



ANALISIS ULASAN GOOGLE MAPS WISATA BAHARI LAMONGAN BERBASIS SERVQUAL MENGGUNAKAN ALGORITMA CLASSIFIER CHAINS

Alviana Imron Nur Halizah¹⁾

¹⁾ Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya, Surabaya, Indonesia
Email: alvianaimron.22039@mhs.unesa.ac.id

Abstract

User reviews on Google Maps are a valuable source of information that can be utilized to evaluate the service quality of a tourist destination based on visitors' actual experiences. However, the sheer volume of reviews makes manual analysis highly inefficient. This study aims to analyze the sentiment of visitor reviews of Wisata Bahari Lamongan (WBL), classify these reviews into SERVQUAL service quality dimensions using the Classifier Chains algorithm, and provide recommendations for service improvement based on Importance Performance Analysis (IPA). The research data was obtained from WBL Google Maps reviews and processed through the Knowledge Discovery in Database (KDD) stages, which included data selection, preprocessing, sentiment analysis using the `w11wo/indonesian-roberta-base-sentiment-classifier` model, SERVQUAL dimension mapping, and multilabel classification using the Classifier Chains algorithm with several base classifiers. The test results indicate that the combination of Classifier Chains with Logistic Regression yielded the best performance compared to other models, achieving a Macro F1-Score of 0.4774 and a Weighted F1-Score of 0.6304. This combination proved highly capable of modeling the interdependencies between SERVQUAL dimensions. Furthermore, the classification results were utilized in IPA and Latent Dirichlet Allocation (LDA) to identify specific service attributes that need to be maintained or improved. This research demonstrates that integrating sentiment analysis, multilabel classification, SERVQUAL, and IPA can produce a more comprehensive evaluation of service quality based on online reviews. Ultimately, it provides strategic recommendations for the management of Wisata Bahari Lamongan to enhance service quality and boost visitor satisfaction.

Keywords: Sentiment Analysis, SERVQUAL, Classifier Chains, Wisata Bahari Lamongan, Importance Performance Analysis.

Abstrak

Ulasan pengguna pada Google Maps merupakan sumber informasi yang dapat dimanfaatkan untuk mengevaluasi kualitas layanan suatu destinasi wisata berdasarkan pengalaman nyata pengunjung. Namun, banyaknya jumlah ulasan menyebabkan proses analisis secara manual menjadi tidak efisien. Penelitian ini bertujuan menganalisis sentimen ulasan pengunjung Wisata Bahari Lamongan (WBL), mengklasifikasikan ulasan ke dalam dimensi kualitas layanan SERVQUAL menggunakan algoritma Classifier Chains, serta memberikan rekomendasi peningkatan layanan berdasarkan Importance Performance Analysis (IPA). Data penelitian diperoleh dari ulasan Google Maps WBL yang melalui tahapan Knowledge Discovery in Database (KDD), meliputi data selection, preprocessing, analisis sentimen menggunakan model `w11wo/indonesian-roberta-base-sentiment-classifier`, pemetaan dimensi SERVQUAL, dan klasifikasi multilabel menggunakan algoritma Classifier Chains dengan beberapa base classifier. Hasil pengujian menunjukkan bahwa kombinasi Classifier Chains dengan Logistic Regression memberikan performa terbaik dibandingkan model lain, dengan nilai Macro F1-Score sebesar 0,4774 dan Weighted F1-Score sebesar 0,6304, sehingga mampu memodelkan keterkaitan antar dimensi SERVQUAL secara lebih baik. Selanjutnya, hasil klasifikasi dimanfaatkan dalam analisis IPA dan Latent Dirichlet Allocation (LDA) untuk mengidentifikasi atribut layanan yang perlu dipertahankan maupun ditingkatkan. Penelitian ini menunjukkan bahwa integrasi analisis sentimen, klasifikasi multilabel, SERVQUAL, dan IPA mampu menghasilkan evaluasi kualitas layanan berbasis ulasan daring secara lebih komprehensif serta memberikan rekomendasi strategis bagi pengelola Wisata Bahari Lamongan dalam meningkatkan kualitas pelayanan dan kepuasan pengunjung.

Kata Kunci: Analisis Sentimen, SERVQUAL, Classifier Chains, Wisata Bahari Lamongan, Importance Performance Analysis.



PENDAHULUAN

Persaingan destinasi wisata di Indonesia semakin meningkat seiring bertambahnya jumlah wisatawan domestik maupun mancanegara. Kondisi tersebut mendorong setiap pengelola destinasi wisata untuk terus meningkatkan kualitas layanan agar mampu mempertahankan daya saing dan memberikan pengalaman terbaik kepada pengunjung (Rizaty, 2024). Salah satu destinasi wisata unggulan di Jawa Timur adalah Wisata Bahari Lamongan (WBL) yang dikelola oleh PT Bumi Lamongan Sejati dan menjadi ikon pariwisata Kabupaten Lamongan. Berdasarkan data Dinas Pariwisata dan Kebudayaan Kabupaten Lamongan, jumlah kunjungan wisatawan pada tahun 2024 mencapai 4,3 juta, dengan sebanyak 382.324 pengunjung berasal dari WBL (Sudjarwo, 2025). Selain memiliki tingkat kunjungan yang tinggi, WBL juga memperoleh lebih dari 5.600 ulasan pada Google Maps dengan rata-rata penilaian 4,6 dari 5 bintang. Besarnya jumlah ulasan tersebut menjadikan Google Maps sebagai sumber informasi yang potensial untuk mengevaluasi kualitas layanan berdasarkan pengalaman nyata pengunjung.

Ulasan digital pada Google Maps berisi opini, kritik, maupun apresiasi terhadap berbagai aspek pelayanan yang diterima pengunjung. Informasi tersebut dapat dimanfaatkan sebagai dasar evaluasi kualitas layanan secara lebih objektif dibandingkan hanya mengandalkan survei atau kuesioner. Namun, banyaknya jumlah ulasan serta bentuk data yang tidak terstruktur menyebabkan proses analisis secara manual menjadi kurang efektif. Oleh karena itu, diperlukan pendekatan berbasis Natural Language Processing (NLP) yang mampu mengolah data teks secara otomatis sehingga informasi penting yang terkandung dalam ulasan dapat diekstraksi secara lebih efisien.

Untuk memahami kecenderungan opini pengunjung, penelitian ini menerapkan analisis sentimen menggunakan model `w11wo/indonesian-roberta-base-sentiment-classifier`. Model tersebut merupakan hasil fine-tuning RoBERTa untuk bahasa Indonesia yang mampu memahami konteks kalimat secara dua arah sehingga menghasilkan klasifikasi sentimen yang lebih akurat dibandingkan pendekatan konvensional [1]. Selain itu, penelitian oleh [2] menunjukkan bahwa model RoBERTa Indonesia memiliki performa yang baik dalam mengklasifikasikan sentimen teks berbahasa Indonesia tanpa memerlukan pelatihan ulang dari awal. Hasil analisis sentimen kemudian dimanfaatkan sebagai dasar dalam proses evaluasi kualitas layanan.

Evaluasi kualitas layanan pada penelitian ini menggunakan model SERVQUAL yang dikembangkan oleh [3]. Model SERVQUAL mengukur kualitas layanan berdasarkan lima dimensi, yaitu tangibles, reliability, responsiveness, assurance, dan empathy [3]. Model ini telah banyak digunakan dalam penelitian kualitas layanan karena mampu menggambarkan kesenjangan antara persepsi pelanggan dan kualitas pelayanan yang diberikan [4]. Penelitian lain juga menunjukkan bahwa penerapan SERVQUAL efektif dalam mengidentifikasi aspek layanan yang perlu dipertahankan maupun ditingkatkan pada sektor pariwisata [5]. Meskipun demikian, satu ulasan pengunjung sering kali mengandung lebih dari satu dimensi

SERVQUAL sehingga diperlukan pendekatan klasifikasi multilabel.

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan algoritma Classifier Chains sebagai metode klasifikasi multilabel. Algoritma ini menghubungkan beberapa binary classifier secara berurutan, di mana hasil prediksi pada label sebelumnya digunakan sebagai fitur tambahan pada proses klasifikasi label berikutnya sehingga mampu menangkap hubungan antar label SERVQUAL [6]. Pendekatan ini terbukti menghasilkan performa yang lebih baik dibandingkan metode Binary Relevance yang mengasumsikan setiap label bersifat independen [7]. Selanjutnya, hasil klasifikasi dimanfaatkan dalam Importance Performance Analysis (IPA) untuk memetakan tingkat kepentingan dan kinerja setiap dimensi layanan, sedangkan Latent Dirichlet Allocation (LDA) digunakan untuk mengidentifikasi topik dominan pada setiap kelompok ulasan sehingga rekomendasi yang dihasilkan menjadi lebih terarah (Royani et al., 2024; Palinggi et al., n.d.).

Berdasarkan telaah penelitian terdahulu, sebagian besar penelitian mengenai evaluasi kualitas layanan masih menggunakan data primer berupa survei atau kuesioner. Sementara itu, penelitian yang memanfaatkan ulasan Google Maps sebagai sumber data, mengintegrasikan analisis sentimen berbasis RoBERTa, klasifikasi multilabel menggunakan Classifier Chains, serta evaluasi kualitas layanan dengan SERVQUAL dan IPA masih sangat terbatas, khususnya pada objek Wisata Bahari Lamongan [10]. Oleh karena itu, penelitian ini bertujuan menganalisis sentimen ulasan Google Maps, mengklasifikasikan ulasan ke dalam dimensi SERVQUAL menggunakan algoritma Classifier Chains, serta menghasilkan rekomendasi strategis bagi pengelola Wisata Bahari Lamongan berdasarkan hasil analisis Importance Performance Analysis.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode Knowledge Discovery in Database (KDD) untuk menganalisis kualitas layanan Wisata Bahari Lamongan (WBL) berdasarkan ulasan pengguna pada Google Maps. Pendekatan KDD dipilih karena menyediakan tahapan sistematis dalam mengolah data teks tidak terstruktur menjadi informasi yang dapat dimanfaatkan sebagai dasar pengambilan keputusan. Tahapan penelitian meliputi data selection, data preprocessing, data transformation, data mining, dan evaluation. Seluruh proses dilakukan menggunakan bahasa pemrograman Python dengan memanfaatkan beberapa pustaka machine learning dan natural language processing.



Gambar 1. Alur Metode Penelitian

3.1 Studi Literatur

Tahap studi literatur dilakukan untuk mengumpulkan teori, konsep, dan penelitian terdahulu yang relevan sebagai landasan penelitian. Literatur yang dikaji meliputi model SERVQUAL, analisis sentimen dan text mining, algoritma Classifier Chains untuk klasifikasi multilabel, serta penelitian terkait evaluasi kualitas layanan berbasis ulasan daring. Kajian ini menjadi dasar dalam membangun model analisis dan menginterpretasikan hasil prediksi berdasarkan dimensi SERVQUAL.

3.2 Data Selection

Tahap data selection bertujuan memperoleh data ulasan yang relevan dengan objek penelitian. Data dikumpulkan dari Google Maps pada lokasi Wisata Bahari Lamongan menggunakan layanan Google Maps Scraper melalui platform Apify. Teknik ini memungkinkan proses pengambilan data dilakukan secara otomatis beserta atribut pendukung seperti nama pengguna, tanggal ulasan, rating, dan isi ulasan. Seluruh ulasan yang tersedia pada halaman Google Maps WBL diambil tanpa pembatasan rentang waktu sehingga diperoleh sebanyak 2.493 ulasan sebagai dataset penelitian. Dataset tersebut menjadi masukan utama pada tahapan berikutnya.

3.3 Data Preprocessing

Tahap preprocessing dilakukan untuk membersihkan dan menstandarkan data teks agar siap diproses oleh model machine learning. Tahapan yang dilakukan terdiri atas:

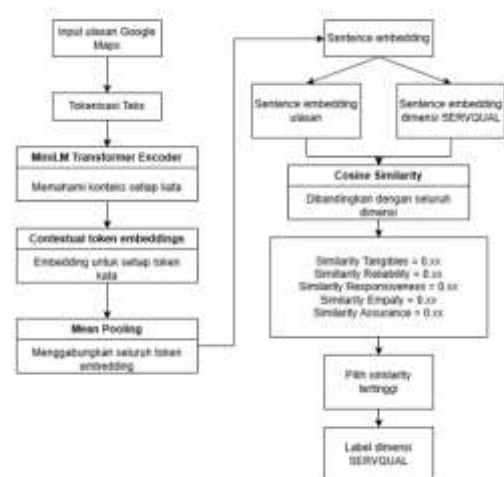
- Tahap Cleaning Adalah tahap menghapus karakter yang tidak relevan seperti tanda baca, angka, URL, emoji, serta karakter khusus lainnya. Setelah itu dilakukan *case folding* dengan mengubah seluruh huruf menjadi huruf kecil (*lowercase*) agar representasi setiap kata menjadi konsisten.
- Tahap tokenizing adalah tahap untuk memisahkan setiap kalimat menjadi kumpulan token atau kata tunggal sehingga memudahkan proses analisis teks pada tahapan berikutnya.
- Tahap Normalisasi Adalah tahap yang dilakukan untuk mengubah kata-kata tidak baku, singkatan,

maupun bahasa informal menjadi bentuk baku sesuai Kamus Besar Bahasa Indonesia (KBBI). Proses ini bertujuan mengurangi variasi penulisan kata yang memiliki makna sama.

- Tahap stopword Removal Adalah tahap yang dilakukan untuk menghilangkan kata-kata umum yang tidak memiliki kontribusi signifikan terhadap proses klasifikasi, seperti "dan", "yang", "di", serta kata penghubung lainnya sehingga hanya kata-kata penting yang dipertahankan.
- Tahap Stemming Adalah Tahap terakhir preprocessing adalah stemming, yaitu mengubah kata berimbuhan menjadi bentuk kata dasar menggunakan algoritma stemming bahasa Indonesia. Hasil proses ini menghasilkan representasi teks yang lebih sederhana sehingga dapat meningkatkan kualitas ekstraksi fitur.

3.4 Data Labelling

Pada tahap data transformation, setiap ulasan terlebih dahulu dianalisis menggunakan model *w11wo/indonesian-roberta-base-sentiment-classifier* untuk memperoleh label sentimen positif, negatif, atau netral. Selanjutnya dilakukan pemetaan dimensi SERVQUAL menggunakan pendekatan semantic similarity berbasis Sentence Transformer. Proses ini membandingkan kemiripan semantik antara isi ulasan dengan deskripsi setiap dimensi SERVQUAL sehingga satu ulasan dapat memperoleh lebih dari satu label. Hasil pemetaan kemudian divalidasi dan dikoreksi secara manual untuk memastikan kesesuaian label terhadap konteks ulasan.



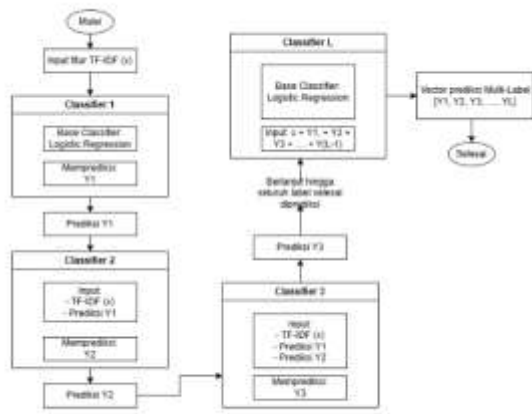
Gambar 2. Alur Kerja Model paraphrase-multilingual-MiniLM-L12-v2

Setelah proses validasi selesai, dataset dipisahkan menjadi data yang berhasil memperoleh label SERVQUAL dan data uncategorized. Data yang memiliki minimal satu label SERVQUAL digunakan sebagai data pelatihan model, sedangkan data uncategorized digunakan sebagai data yang akan diprediksi pada tahap klasifikasi multilabel.



3.5 Data Mining

Tahap data mining dilakukan menggunakan algoritma Classifier Chains untuk mengklasifikasikan ulasan ke dalam lima dimensi SERVQUAL, yaitu tangibles, reliability, responsiveness, assurance, dan empathy.

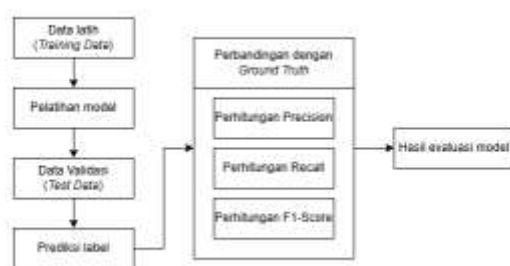


Gambar 3. Alur Pendekatan Classifier Chains

Berbeda dengan metode klasifikasi tunggal, pendekatan ini memungkinkan satu ulasan memiliki lebih dari satu label sehingga lebih sesuai dengan karakteristik data ulasan yang dapat membahas beberapa aspek layanan secara bersamaan.

Sebelum proses pelatihan dilakukan, dataset dibagi menjadi data latih dan data uji menggunakan Multilabel Stratified Shuffle Split dengan proporsi 80:20 agar distribusi setiap label tetap terjaga pada kedua kelompok data. Selanjutnya dilakukan pembangunan model menggunakan beberapa base classifier, yaitu Logistic Regression, Linear Support Vector Machine (Linear SVM), Naïve Bayes, dan Random Forest. Hasil pengujian masing-masing model kemudian dibandingkan untuk menentukan model dengan performa terbaik.

3.6 Evaluasi



Gambar 4. Alur Evaluasi Data

Tahap evaluasi bertujuan mengukur kemampuan model dalam melakukan klasifikasi multilabel. Evaluasi dilakukan menggunakan metrik Subset Accuracy, Hamming Loss, Precision, Recall, dan F1-Score dengan pendekatan micro average, macro average, dan weighted average.

Model dengan hasil evaluasi terbaik selanjutnya digunakan untuk mengklasifikasikan seluruh ulasan yang belum memiliki label SERVQUAL. Hasil klasifikasi tersebut kemudian menjadi dasar dalam analisis Importance Performance Analysis (IPA) untuk mengevaluasi kualitas layanan Wisata Bahari Lamongan. Pada penelitian ini, nilai performance diperoleh dari rata-rata skor sentimen pada setiap dimensi SERVQUAL yang merepresentasikan tingkat kepuasan pengunjung terhadap layanan yang diterima. Sementara itu, nilai importance dihitung berdasarkan proporsi kemunculan masing-masing dimensi SERVQUAL pada seluruh ulasan sehingga menggambarkan tingkat perhatian atau kepentingan pengunjung terhadap aspek layanan tersebut. Selanjutnya, nilai importance dan performance dipetakan ke dalam diagram kartesius IPA yang terdiri atas empat kuadran. Sehingga dapat diidentifikasi atribut layanan yang perlu diprioritaskan untuk diperbaiki maupun dipertahankan.

Selain analisis IPA, penelitian ini juga menerapkan Latent Dirichlet Allocation (LDA) untuk mengidentifikasi topik dominan pada setiap dimensi SERVQUAL. Proses LDA dilakukan dengan mengelompokkan ulasan berdasarkan dimensi SERVQUAL dan kategori sentimennya, kemudian membangun model topik menggunakan representasi TF-IDF hasil preprocessing. Setiap model menghasilkan sekumpulan kata kunci dengan bobot tertinggi yang merepresentasikan topik dominan pada masing-masing kelompok ulasan. Topik tersebut digunakan untuk menginterpretasikan faktor-faktor yang memengaruhi persepsi positif maupun negatif pengunjung sehingga rekomendasi yang dihasilkan tidak hanya menunjukkan dimensi layanan yang perlu ditingkatkan, tetapi juga menjelaskan aspek spesifik yang menjadi perhatian utama pengunjung.

HASIL DAN PEMBAHASAN

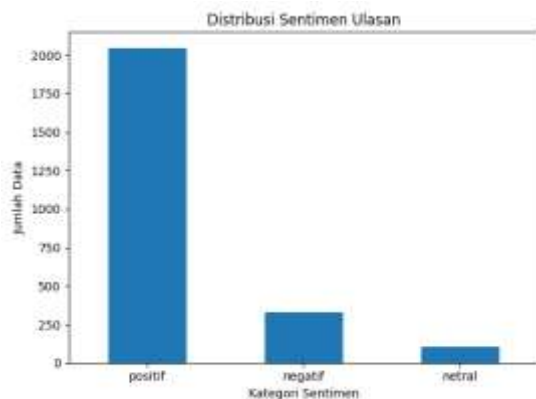
Pada penelitian ini dilakukan analisis kualitas layanan Wisata Bahari Lamongan berdasarkan ulasan Google Maps menggunakan algoritma Classifier Chains. Dataset diperoleh melalui proses web scraping, kemudian diproses melalui tahapan preprocessing, analisis sentimen, pemetaan dimensi SERVQUAL menggunakan semantic similarity, klasifikasi multilabel, serta analisis Importance Performance Analysis (IPA) dan Latent Dirichlet Allocation (LDA) untuk mengevaluasi kualitas layanan dan mengidentifikasi prioritas perbaikannya.

4.1 Hasil Pelabelan Sentimen

Analisis sentimen dilakukan menggunakan model `W11wo/indonesian-roberta-base-sentiment-classifier` yang mengelompokkan ulasan ke dalam tiga kategori, yaitu positif, negatif, dan netral. Hasil klasifikasi menunjukkan bahwa sebagian besar ulasan memiliki sentimen positif, sedangkan sentimen negatif dan netral memiliki proporsi yang lebih kecil. Dominasi sentimen positif menunjukkan bahwa secara umum pengunjung memberikan penilaian yang baik terhadap kualitas layanan Wisata Bahari Lamongan.

Meskipun demikian, ulasan negatif tetap menjadi perhatian penting karena mengandung informasi mengenai

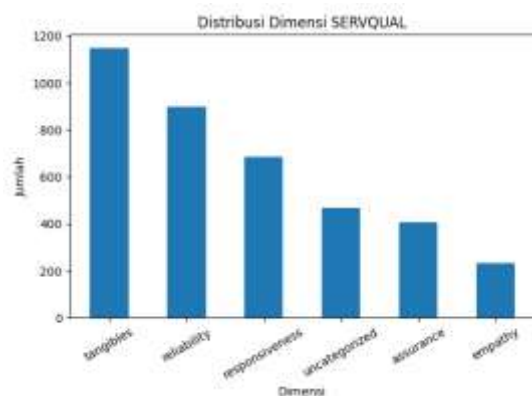
aspek layanan yang masih perlu diperbaiki. Oleh karena itu, hasil analisis sentimen selanjutnya digunakan sebagai dasar dalam proses pemetaan dimensi SERVQUAL dan analisis kualitas layanan.



Gambar 5. Grafik Distribusi Sentimen

4.2 Hasil Pemetaan Dimensi SERVQUAL

Pemetaan dimensi SERVQUAL dilakukan menggunakan model Sentence Transformer paraphrase-multilingual-MiniLM-L12-v2 dengan pendekatan semantic similarity. Setiap ulasan dibandingkan dengan deskripsi lima dimensi SERVQUAL, yaitu Tangibles, Reliability, Responsiveness, Assurance, dan Empathy, sehingga satu ulasan dapat memperoleh lebih dari satu label sesuai konteks pembahasannya.



Gambar 6. Grafik Distribusi Dimensi SERVQUAL

Hasil pemetaan menunjukkan bahwa dimensi Tangibles dan Reliability merupakan dimensi yang paling sering muncul pada ulasan pengunjung. Hal ini mengindikasikan bahwa pengunjung lebih banyak memberikan penilaian terhadap fasilitas fisik, kebersihan, kenyamanan lingkungan, serta keandalan layanan seperti sistem tiket dan operasional wisata. Sebaliknya, dimensi Assurance dan Empathy memiliki jumlah label yang relatif lebih sedikit karena aspek tersebut tidak selalu disebutkan secara eksplisit dalam ulasan.

Untuk meningkatkan kualitas pelabelan, hasil semantic similarity kemudian divalidasi melalui proses koreksi manual (human-in-the-loop). Proses ini menghasilkan perubahan sebanyak 797 label atau sekitar 22,09% dari total label awal sehingga menghasilkan dataset

yang lebih representatif sebagai data pelatihan model klasifikasi multilabel.

4.3 Hasil Klasifikasi Multilabel Menggunakan Classifier Chains

Dataset yang telah memiliki label SERVQUAL selanjutnya digunakan untuk membangun model klasifikasi multilabel menggunakan algoritma Classifier Chains. Dataset dibagi menjadi data latih dan data uji menggunakan Multilabel Stratified Shuffle Split dengan perbandingan 80:20 agar distribusi label tetap seimbang. Representasi fitur menggunakan metode TF-IDF, sedangkan proses klasifikasi membandingkan empat base classifier, yaitu Logistic Regression, Linear Support Vector Machine (Linear SVM), Naïve Bayes, dan Random Forest.

Tabel 1. Perbandingan Model

Model	Subset Accuracy	Micro F1-Score	Macro F1-Score	Weighted F1-Score
Classifier Chains + Logistic Regression	0.3902	0.6171	0.4774	0.6304
Classifier Chains + Linear SVM	0.4119	0.6508	0.4515	0.6219
Classifier Chains + Random Forest	0.4336	0.6516	0.4435	0.6195
Classifier Chains + Naïve Bayes	0.3767	0.6078	0.2679	0.5058

Hasil pengujian menunjukkan bahwa kombinasi Classifier Chains–Logistic Regression memberikan performa terbaik dibandingkan model lainnya. Hal ini menunjukkan bahwa pendekatan Classifier Chains mampu memanfaatkan hubungan antar dimensi SERVQUAL sehingga menghasilkan prediksi multilabel yang lebih akurat dibandingkan pendekatan yang mengasumsikan setiap label bersifat independen.

4.4 Evaluasi Model Klasifikasi

Pemetaan Evaluasi dilakukan menggunakan metrik Subset Accuracy, Hamming Loss, Precision, Recall, dan F1-Score. Berdasarkan hasil pengujian, kombinasi Classifier Chains dengan Logistic Regression memperoleh nilai Macro F1-Score sebesar 0,4774 dan Weighted F1-Score sebesar 0,6304, sehingga dipilih sebagai model terbaik pada penelitian ini. Nilai tersebut menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengklasifikasikan ulasan ke dalam lebih dari satu dimensi SERVQUAL.

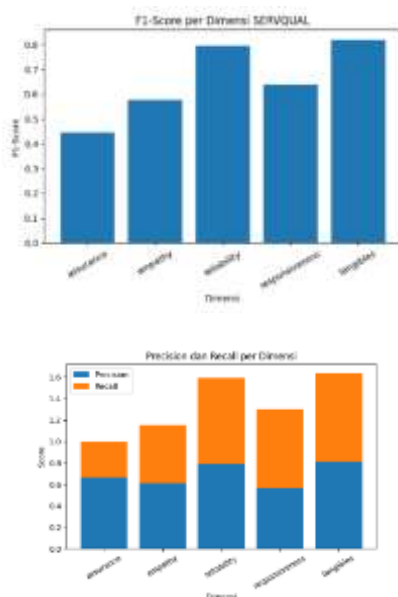


Tabel 2. Hasil Evaluasi Model

Dimensi	Precision	Recall	F1-Score	Support
Assurance	0.67	0.33	0.44	36
Empathy	0.61	0.54	0.58	35
Reliability	0.79	0.80	0.80	206
Responsiveness	0.57	0.74	0.64	53
Tangibles	0.81	0.82	0.82	233

Rata-Rata	Precision	Recall	F1-Score
Micro Avg	0.76	0.76	0.76
Macro Avg	0.69	0.65	0.66
Weighted Avg	0.76	0.76	0.75
Samples avg	0.79	0.80	0.77

Analisis terhadap setiap dimensi menunjukkan bahwa Tangibles dan Reliability memiliki performa klasifikasi yang lebih tinggi dibandingkan dimensi lainnya. Hal ini dipengaruhi oleh jumlah data yang lebih banyak dan karakteristik ulasan yang lebih konsisten. Sebaliknya, dimensi Assurance dan Empathy memiliki performa yang lebih rendah karena jumlah data yang relatif sedikit serta konteks pembahasan yang lebih beragam.



Gambar 7. Grafik Evaluasi Precision, Recall, dan F1-Score

Kurva Receiver Operating Characteristic (ROC) digunakan untuk mengevaluasi kemampuan model dalam membedakan kelas positif dan negatif pada setiap dimensi SERVQUAL. Kurva ROC dibentuk menggunakan nilai probabilitas hasil prediksi (predict_proba) dari model Classifier Chains dengan base classifier Logistic Regression terhadap data uji. Nilai probabilitas tersebut dibandingkan dengan label sebenarnya (ground truth) untuk menghitung True Positive Rate (TPR) dan False Positive Rate (FPR) pada berbagai nilai threshold. Selanjutnya dihitung nilai

Area Under the Curve (AUC) sebagai ukuran kemampuan model dalam membedakan kedua kelas. Nilai AUC berada pada rentang 0–1, di mana nilai yang semakin mendekati 1 menunjukkan kemampuan klasifikasi yang semakin baik. Berikut Adalah nilai AUC pada tiap dimensi.

Tabel 3. Interpretasi grafik roc seluruh dimensi

Dimensi	AUC	Interpretasi Singkat
Empathy	0.85	Model memiliki kemampuan yang baik dan cukup stabil dalam membedakan ulasan yang termasuk dimensi empathy dan non-empathy.
Assurance	0.85	Model mampu mengenali aspek jaminan, keamanan, dan kepercayaan layanan dengan performa yang baik.
Responsive ness	0.90	Model memiliki kemampuan yang sangat baik dalam mengidentifikasi ulasan terkait kecepatan dan ketanggapan layanan.
Reliability	0.84	Model menunjukkan kemampuan yang baik dalam mengenali aspek konsistensi dan keandalan layanan wisata.
Tangibles	0.84	Model cukup baik dalam membedakan ulasan yang berkaitan dengan fasilitas fisik dan kondisi lingkungan wisata.

4.5 Perhitungan Analisis Importance Performance Analysis (IPA)

Hasil klasifikasi SERVQUAL kemudian dimanfaatkan untuk menghitung nilai performance dan importance pada setiap dimensi layanan. Nilai tersebut dipetakan ke dalam diagram kartesius IPA sehingga diperoleh prioritas peningkatan kualitas layanan berdasarkan persepsi pengunjung.

1. Perhitungan Nilai Performance dilakukan dengan mengubah setiap hasil analisis sentimen ke dalam skor numerik, yaitu positif = 3, netral = 2, dan negatif = 1. Skor tersebut digunakan untuk merepresentasikan tingkat kepuasan pengunjung terhadap layanan.

Selanjutnya, setiap ulasan direpresentasikan dalam bentuk biner untuk masing-masing dimensi SERVQUAL. Nilai 1 diberikan apabila suatu dimensi muncul pada ulasan, sedangkan 0 diberikan apabila dimensi tersebut tidak muncul. Apabila suatu dimensi terdapat pada ulasan, maka skor dimensi dihitung menggunakan nilai sentimen yang bersesuaian. Sebaliknya, apabila dimensi tidak muncul, diberikan nilai default sebesar 2 yang merepresentasikan kondisi netral. Dengan demikian, setiap ulasan menghasilkan



nilai *T-score*, *A-score*, *R-score*, *E-score*, dan *Rp-score* untuk dimensi *Tangibles*, *Assurance*, *Reliability*, *Empathy*, dan *Responsiveness*. Nilai *performance* untuk setiap dimensi dihitung sebagai rata-rata skor SERVQUAL menggunakan Persamaan :

$$Performance_{(k,d)} = \frac{\sum Score_{(k,d)}}{n_k}$$

Dengan:

$Score_{(k,d)}$ = skor SERVQUAL dimensi *d*
pada data *k*

n_k = jumlah ulasan pada data *k*

Tabel 4. Perhitungan performance

Dimensi	Average T-Score
Tangibles	2,487548423
Assurance	2,03707803
Reliability	2,441062535
Empathy	2,060874377
Responsiveness	2,050913116

- Perhitungan Nilai Importance dihitung berdasarkan frekuensi kemunculan setiap dimensi SERVQUAL pada seluruh dataset. Mula-mula dihitung jumlah kemunculan masing-masing dimensi, kemudian dibagi dengan jumlah total ulasan. Selanjutnya hasil normalisasi dikalikan dengan jumlah kategori sentimen ($J = 3$) agar berada pada skala yang sama dengan nilai performance. Perhitungan importance ditunjukkan pada Persamaan:

$$Importance_{k,d} = \frac{S_{k,d}}{N}$$

Dengan:

$S_{k,d}$ = jumlah keseluruhan dimensi
SERVQUAL d pada data *k*

N = Total seluruh
ulasan/positif/negatif/netral

J = Jumlah kategori sentiment

Tabel 5. Perhitungan importance

Dimensi	Proses Perkalian	Hasil
Tangibles	$0,645268 \times 3$	1,935805202
Assurance	$0,089651 \times 3$	0,268954068
Reliability	$0,568899 \times 3$	1,706696182
Empathy	$0,094632 \times 3$	0,28389596
Responsiveness	$0,15606 \times 3$	0,468179303

Pada penelitian ini, analisis dilakukan dalam empat skenario, yaitu IPA Global, IPA Sentimen Positif, IPA Sentimen Negatif, dan IPA Sentimen Netral. IPA Global

menggunakan seluruh ulasan, sedangkan analisis berdasarkan sentimen dilakukan pada masing-masing kelompok sentimen untuk mengidentifikasi perbedaan prioritas layanan.

Pada setiap skenario, nilai performance dan importance dihitung menggunakan prosedur yang sama, kemudian dipetakan ke dalam diagram kartesius IPA untuk menentukan prioritas peningkatan kualitas layanan.

4.6 Hasil Analisis IPA

1. Hasil Analisis IPA Global

a. Performance

Tabel 6. Perhitungan performance global

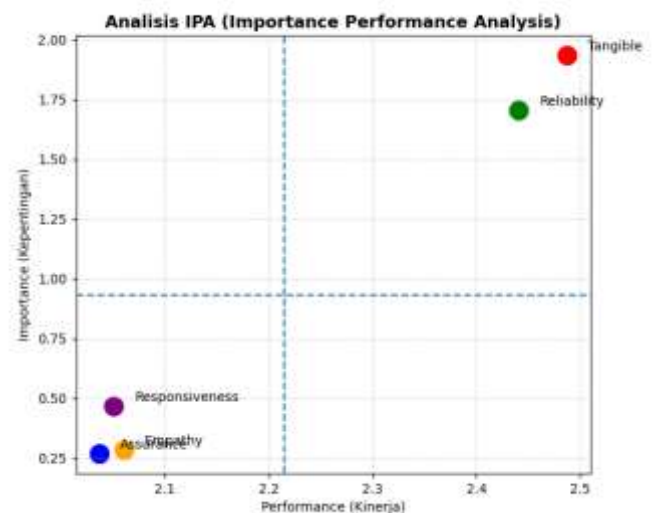
Dimensi	Jumlah
Tangibles	2,487548423
Assurance	2,03707803
Reliability	2,441062535
Empathy	2,060874377
Responsiveness	2,050913116

b. Importance

Tabel 7. Perhitungan importance global

Dimensi	Jumlah
Tangibles	1,935805202
Assurance	0,268954068
Reliability	1,706696182
Empathy	0,28389596
Responsiveness	0,468179303

c. Hasil IPA Global



Gambar 8. Hasil IPA Global

Berdasarkan pemetaan pada Gambar X, dimensi Tangibles dan Reliability berada pada Kuadran II yang menunjukkan bahwa kedua dimensi memiliki tingkat kepentingan dan kinerja yang tinggi sehingga perlu dipertahankan. Sebaliknya, dimensi Assurance, Empathy, dan Responsiveness berada pada Kuadran III yang



mengindikasikan bahwa ketiga dimensi memiliki tingkat kepentingan dan kinerja yang relatif rendah dibandingkan dimensi lainnya.

tingkat kepentingan serta kinerja yang relatif lebih rendah dibandingkan dimensi lainnya, sehingga tidak menjadi fokus utama dalam ulasan positif pengunjung.

2. Hasil Analisis IPA Positif
a. Performance

Tabel 8. Perhitungan performance positif

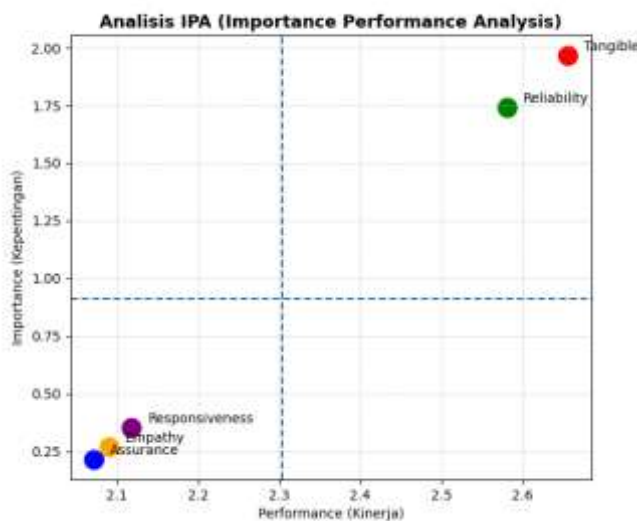
Dimensi	Jumlah
Tangibles	2,655684755
Assurance	2,071705426
Reliability	2,580749354
Empathy	2,090439276
Responsiveness	2,118217054

b. Importance

Tabel 9. Perhitungan importance positif

Dimensi	Jumlah
Tangibles	1,967054264
Assurance	0,215116279
Reliability	1,742248062
Empathy	0,271317829
Responsiveness	0,354651163

c. Hasil IPA Positif



Gambar 9. Hasil IPA Positif

Berdasarkan hasil analisis IPA pada data sentimen positif, dimensi Tangibles dan Reliability berada pada Kuadran II, yang menunjukkan bahwa kedua dimensi memiliki tingkat kepentingan dan kinerja yang tinggi. Hal ini mengindikasikan bahwa fasilitas fisik, kebersihan, kondisi lingkungan, serta keandalan layanan telah memenuhi harapan pengunjung sehingga kualitasnya perlu dipertahankan.

Sementara itu, dimensi Assurance, Empathy, dan Responsiveness berada pada Kuadran III. Posisi ini menunjukkan bahwa aspek jaminan layanan, perhatian terhadap pengunjung, dan ketanggapan layanan memiliki

3. Hasil Analisis IPA Negatif
a. Performance

Tabel 10. Perhitungan performance Negatif

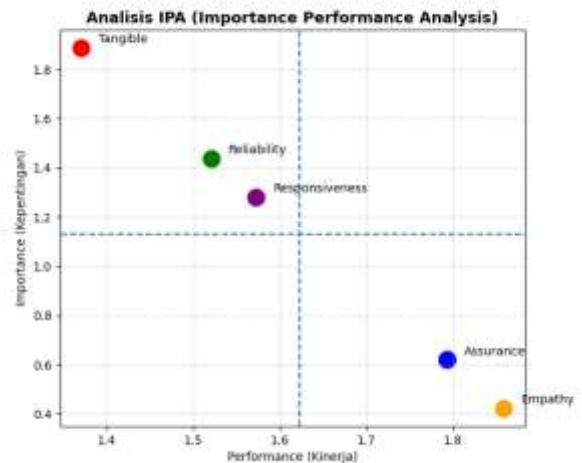
Dimensi	Jumlah
Tangibles	1,370892019
Assurance	1,79342723
Reliability	1,521126761
Empathy	1,85915493
Responsiveness	1,572769953

b. Importance

Tabel 11. Perhitungan Importance Negatif

Dimensi	Jumlah
Tangibles	1,887323944
Assurance	0,61971831
Reliability	1,436619718
Empathy	0,422535211
Responsiveness	1,281690141

c. Hasil IPA Negatif



Gambar 10. Hasil IPA Negatif

Berdasarkan hasil analisis IPA pada data sentimen negatif, dimensi Tangibles, Reliability, dan Responsiveness berada pada Kuadran I. Posisi ini menunjukkan bahwa ketiga dimensi memiliki tingkat kepentingan yang tinggi, namun performanya masih belum memenuhi harapan pengunjung sehingga menjadi prioritas utama untuk diperbaiki. Perbaikan perlu difokuskan pada fasilitas fisik dan kebersihan, keandalan layanan, serta kecepatan dan ketanggapan dalam memberikan pelayanan.

Sementara itu, dimensi Assurance dan Empathy berada pada Kuadran IV, yang menunjukkan bahwa kedua dimensi memiliki tingkat kinerja yang relatif baik dibandingkan tingkat kepentingannya. Hal ini mengindikasikan bahwa aspek jaminan layanan dan



perhatian kepada pengunjung telah berjalan dengan baik, meskipun tidak menjadi fokus utama dalam ulasan negatif pengunjung.

4. Hasil Analisis IPA Netral
a. Performance

Tabel 12. Perhitungan performance netral

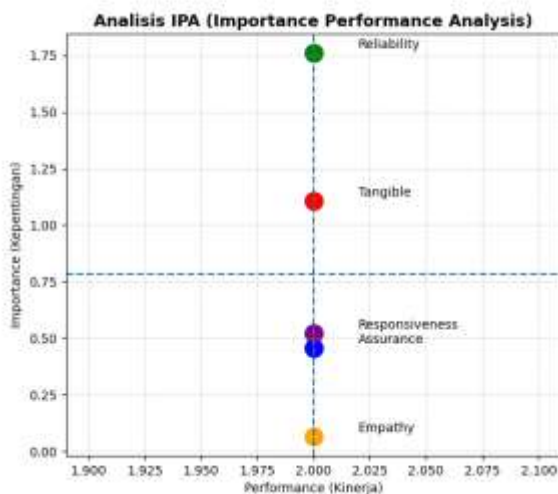
Dimensi	Jumlah
Tangibles	2
Assurance	2
Reliability	2
Empathy	2
Responsiveness	2

b. Importance

Tabel 13. Perhitungan Importance netral

Dimensi	Jumlah
Tangibles	1,108695652
Assurance	0,456521739
Reliability	1,760869565
Empathy	0,065217391
Responsiveness	0,52173913

c. Hasil IPA Netral



Gambar 11. Hasil IPA Netral

Berdasarkan hasil analisis IPA pada data sentimen netral, dimensi Tangibles dan Reliability berada pada Kuadran II, yang menunjukkan bahwa kedua dimensi memiliki tingkat kepentingan yang tinggi dengan kinerja yang telah memenuhi harapan pengunjung. Hal ini mengindikasikan bahwa fasilitas fisik, kondisi lingkungan, serta keandalan layanan perlu dipertahankan agar kualitas layanan tetap konsisten.

Sementara itu, dimensi Assurance, Empathy, dan Responsiveness berada pada Kuadran IV. Posisi ini menunjukkan bahwa ketiga dimensi memiliki tingkat kinerja yang relatif baik dibandingkan tingkat kepentingannya, sehingga aspek jaminan layanan, perhatian

kepada pengunjung, dan ketanggapan layanan telah berjalan dengan baik meskipun tidak menjadi fokus utama dalam ulasan sentimen netral.

4.7 Analisis Gap

Analisis gap dilakukan dengan menghitung selisih antara nilai performance dan importance pada setiap dimensi SERVQUAL. Nilai gap digunakan untuk menggambarkan kesesuaian antara kinerja layanan dan tingkat kepentingan yang dirasakan pengunjung. Nilai gap positif menunjukkan bahwa kinerja layanan telah memenuhi atau melebihi tingkat kepentingan, sedangkan nilai gap negatif menunjukkan bahwa kinerja layanan masih berada di bawah harapan pengunjung sehingga memerlukan perhatian dan perbaikan.

1. Analisis Gap Data Global

Tabel 14. Analisis GAP Data Global

Dimensi	Performance	Importance	Gap (P-I)
Tangibles	2,487548423	1,935805202	0,5517
Assurance	2,03707803	0,268954068	1,7681
Reliability	2,441062535	1,706696182	0,7344
Empathy	2,060874377	0,28389596	1,7770
Responsiveness	2,050913116	0,468179303	1,5827

2. Analisis Gap Data Positif

Tabel 15. Analisis GAP Data Positif

Dimensi	Performance	Importance	Gap (P-I)
Tangibles	2,655684755	1,967054264	0,6886
Assurance	2,071705426	0,215116279	1,8566
Reliability	2,580749354	1,742248062	0,8385
Empathy	2,090439276	0,271317829	1,8191
Responsiveness	2,118217054	0,354651163	1,7635

3. Analisis Gap Data Negatif

Tabel 16. Analisis GAP Data Negatif

Dimensi	Performance	importance	Gap (P-I)
Tangibles	1,370892019	1,887323944	-0,5164
Assurance	1,79342723	0,61971831	1,1737
Reliability	1,521126761	1,436619718	0,0845
Empathy	1,85915493	0,422535211	1,4366
Responsiveness	1,572769953	1,281690141	0,2911

4. Analisis Gap Data Netral

Tabel 18. Analisis GAP Data Netral

Dimensi	Performance	Importance	Gap (P-I)
Tangibles	2	1,108695652	0,8913
Assurance	2	0,456521739	1,5435
Reliability	2	1,760869565	0,2391



Empathy	2	0,065217391	1,9348
Responsiveness	2	0,52173913	1,4783

4.8 Hasil Analisis Topik LDA

Analisis Latent Dirichlet Allocation (LDA) dilakukan untuk mendukung interpretasi hasil Importance Performance Analysis (IPA) pada empat skenario data, yaitu global, sentimen positif, sentimen netral, dan sentimen negatif. Pada setiap skenario, data dipisahkan berdasarkan dimensi SERVQUAL, kemudian direpresentasikan menggunakan TF-IDF dan dimodelkan menggunakan LDA dengan satu topik pada setiap dimensi. Hasil pemodelan berupa tujuh kata kunci dengan bobot tertinggi yang digunakan untuk merepresentasikan karakteristik setiap dimensi SERVQUAL serta mendukung interpretasi hasil analisis IPA.

1. Hasil Analisis Topik LDA Global

Tabel 19. Hasil Analisis Topik LDA Global

Dimensi	Keyword	Interpretasi Topik
Tangibles	tempat, bersih, luas, nyaman, lengkap, libur, pantai	Topik dominan menggambarkan kondisi fisik wisata seperti kebersihan, kelengkapan fasilitas, dan kenyamanan area wisata yang menjadi perhatian utama pengunjung.
Assurance	aman, jaga, nyaman, tiket, kucing,antang, air	Topik berkaitan dengan rasa aman, pengelolaan area, serta pengalaman pengunjung terkait keamanan dan kontrol lingkungan wisata.
Reliability	cocok, libur, senang, tiket, harga, lengkap, lumayan	Topik menggambarkan konsistensi layanan seperti kesesuaian harga, kemudahan tiket, serta kepuasan umum terhadap layanan yang diberikan.
Empathy	ramah, senang, layan, jaga,	Topik menunjukkan

	tugas, bersih, cocok	aspek perhatian dan keramahan petugas dalam melayani pengunjung serta pengalaman emosional pengguna.
Responsi	layan, tiket, jaga, bersih	Topik berkaitan dengan kecepatan dan ketanggapan layanan, termasuk respons petugas dalam membantu pengunjung dan menangani kebutuhan mereka.

2. Hasil LDA Positif

Tabel 20. Hasil Analisis Topik LDA Positif

Dimensi	Keyword	Interpretasi Topik
Tangibles	tempat, luas, bersih, nyaman, lengkap, libur, cocok	Fokus pada fasilitas, bersih, dan nyaman bagi pengunjung.
Assurance	aman, nyaman, tiket,antang, air, parkir, ramah	Berkaitan dengan rasa aman, parkir, dan kenyamanan pengelolaan wisata.
Reliability	cocok, libur, senang, tiket, harga, lengkap, lumayan	Menunjukkan kepuasan terhadap harga, tiket, dan layanan yang konsisten.
Empathy	ramah, senang, layan, bersih, tugas, luas, cocok	Menggambarkan keramahan dan pelayanan yang perhatian kepada pengunjung.
Responsiveness	laya, ramah, tiket, bersih, harga, masuk, antri	Berkaitan dengan kecepatan layanan dan pengelolaan antrian pengunjung.



3. Hasil LDA Negatif

Tabel 21. Hasil Analisis Topik LDA Negatif

Dimensi	Keyword	Interpretasi Topik
Tangibles	rawat, kucing, bersih, panas, sayang, kayak, baik	Menunjukkan kondisi fasilitas dan lingkungan yang kurang nyaman, terutama pada wahana/tempat kucing.
Assurance	jaga, kucing, beli, tiket, pakai, orang, unjung	Berkaitan dengan pengelolaan area, tiket, dan keamanan pengunjung.
Reliability	mahal, tiket, beli, makan, sayang, harga, sepi	Menggambarkan ketidaksesuaian harga, tiket mahal, dan kurangnya pengunjung.
Empathy	ramah, jaga, layan, unjung, makan, tugas, kucing	Berkaitan dengan pelayanan dan perhatian petugas kepada pengunjung.
Responsiveness	rawat, kucing, tiket, jaga, layan, baik, beli	Menunjukkan respons layanan dan penanganan fasilitas yang masih kurang optimal.

4. Hasil LDA Netral

Tabel 22. Hasil Analisis Topik LDA Netral

Dimensi	Keyword	Interpretasi Topik LDA
Tangibles	ruang, butuh	Topik membahas ruang dan kebutuhan fasilitas fisik.
Assurance	jaga, bawa	Berkaitan dengan pengawasan dan keamanan pengunjung.
Reliability	tiket, masuk, harga, terus, hiburan, ribu, paket	Menggambarkan sistem tiket, harga, dan paket layanan wisata.

Responsiveness cash, payment, qris Berkaitan dengan metode pembayaran dan kemudahan transaksi.

KESIMPULAN

Berdasarkan hasil penelitian, pendekatan Classifier Chains dengan menggunakan algoritma Logistic Regression mampu mengklasifikasikan ulasan Google Maps Wisata Bahari Lamongan ke dalam lima dimensi SERVQUAL dan menghasilkan informasi yang dapat digunakan untuk mengevaluasi kualitas layanan. Hasil analisis Importance Performance Analysis (IPA) menunjukkan bahwa pada data global, dimensi Tangibles dan Reliability berada pada Kuadran II sehingga perlu dipertahankan, sedangkan Assurance, Empathy, dan Responsiveness berada pada Kuadran III. Pada analisis berdasarkan sentimen, dimensi Tangibles, Reliability, dan Responsiveness menjadi prioritas utama perbaikan pada data sentimen negatif karena berada pada Kuadran I.

Integrasi hasil IPA dan Latent Dirichlet Allocation (LDA) menunjukkan bahwa persepsi pengunjung terhadap kualitas layanan Wisata Bahari Lamongan didominasi oleh aspek Tangibles dan Reliability, terutama terkait fasilitas fisik, kebersihan, kondisi lingkungan, harga, dan sistem tiket. Sementara itu, pada ulasan negatif ditemukan bahwa aspek fasilitas, keandalan layanan, dan ketanggapan petugas masih menjadi sumber utama keluhan pengunjung. Hasil penelitian ini diharapkan dapat menjadi dasar bagi pengelola Wisata Bahari Lamongan dalam menentukan prioritas peningkatan kualitas layanan sesuai dengan persepsi dan kebutuhan pengunjung.

DAFTAR PUSTAKA

- J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." [Online]. Available: <https://github.com/tensorflow/tensor2tensor>
- Y. O. Sihombing, N. V. Situmorang, B. K. Negara, and J. M. Sutoyo, "Jakarta Timur, 13640 b Badan Kepegawaian Negara," 1364.
- A. Parasuraman, V. A. Zeithaml, and L. L. Berry, "A Conceptual Model of Service Quality and Its Implications for Future Research," J. Mark., vol. 49, no. 4, p. 41, Autumn 1985, doi: 10.2307/1251430.
- M. F. Sanz, E. D. Ferrandis, and J. G. Ferrer, "Service quality scales and tourists with special needs: A systematic review," Sustainability (Switzerland), vol. 11, no. 14, Jul. 2019, doi: 10.3390/su11143844.
- J. Sosial, E. Dan, K. Pertanian, R. Nurani, W. Roessali, and T. Ekowati, "AGROSOCIONOMICS VISITORS SATISFACTION STUDY OF JOLLONG PATI AGRITOURISM IN TERMS OF SERVQUAL DIMENSIONS," 2022, [Online]. Available: <http://ejournal2.undip.ac.id/index.php/agrisocionomic>



- J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier Chains: A Review and Perspectives," Apr. 2020, doi: 10.1613/jair.1.12376.
- J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier Chains: A Review and Perspectives," 2021.
- N. Royani, C. E. Widodo, and B. Warsito, "Topic Modelling Latent Dirichlet Allocation untuk Klasifikasi Komentar pada Layanan Streaming Platform," JST (Jurnal Sains dan Teknologi), vol. 12, no. 3, Jan. 2024, doi: 10.23887/jstundiksha.v12i3.68492.
- Y. Palinggi, U. Kutai Kartanegara, J. Gunung Kombeng No, K. Kutai Kartanegara, and K. Pos, "ANALISIS KEPUASAN PELANGGAN DENGAN IMPORTANCE PERFORMANCE ANALYSIS (IPA)."
- N. Millenia and A. Kholik, "ANALISIS KUALITAS PELAYANAN DI WISATA REKREASI ATV ADVENTURE INDONESIA JAKARTA TIMUR," Journal Olahraga ReKat (Rekreasi Masyarakat, vol. 3, pp. 93–103.