

**ANALISIS SENTIMEN ULASAN PENGUNJUNG MENGENAI OBJEK  
WISATA DI NUSA TENGGARA BARAT MENGGUNAKAN METODE  
SVM (STUDI KASUS: KABUPATEN/KOTA BIMA)**

**SKRIPSI**

Oleh:  
**NURUL FADHILAH**  
**NIM. 200605110023**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2026**

**ANALISIS SENTIMEN ULASAN PENGUNJUNG MENGENAI OBJEK  
WISATA DI NUSA TENGGARA BARAT MENGGUNAKAN METODE  
SVM (STUDI KASUS: KABUPATEN/KOTA BIMA)**

**SKRIPSI**

Diajukan kepada:  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:  
**NURUL FADILAH**  
**NIM. 200605110023**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2026**

## HALAMAN PERSETUJUAN

### ANALISIS SENTIMEN ULASAN PENGUNJUNG MENGENAI OBJEK WISATA DI NUSA TENGGARA BARAT MENGGUNAKAN METODE SVM (STUDI KASUS: KABUPATEN/KOTA BIMA)

#### SKRIPSI

Oleh:

**NURUL FADHILAH**  
**NIM. 200605110023**

Telah Diperiksa dan Disetujui untuk Diuji:  
Tanggal: 15 Juni 2026

Pembimbing I,



Dr. Totok Chamidy, M. Kom  
NIP. 19691222 200604 1 001

Pembimbing II,



Ashri Shabrina Afrah, M.T  
NIP. 19900430 202012 2 003

Mengetahui,

Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono, M.Kom  
NIP. 19841010 201903 1 012

## HALAMAN PENGESAHAN

### ANALISIS SENTIMEN ULASAN PENGUNJUNG MENGENAI OBJEK WISATA DI NUSA TENGGARA BARAT MENGGUNAKAN METODE SVM (STUDI KASUS: KABUPATEN/KOTA BIMA)

#### SKRIPSI

Oleh:  
**NURUL FADHILAH**  
**NIM. 200605110023**

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Sarjana Komputer ( S.Kom )  
Tanggal: 15 Juni 2026

#### Susunan Dewan Penguji

Ketua Penguji : Dr. M. Ainul Yaqin, M.Kom  
NIP. 19761013 200604 1 004

Anggota Penguji I : Fatchurrochman, M.Kom  
NIP. 19700731 200501 1 002

Anggota Penguji II : Dr. Totok Chamidy, M. Kom  
NIP. 19691222 200604 1 001

Anggota Penguji III : Ashri Shabrina Afrah, M.T  
NIP. 19900430 202012 2 003

(  )  
(  )  
(  )  
(  )

Mengetahui dan Mengesahkan,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Supriyono, M.Kom  
NIP. 19841010 201903 1 012

## **PERNYATAAN KEASLIAN TULISAN**

Saya yang bertanda tangan di bawah ini:

Nama : Nurul Fadhilah

NIM : 200605110023

Fakultas / Prodi : Sains dan Teknologi / Teknik Informatika

Judul Skripsi : Analisis Sentimen Ulasan Pengunjung Mengenai  
Objek Wisata Di Nusa Tenggara Barat Menggunakan  
Metode SVM (Studi Kasus: Kabupaten/Kota Bima).

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 15 Juni 2026

Yang membuat pernyataan,



Nurul Fadhilah

NIM. 200605110023

## **MOTTO**

*“Setetes keringat orang tuaku yang keluar, ada seribu langkahku  
untuk maju”*

## **HALAMAN PERSEMBAHAN**

Dengan penuh rasa Syukur ke hadirat Allah SWT atas segala rahmat, taufik, dan hidayah-Nya, Tak lupa, shalawat dan salam selalu tercurah kepada Nabi Muhammad SAW. karya sederhana ini kupersembahkan sebagai wujud terima kasih yang tulus kepada:

Orang Tua Tercinta Bapak Jaelani dan Ibu Sarlah S.Pd, kedua Orang Tua yang hebat selalu menjadi penyemangat bagi penulis dan menjadi sandaran terkuat dari kerasnya dunia. Yang tidak henti-hentinya memberikan kasih sayang dengan penuh cinta selalu memberikan motivasi. Terima kasih untuk semua berkat do'a dan dukungan ibu dan bapak sehingga penulis bisa berada dititik ini. Sehat selalu tolong hiduplah lebih lama lagi.

Untuk diriku sendiri, Nurul Fadhilah telah mampu berusaha dan berjuang sejauh ini. Mampu mengendalikan diri walaupun banyak tekanan dari luar keadaan dan tidak pernah memutuskan untuk menyerah sesulit apapun proses penyusunan skripsi ini.

Kakak tercinta Muhammad Ainul Yakin dan Adikku tercinta Ibnu Majah, terimakasih atas do'a dan dukungannya, yang telah berhasil membawa penulis sampai sejauh ini, sehingga akhirnya mampu menyelesaikan studinya hingga sarjana

Kepada Dosen Pembimbing 1, Bapak Dr. Totok Chamidy M.Kom dan Dosen Pembimbing 2, Ibu Ashri Shabrina Afrah, M.T, serta seluruh Dosen Prodi Teknik Informatika UIN Maulana Malik Ibrahim Malang yang telah membimbing dan mendukung penulis sehingga bisa menyelesaikan skripsi sesuai target.



Teman – teman seperjuangan, lida, eka, upi, putri, adawia, jumi, yanti, isna dan ayu yang selalu kebersamai dalam proses skripsi ini, saling menguatkan, dan berbagi tawa serta air mata. Terima kasih untuk cerita, pelukan hangat, dan semangat yang tak pernah padam.

Untuk seseorang yang belum bisa kutulis dengan jelas namanya disini, namun sudah tertulis jelas di *Lauhul Mahfudz* untukku. Terima kasih sudah menjadi salah satu sumber motivasi penulis dalam menyelesaikan skripsi ini sebagai salah satu bentuk penulis dalam memantaskan diri. Meskipun saat ini penulis tidak tahu keberadaanmu entah di bumi bagian mana dan menggenggam tangan siapa. Seperti kata Bj Habibie “Kalau memang dia dilahirkan untuk saya, kamu jungkir balik pun saya yang dapat”.



## KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT atas segala rahmat, taufik, serta hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Ulasan Pengunjung Mengenai Objek Wisata di Nusa Tenggara Barat Menggunakan Metode *SVM* (Studi Kasus: Kabupaten/Kota Bima)”. Sholawat serta salam semoga tetap tercurahkan kepada junjungan kita Nabi Besar Muhammad SAW yang telah membawa kita dari zaman kegelapan menuju zaman yang terang benderang.

Penyusunan skripsi ini dimaksudkan untuk memenuhi salah satu syarat dalam memperoleh gelar Sarjana Komputer (S.Kom) pada Program Studi Teknik Informatika di Universitas Islam Negeri Maulana Malik Ibrahim. Penulis menyadari bahwa selesainya skripsi ini tidak terlepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj Ilfi Nur Diana, M.Si selaku Rektor UIN Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono selaku Dekan Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.
3. Supriyono, M.Kom, selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Totok Chamidy M.Kom selaku Dosen Pembimbing I dan Ibu Ashri Shabrina Afrah, M.T selaku Dosen Pembimbing II, yang dengan sabar meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan,

masukan, dan koreksi kepada penulis dalam setiap tahap penyusunan skripsi, mulai dari penyusunan proposal hingga penyelesaian akhir.

5. Bapak Dr. M. Ainul Yaqin, M.Kom selaku dosen Penguji I dan Bapak Fatchurrochman dosen Penguji II yang telah memberikan kritik, saran, serta masukan yang sangat membangun demi kesempurnaan skripsi ini.
6. Segenap Dosen dan Staf Administrasi Program Studi Teknik Informatika yang telah memberikan bekal ilmu dan bantuan selama masa perkuliahan.
7. Kedua orang tua tercinta dan keluarga besar yang selalu memberikan doa restu, dukungan moral maupun material, serta motivasi yang tidak terhingga.
8. Rekan-rekan mahasiswa Teknik Informatika angkatan 2020 dan semua pihak yang telah membantu dalam proses penelitian maupun penyusunan dokumen ini.
9. Serta semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah membantu dalam bentuk apa pun, baik secara langsung maupun tidak langsung.

## DAFTAR ISI

|                                                                     |                                     |
|---------------------------------------------------------------------|-------------------------------------|
| <b>HALAMAN PERSETUJUAN .....</b>                                    | <b>ii</b>                           |
| <b>HALAMAN PENGESAHAN.....</b>                                      | <b>Error! Bookmark not defined.</b> |
| <b>PERNYATAAN KEASLIAN TULISAN .....</b>                            | <b>vi</b>                           |
| <b>MOTTO .....</b>                                                  | <b>vi</b>                           |
| <b>HALAMAN PERSEMBAHAN .....</b>                                    | <b>vii</b>                          |
| <b>KATA PENGANTAR.....</b>                                          | <b>ix</b>                           |
| <b>DAFTAR ISI.....</b>                                              | <b>xi</b>                           |
| <b>DAFTAR TABEL .....</b>                                           | <b>xiii</b>                         |
| <b>DAFTAR GAMBAR.....</b>                                           | <b>xiv</b>                          |
| <b>ABSTRAK .....</b>                                                | <b>xv</b>                           |
| <b>ABSTRACT .....</b>                                               | <b>xvi</b>                          |
| <b>ملخص البحث .....</b>                                             | <b>xvii</b>                         |
| <b>BAB I PENDAHULUAN.....</b>                                       | <b>2</b>                            |
| 1.1 Latar Belakang .....                                            | 2                                   |
| 1.2 Rumusan Masalah .....                                           | 7                                   |
| 1.3 Batasan Masalah.....                                            | 7                                   |
| 1.4 Tujuan Penelitian .....                                         | 8                                   |
| 1.5 Manfaat Penelitian .....                                        | 8                                   |
| <b>BAB II STUDI PUSTAKA .....</b>                                   | <b>10</b>                           |
| 2.1 Penelitian Terkait .....                                        | 10                                  |
| 2.2 Perbandingan Penelitian Terdahulu dan Kebaruan Penelitian ..... | 15                                  |
| 2.3 Pariwisata di Kabupaten/Kota Bima .....                         | 15                                  |
| 2.4 Analisis Sentimen .....                                         | 16                                  |
| 2.4.1 Perhitungan Bobot TF-IDF .....                                | 17                                  |
| 2.5 Preprocessing Teks.....                                         | 18                                  |
| 2.5.1 <i>Cleaning</i> .....                                         | 18                                  |
| 2.5.2 <i>Case folding</i> .....                                     | 19                                  |
| 2.5.3 <i>Tokenizing</i> .....                                       | 19                                  |
| 2.5.4 <i>Stopword Removal</i> .....                                 | 19                                  |
| 2.6 <i>Support Vector Machine (SVM)</i> .....                       | 20                                  |
| 2.7 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i> ..... | 21                                  |
| 2.8 <i>Latent Dirichlet Allocation (LDA)</i> .....                  | 22                                  |
| 2.9 <i>Web Scraping</i> .....                                       | 23                                  |
| <b>BAB III METODE PENELITIAN .....</b>                              | <b>24</b>                           |
| 3.1 Pengumpulan Data .....                                          | 24                                  |
| 3.2 Ekstraksi Data.....                                             | 25                                  |
| 3.3 Rancangan Sistem .....                                          | 26                                  |
| 3.3.1 Output Sistem dan Cara Penggunaan .....                       | 27                                  |
| 3.4 <i>Preprocessing Text</i> .....                                 | 28                                  |
| 3.4.1 <i>Cleaning</i> .....                                         | 28                                  |
| 3.4.2 <i>Case folding</i> .....                                     | 28                                  |
| 3.4.3 <i>Tokenizing</i> .....                                       | 29                                  |
| 3.4.4 <i>Stopword Removal</i> .....                                 | 29                                  |

|                                                                                      |           |
|--------------------------------------------------------------------------------------|-----------|
| 3.4.5 Stemming.....                                                                  | 29        |
| 3.5 Fitur Pembobotan .....                                                           | 32        |
| 3.5.1 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....                | 32        |
| 3.5.2 Normalisasi Vektor <i>TF-IDF</i> .....                                         | 33        |
| 3.6 Pembagian Data dan Penerapan SMOTE .....                                         | 34        |
| 3.6.1 Pembagian Data .....                                                           | 34        |
| 3.6.2 Penerapan SMOTE .....                                                          | 34        |
| 3.7 Implementasi <i>Support Vector Machine</i> .....                                 | 35        |
| 3.8 Implementasi <i>Latent Dirichlet Allocation (LDA)</i> .....                      | 42        |
| 3.8.1 Input Data LDA .....                                                           | 42        |
| 3.8.2 Parameter LDA .....                                                            | 43        |
| 3.9 Evaluasi Model.....                                                              | 43        |
| <b>BAB IV HASIL DAN PEMBAHASAN .....</b>                                             | <b>45</b> |
| 4.1 Skenario Uji Coba .....                                                          | 45        |
| 4.1.1 Kegunaan dan Cara Kerja Masing-Masing Metode .....                             | 46        |
| 4.2 Pengumpulan Data .....                                                           | 48        |
| 4.3 Hasil Analisis Data.....                                                         | 50        |
| 4.3.1 Hasil <i>Preprocessing</i> Data pada Wisata Pantai Lariti.....                 | 50        |
| 4.3.2 Hasil Pengujian Model SVM.....                                                 | 51        |
| 4.3.3 Hasil Klasifikasi SVM .....                                                    | 52        |
| 4.3.4 Analisis Topik Keluhan Menggunakan LDA .....                                   | 55        |
| 4.4 Pembahasan.....                                                                  | 57        |
| 4.4.1 Efektivitas SMOTE dalam Menangani Ketidakseimbangan Data .....                 | 57        |
| 4.4.2 Performa Klasifikasi SVM dan Perbandingan<br>dengan Penelitian Terdahulu ..... | 57        |
| 4.4.3 Hasil Pemodelan Topik LDA dan Implikasinya.....                                | 58        |
| 4.5 Ringkasan Temuan.....                                                            | 61        |
| 4.6 Integrasi Islam.....                                                             | 63        |
| <b>BAB V KESIMPULAN DAN SARAN .....</b>                                              | <b>65</b> |
| 5.1 Kesimpulan .....                                                                 | 65        |
| 5.2 Saran .....                                                                      | 67        |
| 5.2.1 Saran Praktis .....                                                            | 67        |
| 5.2.2 Saran untuk Penelitian Selanjutnya .....                                       | 69        |
| <b>DAFTAR PUSTAKA</b>                                                                |           |

## DAFTAR TABEL

|                                                                 |    |
|-----------------------------------------------------------------|----|
| Tabel 2.1 Perbandingan Penelitian.....                          | 14 |
| Tabel 2.2 Perbandingan Penelitian.....                          | 15 |
| Tabel 3.1 Penyajian Data .....                                  | 24 |
| Tabel 3.2 Data <i>Testing</i> .....                             | 25 |
| Tabel 3.3 Data <i>Training</i> .....                            | 25 |
| Tabel 3.4 Contoh <i>Cleaning</i> .....                          | 28 |
| Tabel 3.5 Contoh <i>Case Folding</i> .....                      | 29 |
| Tabel 3.6 Contoh <i>Tokenizing</i> .....                        | 29 |
| Tabel 3.7 Contoh <i>Stopword Removal</i> .....                  | 29 |
| Tabel 3. 8 Contoh <i>Stemming</i> .....                         | 30 |
| Tabel 3.9 Pelabelan Data.....                                   | 31 |
| Tabel 3.10 Contoh Dokumen Ulasan Untuk Perhitungan TF-IDF ..... | 32 |
| Tabel 3.11 Skenario Pembagian Data Latih dan Data Uji .....     | 34 |
| Tabel 3.12 Data Vektor hasil TF-IDF .....                       | 36 |
| Tabel 3.13 Hasil Perhitungan Nilai X dengan kernel .....        | 37 |
| Tabel 3.14 Hasil Keseluruhan <i>Support Vector</i> .....        | 40 |
| Tabel 3.15 Nilai Bias pada <i>Support Vector</i> .....          | 40 |
| Tabel 4.1 Distribusi Sentimen Dataset per Destinasi Wisata..... | 45 |
| Tabel 4.2 Perbandingan Performa Model SVM.....                  | 52 |

## DAFTAR GAMBAR

|                                                                                         |    |
|-----------------------------------------------------------------------------------------|----|
| Gambar 2.1 Ilustrasi Hyperplane dan SVM.....                                            | 21 |
| Gambar 2.2 Ilustrasi Proses Sintesis Sampel pada SMOTE .....                            | 21 |
| Gambar 2.3 Model Grafis LDA (Blei et al., 2003) .....                                   | 22 |
| Gambar 3.1 Diagram Alur Rancang Sistem.....                                             | 26 |
| Gambar 3.2 Tahapan <i>Preprocessing Text</i> .....                                      | 28 |
| Gambar 3.3 Distribusi Kelas Sebelum dan Sesudah Penerapan<br>SMOTE pada Data Latih..... | 35 |
| Gambar 3.4 Grafik <i>Coherence Score</i> untuk Penentuan Jumlah Topik Optimal ...       | 43 |
| Gambar 4.1 Wisata Pantai Lariti.....                                                    | 48 |
| Gambar 4.2 Wisata Gunung Tambora .....                                                  | 49 |
| Gambar 4.3 Wisata Pantai Lawata .....                                                   | 49 |
| Gambar 4.4 Proses Scraping .....                                                        | 49 |
| Gambar 4.5 Hasil Case Folding .....                                                     | 50 |
| Gambar 4. 6 Hasil Cleaning .....                                                        | 50 |
| Gambar 4. 7 Hasil Stopwords Removal.....                                                | 51 |
| Gambar 4. 8 Hasil Stemming .....                                                        | 51 |
| Gambar 4.9 WordCloud Sentimen Negatif.....                                              | 56 |

## ABSTRAK

Fadhilah, Nurul. 2026. **Analisis Sentimen Ulasan Pengunjung Mengenai Objek Wisata Di Nusa Tenggara Barat Menggunakan Metode SVM (Studi kasus: Kabupaten/Kota Bima)**. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Totok Chamidy, M.Kom, (II) Ashri Shabrina Afrah, M.T.

**Kata Kunci:** Analisis Sentimen, SVM, SMOTE, LDA, Pariwisata Bima.

Kabupaten/Kota Bima memiliki kekayaan alam yang melimpah di sektor pariwisata, namun pemanfaatan data untuk strategi pengembangan dan eksposur masih belum optimal. Studi ini mengevaluasi persepsi wisatawan terhadap Pantai Lariti, Pantai Lawata, dan Gunung Tambora melalui analisis sentimen pada ulasan *Google Maps*. Pendekatan yang digunakan mengombinasikan algoritma *Support Vector Machine* (SVM) untuk klasifikasi, *Synthetic Minority Over-sampling Technique* (SMOTE) guna mengatasi ketidakseimbangan data, serta *Latent Dirichlet Allocation* (LDA) untuk memetakan topik pada ulasan negatif. Proses penelitian diawali dengan *web scraping* data selama satu dekade terakhir, yang kemudian diproses melalui tahapan *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Pembobotan kata dilakukan dengan *Term Frequency-Inverse Document Frequency* (TF-IDF). Pengujian performa model menggunakan tiga variasi skenario pembagian data (90:10, 80:20, dan 70:30), di mana SVM dengan *kernel linear* pada rasio 90:10 memberikan hasil terbaik dengan akurasi 84,91%. Penggunaan SMOTE terbukti mampu mendongkrak sensitivitas model dalam mengidentifikasi sentimen negatif sebagai kelas minoritas. Analisis topik melalui LDA mengungkap bahwa ketidakpuasan pengunjung didominasi oleh isu kebersihan (pengelolaan sampah) dan aksesibilitas jalan menuju lokasi wisata. Secara manajerial, hasil ini merekomendasikan pemerintah daerah serta pengelola destinasi untuk fokus pada revitalisasi infrastruktur fisik dan peningkatan standar layanan kebersihan demi memperkuat daya saing pariwisata di wilayah Bima.



## ABSTRACT

Fadhilah, Nurul. 2026. **Sentiment Analysis of Visitor Reviews Regarding Tourist Attractions in West Nusa Tenggara Using the SVM Method (Case Study: Bima Regency/City)**. Thesis. Department of Informatics Engineering, Faculty of Science and Technology, State Islamic University of Maulana Malik Ibrahim Malang. Supervisor: (I) Dr. Totok Chamidy, M.Kom, (II) Ashri Shabrina Afrah, M.T.

**Keywords:** Sentiment Analysis, SVM, SMOTE, LDA, Bima Tourism.

Bima Regency/City boasts abundant natural resources in the tourism sector; however, data utilization for development and exposure strategies remains suboptimal. This study evaluates tourist perceptions of Lariti Beach, Lawata Beach, and Mount Tambora through sentiment analysis of reviews from Google Maps. The approach combines the Support Vector Machine (SVM) algorithm for classification, the Synthetic Minority Over-sampling Technique (SMOTE) to address data imbalance, and Latent Dirichlet Allocation (LDA) to map topics within negative reviews. The research process began with web scraping data from the past decade, which was then processed through stages of cleaning, case folding, tokenizing, stopword removal, and stemming. Word weighting was performed using Term Frequency-Inverse Document Frequency (TF-IDF). Model performance testing involved three variations of data-sharing scenarios (90:10, 80:20, and 70:30), where SVM with a linear kernel at a 90:10 ratio yielded the best results with an accuracy of 84.91%. The implementation of SMOTE was proven to increase the model's sensitivity in identifying negative sentiment as a minority class. Topic analysis using LDA revealed that visitor dissatisfaction was dominated by issues of cleanliness (waste management) and accessibility to tourist sites. From a managerial perspective, these results recommend that local governments and destination managers focus on revitalizing physical infrastructure and improving cleanliness service standards to strengthen tourism competitiveness in the Bima region.

## ملخص البحث

فضيلة، نورول. 2026. تحليل المشاعر في تقييمات الزوار المتعلقة بالمعالم السياحية في غرب نوسا تينجارا باستخدام طريقة SVM (دراسة حالة: مقاطعة/مدينة بيما). رسالة ماجستير. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية في مالانج. المشرفون: (1) د. توتوك شاميدي، ماجستير علوم الحاسوب، (2) أشري شابرينا أفراح، ماجستير هندسة.

الكلمات المفتاحية: تحليل المشاعر، SVM، SMOTE، LDA، السياحة في بيما.

تزخر مقاطعة/مدينة بيما بموارد طبيعية وفيرة في قطاع السياحة، إلا أن استخدام البيانات في استراتيجيات التطوير والترويج لا يزال دون المستوى الأمثل. تقيم هذه الدراسة تصورات السياح عن شاطئ لاريتي، وشاطئ لاواتا، وجبل تامبورا من خلال تحليل المشاعر في التقييمات. خرائط جوجليجمع النهج المستخدم بين الخوارزميات المتجهات الداعمة (SVM) للتصنيف، تقنية أخذ العينات الزائدة للأقلية الاصطناعية (SMOTE) لمعالجة عدم توازن البيانات، بالإضافة إلى تخصيص ديريشلي الكامن (LDA) لرسم خريطة للمواضيع المتعلقة بالتقييمات السلبية. تبدأ عملية البحث باستخراج البيانات من مواقع الويب بالبيانات على مدى العقد الماضي، والتي تمت معالجتها بعد ذلك عبر مراحل لتنظيف، طي الحالات، تقسيم الكلمات إلى رموز، إزالة الكلمات المحظورة، والتجريد يتم تحديد وزن الكلمات بواسطة تردد المصطلح - تردد المستند العكسي (TF-IDF) اختبار أداء النموذج باستخدام ثلاثة سيناريوهات مختلفة لمشاركة البيانات (90:10، 80:20، و70:30)، حيث SVM معنوة خطية حققت نسبة 90:10 أفضل النتائج بدقة بلغت 84.91%. وقد ثبت أن استخدام تقنية SMOTE يزيد من حساسية النموذج في تحديد المشاعر السلبية كفاءة أقلية. وكشف تحليل المواضيع باستخدام خوارزمية LDA أن استياء الزوار يتركز بشكل أساسي على قضايا النظافة (إدارة النفايات) وسهولة الوصول إلى المواقع السياحية. ومن منظور إداري، توصي هذه النتائج الحكومات المحلية ومديري الوجهات السياحية بالتركيز على إعادة تنشيط البنية التحتية المادية وتحسين معايير خدمات النظافة لتعزيز القدرة التنافسية للسياحة في منطقة بيما.

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pariwisata adalah salah satu sektor yang memiliki kontribusi signifikan terhadap perekonomian global. Berdasarkan laporan dari Organisasi Pariwisata Dunia (UNWTO) tahun 2024, sektor pariwisata global diproyeksikan tumbuh sekitar 4-5% per tahun, dengan lebih dari 1,4 miliar wisatawan internasional pada tahun 2023 (Ekstrom, 2024). Di ranah domestik, industri ini menjadi pilar krusial bagi ekonomi daerah di Indonesia melalui peningkatan kontribusi terhadap Produk Domestik Bruto (PDB) yang konsisten. Secara khusus, Pemerintah dari Provinsi Nusa Tenggara Barat (NTB) sendiri telah menjadikan pariwisata sebagai motor utama percepatan pembangunan ekonomi wilayah. Hal ini dibuktikan dengan lonjakan kunjungan wisatawan ke NTB pada tahun 2024 yang mencakup berbagai destinasi, mulai dari kekayaan alam dan budaya hingga situs-situs sejarah.

Data dari Badan Pusat Statistik NTB mencatat bahwa sektor pariwisata di NTB berhasil menyumbang hampir 6% terhadap PDB provinsi pada tahun 2024, mencerminkan tren positif yang terus berkembang (BPS NTB, 2024). Namun demikian, meskipun ada peningkatan signifikan dalam sektor pariwisata di NTB, tidak semua daerah dan objek wisata di NTB mendapatkan perhatian yang sama. Salah satu contohnya adalah Kabupaten/Kota Bima, yang terletak di bagian timur NTB dan memiliki potensi wisata alam yang sangat besar, namun belum sepenuhnya terekspose atau berkembang secara maksimal. Berdasarkan informasi

yang telah resmi diterbitkan oleh Dinas Pariwisata Kabupaten Bima (Disparda), terdata sebanyak 106 objek wisata yang terletak di berbagai wilayah dalam kabupaten (Fajar & Ifantri Ifantri, 2021). Angka tersebut mencakup destinasi-destinasi yang berada dalam pengelolaan pemerintah daerah maupun yang dikelola secara mandiri oleh sektor swasta dan masyarakat sekitar. Sedangkan untuk wilayah Kota Bima, data resmi yang sering dirujuk dalam beberapa laporan statistik daerah (BPS Kota Bima) yang bersumber dari dinas pariwisata dan kebudayaan setempat menyatakan sekitar 32 objek wisata (Rasyad & Fajar, 2021). Destinasi ini mencakup berbagai kategori, mulai dari wisata alam, bahari, hingga sejarah dan budaya. Di antara persebaran data tersebut, objek wisata seperti Pantai Lariti, Pantai Lawata, dan Gunung Tambora menjadi destinasi unggulan yang paling banyak menarik minat wisatawan. Banyaknya jumlah objek wisata ini menuntut adanya pengelolaan dan evaluasi kualitas layanan yang lebih mendalam guna mempertahankan citra positif pariwisata daerah. Kondisi tersebut dapat mengakibatkan ketidakseimbangan dalam jumlah kunjungan wisatawan di daerah tersebut, padahal potensinya dapat menjadi sumber pendapatan yang lebih signifikan bagi masyarakat di sekitar objek wisata. Oleh karena itu, sangatlah krusial untuk memahami persepsi serta pandangan para wisatawan mengenai berbagai objek wisata di Kabupaten/Kota Bima. Pemahaman tersebut sangat diperlukan agar pihak pengelola maupun pemangku kepentingan lainnya mampu merumuskan strategi pengembangan destinasi yang lebih efektif dan tepat sasaran di kawasan tersebut.

Salah satu cara untuk memahami persepsi pengunjung terhadap objek wisata adalah dengan menganalisis ulasan-ulasan yang mereka tinggalkan di platform online seperti *Google Maps* dan *TripAdvisor*. Ulasan tersebut dapat memberikan wawasan berharga mengenai kepuasan pengunjung, kelebihan dan kekurangan dari objek wisata, serta faktor-faktor yang mempengaruhi keputusan mereka untuk berkunjung (Ariansyah et al., 2025). Namun, pengelolaan dan analisis data ulasan pengunjung secara manual sangat memakan waktu dan sumber daya. Oleh sebab itu, implementasi teknologi analisis sentimen muncul sebagai solusi yang efisien untuk membedah data teks tersebut. Melalui pemanfaatan algoritma *Support Vector Machine* (SVM) dalam metode analisis sentimen, ulasan pengunjung dapat diklasifikasikan ke dalam kategori positif, negatif, atau netral secara otomatis. Pendekatan ini memungkinkan perolehan gambaran yang lebih komprehensif mengenai persepsi serta perasaan wisatawan terhadap suatu objek wisata (Syahlan et al., 2023).

Sejumlah studi terdahulu telah memberikan perhatian yang signifikan pada analisis sentimen terhadap berbagai objek wisata populer, khususnya di wilayah Bali dan Yogyakarta (Subarkah et al., 2022; Utama, 2023), namun masih sangat terbatas yang meneliti objek wisata di daerah yang kurang populer, seperti Kabupaten/Kota Bima. Hal ini menciptakan kesenjangan dalam penelitian yang ada, di mana destinasi wisata yang kurang populer dan memiliki potensi besar seperti Kabupaten/Kota Bima belum mendapatkan perhatian yang cukup berdasarkan perspektif analisis sentimen. Guna mengatasi keterbatasan informasi tersebut, penelitian ini menerapkan teknik analisis sentimen melalui metode

*Support Vector Machine* (SVM). Pendekatan ini difokuskan pada pengolahan ulasan pengunjung untuk mengevaluasi berbagai destinasi wisata yang terletak di wilayah Kabupaten serta Kota Bima, Nusa Tenggara Barat.

Analisis sentimen, sebagai salah satu komponen dalam pemrosesan bahasa alami, berfungsi untuk mengenali, mengekstraksi, serta mengukur data subjektif yang terdapat dalam sebuah teks. Teknik ini diterapkan untuk mengidentifikasi dan menguantifikasi informasi subjektif guna memahami kandungan emosional atau opini dalam materi tertulis tersebut. Di sektor pariwisata, metode ini sangat bermanfaat dalam membedah emosi, opini, dan perilaku wisatawan terkait beragam fasilitas di destinasi liburan. *Support Vector Machine* (SVM) telah terbukti menjadi metode yang sangat efisien untuk klasifikasi teks di antara berbagai pendekatan analisis yang ada. Hal ini dikarenakan kemampuannya dalam menjaga stabilitas pada ruang fitur berdimensi tinggi serta efektivitasnya dalam menekan risiko *overfitting*. Seperti ditunjukkan oleh Syahlan et al (2023) dalam penelitiannya mengenai analisis sentimen di taman air mancur Sri Baduga Purwakarta, algoritme SVM berhasil menunjukkan performa dengan tingkat akurasi mencapai 94%. Sedangkan penelitian lain yang dilakukan oleh Tri et al (2022) memperoleh hasil yang baik dalam analisis sentimen objek wisata Danau Toba dengan tingkat Akurasi 83%.

Namun, klasifikasi sentimen semata hanya mampu mengidentifikasi polaritas ulasan tanpa menjelaskan aspek spesifik yang menjadi sumber keluhan. Penelitian ini menerapkan algoritma *Latent Dirichlet Allocation* (LDA) guna menutup celah penelitian yang ada melalui pemodelan topik pada ulasan dengan

sentimen negatif. Dengan kemampuan LDA dalam mengekstraksi berbagai tema laten dari sekumpulan teks, keluhan para pengunjung dapat dipetakan menjadi topik-topik yang lebih konkret, seperti kualitas fasilitas, kebersihan, maupun kondisi infrastruktur, sehingga lebih mudah untuk ditindaklanjuti.

Dalam analisis sentimen ulasan pariwisata, ketidakseimbangan kelas data merupakan tantangan yang lazim muncul, mengingat jumlah ulasan positif sering kali mendominasi ulasan negatif. Fenomena ini berisiko menimbulkan bias pada model klasifikasi. Oleh karena itu, penelitian ini menerapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) pada dataset pelatihan guna menyelaraskan distribusi antar kelas sebelum proses klasifikasi dilakukan.

Penelitian ini mengajukan sebuah kerangka analisis sentimen dengan studi kasus ulasan pengunjung objek wisata di Kabupaten/Kota Bima. Kerangka tersebut mengintegrasikan SVM guna mengklasifikasikan sentimen, SMOTE untuk mengatasi persoalan ketidakseimbangan data, serta LDA untuk memodelkan berbagai topik keluhan. Diharapkan hasil dari penelitian ini mampu menyajikan informasi berbasis data yang bermanfaat bagi pemerintah daerah dan pengelola wisata dalam menyusun strategi pengembangan pariwisata secara lebih efektif dan terarah.

menyimak serta memahami pandangan maupun kontribusi pemikiran dari sesama turut digarisbawahi dalam Al-Qur'an sebagai perwujudan nilai-nilai Islam. Hal ini sebagaimana firman Allah SWT yang terkandung dalam Surah Al-A'raf ayat 204:

وَإِذَا قُرِئَ الْقُرْآنُ فَاسْتَمِعُوا لَهُ ۖ وَأَنْصِتُوا لَعَلَّكُمْ تُرْحَمُونَ

*“Dan apabila dibacakan Al-Qur'an, maka dengarkanlah baik-baik, dan perhatikanlah dengan tenang agar kamu mendapat rahmat.” (Q.S Al A'rof: 204).*

Pentingnya menyimak dengan penuh perhatian serta bersikap tenang ketika Al-Qur'an dilantunkan digarisbawahi dalam ayat ini sebagai jalan untuk mendapatkan rahmat. Prinsip untuk menyimak serta memahami dengan saksama dapat diterapkan pada beragam dimensi kehidupan, seperti saat menerima saran atau tinjauan dari pihak lain, sekalipun konteks asli ayat tersebut merujuk pada aktivitas mendengarkan lantunan Al-Qur'an. Dengan mendengarkan dan memahami ulasan pengunjung, pengelola destinasi wisata dapat meningkatkan kualitas layanan dan fasilitas yang disediakan.

Melalui penerapan analisis sentimen pada ulasan wisatawan, studi ini diproyeksikan mampu menyumbangkan manfaat praktis bagi strategi kemajuan sektor pariwisata di Kabupaten/Kota Bima. Temuan dari penelitian ini tidak hanya menyajikan data yang lebih imparial bagi para pengelola destinasi, namun juga dapat dijadikan instrumen dasar dalam penyusunan kebijakan yang berbasis bukti dan lebih fokus guna mengoptimalkan potensi daya tarik wisata di wilayah Kabupaten/Kota Bima. Salah satu contoh kontribusinya adalah dengan mengidentifikasi aspek-aspek objek wisata yang perlu diperbaiki berdasarkan ulasan negatif yang sering muncul, sehingga dapat meningkatkan kualitas pelayanan dan fasilitas bagi wisatawan.



## 1.2 Rumusan Masalah

- a. Seberapa akurat penerapan metode SVM dalam menganalisis sentimen guna mengidentifikasi kelebihan dan kekurangan dari ulasan pengunjung sebagai dasar perumusan strategi pengembangan pariwisata di Kabupaten/Kota Bima?
- b. Topik keluhan apa saja yang paling dominan muncul dari ulasan bersentimen negatif pada masing-masing objek wisata berdasarkan pemodelan topik LDA?
- c. Bagaimana penerapan SMOTE berkontribusi dalam meningkatkan kemampuan model SVM dalam mendeteksi sentimen negatif sebagai kelas minoritas?

## 1.3 Batasan Masalah

Batasan masalah yang diterapkan dalam penelitian ini adalah sebagai berikut:

- a. Penelitian menggunakan Tableau sebagai tools visualisasi data.
- b. Data bersumber dari ulasan pengunjung pada *Google Maps*, dengan data terbatas pada statistik kunjungan wisatawan 10 tahun terakhir.
- c. Mengakuisisi data ulasan menggunakan teknik web Scraping.
- d. Klasifikasi sentiment menggunakan Algoritma SVM dengan metode TF-IDF sebagai *Term Weighting*.
- e. Pemodelan topik menggunakan algoritma LDA yang diterapkan khusus pada ulasan bersentimen negatif.

#### 1.4 Tujuan Penelitian

- a. Mengukur tingkat Akurasi model SVM dalam mengklasifikasikan sentimen ulasan pengunjung pada objek wisata di Kabupaten/Kota Bima.
- b. Mengidentifikasi topik-topik keluhan utama yang terkandung dalam ulasan bersentimen negatif menggunakan algoritma LDA.
- c. Mengevaluasi pengaruh penerapan SMOTE terhadap performa model SVM dalam mendeteksi kelas sentimen minoritas.
- d. Menyusun rekomendasi pengembangan pariwisata Kabupaten/Kota Bima berbasis temuan analisis data.

#### 1.5 Manfaat Penelitian

- a. Manfaat Teoritis

Penelitian ini berkontribusi pada pengembangan kajian analisis sentimen berbahasa Indonesia, khususnya dalam konteks pariwisata daerah. Integrasi SVM, SMOTE, dan LDA dalam satu kerangka analisis memperkaya literatur mengenai pendekatan hibrida untuk pemrosesan teks ulasan yang tidak seimbang. Di samping itu, bagi penelitian selanjutnya yang akan melakukan implementasi metode sejenis di berbagai objek wisata lain di wilayah Indonesia, penelitian ini pun dapat dijadikan sebagai acuan.

- b. Manfaat Praktis

Hasil penelitian ini memberikan dasar yang kuat bagi pengelola objek wisata serta Dinas Pariwisata Kabupaten/Kota Bima untuk mengambil keputusan strategis dalam memprioritaskan peningkatan layanan dan perbaikan fasilitas, yang didukung oleh data keluhan pengunjung yang akurat.

Dashboard visualisasi yang dihasilkan melalui Tableau juga dapat dimanfaatkan sebagai alat pemantauan persepsi wisatawan secara berkelanjutan, sehingga mendukung pengelolaan pariwisata yang lebih responsif dan berbasis bukti.

Secara lebih terperinci, sistem analisis sentimen ini dirancang untuk dimanfaatkan oleh tiga kelompok pengguna utama. Pertama, Dinas Pariwisata Kabupaten/Kota Bima selaku pemerintah daerah sebagai pemangku kepentingan utama, yang dapat menggunakan hasil analisis ini sebagai dasar pengambilan kebijakan berbasis data (*evidence-based policy*), khususnya dalam mengalokasikan anggaran perbaikan infrastruktur dan kebersihan destinasi wisata. Kedua, pengelola objek wisata dan Kelompok Sadar Wisata (Pokdarwis) yang dapat memanfaatkan output topik keluhan LDA sebagai panduan operasional harian, misalnya dalam membentuk tim satgas kebersihan di titik-titik yang sering dikeluhkan pengunjung. Ketiga, peneliti dan akademisi yang dapat menggunakan kerangka metodologi SVM, SMOTE, dan LDA sebagai referensi untuk penelitian serupa pada destinasi wisata lain di Indonesia.

## **BAB II**

### **STUDI PUSTAKA**

#### **2.1 Penelitian Terkait**

Penelitian lain bisa dijadikan referensi dan juga memberikan panduan yang membantu penulis bisa menyelesaikan penelitian ini. Dalam beberapa tahun terakhir, penelitian mengenai analisis sentimen di sektor pariwisata telah mengalami kemajuan signifikan melalui beragam metodologi untuk mengategorikan ulasan pengunjung di berbagai destinasi wisata. Berbagai penelitian terdahulu telah menawarkan perspektif penting terkait penggunaan algoritma machine learning guna menginterpretasikan pandangan wisatawan terhadap beragam elemen di objek wisata tersebut. Salah satu penelitian yang relevan yaitu oleh (Syahlan et al., 2023) penelitian ini menerapkan algoritma *Support Vector Machine* (SVM) untuk melakukan analisis sentimen pada objek wisata Taman Air Mancur Sri Baduga Purwakarta. Data ulasan yang dikumpulkan melalui platform *Google Maps* diproses menggunakan metode SVM guna mengelompokkan pandangan pengunjung ke dalam tiga kategori utama, yaitu sentimen positif, netral, dan negatif. Hasilnya memperoleh hasil Akurasi yang sangat baik, mencapai 94%, dengan distribusi sentimen yang terbagi menjadi 872 kata positif, 156 kata negatif, dan 127 kata netral. Penelitian ini menegaskan bahwa penggunaan SVM dalam analisis sentimen untuk pariwisata memberikan hasil yang dapat diandalkan dan berguna dalam meningkatkan kualitas pengelolaan destinasi wisata.

Selain itu, penelitian yang dilakukan oleh (Tri et al., 2022) yang menyatakan penelitian lain turut memanfaatkan metode SVM guna melakukan analisis sentimen terkait objek wisata Danau Toba, dengan perolehan tingkat Akurasi sebesar 83%. Hasil evaluasi tambahan menunjukkan nilai *Precision*, *Recall*, serta *f-measure* yang konsisten di angka 84%, 83%, dan 83%. Dalam studi ini, diuji dua varian kernel *SVM*, yakni linear dan *Radial Basis Function (RBF)*, di mana ditemukan bahwa penggunaan *kernel linear* lebih unggul dalam mengolah sentimen wisatawan. Penelitian ini juga menyoroti pentingnya pemilihan kernel yang tepat berdasarkan karakteristik data, serta penggunaan teknik seperti *cross-validation* untuk meningkatkan Akurasi model.

Di sisi lain, penelitian yang dilakukan oleh (Sibyan & Hasanah, 2022) tentang analisis sentimen pada objek wisata Dieng dengan menerapkan algoritma *K-Nearest Neighbor (K-NN)* memperlihatkan hasil yang memuaskan dengan tingkat akurasi sebesar 86%. Walaupun dalam berbagai studi terdahulu SVM lebih sering digunakan, K-NN tetap menunjukkan efektivitasnya dalam membedakan ulasan wisatawan yang bersifat positif maupun negatif. Hal ini menegaskan bahwa penggunaan teknik yang lebih sederhana pun dapat memberikan hasil optimal jika algoritma yang dipilih sesuai dengan kebutuhan analisis.

Pada tahun 2022, Gergoris Kopong Pati dan Elfira Umar (Pati & Umar, 2022) Penelitian tersebut juga melakukan analisis sentimen terhadap ulasan pengunjung Danau Weekuri dengan menerapkan metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor*. Berdasarkan pengujian yang dilakukan menggunakan *Rapidminer*, perbandingan tingkat akurasi menunjukkan bahwa metode *Naïve*

*Bayes Classifier* mencapai 73,47%, sementara metode *K-Nearest Neighbor* menghasilkan akurasi yang lebih tinggi yaitu sebesar 76,53%.

Penelitian Shima Fanisa dan kawan-kawan (Fanissa et al., 2018) dengan judul penelitian “Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode *Naïve Bayes* dan *Seleksi Fitur Query Expansion Ranking*” menjelaskan pentingnya teknik ekstraksi data dalam mengoptimalkan proses dari analisis sentimen dengan cara mengurangi dimensi fitur dan meningkatkan Akurasi klasifikasi, yang menghasilkan analisis sentimen pariwisata Kota Malang yang lebih akurat.

Chawla et al. (2002) Teknik *Synthetic Minority Over-sampling Technique (SMOTE)* diperkenalkan untuk mengatasi masalah ketidakseimbangan kelas pada dataset. Dibandingkan hanya menyalin data yang ada, *SMOTE* menciptakan sampel baru untuk kelas minoritas melalui proses interpolasi antar sampel. Metode ini telah terbukti lebih unggul dalam mengoptimalkan kinerja model klasifikasi jika dibandingkan dengan teknik oversampling konvensional.

Penerapan *SMOTE* dalam studi ini didasari oleh temuan Nugraha et al. (2024) yang menggunakan berbagai algoritma klasifikasi dan teknik *resampling* pada dataset *Heart Failure Prediction*. Hasil penelitian tersebut mengonfirmasi bahwa mengatasi ketidakseimbangan data sebelum fase klasifikasi secara efektif meningkatkan metrik *Recall* untuk kelas minoritas dalam prediksi gagal jantung.

Berikutnya, ulasan mengenai penelitian terdahulu yang membahas pemodelan topic menggunakan LDA seperti, Blei et al. (2003) memperkenalkan *Latent Dirichlet Allocation (LDA)* sebagai sebuah model generatif probabilistik

yang digunakan dalam pemodelan topik di berbagai koleksi dokumen. Dalam pendekatannya, LDA bekerja dengan asumsi bahwa setiap dokumen disusun dari kombinasi beberapa topik, di mana masing-masing topik tersebut dipahami sebagai distribusi pada sekumpulan kosakata. Teknik ini telah diakui secara luas sebagai standar utama guna mengekstraksi topik dari data teks yang bersifat tidak terstruktur.

Subarkah et al. (2022) mengintegrasikan analisis sentimen dengan pemodelan topik pada ulasan objek wisata di Yogyakarta menggunakan model *Neural Network IndoBERTweet*. Penelitian ini menunjukkan bahwa pemodelan topik setelah klasifikasi sentimen menghasilkan wawasan yang lebih mendalam dan *actionable* bagi pengelola destinasi wisata, yang menjadi justifikasi integrasi LDA didalam penelitian ini.

Berdasarkan penelitian-penelitian tersebut, dapat disimpulkan bahwa analisis sentimen menggunakan berbagai algoritma klasifikasi seperti SVM, K-NN, dan *Naïve Bayes* telah diterapkan secara efektif pada objek wisata yang sudah dikenal luas, seperti Yogyakarta, Danau Toba, dan Purwakarta. Namun, sebagian besar penelitian tersebut hanya berfokus pada klasifikasi sentimen tanpa menindaklanjuti hasil klasifikasi untuk mengidentifikasi aspek spesifik yang menjadi sumber keluhan. Selain itu, persoalan ketidakseimbangan kelas data yang umum terjadi pada ulasan pariwisata belum ditangani secara eksplisit dalam sebagian besar penelitian yang ada. Terdapat pula kesenjangan penelitian pada destinasi wisata yang kurang populer namun memiliki potensi besar, seperti Kabupaten/Kota Bima di Nusa Tenggara Barat. Penelitian ini memiliki tujuan untuk

mengisi kesenjangan tersebut melalui integrasi SVM untuk klasifikasi sentimen, SMOTE untuk penanganan ketidakseimbangan kelas, dan LDA untuk pemodelan topik keluhan, sehingga menghasilkan analisis yang *komprehensif* dan rekomendasi Pembangunan atau pengembangan pariwisata yang lebih terarah berbasis data.

Rangkuman penelitian terkait dapat dilihat pada Tabel 2.1 berikut:

Tabel 2.1 Perbandingan Penelitian

| No. | Referensi                        | Metode                                                                                    | Hasil                                                                                                                       | Relevansi                                                                                                                                                                                                                       |
|-----|----------------------------------|-------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | Syahlan, M. S., et al. (2023)    | <i>Support Vector Machine</i>                                                             | Akurasi 94%, data <i>Google Maps</i> wisata Purwakarta                                                                      | Acuan penerapan SVM untuk sentimen wisata                                                                                                                                                                                       |
| 2   | Tri, I., et al. (2022)           | <i>Support Vector Machine</i> (linear dan RBF)                                            | Akurasi 83%, <i>kernel linear</i> lebih unggul                                                                              | Dasar pemilihan kernel linear                                                                                                                                                                                                   |
| 3   | Sibyan, H., & Hasanah, N. (2022) | <i>K-Nearest Neighbor</i>                                                                 | Akurasi 86%                                                                                                                 | Perbandingan algoritma klasifikasi                                                                                                                                                                                              |
| 4   | Pati, G. K., & Umar, E. (2022)   | <i>K-Nearest Neighbor and Naive Bayes Classifier</i>                                      | Naïve Bayes Akurasi 73.47%, K-NN: Akurasi 76.53%                                                                            | Perbandingan metode klasifikasi teks                                                                                                                                                                                            |
| 5   | Fanissa, S., et al. (2018)       | Implementasi <i>Query Expansion Ranking</i> pada Algoritma <i>Naive Bayes</i>             | Meningkatkan Akurasi dengan seleksi fitur                                                                                   | Pentingnya ekstraksi fitur teks                                                                                                                                                                                                 |
| 6   | Chawla et al. (2002)             | SMOTE                                                                                     | Efektif menangani ketidakseimbangan kelas                                                                                   | Dasar penerapan SMOTE pada data latih                                                                                                                                                                                           |
| 7   | (Nugraha et al., 2024)           | Klasifikasi + Resampling                                                                  | Peningkatan <i>Recall</i> kelas minoritas                                                                                   | Acuan penerapan SMOTE dalam penelitian ini                                                                                                                                                                                      |
| 8   | Blei et al. (2003)               | LDA                                                                                       | Model probabilistik pemodelan topik teks                                                                                    | Landasan teoritis LDA                                                                                                                                                                                                           |
| 9   | Subarkah et al. (2022)           | IndoBERTweet + Topic Modeling                                                             | Wawasan lebih mendalam dari ulasan negatif                                                                                  | Justifikasi integrasi LDA setelah klasifikasi SVM                                                                                                                                                                               |
| 10  | Penelitian ini (2025)            | Support Vector Machine (SVM), <i>SMOTE</i> , dan <i>Latent Dirichlet Allocation (LDA)</i> | Akurasi tertinggi 84,91% (skenario 90:10). Berhasil mengidentifikasi topik spesifik keluhan (infrastruktur dan kebersihan). | 84,91% (skenario 90:10). Berhasil mengidentifikasi topik spesifik keluhan (infrastruktur dan kebersihan). Integrasi SVM untuk klasifikasi sentimen, SMOTE untuk data tidak seimbang, dan LDA untuk ekstraksi topik keluhan pada |



| No. | Referensi | Metode | Hasil | Relevansi                         |
|-----|-----------|--------|-------|-----------------------------------|
|     |           |        |       | objek wisata Kabupaten/Kota Bima. |

## 2.2 Perbandingan Penelitian Terdahulu dan Kebaruan Penelitian

Berdasarkan tinjauan pustaka yang telah dilakukan, berikut adalah tabel ringkasan perbedaan antara penelitian ini dengan beberapa penelitian sebelumnya:

Tabel 2. 2 Perbandingan Penelitian

| No. | Referensi                        | Metode                                                                            | Objek Wisata                                  | Perbedaan/Kebaruan                                                                                                                                          |
|-----|----------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | Syahlan, M. S., et al. (2023)    | <i>Support Vector Machine</i>                                                     | Tempat wisata (umum)                          | Fokus pada klasifikasi sentimen umum tanpa penanganan data tidak seimbang.                                                                                  |
| 2   | Tri, I., et al. (2022)           | <i>Support Vector Machine</i> (linear dan RBF)                                    | Danau Weekuri                                 | Menggunakan algoritma probabilitas, objek terbatas pada satu lokasi spesifik.                                                                               |
| 3   | Sibyan, H., & Hasanah, N. (2022) | K-Nearest Neighbor                                                                | Wisata Dieng                                  | Menggunakan algoritma KNN, tidak melakukan analisis topik keluhan.                                                                                          |
| 4   | Penelitian ini (2025)            | Support Vector Machine (SVM), SMOTE, dan <i>Latent Dirichlet Allocation</i> (LDA) | Kabupaten/Kota Bima (Lariti, Lawata, Tambora) | Kebaruan: Mengintegrasikan SMOTE untuk menangani <i>imbalanced data</i> ulasan dan LDA untuk mengekstraksi topik keluhan spesifik sebagai saran manajerial. |

## 2.3 Pariwisata di Kabupaten/Kota Bima

Kabupaten/kota Bima, yang berada di Nusa Tenggara Barat, merupakan daerah dengan potensi wisata alam, budaya, dan sejarah yang sangat besar. Objek wisata di Kabupaten/Kota Bima meliputi destinasi alam seperti Pantai Lariti, Pantai Lawata, dan Gunung Tambora, serta objek wisata budaya dan sejarah seperti Istana Kesultanan Bima, Museum Samparaja, dan Benteng Asakota. Menurut data dari Badan Pusat Statistik NTB (2023), sektor pariwisata di NTB mengalami

peningkatan yang signifikan dengan kenaikan jumlah kunjungan wisatawan sebesar 18,6% pada tahun 2023. Namun, Kabupaten/Kota Bima masih menghadapi berbagai tantangan dalam hal promosi dan infrastruktur, yang membuatnya kurang dikenal dibandingkan dengan daerah lain seperti Lombok.

Sektor pariwisata memiliki potensi signifikan dalam mendorong pendapatan masyarakat lokal, sehingga peningkatan kualitas objek wisata di Kabupaten/Kota Bima menjadi krusial bagi pembangunan ekonomi daerah. Mengingat pentingnya peran tersebut, pemahaman mendalam mengenai persepsi pengunjung terhadap destinasi di wilayah ini sangat diperlukan. Hal ini bertujuan untuk merumuskan kebijakan pengelolaan serta pengembangan pariwisata yang lebih akurat dan efektif di masa mendatang.

## 2.4 Analisis Sentimen

Metode *text mining* yang dikenal sebagai analisis sentimen, atau *opinion mining*, berfungsi untuk mengenali serta mengambil data subjektif dari sebuah teks (Syahlan et al., 2023). Analisis sentimen pada sektor pariwisata diterapkan guna menelaah persepsi, emosi, serta pandangan wisatawan mengenai beragam elemen destinasi, termasuk kualitas pelayanan, sarana prasarana, kemudahan *aksesibilitas*, maupun daya tarik yang ditawarkan (Tri et al., 2022). Tahapan dalam proses analisis sentimen biasanya meliputi pengumpulan data, *preprocessing*, *ekstraksi fitur*, klasifikasi sentimen, serta evaluasi model (Syahlan et al., 2023).

Proses analisis sentimen diawali dengan pengumpulan data, di mana ulasan pengunjung dihimpun dari platform digital seperti *TripAdvisor*, *Google Maps*, ataupun media sosial. Guna memastikan data siap diproses secara optimal, langkah

selanjutnya adalah tahap *preprocessing* yang meliputi *cleaning* (pembersihan data), *stemming*, serta eliminasi *stopwords*. Setelah itu, ekstraksi fitur teks dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengevaluasi signifikansi sebuah kata di dalam dokumen relatif terhadap keseluruhan korpus data yang tersedia (Utama, 2023).

#### 2.4.1 Perhitungan Bobot TF-IDF

Pentingnya sebuah kata dalam suatu dokumen terhadap korpus diukur menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Perhitungan bobot akhir untuk tiap kata dilakukan melalui perkalian antara nilai Term Frequency (TF) dan Inverse Document Frequency (IDF), yang secara matematis ditunjukkan pada Persamaan (2.1):

$$TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (2.1)$$

Keterangan simbol pada Persamaan :

$W(t, d)$ : Bobot *term* (kata)  $t$  pada dokumen  $d$ .

$Tf(t, d)$ : Frekuensi kemunculan *term*  $t$  dalam dokumen  $d$ .

$Idf(t, D)$ : Nilai *Inverse Document Frequency* dari *term*  $t$  pada seluruh korpus  $D$ .

Perhitungan nilai IDF ditujukan guna menetapkan bobot yang lebih besar bagi istilah yang jarang muncul (diskriminatif), sekaligus memberikan bobot yang lebih kecil untuk kata-kata yang frekuensi kemunculannya sangat tinggi di seluruh kumpulan dokumen. Adapun formulasi IDF tersebut dipaparkan dalam Persamaan (2.2):

$$IDF(t, D) = \log(N / (df(t) + 1)) \quad (2.2)$$

Keterangan simbol pada Persamaan (2.2):

$N$ : Jumlah total seluruh dokumen dalam korpus.

$df(t)$ : Jumlah dokumen yang mengandung *term*  $t$  tertentu.

+ 1: Penambahan angka satu (*Laplace Smoothing*) dilakukan untuk menghindari pembagian dengan nol jika suatu kata tidak ditemukan sama sekali dalam korpus.

Dalam hal ini, variabel  $N$  merepresentasikan kuantitas keseluruhan dokumen pada korpus, sementara  $df(t)$  menunjukkan banyaknya dokumen yang memuat term  $t$ .

Tahap berikutnya setelah proses ekstraksi fitur adalah evaluasi model. Penilaian performa model dalam melakukan klasifikasi sentimen ini dilakukan melalui pengukuran metrik-metrik evaluasi, yang meliputi *Akurasi*, *presisi*, *Recall*, serta *F1-Score*.

## 2.5 Preprocessing Teks

Langkah awal yang sangat penting dalam analisis sentimen adalah tahap *preprocessing* data, yang berfungsi mengonversi teks tidak terstruktur menjadi bentuk yang lebih terstandarisasi dan bersih. Dengan melakukan proses ini, kualitas fitur teks dapat ditingkatkan guna mengoptimalkan kinerja algoritma *Support Vector Machine* (SVM) (Hakim, 2021). Menurut (Feldman, n.d.) *text preprocessing* adalah tahapan fundamental di mana sekumpulan karakter teks dipecah dan dibersihkan untuk mengurangi *noise* pada proses klasifikasi. Metode *preprocessing* dalam penelitian ini meliputi:

### 2.5.1 Cleaning

Tahap *cleaning* merupakan langkah untuk menyeleksi teks ulasan dengan menghapus elemen yang kurang informatif bagi analisis klasifikasi. Guna

mengurangi noise, berbagai karakter seperti tanda baca, simbol, angka, tautan URL, hingga emotikon dihilangkan pada proses ini. (Feldman, n.d.).

### **2.5.2 Case folding**

Tujuan utama dari case folding adalah menyeragamkan seluruh karakter dalam dokumen ke dalam bentuk huruf kecil (*lowercase*), mulai dari 'a' hingga 'z'. Proses ini berfungsi untuk mencegah terjadinya duplikasi data serta menjamin agar mesin dapat mengidentifikasi kata yang identik secara konsisten, tanpa terhambat oleh perbedaan penggunaan huruf kapital (Hakim, 2021).

### **2.5.3 Tokenizing**

*Tokenizing* adalah suatu tahapan untuk memecah rangkaian teks atau kalimat menjadi kumpulan kata tunggal yang disebut sebagai token. Dalam proses ini, elemen seperti spasi kosong, angka, dan tanda baca dibersihkan sehingga struktur teks terurai menjadi bagian-bagian individual untuk memudahkan analisis pada tahap pembobotan (Feldman, n.d.) .

### **2.5.4 Stopword Removal**

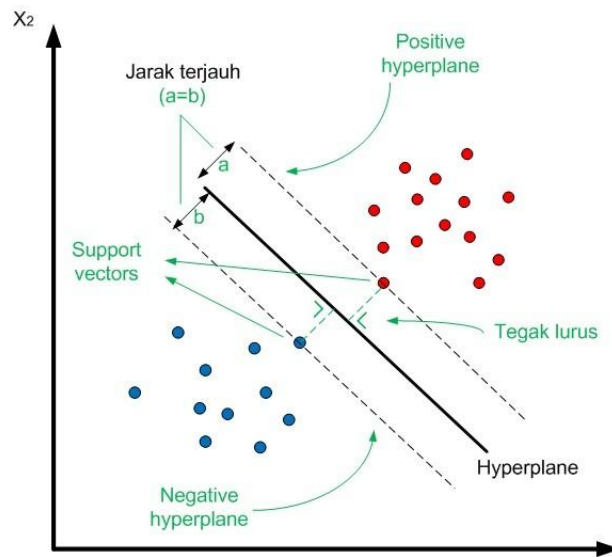
*Stopword removal* merupakan proses mengeliminasi kata-kata umum (*common words*) yang sering muncul tetapi tidak membawa makna sentimen atau nilai informasi yang berarti. Contoh dari kata-kata tersebut adalah kata hubung seperti "di", "yang", "untuk", atau "dan". Dengan melakukan penghapusan ini, model dapat bekerja dengan lebih akurat karena lebih terfokus pada kata-kata kunci yang memiliki bobot sentiment (Hakim, 2021).

## 2.6 *Support Vector Machine (SVM)*

Dalam bidang analisis sentimen, *Support Vector Machine (SVM)* merupakan algoritma pembelajaran mesin yang populer digunakan untuk klasifikasi teks. Algoritma ini beroperasi dengan menentukan hyperplane paling optimal guna menciptakan margin maksimal di antara dua kategori data, seperti sentimen positif dan negatif. Selain itu, SVM mengandalkan strategi *One-vs-Rest (OVR)* atau *One-vs-One (OVO)* ketika menangani kasus klasifikasi yang melibatkan lebih dari dua kelas.

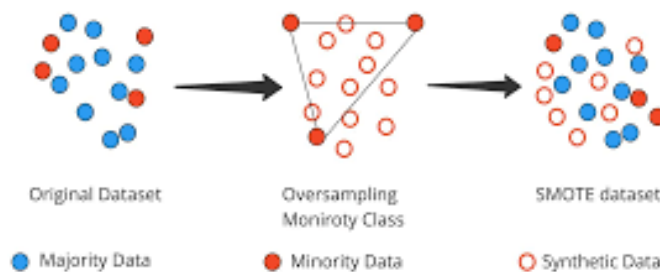
Beberapa kelebihan SVM meliputi kemampuannya untuk mengolah ruang fitur berdimensi tinggi, fleksibilitas dalam penerapan beragam kernel seperti linear, polynomial, serta RBF, dan ketahanan terhadap masalah *overfitting*. Dengan karakteristik tersebut, SVM menjadi metode yang sangat relevan untuk analisis sentimen ulasan wisatawan yang umumnya memiliki tingkat kompleksitas dan dimensi data yang sangat besar.

Penelitian ini akan menggunakan SVM pendekatan *One-vs-Rest (OVR)* dengan *kernel linear* untuk mengklasifikasikan sentimen ulasan pengunjung objek wisata di Kabupaten/Kota Bima, Guna menyajikan deskripsi yang lebih eksplisit terkait sudut pandang pelancong mengenai objek wisata di wilayah tersebut.

Gambar 2.1 Ilustrasi *Hyperplane* dan SVM

## 2.7 Synthetic Minority Over-sampling Technique (SMOTE)

SMOTE adalah teknik resampling yang dikembangkan oleh Chawla et al. (2002) untuk mengatasi permasalahan ketidakseimbangan kelas pada dataset. Alih-alih melakukan duplikasi pada sampel minoritas sebagaimana oversampling konvensional, SMOTE bekerja dengan cara mensintesis sampel baru secara artifisial melalui teknik interpolasi. Proses ini dilakukan dengan memanfaatkan informasi dari sampel-sampel minoritas yang telah ada berdasarkan perhitungan jarak  $k$ -tetangga terdekat (*k-nearest neighbors*).



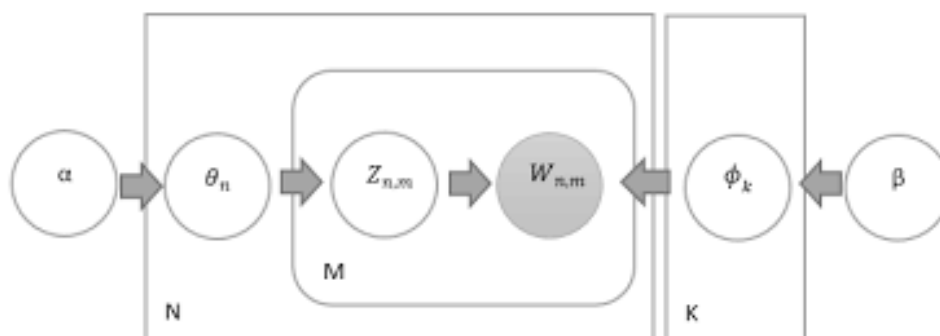
Gambar 2.2 Ilustrasi Proses Sintesis Sampel pada SMOTE

SMOTE diimplementasikan hanya pada data latih dalam studi ini guna menghindari data leakage terhadap data uji. Langkah tersebut diambil untuk menyeimbangkan distribusi kelas pada data latih, memastikan model SVM dapat mengidentifikasi kelas minoritas (sentimen negatif) secara optimal.

## 2.8 Latent Dirichlet Allocation (LDA)

Model probabilistik generatif yang dikenal sebagai *Latent Dirichlet Allocation* (LDA) pertama kali dikemukakan oleh Blei et al. (2003) LDA digunakan dalam pemodelan topik untuk menganalisis koleksi dokumen. Dalam model ini, setiap dokumen dipandang sebagai kombinasi dari berbagai topik laten, di mana setiap topik tersebut didefinisikan sebagai distribusi probabilitas dari kata-kata yang tersedia.

Dalam kerangka formalnya, *Latent Dirichlet Allocation* (LDA) menyajikan dua bentuk distribusi mendasar: distribusi kata terhadap setiap topik ( $\phi$ ) serta distribusi topik yang menyusun setiap dokumen ( $\theta$ ). Proses inferensi dilakukan untuk memperkirakan distribusi-distribusi tersebut berdasarkan dokumen yang diamati. Kata-kata dengan probabilitas tertinggi pada suatu topik digunakan untuk menginterpretasikan tema yang diwakili oleh topik tersebut.



Gambar 2.3 Model Grafis LDA (Blei et al., 2003)



Dalam penelitian ini, LDA diterapkan pada kumpulan ulasan yang telah diklasifikasikan sebagai sentimen negatif oleh model SVM. Guna memberikan informasi spesifik yang dapat ditindaklanjuti oleh pihak pengelola wisata, tujuan utamanya adalah melakukan identifikasi terhadap berbagai topik keluhan yang paling dominan muncul.

## **2.9 Web Scraping**

Dalam penelitian ini, teknik pengambilan data secara otomatis dari halaman web yang dikenal sebagai *web scraping* diterapkan untuk menghimpun ulasan pengunjung melalui ekstensi browser. Proses pengumpulan data tersebut difokuskan pada teks ulasan dari *Google Maps* yang mencakup periode sepuluh tahun terakhir pada tiga destinasi wisata di wilayah Kabupaten dan Kota Bima.

## BAB III

### METODE PENELITIAN

#### 3.1 Pengumpulan Data

Penelitian ini mengumpulkan data dengan cara teknik web scrapping yang mengambil dari *Google Maps* dengan memanfaatkan sebuah aplikasi berbasis web yaitu *Instan data scraper*. Proses pengumpulan data dilakukan pada tanggal 7 November 2025, dengan cakupan batasan data berupa ulasan wisatawan dalam rentang waktu 10 tahun terakhir, kemudian data yang terkumpul disimpan dalam format Excel untuk memudahkan pemrosesan lebih lanjut, ditunjukkan pada Tabel 3.1.

Tabel 3.1 Penyajian Data

| Komentar                                                                                                                                                                                                                                         |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Pantainya Bagus, indah, bisa lihat fenomena laut terbelah juga,.                                                                                                                                                                                 |
| Pantai ini 2 jam dari Kota Bima menyusuri jalan pegunungan berkelok-kelok. Bisa ditempuh dengan kendaraan roda 2 atau roda 4. Jalan sempit tapi masih bisa papasan. Buat yang tidak terbiasa, mungkin akan pusing dan mual dalam perjalanan. â€  |
| Bintang 4 karena pantainya bagus, experience untuk jalan kepulau sebrang menyenangkan. Sayang sekali, pantai yang indah tapi sekeliling masih sangat kotor. Tidak tersedia tempat pembuangan sampah jadi memang harus kesadaran masing-masing â€ |
| Pantainya Kotor karna kurangnya kesadaran Dari wisatawan dlm menjaga kebersihan, selain itu fasilitas empat sampah yg tdk memadai juga menjadi Salah satu penyebab                                                                               |

Setelah proses pengumpulan data, kemudian data dibagi menjadi 2 bagian yaitu: data training dan data testing. Pada data training berfungsi untuk dasar model klasifikasi *Support Vector Machine*, sedangkan data testing berfungsi untuk menilai kualitas atau performa model pada data yang tidak terlihat. Pada proses pemisahan data ini untuk memastikan model dapat menyamaratakan pengetahuan secara efektif untuk mengklasifikasikan ulasan baru.

Tabel 3.2 Data *Testing*

| Komentar                                                                                                          |
|-------------------------------------------------------------------------------------------------------------------|
| Pantainya indah                                                                                                   |
| Pantainya bagus banget                                                                                            |
| Pantainya keren, panorama alamnya indah, perlu penambahan tempat untuk berteduh, akses jalan masih diperbaiki.... |
| Alhamdulillah... rejeki cewe Solehahhhh...                                                                        |
| Pertama kalinya Dateng langsung dikasi rejeki pantai yang terbelah...                                             |
| Keren                                                                                                             |
| The best beach bagian Bima nih                                                                                    |
| Pantai terbelah , Unik                                                                                            |

Tabel 3.3 Data *Training*

| Komentar                                                                                                                                                                                                                                  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Beautiful                                                                                                                                                                                                                                 |
| Pantai alami yang exotic                                                                                                                                                                                                                  |
| Pantainya Bagus, indah, bisa lihat fenomena laut terbelah juga,.                                                                                                                                                                          |
| Lariti Punya Masa depan Cerah.                                                                                                                                                                                                            |
| Lariti yang terletak di desa Soro kecamatan Lambu Kabupaten Bima adalah Pantai â€                                                                                                                                                         |
| Tempat yang indah alami sangat berkelas...                                                                                                                                                                                                |
| Wisata pantai indah di Sape, Lariti. Terdapat banyak spot untuk berfoto ria bersama, ayo datang dan ramaikan. Jangan lupa untuk tetap menjaga kelestarian lingkungan kita dengan menjaga kebersihan, jangan mengotori harta warisan kita. |
| Pantai yg bagus                                                                                                                                                                                                                           |

### 3.2 Ekstraksi Data

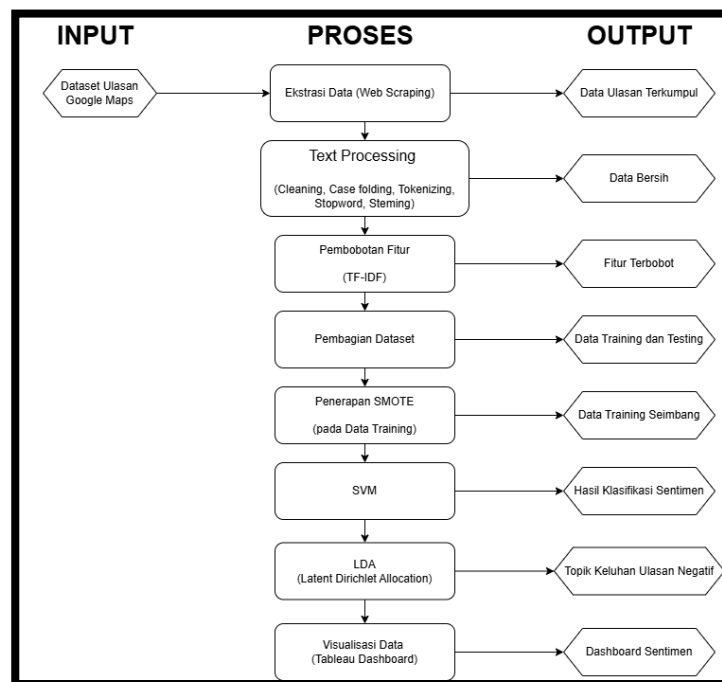
Setelah dilakukan pengumpulan, langkah selanjutnya yaitu melakukan ekstraksi data dengan proses sistematis pengambilan dan pengumpulan data ulasan dari platform *Google Maps* menggunakan teknik web scraping untuk mendapatkan dataset yang komprehensif dan representatif, dilanjutkan dengan penyimpanan data dalam format terstruktur seperti *CSV*, *JSON*, atau database relasional untuk memudahkan pemrosesan tahap selanjutnya. Penelitian Shima Fanisa dan kawan-kawan (Fanissa et al., 2018) dengan judul ‘Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode *Naïve Bayes* dan Seleksi Fitur *Query Expansion Ranking*’ menjelaskan pentingnya teknik ekstraksi data dalam jurnal ini sangat penting untuk mengoptimalkan proses analisis sentiment dengan mengurangi

dimensi fitur dan meningkatkan Akurasi klasifikasi, yang pada akhirnya akan menghasilkan analisis sentiment pariwisata Kota Malang yang lebih akurat.

### 3.3 Rancangan Sistem

Sistem analisis sentimen yang dibangun dalam penelitian ini dirancang dengan alur kerja yang sistematis, mulai dari pengumpulan data hingga visualisasi hasil. Secara garis besar, alur sistem terdiri dari enam tahapan utama: (1) pengumpulan data melalui web scraping, (2) pelabelan data secara manual, (3) preprocessing teks, (4) pembobotan fitur TF-IDF, (5) klasifikasi sentimen menggunakan SVM dengan penanganan ketidakseimbangan data SMOTE, dan (6) pemodelan topik keluhan menggunakan LDA.

Hasil akhir kemudian divisualisasikan menggunakan Tableau. Alur proses analisis sentimen yang dilakukan dalam penelitian ini dapat dilihat pada gambar berikut.



Gambar 3.1 Diagram Alur Rancang Sistem

Setiap tahapan dirancang agar dapat dijalankan secara berurutan dan reproduisibel. Pemisahan penerapan SMOTE hanya pada data latih merupakan keputusan desain yang penting untuk memastikan validitas evaluasi model pada data uji yang mencerminkan kondisi nyata.

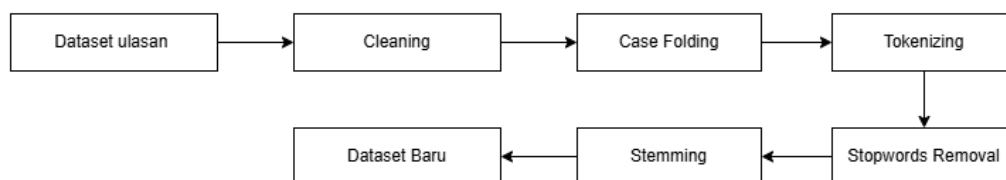
### 3.3.1 Output Sistem dan Cara Penggunaan

Sistem analisis sentimen ini menghasilkan dua bentuk output utama. Pertama, output klasifikasi sentimen berupa label kelas (positif, netral, atau negatif) untuk setiap ulasan pengunjung, disertai metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score* dalam bentuk *Confusion Matrix*. Kedua, output pemodelan topik berupa daftar topik keluhan beserta kata kunci dominan yang divisualisasikan dalam bentuk *wordcloud* dan grafik *Coherence Score*.

Secara praktis, sistem ini dapat dimanfaatkan melalui dua cara. (1) Melalui notebook Python pengguna menjalankan setiap sel secara berurutan mulai dari tahap web *scrapping*, *preprocessing*, pelatihan model SVM, hingga pemodelan topik LDA. Setiap tahap menghasilkan output yang dapat diamati dan disimpan langsung. (2) Melalui dashboard Tableau: hasil klasifikasi sentimen dan topik keluhan yang telah diekspor ke format CSV diimpor ke Tableau untuk ditampilkan dalam bentuk visualisasi interaktif berupa diagram distribusi sentimen per destinasi wisata, grafik tren ulasan dari waktu ke waktu, dan peta topik keluhan yang dapat digunakan oleh pemangku kepentingan sebagai dasar pengambilan keputusan.

### 3.4 Preprocessing Text

*Preprocessing text* berfungsi untuk mempersiapkan data agar data menjadi lebih terstruktur dan siap diolah oleh algoritma klasifikasi teks, dengan sengaja menghindari proses stemming karena dapat menurunkan Akurasi klsifikasi teks bahasa Indonesia (Hidayatullah & dkk, 2016). Pada penelitian ini, *preprocessing* teks melibatkan lima tahap yang dijelaskan dalam gambar 3.2.



Gambar 3. 2 Tahapan *Preprocessing Text*

#### 3.4.1 Cleaning

Dalam proses *preprocessing* data, tahap pembersihan dilakukan untuk Menghilangkan suara seperti angka, simbol, tanda baca, emotikon, dan lainnya.

Tabel 3.4 menunjukkan diagram aliran data untuk membersihkan data.

Tabel 3.4 Contoh Cleaning

| Langkah         | Hasil                                                             |
|-----------------|-------------------------------------------------------------------|
| Teks Asli       | “Pantainya Bagus, indah, bisa lihat fenomena laut terbelah juga.” |
| <i>Cleaning</i> | “pantainya bagus indah bisa lihat fenomena laut terbelah juga”    |

#### 3.4.2 Case folding

Tujuan dari *case folding* adalah mengubah huruf kapital di setiap dokumen menjadi karakter kecil untuk menghilangkan pengulangan pada data sehingga proses klasifikasi dan perhitungan dapat dilakukan dengan lebih baik. Oleh karena itu, huruf kapital dibuang dari teks dan huruf kecil digunakan secara keseluruhan (Sakdiyah, 2023). Contohnya ditunjukkan pada Tabel 3.5

Tabel 3.5 Contoh Case folding

| Langkah             | Hasil                                                              |
|---------------------|--------------------------------------------------------------------|
| Teks Asli           | “Pantainya Bagus, indah, bisa lihat fenomena laut terbelah juga,.” |
| <i>Case folding</i> | “pantainya bagus, indah, bisa lihat fenomena laut terbelah juga,.” |

### 3.4.3 Tokenizing

*Tokenization* merupakan proses untuk memecah teks dokumen menjadi kalimat, kemudian memecahnya menjadi kata-kata. Proses ini juga dilakukan untuk menghilangkan angka, tanda baca dan spasi. *Tokenizing* dan contohnya ditunjukkan pada Tabel 3.6

Tabel 3.6 Contoh *Tokenizing*

| Langkah    | Hasil                                                                                    |
|------------|------------------------------------------------------------------------------------------|
| Teks Asli  | “Pantainya Bagus, indah, bisa lihat fenomena laut terbelah juga,.”                       |
| Tokenizing | ['pantainya', 'bagus', 'indah', 'bisa', 'lihat', 'fenomena', 'laut', 'terbelah', 'juga'] |

### 3.4.4 Stopword Removal

Proses penghapusan kata *stopword* tidak berdampak pada dokumen atau nilai. Akurasi klasifikasi sentimennya. Sebaliknya, kata yang dihapus akan disimpan dalam database kata *stopword* dan akan dibersihkan jika kemudian ditemukan. *Stopword* dan contohnya ditunjukkan pada Tabel 3.7

Tabel 3.7 Contoh *Stopword Removal*

| Langkah                 | Hasil                                                                    |
|-------------------------|--------------------------------------------------------------------------|
| Teks Asli               | “Pantainya Bagus, indah, bisa lihat fenomena laut terbelah juga,.”       |
| <i>Stopword Removal</i> | ['pantainya', 'bagus', 'indah', 'lihat', 'fenomena', 'laut', 'terbelah'] |

### 3.4.5 Stemming

Proses mengubah kata yang berimbuhan menjadi Kembali ke kata awal sesuai dengan aturan stemming. Ini termasuk prefix seperti “me-”, “pe-”, “ber-”, “te-”, lalu infix seperti “-el-”, “em-”, dan sufiks.

Tabel 3.8 Contoh *Stemming*

| Langkah   | Hasil                                                              |
|-----------|--------------------------------------------------------------------|
| Teks Asli | “Pantainya Bagus, indah, bisa lihat fenomena laut terbelah juga..” |
| Stemming  | ['pantai', 'bagus', 'indah', 'lihat', 'fenomena', 'laut', 'belah'] |

Tahapan-tahapan ini didesain untuk menghasilkan data teks yang terstandarisasi, bebas dari elemen yang tidak diperlukan, dan optimal untuk diolah menggunakan teknik klasifikasi *Support Vector Machine* (SVM). *Preprocessing* yang tepat memungkinkan sistem untuk meningkatkan Akurasi dalam mengenali sentimen yang terkandung dalam ulasan pada *Google Maps*.

Dalam upaya mengatasi sentimen campuran atau kompleks, penelitian yang dilakukan Utama (2024) menunjukkan bahwa heuristik sering diterapkan dengan cara menetapkan aturan-aturan khusus. Beberapa aturan heuristik umum yang digunakan dalam analisis sentiment, yaitu:

- a) *Positive Override*: Jika terdapat elemen positif dan negatif dalam satu kalimat, tetapi aspek positif dianggap lebih dominan atau penting, maka keseluruhan kalimat dapat diberi label positif. Aturan ini berguna dalam situasi di mana aspek positif memberikan dampak yang signifikan terhadap kepuasan keseluruhan, meskipun ada elemen negatif kecil.
- b) *Negative Override*: Sebaliknya, jika aspek negatif lebih kuat atau berperan besar dalam memengaruhi persepsi, maka kalimat tersebut dilabeli sebagai negatif, bahkan jika ada elemen positif di dalamnya. Aturan ini membantu menyoroti masalah atau kekurangan signifikan yang dirasakan pengguna.



c) *Netral Override*: Jika terdapat elemen positif dan negatif dalam satu kalimat, tetapi aspek positif dan negatif dianggap setara atau tidak ada yang lebih dominan, maka keseluruhan kalimat dapat diberi label netral. Aturan ini berguna dalam situasi di mana aspek positif dan negatif saling menetralkan satu sama lain atau tidak memiliki dampak yang cukup signifikan untuk menentukan polaritas sentimen secara jelas.

Aturan-aturan ini berfungsi untuk menentukan sentimen berdasarkan kekuatan atau dominasi aspek tertentu dalam kalimat.

Tabel 3.9 Pelabelan Data

| Komentar                                                                                                                                                                                                                                           | Sentimen | Preprocessing komentar                                                                                                                                               |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Pantainya Bagus, indah, bisa lihat fenomena laut terbelah juga,.                                                                                                                                                                                   | positif  | pantai bagus indah lihat fenomena laut belah                                                                                                                         |
| Pantai ini 2 jam dari Kota Bima menyusuri jalan pegunungan berkelok-kelok. Bisa ditempuh dengan kendaraan roda 2 atau roda 4. Jalan sempit tapi masih bisa papasan. Buat yang tidak terbiasa, mungkin akan pusing dan mual dalam perjalanan. â€    | netral   | pantai jam kota bima susur jalan gunung berkelokkelok tempuh kendar roda roda jalan sempit masih papas buat tidak biasa mungkin pusing mual jalan                    |
| Bintang 4 karena pantainya bagus, experience untuk jalan kepulau sebrang menyenangkan. Sayang sekali, pantai yang indah tapi sekeliling masih sangat kotor. Tidak tersedia tempat pembuangan sampah jadi memang harus kesadaran masing-masing â€   | netral   | bintang pantai bagus experience jalan pulau sebrang senang sayang sekali pantai indah keliling sangat kotor sedia tempat buang sampah jadi memang sadar masingmasing |
| Pantainya indah, harus tiba sekitar jam 7 - 9 di lokasi atau pas sebelum air laut pasang, agar bisa melihat fenomena alam jalan di tengah laut menuju pulau kecil di seberangnya. Tempatnya udah ada warung dan lapak-lapak penjual, sisa kita â€  | positif  | pantai indah tiba jam lokasi pas air laut pasang bisa lihat fenomena alam jalan tengah laut tuju pulau kecil seberang tempat udah warung lapaklapak jual sisa        |
| Pantai Lariti sangat unik daripada pantai yang saya temuin selama ini, karena di pantai ini memiliki jalanan terbelah di tengah laut bagaikan cerita Nabi Musa dan Raja Firâ€™Maun... â€                                                           | positif  | pantai lariti sangat unik pantai saya temuin lama karena pantai milik jalan belah tengah laut bagai cerita nabi musa raja firaun                                     |
| Pantai yang Indah, ketika menuju kesana hamparan perbukitan batu yang menawan. Dibali banyak pantai, tapi yang indah seperti ini sedikit. TAPI sayang sekali, tidak didukung oleh pengunjung, sangat disayangkan sampah berserakan dimana mana, â€ | negatif  | pantai indah tuju kesana hampar bukit batu tawan bal banyak pantai yang indah ini sedikit sayang sekali dukung unjung sangat sayang sampah sera mana                 |

### 3.5 Fitur Pembobotan

Setelah melakukan *preprocessing teks*, data diubah ke bentuk numerik melalui proses pembobotan kata. Langkah ini memberikan nilai pada tiap kata yang menunjukkan tingkat kepentingannya dalam analisis. Secara umum, peningkatan jumlah data menghasilkan kumpulan fitur yang lebih komprehensif (kata-kata dengan bobot), sehingga memperkuat kapasitas model dalam mendeteksi berbagai pola dan detail-detail kecil dalam dataset. Pada penelitian ini, digunakan beberapa metode pembobotan fitur teks sebagai berikut:

#### 3.5.1 Term Frequency-Inverse Document Frequency (TF-IDF)

Proses pemberian bobot pada setiap *term* dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)*. Merujuk pada Persamaan (2.1) dan Persamaan (2.2) yang telah dijabarkan pada Bab II, berikut adalah ilustrasi, diberikan contoh perhitungan TF-IDF pada tiga dokumen ulasan sederhana yang telah melalui *preprocessing*:

Tabel 3.10 Contoh Dokumen Ulasan Untuk Perhitungan TF-IDF

| Komentar                                                                                                                                                                                                                                         | Sentimen | Preprocessing komentar                                                                                                                                               |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Pantainya Bagus, indah, bisa lihat fenomena laut terbelah juga,.                                                                                                                                                                                 | positif  | pantai bagus indah lihat fenomena laut belah                                                                                                                         |
| Pantai ini 2 jam dari Kota Bima menyusuri jalan pegunungan berkelok-kelok. Bisa ditempuh dengan kendaraan roda 2 atau roda 4. Jalan sempit tapi masih bisa papasan. Buat yang tidak terbiasa, mungkin akan pusing dan mual dalam perjalanan. â€  | netral   | pantai jam kota bima susur jalan gunung berkelokkelok tempuh kendara roda roda jalan sempit masih papas buat tidak biasa mungkin pusing mual jalan                   |
| Bintang 4 karena pantainya bagus, experience untuk jalan kepulau sebrang menyenangkan. Sayang sekali, pantai yang indah tapi sekeliling masih sangat kotor. Tidak tersedia tempat pembuangan sampah jadi memang harus kesadaran masing-masing â€ | netral   | bintang pantai bagus experience jalan pulau sebrang senang sayang sekali pantai indah keliling sangat kotor sedia tempat buang sampah jadi memang sadar masingmasing |

Misalkan kata “pantai” muncul 1 kali dalam dokumen d1 yang berisi 7 kata, dan kata tersebut muncul di semua dokumen. Maka:

$$TF(pantai, d1) = 1 / 7 = 0,143$$

$$IDF(pantai, D) = \log(3 / (2 + 1)) = \log(1) = 0,000$$

$$TF-IDF(pantai, d1) = 0,143 \times 0,000 = 0,000$$

Sebaliknya, kata “kotor” yang hanya muncul 1 kali dalam dokumen d3 yang berisi 23 kata. Kata tersebut hanya muncul di 1 dokumen dari 3 total dokumen:

$$TF(kotor, d1) = 1 / 23 = 0,043$$

$$IDF(kotor, D) = \log(3 / (1 + 1)) = \log(1,5) \approx 0,176$$

$$TF-IDF(kotor, d1) = 0,043 \times 0,176 = 0,008$$

Hasil ini menunjukkan bahwa kata “kotor” memiliki bobot lebih tinggi dibandingkan “pantai” karena lebih langka dan lebih diskriminatif secara semantik.

### 3.5.2 Normalisasi Vektor TF-IDF

Untuk meningkatkan kinerja klasifikasi, vektor TF-IDF yang dihasilkan perlu dinormalisasi agar skala semua dokumen menjadi seragam. Normalisasi frekuensi kata dalam analisis teks sangat penting untuk memastikan bahwa panjang dokumen tidak mempengaruhi hasil analisis. Dokumen yang lebih panjang cenderung memiliki frekuensi kata yang lebih tinggi, yang dapat menyebabkan bias dalam proses klasifikasi. Dengan melakukan normalisasi, frekuensi kata disesuaikan sehingga ulasan pendek dan panjang diperlakukan secara setara. Sebagai contoh, penelitian yang diterbitkan dalam Jurnal Teknologi Informasi menjelaskan bahwa normalisasi frekuensi kata dilakukan dengan membagi nilai

*frekuensi term* (TF) dengan panjang dokumen, untuk mencegah dokumen yang lebih panjang memiliki nilai TF yang lebih tinggi (Utama, 2023).

### 3.6 Pembagian Data dan Penerapan SMOTE

#### 3.6.1 Pembagian Data

Dataset dibagi menjadi data latih (*training data*) dan data uji (*testing data*) untuk keperluan pelatihan dan evaluasi model. Penelitian ini menguji tiga skenario rasio pembagian data, yaitu 90:10, 80:20, dan 70:30. Pembagian dilakukan secara acak dengan mempertahankan proporsi kelas (*stratified split*) agar distribusi sentimen pada data latih dan data uji tetap representatif.

Tabel 3.11 Skenario Pembagian Data Latih dan Data Uji

| Skenario | Data Latih (%) | Data Uji (%) | Keterangan                               |
|----------|----------------|--------------|------------------------------------------|
| 1        | 90%            | 10%          | Data latih besar, data uji kecil         |
| 2        | 80%            | 20%          | Keseimbangan standar                     |
| 3        | 70%            | 30%          | Data uji lebih besar untuk evaluasi luas |

#### 3.6.2 Penerapan SMOTE

Ketidakseimbangan kelas pada data ulasan pariwisata merupakan tantangan yang lazim terjadi, di mana ulasan positif secara alami mendominasi dibandingkan ulasan negatif. Tanpa penanganan khusus, model SVM cenderung mengabaikan kelas minoritas. Untuk mengatasi hal ini, diterapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) secara eksklusif pada data latih.

SMOTE mensintesis sampel baru pada kelas minoritas dengan cara menginterpolasi antara suatu sampel minoritas dengan salah satu dari  $k$ -tetangga terdekatnya. Persamaan 3.1 sintesis sampel baru adalah:

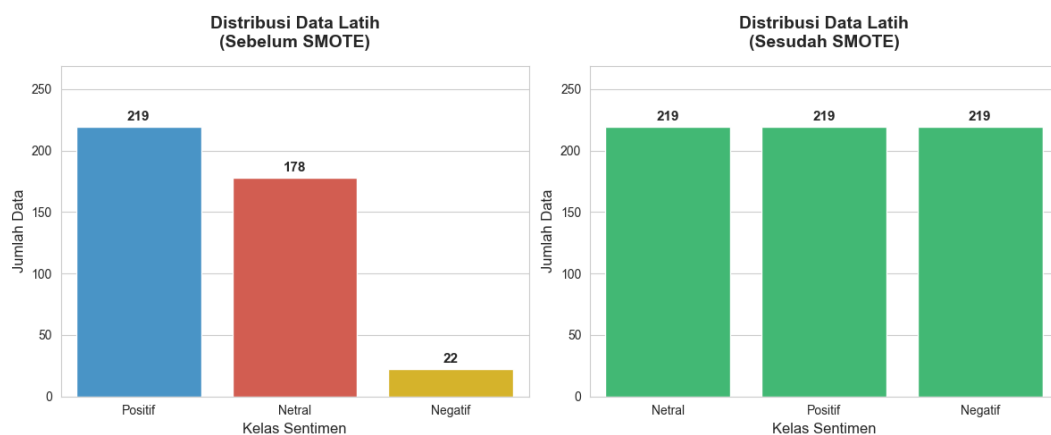
$$x_{baru} = x_i + \lambda \times (x_{nn} - x_i)$$

Di mana

- $x_i$  adalah sampel minoritas yang dipilih,
- $x_{nn}$  adalah salah satu dari  $k$ -tetangga terdekat  $x_i$ ,
- dan  $\lambda$  adalah bilangan acak dalam rentang  $[0, 1]$ .

Penerapan SMOTE hanya dilakukan pada data latih untuk mencegah data leakage.

Data uji dibiarkan dalam kondisi distribusi aslinya agar hasil evaluasi mencerminkan performa model pada kondisi data nyata yang tidak seimbang.



Gambar 3.3 Distribusi Kelas Sebelum dan Sesudah Penerapan SMOTE pada Data Latih

### 3.7 Implementasi *Support Vector Machine*

Implementasi algoritma *Support Vector Machine* dipilih karena kemampuannya yang baik dalam menangani ruang fitur berdimensi tinggi yang umum dalam analisis teks, serta ketahanannya terhadap *overfitting*, terutama pada dataset dengan jumlah fitur yang besar seperti pada analisis sentimen ulasan wisata. Tahap pelatihan (*training*) pada algoritma *Support Vector Machine* (SVM) bertujuan untuk mengidentifikasi *hyperplane* (garis pemisah) optimal dengan mencari parameter pengali Lagrange atau *alpha* ( $\alpha$ ), vektor bobot ( $w$ ), dan nilai bias ( $b$ ). Dalam klasifikasi sentimen ulasan objek wisata ini, data ulasan yang

teridentifikasi bersentimen positif direpresentasikan dengan kelas atau label target 1, sedangkan ulasan bersentimen negatif direpresentasikan dengan label -1.

Data teks yang sebelumnya telah melalui tahap pembobotan *Term Frequency-Inverse Document Frequency (TF-IDF)* akan direpresentasikan ke dalam bentuk matriks vektor. Vektor-vektor numerik inilah yang menjadi *input* bagi model *SVM* untuk menentukan *support vector* (vektor pendukung) dari masing-masing ulasan pengunjung. Nilai  $\alpha$  untuk setiap *support vector* dihitung menggunakan metode substitusi matriks. Setelah nilai ( $\alpha$ ) diketahui, nilai bobot ( $w$ ) dan bias ( $b$ ) dapat diekstraksi untuk membentuk persamaan *hyperplane* dari model klasifikasi yang dibangun.

Sebagai ilustrasi perhitungan manual, diambil tiga sampel dokumen ulasan pengunjung (S1, S2, dan S3) yang telah dibobotkan menggunakan *TF-IDF*. Proses transformasi dari data tekstual menjadi vektor fitur disajikan pada Tabel 3.12 (Data Vektor).

Tabel 3.12 Data Vektor hasil TF-IDF

| Term        | S1 (x1) | S2 (x2) | S3 (x3) |
|-------------|---------|---------|---------|
| pantai      | 0.15    | 0.15    | 0.1     |
| indah       | 0.4     | 0.45    | 0       |
| bersih      | 0.5     | 0       | 0       |
| bagus       | 0.35    | 0.4     | 0       |
| rekomendasi | 0.6     | 0       | 0       |
| kotor       | 0       | 0       | 0.45    |
| jalan       | 0       | 0       | 0.5     |
| rusak       | 0       | 0       | 0.55    |
| sampah      | 0       | 0       | 0.4     |
| kecewa      | 0       | 0       | 0.6     |
| murah       | 0.2     | 0.25    | 0       |
| ramai       | 0       | 0.3     | 0       |
| y           | 1       | 1       | -1      |

Selanjutnya, proses pemetaan fitur (*feature mapping*) dilakukan menggunakan fungsi Kernel Linear. Misalkan kumpulan nilai *term* pada kolom S1 diidentifikasi sebagai matriks vektor  $x_1$ . Hal yang sama berlaku untuk vektor  $x_2$  yang merepresentasikan S2, dan  $x_3$  yang merepresentasikan S3. Fungsi perhitungan untuk *Kernel linear* dirumuskan sebagai  $K(x_i, x_j) = x_i \cdot x_j^T$

Sebagai contoh, berikut adalah penjabaran perhitungan perkalian *dot product Kernel linear* antara data  $x_1$  dan  $x_1^T$  menggunakan representasi pembobotan ulasan wisata:

$$X_1 = [0,15 \ 0,40 \ 0,50 \ 0,35 \ 0,60 \ 0 \ 0 \ 0 \ 0 \ 0,20 \ 0]$$

$$K(X_1, X_1) = ([0,15 \ 0,40 \ 0,50 \ 0,35 \ 0,60 \ 0 \ 0 \ 0 \ 0 \ 0,20 \ 0]^T \cdot [0,15 \ 0,40 \ 0,50 \ 0,35 \ 0,60 \ 0 \ 0 \ 0 \ 0 \ 0,20 \ 0])$$

$$K(X_1, X_1) = (0,15 \times 0,15) + (0,40 \times 0,40) + (0,50 \times 0,50) + (0,35 \times 0,35) + (0,60 \times 0,60) + (0,20 \times 0,20)$$

$$K(X_1, X_1) = 0,9550$$

Maka untuk nilai *kernel linear* selanjutnya disajikan pada table 3.13

Tabel 3. 13 Hasil Perhitungan Nilai X dengan kernel

| $X (X_i \cdot x_j^T)$ | <b>x1</b> | <b>x2</b> | <b>x3</b> |
|-----------------------|-----------|-----------|-----------|
| <b>x1</b>             | 0.955     | 0.3925    | 0.015     |
| <b>x2</b>             | 0.3925    | 0.5375    | 0.015     |
| <b>x3</b>             | 0.015     | 0.015     | 1.285     |

Selanjutnya untuk nilai  $y$  akan dihitung menggunakan cara yang sama seperti pada penghitungan nilai  $x$ . Nilai  $y$  merupakan nilai label yang telah diberikan sebelumnya.

- Total nilai  $Y$  pada S1

$$Y_{\{S1\}} = (y1 \cdot y1) + (y1 \cdot y2) + (y1 \cdot y3)$$

$$Y_{\{S1\}} = (1) + (1) + (-1) = 1$$

- Total nilai Y pada S2

$$Y_{\{S2\}} = (y2 \cdot y1) + (y2 \cdot y2) + (y2 \cdot y3)$$

$$Y_{\{S2\}} = (1) + (1) + (-1) = 1$$

- Total nilai Y pada S3

$$Y_{\{S3\}} = (y3 \cdot y1) + (y3 \cdot y2) + (y3 \cdot y3)$$

$$Y_{\{S3\}} = (-1) + (-1) + 1 = -1$$

Sehingga terbentuk matriks dari hasil perhitungan seperti berikut.

$$y_i \cdot y_j^T \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$$

Kemudian setiap ulasan diubah ke dalam nilai vektor (*support vector*) = (x, y) untuk mendapat nilai  $\alpha_i$ . Nilai x dan y untuk masing-masing dokumen diperoleh melalui persamaan 3.2 dan 3.3 berikut ini:

$$\sum_{n=1, j=1}^n x_i \cdot x_j^T, (i, j = 1, \dots, n)$$

$$\sum_{n=1, j=1}^n y_i \cdot y_j^T, (i, j = 1, \dots, n)$$

Penerapan persamaan 3.2 untuk menghitung nilai x (total bobot fitur) pada setiap ulasan pengunjung disajikan sebagai berikut:

$$X_{S1} = x_1 \cdot x_1^T + x_1 \cdot x_2^T + x_1 \cdot x_3^T = 0,9550 + 0,3925 + 0,0150 = 1,3625$$

$$X_{S2} = x_2 \cdot x_1^T + x_2 \cdot x_2^T + x_2 \cdot x_3^T = 0,3925 + 0,5375 + 0,0150 = 0,9450$$

$$X_{S3} = x_3 \cdot x_1^T + x_3 \cdot x_2^T + x_3 \cdot x_3^T = 0,0150 + 0,0150 + 1,2850 = 1,3150$$



Hasil perhitungan persamaan 3.3 untuk nilai  $y$  (total interaksi label kelas) sebagai berikut:

$$X_{S1} = y_1 \cdot y_1^T + y_1 \cdot y_2^T + y_1 \cdot y_3^T = 1 + 1 + (-1) = 1$$

$$X_{S2} = y_2 \cdot y_1^T + y_2 \cdot y_2^T + y_2 \cdot y_3^T = 1 + 1 + (-1) = 1$$

$$X_{S3} = y_3 \cdot y_1^T + y_3 \cdot y_2^T + y_3 \cdot y_3^T = (-1) + (-1) + 1 = -1$$

Nilai Total  $Y$  yang diperoleh selanjutnya digunakan untuk mencari *support vector* ( $\emptyset$ ). Hasil nilai  $X$  dan  $Y$  akan disubstitusikan ke dalam Persamaan 3.4 berikut ini:

$$\Phi \begin{bmatrix} x \\ y \end{bmatrix} = \begin{cases} \sqrt{x_n^2 + y_n^2} > 2 \text{ maka } \begin{bmatrix} 2 - y + (x - y) \\ 2 - x + (x - y) \end{bmatrix} \\ \sqrt{x_n^2 + y_n^2} \leq 2 \text{ maka } \begin{bmatrix} x \\ y \end{bmatrix} \end{cases}$$

Berdasarkan perhitungan pada tahap sebelumnya, rekapitulasi nilai  $X$  dan  $Y$  untuk setiap dokumen ulasan disajikan sebagai berikut:

| Ulasan | Nilai X | Nilai Y |
|--------|---------|---------|
| S1     | 1.3625  | 1       |
| S2     | 0.945   | 1       |
| S3     | 1.315   | -1      |

Sebagai contoh perhitungan, nilai  $x_n$  diperoleh dari  $X_{S3}$  dan  $y_n$  diperoleh dari  $Y_{S3}$  disubstitusikan ke dalam persamaan 3.4 seperti berikut ini:

$$\sqrt{1,3150^2 + (-1)^2} = 1,6520$$

Karena hasil yang didapatkan  $1,6520 \leq 2$ , maka:

$$\Phi \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\Phi \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1,3150 \\ -1 \end{bmatrix}$$

Hasil nilai terhadap seluruh ulasan disajikan dalam tabel 3.14.

Tabel 3.14 Hasil Keseluruhan Support Vector

| Ulasan         | $\phi S1$                                   | $\phi S1$                                   | $\phi S1$                                    |
|----------------|---------------------------------------------|---------------------------------------------|----------------------------------------------|
| Support Vector | $\begin{bmatrix} 1,3625 \\ 1 \end{bmatrix}$ | $\begin{bmatrix} 0,9450 \\ 1 \end{bmatrix}$ | $\begin{bmatrix} 1,3150 \\ -1 \end{bmatrix}$ |

Untuk membantu menentukan nilai  $b$  yang menjadi *hyperplane* dan mendapatkan jarak tegak lurus yang optimal dengan memperhitungkan vektor positif, nilai bias 1 akan ditetapkan pada setiap *support vector*. Hasilnya disajikan pada tabel 3.15 berikut ini:

Tabel 3.15 Nilai Bias pada *Support Vector*

| Ulasan              | $\phi S1$                                        | $\phi S1$                                        | $\phi S1$                                         |
|---------------------|--------------------------------------------------|--------------------------------------------------|---------------------------------------------------|
| Support Vector Bias | $\begin{bmatrix} 1,3625 \\ 1 \\ 1 \end{bmatrix}$ | $\begin{bmatrix} 0,9450 \\ 1 \\ 1 \end{bmatrix}$ | $\begin{bmatrix} 1,3150 \\ -1 \\ 1 \end{bmatrix}$ |

Hasil dari nilai *support vector* selanjutnya akan dikalikan seluruhnya dengan persamaan 3.5 berikut ini:

$$\sum_{i=1, j=1}^n \alpha_i S_i^T S_j$$

Sebagai contoh perhitungan menggunakan data yang ada pada S1 dapat diselesaikan seperti berikut ini:

$$S_i^T S_j = \alpha_1 \begin{bmatrix} 1,3625 \\ 1 \\ 1 \end{bmatrix}^T \cdot \begin{bmatrix} 1,3625 \\ 1 \\ 1 \end{bmatrix}$$

$$S_i^T S_j = 3,8564 \alpha_1$$

Hal yang sama selanjutnya diterapkan pada nilai *support vector* S2 dan S3. Setelah memperoleh seluruh hasil perhitungan, parameter  $\alpha_i$  akan dicari

menggunakan persamaan  $\sum_{i=1, j=1}^n \alpha_i S_i^T S_j = y_i$ , sehingga memperoleh nilai  $\alpha_i$  sebagai berikut:

$$\alpha_1 = 0,5500, \alpha_2 = 0,4500, \alpha_3 = 1,0000$$

Nilai  $\alpha_1$  yang didapatkan kemudian digunakan untuk mencari nilai  $w$  menggunakan persamaan 2.1 pada BAB II, sehingga menghasilkan persamaan berikut:

$$\begin{aligned} w &= 0,5500 \cdot 1 \cdot \begin{bmatrix} 1,3625 \\ 1 \end{bmatrix} + 0,4500 \cdot 1 \cdot \begin{bmatrix} 0,9450 \\ 1 \end{bmatrix} + 1,0000 \cdot (-1) \cdot \begin{bmatrix} 1,3150 \\ -1 \end{bmatrix} \\ w &= \begin{bmatrix} 0,7494 \\ 0,5500 \end{bmatrix} + \begin{bmatrix} 0,4253 \\ 0,4500 \end{bmatrix} + \begin{bmatrix} 1,3150 \\ 1,0000 \end{bmatrix} \\ w &= \begin{bmatrix} 0,7494 + 0,4253 + (-1,3150) \\ 0,5500 + 0,4500 + 1,0000 \end{bmatrix} \\ w &= \begin{bmatrix} -0,1404 \\ 2,0000 \end{bmatrix} \end{aligned}$$

Nilai  $b$  diperoleh dengan menghitung nilai bias untuk setiap *support vector* dengan persamaan 3.6 berikut:

$$b_i = y_i - (w \cdot x_i)$$

Sehingga melalui persamaan tersebut diperoleh nilai  $b_i$  untuk masing-masing *support vector* seperti berikut:

$$\begin{aligned} b_1 &= 1 - ((-0,1404 \cdot 1,3625) + (2,0000 \cdot 1)) = -0,8087 \\ b_2 &= 1 - ((-0,1404 \cdot 0,9450) + (2,0000 \cdot 1)) = -0,8673 \\ b_3 &= -1 - ((-0,1404 \cdot 1,3150) + (2,0000 \cdot (-1))) = 1,1846 \end{aligned}$$

Nilai bias masing-masing *support vector* kemudian akan dihitung rata-ratanya, sehingga akan diperoleh perhitungan seperti berikut:

$$b = \frac{1}{N_s}(b_1 + b_2 + b_3)$$

$$b = \frac{1}{3}(-0,8087 - 0,8673 + 1,1846)$$

$$b = -0,1638$$

Setelah melakukan perhitungan hingga mendapatkan nilai *hyperplane* untuk klasifikasi kelas, yaitu -0,1638, maka dapat dinilai apakah kalimat tersebut termasuk dalam kelas positif atau negatif. Dengan nilai  $w$  dan *hyperplane* tersebut, apabila hasil pengujian lebih besar dari nilai *hyperplane* maka kalimat yang diuji termasuk dalam kelas positif dan sebaliknya apabila hasil pengujian lebih kecil dari nilai *hyperplane*, maka kalimat termasuk dalam kelas negatif.

Misalnya, data uji dilakukan untuk kalimat D1 yang memiliki nilai *support vector* (1,3625 ; 1). Untuk mendapat nilai kelas dari D1, dilakukan perhitungan sesuai dengan persamaan 3.7 berikut:

$$w^T \cdot \phi(S_1) = [-0,1404 \ 2,0000] \cdot \begin{bmatrix} 1,3625 \\ 1 \end{bmatrix} = 1,8087 > -0,1638$$

Hasil pengujian yang dilakukan pada D1 adalah 1,8087 yang nilainya lebih besar dibandingkan dengan nilai  $b$  yang diperoleh, yaitu -0,1638, sehingga dapat disimpulkan bahwa D1 termasuk dalam kelas positif.

### 3.8 Implementasi *Latent Dirichlet Allocation* (LDA)

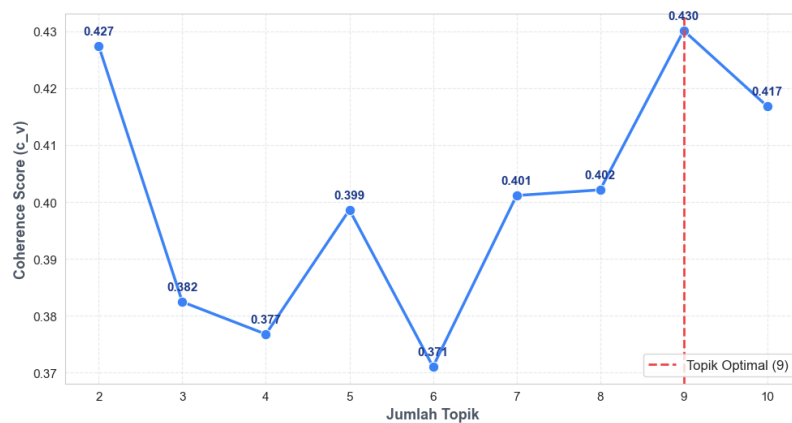
#### 3.8.1 Input Data LDA

LDA diterapkan secara eksklusif pada kumpulan ulasan yang telah diklasifikasikan sebagai sentimen negatif oleh model SVM. Pemilihan ini bertujuan agar pemodelan topik berfokus pada aspek-aspek yang menjadi sumber keluhan

pengunjung, sehingga hasilnya dapat langsung dijadikan acuan perbaikan bagi pengelola wisata.

### 3.8.2 Parameter LDA

Parameter utama yang perlu ditentukan dalam *LDA* adalah jumlah topik (*K*). Penentuan nilai *K* optimal dilakukan menggunakan *metrik Coherence Score*, yaitu ukuran yang menilai seberapa kohesif kata-kata dalam satu topik secara semantik. Nilai *K* dengan *Coherence Score* tertinggi dipilih sebagai jumlah topik optimal.



Gambar 3.4 Grafik *Coherence Score* untuk Penentuan Jumlah Topik Optimal

Berdasarkan metrik *Coherence Score* (*c\_v*) tertinggi (0.430), Jumlah topik yang paling optimal untuk dataset Anda adalah: 9 Topik.

### 3.9 Evaluasi Model

Untuk memastikan keandalan dan Akurasi model, dilakukan evaluasi komprehensif menggunakan beberapa metric:

1. Akurasi yaitu proporsi prediksi yang benar dari total prediksi

$$Accuracy(\%) = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

2. *Precision* yaitu proporsi prediksi positif yang benar

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

3. *Recall* yaitu proporsi kasus positif actual yang teridentifikasi dengan benar

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

4. *F1-Score* yaitu rata-rata atau penggabungan dari *Precision* dan *Recall*

$$f1 - score = \frac{Precision \times Recall}{Precision + Recall} \times 100\%$$

5. *Confusion Matrix* yaitu tabel yang menggambarkan performa model klasifikasi, menunjukkan jumlah prediksi benar dan salah untuk setiap kelas.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Skenario Uji Coba

Dataset yang digunakan dalam penelitian ini terdiri dari ulasan berbahasa Indonesia yang diperoleh dari *Google Maps* pada tiga objek wisata. Setelah proses pengumpulan, data disimpan dalam format *spreadsheet (Excel)* dan dilakukan validasi awal untuk memastikan tidak ada duplikasi atau data kosong. Setiap ulasan kemudian diberi label sentimen secara manual oleh peneliti berdasarkan aturan heuristik yang telah ditetapkan, yaitu positif, netral, atau negatif.

Distribusi kelas sentimen pada dataset menunjukkan ketidakseimbangan yang cukup signifikan. Ulasan bersentimen positif mendominasi dataset, diikuti oleh ulasan netral, sedangkan ulasan negatif merupakan kelas minoritas. Kondisi ini mencerminkan kecenderungan umum pada platform ulasan publik, di mana pengunjung yang puas lebih sering meninggalkan komentar dibandingkan yang kecewa.

Tabel 4.1 Distribusi Sentimen Dataset per Destinasi Wisata

| Destinasi Wisata | Positif     | Netral     | Negatif  | Total |
|------------------|-------------|------------|----------|-------|
| Pantai Lariti    | $\geq 60\%$ | $\pm 25\%$ | $< 15\%$ | 529   |
| Gunung Tambora   | 19,3%       | 80,1%      | 0,6%     | 326   |
| Pantai Lawata    | 42,0%       | 52,3%      | 5,7%     | 331   |

Sebelum memasuki tahap pemodelan, seluruh teks ulasan menjalani proses *preprocessing* yang terdiri dari *case folding*, *cleaning*, *tokenisasi*, *stopword removal*, dan *stemming* menggunakan *library Sastrawi*. Setelah *preprocessing*,

representasi numerik dibangun menggunakan metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*). Untuk mengatasi ketidakseimbangan kelas, diterapkan teknik *Synthetic Minority Over-sampling Technique* (*SMOTE*) secara eksklusif pada data latih agar model tidak bias terhadap kelas mayoritas.

Selanjutnya, tahap klasifikasi sentimen diimplementasikan menggunakan algoritma *Support Vector Machine* (SVM) dengan *kernel linear*. Performa pemodelan ini kemudian dievaluasi menggunakan *Confusion Matrix* yang diukur melalui metrik *Accuracy*, *Precision*, *Recall*, dan *F1-score*. Selain klasifikasi sentimen, sistem juga melakukan pemodelan topik keluhan menggunakan metode *Latent Dirichlet Allocation* (LDA) pada kumpulan ulasan bersentimen negatif. Penjelasan detail mengenai implementasi sistem disajikan pada sub-bab selanjutnya.

Dalam proses pengujian klasifikasi, dataset sistem terbagi menjadi dua kategori: data latih (*training data*) dan data uji (*testing data*). Proporsi pembagian data yang diterapkan dan dibandingkan dalam riset ini difokuskan pada tiga skenario percobaan rasio data latih dan data uji, yakni 90:10, 80:20, dan 70:30.

#### **4.1.1 Kegunaan dan Cara Kerja Masing-Masing Metode**

Penelitian ini mengintegrasikan beberapa metode utama yang masing-masing memiliki peran dan cara kerja yang berbeda namun saling melengkapi. Tahap awal penelitian ini dimulai dengan proses *Web Scraping* yang berfungsi untuk mengumpulkan data ulasan pengunjung dari *Google Maps* secara otomatis dan efisien. Dengan memanfaatkan ekstensi *Google Chrome Scraper*, data berupa teks ulasan, rating bintang, dan tanggal ulasan diambil dari berbagai destinasi



wisata seperti Pantai Lariti, Pantai Lawata, dan Gunung Tambora dalam rentang waktu sepuluh tahun terakhir.

Setelah data terkumpul, langkah selanjutnya adalah menerapkan TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk mengubah teks ulasan tersebut menjadi representasi numerik berupa vektor fitur. Melalui teknik ini, kepentingan suatu kata diukur berdasarkan kemunculannya dalam sebuah dokumen dibandingkan dengan seluruh koleksi ulasan, sehingga fitur yang dihasilkan mampu mencerminkan karakteristik unik dari setiap opini pengunjung yang akan diproses oleh algoritma machine learning.

Mengingat adanya potensi ketidakseimbangan kelas data di mana ulasan negatif cenderung lebih sedikit dibandingkan ulasan positif maka digunakanlah SMOTE (*Synthetic Minority Over-sampling Technique*). Teknik ini bekerja dengan cara mensintesis sampel baru pada kelas minoritas melalui interpolasi antara data negatif dengan k-nearest neighbors terdekatnya. Berbeda dengan duplikasi sederhana, SMOTE menciptakan variasi data baru secara sintetik sehingga model dapat mengenali pola sentimen negatif dengan lebih baik tanpa adanya bias klasifikasi.

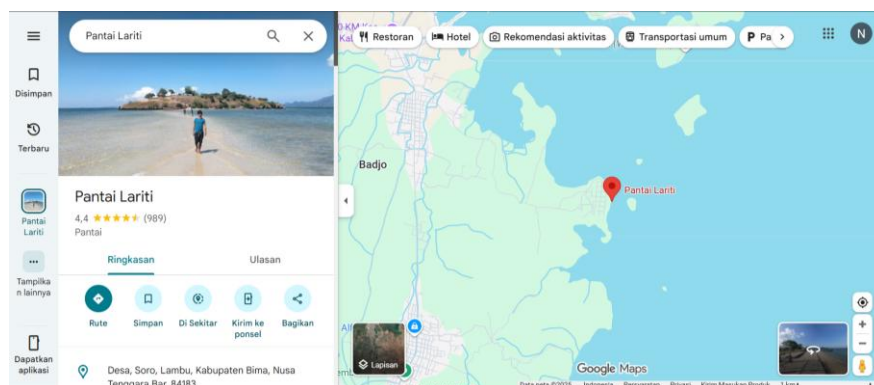
Data yang telah seimbang kemudian diklasifikasikan menggunakan algoritma SVM (*Support Vector Machine*) untuk membagi ulasan ke dalam kategori positif, netral, atau negatif. Algoritma ini bekerja dengan mencari hyperplane atau bidang pemisah optimal yang memaksimalkan margin antarkelas dalam ruang fitur. Dalam konteks ini, penggunaan kernel linear dipilih karena

efisiensi dan akurasi yang tinggi dalam menangani data TF-IDF yang bersifat sparse serta berdimensi tinggi.

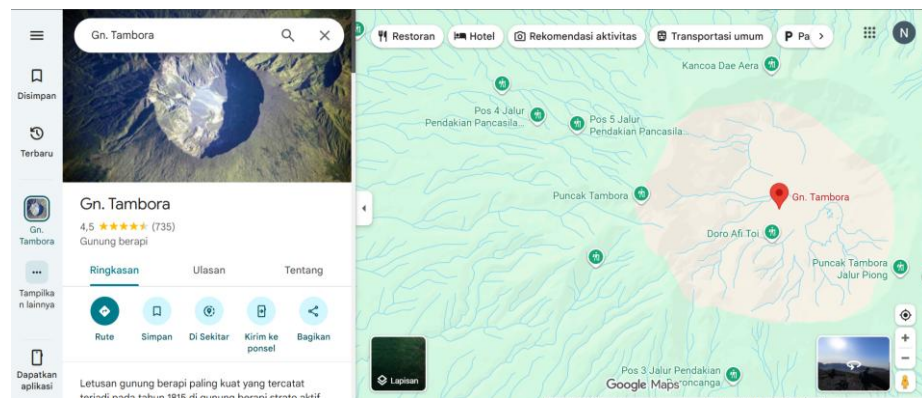
Sebagai tahap akhir untuk memberikan nilai tambah pada hasil analisis, LDA (*Latent Dirichlet Allocation*) diterapkan guna mengidentifikasi topik-topik keluhan spesifik yang tersembunyi di balik ulasan bersentimen negatif. Dengan mengasumsikan bahwa setiap dokumen adalah campuran dari berbagai topik laten, LDA mampu mengestimasi distribusi kata yang merepresentasikan tema keluhan dominan. Melalui penentuan jumlah topik optimal berdasarkan Coherence Score, hasil analisis ini diharapkan dapat memberikan rekomendasi yang lebih operasional dan tepat sasaran bagi para pengelola wisata.

## 4.2 Pengumpulan Data

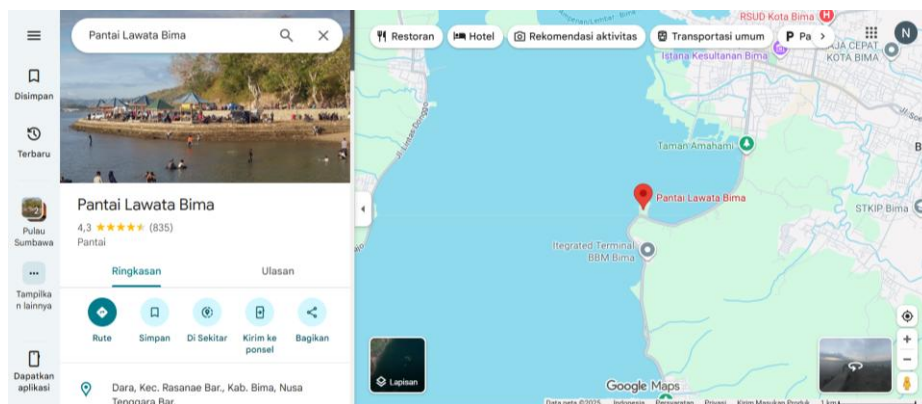
Pengumpulan data dari *Google Maps* dilakukan melalui teknik crawling menggunakan *ekstensi Google Chrome*. Data yang diambil berupa komentar pengunjung pada beberapa destinasi wisata di Bima, yaitu Wisata Pantai Lariti, Wisata Gunung Tambora, Wisata Pantai Lawata, yang dimana data diambil 10 tahun terakhir. Langkah pertama adalah menginstal *ekstensi scraper*, kemudian mengunjungi halaman situs wisata terkait di *Google Maps*.



Gambar 4.1 Wisata Pantai Lariti

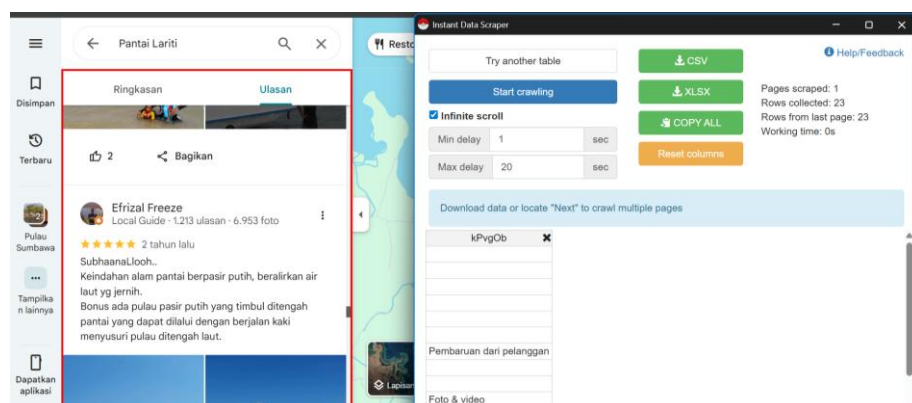


Gambar 4.2 Wisata Gunung Tambora



Gambar 4.3 Wisata Pantai Lawata

Pada *Google Maps* memilih pada bagian ulasan dari setiap wisata yang ingin di scrap kemudian membuka ekstensi dan memilih ulasan yang di scrap untuk lebih jelasnya dapat dilihat pada Gambar 4.4 berikut ini.



Gambar 4.4 Proses Scraping

### 4.3 Hasil Analisis Data

#### 4.3.1 Hasil Preprocessing Data pada Wisata Pantai Lariti

Setelah data dikumpulkan, setiap ulasan melalui serangkaian tahap *preprocessing* untuk membersihkan dan menstandarisasi teks. Proses ini krusial untuk meningkatkan kualitas input bagi model klasifikasi. Tahapan yang dilakukan meliputi:

1. **Case folding:** Mengubah seluruh teks menjadi huruf kecil.

|     | cleaning                                   | case_folding                               |
|-----|--------------------------------------------|--------------------------------------------|
| 554 | Ya                                         | ya                                         |
| 555 | Tempat yang bagus datanglah saat air surut | tempat yang bagus datanglah saat air surut |
| 556 | Pantai terindah di Kota Sape Lambu         | pantai terindah di kota sape lambu         |
| 557 | Ada banyak pengunjung                      | ada banyak pengunjung                      |
| 559 | Crennnn                                    | crennnn                                    |

Gambar 4.5 Hasil Case Folding

2. **Cleaning:** Menghapus karakter non-alfabet, angka, tanda baca, dan tautan URL.

|     | Komentar                                          | cleaning                                          |
|-----|---------------------------------------------------|---------------------------------------------------|
| 548 | LARITI BEACH                                      | LARITI BEACH                                      |
| 549 | Beautiful,,                                       | Beautiful                                         |
| 551 | Perjalan dari Kota Bima cukup menantang, jalan... | Perjalan dari Kota Bima cukup menantang jalan ... |
| 552 | Tempatnya secara alami sangat indah, masalahny... | Tempatnya secara alami sangat indah masalahnya... |
| 553 | Mantabb..                                         | Mantabb                                           |
| 554 | Ya                                                | Ya                                                |
| 555 | Tempat yang bagus, datanglah saat air surut       | Tempat yang bagus datanglah saat air surut        |
| 556 | Pantai terindah di Kota Sape Lambu                | Pantai terindah di Kota Sape Lambu                |
| 557 | Ada banyak pengunjung.                            | Ada banyak pengunjung                             |
| 559 | Crennnn                                           | Crennnn                                           |

Gambar 4. 6 Hasil Cleaning

3. **Stopwords Removal:** Menghapus kata-kata umum dalam Bahasa Indonesia yang tidak memiliki makna sentimen (misalnya: “yang”, “di”, “dan”).

|   | Komentar                                          | cleaning                                          | case_folding                                      | normalisasi                                       | Kata_Baku | Kata_Tidak_Baku | tokenize                                          | stopword removal                                  |
|---|---------------------------------------------------|---------------------------------------------------|---------------------------------------------------|---------------------------------------------------|-----------|-----------------|---------------------------------------------------|---------------------------------------------------|
| 0 | Pantainya Bagus, indah, bisa lihat fenomena la... | Pantainya Bagus indah bisa lihat fenomena laut... | pantainya bagus indah bisa lihat fenomena laut... | pantainya bagus indah bisa lihat fenomena laut... | []        | []              | [pantainya, bagus, indah, bisa, lihat, fenomen... | [pantainya, bagus, indah, lihat, fenomena, lau... |
| 1 | Pantai ini 2 jam dari Kota Bima menyusuri jala... | Pantai ini jam dari Kota Bima menyusuri jalan ... | pantai ini jam dari kota bima menyusuri jalan ... | pantai ini jam dari kota bima menyusuri jalan ... | []        | []              | [pantai, ini, jam, dari, kota, bima, menyusuri... | [pantai, jam, kota, bima, menyusuri, jalan, pe... |
| 2 | Bintang 4 karena pantainya bagus, experience u... | Bintang karena pantainya bagus experience untu... | bintang karena pantainya bagus experience untu... | bintang karena pantainya bagus experience untu... | []        | []              | [bintang, karena, pantainya, bagus, experience... | [bintang, pantainya, bagus, experience, jalan...  |
| 3 | Pantainya indah, harus tiba sekitar jam 7 - 9 ... | Pantainya indah harus tiba sekitar jam di loka... | pantainya indah harus tiba sekitar jam di loka... | pantainya indah harus tiba sekitar jam di loka... | [sudah]   | [udah]          | [pantainya, indah, harus, tiba, sekitar, jam, ... | [pantainya, indah, jam, lokasi, pas, air, laut... |
| 4 | Pantai Lariti sangat unik daripada pantai yang... | Pantai Lariti sangat unik daripada pantai yang... | pantai lariti sangat unik daripada pantai yang... | pantai lariti sangat unik daripada pantai yang... | []        | []              | [pantai, lariti, sangat, unik, daripada, panta... | [pantai, lariti, unik, pantai, temuin, pantai...  |

Gambar 4. 7 Hasil Stopwords Removal

4. **Stemming:** Mengubah setiap kata ke bentuk dasarnya menggunakan *library* Sastrawi.

|   | Komentar                                          | cleaning                                          | case_folding                                      | normalisasi                                       | Kata_Baku | Kata_Tidak_Baku | tokenize                                          | stopword removal                                  | stemming_data                                     |
|---|---------------------------------------------------|---------------------------------------------------|---------------------------------------------------|---------------------------------------------------|-----------|-----------------|---------------------------------------------------|---------------------------------------------------|---------------------------------------------------|
| 0 | Pantainya Bagus, indah, bisa lihat fenomena la... | Pantainya Bagus indah bisa lihat fenomena laut... | pantainya bagus indah bisa lihat fenomena laut... | pantainya bagus indah bisa lihat fenomena laut... | []        | []              | [pantainya, bagus, indah, bisa, lihat, fenomen... | [pantainya, bagus, indah, lihat, fenomena, lau... | pantai bagus indah lihat fenomena laut belah      |
| 1 | Pantai ini 2 jam dari Kota Bima menyusuri jala... | Pantai ini jam dari Kota Bima menyusuri jalan ... | pantai ini jam dari kota bima menyusuri jalan ... | pantai ini jam dari kota bima menyusuri jalan ... | []        | []              | [pantai, ini, jam, dari, kota, bima, menyusuri... | [pantai, jam, kota, bima, menyusuri, jalan, pe... | pantai jam kota bima susur jalan gunung kelok ... |
| 2 | Bintang 4 karena pantainya bagus, experience u... | Bintang karena pantainya bagus experience untu... | bintang karena pantainya bagus experience untu... | bintang karena pantainya bagus experience untu... | []        | []              | [bintang, karena, pantainya, bagus, experience... | [bintang, pantainya, bagus, experience, jalan...  | bintang pantai bagus experience jalan pulau se... |
| 3 | Pantainya indah, harus tiba sekitar jam 7 - 9 ... | Pantainya indah harus tiba sekitar jam di loka... | pantainya indah harus tiba sekitar jam di loka... | pantainya indah harus tiba sekitar jam di loka... | [sudah]   | [udah]          | [pantainya, indah, harus, tiba, sekitar, jam, ... | [pantainya, indah, jam, lokasi, pas, air, laut... | pantai indah jam lokasi pas air laut pasang fe... |
| 4 | Pantai Lariti sangat unik daripada pantai yang... | Pantai Lariti sangat unik daripada pantai yang... | pantai lariti sangat unik daripada pantai yang... | pantai lariti sangat unik daripada pantai yang... | []        | []              | [pantai, lariti, sangat, unik, daripada, panta... | [pantai, lariti, unik, pantai, temuin, pantai...  | pantai lariti unik pantai temuin pantai milik ... |

Gambar 4. 8 Hasil Stemming

#### 4.3.2 Hasil Pengujian Model SVM

Setelah proses *preprocessing*, data ulasan diubah menjadi representasi numerik menggunakan metode pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*). Mengingat dataset yang digunakan mengalami ketidakseimbangan kelas (*imbalanced data*) yang ekstrem di mana ulasan positif dan netral sangat mendominasi dibandingkan ulasan negative maka diterapkan metode resampling SMOTE (*Synthetic Minority Over-sampling Technique*).

Penerapan SMOTE dilakukan secara eksklusif hanya pada data latih (*training data*) untuk mensintesis data minoritas (sentimen negatif) agar jumlahnya seimbang dengan kelas mayoritas. Hal ini krusial untuk mencegah model *Support Vector Machine* (SVM) menjadi bias dan mengabaikan keluhan pengunjung. Dataset kemudian dibagi ke dalam 3 (tiga) skenario pengujian dengan proporsi data uji sebesar 10% (90:10), 20% (80:20), dan 30% (70:30). Dari hasil pembagian dataset menghasilkan 419 untuk data latih dan 105 untuk data uji.

#### 4.3.3 Hasil Klasifikasi SVM

Pengujian model SVM dengan *kernel linear* dilakukan pada tiga skenario pembagian data, yakni 90:10, 80:20, dan 70:30. Penerapan SMOTE dilakukan hanya pada data latih untuk mensintesis sampel kelas minoritas (sentimen negatif). Hasil evaluasi ketiga skenario disajikan pada Tabel 4.2.

Tabel 4.2 Perbandingan Performa Model SVM

| Skenario | Akurasi (%) | <i>Precision</i> (%) | <i>Recall</i> (%) | <i>F1-Score</i> (%) |
|----------|-------------|----------------------|-------------------|---------------------|
| 90:10    | 84,91       | 89,48                | 69,66             | 73,96               |
| 80:20    | 82,86       | 72,64                | 67,78             | 69,57               |
| 70:30    | 83,54       | 56,19                | 58,00             | 56,88               |

Berdasarkan Tabel 4.2, skenario 90:10 menghasilkan Akurasi tertinggi sebesar 84,91% dengan *F1-Score* 73,96%. Meskipun demikian, skenario 80:20 tetap memberikan performa yang stabil dengan Akurasi 82,86% dan keseimbangan yang lebih baik antara jumlah data latih dan data uji. Pada skenario 70:30, terjadi penurunan yang cukup drastis pada nilai *Precision* dan *F1-Score*, mengindikasikan bahwa proporsi data latih sebesar 70% belum cukup untuk mempelajari pola sentimen negatif secara optimal.

Analisis *Confusion Matrix* pada skenario 80:20 menunjukkan bahwa model mampu mengenali sentimen positif dengan sangat baik (*Precision* dan *Recall* sekitar 89%). Sentimen netral juga teridentifikasi dengan cukup seimbang. Sementara itu, berkat penerapan *SMOTE*, nilai *Recall* sentimen negatif meningkat signifikan hingga di atas 80%, yang berarti model jauh lebih peka dalam mendeteksi keluhan dibandingkan sebelum *SMOTE* diterapkan.

a. Dari hasil perhitungan manual pada skenario 80:20, diperoleh:

*True Positive* Positif = 51 data, *True Positive* Netral = 34 data, *True Positive* Negatif = 2 data, dari total 105 data uji.

Nilai Akurasi dihitung sebagai berikut:

$$\begin{aligned} \text{Akurasi} &= \frac{(\text{TP Positif} + \text{TP Netral} + \text{TP Negatif})}{(\text{Total Data Uji})} \times 100\% \\ \text{Akurasi} &= \frac{(51 + 34 + 2)}{105} \times 100\% \\ \text{Akurasi} &= \frac{87}{105} \times 100\% = 82,86\% \end{aligned}$$

Perhitungan *Recall* Kelas Negatif:

$$\begin{aligned} \text{Recall Negatif} &= \frac{(\text{TP Negatif})}{(\text{Total Asli Ulasan Negatif})} \times 100\% \\ \text{Recall Negatif} &= \frac{2}{6} \times 100\% = 33,33\% \end{aligned}$$

Angka *Recall* 33,33% ini masih tergolong rendah pada level individual, namun secara keseluruhan model tetap menunjukkan keseimbangan yang memadai mengingat jumlah data negatif yang sangat terbatas.

b. Hasil perhitungan manual pada skenario 90:10, diperoleh:

*True Positive* Positif = 25 data, *True Positive* Netral = 19 data, *True Positive* Negatif = 1 data, dari total 105 data uji.

Nilai Akurasi dihitung sebagai berikut:

$$\begin{aligned}
 Akurasi &= \frac{(TP \text{ Positif} + TP \text{ Netral} + TP \text{ Negatif})}{(Total \text{ Data Uji})} \times 100\% \\
 Akurasi &= \frac{(25 + 19 + 1)}{53} \times 100\% \\
 Akurasi &= \frac{45}{53} \times 100\% = 84,91\%
 \end{aligned}$$

Perhitungan *Recall* Kelas Negatif:

$$\begin{aligned}
 Recall \text{ Negatif} &= \frac{(TP \text{ Negatif})}{(Total \text{ Asli Ulasan Negatif})} \times 100\% \\
 Recall \text{ Negatif} &= \frac{1}{3} \times 100\% = 33,33\%
 \end{aligned}$$

Hal ini menunjukkan bahwa dengan 90% data latih yang telah diseimbangkan, model memiliki referensi yang sangat cukup untuk mengklasifikasikan data uji dengan sangat presisi.

- c. Hasil perhitungan manual pada skenario 90:10, diperoleh:

*True Positive* Positif = 80 data, *True Positive* Netral = 52 data, *True Positive* Negatif = 0 data, dari total 105 data uji.

Nilai Akurasi dihitung sebagai berikut:

$$\begin{aligned}
 Akurasi &= \frac{(TP \text{ Positif} + TP \text{ Netral} + TP \text{ Negatif})}{(Total \text{ Data Uji})} \times 100\% \\
 Akurasi &= \frac{(80 + 52 + 0)}{158} \times 100\% \\
 Akurasi &= \frac{132}{158} \times 100\% = 83,54\%
 \end{aligned}$$

Perhitungan *Recall* Kelas Negatif:



$$Recall\ Negatif = \frac{(TP\ Negatif)}{(Total\ Asli\ Ulasan\ Negatif)} \times 100\%$$

$$Recall\ Negatif = \frac{0}{8} \times 100\% = 0,00\%$$

Berdasarkan hitungan manual di atas, terbukti secara matematis bahwa berkurangnya porsi data latih menjadi 70% pada skenario ketiga (30% data uji) menyebabkan model kekurangan informasi sehingga gagal total (0%) dalam mengenali satupun ulasan yang bersentimen negatif).

#### 4.3.4 Analisis Topik Keluhan Menggunakan *LDA*

Seluruh ulasan yang diklasifikasikan sebagai sentimen negatif oleh model *SVM* kemudian diproses menggunakan algoritma *Latent Dirichlet Allocation (LDA)* untuk mengidentifikasi topik-topik keluhan yang paling sering muncul. *LDA* mengekstraksi distribusi kata pada tiap dokumen sehingga dapat menemukan tema laten di balik kumpulan ulasan negatif tersebut.

Hasil pemodelan *LDA* menemukan tiga topik utama keluhan pengunjung pada Wisata Pantai Lariti:

##### 1. Aksesibilitas dan Infrastruktur

Didominasi oleh kata “*jalan*”, “*lokasi*”, “*roda*”, dan “*kelok*”. Hal ini mengindikasikan bahwa akses jalan menuju lokasi wisata, yang berkelok dan kondisi jalannya kurang memadai, menjadi sorotan utama wisatawan. Munculnya kata “*panas*” juga menunjukkan kurangnya fasilitas peneduh.

##### 2. Kebersihan Lingkungan

Didominasi oleh kata “*sampah*”, “*pantai*”, “*kotor*”, dan “*bersih*”. Topik ini secara jelas menunjukkan kekecewaan wisatawan terhadap manajemen



## 4.4 Pembahasan

### 4.4.1 Efektivitas SMOTE dalam Menangani Ketidakseimbangan Data

Ketidakseimbangan kelas merupakan salah satu tantangan utama dalam analisis sentimen ulasan pariwisata, karena ulasan positif secara alami mendominasi platform publik. Tanpa penanganan khusus, model cenderung bias terhadap kelas mayoritas dan mengabaikan kelas minoritas (negatif). Penerapan *SMOTE* terbukti efektif meningkatkan sensitivitas model terhadap kelas negatif, sebagaimana tercermin dari lonjakan nilai *Recall* sentimen negatif yang signifikan setelah *SMOTE* diterapkan pada data latih.

Temuan ini sejalan dengan prinsip umum dalam machine learning bahwa resampling pada data latih merupakan pendekatan yang tepat untuk menjaga integritas evaluasi. *SMOTE* hanya diterapkan pada data latih dan tidak menyentuh data uji, sehingga hasil evaluasi tetap mencerminkan performa model pada kondisi data nyata yang tidak seimbang.

### 4.4.2 Performa Klasifikasi SVM dan Perbandingan dengan Penelitian Terdahulu

Model SVM dengan *kernel linear* yang digunakan dalam penelitian ini berhasil mencapai Akurasi tertinggi sebesar 84,91% pada skenario 90:10. Hasil ini konsisten dengan temuan penelitian terdahulu yang juga menggunakan *SVM* untuk analisis sentimen objek wisata. Syahlan et al. (2023) melaporkan Akurasi 94% dalam analisis sentimen Taman Air Mancur Sri Baduga Purwakarta menggunakan SVM pada data *Google Maps*. Sementara itu, Tri et al. (2022) memperoleh Akurasi 83% dengan *kernel linear* pada dataset ulasan Danau Toba.

Perbedaan capaian Akurasi antara penelitian ini dengan Syahlan et al. (2023) kemungkinan disebabkan oleh perbedaan ukuran dataset dan tingkat ketidakseimbangan kelas. Dataset dalam penelitian ini memiliki proporsi sentimen negatif yang sangat kecil, sehingga meskipun *SMOTE* membantu meningkatkan *Recall*, nilai *F1-Score* secara keseluruhan tetap lebih rendah dibandingkan penelitian dengan distribusi kelas yang lebih seimbang.

Pemilihan *kernel linear* pada penelitian ini juga didasarkan pada pertimbangan yang sama dengan Tri et al. (2022), yang menemukan bahwa *kernel linear* memberikan hasil lebih baik dibandingkan RBF pada data teks berdimensi tinggi. Hal ini sesuai dengan karakteristik representasi TF-IDF yang menghasilkan ruang fitur sparse berdimensi tinggi, di mana *kernel linear* terbukti lebih efisien secara komputasi dan cenderung lebih tepat dalam menemukan *hyperplane* optimal.

#### 4.4.3 Hasil Pemodelan Topik LDA dan Implikasinya

Penggunaan LDA sebagai pelengkap SVM dalam penelitian ini memberikan nilai tambah yang nyata. Jika SVM hanya mengklasifikasikan sentimen ke dalam kategori positif, netral, atau negatif, maka LDA mampu menjawab pertanyaan lebih dalam: mengapa suatu ulasan bersifat negatif, dengan cara mengidentifikasi topik-topik keluhan yang spesifik.

Dua keluhan utama yang berhasil diidentifikasi, yaitu infrastruktur akses jalan dan kebersihan lingkungan, merupakan temuan yang dapat dimaknai secara langsung. Kondisi jalan menuju Pantai Lariti yang berkelok dan dinilai kurang memadai menjadi hambatan nyata yang memengaruhi kenyamanan perjalanan wisatawan. Sementara itu, masalah kebersihan, khususnya pengelolaan sampah,

mencerminkan tantangan yang umum dihadapi objek wisata alam di Indonesia yang belum ditunjang infrastruktur pengelolaan sampah yang memadai.

Temuan ini menegaskan bahwa analisis sentimen sebaiknya tidak berhenti di tahap klasifikasi saja, melainkan perlu dilanjutkan dengan pemodelan topik agar rekomendasi yang dihasilkan lebih konkret dan dapat langsung ditindaklanjuti oleh pemangku kepentingan.

Untuk memahami bagaimana kedua temuan keluhan tersebut dapat muncul, perlu dijelaskan terlebih dahulu bagaimana SVM dan LDA bekerja secara bertahap dalam penelitian ini. SVM bekerja pada tahap pertama dengan membaca setiap ulasan dan menentukan apakah ulasan tersebut termasuk sentimen positif, netral, atau negatif. Ulasan yang dianggap mengandung kekecewaan, ketidakpuasan, atau penilaian buruk terhadap suatu aspek wisata akan diberi label negatif. Selanjutnya, hanya ulasan berlabel negatif inilah yang dikumpulkan dan diteruskan ke LDA sebagai bahan analisis, sementara ulasan positif dan netral tidak diikutsertakan. Pemisahan ini sengaja dilakukan agar LDA benar-benar fokus menggali sumber-sumber keluhan, bukan memproses ulasan secara keseluruhan.

Setelah kumpulan ulasan negatif tersebut diterima, LDA mulai bekerja melalui beberapa tahap berikut:

### **1. Pemecahan ulasan menjadi kata-kata**

Setiap ulasan negatif dipecah menjadi kata-kata individual tanpa memperhatikan urutan kalimat. Misalnya, ulasan "*jalannya rusak dan kotor sekali*" akan dipecah menjadi token: *["jalan", "rusak", "kotor"]*. Tahap ini menghasilkan kumpulan kata dari seluruh ulasan negatif yang siap dianalisis.

## **2. Pembangunan kamus kosakata**

Seluruh kata unik dari semua ulasan negatif dikumpulkan menjadi satu kamus (*vocabulary*). Kamus inilah yang menjadi ruang kata tempat LDA bekerja. Setiap kata dalam kamus memiliki peluang untuk menjadi bagian dari topik manapun.

## **3. Penentuan jumlah topik (K)**

Sebelum proses ini berjalan, peneliti harus menentukan terlebih dahulu berapa jumlah topik yang ingin ditemukan. Penentuan ini dilakukan dengan mencoba beberapa nilai K, lalu memilih nilai yang menghasilkan *Coherence Score* tertinggi. *Coherence Score* mengukur seberapa erat hubungan makna antar kata-kata teratas dalam satu topik. Semakin tinggi nilainya, semakin masuk akal topik yang terbentuk secara semantis.

## **4. Inisialisasi acak**

Pada tahap awal, LDA secara acak menetapkan setiap kata dalam setiap ulasan ke salah satu dari K topik yang telah ditentukan. Penetapan ini masih belum bermakna karena bersifat acak, namun menjadi titik awal yang akan terus diperbaiki pada tahap berikutnya.

## **5. Proses iteratif untuk menemukan pola**

Inilah inti dari cara LDA menemukan topik. LDA berulang kali menelusuri setiap kata di setiap ulasan dan mengevaluasi ulang. Evaluasi dilakukan berdasarkan dua pertimbangan yaitu seberapa sering topik tersebut muncul dalam ulasan yang sedang dianalisis, dan seberapa sering kata tersebut muncul bersama kata-kata lain di bawah topik yang sama di seluruh corpus. Proses ini diulang

ratusan hingga ribuan kali hingga penugasan kata ke topik menjadi stabil dan tidak banyak berubah. Di sinilah topik secara alami *muncul dari data*.

## 6. Terbentuknya dua distribusi utama sebagai output

Setelah proses iteratif selesai, LDA menghasilkan dua informasi utama. Pertama, distribusi topik per ulasan, yang menunjukkan seberapa besar proporsi setiap topik dalam tiap ulasan, misalnya: *ulasan A mengandung 70% topik jalan dan 30% topik kebersihan*. Kedua, distribusi kata per topik, yang menunjukkan kata-kata apa yang paling sering dan paling kuat merepresentasikan setiap topik, misalnya: *topik 1 didominasi kata "jalan", "rusak", "berkelok"*, sedangkan *topik 2 didominasi kata "sampah", "kotor", "bau"*.

## 7. Interpretasi manual oleh peneliti

LDA hanya menghasilkan kelompok-kelompok kata, bukan nama topik. Peneliti kemudian membaca kata-kata teratas dari setiap topik dan memaknainya secara kontekstual. Pada Pantai Lariti, kelompok kata *"jalan", "rusak", "berkelok", "akses"* diinterpretasikan sebagai topik infrastruktur akses jalan, sementara kelompok kata *"sampah", "kotor", "bau", "buang"* diinterpretasikan sebagai topik kebersihan lingkungan. Nama topik inilah yang menjadi temuan akhir yang dilaporkan dalam penelitian.

## 4.5 Ringkasan Temuan

Secara keseluruhan, kombinasi SVM, SMOTE, dan LDA menghasilkan kerangka analisis yang komprehensif untuk memahami persepsi wisatawan di Kabupaten/Kota Bima. SMOTE berhasil mengatasi bias model akibat ketidakseimbangan kelas, SVM memberikan klasifikasi sentimen yang stabil

dengan Akurasi di atas 82% pada seluruh skenario, dan LDA mengungkap dua topik keluhan prioritas yang dapat langsung ditindaklanjuti, yaitu infrastruktur akses jalan dan manajemen kebersihan. Temuan ini diharapkan dapat menjadi landasan berbasis data bagi Dinas Pariwisata Kabupaten/Kota Bima dalam merumuskan strategi perbaikan destinasi wisata yang lebih terarah.

Setelah dilakukan pengerjaan yang sama pada Wisata Pantai Lawata, diperoleh hasil bahwa Akurasi 82,32% dan LDA mengungkap tiga topik keluhan yaitu kebersihan dan kenyamanan, pengelolaan sampah serta biaya dan fasilitas parkir. Hal ini menunjukkan bahwa Pantai Lawata adalah destinasi yang mudah diakses dari pusat kota Bima, sehingga keluhan infrastruktur jalan tidak muncul. Keluhan berpusat pada tiga hal: kebersihan pantai, pengelolaan sampah, dan biaya parkir yang tidak sebanding dengan kondisi fasilitas.

Adapun pada Wisata Gunung Tambora, hasil analisis menunjukkan bahwa Akurasi 87,88% serta mengungkap 2 topik keluhan yaitu karakter fisik gunung yang dimana gunung ini dikenal dengan gunung yang masih aktif serta kondisi medan pendakian yang cukup tinggi dan berat dan topik keluhan yang kedua yaitu kondisi iklim. Temuan ini mengindikasikan bahwa jumlah ulasan negatif Gunung Tambora hanya 2 ulasan, sehingga hasil LDA sangat tidak representatif dan topik yang dihasilkan cenderung tumpang tindih (Topik 2 dan 3 identik).

Dengan demikian, ketiga destinasi wisata di Kabupaten/Kota Bima menunjukkan pola yang berbeda dalam hal keluhan pengunjung, di mana aspek Kebersihan lingkungan menjadi keluhan yang muncul di dua dari tiga destinasi (Lariti dan Lawata), menjadikannya aspek paling dominan secara keseluruhan.



menjadi perhatian utama yang perlu ditindaklanjuti oleh pengelola wisata dan pemerintah daerah.

#### 4.6 Integrasi Islam

Dari perspektif nilai-nilai Islam, penelitian ini secara tidak langsung berkontribusi pada upaya perbaikan dua aspek yang mendapat perhatian khusus dalam ajaran Islam. Pertama, masalah kebersihan yang teridentifikasi sebagai keluhan utama wisatawan berkaitan erat dengan prinsip thaharah (kebersihan) yang ditekankan dalam Islam. Sebagaimana sabda Rasulullah SAW:

النَّظَافَةُ مِنَ الْإِيمَانِ

*“Kebersihan adalah sebagian dari iman.” (HR. Muslim)*

Sistem deteksi keluhan berbasis data ini dapat menjadi alat bantu yang memfasilitasi pengelola wisata untuk segera mengambil tindakan pembenahan. Kedua, keluhan terkait akses jalan yang rusak berkaitan dengan konsep menyingkirkan gangguan dari jalan sebagai salah satu bentuk sedekah, sebagaimana Rasulullah SAW bersabda:

..وإِمَاطَةُ الْأَذَى عَنِ الطَّرِيقِ صَدَقَةٌ

*“...dan menyingkirkan gangguan (duri, rintangan) dari jalan Adalah sedekah.” (HR. Bukhari dan Muslim).*

Dengan mengidentifikasi secara otomatis keluhan infrastruktur melalui sistem analisis sentimen, pemerintah daerah memperoleh dasar data yang kuat untuk memprioritaskan perbaikan fasilitas, yang dalam konteks Islam bernilai amal kebajikan

Selain itu, penelitian ini juga mencerminkan prinsip mendengarkan masukan dari sesama sebagaimana dianjurkan dalam Al-Qur'an Surah Al-A'raf ayat 204, yang menekankan pentingnya memperhatikan dengan seksama. Allah SWT berfirman:

وَإِذَا قُرِئَ الْقُرْآنُ فَاسْتَمِعُوا لَهُ ۖ وَأَنْصِتُوا لَعَلَّكُمْ تُرْحَمُونَ ﴿٢٠٤﴾

*“Dan apabila dibacakan Al-Qur'an, maka dengarkanlah baik-baik, dan perhatikanlah dengan tenang agar kamu mendapat rahmat.”*

Meskipun konteks ayat tersebut berkaitan dengan mendengarkan bacaan Al-Qur'an, prinsip untuk mendengarkan dengan baik dan memahami dapat diaplikasikan dalam berbagai aspek kehidupan, termasuk dalam menerima masukan atau ulasan dari orang lain. Teknologi analisis sentimen pada dasarnya adalah bentuk modern dari upaya sistematis untuk “mendengar” suara pengunjung secara masif dan terstruktur guna meningkatkan kualitas layanan dan fasilitas yang disediakan.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil penelitian mengenai analisis sentimen ulasan pengunjung objek wisata di Kabupaten/Kota Bima menggunakan metode *Support Vector Machine* (SVM) dan *Latent Dirichlet Allocation* (LDA), dapat ditarik kesimpulan sebagai berikut.

##### 1. Penanganan Ketidakseimbangan Data

Penerapan teknik SMOTE pada data latih terbukti efektif dalam mengatasi ketimpangan proporsi kelas sentimen. Model *SVM* yang dilatih dengan data hasil *SMOTE* menunjukkan peningkatan nilai *Recall* pada kelas minoritas (sentimen negatif) secara signifikan, sehingga hasil klasifikasi menjadi lebih seimbang dan tidak bias terhadap kelas mayoritas.

##### 2. Performa Klasifikasi SVM

Model SVM dengan *kernel linear* berhasil mengklasifikasikan sentimen ulasan pengunjung ke dalam kategori positif, netral, dan negatif dengan tingkat Akurasi yang stabil. Pada Wisata Pantai Lariti, Akurasi tertinggi dicapai pada skenario 90:10 sebesar 84,91%, sementara skenario 80:20 menunjukkan keseimbangan yang baik antara data latih dan data uji dengan Akurasi 82,8

### 3. Identifikasi Topik Keluhan (LDA)

Pemodelan topik menggunakan LDA pada ulasan bersentimen negatif berhasil mengidentifikasi akar permasalahan yang dikeluhkan wisatawan. Pada Wisata Pantai Lariti, dua topik dominan yang ditemukan adalah kondisi infrastruktur akses jalan dan kebersihan lingkungan, khususnya pengelolaan sampah.

### 4. Visualisasi Data

Penerapan visualisasi data dalam sistem ini berhasil menterjemahkan hasil kodingan yang rumit menjadi informasi yang mudah dipahami. Kesimpulannya, model SVM dan SMOTE yang diusulkan sudah sangat akurat, dan hasilnya menunjukkan secara jelas bahwa perbaikan akses jalan serta pengelolaan sampah adalah prioritas utama untuk memajukan pariwisata di Bima.

### 5. Implikasi Manajerial

Integrasi SVM dan LDA menghasilkan kerangka analitik yang dapat digunakan secara langsung oleh Dinas Pariwisata Kabupaten/Kota Bima sebagai dasar perumusan kebijakan perbaikan fasilitas wisata berbasis data. Dua prioritas utama yang perlu segera ditindaklanjuti adalah perbaikan akses jalan dan peningkatan sistem manajemen kebersihan di kawasan wisata.

### 6. Cara memanfaatkan sistem

Hasil penelitian ini dapat diakses melalui dashboard visualisasi Tableau yang memuat distribusi sentimen per destinasi, tren ulasan dari waktu ke waktu, dan peta topik keluhan. Dashboard ini dapat diperbarui secara berkala oleh Dinas

Pariwisata dengan mengimpor data ulasan baru yang telah diproses oleh sistem ke dalam file yang terhubung dengan Tableau.

Adapun tindak lanjut yang direkomendasikan berdasarkan temuan penelitian ini adalah sebagai berikut.

1. Dalam jangka pendek, pembentukan tim satgas kebersihan mingguan di Pantai Lariti dan Pantai Lawata, serta pemasangan papan imbauan pengelolaan sampah yang edukatif dan menarik.
2. Dalam jangka menengah, pengajuan anggaran perbaikan jalan akses menuju Pantai Lariti kepada Dinas Pekerjaan Umum Kabupaten Bima dengan menggunakan data keluhan pengunjung sebagai justifikasi berbasis bukti.
3. Dalam jangka panjang, pengembangan sistem pemantauan sentimen otomatis yang berjalan secara berkelanjutan, sehingga pemerintah daerah dapat memantau perubahan persepsi wisatawan secara real-time setelah setiap intervensi perbaikan dilakukan.

## **5.2 Saran**

Berdasarkan hasil penelitian, terdapat beberapa saran yang ditujukan kepada pihak terkait maupun penelitian selanjutnya.

### **5.2.1 Saran Praktis**

Berdasarkan hasil analisis sentimen dan pemodelan topik yang telah dilakukan, peneliti merumuskan beberapa rekomendasi strategis bagi pihak-pihak terkait sebagai berikut:

## 1. Bagi Pemerintah Daerah dan Pengelola Wisata (Pokdarwis)

### a. Revitalisasi Infrastruktur

Mengingat topik keluhan dominan pada aksesibilitas, diharapkan pemerintah daerah memprioritaskan perbaikan kualitas jalan menuju destinasi jarak jauh seperti Pantai Lariti. Dalam teori *Marketing Management* (Philip Kotler, Kevin Keller, Mairead Brady, Malcolm Goodman, n.d.), *elemen Place* (saluran distribusi dan akses) sangat krusial, infrastruktur jalan yang memadai akan menurunkan hambatan fisik wisatawan untuk berkunjung, sehingga meningkatkan daya tarik destinasi.

### b. Manajemen Kebersihan Digital

Memanfaatkan hasil analisis sentimen ini sebagai instrumen monitoring real-time untuk membentuk tim satgas kebersihan di titik kritis. Mengacu pada dimensi Tangibles dalam teori *Service Quality (ServQual)* oleh (Parasuraman et al., 1985), kebersihan fasilitas fisik merupakan indikator utama yang memengaruhi kepuasan pengunjung. Peningkatan standar kebersihan akan secara langsung mereduksi sentimen negatif di masa mendatang.

### c. Strategi Pemasaran

Menonjolkan sisi unik "wisata geologi" dan "wisata minat khusus" untuk mengubah citra Bima menjadi destinasi tujuan utama. Berdasarkan teori Strategi Diferensiasi (Porter, n.d.), destinasi harus memiliki nilai unik yang membedakannya dari kompetitor. Temuan sentimen positif mengenai

fenomena laut terbelah dapat dijadikan landasan Positioning untuk memperkuat identitas pariwisata Bima.

d. Transparansi Layanan

Menata ulang sistem retribusi dan parkir agar lebih transparan guna menghindari persepsi biaya yang tidak proporsional. Sesuai teori Price (Nurhayaty, 2022), harga yang dibayarkan pengunjung harus selaras dengan nilai yang dirasakan (*perceived value*). Ketidakseimbangan antara biaya dan kualitas fasilitas akan memicu ketidakpuasan yang tercermin pada ulasan negatif.

## 2. Pemanfaatan *Dashboard Visualisasi*

Hasil analisis yang telah divisualisasikan melalui Tableau disarankan untuk dimanfaatkan secara berkala oleh Dinas Pariwisata sebagai alat pemantauan persepsi wisatawan. Pembaruan data ulasan secara periodik akan menjaga relevansi informasi sebagai dasar pengambilan keputusan.

### 5.2.2 Saran untuk Penelitian Selanjutnya

Berdasarkan hasil penelitian dan keterbatasan yang ditemukan, arah pengembangan penelitian selanjutnya disusun dalam *road map* sebagai berikut:

1. Mengembangkan Model dengan Pendekatan *Deep Learning*

Mengingat ulasan pengunjung sering mengandung bahasa informal, singkatan, dan kata-kata tidak baku, penelitian berikutnya dapat mengimplementasikan pendekatan deep learning menggunakan *pre-trained language* model seperti IndoBERT. Pendekatan ini berpotensi menangkap

konteks semantik kalimat secara lebih akurat dibandingkan representasi TF-IDF yang digunakan dalam penelitian ini

## 2. Menerapkan *Aspect-Based Sentiment Analysis* (ABSA)

Untuk mengidentifikasi sentimen pada aspek spesifik seperti fasilitas, harga, dan pelayanan secara terpisah agar rekomendasi lebih operasional..

## 3. Pembaruan Data Secara Berkala

Membangun sistem pemantauan sentimen otomatis dan berkelanjutan (Real-time monitoring system) yang mengintegrasikan data dari berbagai platform (Google Maps, TripAdvisor, Traveloka) sebagai alat bantu pengambilan keputusan berbasis data bagi Dinas Pariwisata.



## DAFTAR PUSTAKA

- Ariansyah, K., Wahyu, J., Prawiro, H., & Sanjaya, R. D. (2025). *Pengaruh Ulasan Online Terhadap Keputusan Wisatawan dalam Memilih Hotel*. 2, 1–8.
- BPS NTB. (2024). *Perkembangan Pariwisata Nusa Tenggara Barat*. <https://ntb.bps.go.id/id/pressrelease/2024/05/02/988/ntb--maret-2024--perkembangan-pariwisata.html>
- Ekstrom, S. (2024). 2024 Edition International Tourism Highlights. *Learn Tourism*. <https://doi.org/10.18111/9789284425808>
- Fajar, M. R. Al, & Ifantri Ifantri. (2021). *Strategi Promosi Pengembangan Pariwisata Di Kabupaten Bima Di Tinjau Dari Perspektif Ekonomi Syariah*.
- Fanissa, S., Fauzi, M. A., & Adinugroho, S. (2018). Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(8), 2766–2770. <http://j-ptiik.ub.ac.id>
- Feldman, S. (n.d.). *Advanced Approaches in Analyzing Unstructured Data*.
- Hakim, B. (2021). ANALISA SENTIMEN DATA TEXT PREPROCESSING PADA DATA MINING DENGAN MENGGUNAKAN MACHINE LEARNING DATA TEXT PRE-PROCESSING SENTIMENT ANALYSIS IN DATA MINING USING MACHINE LEARNING School of Computer Science and Technology , Harbin Institute of Technology. *Journal of Business and Audit Information Systems*, 4(2), 16–22.
- Hidayatullah, F., & dkk. (2016). Penerapan Text Mining dalam Klasifikasi Judul Skripsi. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi) Agustus*, 1907–5022.
- Nugraha, W., Syarif, M., Bina, U., Informatika, S., Tree, D., & Forest, R. (2024). *Evaluasi Performa Algoritma Klasifikasi dalam Prediksi Gagal Jantung: Studi Kasus Dataset Heart Failure Prediction*. 23(4), 897–908.
- Nurhayaty, M. (2022). Strategi Campuran Pemasaran (Produk, Harga, Tempat, Promosi, Sumber Daya Manusia, Proses, Bukti Fisik). *Jurnal Media Teknologi*, 8(2), 119–127.
- Parasuraman, A., Zeithaml, V. a, & Berry, L. L. (1985). A Conceptual Model of Service Quality and Its Implications for Future Research. In *Journal of Marketing Research* (Vol. 49, pp. 41–50).
- Pati, G. K., & Umar, E. (2022). Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Danau Weekuri Menggunakan Metode Naive Bayes Classifier Dan K-Nearest Neighbor. *Jurnal Media Informatika Budidarma*, 6(4), 2309. <https://doi.org/10.30865/mib.v6i4.4635>

- Philip Kotler, Kevin Keller, Mairead Brady, Malcolm Goodman, T. H. (n.d.). *Marketing Management*. <https://books.google.co.id/books?id=-2hDwAAQBAJ&printsec=frontcover&hl=id#v=onepage&q&f=false>
- Porter, M. E. (n.d.). *Competitive Advantage*. <https://www.perlego.com/book/780037/competitive-advantage-creating-and-sustaining-superior-performance-pdf>
- Rasyad, M., & Fajar, A. (2021). *Strategi Promosi Pengembangan Pariwisata Di Kabupaten Bima*. 4, 77–95.
- Sibyan, H., & Hasanah, N. (2022). Analisis Sentimen Ulasan Pada Wisata Dieng Dengan Algoritma K-Nearest Neighbor (K-Nn). *Jurnal Penelitian Dan Pengabdian Kepada Masyarakat UNSIQ*, 9(1), 38–47. <https://doi.org/10.32699/ppkm.v9i1.2218>
- Subarkah, M. Z., Hilda, M., & Zukhronah, E. (2022). Analisis Sentimen Review Tempat Wisata Pada Data Online Travel Agency Di Yogyakarta Menggunakan Model Neural Network IndoBERTweet Fine Tuning. *Seminar Nasional Official Statistics*, 2022(1), 543–552. <https://doi.org/10.34123/semnasoffstat.v2022i1.1246>
- Syahlan, M. S., Irmayanti, D., & Alam, S. (2023). ANALISIS SENTIMEN TERHADAP TEMPAT WISATA DARI KOMENTAR PENGUNJUNG DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM). *Simtek : Jurnal Sistem Informasi Dan Teknik Komputer*, 8(2), 315–319. <https://doi.org/10.51876/simtek.v8i2.281>
- Tri, I., Sihalohe, P., Ratnawati, D. E., & Rahayudi, B. (2022). *Analisis Sentimen Objek Wisata Danau Toba berdasarkan Ulasan Pengunjung menggunakan Algoritma Support Vector Machine*. 6(9), 4204–4209. <http://j-ptiik.ub.ac.id>
- Utama, D. A. (2023). *Gunungkidul Menggunakan Metode Naive Bayes Sentiment Analysis of Tourism Places in Gunungkidul District Using Naive Bayes Method Based on Criticism of*. 2(September), 1102–1110.