Tanjung Lesung karena objek wisata yang diteliti adalah destinasi prioritas yang mencakup pantai, laut, dan danau.
3. Hasil ulasan kelima objek wisata kelautan diperoleh dari situs Google Maps.
4. Analisis sentimen mencakup pengelompokan menjadi dua kategori, yaitu positif dan negatif.
5. Tidak dibahas analisis sentimen kelompok netral.
6. Naïve Bayes *Classifier* menggunakan *laplace correction* untuk menghilangkan *zero probability* dan meningkatkan performa model

## 1.4 Tujuan

Berdasarkan uraian rumusan masalah, maka tujuan yang dicapai adalah menghasilkan sebuah klasifikasi analisis sentimen pada destinasi prioritas di Indonesia untuk mengetahui pandangan pengunjung terhadap beberapa destinasi tersebut, serta mengidentifikasi aspek-aspek yang perlu diperbaiki oleh pengelola dalam rangka meningkatkan kualitas layanan dan kenyamanan wisata.

## 1.5 Manfaat

Dari penelitian ini, diharapkan dapat memanfaatkan analisis sentimen sebagai alat pengukuran untuk destinasi prioritas di Indonesia guna memperbaiki persepsi dan penilaian terhadap pariwisata Indonesia secara keseluruhan.

## BAB II

### LANDASAN TEORI

#### 2.1 Penelitian Sebelumnya

Tabel 2. 1. Referensi Penelitian Terdahulu

Judul	Penulis	Hasil	Persentase Akurasi
Penerapan Algoritma Sentimen Analysis dan Naïve Bayes terhadap opini pengunjung di tempat wisata pantai Pintu Kota, Kota Ambon	Julius Chrisostomus Aponno (2022)	Tools : RapidMiner Testing : <i>Confusion Matrix</i> Sumber Data : Google Maps Subjek : Pantai Pintu Kota	Naïve Bayes : 90,65% Positive : 97,22% Negative : 85,42%
Analisis Sentimen Ulasan Wisatawan Pantai Gandoriah Pariaman Sumatra Barat Menggunakan Klasifikasi Naïve Bayes	Amannia Zeze Barki, Ikhsan Gunawan P, S.ST (2024)	Tools : - Testing : - Sumber Data : <i>Google Maps</i> Subjek : Pantai Gondoriah	Akurasi : 82% Presicion : 63% Recall : 42% F1-score : 43%
Analisis Sentimen Objek Wisata Di Provinsi Sulawesi Selatan Berdasarkan Ulasan Pengunjung Menggunakan Metode Random Forest Classifier	Aisyah Nurul Izza, Dian Eka Ratnawati, Widhy Hayuhardhika Nugraha Putra (2022)	Tools : - Testing : <i>Confusion Matrix</i> Sumber Data : TripAdvisor Objek : Tempat Wisata di Sulawesi Selatan seperti Pantai Tanjung Bira, Wisata Alam Malino, dan Taman Nasional Bantimurung	Akurasi : 82% Precision : 86% Recall : 86% F1-score : 89%

Berikut adalah perbedaan antara penelitian yang dilakukan oleh peneliti dengan penelitian sebelumnya :

1. Data yang digunakan dalam penelitian ini berasal dari ulasan di Google Maps.
2. Data yang digunakan pada penelitian terdahulu hanya 1 jenis pantai di daerah tertentu. Pada penelitian ini, data yang digunakan adalah 5 destinasi prioritas yang dinobatkan oleh Kemenparekraf yaitu Pulau Morotai, Tanjung Kelayang, Danau Toba, Wakatobi, dan Tanjung Lesung.
3. Untuk mengukur tingkat keakuratan data yang diteliti, peneliti menggunakan teknik validasi *10-Fold Cross-Validation*, dan akurasi algoritma akan dievaluasi menggunakan *confusion matrix* serta *classification report*.

## 2.2 Sektor Wisata Alam Pantai

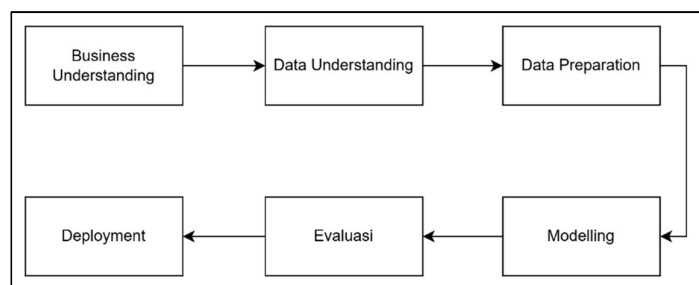
Indonesia adalah negara kepulauan terbesar di dunia yang terletak di garis khatulistiwa. Wilayah perairannya mencakup lebih dari 75% dari keseluruhan luas negara, yaitu sekitar 5,8 juta kilometer persegi, dengan total 17.500 pulau dan garis pantai sepanjang 81.000 km. Posisinya yang berada di zona khatulistiwa membuat Indonesia memiliki kekayaan sumber daya alam yang melimpah, termasuk pesona dan potensi pariwisata yang besar. Potensi wisata ini bisa dikembangkan untuk menjadi salah satu sumber devisa bagi negara (Buditiawan & Harmono, 2020).

Salah satu jenis destinasi wisata yang diminati adalah wisata alam, yaitu jenis wisata yang memanfaatkan keindahan lingkungan sekitar sebagai daya tarik utamanya. Di tempat-tempat ini, kita bisa menikmati panorama alam yang menakjubkan, dengan kegiatan seperti mendaki gunung atau berkemah di alam terbuka (Lasari, Kristiawati, & Afif, 2023).

Sektor pariwisata merupakan salah satu industri yang terus berkembang. Wisata pantai adalah salah satu jenis wisata yang sering dikembangkan untuk mendorong perekonomian daerah. Dengan kemajuan teknologi dan transportasi menjadi semakin mudah, sehingga membuka peluang besar bagi pengembangan sektor wisata ini (Damayanti, Widiastuti, Riandhana, & Ashad, 2022).

## 2.3 CRISP-DM

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) adalah kerangka kerja yang digunakan dalam proses pemodelan data. Kerangka ini dirancang untuk meningkatkan efisiensi biaya, keandalan, dan kecepatan dalam menyelesaikan proyek data mining skala besar (Nurhazizah & Puspitasari, 2023). Tahapan CRISP-DM meliputi 6 langkah seperti pada gambar 2.1



Gambar 2. 1. *Framework* CRISP-DM

**a. *Business Understanding***

Tahap ini dimulai dengan memahami kebutuhan dan tujuan bisnis. Langkah ini melibatkan transformasi pemahaman menjadi definisi masalah yang relevan dalam konteks data *mining*, diikuti perencanaan penyusunan strategi untuk mencapai tujuan yang ditetapkan dalam proses data *mining* (Hardiani & Putri, 2024).

**b. *Data Understanding***

Tahap ini dimulai dengan mengumpulkan data, kemudian dilanjutkan dengan proses analisis eksplorasi untuk memahami lebih dalam karakteristik data. Langkah ini mencakup pencarian wawasan awal, identifikasi kualitas data, serta penemuan bagian-bagian data yang menarik yang dapat digunakan untuk merumuskan hipotesis dan mengungkap informasi tersembunyi (Murti, 2024).

**c. *Data Preparation***

Pada tahap ini, memastikan data siap untuk digunakan dalam proses pemodelan menjadi hal yang sangat penting. Masalah terkait kualitas data dapat diatasi pada fase ini. Kualitas data yang diolah sangat bergantung pada model yang akan digunakan, sehingga data dapat diproses dan disesuaikan berdasarkan kebutuhan spesifik dari model tersebut (Kusuma & Wicaksono, 2023).

**d. *Modelling***

*Modelling* adalah salah satu tahap dalam CRISP-DM yang melibatkan pemilihan teknik data *mining* serta penentuan algoritma yang paling sesuai untuk diterapkan. Tahap ini juga mencakup pengujian dan penyesuaian model untuk memastikan performa yang optimal dalam menjawab permasalahan yang telah didefinisikan (Wahyudi, Sa'adah, & Puspitasari, 2023).

**e. *Evaluation***

Tahap evaluasi digunakan untuk mengukur dan menilai kinerja model yang telah dihasilkan pada tahap sebelumnya. Dalam proses data *mining*, tahap ini bertujuan untuk menentukan model yang paling optimal untuk digunakan (Permatasari, Dermawan, Iqbal, & Kurniawan, 2024).

#### **f. *Deployment***

Tahap *deployment*, atau perencanaan penerapan model, merupakan bagian paling bernilai dalam proses CRISP-DM. Perencanaan untuk tahap ini sebenarnya dimulai sejak fase *Business Understanding* dan mencakup tidak hanya cara memanfaatkan model untuk menghasilkan nilai, tetapi juga bagaimana mengonversi skor menjadi keputusan yang efektif serta mengintegrasikan keputusan tersebut ke dalam sistem operasional (Pratama, Nurcahyo, & Firgia, 2023).

### **2.4 Analisis Sentimen**

Menurut Amrullah, Anas, & Hidayat (2020), analisis sentimen adalah proses komputasional yang mempelajari opini, sentimen, dan emosi yang diekspresikan dalam teks. Tujuan dari analisis ini adalah untuk mengkategorikan polaritas teks dalam suatu kalimat atau dokumen, sehingga dapat menentukan apakah pendapat yang disampaikan bersifat positif, negatif, atau netral.

Menurut Ardras & Voutama (2023), analisis sentimen adalah penelitian yang menggunakan opini mengenai suatu topik untuk menilai apakah opini tersebut bernilai positif, negatif, atau netral. Opini-opini ini diperoleh melalui pengumpulan data dari dokumen, situs web, atau media sosial.

Jadi tujuan analisis sentimen adalah untuk memahami pandangan, perasaan, atau sikap yang terkandung dalam teks, apakah bersifat positif, negatif, atau netral. Proses ini melibatkan analisis dan ekstraksi sentimen atau opini dari berbagai sumber data, seperti ulasan produk, komentar pengguna, dan media sosial (Sari & Suryono, 2024).

### **2.5 Web Scraping**

*Web scraping* adalah proses otomatis untuk mengumpulkan data terstruktur dari web menggunakan aplikasi atau kode pemrograman khusus. Data yang dikumpulkan bisa berasal dari ribuan hingga miliaran sumber di internet. Proses otomatis ini dapat disesuaikan untuk memperoleh informasi yang relevan dan meminimalkan data yang tidak diperlukan (Rudini, Purnama, & Khan, 2023).



## 2.6 SerpAPI

SerpApi adalah layanan yang memfasilitasi pengambilan data dari hasil pencarian mesin telusur secara terstruktur dan otomatis. Layanan ini memungkinkan akses data berbagai sumber dan menyajikannya dalam format yang mudah digunakan. Tersedia pula opsi Web Scraping SerpApi untuk mendukung proses pengambilan data secara efisien (Ma'ady, Rizaldy, Satria, & Anaking, 2023).

## 2.7 Data Cleaning

Data *Cleaning* merupakan tahap dalam analisis sentimen yang bertujuan untuk mengurangi gangguan atau elemen yang tidak relevan dalam dataset, seperti tautan (URL), simbol "@" yang digunakan untuk *mention*, serta karakter lain yang tidak memberikan kontribusi terhadap analisis sentimen. Proses ini juga mencakup penghapusan baris-baris data yang bersifat duplikat (Ryandi, Pratiwi, & Sari, 2025).

Data *cleaning* dilakukan untuk menghasilkan dataset yang bersih dan siap digunakan pada tahap data *mining*. Proses ini bertujuan untuk memastikan bahwa data yang tersedia berisi nilai-nilai yang relevan, bebas dari nilai yang hilang (*missing value*), serta tidak mengandung data yang berulang (Supoyo & Prasetyaningrum, 2022).

## 2.8 Text Mining

Algoritma *Text Mining* adalah algoritma yang digunakan untuk menggali data guna memenuhi kebutuhan informasi dengan memanfaatkan metode seperti *machine learning*, *data mining*, pemrosesan bahasa alami, manajemen pengetahuan, serta pencarian informasi. Metode *text mining* melibatkan proses awal pada dokumen seperti pengkategorian teks, ekstraksi informasi, dan ekstraksi kata. Teknik ini berguna untuk memperoleh informasi dari sumber data melalui identifikasi dan eksplorasi pola yang menarik (Ridwansyah, 2022).

Dokumen teks harus melalui tahap prapemrosesan sebelum digunakan dalam berbagai operasi teks. Tahap ini bertujuan untuk menstrukturkan teks dan mengekstraksi fitur-fitur tertentu. Secara umum, ada empat langkah dalam prapemrosesan, yaitu tokenisasi, penghapusan *stop-word*, konversi ke huruf kecil (*case folding*), dan *stemming*. Hasil dari prapemrosesan ini berpengaruh terhadap



tingkat akurasi klasifikasi dan waktu komputasi yang dibutuhkan (Rosid, Fitriani, Astutik, Mulloh, & Gozali, 2020).

**a. Case Folding**

*Case folding* adalah langkah standarisasi yang mengubah semua elemen dalam teks menjadi huruf kecil. Proses ini juga melibatkan penghapusan karakter non-huruf yang tidak relevan dalam pengolahan teks, seperti angka dan tanda baca (Rofiqi & Akbar, 2024).

**b. Tokenizing**

Proses *tokenizing* adalah langkah untuk memecah kalimat menjadi kata-kata. Proses ini dimulai dengan mengubah semua kata dalam kalimat menjadi huruf kecil, mengganti karakter selain huruf dan angka dengan spasi, dan kemudian memisahkan kalimat menjadi kata-kata individual (Hanif, Farid, & Hasanah, 2023).

**c. Stop-Word Removal**

*Stopword Removal* (penghapusan kata bantu) adalah proses menghilangkan kata-kata yang umumnya tidak memberikan nilai informasi signifikan dalam analisis teks, seperti "dan", "atau", "yang", dan sejenisnya. Langkah ini biasanya dilakukan setelah proses tokenisasi, di mana teks telah dipecah menjadi kata-kata terpisah. Selanjutnya, setiap kata yang dihasilkan akan diperiksa, dan kata-kata yang termasuk dalam daftar stop words akan dihapus (Toresa, Sitorus, Muzdalifah, Wiza, & Syelly, 2024).

**d. Stemming**

*Stemming* adalah proses mengenali dan menguraikan kata-kata ke bentuk dasarnya. Dalam langkah *stemming*, setelah data komentar melewati proses *stop-word removal*, kata-kata yang telah diubah ke bentuk dasar akan disesuaikan menjadi bentuk kata baku yang benar dalam bahasa Indonesia (Firmansyah & Lestariningsih, 2024).

## 2.9 Naïve Bayes

Algoritma Naïve Bayes adalah salah satu metode yang digunakan dalam klasifikasi untuk memperkirakan probabilitas kejadian di masa depan berdasarkan data di masa lalu. Algoritma ini dikembangkan dari teori yang diperkenalkan Thomas Bayes. Naïve Bayes menggunakan fungsi statistik sederhana berdasarkan teorema Bayes dengan asumsi bahwa kehadiran suatu fitur dalam suatu kelas dianggap independen atau tidak bergantung pada fitur lainnya (Aponno , 2022).

Berdasarkan penelitian yang dilakukan oleh Mahendra, Sunarya, dan Wirawan (2025), tiga metode Naïve Bayes yaitu Multinomial, Bernoulli, dan Gaussian dibandingkan untuk mengklasifikasikan kategori keluhan pada aplikasi PRO Denpasar. Dari delapan skenario eksperimen, Multinomial Naïve Bayes menunjukkan performa terbaik, unggul dalam tujuh eksperimen. Sebaliknya, Gaussian Naïve Bayes memiliki performa terendah. Oleh karena itu, dalam penelitian ini akan digunakan metode Multinomial Naïve Bayes.

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Di mana :

X : Data dengan kelas belum diketahui

H : Hipotesis data merupakan suatu kelas spesifik

$P(H|X)$  : Probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas)

$P(H)$  : Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$  : Probabilitas hipotesis X berdasarkan kondisi H

$P(X)$  : Probabilitas X

Dalam penelitian ini, *Laplace Correction* digunakan sebagai metode untuk (2) mencegah munculnya nilai nol apabila data uji tidak terdapat dalam data latih. Pada setiap perhitungan, metode ini hanya menambahkan nilai satu sebagai penyesuaian. Hasilnya, *Laplace Correction* menunjukkan kinerja yang lebih unggul dibandingkan metode-metode lain (Lemantara, 2024). Berikut adalah rumus *Laplace Correction* :

$$P(Mk|Cl) = \frac{X_{kata} + 1}{X + X_{total}} \quad (3)$$

Di mana :

$P(Mk)$  : Probabilitas Munculnya Kata

$Cl$  : Kategori Kelas

$P(Mk|Cl)$  : Probabilitas munculnya kata ( $Mk$ ) pada kategori kelas tentu ( $Cl$ )

$X_{kata}$  : Nilai munculnya kata pada kategori  $Cl$

$X$  : Jumlah seluruh kata pada kategori  $Cl$

$X_{total}$  : Jumlah seluruh kata unik

## 2.10 Confusion Matrix

Menurut A'ayunnisa et al (2022) *confusion matrix* visualisasi evaluasi kinerja model klasifikasi yang menunjukkan perbandingan antara kelas sebenarnya dan kelas yang diprediksi. Matriks ini berbentuk 2x2 dan digunakan untuk menghitung tingkat ketepatan algoritma yang diterapkan. Berikut adalah tabel *confusion matrix*

Tabel 2. 2. *Confusion Matrix*

Class Predicted	Actual Class	
	Class = Yes	Class = No
	TP ( <i>True Positive</i> )	FP ( <i>False Positive</i> )
Class = Yes	FN ( <i>False Negative</i> )	TN ( <i>True Negative</i> )
Class = No		

Dimana :

*True Positive* (TP) : Jika data yang diprediksi bernilai positif dan cocok dengan nilai aktual

*False Negative* (FN) : Jika data yang diprediksi bernilai negatif dan tidak cocok dengan nilai aktual

*False Positive* (FP) : Jika data yang diprediksi tidak sesuai dengan nilai aktual

*True Negative* (TN) : Jika data yang diprediksi bernilai negatif dan cocok dengan nilai aktual

Akurasi adalah tingkat kecocokan antara hasil prediksi dengan nilai sebenarnya. Berikut adalah rumus akurasi *Confusion Matrix* :

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

Menurut Hidayah & Dodiman (2024), untuk menilai performa metode yang digunakan, digunakan AUC (*Area Under Curve*). AUC memiliki nilai diagnostik yang mencerminkan tingkat akurasi model yaitu :

1. Akurasi 0,90 – 1,00 = klasifikasi sangat baik
2. Akurasi 0,80 – 0,90 = klasifikasi baik
3. Akurasi 0,70 – 0,80 = klasifikasi cukup baik
4. Akurasi 0,60 – 0,70 = klasifikasi buruk
5. Akurasi 0,50 – 0,60 = klasifikasi sangat buruk

### 2.11 *K-Fold Cross-Validation*

*K-Fold Cross Validation* adalah metode evaluasi dalam *machine learning* yang membagi data menjadi *k* bagian, kemudian melatih dan menguji model secara bergantian sebanyak *k* kali. Teknik ini membantu memastikan hasil evaluasi yang lebih stabil, mengurangi bias, dan menghindari overfitting. Beberapa penelitian sebelumnya menunjukkan bahwa penggunaan *K-Fold Cross Validation* mampu meningkatkan akurasi model klasifikasi seperti *Decision Tree*, Naïve Bayes, SVM, dan *Neural Network* (Wijiyanto, Pradana, Sopingi, & Atina, 2024).

### 2.12 *Augmentasi Data*

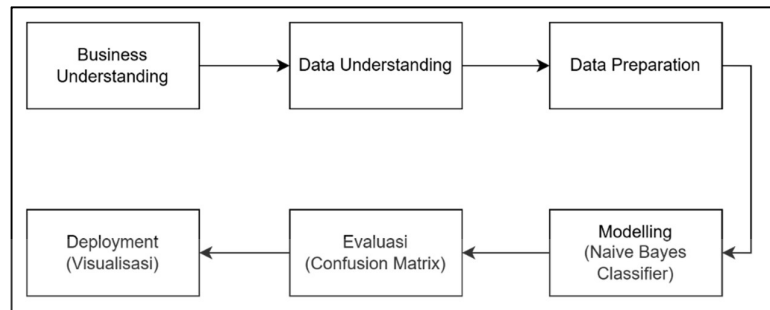
Augmentasi data merupakan suatu teknik yang digunakan untuk menambah dan memperkaya dataset dengan menciptakan variasi dari data yang sudah tersedia. Teknik ini tidak hanya bertujuan untuk memperbesar jumlah data pelatihan, tetapi juga untuk membantu model menjadi lebih *robust* dalam menghadapi berbagai kondisi lingkungan yang beragam (Amrulloh, Sari, & Padilah, 2024).

Dalam klasifikasi teks, tantangan utamanya terletak pada upaya menghasilkan teks baru tanpa mengubah label aslinya. Salah satu metode augmentasi teks yang dapat digunakan adalah penggantian sinonim dengan bantuan tesaurus (Rahma & Suadaa, 2023).

## BAB III

### METODOLOGI PENELITIAN

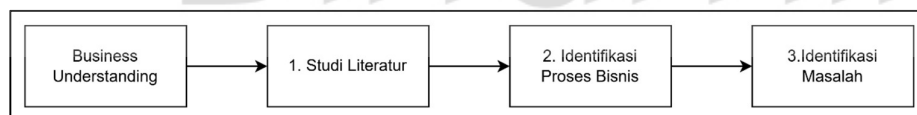
Bagian ini menjelaskan langkah-langkah yang ditempuh untuk menyelesaikan penelitian. Penelitian ini menggunakan metodologi CRISP-DM seperti pada gambar 3.1.



Gambar 3. 1. Metode CRISP – DM

#### 3.1 *Business Understanding*

Tahap *Business Understanding* meliputi :



Gambar 3. 2. Tahap *Business Understanding*

Gambar 3.2 menunjukkan tahapan *Business Understanding* yang dilakukan penulis dalam penelitian ini. Adapun rincian dari tahap awal adalah sebagai berikut:

##### 3.1.1 Studi Literatur

Studi literatur menjadi langkah pertama dalam metode penelitian ini. Pada tahap ini, penulis melakukan tinjauan terhadap penelitian-penelitian sebelumnya, termasuk Tugas Akhir (TA), jurnal, dan sumber lainnya. Setelah menyelesaikan tinjauan tersebut, penulis dapat menetapkan studi kasus serta langkah-langkah yang akan digunakan dalam pelaksanaan penelitian ini.

### 3.1.2 Identifikasi Proses Bisnis

Pengelola tempat wisata pantai di Indonesia perlu melakukan analisis sentimen untuk memahami persepsi dan pengalaman pengunjung yang terekam melalui ulasan di platform online. Dalam proses bisnis yang melibatkan promosi, pengelolaan fasilitas, dan tanggapan terhadap umpan balik, sering kali terjadi kendala dalam mengidentifikasi area yang memerlukan perbaikan. Saat ini, pengelola wisata masih menghadapi tantangan dalam mengelola ulasan atau review dari berbagai platform online secara efektif. Banyak ulasan yang tersebar di berbagai kanal seperti Google Reviews, sering kali tidak terstruktur, sehingga menyulitkan pengelola untuk mengidentifikasi isu utama yang perlu diperbaiki.

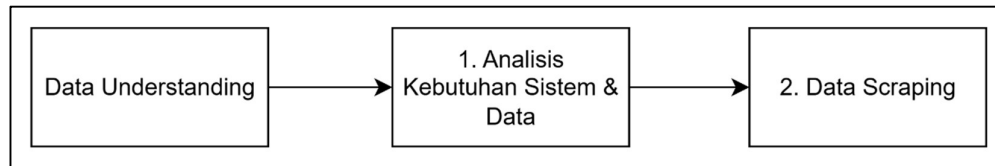
Analisis sentimen memungkinkan pengelola untuk mendapatkan wawasan mendalam mengenai keluhan atau kepuasan pengunjung, sehingga bisa mengoptimalkan layanan dan fasilitas yang ditawarkan. Selain itu, hasil analisis ini dapat membantu merancang strategi pemasaran yang lebih efektif, meningkatkan kualitas layanan, dan merespons keluhan dengan cepat, sehingga dapat meningkatkan kepuasan dan loyalitas wisatawan serta memperkuat reputasi destinasi wisata pantai.

### 3.1.3 Identifikasi Masalah

Berdasarkan analisis proses bisnis pengelola tempat wisata, terdapat beberapa masalah yang diidentifikasi yang memerlukan solusi melalui analisis sentimen. Pertama, pengelola sering kesulitan memahami secara cepat persepsi pengunjung mengenai kualitas fasilitas dan layanan yang disediakan karena ulasan yang tersebar di berbagai platform sulit dikompilasi dan diinterpretasi secara manual. Kedua, kurangnya pemantauan yang efektif terhadap opini negatif membuat pengelola lambat dalam merespons keluhan, yang dapat berdampak pada citra destinasi dan menurunkan minat pengunjung. Ketiga, strategi pemasaran yang kurang tepat karena tidak berdasarkan pada data nyata mengenai kebutuhan dan preferensi pengunjung menyebabkan kurang optimalnya promosi yang dilakukan. Oleh karena itu, diperlukan analisis sentimen untuk mengidentifikasi tren dan sentimen pengunjung, sehingga pengelola dapat meningkatkan kualitas layanan dan menyusun strategi pemasaran yang lebih akurat dan responsif.

### 3.2 *Data Understanding*

Tahapan ini menjelaskan tentang tahap pengumpulan data untuk menyelesaikan penelitian seperti pada gambar 3.3.



Gambar 3. 3 Tahap *Data Understanding*

Gambar 3.3 memperlihatkan tahapan *Data Understanding* yang telah dilakukan oleh penulis dalam penelitian ini. Rincian dari tahapan awal tersebut dijelaskan sebagai berikut:

#### 3.2.1 Analisis Kebutuhan Sistem dan Data

##### a. Kebutuhan Perangkat Lunak

Kebutuhan perangkat lunak mencakup berbagai komponen perangkat lunak yang akan digunakan dalam penelitian ini. Berikut adalah daftar perangkat lunak yang digunakan:

- i. *Tools* yang digunakan dalam melakukan *Scrapping Web* dari Google Maps dan klasifikasi Naïve Bayes adalah Python.
- ii. Di dalam proses *Web Scrapping*, digunakan layanan SerpApi untuk mendapatkan API yang bisa mengakses data dari Google Maps.

##### b. Kebutuhan Perangkat Keras

Kebutuhan perangkat keras merupakan komponen fisik yang membentuk sistem komputer terstruktur, di mana terdapat berbagai perangkat keras pendukung yang berperan penting dalam kelancaran jalannya sistem tersebut. Berikut adalah daftar perangkat keras yang digunakan :

- i. *CPU (Central Processing Unit)* : Intel Core I5-10200U
- ii. *RAM (Random Accesss Memory)* : 8 GB
- iii. *GPU (Graphic Processing Unit)* : AMD Radeon

### c. Kebutuhan Data

Kebutuhan data mencakup berbagai jenis data yang akan digunakan dalam penelitian ini. Berikut adalah daftar kebutuhan data yang diperlukan:

- i. Nama penulis
- ii. *Rating* yang diberikan
- iii. *Review*/ulasan yang ditulis
- iv. Tanggal penulisan *review*

### 3.2.2 Data Scraping

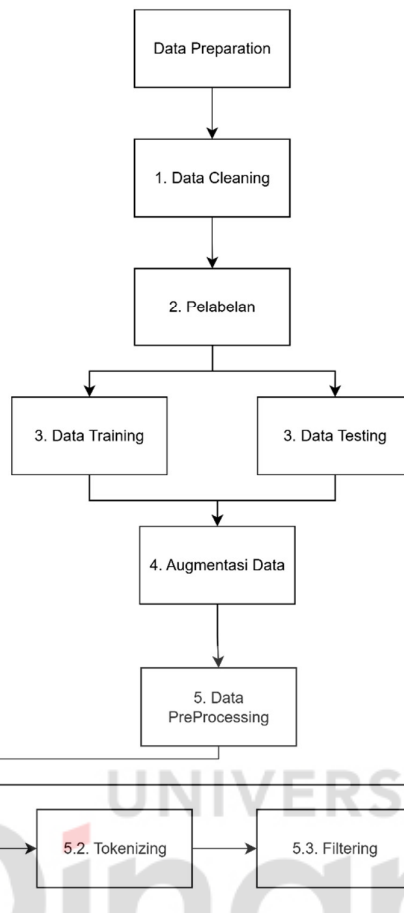
Proses pengambilan data dalam penelitian ini dilakukan melalui teknik data *scraping*. Penambahan data ini menggunakan *software* Visual Studio Code dengan bahasa pemrograman Python untuk mengekstrak data dari situs Google Maps. Langkah-langkah pengambilan data menggunakan Python dijelaskan sebagai berikut:

1. Mencari website yang sudah ditentukan dengan menggunakan URL *website* tersebut yaitu maps.google.com
2. Lalu di dalam Visual Studio Code menginstall *library* *serpAPI*, *google-search-results* dan *pandas*
3. Lalu dilanjutkan dengan code untuk mengambil data yang digunakan untuk mendukung penelitian ini yaitu data *rating*, *review*, nama penulis, dan tanggal.
4. Data yang sudah diambil akan dirubah dan dimasukkan ke dalam file csv.
5. Hasil penambahan berupa 5 sumber berisi *review* dari Pulau Morotai, Pantai Tanjung Kelayang, Danau Toba, Wakatobi, dan Pantai Tanjung Lesung.

### 3.3 Data Preparation

Tahapan ini menjelaskan tentang tahap *preprocessing* data seperti pada gambar 3.4.





Gambar 3. 4 Tahap *Data Preparation*

Gambar 3.4 menunjukkan tahapan *Data Preparation* yang dilakukan penulis dalam penelitian ini. Adapun rincian tahap awal tersebut adalah sebagai berikut:

### 3.3.1 *Data Cleaning*

Tahapan awal dalam penelitian ini dimulai dengan proses data *cleaning*. Tahapan ini memiliki peran krusial dalam memastikan bahwa data yang akan dianalisis dalam kondisi bersih, relevan, dan benar-benar merepresentasikan opini masyarakat. Jika proses pembersihan ini diabaikan, maka hasil analisis sentimen dapat menjadi tidak akurat atau bias akibat adanya unsur-unsur yang tidak penting, seperti simbol, angka yang tidak bermakna, atau kata-kata umum yang tidak memiliki nilai informasi. Proses data *cleaning* ini terdiri dari tiga bagian utama, yaitu:

**A. Penghapusan data kosong (*null*)**

Langkah ini bertujuan untuk menghapus entri atau kolom yang memiliki nilai kosong (*null* atau NaN). Data kosong dapat menyebabkan gangguan dalam analisis karena dapat menghambat proses perhitungan, pengolahan model *machine learning*, atau bahkan menyebabkan error saat visualisasi. Selain itu, data yang kosong tidak memberikan informasi berarti. Contohnya adalah ulasan dengan kolom yang kosong atau hanya berisi spasi tidak dapat dianalisis, sehingga perlu dihapus.

**B. Eliminasi ulasan yang terdiri kurang dari 3 kata**

Ulasan yang sangat pendek misalnya hanya satu atau dua kata seperti "bagus", "oke", "tidak", atau hanya berisi emotikon seperti "👍", "😊" cenderung tidak memberikan konteks atau informasi yang cukup untuk dianalisis. Oleh karena itu, ulasan-ulasan tersebut dihapus untuk meningkatkan kualitas dan relevansi data yang digunakan dalam proses analisis. Proses ini dilakukan secara manual di dalam file excel hasil dari *web scraping*.

**C. Eliminasi ulasan yang tidak relevan**

Ulasan yang tidak membahas topik utama seperti sektor pariwisata—misalnya ulasan yang menyinggung produk lain, isu politik, atau hanya berisi spam—dianggap tidak mendukung tujuan analisis. Oleh karena itu, ulasan semacam ini disaring agar hanya data yang benar-benar relevan dan berkualitas yang digunakan dalam proses analisis lanjutan. Proses ini dilakukan secara manual di dalam file excel hasil dari *web scraping*.

### 3.3.2 Pelabelan

Proses pelabelan atau klasifikasi atribut ditentukan berdasarkan penilaian subjektif dari peneliti. Data ulasan dikumpulkan dari Google Maps, lalu secara manual dikelompokkan menjadi dua kategori: sentimen positif untuk ulasan yang bernada positif dan sentimen negatif untuk ulasan yang bernada negatif. Kategori tersebut ditentukan secara manual tanpa menggunakan aplikasi otomatis. Tahapan pelabelan ini dapat dilihat pada tabel di bawah :

Tabel 3. 1. Contoh Tahap Pelabelan

Kategori	Review
Negatif	Semua fasilitas mahal untuk ukuran wisatawan lokal
Positif	Tempat sangat indah sekali bersih. Parkir luas. Sangat tertata. Recommed pokonya
Positif	View pantainya bagus banget 😊👍 ...
Negatif	Pantainya kotor . Banyak lumpur . Fasilitas air bersih kurang
Positif	Cocok banget untuk camping bersama keluarga, di temani pemandangan pantai dan pohon pinus....rekomended...

### 3.3.3 Pembagian Data *Training* dan Data *Testing*

Bagian ini mencakup proses pembagian data pelatihan (*training*) dan pengujian (*testing*). Data yang sudah diberi label dipisah menjadi data *training* dan *testing* secara acak dengan rasio 70-30 di mana 70% data digunakan untuk melatih model dan 30% sisanya digunakan untuk menguji performa model.

### 3.3.4 Augmentasi Data

Tahap augmentasi data dilakukan untuk mengatasi ketidakseimbangan jumlah ulasan positif dan negatif, di mana ulasan negatif jumlahnya jauh lebih sedikit dibandingkan ulasan positif atau sebaliknya. Ketimpangan ini menyebabkan model cenderung bias terhadap kelas mayoritas, sehingga menurunkan akurasi klasifikasi **sentimen** secara keseluruhan. Untuk mengurangi dampak ketidakseimbangan tersebut, digunakan teknik augmentasi berbasis *synonym replacement*.

Teknik ini bekerja dengan cara mengganti beberapa kata dalam ulasan dengan sinonimnya yang memiliki makna serupa, sehingga menghasilkan kalimat baru tanpa mengubah sentimen asli dari ulasan tersebut. Dalam penelitian ini, proses pencarian sinonim dilakukan menggunakan *library* tesaurus Bahasa Indonesia dari repositori GitHub victoriasovereigne/tesaurus. *Library* ini menyediakan fungsi `getSinonim()` yang memungkinkan pencarian daftar sinonim dari sebuah kata dalam Bahasa Indonesia, berdasarkan data tesaurus yang terverifikasi.

Sebagai contoh, ulasan negatif seperti “Pelayanan di tempat ini **sangat lambat** dan tidak profesional.” dapat diubah menjadi “Pelayanan di tempat ini **amat lelet dan tidak profesional.**” Dalam contoh tersebut, kata "sangat" diganti dengan "amat", dan "lambat" diganti dengan "lelet", yang keduanya merupakan sinonim dalam konteks negatif. Meskipun terdapat perubahan kata, makna dan sentimen negatif dari ulasan tetap dipertahankan.

Proses augmentasi dilakukan dengan mengacak urutan kata terlebih dahulu, lalu mengecek satu per satu apakah kata tersebut memiliki sinonim menggunakan library tesaurus. Jika tersedia, salah satu sinonim dipilih secara acak untuk menggantikan kata tersebut, hingga batas maksimum jumlah kata yang boleh diganti tercapai. Kalimat hasil augmentasi hanya disimpan apabila berbeda dari kalimat asli, guna memastikan keragaman dalam data.

Dengan strategi ini, variasi linguistik dalam data latih meningkat dan distribusi antara kelas positif dan negatif menjadi lebih seimbang. Hal ini membantu model Naïve Bayes dalam mengenali sentimen secara lebih akurat, karena model tidak hanya belajar dari satu bentuk kalimat tetapi juga dari berbagai ekspresi yang berbeda dengan makna yang tetap konsisten.

### 3.3.5 Data Preprocessing

Tahap berikutnya adalah data *preprocessing*, yang dimulai dengan ekstraksi data. Tujuan dari ekstraksi data adalah untuk mengubah data hasil penambangan yang tidak terstruktur menjadi format yang lebih teratur berupa kata dasar. Pada penelitian ini, data *preprocessing* berfokus pada pengolahan data tidak terstruktur menjadi kata dasar dan mencakup empat langkah, yaitu *case folding*, *tokenizing*, *stop-word removal*, dan *stemming*. Tahapan data *preprocessing* yang dilakukan menggunakan Python adalah sebagai berikut:

1. Mengimpor *library* yang dibutuhkan seperti *pandas*, *nlTK*, dan *Sastrawi*.
2. Membaca data yang sudah di ambil pada tahap *preprocessing* berupa file excel atau CSV.
3. Menghapus teks yang tidak diperlukan seperti angka, tanda baca, emotikon ataupun teks lainnya yang tidak relevan.
4. *Case Folding* : tahap yang berfungsi untuk mengonversi semua huruf kapital menjadi huruf kecil.
5. *Tokenizing* : tahap yang berfungsi untuk menghapus karakter atau simbol selain huruf A hingga Z.
6. *Stop-Word Removal* : tahap yang berfungsi untuk menghilangkan kata-kata yang tidak relevan, seperti kata hubung, kata ganti, dan kata depan dengan menggunakan library *Sastrawi*.

7. *Stemming* : tahap yang berfungsi mengubah kata berimbuhan menjadi bentuk kata dasarnya. Langkah ini akan menggunakan *library* Sastrawi
8. Menyimpan data hasil *preprocess* ke dalam file, dapat berupa excel atau CSV.

Hasil dari proses *text preprocessing* adalah data ulasan yang telah diproses dengan baik dan diekspor ke dalam format Excel, yang mencakup 5 sumber, yaitu Pulau Morotai, Pantai Tanjung Kelayang, Danau Toba, Wakatobi, dan Pantai Tanjung Lesung. Contoh hasil dari *text preprocessing* akan dijelaskan berikut ini.

### 1. *Case Folding*

Contoh hasil dari proses *Case Folding* dapat dilihat pada tabel di bawah ini:

Tabel 3. 2. Hasil *Case Folding*

Sebelum	Sesudah
Semua fasilitas mahal untuk ukuran wisatawan lokal	semua fasilitas mahal untuk ukuran wisatawan lokal
Tempat sangat indah sekali bersih. Parkir luas. Sangat tertata. Recommed pokonya	tempat sangat indah sekali bersih. parkir luas. sangat tertata. recommed pokonya
View pantainya bagus banget 😊😍 ...	view pantainya bagus banget 😊😍 ...
Pantainya kotor . Banyak lumpur . Fasilitas air bersih kurang	pantainya kotor. banyak lumpur. fasilitas air bersih kurang.
Cocok banget untuk camping bersama keluarga, di temani pemandangan pantai dan pohon pinus....rekomended...	cocok banget untuk camping bersama keluarga, di temani pemandangan pantai dan pohon pinus....rekomended...

### 2. *Tokenizing*

Contoh hasil dari proses *Tokenizing* dapat dilihat pada tabel di bawah ini:

Tabel 3. 3. Hasil *Tokenizing*

Sebelum	Sesudah
semua fasilitas mahal untuk ukuran wisatawan lokal	[semua], [fasilitas], [mahal], [untuk], [ukuran], [wisatawan], [lokal]
tempat sangat indah sekali bersih. parkir luas. sangat tertata. recommed pokonya	[tempat], [sangat], [indah], [sekali], [bersih], [parker], [luas], [sangat], [tertata], [recommed], [pokonya]
view pantainya bagus banget 😊😍 ...	[view], [pantainya], [bagus], [banget], [😊], [😍]
pantainya kotor. banyak lumpur. fasilitas air bersih kurang.	[pantainya], [kotor], [banyak], [lumpur], [fasilitas], [air], [bersih], [kurang]
cocok banget untuk camping bersama keluarga, di temani pemandangan pantai dan pohon pinus....rekomended...	[cocok], [banget], [untuk], [camping], [bersama], [keluarga], [di], [temani], [pemandangan], [pantai], [dan], [pohon], [pinus], [rekomended]

### 3. *Stop-Word Removal*

Contoh hasil dari proses *Stop-Word Removal* dapat dilihat pada tabel di bawah ini:

Tabel 3. 4. Hasil *Stop-Word Removal*

Sebelum	Sesudah
[semua], [fasilitas], [mahal], [untuk], [ukuran], [wisatawan], [lokal]	[fasilitas], [mahal], [ukuran], [wisatawan], [lokal]
[tempat], [sangat], [indah], [sekali], [bersih], [parker], [luas], [sangat], [tertata], [recommmed], [pokonya]	[tempat], [indah], [bersih], [parkir], [luas], [tertata], [recommmed]
[view], [pantainya], [bagus], [banget], [😊], [😄]	[view], [pantainya], [bagus], [banget]
[pantainya], [kotor], [banyak], [lumpur], [fasilitas], [air], [bersih], [kurang]	[pantainya], [kotor], [lumpur], [fasilitas], [air], [bersih], [kurang]
[cocok], [banget], [untuk], [camping], [bersama], [keluarga], [di], [temani], [pemandangan], [pantai], [dan], [pohon], [pinus], [rekomended]	[cocok], [banget], [camping], [bersama], [keluarga], [temani], [pemandangan], [pantai], [pohon], [pinus], [rekomended]

### 4. *Stemming*

Contoh hasil dari proses *Stemming* dapat dilihat pada tabel di bawah ini:

Tabel 3. 5. Hasil *Stemming*

Sebelum	Sesudah
[fasilitas], [mahal], [ukuran], [wisatawan], [lokal]	fasilitas mahal ukur wisata lokal
[tempat], [indah], [bersih], [parkir], [luas], [tertata], [recommmed]	tempat indah bersih parkir luas tata rekomend
[view], [pantainya], [bagus], [banget]	view pantai bagus banget
[pantainya], [kotor], [lumpur], [fasilitas], [air], [bersih], [kurang]	pantai kotor lumpur fasilitas air, bersih kurang
[cocok], [banget], [camping], [bersama], [keluarga], [temani], [pemandangan], [pantai], [pohon], [pinus], [rekomended]	cocok banget camp sama keluarga teman pandang pantai pohon pinus rekomend

### 3.4 *Modelling*

Langkah berikutnya adalah melakukan klasifikasi data menggunakan Naïve Bayes *Classifier*, yang bertujuan untuk mengklasifikasikan data dengan menghitung frekuensi kemunculan setiap kata dalam kalimat, apakah termasuk kategori positif atau negatif. Pada penelitian ini, klasifikasi data berfungsi untuk menentukan kelas data menjadi positif atau negatif.

Langkah-langkah atau model yang akan diterapkan oleh Python untuk melakukan analisis sentimen adalah sebagai berikut:

1. Import *Training Data* dan *Testing Data*

Pada tahap ini, data *training* dan *testing* diimpor dari file CSV atau format lain yang telah melalui proses *preprocessing*. Pembagian data dilakukan secara manual di Excel, lalu data dimuat ke dalam Python menggunakan pustaka *pandas*.

2. Proses Naïve Bayes

Dalam proses Naïve Bayes, langkah yang dilakukan adalah memasukkan kode untuk membangun model Naïve Bayes sebagai berikut :

- a) Mengimpor *library* *pandas*, *scikit-learn*, *matplotlib*, *seaborn*, dan *openpyxl*.
- b) Membaca dan memuat data *training* dan data *testing* menggunakan pustaka *pandas*. Dilakukan juga validasi untuk memastikan kolom tersebut ada pada dataset.
- c) Label sentimen dikonversi menjadi format numerik menggunakan *LabelEncoder* dari *sklearn*.
- d) Data teks pada kolom 'Review' diubah menjadi representasi numerik dengan *CountVectorizer*.
- e) Model Multinomial Naïve Bayes dilatih menggunakan data *training*. Model digunakan untuk memprediksi data *testing*.

3. Evaluasi Naïve Bayes

Pada tahap evaluasi Naïve Bayes, hal yang dilakukan yaitu melakukan evaluasi yang memerlukan operator sebagai berikut :

- a) Evaluasi dilakukan menggunakan *Confusion Matrix*, *Classification Report*, dan *10-fold cross-validation*.
- b) Visualisasi *Confusion Matrix* menggunakan *seaborn*.

Berikut adalah gambaran proses yang dilakukan dalam metode Naïve Bayes:

Dalam tahap ini, data yang digunakan harus sudah melewati proses *preprocessing*. Berikut adalah hasil dari text preprocessing yang telah dilakukan:

Tabel 3. 6. Hasil *Text Preprocessing* dan Jumlah Kata per Ulasan

Kategori	Review	Jumlah Kata
Negatif	fasilitas mahal ukur wisata lokal	5
Positif	tempat indah bersih parkir luas tata rekomend	7
Positif	view pantai bagus banget	4
Negatif	pantai kotor lumpur fasilitas air bersih kurang	7
Positif	cocok banget camp sama keluarga teman pandang pantai pohon pinus rekomend	11

Setelah proses *preprocessing* selesai, langkah berikutnya adalah menghitung probabilitas untuk ulasan positif dan negatif. Dari total 5 ulasan yang digunakan, probabilitas ulasan positif adalah sebesar  $P(Positif) = \frac{3}{5} = 0,6$  dan probabilitas ulasan negatif adalah sebesar  $P(Negatif) = \frac{2}{5} = 0,4$ . Selanjutnya, setiap kata dalam review dibagi dan dianalisis. Berdasarkan data tersebut, jumlah kata yang terdapat dalam dataset adalah sebanyak 34 kata dengan kata dari ulasan positif sebanyak 22 dan ulasan negatif sebanyak 12 kata. Langkah berikutnya adalah menghitung probabilitas setiap kata menggunakan rumus Naïve Bayes berdasarkan frekuensi kemunculannya dalam data training.

Tabel 3. 7. Frekuensi Probabilitas Kata

No.	Kata	Training Negatif	Training Positif
1	indah	0	1
2	pandang	0	1
3	cocok	0	1
4	tata	0	1
5	kurang	1	0
6	lumpur	1	0
7	lokal	1	0
8	parkir	0	1
9	banget	0	2
10	bersih	1	1
11	rekomend	0	2
12	bagus	0	1
13	kotor	1	0
14	air	1	0
15	pinus	0	1
16	pohon	0	1
17	ukur	1	0
18	tempat	0	1



No.	Kata	Training Negatif	Training Positif
19	sama	0	1
20	wisata	1	0
21	luas	0	1
22	keluarga	0	1
23	teman	0	1
24	mahal	1	0
25	camp	0	1
26	view	0	1
27	fasilitas	2	0
28	pantai	1	2

Berdasarkan tabel 3.7, nilai probabilitas dapat dicari dengan menggunakan rumus *laplace correction* yaitu  $P(Mk|Cl)$  sebagai berikut :

$$P(\text{indah}|\text{negatif}) = (0+1) / (12+28) = 0,025$$

$$P(\text{indah}|\text{positif}) = (1+1) / (22+28) = 0,04$$

$$P(\text{bersih}|\text{negatif}) = (1+1) / (12+28) = 0,05$$

$$P(\text{bersih}|\text{positif}) = (1+1) / (22+28) = 0,04$$

$$P(\text{pantai}|\text{negatif}) = (1+1) / (12+28) = 0,05$$

$$P(\text{pantai}|\text{positif}) = (2+1) / (22+28) = 0,06$$

Berdasarkan hasil perhitungan dengan menggunakan rumus  $P(Mk|Cl)$ , maka didapati hasil sebagai berikut :

Tabel 3. 8. Hasil Perhitungan Rumus *Laplace Correction*

No.	Kata	$X_k \text{Negatif}$	$X_k \text{Positif}$	$P(Mk Cl)$ Negatif	$P(Mk Cl)$ Positif
1	indah	0	1	0,025	0,04
2	pandang	0	1	0,025	0,04
3	cocok	0	1	0,025	0,04
4	tata	0	1	0,025	0,04
5	kurang	1	0	0,05	0,02
6	lumpur	1	0	0,05	0,02
7	lokal	1	0	0,05	0,02
8	parkir	0	1	0,025	0,04
9	banget	0	2	0,025	0,06
10	bersih	1	1	0,05	0,04
11	rekomend	0	2	0,025	0,06
12	bagus	0	1	0,025	0,04
13	kotor	1	0	0,05	0,02
14	air	1	0	0,05	0,02
15	pinus	0	1	0,025	0,04
16	pohon	0	1	0,025	0,04
17	ukur	1	0	0,05	0,02
18	tempat	0	1	0,025	0,04

No.	Kata	$X_k Negatif$	$X_k Positif$	$P(M_k Cl)$ Negatif	$P(M_k Cl)$ Positif
19	sama	0	1	0,025	0,04
20	wisata	1	0	0,05	0,02
21	luas	0	1	0,025	0,04
22	keluarga	0	1	0,025	0,04
23	teman	0	1	0,025	0,04
24	mahal	1	0	0,05	0,02
25	camp	0	1	0,025	0,04
26	view	0	1	0,025	0,04
27	fasilitas	2	0	0,075	0,02
28	pantai	1	2	0,05	0,06

Dari hasil perhitungan probabilitas setiap kata yang telah dilakukan pada tabel 3.8, dapat diketahui probabilitas masing-masing kata. Data pelatihan sebelumnya dapat digunakan untuk menganalisis hasil uji model dengan mengalikan semua nilai probabilitas tersebut. Berikut adalah data pengujian yang akan digunakan : “Rekomended... pantainya bersih dan ada pohon pinusnya ❤️❤️”. Dan setelah dilakukan *text preprocessing*, hasilnya “rekomend pantai bersih pohon pinus”. Berikut adalah perhitungan probabilitas setiap kata :

Positif

$$\begin{aligned}
 &= P(\text{positif}) \times P(\text{rekomend}|\text{positif}) \times P(\text{pantai}|\text{positif}) \times P(\text{bersih}|\text{positif}) \times \\
 &P(\text{pohon}|\text{positif}) \times P(\text{pinus}|\text{positif}) \\
 &= 0,6 \times 0,06 \times 0,06 \times 0,04 \times 0,04 \times 0,04 \\
 &= 1,3824 \times 10^{-7}
 \end{aligned}$$

Negatif

$$\begin{aligned}
 &= P(\text{negatif}) \times P(\text{rekomend}|\text{negatif}) \times P(\text{pantai}|\text{negatif}) \times P(\text{bersih}|\text{negatif}) \times \\
 &P(\text{pohon}|\text{negatif}) \times P(\text{pinus}|\text{negatif}) \\
 &= 0,4 \times 0,025 \times 0,05 \times 0,05 \times 0,025 \times 0,025 \\
 &= 0,1562 \times 10^{-7}
 \end{aligned}$$

Perhitungan di atas merupakan langkah dalam klasifikasi sentimen berdasarkan nilai probabilitas, dimana hasil tertinggi sebesar  $1,3824 \times 10^{-7}$ . Dengan demikian, ulasan tersebut dikategorikan sebagai ulasan "positif".

### 3.5 Evaluasi (*Confusion Matrix*)

Evaluasi dilakukan untuk menilai seberapa akurat model dalam melakukan klasifikasi data. Tingkat akurasi model akan dihitung dengan memanfaatkan *confusion matrix*, yang memberikan gambaran tentang performa prediksi dibandingkan dengan nilai aktual. Untuk memastikan keandalan model, teknik validasi juga diterapkan pada data.

Penelitian ini menggunakan metode *10-fold cross-validation*, di mana data dibagi menjadi 10 bagian. Prosesnya dilakukan secara iteratif, di mana pada setiap iterasi, satu *fold* digunakan sebagai data uji, sementara 9 *fold* lainnya digunakan sebagai data training. Proses ini dilakukan 10 kali sehingga setiap *fold* berperan sebagai data uji sekali, dan hasilnya dirata-rata untuk evaluasi performa model yang lebih stabil dan mengurangi risiko bias.

### 3.6 Deployment (Visualisasi)

Tahap visualisasi data dilakukan dengan memanfaatkan *Wordcloud*, yaitu metode yang menampilkan kata-kata yang paling sering muncul dalam kumpulan teks. Ukuran huruf pada *Wordcloud* menunjukkan seberapa sering kata tersebut muncul, semakin besar ukuran font, semakin tinggi frekuensinya. Data yang digunakan dalam visualisasi ini adalah data yang telah diklasifikasikan ke dalam sentimen positif dan negatif, namun belum melalui proses pengolahan dengan algoritma Naïve Bayes. Oleh karena itu, *Wordcloud* yang dihasilkan dibedakan berdasarkan kategori sentimen tersebut untuk mengidentifikasi aspek-aspek yang menjadi daya tarik utama maupun yang perlu diperbaiki.

## BAB IV

### HASIL DAN PEMBAHASAN

Bab ini akan membahas hasil analisis dari pengolahan data ulasan yang diambil dari Google Maps terkait lima destinasi prioritas di Indonesia, yaitu Pulau Morotai, Tanjung Kelayang, Danau Toba, Wakatobi, dan Tanjung Lesung.

#### 4.1 Analisis

Analisis diawali dengan mengidentifikasi masalah yang menjadi dasar untuk memulai dan melaksanakan proses tersebut. Penulis menemukan bahwa terdapat beberapa masalah dalam proses bisnis pengelolaan tempat wisata yang memerlukan analisis sentimen sebagai solusi. Kesulitan dalam memahami persepsi pengunjung secara cepat, kurangnya pemantauan opini negatif yang memperlambat respons terhadap keluhan, serta strategi pemasaran yang kurang optimal karena tidak berbasis data pengunjung menjadi tantangan utama. Analisis sentimen diperlukan untuk mengidentifikasi tren dan sentimen pengunjung guna meningkatkan layanan dan menyusun strategi pemasaran yang lebih efektif. Oleh karena itu, penelitian ini ditujukan untuk pengelola tempat wisata mengidentifikasi tren dan sentimen pengunjung, sehingga mereka dapat meningkatkan kualitas layanan dan menyusun strategi pemasaran yang lebih efektif.

Analisis tugas akhir ini dikerjakan dengan menggunakan Python. Di dalamnya digunakan SerpAPI untuk *web scraping* dan Naïve Bayes sebagai model yang diterapkan pada data.

#### 4.2 Hasil Web Scraping

Setelah tahap analisis selesai, proses dilanjutkan dengan penambahan data atau *web scraping*. Penambahan data dilakukan untuk melakukan *scraping* data dari sumber yang digunakan dalam penelitian ini. Alat yang digunakan untuk melakukan penambahan data adalah Python dengan data yang diambil adalah nama, *rating*, *review*, dan tanggal. Hasil *web scraping* berupa jumlah ulasan dapat dilihat pada tabel 4.1.

Tabel 4. 1 Hasil *Web Scraping* dari tiap Destinasi

Nama Destinasi	Jumlah Review
Pulau Morotai	316
Tanjung Kelayang	1980
Danau Toba	2380
Wakatobi	1940
Tanjung Lesung	1890

### 4.3 Data Cleaning

Setelah proses *web scraping* selesai, langkah berikutnya adalah data *cleaning*. Tahap ini dilakukan untuk membersihkan data dari elemen-elemen yang tidak relevan seperti tanda baca, angka, simbol, dan kata-kata umum yang tidak memiliki makna dalam analisis sentimen. Dengan data *cleaning*, data yang diperoleh menjadi lebih terstruktur dan siap untuk dianalisis secara akurat oleh model klasifikasi.

#### A. Penghapusan data kosong (*null*)

Dalam langkah ini dilakukan penghapusan data kosong (*null* atau NaN) pada kolom ulasan. Baris data yang tidak memuat isi ulasan atau hanya berisi spasi dianggap tidak memberikan kontribusi informasi yang berarti. Keberadaan data semacam ini dapat mengganggu proses analisis, baik dalam perhitungan statistik, pelatihan model *machine learning*, maupun saat visualisasi data. Jumlah entri kosong dari masing-masing destinasi disajikan pada Tabel 4.2.

Tabel 4. 2 Jumlah data kosong pada tiap destinasi

Nama Destinasi	Jumlah Data Kosong
Pulau Morotai	158
Tanjung Kelayang	1013
Danau Toba	1072
Wakatobi	1175
Tanjung Lesung	874

#### B. Eliminasi ulasan yang terdiri kurang dari tiga kata

Ulasan sangat pendek (hanya terdiri dari satu atau dua kata) atau hanya berisi emotikon tidak memberikan konteks yang cukup untuk dianalisis sehingga sulit untuk diproses oleh model Naïve Bayes. Oleh karena itu, ulasan semacam ini dihapus guna menjaga kualitas dan relevansi data. Jumlah ulasan kurang dari tiga kata pada tiap destinasi ditampilkan pada tabel 4.3.

Tabel 4. 3 Jumlah Ulasan Kurang dari 3 Kata pada Tiap Destinasi

<b>Nama Destinasi</b>	<b>Jumlah Ulasan Kurang dari 3 Kata</b>
Pulau Morotai	29
Tanjung Kelayang	210
Danau Toba	360
Wakatobi	250
Tanjung Lesung	254

### C. Eliminasi ulasan yang tidak relevan

Ulasan yang tidak relevan dengan sektor pariwisata—seperti yang membahas hal lain, isu politik, atau mengandung spam—tidak mendukung tujuan analisis sentimen. Untuk menjaga fokus dan kualitas data, ulasan semacam ini disaring dan dihapus. Jumlah ulasan yang tidak relevan ditampilkan pada Tabel 4.4.

Tabel 4. 4 Jumlah Ulasan Tidak Relevan pada Tiap Destinasi

<b>Nama Destinasi</b>	<b>Jumlah Ulasan Tidak Relevan</b>
Pulau Morotai	33
Tanjung Kelayang	329
Danau Toba	275
Wakatobi	118
Tanjung Lesung	235

Jumlah total ulasan pada lima destinasi setelah melalui tahap *cleaning* ditampilkan pada tabel 4.5.

Tabel 4. 5 Jumlah *Review* Setiap Destinasi

<b>Nama Destinasi</b>	<b>Jumlah Review</b>	<b>Data Kosong</b>	<b>Ulasan Kurang dari 3 Kata</b>	<b>Ulasan Tidak Relevan</b>	<b>Jumlah Akhir</b>
Pulau Morotai	316	158	29	33	96
Tanjung Kelayang	1980	1013	210	329	428
Danau Toba	2380	1072	360	275	673
Wakatobi	1940	1175	250	118	397
Tanjung Lesung	1890	874	254	235	527

## 4.4 Pelabelan

Setelah proses penambangan data selesai, langkah berikutnya adalah pelabelan data. Proses ini dilakukan secara manual berdasarkan subjektivitas penulis. Sebelum pelabelan dimulai, data hasil penambangan diekstrak dari Excel dan dipindahkan ke lembar kerja terpisah untuk mempermudah proses pelabelan.

Selanjutnya, setiap *review* dalam data tersebut diberikan label, dengan kategori yang ditentukan yaitu "positif" dan "negatif." Hasil pelabelan dapat dilihat pada gambar 4.1.

1	Review	Sentimen
1	Tempat yang kaya akan alam dan kaya akan narasi sejarah. Menjadi salah satu tempat persinggahan Jepang dan Amerika saat Perang Dunia II. Selain itu pulau-pulau kecil di sekitar Morotai juga sangat amat indah seperti Pantai Dodola dan Pantai Tabailenge.	
2	Sayang pemerintah daerah kurang mengencangkan wisata ke warga setempat, museum PD II dari pemerintah terbengkalai, banyak sampah berserakan.	Negatif
3	Padahal potensi pulau ini sangat tinggi!!	Positif
4	Suasana pulau yang asri dan tenang dengan pemandangan alam yang indah	
5	Keren banget, sebagai destinasi kelas dunia, Morotai pantas sekali. Daya tarik atraksinya ada di bahari atau marine tourism. Punya pantai bagus, pasir putih, laut jernih, feasibility tinggi, langit cerah, lengkap dengan banyak pulau kecil di Samodera Pacific yang bagus bagus. Tinggal akses dan amenitas yang masih belum cukup dan harus diperkuat.	Positif
6	Morotai adalah sebuah pulau di bibir samudera Pasifik, yang terletak di kabupaten Pulau Morotai, Provinsi Maluku Utara. Pulau ini memiliki destinasi wisata laut yang cukup indah, selain itu pulau morotai juga merupakan sejarah bisu perang dunia kedua. Dimana Jenderal Mac.Arthur Douglas pernah mendaratkan kakinya bersama pasukan sekutu di pulau ini untuk menyerang Jepang, dengan strategy leapfrog "lompatan katak"	Positif
7	Bagus sekali pemandangannya.. Bersih.. Asri sejuk..	Positif
8	Tempat yang menyimpan sejarah perang dunia ke II banyak peninggalan Jepang masih tersebar di Pulau ini. Tempat dimana jendral Nakamura bersembunyi hingga perang II berakhir. Menyimpan banyak spot diving yang masih asri. Salah satu destinasi wisata yang jauh dari hinggar	Positif

Gambar 4. 1 Contoh Hasil Pelabelan

#### 4.5 Pembagian Data *Training* dan Data *Testing*

Seluruh data ulasan yang telah melalui tahap *cleaning* dapat dilihat pada tabel 4.5, yang kemudian dibagi secara manual menjadi data pelatihan dan data pengujian dengan rasio 70-30, dengan hasil yang ditampilkan pada tabel 4.6.

Tabel 4. 6. Pembagian Data *Training* dan Data *Testing*

	Training		Testing		Total Training	Total Testing
	Positif	Negatif	Positif	Negatif		
Morotai	57	9	25	5	66	30
Tanjung Kelayang	279	22	118	9	30	127
Toba	448	23	192	10	471	202
Wakatobi	268	10	115	4	278	119
Tanjung Lesung	270	99	115	43	369	158

#### 4.6 Augmentasi Data

Augmentasi data dalam analisis sentimen bertujuan untuk meningkatkan kemampuan model dalam menggeneralisasi data baru serta memperkuat ketahanannya terhadap variasi input. Pada penelitian ini, augmentasi diterapkan sebagai solusi atas ketidakseimbangan jumlah data antara sentimen positif dan negatif, di mana jumlah ulasan negatif jauh lebih sedikit dibandingkan ulasan positif. Ketimpangan ini dapat menyebabkan model cenderung bias terhadap kelas mayoritas, sehingga berpotensi menurunkan akurasi prediksi, khususnya pada kelas minoritas. Contoh hasil augmentasi dapat dilihat pada gambar 4.2.

1	Review Asli	Hasil Augmentasi 1	Hasil Augmentasi 2
2	Infrastruktur belum lengkap	prasarana belum afdal	prasarana belum genap
3	Keren tapi kurang ramai & tidak ada jajanan	elok tapi terbatas ramai & tidak ada jajanan	elok tapi minim ramai & tidak ada jajanan
4	Memang tepat julukannya "Surga di Bawah Laut", tapi di daratannya, Sarana dan Prasarana masih sgt terbatas, mulai dari Penerbangan, Hotel dan transportasi.	masuk akal kena julukannya "Surga di Bawah Laut", tapi di daratannya, Sarana dan Prasarana masih sgt terbatas, mulai dari Penerbangan, Hotel dan transportasi.	dasar langsung julukannya "Surga di Bawah Laut", tapi di daratannya, Sarana dan Prasarana masih sgt terbatas, mulai dari Penerbangan, Hotel dan transportasi.

Gambar 4. 2 Contoh Hasil Augmentasi

Augmentasi diterapkan pada data *training* kelima destinasi yang mengalami ketidakseimbangan jumlah data, khususnya karena jumlah data negatif yang lebih sedikit. Hasil dari proses augmentasi tersebut disajikan dalam tabel 4.7

Tabel 4. 7 Jumlah Data Hasil Augmentasi

	Jumlah Data <i>Train</i>		Jumlah Augmentasi Data	Hasil Akhir	
	Positif	Negatif		Positif	Negatif
Morotai	57	9	45	57	54
Tanjung Kelayang	279	22	242	279	264
Toba	448	23	414	448	437
Wakatobi	268	10	250	268	260
Tanjung Lesung	270	99	99	270	198

## 4.7 Data Preprocessing

Langkah berikutnya adalah melakukan data *preprocessing*, yaitu mengolah data hasil penambangan yang masih dalam format tidak terstruktur agar menjadi bentuk kata dasar. Proses ini dilakukan menggunakan Python dan mencakup empat tahap utama: *case folding*, *tokenizing*, *stop word removal*, dan *stemming*. Setelah semua tahapan *preprocessing* selesai, data yang telah diproses kemudian disimpan kembali ke dalam file Excel.

### 4.7.1 Case Folding

```
# Case folding
text = text.lower()

# Hapus karakter non-alfabet dan angka
text = re.sub(r'^a-zA-Z\s', ' ', text)

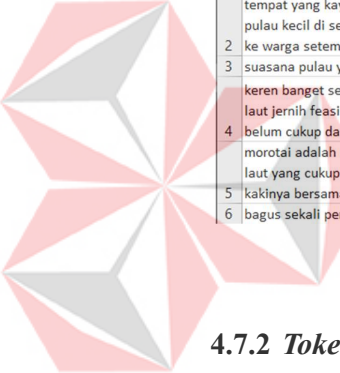
# Hapus spasi ganda
text = re.sub(r'\s+', ' ', text).strip()
```

Gambar 4. 3 Tahap Case Folding



Tahap *case folding*, seperti pada Gambar 4.3, dilakukan dengan mengubah seluruh huruf menjadi kecil menggunakan fungsi *lower text* untuk menyeragamkan format penulisan. Selanjutnya, karakter non-alfabet seperti angka, tanda baca, dan simbol dihapus agar teks lebih bersih dan fokus pada informasi penting.

Langkah selanjutnya adalah membersihkan spasi yang berlebihan akibat penghapusan karakter sebelumnya. Fungsi khusus digunakan untuk merapikan teks agar tidak mengandung spasi ganda atau spasi di awal dan akhir kalimat. Misalnya, ulasan “Pulau kecil dengan keindahan yang alami, 🍌” akan diubah menjadi “pulau kecil dengan keindahan yang alami” setelah melalui proses ini. Proses ini dilakukan untuk memastikan data teks menjadi lebih konsisten dan siap untuk tahap analisis berikutnya. Hasil *case folding* dapat dilihat pada gambar 4.4.



	case_folding
1	tempat yang kaya akan alam dan kaya akan narasi sejarah menjadi salah satu tempat persinggahan jepang dan amerika saat perang dunia ii selain itu pulau
2	pulau kecil di sekitar morotai juga sangat amat indah seperti pantai dodola dan pantai tabailenge sayang pemerintah daerah kurang mengencarkan wisata
3	ke warga setempat museum pd ii dari pemerintah terbengkalai banyak sampah berserakan padahal potensi pulau ini sangat tinggi
3	suasana pulau yang asri dan tenang dengan pemandangan alam yang indah
4	keren banget sebagai destinasi kelas dunia morotai pantas sekali daya tarik atraksinya ada di bahari atau marine tourism punya pantai bagus pasir putih
4	laut jernih feasibility tinggi langit cerah lengkap dengan banyak pulau kecil di samodera pacific yang bagus bagus tinggal akses dan amenitas yang masih
4	belum cukup dan harus diperkuat
5	morotai adalah sebuah pulau di bibir samudera pasifik yang terletak di kabupaten pulau morotai provinsi maluku utara pulau ini memiliki destinasi wisata
5	laut yang cukup indah selain itu pulau morotai juga merupakan sejarah bisu perang dunia kedua dimana jenderal mac arthur douglas pernah mendaratkan
5	kakinya bersama pasukan sekutu di pulau ini untuk menyerang jepang dengan strategy leapfrog lompatan katak
6	bagus sekali pemnadangannya bersih asri sejuk

Gambar 4. 4 Hasil *Case Folding*

#### 4.7.2 Tokenizing

```
# Tokenizing
tokens = word_tokenize(text)
```

Gambar 4. 5 Tahap *Tokenizing*

Pada tahap *tokenizing*, seperti ditunjukkan gambar 4.5, teks yang sudah dibersihkan dipecah menjadi kata-kata atau token individual menggunakan fungsi *word tokenizing*. Proses ini memudahkan analisis secara lebih mendalam terhadap tiap kata dalam ulasan, seperti penghitungan frekuensi, identifikasi kata penting, atau klasifikasi sentimen.

Sebagai contoh, teks “pulau kecil dengan keindahan yang alami” akan dipecah menjadi token seperti [“pulau”, “kecil”, “dengan”, “keindahan”, “yang”, “alami”]. Proses ini menjadi dasar untuk tahapan pengolahan teks berikutnya. Hasil *tokenizing* dapat dilihat pada gambar 4.6.

1	tokenized
	['tempat', 'yang', 'kaya', 'akan', 'alam', 'dan', 'kaya', 'akan', 'narasi', 'sejarah', 'menjadi', 'salah', 'satu', 'tempat', 'persinggahan', 'jepang', 'dan', 'amerika', 'saat', 'perang', 'dunia', 'ii', 'selain', 'itu', 'pulau', 'pulau', 'kecil', 'di', 'sekitar', 'morotai', 'juga', 'sangat', 'amat', 'indah', 'seperti', 'pantai', 'dodola', 'dan', 'pantai', 'tabailenge', 'sayang', 'pemerintah', 'daerah', 'kurang', 'menggencarkan', 'wisata', 'ke', 'warga', 'setempat', 'museum', 'pd', 'ii', 'dari', 'pemerintah', 'terbengkalai', 'banyak', 'sampah', 'berserakan', 'padahal', 'potensi', 'pulau', 'pulau', 'ini', 'sangat', 'tinggi']
2	['suasana', 'pulau', 'yang', 'asri', 'dan', 'tenang', 'dengan', 'pemandangan', 'alam', 'yang', 'indah']
3	['keren', 'banget', 'sebagai', 'destinasi', 'kelas', 'dunia', 'morotai', 'pantas', 'sekali', 'daya', 'tarik', 'atraksinya', 'ada', 'di', 'bahari', 'atau', 'marine', 'tourism', 'punya', 'pantai', 'bagus', 'pasir', 'putih', 'laut', 'jernih', 'feasibility', 'tinggi', 'langit', 'cerah', 'lengkap', 'dengan', 'banyak', 'pulau', 'kecil', 'di', 'samodera', 'pacific', 'yang', 'bagus', 'bagus', 'tinggal', 'akses', 'dan', 'amenitas', 'yang', 'masih', 'belum', 'cukup', 'dan', 'harus', 'diperkuat']
4	

Gambar 4. 6 Hasil *Tokenizing*

#### 4.7.3 Stop Word Removal

```
#Stopword removal menggunakan Sastrawi
tokens = [w for w in tokens if w not in stop_words]
```

Gambar 4. 7 Tahap *Stop Word Removal*

Pada tahap *stopword removal* seperti pada gambar 4.7, kata-kata umum yang kurang bermakna seperti “dan”, “yang”, atau “di” dihapus dari kumpulan kata menggunakan daftar *stopwords* bahasa Indonesia dari library Sastrawi. Tujuan dari proses ini adalah untuk memfokuskan analisis pada kata-kata penting yang lebih relevan dalam penentuan sentimen atau makna teks.

Misalnya, dari kalimat token [“pulau”, “kecil”, “dengan”, “keindahan”, “yang”, “alami”], kata “dengan” dan “yang” akan dihapus sehingga tersisa [“pulau”, “kecil”, “keindahan”, “alami”]. Hasil *stopword removal* dapat dilihat pada gambar 4.8.

1	stopword_removed
	['kaya', 'alam', 'kaya', 'narasi', 'sejarah', 'salah', 'persinggahan', 'jepang', 'amerika', 'perang', 'dunia', 'ii', 'pulau', 'pulau', 'morotai', 'indah', 'pantai', 'dodola', 'pantai', 'tabailenge', 'sayang', 'pemerintah', 'daerah', 'menggencarkan', 'wisata', 'warga', 'museum', 'pd', 'ii', 'pemerintah', 'terbengkalai', 'sampah', 'berserakan', 'potensi', 'pulau']
2	['suasana', 'pulau', 'asri', 'tenang', 'pemandangan', 'alam', 'indah']
3	['keren', 'banget', 'destinasi', 'kelas', 'dunia', 'morotai', 'daya', 'tarik', 'atraksinya', 'bahari', 'marine', 'tourism', 'pantai', 'bagus', 'pasir', 'putih', 'laut', 'jernih', 'feasibility', 'langit', 'cerah', 'lengkap', 'pulau', 'samodera', 'pacific', 'bagus', 'bagus', 'tinggal', 'akses', 'amenitas', 'diperkuat']
4	

Gambar 4. 8 Hasil *Stop Word Removal*

#### 4.7.4 Stemming

```
#Stemming
stemmed_tokens = [stemmer.stem(w) for w in tokens]
```

Gambar 4. 9 Tahap *Stemming*

Pada tahap *stemming* seperti pada gambar 4.9, setiap kata diubah ke bentuk dasarnya menggunakan fungsi *stemmer* dari library Sastrawi agar variasi kata yang memiliki makna sama dapat dikenali sebagai satu kata tunggal. Proses ini membantu menyederhanakan teks dan meningkatkan konsistensi data dalam analisis.

Misalnya, dari token hasil *stopword removal* seperti [“pulau”, “kecil”, “keindahan”, “alami”], kata-kata tersebut akan diubah ke bentuk dasar sesuai aturan bahasa Indonesia, misalnya “keindahan” menjadi “indah” sehingga hasilnya menjadi [“pulau”, “kecil”, “indah”, “alami”]. Hasil *stemming* dapat dilihat pada gambar 4.10.

1	stemmed
	['kaya', 'alam', 'kaya', 'narasi', 'sejarah', 'salah', 'singgah', 'jepang', 'amerika', 'perang', 'dunia', 'ii', 'pulau', 'pulau', 'morotai', 'indah', 'pantai', 'dodola', 'pantai', 'tabailenge', 'sayang', 'perintah', 'daerah', 'gencar', 'wisata', 'warga', 'museum', 'pd', 'ii', 'perintah', 'bengkalai', 'sampah', 'sera', 'potensi', 'pulau']
2	['suasana', 'pulau', 'asri', 'tenang', 'pandang', 'alam', 'indah']
3	['keren', 'banget', 'destinasi', 'kelas', 'dunia', 'morotai', 'daya', 'tarik', 'atraksi', 'bahari', 'marine', 'tourism', 'pantai', 'bagus', 'pasir', 'putih', 'laut', 'jernih', 'feasibility', 'langit', 'cerah', 'lengkap', 'pulau', 'samodera', 'pacific', 'bagus', 'bagus', 'tinggal', 'akses', 'amenitas', 'kuat']
4	

Gambar 4. 10 Hasil *Stemming*

#### 4.7.5 Hasil *Preprocessing*

```
return ' '.join(corrected_tokens)

# Terapkan preprocessing ke kolom teks
df['processed_text'] = df[text_column].apply(preprocess_text)
```

Gambar 4. 11 Tahap Akhir *Preprocessing*

Setelah melewati tahapan *preprocessing*, token-token kata yang sudah diproses digabung kembali menjadi kalimat utuh dengan menggunakan fungsi *return* dan *join* seperti pada gambar 4.11. Hasil penggabungan ini kemudian dikembalikan sebagai output dari fungsi *preprocessing* untuk setiap baris data teks.

Selanjutnya, seluruh proses *preprocessing* diterapkan secara otomatis ke kolom teks dalam dataset menggunakan fungsi khusus yang memproses setiap entri teks. Sebagai contoh, teks awal “Pulau kecil dengan keindahan yang alami, 🍌” setelah melalui tahapan *preprocessing* menjadi “pulau kecil indah alami”. Dengan demikian, data teks yang telah bersih dan konsisten siap digunakan untuk analisis lebih lanjut. Hasil akhir tahap *preprocessing* dapat dilihat pada Gambar 4.12.

1	hasil_akhir
2	kaya alam kaya narasi sejarah salah singgah jepang amerika perang dunia ii pulau pulau morotai indah pantai dodola pantai tabailenge sayang perintah daerah gencar wisata warga museum pd ii perintah bengkalai sampah sera potensi pulau
3	suasana pulau asri tenang pandang alam indah
4	keren banget destinasi kelas dunia morotai daya tarik atraksi bahari marine tourism pantai bagus pasir putih laut jernih feasibility langit cerah lengkap pulau samodera pacific bagus bagus tinggal akses amenitas kuat

Gambar 4. 12 Hasil Akhir *Preprocessing*

#### 4.8 Implementasi Naïve Bayes

Naïve Bayes adalah metode yang digunakan untuk memprediksi kemungkinan kemunculan suatu kelas. Langkah awal dalam proses klasifikasi adalah mengimpor data *training* dan *testing* dari lima sumber yang telah ditentukan.

Selanjutnya, data *training* dan *testing* diambil dari folder tertentu untuk diproses. Setelah data tersedia, dilakukan pengecekan apakah kolom *review* dan sentimen ada. Jika keduanya tersedia, maka kolom sentimen akan digunakan sebagai label acuan dalam proses prediksi menggunakan algoritma *Naïve Bayes*.

##### A. Membangun model Naïve Bayes

Model Naïve Bayes adalah algoritma klasifikasi berbasis probabilistik yang bekerja dengan prinsip Teorema Bayes, yaitu menghitung kemungkinan suatu data termasuk ke dalam kelas tertentu berdasarkan informasi dari fitur-fiturnya. Pada penerapannya, model ini mengasumsikan bahwa setiap fitur (kata-kata dalam *review*) bersifat independen terhadap fitur lainnya, meskipun pada kenyataannya tidak selalu demikian. Asumsi inilah yang disebut sebagai naïve (naif).

Model Multinomial Naïve Bayes digunakan karena cocok untuk data teks berbasis frekuensi kata dengan menggunakan fungsi *Count Vectorizer*. Proses *fit* melatih model menggunakan data latih ( $X_{\text{train}}$ ,  $y_{\text{train}}$ ) dimana  $X$  merupakan *review* dan  $y$  merupakan label, sedangkan *predict* digunakan untuk memprediksi kelas dari data uji ( $X_{\text{test}}$ ) berdasarkan kemungkinan tertinggi. Langkah Naïve Bayes di Python dapat dilihat pada lampiran 1. Secara teoritis, Naïve Bayes dengan *Laplace Correction* menghitung probabilitas suatu *review* termasuk dalam kelas positif atau negatif dengan rumus dapat dilihat pada poin 2.8.

Pada tahap ini dilakukan perhitungan manual pada data *train* salah satu destinasi yang sudah di augmentasi, yaitu Pulau Morotai. Berdasarkan data yang sudah diolah, terdapat 111 ulasan yang terdiri dari 57 ulasan berlabel positif dan 54 ulasan berlabel negatif. Dari total 111 ulasan yang digunakan, probabilitas ulasan positif adalah sebesar  $P(Positif) = \frac{57}{111} = 0,513$  dan probabilitas ulasan negatif adalah sebesar  $P(Negatif) = \frac{54}{111} = 0,486$ . Berdasarkan data tersebut, jumlah kata dalam dataset terdiri dari 512 kata yang berasal dari ulasan positif dan 665 kata dari ulasan negatif. Secara keseluruhan, terdapat 348 kata unik dalam dataset.

Langkah berikutnya adalah menghitung probabilitas tiap kata dengan rumus Naïve Bayes berdasarkan frekuensi kata di data *training*. Pada tahap ini, digunakan satu *review* data *testing* untuk dihitung probabilitasnya secara manual yaitu: **“Pulau tropis yang kaya akan sejarah ini benar-benar layak untuk dikunjungi.”** Setelah melalui proses *preprocessing*, kalimat tersebut berubah menjadi: **“pulau tropis kaya sejarah benar benar layak kunjung.”**

Tabel 4. 8 Frekuensi Probabilitas Kata pada Data *Testing*

No.	Kata	Training Positif	Training Negatif
1	pulau	22	20
2	kaya	1	2
3	sejarah	7	6
4	layak	1	6
5	kunjung	7	0

Tabel 4.8 berisikan kata-kata berdasarkan *review* yang digunakan untuk testing pada perhitungan manual Naïve Bayes. Nilai probabilitas dapat dicari dengan menggunakan rumus *laplace correction* yaitu  $P(Mk|Cl)$  sebagai berikut :

$$P(\text{pulau}|\text{negatif}) = (20+1) / (665+348) = 0.020730503$$

$$P(\text{pulau}|\text{positif}) = (22+1) / (512+348) = 0.026744186$$

$$P(\text{kaya}|\text{negatif}) = (2+1) / (665+348) = 0.0029615$$

$$P(\text{kaya}|\text{positif}) = (1+1) / (512+348) = 0.002325581$$

Berdasarkan hasil perhitungan dengan menggunakan rumus  $P(Mk|Cl)$ , maka didapati hasil sebagai berikut :

Tabel 4. 9 Hasil Testing dengan *Laplace Correction*

No.	Kata	$X_k Negatif$	$X_k Positif$	$P(Mk Cl)$ Positif	$P(Mk Cl)$ Negatif
1	pulau	22	20	0.026744186	0.020730503
2	kaya	1	2	0.002325581	0.0029615
3	sejarah	7	6	0.009302326	0.006910168
4	layak	1	6	0.002325581	0.006910168
5	kunjung	7	0	0.009302326	0.000987167

Berdasarkan Tabel 4.9, diperoleh probabilitas tiap kata dari data pelatihan. Nilai-nilai probabilitas tersebut kemudian dikalikan untuk memperoleh hasil prediksi dari data uji. Berikut merupakan perhitungan probabilitas masing-masing kata pada kalimat “**pulau tropis kaya sejarah benar benar layak kunjung.**”

Positif

$$\begin{aligned}
 &= P(\text{positif}) \times P(\text{pulau}|\text{positif}) \times P(\text{kaya}|\text{positif}) \times P(\text{sejarah}|\text{positif}) \times \\
 &P(\text{layak}|\text{positif}) \times P(\text{kunjung}|\text{positif}) \\
 &= 0,513 \times 0.026744186 \times 0,002325581 \times 0,009302326 \times 0,002325581 \times 0,009302326 \\
 &= 6.4225 \times 10^{-12}
 \end{aligned}$$

Negatif

$$\begin{aligned}
 &= P(\text{negatif}) \times P(\text{pulau}|\text{negatif}) \times P(\text{kaya}|\text{negatif}) \times P(\text{sejarah}|\text{negatif}) \times \\
 &P(\text{layak}|\text{negatif}) \times P(\text{kunjung}|\text{negatif}) \\
 &= 0,486 \times 0,020730503 \times 0,0029615 \times 0,006910168 \times 0,006910168 \times 0,000987167 \\
 &= 1.4066 \times 10^{-12}
 \end{aligned}$$

Setelah itu, hasil dari perkalian probabilitas positif dan negatif dinormalisasi untuk mengetahui persentase probabilitasnya.

$$\text{Probabilitas Positif} = \frac{6.4225 \times 10^{-12}}{(1.4066 \times 10^{-12} + 6.4225 \times 10^{-12})} = 0.820$$

$$\text{Probabilitas Negatif} = \frac{1.4066 \times 10^{-12}}{(1.4066 \times 10^{-12} + 6.4225 \times 10^{-12})} = 0.179$$



Perhitungan di atas merupakan bagian dari proses klasifikasi sentimen berdasarkan nilai probabilitas yang diperoleh, di mana probabilitas positif mendapat nilai tertinggi sebesar 0,820. Oleh karena itu, ulasan diklasifikasikan ke dalam sentimen "positif". Hal ini sesuai dengan hasil naïve bayes di python yang ditunjukkan pada gambar 4.13.

1	Review	Prob_Negatif	Prob_Positif	Prediksi	Sentimen Asli
2	Pulau tropis yang kaya akan sejarah ini benar-benar layak untuk dik	0.179166157	0.820833843	Positif	Positif
3	Asri namun lumayan panas ☹️	0.704705425	0.295294575	Negatif	Negatif
4	Bawa baby oil untuk menghindari sengatan agas	0.616480285	0.383519715	Negatif	Negatif
5	Kurangnya fasilitas umum di kota..Cafe2 yang kurang bahkan tidak a	0.998439055	0.001560945	Negatif	Negatif
6	Kota sejarah perang dunia kedua ketika menjadi pangkalan militer !	0.263829524	0.736170476	Positif	Negatif
7	Tidak terawat sama sekali, pemda sangat abai	0.034454222	0.965545778	Positif	Negatif

Gambar 4. 13 Hasil Naive Bayes Python

Selain itu, diperoleh pula informasi mengenai proporsi hasil distribusi sentimen yang diperoleh dari proses klasifikasi sentimen terhadap kelima destinasi wisata yang menjadi fokus penelitian. Hasil proporsi distribusi sentimen tersebut memberikan gambaran mengenai bagaimana persepsi masyarakat terhadap masing-masing destinasi. Rincian lengkap mengenai proporsi distribusi sentimen untuk setiap destinasi dapat dilihat pada Tabel 4.10.

Tabel 4. 10 Persentase Hasil Distribusi Sentimen Naive Bayes

Kriteria	Morotai	Tanjung Kelayang	Toba	Wakatobi	Tanjung Lesung
Jumlah Data Positif	27	121	195	116	114
Jumlah Data Negatif	3	6	7	3	44
Persentase Data Positif	86.67%	95.28%	96.53%	97.48%	72.15%
Persentase Data Negatif	13.33%	4.72%	3.47%	2.52%	27.85%

## B. Evaluasi dengan *Confusion Matrix*

*Confusion matrix* adalah tabel yang menunjukkan performa model klasifikasi dengan cara membandingkan hasil prediksi model terhadap data sebenarnya. Matriks ini terdiri dari empat komponen utama: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), yang menggambarkan jumlah prediksi yang benar dan salah untuk setiap kelas. Code *confusion matrix* dapat dilihat pada Gambar L1.7.

Hasil tersebut kemudian divisualisasikan menggunakan *heatmap* dengan sumbu x menunjukkan label hasil prediksi dan sumbu y menunjukkan label sebenarnya, sehingga memudahkan pemahaman terhadap performa model. Visualisasi *confusion matrix* ini dapat dilihat pada gambar 4.14.



Gambar 4. 14 Hasil *Confusion Matrix*

### C. Evaluasi dengan *Classification Report*

*Classification report* adalah metode evaluasi yang digunakan untuk mengukur performa model klasifikasi dengan menghitung metrik-metrik utama seperti *precision*, *recall*, dan *f1-score* untuk setiap kelas. Laporan ini bekerja dengan membandingkan hasil prediksi model dengan label sebenarnya dari data uji. Code *classification report* dapat dilihat pada Gambar L1.8. Berdasarkan perbandingan tersebut, *classification report* menghitung:

1. *Precision* yaitu seberapa banyak prediksi suatu kelas yang benar dibandingkan seluruh prediksi pada kelas tersebut. Rumusnya :

$$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)}$$

2. *Recall* yaitu seberapa banyak data yang benar-benar termasuk dalam suatu kelas yang berhasil diprediksi dengan tepat oleh model. Rumusnya :

$$Recall = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)}$$



3. *F1-score*, yaitu rata-rata antara *precision* dan *recall* yang memberikan gambaran menyeluruh terhadap keseimbangan keduanya. Rumusnya :

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Contoh perhitungan manualnya adalah sebagai berikut :

Pada gambar 4.14 didapati hasil dari *Confusion Matrix* untuk pulau Morotai yaitu *True Positive* dengan skor 24, *True Negative* dengan skor 3, *False Positive* dengan skor 2 dan *False Negative* dengan skor 1. Jika dihitung menggunakan rumus *Classification Report* maka didapati hasil :

$$1. \quad Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)}$$

$$Precision = \frac{24}{24+2}$$

$$Precision = 0.9231$$

$$2. \quad Recall = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)}$$

$$Recall = \frac{24}{24+1}$$

$$Recall = 0.96$$

$$3. \quad F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$F1 - score = 2 \times \frac{0.9230 \times 0.96}{0.9230 + 0.96}$$

$$F1 - score = 0.9412$$

Hasil *Classification Report* untuk Pulau Morotai pada Gambar 4.15 menunjukkan kesesuaian antara perhitungan manual dan Python, menandakan bahwa metrik evaluasi seperti *precision*, *recall*, dan *f1-score* yang dihitung Python akurat dan dapat diandalkan dalam analisis kinerja model klasifikasi sentimen.

Classification Report:				
	precision	recall	f1-score	support
Negatif	0.7500	0.6000	0.6667	5
Positif	0.9231	0.9600	0.9412	25
accuracy			0.9000	30
macro avg	0.8365	0.7800	0.8039	30
weighted avg	0.8942	0.9000	0.8954	30

Gambar 4. 15 Hasil *Classification Report*

#### D. Pengujian 10-fold cross-validation

10-Fold Cross-Validation adalah metode evaluasi model yang membagi data latih menjadi 10 bagian (*fold*), lalu melatih dan menguji model sebanyak 10 kali dengan kombinasi bagian yang berbeda. Code 10-fold cross-validation dapat dilihat pada Gambar L1.9. Proses ini menghasilkan 10 nilai akurasi yang kemudian dikonversi ke dalam DataFrame untuk ditampilkan. Selanjutnya, dihitung rata-rata akurasi sebagai ukuran performa umum model, dan deviasi standarnya digunakan untuk mengetahui seberapa konsisten model terhadap variasi data. Hasil 10-fold cross-validation dapat dilihat pada gambar 4.16.

```

Hasil 10-Fold Cross Validation:
Akurasi
0 0.888889
1 0.777778
2 0.888889
3 0.888889
4 0.750000
5 1.000000
6 1.000000
7 1.000000
8 0.875000
9 1.000000

Rata-rata Akurasi: 0.9069
Deviasi Standar Akurasi: 0.0883

```

Gambar 4. 16 Hasil Evaluasi 10-fold cross-validation

#### 4.9 Hasil Evaluasi Naïve Bayes

Pemodelan sentimen menggunakan algoritma Naïve Bayes pada review lima destinasi prioritas di Indonesia menghasilkan prediksi probabilitas untuk klasifikasi positif (mendukung) dan negatif (tidak mendukung). Evaluasi dilakukan melalui *confusion matrix* dan *classification report* yang menunjukkan akurasi model berdasarkan metrik *precision*, *recall*, dan *f1-score*.

#### 4.9.1 Hasil *Classification Report*

*Classification Report* menyajikan ringkasan hasil evaluasi performa model Naïve Bayes terhadap setiap kelas dalam data. *Precision* menunjukkan ketepatan prediksi positif, *recall* mengukur kemampuan mendeteksi semua data relevan, dan *F1-score* mencerminkan keseimbangan keduanya. Pada tabel 4.11 disajikan hasil *Classification Report* dari kelima destinasi.

Tabel 4. 11 Hasil *Classification Report* kelima destinasi

Kriteria	Morotai	Tanjung Kelayang	Toba	Wakatobi	Tanjung Lesung
<i>Precision</i>	92.31%	96.69%	96.91%	98.28%	85.96%
<i>Recall</i>	96.00%	99.15%	97.92%	99.13%	85.22%
<i>F1 – score</i>	94.12%	97.91%	97.41%	98.70%	85.59%
<i>Support</i> (positif)	25	118	192	115	115
<i>Support</i> (negatif)	5	9	10	4	43
<i>Accuracy</i>	90.00%	96.06%	95.05%	97.48%	79.11%

#### 4.9.2 Hasil 10-fold cross-validation

Setelah pemodelan, langkah berikutnya adalah menguji validitas dengan menggunakan teknik *10-fold cross-validation*, di mana data secara otomatis dibagi menjadi data *training* dan data *testing*, pada setiap iterasi. Proporsi pembagian data tersebut dapat dilihat pada Tabel 4.6. Sementara itu, Tabel 4.11 menyajikan hasil uji validitas model Naïve Bayes berdasarkan evaluasi dari proses *cross-validation*.

Tabel 4. 12 Hasil 10- fold cross-validation

Fold ke	Morotai	Tanjung Kelayang	Toba	Wakatobi	Tanjung Lesung
0	0.888889	0.914286	0.932203	0.921053	0.957447
1	0.777778	0.857143	0.949153	0.973684	0.978723
2	0.888889	0.914286	1.000000	0.947368	0.936170
3	0.888889	0.971429	0.983051	0.973684	0.787234
4	0.750000	0.942857	0.915254	1.000000	0.851064
5	1.000000	0.970588	0.949153	0.973684	0.893617
6	1.000000	0.970588	0.982759	0.973684	0.956522
7	1.000000	0.911765	0.948276	0.947368	0.956522
8	0.875000	1.000000	0.948276	0.972973	0.891304
9	1.000000	0.941176	0.982759	0.783784	0.804348
AVG	0.9069	0.9394	0.9591	0.9467	0.9013
Deviasi	0.0883	0.0393	0.0254	0.0580	0.0645

#### 4.10 Hasil Akhir Naïve Bayes

Hasil akhir dari penerapan metode Naïve Bayes yang telah dilakukan ditunjukkan pada Tabel 4.13, berdasarkan evaluasi yang dilakukan melalui *classification report* serta pengujian menggunakan metode *10-fold cross-validation*.

Tabel 4. 13. Hasil Akhir Naive Bayes

	Morotai	Kelayang	Toba	Wakatobi	Lesung
<i>Accuracy</i>	90.69 %	93.94%	95.91%	94.67%	90.13%
<i>Precision</i>	92.31 %	96.69%	96.91%	98.28%	85.96%
<i>Recall (Sens)</i>	96.00 %	99.15%	97.92%	99.13%	85.22%
<i>Positive(F1)</i>	94.12 %	97.91%	97.41%	98.70%	85.59%
<i>Negative(F1)</i>	66.67 %	66.67%	44.44%	57.14%	62.07%
<i>Support (positif)</i>	25	118	192	115	115
<i>Support (negatif)</i>	5	9	10	4	43
<i>Jumlah Data Positif (Klasifikasi)</i>	27	121	195	116	114
<i>Jumlah Data Negatif (Klasifikasi)</i>	3	6	7	3	44
<i>Persentase Positif Data</i>	86.67%	95.28%	96.53%	97.48%	72.15%
<i>Persentase Negatif Data</i>	13.33%	4.72%	3.47%	2.52%	27.85%

Dari tabel 4.13 dapat diketahui bahwa:

1. Akurasi diambil dari rata-rata *10-fold cross-validation* untuk mencerminkan generalisasi model, sedangkan *precision*, *recall*, dan *F1-score* dari *classification report* untuk menilai performa tiap kelas secara rinci.
2. Rata-rata akurasi model sebesar 93,07%, dengan akurasi tertinggi pada Danau Toba (95,91%) dan terendah pada Pantai Tanjung Lesung (90,13%). Hasil ini menunjukkan bahwa model memiliki performa klasifikasi yang baik dan andal dalam memprediksi sentimen pengunjung terhadap destinasi wisata.
3. Rata-rata *precision* lima destinasi sebesar 94,03%, dengan Wakatobi tertinggi (98,28%) dan Pantai Tanjung Lesung terendah (85,96%), menunjukkan ketepatan tinggi model dalam mengidentifikasi sentimen positif.
4. Rata-rata nilai *recall* kelima destinasi mencapai 95,48%, dengan Pantai Tanjung Kelayang tertinggi (99,15%) dan Pantai Tanjung Lesung terendah (85,22%). Hasil ini menunjukkan bahwa model cukup efektif dalam mengenali sebagian besar ulasan yang relevan, khususnya sentimen positif.

5. Rata-rata *F1-score* positif sebesar 94,75%, dengan Wakatobi tertinggi (98,70%) dan Pantai Tanjung Lesung terendah (85,59%). Untuk sentimen negatif, rata-rata *F1-score* hanya 59,40%, tertinggi pada Pulau Morotai dan Pantai Tanjung Kelayang (66,67%) serta terendah pada Danau Toba (44,44%). Hasil menunjukkan model sangat baik dalam mengenali sentimen positif, tapi masih kurang optimal dalam mendeteksi sentimen negatif.
6. Berdasarkan hasil klasifikasi Naïve Bayes, proporsi sentimen positif tertinggi diraih oleh Wakatobi (97,48%). Sementara itu, Pantai Tanjung Lesung mencatat proporsi sentimen negatif tertinggi sebesar 27,85%, menunjukkan persepsi pengunjung yang lebih beragam atau pengalaman negatif yang lebih dominan.

#### 4.11 Visualisasi

Tahap visualisasi data dilakukan melalui salah satu metode yaitu *Word Cloud*. Visualisasi menggunakan *Word Cloud* bertujuan untuk menampilkan kata-kata yang paling sering muncul dalam ulasan yang telah dianalisis sebelumnya, di mana ukuran font setiap kata mencerminkan frekuensi kemunculannya.

##### 4.11.1 Visualisasi Pulau Morotai



Gambar 4. 17 Hasil *Wordcloud* Pulau Morotai

Gambar 4.17 menunjukkan kata-kata seperti “indah”, “bagus”, dan “alam” paling menonjol dalam ulasan positif, menggambarkan pengalaman menyenangkan di Pulau Morotai. Kata lain seperti “laut”, “selam”, dan “libur” mengindikasikan daya tarik utama destinasi ini adalah keindahan bawah laut dan wisata alam.

#### 4.11.2 Visualisasi Pantai Tanjung Kelayang



Sementara itu, pada bagian sentimen negatif, kata yang paling sering muncul adalah “sampah”, diikuti oleh “kotor”, “karang”, dan “awat”. Dominasi kata “sampah” dan “kotor”, mencerminkan keluhan yang serius dari pengunjung terkait buruknya kebersihan di beberapa lokasi dan tempat. Kata “karang” dalam konteks ini muncul karena adanya keluhan tentang kondisi karang yang tajam atau kurang terjaga. Kata-kata lain seperti “harga”, “mahal”, dan “sepi” menunjukkan kekhawatiran terhadap aspek keamanan, biaya, dan kurangnya pengunjung.



#### 4.11.3 Visualisasi Danau Toba

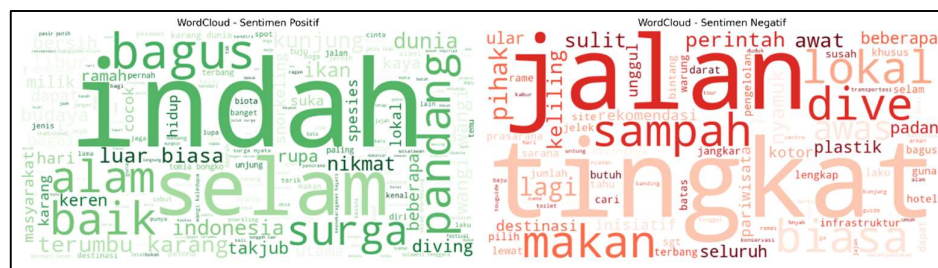


Gambar 4. 19 Hasil *WordCloud* Danau Toba

Berdasarkan Gambar 4.19, kata dominan dalam ulasan positif adalah “indah”, “pandang”, dan “alam”, menunjukkan panorama dan kondisi alam menjadi daya tarik utama. Kata seperti “bagus”, “luar biasa”, “sejuk”, dan “tenang” juga menandakan apresiasi wisatawan terhadap suasana yang nyaman dan tenang.

Sementara itu, dalam sentimen negatif, kata yang paling mencolok adalah “kotor” dan “letus”, yang menandakan keluhan utama terkait kebersihan dan kemungkinan masalah erupsi gunung berapi. Kata-kata lain seperti “jelek”, “mahal”, “sampah”, dan “ganggu” menunjukkan adanya ketidakpuasan terhadap kondisi visual, harga yang dirasa tinggi, kebersihan lingkungan, serta kenyamanan.

#### 4.11.4 Visualisasi Wakatobi



Gambar 4. 20 Hasil *WordCloud* Wakatobi

Berdasarkan gambar 4.20, kata-kata dominan dalam ulasan positif adalah “indah”, “selam”, dan “alam”, menunjukkan bahwa aspek keindahan alam bawah laut menjadi daya tarik utama destinasi wisata ini. Kata-kata seperti “bagus”, “baik”, “pandang”, dan “surga” juga mencerminkan kesan mendalam pengunjung

terhadap pemandangan yang luar biasa dan pengalaman menyelam yang memuaskan.

Di sisi lain, dalam sentimen negatif, kata yang paling menonjol adalah “jalan”, “tingkat”, dan “makan”. Hal ini menunjukkan adanya keluhan terhadap fasilitas umum yang perlu ditingkatkan seperti akses jalan yang kurang memadai serta pengalaman kuliner yang tidak sesuai harapan. Kemunculan kata “sampah”, “kotor”, dan “plastik” juga mencerminkan permasalahan kebersihan yang cukup mengganggu kenyamanan pengunjung.

#### 4.11.5 Visualisasi Pantai Tanjung Lesung



Gambar 4. 21 Hasil *WordCloud* Pantai Tanjung Lesung

Berdasarkan gambar 4.21, dalam ulasan sentimen positif, kata-kata yang paling dominan adalah “bagus”, “bersih”, dan “indah”. Ketiganya menunjukkan bahwa pengunjung sangat mengapresiasi kebersihan, keindahan alam, serta kualitas tempat wisata secara keseluruhan. Kata “pandang”, “jernih”, dan “pasir putih” semakin menegaskan daya tarik visual menjadi kekuatan utama pada destinasi ini.

Sementara itu, dalam sentimen negatif, kata “tiket”, “mahal”, dan “makan” mendominasi, yang menunjukkan bahwa pengunjung merasa harga tiket masuk dan makanan tergolong tinggi. Kata “kotor”, “toilet”, dan “fasilitas” mencerminkan keluhan terhadap kebersihan dan sarana pendukung yang dinilai belum memadai. Kata “jauh”, “parkir”, dan “kecil” juga menunjukkan adanya kendala terkait jarak, aksesibilitas, serta ukuran atau kapasitas yang tidak sesuai harapan.



#### 4.11.6 Hasil Akhir Visualisasi

Hasil dari tahap visualisasi menunjukkan bahwa :

1. Daya tarik utama yang sering disebut dalam sentimen positif mencakup keindahan alam, kebersihan, dan aktivitas wisata seperti menyelam dan menikmati panorama. Kata-kata seperti "*indah*", "*bagus*", dan "*alam*" konsisten muncul di semua destinasi.
2. Keluhan utama yang muncul dalam sentimen negatif berkisar pada isu kebersihan ("*sampah*", "*kotor*"), fasilitas umum ("*toilet*", "*parkir*", "*akses jalan*"), dan harga ("*mahal*", "*tiket*"). Beberapa destinasi juga mendapatkan kritik terhadap pengelolaan dan perhatian terhadap infrastruktur.
3. Berdasarkan hasil visualisasi, Pulau Morotai mendapat keluhan dominan mengenai isu "*panas*", "*potensi*", dan "*fasilitas*" yang menunjukkan kurangnya kenyamanan serta infrastruktur yang belum memadai; Pantai Tanjung Kelayang dikeluhkan karena isu "*sampah*", "*kotor*", dan "*mahal*" yang mencerminkan masalah kebersihan dan harga; Danau Toba mendapat keluhan atas kata "*kotor*", "*letus*", dan "*mahal*" yang mengindikasikan persoalan kebersihan, potensi bencana, dan biaya; Wakatobi menunjukkan keluhan terhadap "*jalan*", "*makan*", dan "*sampah*" yang menandakan kurangnya fasilitas umum dan pengelolaan lingkungan; sedangkan Pantai Tanjung Lesung dikritik melalui kata "*tiket*", "*toilet*", dan "*parkir*" yang mencerminkan keluhan terhadap harga, kebersihan, dan aksesibilitas lokasi.
4. Secara keseluruhan, analisis ini menunjukkan bahwa meskipun destinasi wisata di Indonesia sangat potensial dan diapresiasi wisatawan, tetap diperlukan peningkatan fasilitas, pengelolaan kebersihan, dan perbaikan aksesibilitas untuk meningkatkan kepuasan pengunjung secara menyeluruh.

## **BAB V**

### **PENUTUP**

Bab ini menyajikan hasil penelitian, dengan subbab kesimpulan merangkum temuan utama dan subbab saran memberikan rekomendasi untuk pengembangan penelitian di masa mendatang.

#### **5.1 Kesimpulan**

Berdasarkan hasil penelitian , dapat ditarik kesimpulan bahwa:

1. Penelitian ini berhasil membangun model klasifikasi sentimen ulasan destinasi wisata prioritas di Indonesia berbasis Naive Bayes dengan akurasi rata-rata 93,07%, menunjukkan bahwa hasil klasifikasi dapat diandalkan untuk menggambarkan persepsi pengunjung secara akurat. Dengan demikian, baik dominasi sentimen positif maupun keluhan yang sering muncul mencerminkan opini nyata pengunjung terhadap destinasi wisata.
2. Pandangan pengunjung tergambarkan melalui visualisasi *Wordcloud*, di mana kemunculan kata-kata dominan dalam ulasan negatif, seperti “sampah” pada Danau Toba, menjadi indikator aspek yang perlu diperbaiki dengan adanya peningkatan kebersihan demi kenyamanan wisatawan.

#### **5.2 Saran**

Beberapa rekomendasi yang dapat dijadikan acuan untuk pengembangan penelitian selanjutnya antara lain sebagai berikut:

1. Penelitian selanjutnya dapat mempertimbangkan penggunaan algoritma *machine learning* lain untuk dibandingkan dengan Naïve Bayes, khususnya dalam mengatasi permasalahan data yang tidak seimbang.
2. Penelitian selanjutnya disarankan mengeksplorasi metode augmentasi data lain seperti *Fasttext* ataupun metode lainnya untuk meningkatkan performa model secara lebih efektif.

## DAFTAR PUSTAKA

- A'ayunnisa, N., Salim, Y., & Azis, H. (2022). Analisis performa metode Gaussian Naïve Bayes untuk klasifikasi citra tulisan tangan karakter arab. *Indonesian Journal of Data and Science(IJODAS)*, 3(3), 115-121. doi:<https://doi.org/10.56705/ijodas.v3i3.54>
- Amrullah, A. Z., Anas, A. S., & Hidayat, M. A. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal BITE*, 40-44. doi:<https://doi.org/10.30812/bite.v2i1.804>
- Amrulloh, I. T., Sari, B. N., & Padilah, T. N. (2024). Evaluasi Augmentasi Data Pada Deteksi Penyakit Daun Tebu Dengan YOLOv8. *JATI(Jurnal Mahasiswa Teknik Informatika)*, 7547-7552. doi:<https://doi.org/10.36040/jati.v8i4.10267>
- Aponno, J. C. (2022). Penerapan Algoritma Sentiment Analysis Dan Naïve Bayes Terhadap Opini Pengunjung Di Tempat Wisata Pantai Pintu Kota, Kota Ambon. *Jurnal Teknik Informatika dan Sistem Informasi*, 3180-3188. doi:<https://doi.org/10.35957/jatinsi.v9i4.2697>
- Ardras, D. W., & Voutama, A. (2023). Analisis Sentimen Anti Lgbt Di Indonesiamelalui Media Sosial Twitter. *Jurnal Teknik (Jurnal Fakultas Teknik Universitas Islam Lamongan)*, 23-28. doi:<https://doi.org/10.30736/jt.v15i1.926>
- Badan Pusat Statistik. (2024, Oktober 1). *Badan Pusat Statistik*. Retrieved from Berita Resmi Statistik: <https://www.bps.go.id/id/pressrelease/2024/10/01/2355/kunjungan-wisatawan-mancanegara--wisman--pada-agustus-2024-mencapai-1-34-juta-kunjungan--naik-18-30-persen-year-on-year--y-on-y--.html>
- Buditiawan, K., & Harmono. (2020). Strategi Pengembangan Destinasi Pariwisata Kabupaten Jember. *Jurnal Kebijakan Pembangunan*, 37-50. doi:<https://doi.org/10.47441/jkp.v15i1.50>
- Damayanti, A. P., Widiastuti, I. M., Riandhana, T., & Ashad, M. (2022). Pemberdayaan Pemuda Sadar Wisata Dalam Pengelolaan Kawasan Pantai Bambahano. *Maslahat: Jurnal Pengabdian Kepada Masyarakat*, 3(1), 1-9.
- Firmansyah, D. R., & Lestariningsih, E. (2024). Analisis Sentimen Ulasan Aplikasi Smart Campus Unisbank di Google Playstore Menggunakan Algoritma Naive Bayes. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, 8(2), 498-507. doi:<https://doi.org/10.35870/jtik.v8i2.1882>

- Hanif, A. J., Farid, M. N., & Hasanah, B. (2023). Penerapan Natural Language Processing untuk Klasifikasi Bidang Minat berdasarkan Judul Tugas Akhir. *Jurnal Sistim Informasi dan Teknologi*, 41-49. doi:<https://doi.org/10.37034/jsisfotek.v5i1.196>
- Hardiani, T., & Putri, R. N. (2024). Implementasi Metode Naïve BayesClassifier Untuk Klasifikasi StuntingPada Balita. *Digital Transformation Technology (Digitech)*, 4(1), 621-627. doi:<https://doi.org/10.47709/digitech.v4i1.4481>
- Hidayah, N., & Dodiman. (2024). Implementasi Algoritma Multinomial Naïve Bayes, TF-IDF dan Confusion Matrix dalam Pengklasifikasian Saran Monitoring dan Evaluasi Mahasiswa Terhadap Dosen Teknik Informatika Universitas Dayanu Ikhsanuddin. *JURNAL AKADEMIK PENDIDIKAN MATEMATIKA*, 10(1), 8-15. doi:<https://doi.org/10.55340/japm.v10i1.1491>
- Kemenparekraf/Baparekraf. (2023, 08 10). *Profil Kementerian Pariwisata dan Ekonomi Kreatif/ Badan Pariwisata dan Ekonomi Kreatif*. Retrieved from Kementerian Pariwisata dan Ekonomi Kreatif/Badan Pariwisata dan Ekonomi Kreatif Republik Indonesia: <https://www.kemenparekraf.go.id/profil/profil-lembaga>
- Kemenparekraf/Baparekraf RI. (2020, 10 16). *Mengenal 10 Destinasi Prioritas Pariwisata Indonesia*. Retrieved from Kementerian Pariwisata dan Ekonomi Kreatif/Badan Pariwisata dan Ekonomi Kreatif Republik Indonesia: <https://kemenparekraf.go.id/rumah-difabel/Mengenal-10-Destinasi-Prioritas-Pariwisata-Indonesia>
- Kusuma, D. A., & Wicaksono, A. D. (2023). Analisis Klastering Dampak Lingkungan Berdasarkan Konsumsi Energi Perusahaan Berbasis Industri 4.0 Menggunakan Metode Crisp-Dm. *Positif: Jurnal Sistem dan Teknologi Informasi*, 9(2), 1-6. doi:<https://doi.org/10.31961/positif.v9i2.2050>
- Lasari, T., Kristiawati, E., & Afif, A. (2023). Analisis Kesiapan Penerapan Digitalisasi Akuntansi Usaha Mikro Kecil Sektor Pariwisata Di Kabupaten Mempawah. *Jurnal Akuntansi, Auditing dan Investasi (JAADI)*, 3(1), 1-6.
- Lemantara, J. (2024). Rancang bangun aplikasi hipertensi.edu sebagai media edukasi dan diagnosis penyakit hipertensi menggunakan metode naïve bayes dengan laplace correction. *INFOTECH: Jurnal Informatika & Teknologi*, 146-160. doi:<https://doi.org/10.37373/infotech.v5i1.1197>
- Ma'ady, M. N., Rizaldy, D. D., Satria, R. F., & Anaking, P. (2023). SPARRING: Sistem Rekomendasi Peneliti Terintegrasi Google Scholarvia SerpAPI dan Latent Dirichlet Allocationpada Konteks Perguruan Tinggi. *JTMI : Jurnal Teknologi dan Manajemen Informatika*, 9(2), 161-171. doi:<https://doi.org/10.26905/jtmi.v9i2.11111>
- Mahendra, I. B., Wirawan, I. M., & Sunarya, I. G. (2025, Maret). Comparison of Multinomial, Bernoulli, and Gaussian Naïve Bayes for Complaint

Classification in Pro Denpasar Application. *JUITA: Jurnal Informatika*, 13(1), 77-86. doi:<https://doi.org/10.30595/juita.v13i1.24828>

Makarim, I. A., & Pradana, B. I. (2022). Pengaruh Pemasaran Digital, Word Of Mouth, Dan Online Customer Review Terhadap Minat Kunjungan Wisatawan. *Jurnal Manajemen Pemasaran Dan Perilaku Konsumen*, 1(1), 139-146. doi:<http://dx.doi.org/10.21776/jmppk>

Mukaffi, Z., & Haryanto, T. (2022, Mei). Dampak Sektor Pariwisata Terhadap Pertumbuhan Ekonomi di Kabupaten Banyuwangi. *TOBA (Journal of Tourism, Hospitality and Destination)*, 1(2), 38-43. doi:[10.55123/toba.v1i2.356](https://doi.org/10.55123/toba.v1i2.356)

Murti, E. W. (2024). Analisis Dan Perbandingan Algoritma Prediksi Dalam Mengetahui Perkiraan Peningkatan Jumlah Kasus COVID-19 Di Kabupaten Boyolali Dengan Metodologi CRISP-DM. *JIKES: JURNAL ILMU KESEHATAN*, 3(1), 24-34.

Nugroho, D. S., Hanif, I. F., Hasbi, M. A., Fredianto, Saputra, A. M., & Zildjian, R. (2024, July 3). Analisis Sentimen Dugaan Pelanggaran Pemilu 2024 Berdasarkan Tweet Menggunakan Algoritma Naïve Bayes Classifier. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1169-1176. doi:<https://doi.org/10.57152/malcom.v4i3.1496>

Nurhazizah, E. D., & Puspitasari, I. (2023). Opinion Mining Fungsi Kpi (Key Performance Indikator) Dengan Algoritma Naïve Bayes Classifier Dan Support Vector Machine (Svm). *Jurnal Cahaya Mandalika (JCM)*, 3(2), 290-302. doi:<https://doi.org/10.36312/jcm.v3i2.1534>

Permatasari, I. A., Dermawan, B. A., I. M., & Kurniawan, D. E. (2024). Classification of COVID-19 Aid Recipients in Kasomalang District Using the K-Nearest Neighbor Method. *Journal of Applied Informatics and Computing (JAIC)*, 8(1), 133~139. doi:<https://doi.org/10.30871/jaic.v8i1.3279>

Pratama, A., Nurcahyo, A. C., & Firgia, L. (2023). Penerapan Machine Learning dengan Algoritma Logistik Regresi untuk Memprediksi Diabetes. *SEMINAR NASIONAL CORISINDO*, 116-121.

Rahma, I. A., & Suadaa, L. H. (2023, Desember 6). Penerapan Text Augmentation Untuk Mengatasi Data Yang Tidak Seimbang Pada Klasifikasi Teks Berbahasa Indonesia Studi Kasus: Deteksi Judul Clickbait dan Komentar Hate Speech. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 1329-1340. doi:<https://doi.org/10.25126/jtiik.2023107325>

Ridwansyah, T. (2022). Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier. *KLIK: Kajian Ilmiah*

*Informatika dan Komputer*, 178-185.  
doi:<https://doi.org/10.30865/klik.v2i5.362>

Rofiqi, L., & Akbar, M. (2024). Analisis Sentimen Terkait RUU Perampasan Aset dengan Support Vector Machine. *JEKIN (Jurnal Teknik Informatika)*, 530-538. doi:<https://doi.org/10.58794/jekin.v4i3.824>

Rosid, M. A., Fitrani, A. S., Astutik, I. R., Mulloh, N. I., & Gozali, H. A. (2020). Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi. *IOP Conf. Series: Materials Science and Engineering*, 1-6. doi:[10.1088/1757-899X/874/1/012017](https://doi.org/10.1088/1757-899X/874/1/012017)

Rudini, D., Purnama, D. G., & Khan, A. A. (2023, November). Penggunaantechnik Web Scraping Dalam Aplikasi Pengambilan Data Dari Google Maps Untuk Menunjang Digital Marketing. *Lentera: Multidisciplinary Studies*, 2(1), 10-19. doi:<https://doi.org/10.57096/lentera.v2i1.61>

Ryandi, F. A., Pratiwi, D., & Sari, S. (2025). Analisis Sentimen Masyarakat Di Media Sosial X Terhadap Kemenkes Dengan Naive Bayes dan SVM. *Jurnal Sains dan Teknologi*, 7(1), 1-6. doi:<https://doi.org/10.55338/saintek.v7i1.4615>

Sari, P. K., & Suryono, R. R. (2024). Komparasi Algoritma Support Vector Machine Dan Random Forest Untuk Analisis Sentimen Metaverse. *Jurnal MNEMONIC*, 31-39. doi:<https://doi.org/10.36040/mnemonic.v7i1.8977>

Simangunsong, W. S., & Widyanti, N. W. (2023, 4 18). *64 Persen Orang Indonesia Pilih Wisata Bahari, Ini Pantai Favorit*. Retrieved from KOMPAS: [https://travel.kompas.com/read/2023/04/18/163600927/64-persen-orang-indonesia-pilih-wisata-bahari-ini-pantai-favorit?page=all#google\\_vignette](https://travel.kompas.com/read/2023/04/18/163600927/64-persen-orang-indonesia-pilih-wisata-bahari-ini-pantai-favorit?page=all#google_vignette)

Supoyo, A., & Prasetyaningrum, P. T. (2022). Analisis Data Mining Untuk Memprediksi Lama Perawatan Pasien Covid-19 Di DIY. *Bianglala Informatika*, 10(1), 21-29.

Syahrani, M. (2024, September 19). *Kata Data*. Retrieved from Pemerintah Raup Devisa Pariwisata Rp 113,69 Triliun hingga Juli 2024: <https://katadata.co.id/berita/nasional/66ebf16deaddf/pemerintah-raup-devisa-pariwisata-rp-113-69-triliun-hingga-juli-2024>

Toresa, D., Sitorus, S. R., Muzdalifah, I., Wiza, F., & Syelly, R. (2024, July). Analisis Sentimen Terhadap Ulasan Penggunaan Dompot Digital DANA Menggunakan Metode Klasifikasi Support Vector Machine. *Technologica*, 3(2), 64-74. doi:<https://doi.org/10.55043/technologica.v3i2.163>

Wahyudi, T., Sa'adah, N., & Puspitasari, D. (2023). Penerapan Metode K-Means Pada Data Penjualan Untuk Mendapatkan Produk Terlaris Di PT. Titian Nusantara Boga. *Jurnal Sains dan Teknologi*, 5(1), 228-236. doi:<https://doi.org/10.55338/saintek.v5i1.1379>

Wijiyanto, Pradana, A. I., Sopingi, & Atina, V. (2024, Mei). Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa. *Jurnal Algoritma*, 21(1), 239-248. doi:<https://doi.org/10.33364/algoritma/v.21-1.1618>



UNIVERSITAS  
**Dinamika**