

## Implementasi Metode XLM-RoBERTa Untuk Analisis Sentimen Ulasan Pengunjung Objek Wisata Kerta Gosa Klungkung

I Dewa Gede Mahesta Parawangsa<sup>1</sup>, Gede Surya Mahendra<sup>2</sup>, Rama Ngurah Putera Pinatih<sup>3</sup>

<sup>1,2,3</sup> Program Studi Sistem Informasi, Fakultas Teknik dan Kejuruan, Universitas Pendidikan Ganesha

[mahesta@undiksha.ac.id](mailto:mahesta@undiksha.ac.id)<sup>1</sup>, [gmauhendra@undiksha.ac.id](mailto:gmauhendra@undiksha.ac.id)<sup>2</sup>, [rama.putera@undiksha.ac.id](mailto:rama.putera@undiksha.ac.id)<sup>3</sup>

### Abstrak

*Pasca pandemi, pemulihhan sektor pariwisata di Kabupaten Klungkung, khususnya pada Objek Wisata Kerta Gosa, sangat bergantung pada kemampuan pengelola dalam merespons umpan balik pengunjung secara akurat. Saat ini, ribuan ulasan yang tersebar di platform digital seperti Google Maps dan TripAdvisor menjadi aset data krusial, namun pemanfaatannya terhambat oleh volume data yang besar serta karakteristik linguistik yang tidak terstruktur. Tantangan utama muncul dari fenomena percampuran kode bahasa (code-mixing) antara Bahasa Indonesia dan Inggris yang menyulitkan metode klasifikasi konvensional. Penelitian ini bertujuan untuk mengatasi hambatan tersebut dengan mengimplementasikan analisis sentimen otomatis menggunakan arsitektur Deep Learning berbasis XLM-RoBERTa. Model transformer multibahasa ini dipilih secara spesifik karena keunggulannya dalam mengekstraksi representasi semantik lintas bahasa tanpa memerlukan tahapan penerjemahan awal, sehingga nuansa emosi asli pengunjung tetap terjaga. Melalui pendekatan metodologi SEMMA, dataset diproses dengan teknik normalisasi teks adaptif dan dilatih menggunakan skema pembagian data terstratifikasi. Hasil eksperimen menunjukkan performa model yang solid dengan capaian akurasi pengujian sebesar 91.37%, serta nilai F1-Score mencapai 87.26%. Analisis lebih lanjut pada ulasan negatif secara spesifik menyoroti dominasi keluhan terkait transparansi layanan tiket dan interaksi pemandu wisata, yang menjadi area prioritas bagi manajemen untuk pembenahan. Temuan ini menegaskan bahwa XLM-RoBERTa merupakan solusi yang efisien dan robust untuk sistem pemantauan kualitas layanan pariwisata berbasis data sentimen.*

**Kata kunci:** Analisis Sentimen, XLM-RoBERTa, Pariwisata Bali, Kerta Gosa, Deep Learning

### 1. Latar Belakang

Pariwisata merupakan sektor kunci yang menggerakkan ekonomi daerah dan Kabupaten Klungkung, Bali memiliki peran penting dalam peta wisata Bali. Salah satu daya tarik utama di wilayah ini adalah Kerta Gosa yang merupakan bangunan bersejarah dan menjadi kebanggaan daerah Kabupaten Klungkung [1]. Namun pasca pandemi Covid-19, objek wisata ini menghadapi tantangan besar untuk memulihkan tingkat kunjungan di tengah persaingan destinasi yang semakin ketat. Di era digital saat ini pengalaman para pengunjung yang datang terekam secara luas di internet. Wisatawan kini aktif membagikan pendapat mereka pada platform Google Maps dan TripAdvisor yang menjadi sumber informasi utama bagi calon wisatawan lokal maupun mancanegara [2] ,[3]. Ulasan yang diposting secara positif atau negatif di platform ini memiliki dampak besar terhadap popularitas dan keputusan orang lain untuk berkunjung ke Kerta Gosa. Ribuan ulasan ini membentuk kumpulan data besar yang terus bertambah seiring waktu dan menjadi aset informasi yang sangat krusial [4].

Data ulasan dari berbagai platform tersebut seharusnya dapat dimanfaatkan secara maksimal oleh Dinas Kebudayaan dan Pariwisata serta pengelola objek wisata. Informasi yang terkandung dalam ulasan ini memiliki nilai signifikan untuk evaluasi layanan dan pengembangan destinasi agar pengalaman pengunjung meningkat [5]. Namun kendala utama muncul karena belum adanya analisis yang menyeluruh terhadap ribuan komentar tersebut. Melakukan pemantauan manual terhadap data teks yang sangat banyak tentu tidak efektif dan memakan waktu lama. Kurangnya pemahaman mendalam tentang sentimen ini menjadi penghambat bagi pengelola untuk melakukan perbaikan yang tepat sasaran pada objek wisata Kerta Gosa.

Tantangan dalam mengolah data ini menjadi semakin rumit ketika melihat karakteristik ulasan yang sangat beragam dari segi bahasa. Populasi data ulasan Kerta Gosa terdiri dari berbagai bahasa termasuk Bahasa Indonesia dan Inggris serta bahasa asing lainnya mengingat asal wisatawan yang global. Seringkali ditemukan ulasan di

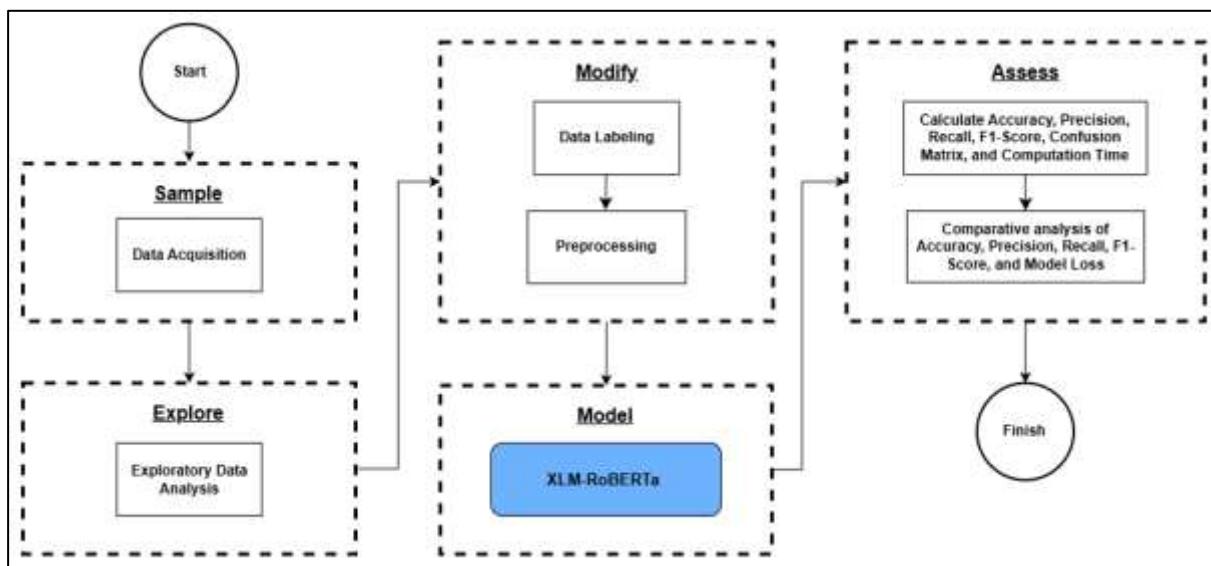
Google Maps maupun TripAdvisor yang mencampuradukkan kosa kata antar bahasa atau menggunakan istilah tidak baku dalam satu kalimat [6]. Kondisi data yang multibahasa dan tidak terstruktur ini menyulitkan proses identifikasi sentimen jika hanya mengandalkan penerjemahan biasa atau model yang hanya mengerti satu bahasa [7]. Kegagalan dalam menangkap konteks bahasa yang bercampur ini dapat menyebabkan kesalahan fatal dalam menyimpulkan apakah sebuah ulasan bernada positif atau negatif.

Untuk mengatasi masalah klasifikasi teks para peneliti awalnya mengandalkan metode *Machine Learning* seperti *Support Vector Machine* (SVM) dan *Naive Bayes* karena efisiensinya [8], [9]. Namun metode ini terbatas dalam memahami konteks kalimat yang rumit [10], [11], [12]. Hal ini memicu peralihan ke *Deep Learning* seperti *Long Short-Term Memory* (LSTM) yang lebih baik menangkap urutan kata meski komputasinya cenderung lambat [13]. Kemajuan teknologi kemudian melahirkan arsitektur *Transformer* yang efisien memproses data paralel dan menjadi fondasi model BERT [14]. Walaupun mampu memahami konteks dua arah BERT dinilai masih *undertrained* sehingga disempurnakan oleh varian RoBERTa yang lebih stabil berkat data latih yang lebih besar [15]. Kendati demikian arsitektur optimal ini tetap memerlukan adaptasi pada kasus ulasan Kerta Gosa. Kompleksitas campuran bahasa di dalamnya menuntut pendekatan yang tidak hanya cerdas secara kontekstual tetapi juga mampu menembus batasan linguistik yang selama ini menjadi kelemahan model satu bahasa.

Menjawab kebutuhan akan kemampuan lintas bahasa tersebut penelitian ini mengajukan penerapan metode XLM-RoBERTa sebagai solusi teknis utama. Model ini secara cerdas menggabungkan arsitektur RoBERTa yang kokoh dengan strategi pelatihan *Cross-lingual Language Model* atau XLM yang dirancang khusus untuk mempelajari representasi makna dari ratusan bahasa secara bersamaan [16]. Integrasi ini menjadi kunci penting karena memungkinkan sistem untuk memahami data ulasan yang mencampur Bahasa Indonesia, Inggris dan bahasa lainnya secara natural tanpa memerlukan proses penerjemahan manual yang berisiko mengubah konteks asli. Dengan menerapkan model ini pada data bersumber dari Google Maps dan TripAdvisor diharapkan pola sentimen pengunjung dapat dikenali dengan presisi tinggi. Hasil analisis yang akurat ini nantinya akan menjadi landasan strategis yang valid bagi pemangku kepentingan dalam merumuskan kebijakan pengembangan pariwisata Klungkung maupun Bali di masa depan.

## 2. Metode Penelitian

Penelitian ini menerapkan kerangka kerja data mining SEMMA yang mencakup tahapan *Sample*, *Explore*, *Modify*, *Model* [17]. Pendekatan ini menjamin alur kerja yang sistematis mulai dari tahap persiapan data hingga evaluasi model, sebagaimana divisualisasikan pada Gambar 1 berikut.



Gambar 1. Metode Penelitian

### 2.1. Sample

Tahap awal penelitian melibatkan proses pengumpulan data ulasan pengguna. Dataset sekunder yang digunakan dalam penelitian ini terdiri dari total 1.621 data ulasan teks mentah. Dataset ini merupakan data publik yang diperoleh melalui proses scraping dari dua platform, yaitu Google Maps dan TripAdvisor Kerta Gosa. Sampel ini

dipilih karena dianggap representatif dalam menggambarkan persepsi pengunjung terhadap Objek Wisata Kerta Gosa, mencakup spektrum opini yang luas mulai dari aspek keindahan arsitektur, nilai sejarah, hingga kondisi fasilitas di lokasi wisata.

## 2.2. Explore

Tahap eksplorasi difokuskan pada telaah mendalam terhadap dataset mentah guna memetakan pola linguistik yang terbentuk. Melalui visualisasi data, dilakukan identifikasi terhadap proporsi kelas sentimen, dominasi kata kunci, serta keragaman bahasa yang digunakan oleh pengunjung. Ditemukan bahwa ulasan memiliki karakteristik yang sangat beragam, mulai dari penggunaan bahasa formal, bahasa campuran antara Indonesia, Inggris dan yang lain, hingga penggunaan singkatan dan *emoji*. Untuk menstandarisasi input bagi model XLM-RoBERTa, seluruh data teks melalui tahapan *preprocessing* untuk membersihkan *noise* tanpa menghilangkan konteks aslinya.

## 2.3. Modify

Tahap modifikasi mencakup serangkaian proses persiapan data yang krusial sebelum masuk ke pemodelan. Adapun rincian langkah dalam fase ini meliputi.

### 2.3.1 Tahap Labeling

- a. Pelabelan Data: Pada tahap ini, dataset ulasan yang telah bersih diberikan anotasi sentimen berdasarkan muatan opini yang terkandung di dalamnya. Proses pelabelan dilakukan secara manual oleh peneliti dengan merujuk pada skala penilaian yang diberikan pengguna, di mana ulasan dengan rating bintang 4 hingga 5 diklasifikasikan sebagai kelas Positif, rating 3 sebagai Netral, dan rating 1 hingga 2 sebagai Negatif. Untuk memastikan reliabilitas data, label awal tersebut kemudian diverifikasi ulang (*cross-check*) dengan meninjau konteks narasi secara menyeluruh guna memvalidasi kesesuaian antara angka rating dan sentimen teks, sehingga inkonsistensi atau bias subjektif dapat diminimalkan sebelum proses pemodelan.

### 2.3.2 Tahap Pre-processing

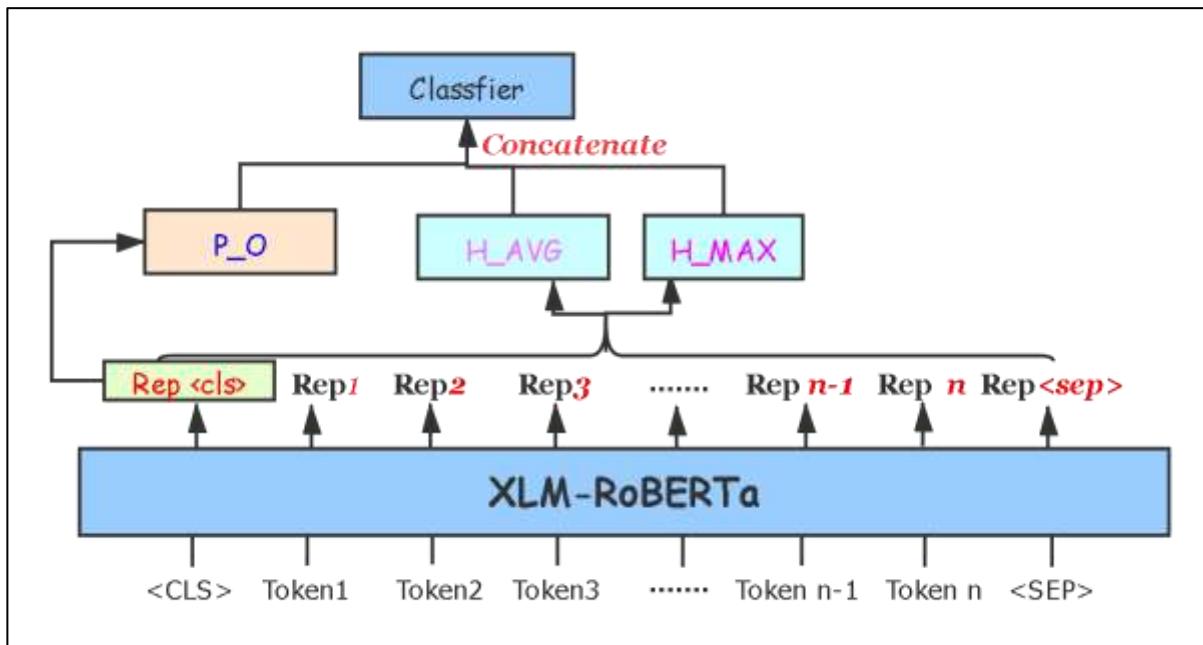
- a. *Data Cleaning*: Proses ini berfokus pada penghapusan elemen-elemen noise yang tidak memberikan kontribusi pada konteks sentimen. Langkah-langkah yang dilakukan meliputi penghapusan tautan (URL), mention pengguna, serta karakter khusus atau simbol yang tidak relevan (seperti @, #, \$, %). Namun, untuk menjaga konteks emosi dalam pendekatan deep learning, penggunaan tanda baca tertentu mungkin dipertahankan sesuai kebutuhan model.
- b. Penyeragaman Karakter: Seluruh teks ulasan dikonversi menjadi huruf kecil (lowercase). Tujuannya adalah untuk menstandarisasi input sehingga mesin tidak memperlakukan kata yang sama sebagai dua entitas fitur yang berbeda, yang dapat memengaruhi efisiensi pelatihan.
- c. Normalisasi Teks: Mengingat data ulasan sering menggunakan bahasa gaul, singkatan, atau kesalahan ketik, dilakukan normalisasi untuk mengubah kata-kata tidak baku menjadi bentuk baku sesuai kaidah bahasa yang benar agar model dapat memahaminya dengan lebih baik.
- d. Pembagian Data: Untuk mengevaluasi performa model secara objektif, dataset dibagi menjadi tiga subset proporsional, yaitu data pelatihan, data validasi, dan data pengujian dengan rasio 70:20:10.
- e. Tokenisasi dan Encoding Input: Berbeda dengan metode tradisional yang menggunakan pembobotan kata, penelitian ini memanfaatkan mekanisme Tokenisasi berbasis SentencePiece yang merupakan standar dari model XLM-RoBERTa. Pada tahap ini, teks dipecah menjadi unit-unit sub-kata (sub-word units) dan diubah menjadi format numerik berupa Input IDs dan Attention Masks. Metode ini memungkinkan model untuk menangani kata yang tidak dikenali (out-of-vocabulary) serta memahami konteks lintas bahasa (multilingual) dari ulasan yang bercampur antara Bahasa Indonesia dan Inggris secara lebih efektif dibandingkan metode ekstraksi fitur konvensional.

## 2.4 Model

Fase ini berfokus pada pengembangan dan pelatihan model *deep learning* yang dikhususkan untuk pemrosesan bahasa alami. Keputusan penggunaan arsitektur ini merujuk pada keberhasilan riset sebelumnya yang terbukti andal dalam mengolah data teks yang kompleks. Berikut adalah arsitektur utama yang digunakan dalam penelitian ini.

#### 2.4.1 XLM-RoBERTa

XLM-RoBERTa adalah model *transformer* multibahasa dari Facebook AI yang digunakan dalam penelitian ini untuk mengklasifikasikan sentimen ulasan ke dalam kategori positif, negatif, atau netral. Keunggulan utamanya terletak pada kemampuan memproses teks dengan campuran bahasa Indonesia dan Inggris secara langsung tanpa perlu penerjemahan, sehingga nuansa emosi asli pengunjung dapat dipahami lebih akurat melalui proses *fine-tuning* [7]. Untuk memberikan gambaran visual mengenai struktur model ini, berikut merupakan ilustrasi dari arsitektur XLM-RoBERTa.



Gambar 2. Arsitektur XLM-RoBERTa

Implementasi arsitektur sebagaimana terlihat pada Gambar 2 dikonfigurasi dengan sejumlah parameter spesifik pada tahap *training*. Konfigurasi ini mengadopsi pengaturan hyperparameter yang telah terbukti optimal dalam penelitian terbaru oleh Masaling & Suhartono pada tahun 2024, yang menunjukkan stabilitas konvergensi yang baik pada model XLM-RoBERTa menggunakan optimizer AdamW [18]. Rincian parameter tuning selengkapnya disajikan pada Tabel 1 berikut.

Tabel 1. Parameter Model XLM-Roberta

Parameter	Deskripsi
<b>Arsitektur Model</b>	<i>FacebookAI/xlm-roberta-base</i>
<b>Input Size</b>	Max Sequence Length 128 Token
<b>Optimizer</b>	AdamW
<b>Learning Rate</b>	0.00001
<b>Batch Size</b>	32
<b>Weight Decay</b>	0.001
<b>Activation Function</b>	<i>GELU (Hidden Layers)</i>
<b>Epoch</b>	10 Epoch
<b>Early Stopping</b>	Patience 3 Epoch (F1-Score)
<b>Loss Function</b>	Weighted Cross-Entropy Loss
<b>Struktur Transformer</b>	12 Hidden Layers, 12 Attention Heads, 768 Hidden Size

#### 2.5 Assess

Tahap evaluasi bertujuan untuk mengukur kinerja model menggunakan dataset pengujian berdasarkan parameter utama seperti Akurasi, Presisi, Recall, dan F1-score. Selain pengukuran statistik tersebut, dilakukan pula pengujian

komparatif pada input kalimat untuk melihat perbedaan hasil antara metode yang terintegrasi dengan kamus multibahasa dan yang tidak. Langkah ini dimaksudkan untuk menilai sejauh mana integrasi kamus tersebut berkontribusi terhadap pemahaman konteks, sehingga dapat dipastikan bahwa model mampu memberikan respons yang lebih relevan dan akurat dalam menangani data yang bersifat multibahasa.

Evaluasi diawali dengan metrik *accuracy*, yang berfungsi untuk mengukur ketepatan model secara menyeluruh dalam memprediksi kelas data. Metrik ini mengukur rasio jumlah prediksi yang benar baik positif, negatif, maupun netral dibandingkan dengan total keseluruhan data uji. Perhitungan Akurasi ditampilkan pada persamaan (1).

$$\text{Akurasi} = (TP + TN) / (TP + FP + TN + FN) \quad (1)$$

Selanjutnya, *Precision* digunakan untuk mengukur tingkat ketepatan prediksi model terhadap kelas positif. Metrik ini menghitung rasio antara prediksi positif yang benar dibandingkan dengan keseluruhan data yang diprediksi positif oleh model. Nilai presisi yang tinggi menunjukkan bahwa model memiliki tingkat kesalahan yang rendah dalam memberikan label positif palsu. Perhitungan metrik Presisi dapat dilihat pada persamaan (2).

$$\text{Presisi} = TP / (TP + FP) \quad (2)$$

Metrik berikutnya adalah *Recall*, yang berfungsi untuk mengukur sensitivitas model dalam mengenali informasi yang relevan. Metrik ini menghitung rasio prediksi positif yang benar dibandingkan dengan keseluruhan data yang sebenarnya berlabel positif. *Recall* sangat berguna untuk mengetahui seberapa banyak data relevan yang berhasil ditemukan kembali oleh model. Perhitungan *Recall* ditunjukkan pada persamaan (3).

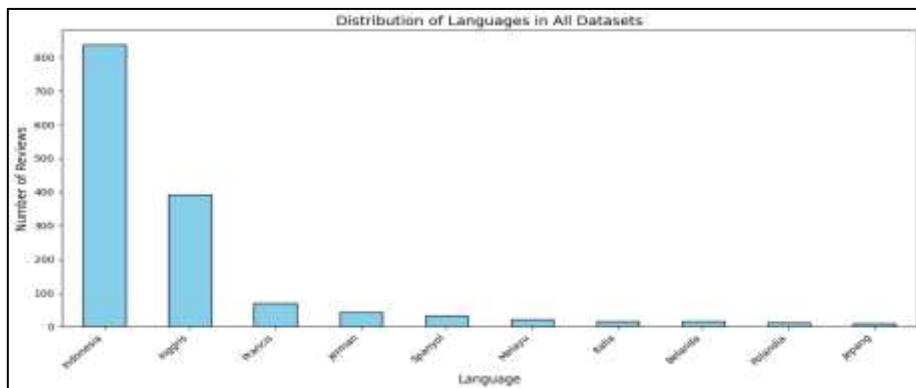
$$\text{Recall} = TP / (TP + FN) \quad (3)$$

Terakhir, *F1-Score* diterapkan untuk mengukur performa model secara komprehensif, mengingat metrik ini merepresentasikan titik keseimbangan antara *Presisi* dan *Recall*. *F1-Score* menjadi indikator penting ketika dataset memiliki distribusi kelas yang tidak seimbang, karena memberikan gambaran tunggal mengenai kualitas prediksi model. Perhitungan dari *F1-Score* ditampilkan pada persamaan (4) berikut.

$$\text{F1-Score} = 2 (\text{Presisi} \times \text{Recall}) / (\text{Presisi} + \text{Recall}) \quad (4)$$

### 3. Hasil dan Diskusi

Eksperimen diawali dengan eksplorasi mendalam terhadap karakteristik dataset guna memahami tantangan linguistik yang dihadapi. Berdasarkan hasil pembersihan data diperoleh total 1452 ulasan valid yang siap diproses. Distribusi linguistik menunjukkan dominasi Bahasa Indonesia sebanyak 837 sampel diikuti oleh Bahasa Inggris sebagai kontributor terbesar kedua dengan jumlah 392 sampel. Selain itu dataset juga memuat keberagaman bahasa asing lainnya yang meliputi Prancis sebanyak 70 ulasan, Jerman 43 ulasan, Spanyol 33 ulasan, Melayu 22 ulasan, Italia serta Belanda masing-masing 16 ulasan, Polandia 13 ulasan, hingga Jepang sebanyak 10 ulasan. Temuan krusial pada tahap ini adalah tingginya frekuensi *code-mixing* dan penggunaan istilah tidak baku dalam struktur kalimat yang sangat ringkas, dengan rata-rata panjang ulasan hanya 12,91 kata. Untuk memberikan gambaran yang lebih komprehensif mengenai keragaman linguistik tersebut, visualisasi distribusi bahasa pada dataset disajikan pada Gambar 3 berikut.



Gambar 3. Visualisasi Grafik Jenis Bahasa

DOI: <https://doi.org/10.31004/riggs.v4i4.5368>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Berdasarkan visualisasi pada Gambar 3, terlihat pola distribusi yang sangat timpang atau *long-tail* di mana Bahasa Indonesia dan Inggris mendominasi sebagian besar proporsi data dibandingkan bahasa lainnya. Kesenjangan jumlah sampel yang signifikan antara kelompok bahasa utama dan bahasa minoritas seperti Jepang atau Polandia ini menghadirkan tantangan nyata bagi model dalam menjaga konsistensi prediksi. Kondisi tersebut menegaskan urgensi penerapan arsitektur *transformer* multibahasa yang tidak hanya andal pada bahasa populer, tetapi juga memiliki kemampuan generalisasi yang kuat untuk menangkap representasi fitur dari bahasa dengan sumber daya rendah atau *low-resource languages* yang muncul dalam dataset.

Untuk menjamin evaluasi kinerja model XLM-RoBERTa yang objektif, dataset tersebut didistribusikan ke dalam tiga subset fungsional menggunakan rasio skema 70:20:10. Pembagian ini menghasilkan 1.016 data *train* yang digunakan untuk proses pembelajaran model, 290 data *validation* untuk memantau konvergensi selama pelatihan, dan 146 data *test* yang diisolasi. Teknik pembagian ini dilakukan secara terstratifikasi untuk memastikan proporsi sebaran kelas sentimen tetap konsisten di setiap bagian, sehingga mencegah kebocoran data dan memastikan validitas hasil pengujian pada data baru.

Pelatihan model dilakukan menggunakan perangkat komputasi lokal Lenovo Legion 5 15IRX9. Spesifikasi perangkat keras yang digunakan meliputi prosesor 13th Gen Intel(R) Core(TM) i7-13650HX, memori RAM 24 GB DDR5, serta NVIDIA GeForce RTX 4060 Laptop GPU dengan VRAM 8 GB. Proses training dijalankan sebanyak 10 epoch dengan menerapkan dua callbacks, yaitu ReduceLROnPlateau untuk penyesuaian learning rate dinamis dan EarlyStopping dengan patience 3 epoch. *EarlyStopping* bertugas memantau *validation loss* dan menghentikan iterasi jika hasilnya tidak membaik demi menjaga generalisasi model. Dengan demikian, model optimal tetap terjaga, dan hasil akurasinya pada data uji dapat dilihat pada Tabel 2 berikut.

Tabel 2 Hasil Performa XLM-RoBERTa

Metrik Evaluasi	Nilai (%)
<i>Accuracy</i>	91.37
<i>Precision</i>	83.51
<i>Recall</i>	91.37
<i>F1-Score</i>	87.26

Berdasarkan hasil kuantitatif pada Tabel 2, model XLM-RoBERTa menunjukkan kemampuan generalisasi yang cukup solid terhadap data ulasan yang belum pernah dilihat sebelumnya (data *testing*). Tingginya nilai akurasi yang diperoleh mengindikasikan model berhasil mempelajari pola linguistik yang kompleks, termasuk struktur kalimat multibahasa yang menjadi ciri khas ulasan di Kerta Gosa. Kemampuan XLM-RoBERTa menangkap konteks *code-mixing* terbukti efektif karena model tidak memerlukan penerjemahan yang seringkali mendistorsi makna slang lokal seperti 'matur suksma' atau 'bestie'.

Sebagaimana terlihat pada tahap eksplorasi data, dominasi sentimen kelas Positif yang mencapai 91.7% dari total populasi data memberikan bias alami pada model untuk lebih mudah mengenali ulasan bernada baik. Sementara itu, tantangan terbesar terletak pada identifikasi kelas minoritas, khususnya sentimen Negatif yang hanya berjumlah 2.2% dan Netral sebesar 6.2%. Meskipun model mampu menangkap konteks ulasan positif dengan sangat presisi, performa pada kelas minoritas mungkin tidak setinggi kelas mayoritas karena minimnya variasi contoh kalimat negatif selama proses pelatihan. Secara spesifik, analisis kesalahan menunjukkan bahwa ketimpangan ini memicu kecenderungan model untuk melakukan generalisasi berlebih, di mana ulasan negatif yang ambigu atau mengandung sindiran halus sering kali salah diklasifikasikan menjadi sentimen positif. Hal ini wajar terjadi pada kasus *real-world dataset* pariwisata di mana pengunjung cenderung memberikan ulasan positif jika merasa puas.

Selain mengukur performa model melalui angka-angka metrik di atas, penelitian ini juga menggali lebih dalam substansi ulasan untuk memahami konteks apa yang sebenarnya menjadi sorotan utama wisatawan. Angka akurasi yang tinggi tidak akan bermakna banyak bagi strategi pengembangan pariwisata jika tidak disertai pemahaman mengenai topik spesifik yang dibicarakan. Oleh karena itu, dilakukan ekstraksi fitur berbasis frekuensi kata untuk memetakan dominasi topik pada setiap kelas sentimen. Visualisasi distribusi kata-kata yang paling sering muncul dalam ulasan pengunjung Kerta Gosa ditampilkan melalui representasi *Word Cloud* sebagaimana terlihat pada Gambar 4 berikut.



Gambar 4. Word Cloud Sentimen Kerta Gosa

Visualisasi pada Gambar 4 mempresentasikan distribusi kata kunci dominan yang terbagi ke dalam tiga kategori sentimen. Pada ulasan positif, narasi pengunjung terlihat konsisten dengan fokus utama pada apresiasi fisik dan nilai historis, yang diwakili oleh frekuensi tinggi kata-kata deskriptif seperti "museum", "history", dan "beautiful". Struktur ini menegaskan bahwa kepuasan utama pengunjung bersumber langsung dari substansi dan estetika objek wisata itu sendiri. Namun, pola yang berbeda mulai terlihat pada kategori non-positif, di mana terjadi pergeseran topik dari deskripsi visual menuju aspek pengalaman operasional.

Pada analisis *Word Cloud* untuk sentimen negatif terlihat bahwa keluhan pengunjung tidak tersebar secara acak, melainkan mengerucut pada mekanisme akses dan interaksi personal di lokasi. Kata kunci seperti "tiket", "loket", "masuk", serta "guide" dan "tip" muncul dengan ukuran yang sangat dominan. Dominasi kata-kata ini mengindikasikan bahwa meskipun pengunjung mengagumi nilai sejarah Kerta Gosa, pengalaman mereka terganggu oleh faktor ketidaknyamanan dalam proses pembelian tiket masuk dan adanya indikasi gangguan dari oknum pemandu layanan yang berkaitan dengan pungutan biaya tambahan. Temuan ini menjadi sinyal valid bagi pengelola bahwa prioritas pemberian sebaiknya difokuskan pada transparansi layanan di pintu masuk serta penertiban standar pelayanan petugas di lapangan.

#### 4. Kesimpulan

Penelitian ini membuktikan efektivitas XLM-RoBERTa dalam menangani analisis sentimen multibahasa di Objek Wisata Kerta Gosa tanpa memerlukan penerjemahan eksternal, dengan capaian Akurasi 91,37% dan F1-Score 87,26%. Selain kinerja klasifikasi yang solid, temuan kualitatif menyoroti dominasi kata kunci "tiket", "loket", dan "guide" pada ulasan negatif, yang mengindikasikan adanya keluhan krusial terkait transparansi biaya dan standar pelayanan petugas. Namun, evaluasi mendalam menunjukkan bahwa performa model pada kelas minoritas (negatif dan netral) masih terbatas akibat ketimpangan distribusi data yang sangat didominasi oleh sentimen positif, sehingga model cenderung kurang sensitif dalam menangkap nuansa keluhan yang tersirat. Berdasarkan keterbatasan distribusi kelas yang teridentifikasi dalam penelitian ini, studi selanjutnya sangat disarankan untuk menerapkan strategi penanganan data yang lebih proaktif guna mengatasi dominasi sentimen positif. Penggunaan teknik *class weighting* pada *loss function* atau penerapan augmentasi data sintetik pada kelas minoritas perlu dieksplorasi secara mendalam agar model tidak bias terhadap kelas mayoritas. Selain itu, upaya untuk memperkaya model sebaiknya dilakukan dengan memperluas cakupan dataset, tidak hanya dari segi kuantitas tetapi juga variasi sumber data. Integrasi ulasan dari berbagai platform digital tambahan akan memberikan eksposur linguistik yang lebih beragam bagi model, sehingga kemampuan generalisasinya dalam mengenali pola kalimat yang kompleks dan tidak baku dapat ditingkatkan secara signifikan. Selanjutnya, untuk menjembatani kesenjangan antara eksperimen akademis dan implementasi industri, penelitian di masa depan perlu melangkah lebih jauh dengan menerapkan *cost-benefit analysis* dalam pemilihan model. Evaluasi ini tidak hanya terpaku pada metrik akurasi semata, melainkan menyeimbangkan *trade-off* antara peningkatan performa deteksi dengan biaya komputasi serta spesifikasi perangkat keras yang dibutuhkan. Pendekatan ini akan memberikan parameter yang lebih terukur dan realistik bagi pengembang sistem dalam menentukan arsitektur mana yang paling efisien, ekonomis, dan *feasible* untuk diintegrasikan ke dalam dasbor pemantauan pariwisata yang berjalan secara *real-time*.

#### Referensi

- [1] Dinas Kebudayaan dan Pariwisata Kabupaten Klungkung, "Kerta Gosa"  
<Https://Tes.Klungkungkab.Go.Id/Halaman/Detail/Kertagosa-Klungkung>.

- [2] I. Nyoman Saputra Wahyu Wijaya, K. Agus Seputra, and N. Putu Novita Puspa Dewi, "Fine Tuning Model IndoBERT untuk Analisis Sentimen Berita Pariwisata Indonesia," *Jurnal Pendidikan Teknologi dan Kejuruan*, vol. 22, no. 2, 2025, [Online]. Available: <https://www.detik.com/search/searchall?query=wisata&siteid=3&sortby=time&fromdate=01/01/2022&>
- [3] G. Aditra Pradnyana, I. Gede, and M. Darmawiguna, "Web-Based System for Bali Tourism Sentiment Analysis during The Covid-19 Pandemic using Django Web Framework and Naive Bayes Method," 2021.
- [4] G. S. Mahendra and K. Y. Ernanda Aryanto, "SPK Penentuan Lokasi ATM Menggunakan Metode AHP dan SAW," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, pp. 49–56, Apr. 2019, doi: 10.25077/teknosi.v5i1.2019.49-56.
- [5] J. Rhena, K. Kraugusteliana, and Hamzar, "Embracing Digitalization in Tourism: Strategic Approaches for Global Competitiveness in the Digital Economy Era," *Indo-Fintech Intellectuals: Journal of Economics and Business*, vol. 4, no. 2, pp. 461–472, Jun. 2024, doi: 10.54373/ifiweb.v4i2.1282.
- [6] M. A. Hasan, "Ensemble Language Models for Multilingual Sentiment Analysis," Mar. 2024, [Online]. Available: <http://arxiv.org/abs/2403.06060>
- [7] A. Gaurav, B. B. Gupta, S. Sharma, R. Bansal, and K. T. Chui, "XLM-RoBERTa Based Sentiment Analysis of Tweets on Metaverse and 6G," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 902–907. doi: 10.1016/j.procs.2024.06.110.
- [8] I. B. Manuaba, G. R. Dantes, and G. Indrawan, "Analisis Sentimen Data Provider Layanan Internet Pada Twitter Menggunakan Support Vector Machine Dengan Penambahan Algoritma Levenshtein Distance," *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 5, no. 2, pp. 9–17, Mar. 2022, doi: 10.47970/siskom-kb.v5i2.261.
- [9] I. P. Dedy, W. Darmawan, G. Aditra Pradnyana, I. Bagus, and N. Pascima, "Optimasi Parameter Support Vector Machine Dengan Algoritma Genetika Untuk Analisis Sentimen Pada Media Sosial Instagram," 2023, [Online]. Available: <https://doi.org/10.31598/jimik.v4i1.157>
- [10] I. G. Hendrayana, D. G. H. Divayana, and M. W. A. Kesiman, "Komparasi Metode SVM, K-NN dan NBC pada Analisis Sentimen," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 4, no. 1, pp. 191–198, Jan. 2023, doi: 10.35870/jimik.v4i1.157.
- [11] K. Pramayasa, I. Md, D. Maysanjaya, G. Ayu, and A. Diatri Indradewi, "Analisis Sentimen Program Mbkm Pada Media Sosial Twitter Menggunakan KNN Dan SMOTE," 2023, [Online]. Available: <https://doi.org/10.31598/jitet.v13i3S1.8145>.
- [12] I. P. A. V. Vidyananta and K. T. D. Dermawani, "Perbandingan Algoritma K-Means Dan K-Medoids Dalam Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Keadaan Sekolah Dasar," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3S1, Oct. 2025, doi: 10.23960/jitet.v13i3S1.8145.
- [13] N. K. T. A. Saputri, I. G. A. Gunadi, and I. M. G. Sunarya, "Analisis Sentimen Pelayanan Daring di Fakultas Teknik dan Kejuruan Universitas Pendidikan Ganesha Menggunakan Algoritma Naïve Bayes dan LSTM," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, pp. 1120–1129, Jul. 2024, doi: 10.57152/malcom.v4i3.1336.
- [14] A. P. Al Aufar and A. Romadhony, "Aspect-based Sentiment Analysis on Beauty Product Reviews using BERT and Long Short-Term Memory," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 14, no. 2, pp. 364–373, Jul. 2025, doi: 10.23887/janapati.v14i2.94392.
- [15] B. M. G. V. Sampreetha, and K. U. Meghana, "Sentiment Analysis for Low-Resource Languages Using a Hybrid XLM-Roberta and Bi-LSTM Model," in *2025 International Conference on Networks & Advances in Computational Technologies (NetACT)*, IEEE, Aug. 2025, pp. 1–6. doi: 10.1109/NetACT65906.2025.11188922.
- [16] W. A. Alvi, R. Talukdar, M. T. Hossain, and M. S. Sayed, "A Comparative Analysis of BERT, RoBERTa, and XLM-RoBERTa for Bengali SMS Multiclass Spam Detection," 2025.
- [17] H. J. G. Palacios, R. A. J. Toledo, G. A. H. Pantoja, and Á. A. M. Navarro, "A comparative between CRISP-DM and SEMMA through the construction of a MODIS repository for studies of land use and cover change," *Advances in Science, Technology and Engineering Systems*, vol. 2, no. 3, pp. 598–604, 2017, doi: 10.25046/aj020376.
- [18] N. A. P. Masaling and D. Suhartono, "Utilizing RoBERTa and XLM-RoBERTa pre-trained model for structured sentiment analysis," *International Journal of Informatics and Communication Technology*, vol. 13, no. 3, pp. 410–421, Dec. 2024, doi: 10.11591/ijict.v13i3.pp410-421.