



Penerapan Algoritma BERT dalam Analisis Sentimen Opini Publik terhadap Destinasi Wisata dengan Metode CRISP-DM

Saikin¹, Mohammad Taufan Asri Zaen^{2*}, Sofiansyah Fadli³, Hairul Fahmi⁴

^{1,2}Program Studi Sistem Informasi, STMIK Lombok

^{3,4}Program Studi Teknik Informatika STMIK Lombok

leken.apache@gmail.com, opanzain@gmail.com*, sofiansyah182@gmail.com, iroel.ami@gmail.com

Abstrak

Penelitian ini bertujuan untuk menganalisis opini publik terhadap destinasi wisata di Pulau Lombok menggunakan algoritma BERT (Bidirectional Encoder Representations from Transformers) dengan pendekatan metodologi CRISP-DM (Cross Industry Standard Process for Data Mining). Pulau Lombok dipilih sebagai studi kasus karena memiliki potensi pariwisata yang tinggi, namun pemanfaatan analisis sentimen berbasis pembelajaran mendalam terhadap ulasan wisatawan masih relatif terbatas. Data penelitian diperoleh dari platform Google Maps sebanyak 5.945 ulasan wisatawan, yang setelah melalui tahap pembersihan data (data cleaning) menghasilkan 4.972 data ulasan yang valid. Proses analisis mengikuti tahapan CRISP-DM hingga tahap evaluasi, meliputi pemahaman bisnis, pemahaman data, persiapan data, pemodelan, dan evaluasi. Model BERT dilatih melalui proses fine-tuning dengan penerapan weighted loss dan weighted random sampler untuk mengatasi permasalahan ketidakseimbangan kelas sentimen. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 0,9209, Weighted F1-Score sebesar 0,9054, dan Macro F1-Score sebesar 0,6293. Kinerja terbaik diperoleh pada kelas sentimen positif dengan nilai F1-score sebesar 0,96, diikuti oleh kelas negatif dengan nilai 0,71, sementara kelas netral menunjukkan performa terendah dengan nilai F1-score sebesar 0,22. Temuan ini menunjukkan bahwa algoritma BERT efektif dalam mengklasifikasikan sentimen ulasan berbahasa Indonesia pada domain pariwisata, meskipun masih terdapat tantangan dalam menangani kelas minoritas. Penelitian ini berkontribusi secara akademik dalam pengembangan analisis sentimen berbasis transformer dan secara praktis memberikan wawasan bagi pemangku kepentingan dalam meningkatkan kualitas layanan serta strategi pengelolaan destinasi wisata di Pulau Lombok.

Kata kunci: BERT, Analisis Sentimen, CRISP-DM, Opini Publik, Destinasi Wisata, Lombok

1. Latar Belakang

Pariwisata merupakan aktivitas yang diminati oleh berbagai kalangan, baik secara individu maupun dalam kelompok [1] karena Berwisata merupakan salah satu kebutuhan yang perlu dipenuhi oleh setiap kalangan, karena wisata berfungsi sebagai sarana untuk menghilangkan rasa penat [2]. Kebutuhan akan berwisata tentu harus didukung oleh keberadaan lokasi wisata yang memadai serta mampu memberikan pengalaman yang berkesan bagi para wisatawan. Oleh karena itu, pengelolaan dan pengembangan destinasi wisata yang berkualitas menjadi hal penting untuk meningkatkan daya tarik wisata serta kepuasan pengunjung. Perkembangan teknologi informasi dapat dimanfaatkan untuk memperoleh informasi mengenai berbagai tempat wisata. Melalui teknologi ini, wisatawan dapat menemukan ulasan dan opini terkait destinasi yang telah mereka kunjungi. Opini tersebut dapat menjadi bahan evaluasi bagi para pemangku kepentingan (stakeholder) dalam meningkatkan kualitas pelayanan, fasilitas, dan pengalaman wisatawan [3].

Pulau Lombok memiliki beragam destinasi wisata yang dapat memenuhi kebutuhan wisatawan, mulai dari pantai yang indah, gugusan Gili yang eksotis, hingga pegunungan yang menantang untuk dijelajahi. Keindahan alam yang dimiliki Pulau Lombok menjadikannya salah satu tujuan wisata favorit, baik bagi wisatawan domestik maupun mancanegara. Selain itu, fasilitas dan infrastruktur yang terus dikembangkan semakin mendukung kenyamanan serta pengalaman berwisata yang menyenangkan. Menurut data Badan Pusat Statistik (BPS) Nusa Tenggara Barat (NTB) jumlah kunjungan wisatawan dari beberapa kabupaten dan kota yakni kota mataram jumlah wisatawan 1.480.711, kabupaten Lombok Barat 1.087.989, kabupaten Lombok Tengah 782.809, kabupaten Lombok Timur 662.477 dan kabupaten Lombok Utara 376.138. Dengan jumlah wisatawan tersebut, angka ini tergolong kecil bagi suatu daerah yang terkenal dengan destinasi wisata yang indah. Oleh karena itu,

perlu dilakukan analisis guna memahami faktor-faktor yang memengaruhi minat wisatawan. Salah satu pendekatan yang dapat digunakan adalah analisis opini publik melalui text analysis, yang berperan penting dalam mengidentifikasi tren, preferensi, serta masukan dari wisatawan. Hasil analisis ini dapat menjadi dasar bagi pengembangan strategi pariwisata yang lebih efektif guna meningkatkan daya tarik dan kepuasan wisatawan.

Berdasarkan penelitian terdahulu, belum banyak studi yang secara khusus menggabungkan metode CRISP-DM dengan algoritma BERT untuk analisis sentimen opini wisatawan berbahasa Indonesia. Sebagian besar penelitian sebelumnya masih berfokus pada algoritma konvensional seperti SVM, Naive Bayes, atau Decision Tree, yang memiliki keterbatasan dalam memahami konteks bahasa alami secara mendalam. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model analisis sentimen berbasis BERT yang diterapkan dalam kerangka metodologi CRISP-DM guna mengidentifikasi persepsi publik terhadap destinasi wisata di Pulau Lombok. Pendekatan ini diharapkan dapat memberikan hasil klasifikasi sentimen yang lebih akurat dan komprehensif dibandingkan pendekatan tradisional.

Adapun rumusan masalah dalam penelitian ini adalah: (1) bagaimana opini publik terhadap destinasi wisata di Pulau Lombok dapat diidentifikasi melalui analisis sentimen berbasis teks, dan (2) seberapa baik performa algoritma BERT dalam mengklasifikasikan sentimen positif, negatif, dan netral pada data ulasan wisata. Tujuan penelitian ini adalah untuk membangun dan mengevaluasi model analisis sentimen berbasis BERT yang mampu merepresentasikan opini publik secara akurat, serta memberikan gambaran persepsi wisatawan terhadap kualitas dan daya tarik destinasi wisata di Pulau Lombok. Analisis teks digunakan untuk mengetahui opini publik terhadap destinasi wisata di Pulau Lombok dengan menerapkan algoritma BERT melalui pendekatan metodologi CRISP-DM (Cross-Industry Standard Process Model for Data Mining), yaitu standar terbuka lintas industri dalam proses penambangan data[4]. Pendekatan CRISP-DM dapat digunakan untuk menganalisis komentar wisatawan terhadap destinasi wisata tertentu [5] yang memiliki 6 tahapan yaitu Business Understanding (Pemahaman Bisnis), Data Understanding (Pemahaman Data), Data Preparation (Persiapan Data), Modeling (Pemodelan), Evaluation (Evaluasi) dan Deployment (Penerapan)[6]. Data teks yang digunakan berasal dari ulasan (review) di Google Maps, salah satu fitur yang memungkinkan pengguna internet untuk memberikan komentar dan penilaian terhadap suatu objek. Melalui fitur ini, pengguna dapat menuliskan ulasan serta memberikan peringkat terhadap berbagai objek, termasuk tempat wisata [7].

[Penelitian terdahulu pada tahun 2023 telah melakukan analisis sentimen terhadap ulasan objek wisata di Ubud, Bali, dengan menggunakan algoritma SVM. Hasil analisis menunjukkan bahwa dari 669 data uji, terdapat 551 ulasan bernilai positif dan 116 ulasan bernilai negative [1]. Penelitian berikutnya yang dilakukan pada tahun 2023 membahas penerapan metode CRISP-DM dalam klasifikasi sentimen dan analisis perilaku pembelian layanan akomodasi hotel berbasis algoritma Decision Tree. Hasil penelitian ini menunjukkan bahwa metode CRISP-DM dengan algoritma Decision Tree sangat relevan untuk mengoptimalkan performa Customer Relationship Management (CRM) di industri perhotelan [5]. Penelitian berikutnya dilakukan pada tahun 2024 melakukan analisis kerja untuk mendeteksi berita hoaks dengan menggunakan algoritma BERT. Hasil evaluasi menunjukkan bahwa model klasifikasi BERT memiliki akurasi sebesar 76% pada data validasi dalam mengklasifikasikan berita hoaks. Hal ini mencerminkan performa model Machine Learning dalam mendeteksi berita hoaks. F1-Score untuk label positif tercatat sebesar 0,72, dengan presisi 0,87 dan recall 0,62 untuk label positif. Sementara itu, hasil evaluasi model klasifikasi BERT Multilingual menunjukkan akurasi yang lebih rendah dibandingkan model klasifikasi BERT, yakni sebesar 63% pada data validasi. Hal ini menunjukkan bahwa performa model Machine Learning dalam mengklasifikasikan berita hoaks masih dapat ditingkatkan [8]. Analisis sentimen merupakan salah satu teknik yang dapat digunakan untuk menganalisis komentar publik terhadap destinasi wisata. Namun, terdapat beberapa permasalahan dalam proses analisis sentimen tersebut, antara lain data yang tidak terstruktur, rendahnya akurasi hasil klasifikasi, serta keterbatasan analisis sentimen yang hanya menghasilkan kategori sentimen positif dan negative [9][10]. Berdasarkan penelitian sebelumnya, penelitian selanjutnya adalah menganalisis opini publik terhadap destinasi wisata menggunakan algoritma BERT dengan menerapkan metode CRISP-DM. BERT adalah deep learning model Pemahaman Konteks Dua Arah (Bidirectional Context Understanding): Berbeda dengan model seperti RNN dan LSTM yang memproses teks secara sekuensial (kiri ke kanan atau sebaliknya), BERT membaca seluruh urutan kata secara simultan. Pendekatan dua arah ini memungkinkan BERT untuk memahami konteks kata berdasarkan informasi sebelum dan sesudahnya, sehingga menghasilkan representasi teks yang lebih kaya dan akurat[11][12]. CRISP-DM merupakan metode standar dalam data mining dan analisis data yang terdiri dari enam tahapan utama[13][14]. Dalam penelitian ini, penerapan metode CRISP-DM dibatasi hanya sampai pada tahap evaluasi. Tahap deployment (penerapan atau implementasi) tidak dilakukan karena fokus penelitian terletak pada pengukuran kinerja model yang telah dibangun. Dengan demikian, hasil evaluasi digunakan untuk menilai

sejauh mana model mampu memberikan hasil analisis yang akurat dan sesuai dengan tujuan penelitian, tanpa melanjutkan ke tahap penerapan sistem secara nyata.

Secara ilmiah, penelitian ini berkontribusi terhadap pengembangan bidang Natural Language Processing (NLP) berbahasa Indonesia, khususnya dalam penerapan model transformer seperti BERT pada domain pariwisata yang masih terbatas kajiannya. Sementara itu, secara praktis, hasil penelitian ini diharapkan dapat menjadi dasar bagi pemerintah daerah dan pengelola destinasi wisata dalam merancang strategi peningkatan pelayanan, promosi, dan pengelolaan destinasi wisata berbasis data opini publik. Dengan demikian, penelitian ini tidak hanya memiliki nilai akademik tetapi juga implikasi langsung bagi pengembangan sektor pariwisata di Pulau Lombok.

2. Metode Penelitian

Penelitian ini menerapkan algoritma BERT dengan pendekatan metodologi CRISP-DM, yakni proses standar terbuka lintas industri untuk penambangan data. Metode ini banyak digunakan karena penerapannya yang efektif serta memiliki tahapan yang mudah diaplikasikan[4].



Gambar 1. Tahapan Penelitian Dengan CRISP-DM

Menjelaskan metode penelitian dan teknik penelitian yang digunakan. Jelaskan secara ringkas, tetapi tetap akurat seperti ukuran, volume, replikasi, dan teknik pengerjaan. Untuk metode baru harus dijelaskan secara rinci bagi peneliti lain untuk mereproduksi percobaan. Sementara metode yang sudah mapan dapat dijelaskan dengan memilih referensi[4-6].

2.1. Business Understanding (Pemahaman Bisnis)

Dalam penelitian ini, isu utama yang menjadi fokus adalah bagaimana opini publik yang tersebar di berbagai platform ulasan (khususnya Google Maps) dapat dimanfaatkan untuk mengevaluasi kualitas dan daya tarik destinasi wisata di Pulau Lombok. Data ulasan wisatawan yang bersifat teks ini mengandung informasi berharga mengenai kepuasan, keluhan, maupun saran yang dapat digunakan sebagai dasar pengambilan keputusan strategis oleh pengelola dan pemangku kepentingan di sektor pariwisata. Mengidentifikasi kebutuhan informasi pemangku kepentingan (dinas pariwisata, pengelola destinasi, pelaku usaha lokal) terkait opini publik tentang destinasi wisata di Pulau Lombok. Menentukan tujuan analisis sentimen, yaitu untuk mengetahui distribusi sentimen positif, negatif, maupun netral terhadap berbagai destinasi wisata. Merumuskan indikator keberhasilan, seperti tingkat akurasi dan F1-score model BERT pada klasifikasi sentimen serta kemampuan hasil analisis untuk memberikan masukan yang dapat diimplementasikan dalam peningkatan kualitas pelayanan, fasilitas, dan strategi pemasaran destinasi wisata. Memastikan hasil penelitian dapat digunakan sebagai dasar pengambilan kebijakan atau strategi yang lebih tepat untuk meningkatkan daya tarik wisata dan kepuasan wisatawan.

2.2. Data Understanding (Pemahaman Data)

Tahap Data Understanding bertujuan untuk memperoleh pemahaman yang mendalam terhadap karakteristik data yang akan digunakan dalam penelitian. Pada penelitian ini, data yang dianalisis berupa teks ulasan (review) destinasi wisata di Pulau Lombok yang diperoleh dari platform Google Maps. Data tersebut berisi komentar dan penilaian yang diberikan pengguna internet terhadap berbagai objek wisata, sehingga mencerminkan opini publik yang dapat dimanfaatkan untuk evaluasi kualitas destinasi wisata.

Sumber data. Data penelitian ini diperoleh dari ulasan destinasi wisata di Pulau Lombok yang tersedia pada platform Google Maps. Platform ini dipilih karena memiliki cakupan pengguna yang luas serta menyediakan fitur

ulasan yang memuat pengalaman, penilaian, dan komentar wisatawan terhadap berbagai objek wisata. Data yang diambil mencakup teks ulasan beserta informasi pendukung yang tersedia secara publik, seperti nama objek wisata, rating bintang, dan tanggal ulasan. Pemanfaatan Google Maps sebagai sumber data diharapkan dapat memberikan gambaran representatif mengenai opini publik untuk dianalisis sentimennya.

Jumlah data. Data yang diperoleh dari Google Maps berjumlah 5.945 baris dan terdiri atas lima fitur, yaitu nama objek wisata, nama pengulas, rating, teks ulasan, waktu, dan foto. Fitur 'objek wisata' memuat nama-nama destinasi yang dikaji, 'rating' menunjukkan penilaian pengguna pada skala 1–5 yang tersedia di Google Maps, sedangkan fitur 'teks' berisi ulasan atau review yang ditulis pengguna.

Analisi data

```
=====
STATISTIK DESKRIPTIF RATING
=====
count      5945.000000
mean       4.542809
std        0.899693
min        1.000000
25%        4.000000
50%        5.000000
75%        5.000000
max        5.000000
```

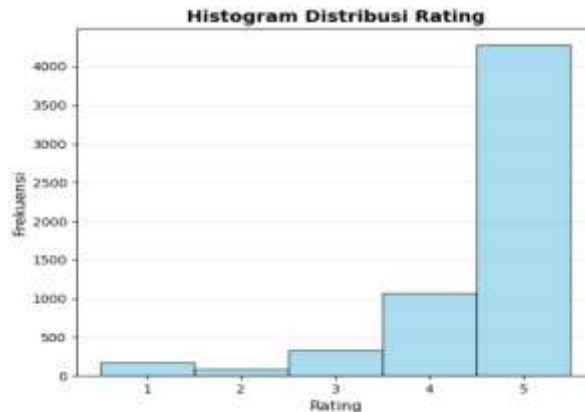
Gambar 2. Distribusi Rating Secara Keseluruhan

Berdasarkan hasil analisis deskriptif, jumlah data rating yang diperoleh sebanyak 5.945 ulasan dengan nilai rata-rata (mean) sebesar 4,54 dan standar deviasi 0,89. Nilai rating minimum adalah 1 dan maksimum 5, dengan median (50%) berada pada nilai 5. Hal ini menunjukkan bahwa sebagian besar ulasan cenderung memberikan penilaian tinggi.

Frekuensi setiap rating:		Persentase setiap rating:	
rating		rating	
1	177	1	2.98
2	90	2	1.51
3	336	3	5.65
4	1068	4	17.96
5	4274	5	71.89

Gambar 3. Menghitung frekuensi setiap rating

Distribusi frekuensi rating memperlihatkan bahwa rating 5 mendominasi dengan jumlah 4.274 ulasan atau sekitar 71,89% dari total data. Rating 4 menempati posisi kedua dengan 1.068 ulasan (17,96%). Sementara itu, rating rendah seperti 1, 2, dan 3 hanya menyumbang persentase kecil masing-masing sebesar 2,98%, 1,51%, dan 5,65%.



Gambar 4. Visualisasi Distribusi Rating

Visualisasi dalam bentuk histogram semakin memperjelas bahwa distribusi rating sangat terpusat pada nilai tinggi (4 dan 5). Artinya, secara umum pengguna memberikan pengalaman yang positif terhadap destinasi wisata yang diulas dalam dataset

Deteksi Spam/Noise

```
Total review: 5945
Review bersih (spam score = 0): 3152 (53.02%)
Review terindikasi spam (spam score > 0): 2793 (46.98%)

Breakdown berdasarkan kategori:
- High Risk Spam: 145 (2.44%)
- Medium Risk Spam: 249 (4.19%)
- Low Risk Spam: 2399 (40.35%)
- Clean: 3152 (53.02%)
```

Gambar 5. Deteksi Spam/Noise

Berdasarkan hasil deteksi spam/noise pada 5.945 ulasan, ditemukan bahwa sebanyak 3.152 ulasan (53,02%) termasuk kategori review bersih (clean), sedangkan 2.793 ulasan (46,98%) terindikasi sebagai spam dengan tingkat risiko yang berbeda. Dari total review yang terindikasi spam, terdapat 145 ulasan (2,44%) yang masuk dalam kategori High Risk Spam, 249 ulasan (4,19%) dalam kategori Medium Risk Spam, dan 2.399 ulasan (40,35%) tergolong Low Risk Spam. Hasil ini menunjukkan bahwa hampir setengah dari data ulasan yang diperoleh mengandung potensi noise atau spam, dengan dominasi terbesar pada kategori Low Risk Spam. Dengan demikian, proses pembersihan data (data cleaning) sangat penting dilakukan agar analisis sentimen yang dihasilkan lebih akurat dan bebas dari bias yang ditimbulkan oleh ulasan spam

2.3. Data Preparation (Persiapan Data)

Cleaning dan filtering dataset

```
📊 Data Original: 5945 baris
📊 Data Setelah Cleaning: 4972 baris
📊 Data yang dihapus: 973 baris (16.37%)
```

Gambar 6. Cleaning dan Filtering data

Pada tahap data cleaning dan filtering, dilakukan proses pembersihan terhadap data ulasan yang telah dikumpulkan. Dari total 5.945 baris data asli, setelah melalui proses pembersihan, jumlah data yang dapat digunakan berkurang menjadi 4.972 baris. Artinya, terdapat 973 baris data (16,37%) yang dihapus karena tidak memenuhi kriteria analisis. Penghapusan ini dilakukan untuk memastikan kualitas data tetap terjaga, misalnya dengan menghilangkan entri yang duplikat, data yang kosong, ulasan yang tidak relevan, atau ulasan yang terlalu pendek dan tidak informatif. Dengan demikian, data yang tersisa diharapkan lebih bersih, valid, dan representatif untuk digunakan pada tahap analisis selanjutnya.

Tokenization BERT

```
Tokenization Info:
• Tokenizer: IndoBERT / Multilingual BERT
• Max Length: 128 tokens
• Avg Token Length: 26.2 tokens
• Reviews > max_length: 0 (0.00%)
```

Gambar 7. Tokenize

Pada tahap tokenization, data ulasan diproses menggunakan IndoBERT / Multilingual BERT Tokenizer untuk mengubah teks menjadi representasi token yang dapat dipahami oleh model BERT. Panjang maksimum token

ditetapkan sebesar 128 tokens, dengan rata-rata panjang ulasan sebesar 26,2 tokens. Hasil tokenisasi menunjukkan bahwa tidak ada ulasan yang melebihi batas maksimum panjang token (0,00%). Hal ini mengindikasikan bahwa seluruh data ulasan dapat diproses tanpa perlu dilakukan truncation (pemotongan teks) ataupun padding berlebih. Dengan demikian, proses tokenisasi dapat berjalan secara optimal dan tetap mempertahankan informasi penting dalam setiap ulasan.

Labeling Data

```
Label Sentimen:  
• Negatif: 218 (4.4%)  
• Netral: 281 (5.7%)  
• Positif: 4473 (90.0%)
```

Gambar 8 Labeling data

Hasil pelabelan sentimen menunjukkan bahwa mayoritas ulasan wisata cenderung bernada positif dengan jumlah 4.473 ulasan (90,0%). Sementara itu, terdapat 281 ulasan netral (5,7%) dan hanya 218 ulasan negatif (4,4%). Distribusi ini mengindikasikan bahwa secara umum, opini publik terhadap destinasi wisata yang diteliti lebih dominan memberikan pengalaman baik dan kepuasan yang tinggi. Proporsi ulasan negatif relatif kecil, sehingga dapat disimpulkan bahwa persepsi wisatawan terhadap destinasi tersebut berada dalam kategori sangat baik.

Splitting Data

```
Dataset Split:  
• Train: 3480 samples (70%)  
• Validation: 746 samples (15%)  
• Test: 746 samples (15%)
```

Gambar 9. Data Splitting

Untuk mendukung proses pelatihan model, data ulasan dibagi menjadi tiga bagian utama, yaitu data latih (training set), data validasi (validation set), dan data uji (test set). Pembagian dilakukan dengan proporsi 70% : 15% : 15%. Secara rinci, terdapat 3.480 sampel (70%) yang digunakan sebagai data latih untuk membangun model. Selanjutnya, 746 sampel (15%) digunakan sebagai data validasi untuk menguji kinerja model selama proses pelatihan dan mencegah terjadinya overfitting. Sementara itu, 746 sampel (15%) sisanya digunakan sebagai data uji guna mengevaluasi performa model secara keseluruhan pada data yang belum pernah dilihat sebelumnya.

2.4. Pemodelan Data

Tahap ini berfokus pada membangun model analisis sentimen menggunakan algoritma BERT (Bidirectional Encoder Representations from Transformers). Setelah data sudah melalui proses pembersihan, pelabelan, dan tokenisasi, maka tahap pemodelan dilakukan dengan langkah-langkah utama berikut:

Model BERT/IndoBERT

Pada bagian ini, digunakan model berbasis transformer seperti BERT atau IndoBERT untuk memahami konteks bahasa Indonesia dari ulasan wisata. Model ini mampu menangkap makna kata berdasarkan konteks kalimat secara dua arah (bidirectional).

Fine-tuning Model

Proses fine-tuning dilakukan dengan melatih model BERT menggunakan dataset ulasan yang telah diberi label sentimen (positif, negatif, netral). Tujuannya agar model dapat menyesuaikan diri dengan karakteristik data dan domain pariwisata yang digunakan dalam penelitian.

Hyperparameter Tuning

Langkah ini bertujuan untuk mengoptimalkan performa model dengan menyesuaikan parameter penting, seperti learning rate, batch size, dan epoch. Pemilihan parameter yang tepat akan mempengaruhi akurasi dan kecepatan pelatihan.

Pelacakan Arsitektur

Meliputi pemantauan struktur dan performa model selama proses pelatihan — seperti nilai loss, accuracy, serta hasil validasi. Proses ini memastikan bahwa model tidak mengalami overfitting dan mampu melakukan generalisasi dengan baik pada data baru.

2.5 Evaluasi Model

Tahap ini bertujuan untuk mengukur dan menilai kinerja model BERT yang telah dibangun pada tahap pemodelan, serta memastikan bahwa hasilnya sesuai dengan tujuan bisnis dan penelitian.

Akurasi, Precision, Recall

Akurasi menunjukkan seberapa sering model memberikan prediksi yang benar. Precision mengukur ketepatan model dalam memprediksi sentimen positif atau negatif. Recall menunjukkan kemampuan model dalam menangkap seluruh data yang relevan (misalnya semua ulasan yang benar-benar positif).

F1-score & Confusion Matrix

F1-score merupakan kombinasi dari precision dan recall untuk memberikan ukuran kinerja yang lebih seimbang, terutama jika jumlah data tiap kelas tidak seimbang. Confusion Matrix digunakan untuk melihat rincian jumlah prediksi benar dan salah untuk tiap kelas (positif, negatif, netral), sehingga memudahkan analisis kesalahan model.

Validasi dengan Tujuan Bisnis

Hasil evaluasi tidak hanya dilihat dari aspek teknis, tetapi juga dikaitkan kembali dengan tujuan awal penelitian—misalnya, apakah model dapat membantu pihak pengelola destinasi wisata memahami opini publik dan meningkatkan strategi pelayanan atau promosi.

3. Hasil dan Diskusi

3.1. Modeling

Dataset yang digunakan dalam penelitian ini terdiri dari total 4.972 sampel ulasan. Data tersebut dibagi menjadi tiga bagian utama untuk mendukung proses pelatihan dan evaluasi model. Sebanyak 3.480 sampel (sekitar 70%) digunakan sebagai training set, yaitu data yang digunakan untuk melatih model BERT agar dapat mengenali pola sentimen dalam teks. Selanjutnya, 746 sampel (sekitar 15%) dialokasikan sebagai validation set yang berfungsi untuk memantau kinerja model selama proses pelatihan dan mencegah terjadinya overfitting. Sementara itu, 746 sampel lainnya (sekitar 15%) digunakan sebagai test set, yang dipakai untuk mengukur performa akhir model setelah pelatihan selesai. Pembagian data ini dilakukan secara proporsional agar hasil yang diperoleh lebih representatif dan dapat mencerminkan kemampuan model dalam memprediksi sentimen pada data baru.

Distribusi Kelas

```
Training Set Distribution:
Negatif (0): 152 samples ( 4.37%)
Netral (1): 197 samples ( 5.66%)
Positif (2): 3131 samples (89.97%)
```

Gambar 10 Distribusi Kelas

Distribusi data pada training set menunjukkan adanya ketidakseimbangan jumlah sampel antar kelas sentimen. Dari total 3.480 data pelatihan, terdapat 152 sampel (4,37%) dengan label negatif, 197 sampel (5,66%) dengan label netral, dan 3.131 sampel (89,97%) dengan label positif. Distribusi ini memperlihatkan bahwa sebagian besar ulasan memiliki sentimen positif terhadap destinasi wisata, sedangkan ulasan dengan sentimen negatif dan netral jumlahnya relatif sedikit. Ketidak seimbangan ini dapat memengaruhi kinerja model, sehingga perlu diperhatikan dalam proses pelatihan, misalnya dengan menerapkan teknik class weighting atau resampling agar model tidak bias terhadap kelas dengan jumlah data terbanyak.

Class Weights untuk Weighted Loss

Pada metode ini, diberikan bobot (weight) yang lebih besar untuk kelas dengan jumlah data yang lebih sedikit agar kesalahan prediksi pada kelas minoritas mendapat penalti lebih tinggi. Bobot yang digunakan masing-masing adalah 7,6316 untuk kelas negatif, 5,8883 untuk kelas netral, dan 0,3705 untuk kelas positif. Dengan pendekatan ini, model diharapkan dapat belajar secara lebih adil terhadap setiap kelas.

Weighted Random Sampler

Selain pemberian bobot, digunakan pula teknik Weighted Random Sampler untuk melakukan oversampling pada kelas minoritas. Strategi ini memastikan bahwa setiap batch data pelatihan memiliki distribusi kelas yang lebih seimbang, sehingga model memperoleh kesempatan yang sama dalam mempelajari semua jenis sentimen

Konfigurasi Hyperparameter

```
Hyperparameter Configuration:
-----
max_len           : 128
batch_size        : 16
learning_rate     : 2e-05 (scientific notation)
epochs            : 5
dropout           : 0.3
```

Gambar 11. Konfigurasi Hyperparameter

Dalam proses pelatihan model BERT untuk analisis sentimen, dilakukan konfigurasi hyperparameter guna mengoptimalkan kinerja model. Hyperparameter merupakan parameter yang nilainya ditentukan sebelum proses pelatihan dimulai dan berpengaruh terhadap performa serta stabilitas model. `max_len = 128`, yang berarti panjang maksimum setiap urutan teks (jumlah token) dibatasi hingga 128 token. Batas ini digunakan agar proses komputasi menjadi lebih efisien tanpa kehilangan terlalu banyak informasi penting dari ulasan. `batch_size = 16`, menunjukkan bahwa dalam setiap iterasi pelatihan, model memproses 16 sampel sekaligus. Ukuran batch ini dipilih untuk menyeimbangkan antara efisiensi komputasi dan kestabilan pembaruan bobot model. `learning_rate = 2e-05`, atau 0,00002 dalam notasi desimal, digunakan sebagai laju pembelajaran yang relatif kecil agar model dapat melakukan penyesuaian bobot secara bertahap dan menghindari konvergensi prematur. `epochs = 5`, menandakan bahwa seluruh dataset pelatihan dilalui sebanyak lima kali untuk memastikan model memperoleh pemahaman yang cukup terhadap pola data tanpa mengalami overfitting. `dropout = 0.3`, yaitu tingkat dropout sebesar 30% yang digunakan untuk mencegah overfitting dengan cara menonaktifkan sebagian neuron secara acak selama proses pelatihan.

Training

Training model BERT dengan konfigurasi loss function menggunakan Weighted Cross Entropy Loss, yang dirancang untuk menangani ketidak seimbangan kelas pada dataset. Model BERT yang diinisialisasi memiliki total 124.443.651 parameter, seluruhnya bersifat trainable, artinya semua parameter diperbarui selama proses pelatihan. Optimizer yang digunakan adalah AdamW dengan learning rate sebesar $2e-05$, sedangkan scheduler yang diterapkan adalah Linear Warmup and Decay dengan total 1.090 langkah pembaruan (steps). Pelatihan dilakukan selama 5 epoch, dan hasilnya menunjukkan peningkatan kinerja model dari waktu ke waktu. Pada epoch pertama, nilai train loss tercatat 0.3565 dengan train accuracy sebesar 0.6112, sedangkan validation accuracy masih rendah yaitu 0.4115. Namun, seiring bertambahnya epoch, akurasi model meningkat signifikan.

Pada epoch kedua, akurasi pelatihan naik menjadi 0.9158, dan akurasi validasi mencapai 0.8405. Peningkatan ini menunjukkan bahwa model mulai memahami pola data dengan lebih baik. Kemudian pada epoch ketiga, akurasi pelatihan meningkat lagi menjadi 0.9739, dengan validation accuracy sebesar 0.8981, menandakan model semakin stabil dan generalisasi mulai terbentuk dengan baik. Pada epoch keempat, model mencatat train accuracy sebesar 0.9920 dan validation accuracy 0.8968, yang menunjukkan bahwa model hampir mencapai konvergensi. Akhirnya, pada epoch kelima, model memperoleh hasil terbaik dengan train accuracy 0.9960 dan validation accuracy 0.9048, serta validation loss sebesar 2.5204. Nilai akurasi validasi di atas 90% menandakan bahwa model telah mampu mengenali pola sentimen dengan sangat baik.

- Best Val Accuracy : 0.9048
- Training Time : 126.08 minutes

Gambar 12. Hasil Training Model

Hasil akhir dari proses pelatihan model BERT menunjukkan bahwa model mencapai nilai akurasi validasi terbaik (Best Validation Accuracy) sebesar 0.9048 atau sekitar 90,48%. Nilai ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengklasifikasikan sentimen ulasan pada data yang tidak digunakan selama pelatihan. Selain itu, total waktu yang dibutuhkan untuk menyelesaikan seluruh proses pelatihan adalah sekitar 126,08 menit. Durasi ini mencakup seluruh lima epoch pelatihan dengan penerapan Weighted Loss dan Weighted Sampler untuk menangani ketidakseimbangan data. Secara keseluruhan, hasil tersebut menunjukkan bahwa konfigurasi hyperparameter dan strategi penanganan imbalanced dataset yang diterapkan sudah efektif, karena model tidak hanya mencapai akurasi tinggi tetapi juga tetap stabil dalam waktu pelatihan yang relatif efisien.

- Test Accuracy : 0.9209
- Test Loss : 2.1487
- Macro F1-Score : 0.6293
- Weighted F1-Score : 0.9054

Gambar 13. Hasil Testing Model

Hasil pengujian model pada data uji (test set) menunjukkan bahwa model memiliki performa yang cukup baik dengan nilai akurasi sebesar 0,9209. Artinya, sekitar 92% dari data uji berhasil diklasifikasikan dengan benar oleh model. Nilai loss sebesar 2,1487 menunjukkan masih terdapat beberapa kesalahan prediksi yang perlu diperbaiki, meskipun secara umum performa model sudah cukup stabil. Dari sisi metrik evaluasi lainnya, diperoleh Macro F1-Score sebesar 0,6293, yang menggambarkan bahwa performa model belum seimbang di seluruh kelas, terutama pada kelas dengan jumlah data yang relatif sedikit. Sementara itu, Weighted F1-Score sebesar 0,9054 menunjukkan bahwa secara keseluruhan model memberikan hasil yang sangat baik ketika mempertimbangkan proporsi data pada tiap kelas. Dengan demikian, dapat disimpulkan bahwa model memiliki kemampuan generalisasi yang tinggi terhadap data mayoritas, namun masih memerlukan perbaikan dalam menangani kelas minoritas agar performanya lebih merata di seluruh kategori.

3.2. Evaluasi Model



Gambar 13. Validation Loss

Berdasarkan hasil pelatihan, grafik Training & Validation Loss menunjukkan bahwa nilai train loss mengalami penurunan yang signifikan pada setiap epoch hingga mendekati nol, menandakan bahwa model mampu mempelajari pola pada data latih dengan sangat baik. Namun, validation loss justru mengalami peningkatan seiring bertambahnya epoch, yang mengindikasikan adanya gejala overfitting, di mana model terlalu menyesuaikan diri terhadap data latih sehingga performanya pada data validasi menurun.



Gambar 14. Validation Accuracy

Sementara itu, pada grafik Training & Validation Accuracy, terlihat bahwa train accuracy meningkat tajam dan mencapai hampir 100%, sedangkan validation accuracy juga meningkat pada awal pelatihan namun kemudian cenderung stagnan di sekitar 90%. Pola ini kembali memperkuat indikasi bahwa model sangat baik dalam mengenali data latih tetapi kurang optimal dalam melakukan generalisasi terhadap data baru.

```

-----
              precision    recall  f1-score   support

   Negatif      0.81      0.64      0.71      33
    Netral      0.46      0.14      0.22      42
    Positif      0.93      0.98      0.96     671

 accuracy              0.92      746
 macro avg           0.73      0.59      0.63      746
 weighted avg        0.90      0.92      0.91      746

Per-Class Metrics:
Negatif : Precision=0.8077, Recall=0.6364, F1=0.7119
Netral   : Precision=0.4615, Recall=0.1429, F1=0.2182
Positif  : Precision=0.9335, Recall=0.9836, F1=0.9579

Macro F1-Score   : 0.6293
Weighted F1-Score : 0.9054
    
```

Gambar 15. Clasification Report

Berdasarkan hasil pengujian lebih lanjut terhadap data uji, diperoleh metrik evaluasi yang menunjukkan performa model secara per kelas. Model mencapai akurasi keseluruhan sebesar 0,92, yang menandakan bahwa sebagian besar prediksi sudah sesuai dengan label sebenarnya. Namun, jika dilihat berdasarkan metrik precision, recall, dan F1-score per kelas, terlihat bahwa performa model belum merata. Pada kelas positif, model menunjukkan kinerja yang sangat baik dengan nilai precision sebesar 0,93, recall 0,98, dan F1-score 0,96, yang berarti model mampu mengenali sebagian besar data positif dengan tingkat kesalahan yang sangat rendah. Sebaliknya, pada kelas netral, nilai F1-score hanya sebesar 0,22, menandakan bahwa model mengalami kesulitan dalam mengenali data pada kelas ini, kemungkinan karena jumlah data yang tidak seimbang atau kemiripan pola dengan kelas lain. Untuk kelas negatif, performanya berada di tingkat menengah dengan F1-score 0,71, yang masih menunjukkan kemampuan prediksi cukup baik. Secara keseluruhan, nilai Macro F1-Score sebesar 0,6293 menunjukkan bahwa rata-rata kinerja model di semua kelas masih dapat ditingkatkan, sedangkan Weighted F1-Score sebesar 0,9054 mengindikasikan bahwa model bekerja sangat baik ketika mempertimbangkan distribusi data antar kelas. Temuan ini konsisten dengan pola pada grafik loss dan accuracy sebelumnya, di mana model cenderung overfitting pada kelas dengan jumlah data dominan, khususnya kelas positif.

4. Kesimpulan

Penelitian ini bertujuan untuk menganalisis opini publik terhadap destinasi wisata di Pulau Lombok dengan menerapkan algoritma BERT (Bidirectional Encoder Representations from Transformers) melalui pendekatan metodologi CRISP-DM (Cross Industry Standard Process for Data Mining). Pendekatan ini digunakan karena mampu memberikan kerangka kerja sistematis mulai dari tahap pemahaman bisnis hingga evaluasi model, sehingga proses analisis dapat dilakukan secara terarah dan komprehensif. Hasil penelitian menunjukkan bahwa model BERT memiliki kinerja yang baik dalam mengklasifikasikan sentimen ulasan wisata dengan nilai akurasi sebesar 0,9209, Weighted F1-Score sebesar 0,9054, serta Macro F1-Score sebesar 0,6293. Nilai-nilai tersebut mengindikasikan bahwa model telah mampu mengenali pola sentimen dengan tingkat ketepatan yang tinggi, meskipun distribusi kinerja antar kelas belum seimbang. Kelas positif memperoleh hasil terbaik dengan F1-score sebesar 0,96, diikuti oleh kelas negatif dengan 0,71, sedangkan kelas netral menunjukkan hasil yang relatif rendah sebesar 0,22 akibat ketidakseimbangan data (class imbalance). Berdasarkan hasil training dan validation, model menunjukkan adanya kecenderungan overfitting, di mana training loss menurun tajam sementara validation loss meningkat pada setiap epoch. Hal ini mengindikasikan bahwa model telah mempelajari pola data latihan dengan sangat baik, namun masih perlu peningkatan kemampuan generalisasi terhadap data baru. Secara keseluruhan, penelitian ini membuktikan bahwa penerapan algoritma BERT efektif digunakan dalam analisis sentimen berbasis teks berbahasa Indonesia, khususnya untuk mengidentifikasi persepsi wisatawan terhadap destinasi wisata. Hasil analisis yang diperoleh dapat menjadi dasar bagi pemangku kepentingan, seperti pengelola destinasi wisata dan pemerintah daerah, dalam merumuskan strategi peningkatan kualitas layanan, fasilitas, serta promosi pariwisata yang lebih tepat sasaran. Sebagai tindak lanjut, penelitian mendatang disarankan untuk menerapkan teknik penyeimbangan data (data balancing), seperti SMOTE atau oversampling, serta mengeksplorasi model transformer lain yang lebih adaptif guna meningkatkan akurasi prediksi pada kelas minoritas dan mengurangi potensi overfitting pada model.

Referensi

1. I. W. B. Suryawan, N. W. Utami, and K. Q. Fredlina, "Analisis Sentimen Review Wisatawan pada Objek Wisata Ubud Menggunakan Algoritma Support Vector Machine," *J. Inform. Teknol. dan Sains*, vol. 5, no. 1, pp. 133–140, 2023.
2. R. Yusuf and M. Veranita, "Minat Berwisata Kaum Milenial Di Era New Normal," *J. Kepariwisata Indonesia. J. Penelit. dan Pengemb. Kepariwisata Indonesia*, vol. 15, no. 2, pp. 158–167, 2021, doi: 10.47608/jki.v15i22021.158-167.
3. F. A. Hizham, C. K. Murni, and M. Qori'atunnadiyah, "Uji Klasifikasi Algoritma Naive Bayes Classification dalam Analisis Sentimen Ulasan Puncak B29 Lumajang," *Progresif J. Ilm. Komput.*, vol. 20, no. 1, p. 361, 2024, doi: 10.35889/progresif.v20i1.1618.
4. R. Mas, R. W. Panca, K. Atmaja1, and W. Yustanti2, "Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers)," *Jeisbi*, vol. 02, no. 3, pp. 55–62, 2021.
5. Y. A. Singgalen, "Penerapan CRISP-DM dalam Klasifikasi Sentimen dan Analisis Perilaku Pembelian Layanan Akomodasi Hotel Berbasis Algoritma Decision Tree (DT)," *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, p. 237, 2023, doi: 10.30865/json.v5i2.7081.
6. N. C. Sastya and I. Nugraha, "Penerapan Metode CRISP-DM dalam Menganalisis Data untuk Menentukan Customer Behavior di MeatSolution," *Unistek*, vol. 10, no. 2, pp. 103–115, 2023, doi: 10.33592/unistek.v10i2.3079.
7. D. C. Rahmadani, S. Khomsah, and M. Y. Fathoni, "Analisis Emosi Wisatawan Menggunakan Metode Lexicon Text Analysis," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 13–27, 2024, doi: 10.28932/jutisi.v10i1.6690.
8. A. R. Hanum et al., "Mendeteksi Berita Hoaks Performance Analysis of the Bert Text Classification Algorithm," vol. 11, no. 3, pp. 537–546, 2024, doi: 10.25126/jtiik938093.
9. Y. Setiowati and A. Helen, "Klasifikasi Analisis Sentimen Mengenai Hotel Di Yogyakarta," *SCAN - J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, 2018, doi: 10.33005/scan.v13i1.1052.
10. O. Somantri and D. Dairoh, "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 2, p. 191, 2019, doi: 10.26418/jp.v5i2.32661.
11. A. S. Talaat, "Sentiment analysis classification system using hybrid BERT models," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00781-w.
12. U. Muhammadiyah Mataram, P. Hafidzah, S. Maryani, B. Yuliatin Ihsani, A. Kurniamala Niswariyana, and P. Bahasa, "Seminar Nasional Paedagoria Penerapan Deep Learning dalam Menganalisis Sentimen di Media Sosial," vol. 4, pp. 328–339, 2024.
13. V. Plotnikova, M. Dumas, and F. Milani, "Adaptations of data mining methodologies: A systematic literature review," *PeerJ Comput. Sci.*, vol. 6, pp. 1–43, 2020, doi: 10.7717/PEERJ-CS.267.
14. R. Wirth and J. Hipp, "CRISP-DM: towards a standard process model for data mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 29–39," *Proc. Fourth Int. Conf. Pract. Appl. Knowl. Discov. Data Min.*, no. 24959, pp. 29–39, 2000.
15. S. Cahyaningtyas, D. Hatta Fudholi, and A. Fathan Hidayatullah, "DeepLearning for Aspect-Based Sentiment Analysis on Indonesian Hotels Reviews," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 3, 2021, doi: 10.22219/kinetik.v6i3.1300.]