

Analisis Spasial Kepuasan Layanan Puskesmas Kota Denpasar Menggunakan TF-IDF dan SVM

I Komang Dharmendra^{*1}, Anak Agung Istri Indryarthi Dewi², Anak Agung Ngurah Putra Riana Prasetya³

¹Fakultas Informatika dan Komputer, Institut Teknologi dan Bisnis STIKOM Bali

²⁻³Fakultas Farmasi dan Ilmu Kesehatan, Universitas Pendidikan Nasional

E-mail: dharmendra@stikom-bali.ac.id¹, g.ind1296@gmail.com²,
agungputra@uniknas.ac.id³

Abstrak. Layanan kesehatan primer melalui puskesmas memiliki peran strategis dalam sistem kesehatan perkotaan, sehingga evaluasi kepuasan masyarakat terhadap kualitas layanan menjadi aspek penting dalam perbaikan berkelanjutan. Penelitian ini bertujuan untuk menganalisis kepuasan layanan puskesmas di Kota Denpasar melalui integrasi analisis sentimen ulasan Google Maps menggunakan metode *Term Frequency–Inverse Document Frequency* dan algoritma *Support Vector Machine*, serta mengaitkannya dengan analisis spasial berbasis wilayah kecamatan. Data penelitian terdiri atas 2.681 ulasan pengguna dari 11 puskesmas yang tersebar di empat kecamatan. Hasil pengujian menunjukkan bahwa model TF-IDF *unigram* dengan *Linear Support Vector Classifier* menghasilkan kinerja terbaik dengan nilai akurasi 88,1% dan f1-score makro sebesar 0,8534. Analisis spasial mengungkap adanya variasi tingkat kepuasan layanan antar kecamatan, di mana Denpasar Utara menunjukkan tingkat kepuasan tertinggi, sedangkan Denpasar Selatan berada pada kategori sedang dan memerlukan prioritas perbaikan. Selain itu, analisis kata dominan menunjukkan bahwa kualitas pelayanan, sikap petugas, dan efisiensi waktu menjadi faktor utama yang memengaruhi persepsi kepuasan dan ketidakpuasan masyarakat. Penelitian ini menyimpulkan bahwa integrasi analisis sentimen berbasis *machine learning* dan analisis spasial mampu memberikan gambaran komprehensif mengenai kualitas layanan puskesmas dan dapat dimanfaatkan sebagai dasar pengambilan kebijakan peningkatan layanan kesehatan primer secara lebih terarah.

Kata kunci: analisis sentimen; puskesmas; TF-IDF; *Support Vector Machine*; analisis spasial

Abstract. Primary healthcare services delivered through community health centers play a strategic role in urban health systems, making the evaluation of public satisfaction with service quality a crucial aspect of continuous improvement. This study aims to analyze satisfaction with community health center services in Denpasar City by integrating sentiment analysis of Google Maps reviews using the Term Frequency–Inverse Document Frequency method and the Support Vector Machine algorithm, combined with spatial analysis at the sub-district level. The dataset consists of 2,681 user reviews collected from 11 community health centers across four sub-districts. The results indicate that the unigram TF-IDF model with a Linear Support Vector Classifier achieves the best performance, with an accuracy of 88.1% and a macro f1-score of 0.8534. Spatial analysis reveals variations in service satisfaction among sub-districts, with North Denpasar showing the highest satisfaction level, while South Denpasar falls into the moderate category and requires priority improvements. Furthermore, dominant word analysis indicates that service quality, staff attitude, and time efficiency are the main factors influencing both

satisfaction and dissatisfaction perceptions. This study concludes that integrating machine learning-based sentiment analysis with spatial analysis provides a comprehensive overview of community health center service quality and can support evidence-based policymaking for more targeted improvements in primary healthcare services.

Keywords: sentiment analysis; community health centers; TF-IDF; Support Vector Machine; spatial analysis.

1. Pendahuluan

Layanan kesehatan primer melalui puskesmas merupakan fondasi utama sistem kesehatan di Indonesia karena berperan strategis dalam upaya promotif dan preventif, sekaligus sebagai garda terdepan pelayanan kuratif bagi masyarakat [1]. Kualitas layanan puskesmas tidak hanya ditentukan oleh ketersediaan fasilitas dan tenaga kesehatan, tetapi juga oleh persepsi dan pengalaman pengguna layanan [2]. Persepsi masyarakat terhadap mutu layanan menjadi indikator penting dalam evaluasi kinerja fasilitas kesehatan, karena secara langsung mencerminkan tingkat kepuasan, kepercayaan, serta keberterimaan layanan yang diberikan [3]. Namun demikian, evaluasi kepuasan layanan puskesmas masih sering dilakukan secara konvensional melalui survei terbatas, sehingga kurang mampu menangkap dinamika opini publik secara luas dan berkelanjutan [4].

Perkembangan platform digital telah membuka peluang baru dalam pengukuran kepuasan layanan berbasis ulasan daring [5]. Ulasan pengguna pada platform seperti Google Maps menyediakan data teks dalam jumlah besar yang merepresentasikan pengalaman langsung masyarakat terhadap layanan kesehatan secara *real-time*. Analisis sentimen terhadap ulasan digital tersebut memungkinkan identifikasi kecenderungan opini positif dan negatif secara lebih objektif dan efisien dibandingkan metode survei tradisional. Seiring dengan meningkatnya volume data teks, pendekatan *machine learning* menjadi relevan untuk melakukan klasifikasi sentimen secara otomatis dan konsisten [6].

Berbagai penelitian menunjukkan bahwa kombinasi *Term Frequency–Inverse Document Frequency* (TF-IDF) [7] dan *Support Vector Machine* (SVM) [8] merupakan pendekatan yang efektif untuk analisis sentimen teks pendek, khususnya pada bahasa Indonesia. TF-IDF mampu merepresentasikan teks berdimensi tinggi secara informatif [9], sementara SVM, terutama dengan kernel linear, terbukti unggul dalam menangani data teks yang bersifat *sparse* [10]. Pendekatan ini telah diaplikasikan secara luas pada domain layanan publik dan media sosial dengan capaian akurasi dan *F1-score* yang relatif tinggi. Meskipun demikian, sebagian besar penelitian masih berfokus pada evaluasi performa model klasifikasi, tanpa mengaitkan hasil sentimen dengan dimensi spasial wilayah layanan.

Dalam konteks layanan kesehatan, analisis spasial umumnya difokuskan pada aspek aksesibilitas geografis [11], distribusi fasilitas [12], dan jangkauan pelayanan [13]. Kajian yang mengintegrasikan hasil analisis sentimen berbasis *machine learning* dengan pemetaan spasial kepuasan layanan masih sangat terbatas, khususnya pada level fasilitas kesehatan primer di wilayah perkotaan. Padahal, integrasi kedua pendekatan tersebut berpotensi memberikan gambaran yang lebih komprehensif mengenai disparitas kualitas layanan antar wilayah administratif, sekaligus menjadi dasar pengambilan kebijakan berbasis bukti.

Kota Denpasar sebagai ibu kota Provinsi Bali memiliki karakteristik perkotaan dengan kepadatan penduduk dan mobilitas masyarakat yang tinggi, serta sistem layanan kesehatan primer yang tersebar di empat kecamatan, yaitu Denpasar Utara, Denpasar Selatan, Denpasar Timur, dan Denpasar Barat. Secara keseluruhan, Kota Denpasar didukung oleh 11 puskesmas yang melayani kebutuhan kesehatan masyarakat perkotaan dengan karakteristik sosial dan beban layanan yang berbeda antar kecamatan [14]. Variasi jumlah puskesmas, intensitas kunjungan, serta kondisi sosial wilayah berpotensi menimbulkan perbedaan tingkat kepuasan masyarakat terhadap layanan yang diberikan. Namun hingga saat ini, pemetaan kepuasan layanan puskesmas di Kota Denpasar yang berbasis persepsi publik dan dianalisis secara spasial masih belum tersedia secara sistematis.

Berdasarkan celah penelitian tersebut, penelitian ini bertujuan untuk mengintegrasikan analisis sentimen ulasan Google Maps menggunakan pendekatan TF-IDF dan SVM dengan analisis spasial kepuasan layanan puskesmas di Kota Denpasar. Penelitian ini tidak hanya berfokus pada evaluasi performa model klasifikasi sentimen, tetapi juga pada pemetaan disparitas kepuasan layanan antar

kecamatan dan antar puskesmas. Dengan demikian, hasil penelitian diharapkan dapat memberikan kontribusi ilmiah dalam pengembangan pendekatan analisis sentimen berbasis spasial, sekaligus menghasilkan rekomendasi berbasis data bagi pemangku kebijakan dalam merumuskan prioritas peningkatan kualitas layanan kesehatan primer secara lebih terarah dan berkelanjutan.

2. Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksplanatori, yang mengintegrasikan analisis sentimen berbasis *machine learning* dan analisis spasial untuk mengevaluasi kepuasan layanan puskesmas di Kota Denpasar. Alur metodologi penelitian mencakup tahapan pengumpulan data, praproses teks, klasifikasi sentimen menggunakan TF-IDF dan SVM, evaluasi performa model, serta agregasi dan analisis spasial hasil klasifikasi pada tingkat puskesmas dan kecamatan.

2.1. Pengumpulan Data

Data penelitian diperoleh dari ulasan pengguna pada platform Google Maps terhadap seluruh puskesmas yang berada di Kota Denpasar. Secara administratif, wilayah penelitian mencakup empat kecamatan, yaitu Denpasar Utara, Denpasar Selatan, Denpasar Timur, dan Denpasar Barat, dengan total 11 puskesmas aktif. Proses pengambilan data dilakukan melalui *web scraping* menggunakan Google Places API dengan kueri nama resmi masing-masing puskesmas. Data yang dikumpulkan terdiri atas teks ulasan, rating bintang (skala 1–5), serta metadata lokasi yang meliputi nama puskesmas dan kecamatan. Dari proses ini diperoleh 2.681 ulasan unik yang merepresentasikan persepsi masyarakat terhadap layanan puskesmas di wilayah penelitian.

2.2. Pelabelan Sentimen

Pelabelan sentimen dilakukan secara otomatis menggunakan pendekatan berbasis rating bintang sebagai proxy kepuasan pengguna [15]. Ulasan dengan rating 4 dan 5 dikategorikan sebagai sentimen positif, sedangkan ulasan dengan rating 1 hingga 3 dikategorikan sebagai sentimen negatif. Skema pelabelan biner ini menghasilkan distribusi kelas sebesar 64% ulasan positif (1.719 ulasan) dan 36% ulasan negatif (962 ulasan). Pendekatan ini dipilih untuk menjaga konsistensi label pada skala data besar serta menghindari subjektivitas anotasi manual. Tabel 1 menunjukkan sebaran data sentimen.

Tabel 1. Sebaran Data Sentiment

| Kecamatan | Bintang | | | | | Sentimen | | Total |
|------------------|---------|----|----|----|-----|----------|---------|-------|
| | 1 | 2 | 3 | 4 | 5 | Negatif | Positif | |
| Denpasar Utara | 227 | 17 | 20 | 39 | 478 | 264 | 517 | 781 |
| Denpasar Timur | 147 | 16 | 11 | 32 | 247 | 174 | 279 | 453 |
| Denpasar Selatan | 251 | 40 | 41 | 85 | 595 | 332 | 680 | 1012 |
| Denpasar Barat | 148 | 18 | 26 | 35 | 208 | 192 | 243 | 435 |

2.3. Praproses

Sebelum dilakukan vektorisasi, teks ulasan melalui tahapan praproses untuk meningkatkan kualitas representasi fitur. Tahapan tersebut meliputi normalisasi huruf menjadi *lowercase*, penghapusan karakter non-alfabetik dan tanda baca, penghapusan *stopword* bahasa Indonesia, serta proses *stemming* menggunakan Sastrawi [16]. Praproses ini bertujuan untuk mengurangi *noise*, menyatukan variasi kata, dan meminimalkan redundansi fitur, sehingga model klasifikasi dapat mempelajari pola sentimen secara lebih efektif [17]. Contoh proses praproses ditunjukkan pada tabel 2, dimana terdapat kalimat sebelum diterapkan praproses, dan kalimat yang sudah mengalami praproses.

Tabel 2. Contoh Praproses Sentiment

| Proses | Kalimat |
|-------------------|--|
| Sebelum Praproses | Melayani pengobatan TB hingga selesai selama 6 bulan dan gratis dengan bpjs. Dipantau oleh tenaga ahli yang baik dan ramah. Terima kasih |

| Proses | Kalimat |
|-------------------|---|
| Setelah Praproses | layan obat tb selesai gratis bpjs pantau tenaga ahli ramah terima kasih |

2.4. Representasi fitur menggunakan TF-IDF

Teks ulasan yang telah dipraproses direpresentasikan menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)*. Metode ini memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam satu dokumen dan kelangkaannya pada keseluruhan korpus [18]. Pendekatan *TF-IDF* terdiri atas dua komponen utama, yaitu *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)*, yang kemudian dikombinasikan untuk menghasilkan bobot akhir suatu kata.

2.4.1. Term Frequency (TF)

Term Frequency (TF) mengukur seberapa sering suatu kata muncul dalam sebuah dokumen. Nilai *TF* mencerminkan tingkat kepentingan kata tersebut di dalam dokumen tertentu. Semakin sering sebuah kata muncul, maka semakin besar nilai *TF*-nya. Yang ditunjukkan pada persamaan (1).

$$TF(t, d) = f(t, d) / (\sum f(*, d)) \quad (1)$$

$TF(t, d)$ menyatakan frekuensi kemunculan kata t dalam dokumen d , yang dinormalisasi dengan jumlah seluruh kata dalam dokumen tersebut. Normalisasi ini bertujuan untuk menghindari bias akibat perbedaan panjang dokumen.

2.4.2. Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) digunakan untuk mengukur tingkat keunikan suatu kata dalam keseluruhan kumpulan dokumen. Kata yang muncul di banyak dokumen dianggap kurang informatif, sehingga bobotnya diperkecil. Sebaliknya, kata yang hanya muncul pada sedikit dokumen akan memiliki nilai *IDF* yang lebih tinggi.

$$IDF(t) = \log (N/df(t)) \quad (2)$$

Pada persamaan (2) menunjukkan $IDF(t)$ dihitung sebagai logaritma dari perbandingan antara jumlah total dokumen (N) dan jumlah dokumen yang mengandung kata t (*document frequency*). Penggunaan fungsi logaritma bertujuan untuk meredam pengaruh nilai ekstrem dan menjaga stabilitas bobot.

2.4.3. Term Frequency - Inverse Document Frequency (TF-IDF)

Bobot akhir *TF-IDF* diperoleh dengan mengalikan nilai *TF* dan *IDF*. Kombinasi ini menghasilkan representasi yang menekankan kata-kata yang penting dalam suatu dokumen, namun tidak bersifat umum di seluruh korpus.

$$TFIDF(t, d) = TF(t, d) \times IDF(t) \quad (3)$$

Nilai *TF-IDF* pada persamaan (3) akan tinggi apabila suatu kata sering muncul dalam dokumen tertentu, tetapi jarang muncul dalam dokumen lain. Dengan karakteristik tersebut, *TF-IDF* efektif dalam menangkap kata-kata yang bersifat diskriminatif untuk keperluan klasifikasi teks.

Dalam konteks analisis sentimen ulasan layanan Puskesmas, *TF-IDF* memungkinkan model untuk menonjolkan kata atau frasa yang merepresentasikan pengalaman pengguna, baik yang bersifat positif maupun negatif. Representasi ini sangat sesuai dikombinasikan dengan algoritma *Support Vector Machine*, karena menghasilkan vektor fitur berdimensi tinggi dan jarang (*sparse*), yang merupakan karakteristik umum data teks ulasan daring.

2.5. Pemodelan Klasifikasi Sentimen

Model klasifikasi sentimen dibangun menggunakan algoritma *Support Vector Machine (SVM)*. *Support Vector Machine (SVM)* merupakan salah satu algoritma pembelajaran mesin terawasi (*supervised learning*) yang digunakan untuk tugas klasifikasi maupun regresi [19]. Pada permasalahan klasifikasi, *SVM* bertujuan untuk menemukan sebuah hiperbidang (*hyperplane*) optimal yang dapat memisahkan data dari dua kelas secara maksimal. Konsep utama *SVM* adalah memaksimalkan jarak pemisah (*margin*) antara hiperbidang dan titik data terdekat dari masing-masing kelas, yang dikenal sebagai *support vectors* s[8].

Dua pendekatan utama digunakan, yaitu *Linear Support Vector Classifier (LinearSVC)* dan *SVM* dengan kernel radial basis function (*RBF*). Pemodelan dilakukan dalam bentuk *pipeline* yang mengintegrasikan tahap ekstraksi fitur *TF-IDF* dan klasifikasi, sehingga proses pelatihan dan evaluasi dapat berjalan secara terstruktur.

$$f(x) = \text{sign}(w \cdot x + b) \quad (4)$$

Persamaan (4) merupakan fungsi keputusan *SVM* untuk klasifikasi dua kelas. Vektor x merepresentasikan data masukan yang telah diubah ke dalam bentuk fitur numerik, sedangkan w dan b masing-masing menyatakan vektor bobot dan bias yang dipelajari selama proses pelatihan. Fungsi *sign* digunakan untuk menentukan kelas keluaran berdasarkan posisi data terhadap hiperbidang pemisah. Apabila nilai $f(x)$ bernilai positif, maka data diklasifikasikan ke kelas positif, sedangkan nilai negatif menunjukkan keanggotaan pada kelas negatif.

2.6. Pembagian data dan optimasi parameter

Data dibagi menjadi data latih dan data uji dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian, menggunakan teknik *stratified sampling* untuk menjaga proporsi kelas sentimen. Optimasi parameter model dilakukan menggunakan *Grid Search Cross Validation* dengan skema validasi silang sebanyak lima lipatan. Parameter yang dioptimasi meliputi nilai regularisasi pada *SVM* serta parameter kernel pada model nonlinier.

2.7. Evaluasi model

Kinerja model dievaluasi menggunakan metrik *f1-score* makro (*f1_macro*), yang dipilih untuk memberikan bobot yang seimbang pada setiap kelas sentimen. Selain itu, evaluasi juga didukung oleh *classification report* dan *confusion matrix* untuk memberikan gambaran rinci terkait performa model dalam mengklasifikasikan sentimen positif dan negatif.

$$F1_{macro} = 1/(K \sum_i 1 K^i) (2 \times P_i \times R_i) / (P_i + R_i) \quad (5)$$

Nilai *f1_macro* yang ditampilkan pada persamaan (5) diperoleh dengan menghitung rata-rata *f1-score* dari setiap kelas sentimen, di mana P_i dan R_i masing-masing menyatakan nilai *precision* dan *recall* pada kelas ke- i , serta K menunjukkan jumlah kelas. Pendekatan ini memberikan bobot yang sama pada setiap kelas, sehingga evaluasi kinerja model tidak dipengaruhi oleh ketidakseimbangan distribusi data sentimen.

$$CM = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \quad (6)$$

Persamaan *Confusion matrix* pada persamaan (6) merepresentasikan perbandingan antara hasil prediksi dan kelas aktual, di mana *TP* menunjukkan jumlah sentimen positif yang terklasifikasi dengan benar, *FP* menyatakan sentimen negatif yang salah diklasifikasikan sebagai positif, *FN* menunjukkan sentimen positif yang salah diklasifikasikan sebagai negatif, dan *TN* merepresentasikan sentimen negatif yang terklasifikasi dengan benar. Matriks ini menjadi dasar perhitungan metrik evaluasi seperti *precision*, *recall*, dan *f1-score*.

2.8. Analisis spasial kepuasan per kecamatan

Hasil klasifikasi sentimen pada data uji selanjutnya diagregasikan berdasarkan kecamatan. Analisis ini mencakup perhitungan proporsi sentimen positif dan negatif, jumlah ulasan, serta rata-rata rating pada masing-masing kecamatan. Agregasi spasial ini digunakan untuk mengidentifikasi perbedaan tingkat kepuasan layanan Puskesmas antar wilayah kecamatan di Kota Denpasar, serta sebagai dasar analisis lanjutan terkait pola kepuasan masyarakat secara geografis. Persamaan untuk menghitung Metrik spasial dapat dilihat pada persamaan (7).

$$P_{\text{positif}} = N_{\text{positif}} / N_{\text{total}} \quad (7)$$

Nilai P pada persamaan (8) menyatakan proporsi sentimen positif, yang dihitung sebagai perbandingan antara jumlah ulasan dengan sentimen positif dan total ulasan pada suatu wilayah.

$$\text{Sentiment Score} = P_{\text{positif}} - P_{\text{negatif}} \quad (8)$$

Sentiment Score merupakan ukuran agregat yang digunakan untuk menggambarkan kecenderungan umum sentimen masyarakat terhadap suatu objek atau wilayah. Nilai ini diperoleh dari selisih antara proporsi sentimen positif (P_{positif}) dan proporsi sentimen negatif (P_{negatif}). Apabila nilai *Sentiment Score* bernilai positif, maka sentimen positif lebih dominan dibandingkan sentimen negatif, yang mengindikasikan tingkat kepuasan masyarakat yang relatif baik. Sebaliknya, nilai yang mendekati nol atau bernilai negatif menunjukkan adanya keseimbangan atau dominasi sentimen negatif, sehingga dapat diinterpretasikan sebagai indikasi ketidakpuasan atau perlunya peningkatan kualitas layanan sesuai tabel 3.

Tabel 3. Interpretasi *Sentiment Score*

| Kategori | Proporsi Positif | Interpretasi |
|---------------|------------------|---------------------|
| Sangat Tinggi | $\geq 75\%$ | Best Practice |
| Tinggi | 65-74% | Monitoring |
| Sedang | 55-64% | Prioritas Perbaikan |
| Rendah | $< 55\%$ | Intervensi Mendesak |

Berdasarkan hasil perhitungan *sentiment score* pada masing-masing kecamatan, tingkat kepuasan layanan Puskesmas selanjutnya diklasifikasikan ke dalam beberapa kategori untuk memudahkan interpretasi dan penarikan implikasi kebijakan yang terlihat pada tabel 3. Klasifikasi tersebut mengacu pada proporsi sentimen positif. Pendekatan kategorisasi ini memungkinkan hasil analisis sentimen disajikan secara lebih sistematis dan aplikatif dalam mendukung evaluasi serta pengambilan keputusan dalam peningkatan pelayanan.

2.9. Klasifikasi Sentimen dengan *TF-IDF + SVM*

Pembagian data: 80% train (2.145 ulasan), 20% test (537 ulasan), stratified split (random_state=42) untuk mempertahankan proporsi kelas. Evaluasi menggunakan classification report, confusion matrix, dan F1-macro test set.

2.10. Analisis Spasial Kepuasan Layanan

Hasil klasifikasi sentimen dari model terbaik diagregasi secara hierarkis pada tingkat puskesmas dan kecamatan. Metrik spasial yang dihitung meliputi proporsi sentimen positif, proporsi sentimen negatif, *sentiment score*, serta rata-rata rating. Berdasarkan proporsi sentimen positif, tingkat kepuasan layanan diklasifikasikan ke dalam kategori sangat tinggi, tinggi, sedang, dan rendah. Analisis spasial dilakukan untuk mengidentifikasi disparitas kepuasan antar wilayah, serta untuk mendeteksi area prioritas perbaikan layanan dan wilayah dengan praktik terbaik yang dapat direplikasi.

3. Hasil dan Pembahasan

Bagian ini menyajikan hasil penelitian yang diperoleh dari proses klasifikasi sentimen ulasan layanan Puskesmas di Kota Denpasar serta analisis yang dilakukan berdasarkan hasil tersebut. Penyajian hasil mencakup evaluasi kinerja model klasifikasi sentimen, distribusi sentimen positif dan negatif, serta analisis kepuasan layanan yang dikaji berdasarkan wilayah kecamatan. Hasil penelitian disajikan dalam bentuk tabel dan ringkasan statistik untuk memudahkan pembaca dalam memahami pola dan temuan utama yang diperoleh.

3.1. Pembangunan dan Pengujian Model

Penyusunan model klasifikasi sentimen dilakukan dengan memanfaatkan *pipeline* pada pustaka *scikit-learn* yang mengintegrasikan tahap ekstraksi fitur menggunakan *TF-IDF* dan tahap klasifikasi menggunakan *Support Vector Machine (SVM)*. Proses pelatihan dan pemilihan model terbaik dilakukan melalui *GridSearchCV* dengan skema validasi silang sebanyak lima lipatan (*5-fold cross-validation*). Metrik evaluasi yang digunakan pada tahap ini adalah *f1-score* makro (*f1_macro*), yang dipilih untuk memberikan penilaian kinerja yang seimbang pada setiap kelas sentimen. Dalam penelitian ini dirancang tiga varian eksperimen secara sistematis untuk membandingkan pengaruh konfigurasi fitur dan jenis kernel *SVM* yang ditampilkan pada tabel 4.

Tabel 4. Tiga Varian Eksperimen

| Varian | N-gram | Max Features | TF-IDF Parameters | Kernel SVM | Grid Search Parameters |
|--------|------------------|--------------|---|------------|---|
| V1 | Unigram (1,1) | 10.000 | Default | LinearSVC | $C \in \{0.1, 1, 10\}$ |
| V2 | Uni+Bigram (1,2) | 20.000 | sublinear_tf=True, min_df=2, max_df=0.9 | LinearSVC | $C \in \{0.1, 1, 10\}$ |
| V3 | Uni+Bigram (1,2) | 20.000 | Default | RBF SVC | $C \in \{0.1, 1, 10\}, \gamma \in \{10^{-3}, 10^{-4}\}$ |

Data ulasan dibagi menjadi data latih (80%) dan data uji (20%) dengan total masing-masing 2.145 dan 537 ulasan menggunakan teknik stratified split (`random_state = 42`) untuk menjaga proporsi kelas sentimen. Evaluasi model dilakukan pada data uji menggunakan classification report, confusion matrix, dan nilai macro F1-score guna menilai kinerja klasifikasi secara menyeluruh.

3.2. Analisa Performa Model

Tabel 5. Hasil Pengujian

| Varian | Konfigurasi Utama | F1-macro CV | Akurasi Test | F1 Negatif | F1 Positif | Confusion Matrix (TN, FP, FN, TP) |
|--------|------------------------------|-------------|--------------|------------|------------|-----------------------------------|
| V1 | Unigram + LinearSVC (C=1) | 0.8534 | 88.1% | 0.82 | 0.91 | (148, 45, 19, 325) |
| V2 | Uni+Bigram + LinearSVC (C=1) | 0.8470 | 88.1% | 0.82 | 0.91 | (148, 45, 22, 322) |
| V3 | Uni+Bigram + RBF SVC | 0.3907 | 64.1% | 0.00 | 0.78 | (0, 193, 0, 344) |

Berdasarkan hasil pengujian pada tabel 5, varian V1 yang menggunakan representasi *TF-IDF* unigram dengan *LinearSVC* dan parameter regularisasi $C = 1$ menunjukkan kinerja paling optimal. Model ini mencapai nilai *f1-macro* sebesar 0,8534 pada tahap validasi silang dan akurasi sebesar 88,1% pada data uji. Secara lebih rinci, model mampu mengklasifikasikan ulasan negatif dengan nilai *f1-score* sebesar 0,82 dan ulasan positif sebesar 0,91. Pola kesalahan yang ditunjukkan oleh *confusion matrix* mengindikasikan adanya 19 kesalahan tipe *false negative* dan 45 kesalahan tipe *false positive*, yang masih berada dalam batas wajar untuk data teks ulasan yang bersifat subjektif.

Varian V2, yang menambahkan fitur bigram pada representasi *TF-IDF*, menghasilkan performa yang relatif sebanding dengan varian V1, namun tanpa peningkatan yang signifikan. Nilai *f1-macro* dan akurasi yang diperoleh hampir identik, sementara distribusi kesalahan klasifikasi juga menunjukkan pola yang serupa. Temuan ini mengindikasikan bahwa penggunaan unigram sudah cukup representatif untuk menangkap informasi sentimen dalam ulasan Puskesmas yang umumnya bersifat singkat dan langsung, sehingga penambahan bigram tidak memberikan kontribusi tambahan yang berarti terhadap kinerja model.

Sebaliknya, varian V3 yang menggunakan *SVM* dengan kernel *radial basis function* menunjukkan performa yang jauh lebih rendah. Model ini menghasilkan nilai *f1-macro* yang rendah dan gagal mengklasifikasikan kelas negatif, yang tercermin dari nilai *f1-score* negatif sebesar 0,00. Seluruh data uji cenderung diprediksi sebagai kelas positif, sehingga *confusion matrix* hanya didominasi oleh nilai *true positive* dan *false positive*. Kondisi ini mengindikasikan terjadinya kegagalan pemisahan kelas, yang umum terjadi pada penerapan kernel nonlinier pada data teks berdimensi tinggi dan bersifat jarang, khususnya ketika distribusi kelas tidak sepenuhnya seimbang.

Secara keseluruhan, hasil analisis ini menunjukkan bahwa kombinasi *TF-IDF* unigram dan *LinearSVC* merupakan pendekatan yang paling stabil dan efektif untuk klasifikasi sentimen ulasan layanan Puskesmas. Oleh karena itu, model pada varian V1 selanjutnya digunakan sebagai dasar analisis kepuasan layanan berbasis wilayah kecamatan pada tahap penelitian berikutnya.

3.3. Analisis Spasial Kepuasan per Kecamatan

Analisis spasial dilakukan untuk mengidentifikasi tingkat kepuasan layanan Puskesmas berdasarkan wilayah kecamatan di Kota Denpasar. Hasil klasifikasi sentimen menggunakan model terbaik (*TF-IDF*

unigram dan *LinearSVC*) diagregasikan pada tingkat kecamatan dengan mempertimbangkan jumlah Puskesmas, jumlah ulasan, proporsi sentimen positif dan negatif, nilai rata-rata rating, serta *sentiment score*. Pendekatan ini bertujuan untuk memberikan gambaran komparatif mengenai persepsi masyarakat terhadap kualitas layanan kesehatan di masing-masing wilayah. Tabel 6 menunjukkan *sentiment score* setiap kecamatan.

Tabel 6. Kepuasan per Kecamatan

| Kecamatan | Jumlah Puskesmas | Jumlah Review | SVM | | Rata-rata Rating | Kategori | Sentiment Score |
|------------------|------------------|---------------|----------|----------|------------------|----------|-----------------|
| | | | %Positif | %Negatif | | | |
| Denpasar Utara | 3 | 650 | 72.3% | 27.7% | 4.12 | Tinggi | +0.446 |
| Denpasar Timur | 2 | 611 | 70.2% | 29.8% | 4.05 | Tinggi | +0.404 |
| Denpasar Barat | 2 | 420 | 68.6% | 31.4% | 3.98 | Tinggi | +0.372 |
| Denpasar Selatan | 4 | 950 | 65.1% | 34.9% | 3.85 | Sedang | +0.302 |

Berdasarkan hasil analisis pada tabel 6, Berdasarkan analisis, Kecamatan Denpasar Utara menunjukkan tingkat kepuasan tertinggi dengan proporsi sentimen positif 72,3%, sentiment score +0,446, dan rata-rata rating 4,12, yang mengindikasikan layanan Puskesmas telah memenuhi ekspektasi masyarakat dan dapat menjadi acuan praktik layanan yang baik. Kecamatan Denpasar Timur dan Denpasar Barat juga berada pada kategori kepuasan tinggi dengan proporsi sentimen positif masing-masing 70,2% dan 68,6%, meskipun masih terdapat ruang perbaikan kecil. Sementara itu, Kecamatan Denpasar Selatan berada pada kategori kepuasan sedang dengan proporsi sentimen positif 65,1% dan sentiment score +0,302, yang menunjukkan perlunya perhatian khusus mengingat jumlah Puskesmas dan ulasan yang paling banyak. Secara keseluruhan, analisis spasial menunjukkan adanya variasi tingkat kepuasan antar kecamatan di Kota Denpasar, sehingga analisis sentimen berbasis wilayah penting sebagai dasar evaluasi dan perumusan kebijakan peningkatan layanan kesehatan yang lebih terarah.

3.4. Analisis Kepuasan Layanan Berdasarkan Puskesmas

Analisis kepuasan layanan pada tingkat Puskesmas dilakukan untuk memberikan gambaran rinci kinerja tiap unit layanan serta mengidentifikasi Puskesmas dengan tingkat kepuasan tertinggi yang dapat dijadikan acuan bagi fasilitas lain, mengingat adanya variasi kualitas layanan baik antar kecamatan maupun antar Puskesmas dalam wilayah yang sama.

Tabel 7. Analisis Kepuasan Layanan 5 Puskesmas Tertinggi

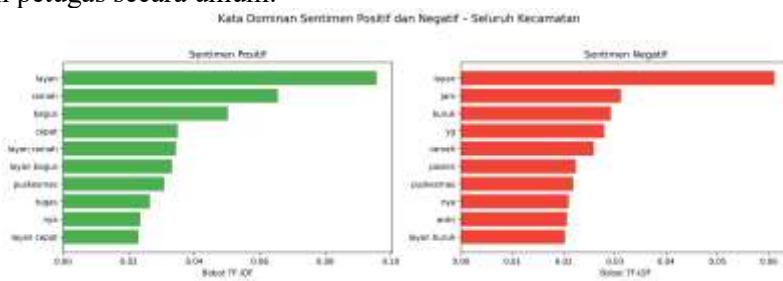
| Rank | Puskesmas | Kecamatan | % Positif SVM | Rata-rata Rating | Jumlah Review | Sentiment Score |
|------|-----------------------------|----------------|---------------|------------------|---------------|-----------------|
| 1 | Puskesmas Denpasar Utara I | Denpasar Utara | 82.5% | 4.35 | 245 | +0.650 |
| 2 | Puskesmas Denpasar Timur I | Denpasar Timur | 78.3% | 4.28 | 189 | +0.566 |
| 3 | Puskesmas Denpasar Utara II | Denpasar Utara | 76.1% | 4.15 | 312 | +0.522 |
| 4 | Puskesmas Denpasar Barat I | Denpasar Barat | 74.2% | 4.08 | 156 | +0.484 |
| 5 | Puskesmas Denpasar Timur II | Denpasar Timur | 72.9% | 4.02 | 223 | +0.458 |

Dari 11 Puskesmas yang dianalisis, lima Puskesmas dengan proporsi sentimen positif tertinggi ditampilkan pada Tabel 7 berdasarkan klasifikasi Support Vector Machine. Puskesmas Denpasar Utara I menempati peringkat tertinggi dengan proporsi sentimen positif 82,5%, sentiment score +0,650, dan rata-rata rating 4,35, yang menunjukkan kepuasan layanan sangat tinggi. Puskesmas Denpasar Timur I

dan Denpasar Utara II juga menunjukkan dominasi sentimen positif di atas 75% dengan jumlah ulasan yang cukup besar. Puskesmas Denpasar Barat I dan Denpasar Timur II melengkapi lima besar dengan proporsi sentimen positif di atas 70%, mencerminkan kinerja layanan yang baik. Variasi sentimen score dan rating menunjukkan perbedaan persepsi masyarakat terhadap kualitas layanan. Secara keseluruhan, analisis ulasan masyarakat mampu mengidentifikasi unit layanan dengan performa unggul dan dapat dimanfaatkan sebagai dasar perumusan strategi peningkatan kualitas layanan Puskesmas berbasis data.

3.5. Analisa Kata Dominan

Visualisasi kata dominan menunjukkan bahwa kepuasan masyarakat terhadap layanan Puskesmas di Kota Denpasar terutama dipengaruhi oleh kualitas pelayanan, sikap petugas yang ramah, serta kecepatan layanan, yang tercermin dari tingginya bobot kata *layan*, *ramah*, *bagus*, dan *cepat* pada sentimen positif. Kemunculan frasa seperti *layan ramah* dan *layan bagus* menegaskan pentingnya interaksi langsung antara petugas dan pasien, sementara kata *puskesmas* dan *tugas* menunjukkan penilaian positif terhadap institusi dan peran petugas secara umum.

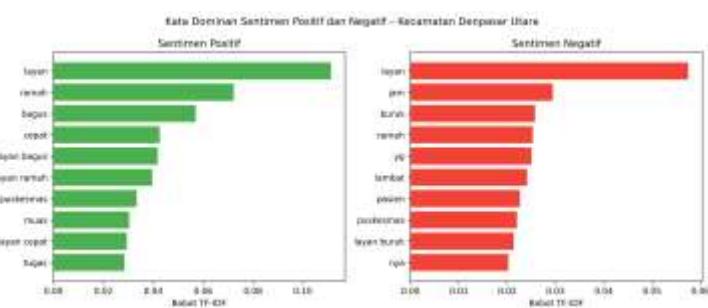


Gambar 1. Kata Dominan pada Sentimen

Pada sentimen negatif, kata *layan* kembali muncul dengan konteks berbeda seperti *layan buruk* dan *antri*. Kemunculan kata *jam*, *antri*, dan *buruk* menunjukkan bahwa ketidakpuasan pengguna terutama dipicu oleh lamanya waktu pelayanan, antrean, dan kualitas layanan yang tidak konsisten. Munculnya kata *ramah* pada sentimen negatif juga mengindikasikan adanya keluhan terhadap sikap petugas. Secara keseluruhan, hasil ini menegaskan bahwa aspek pelayanan menjadi faktor utama pembentuk sentimen masyarakat, sehingga peningkatan sikap petugas dan efisiensi waktu pelayanan perlu menjadi prioritas dalam upaya meningkatkan kepuasan layanan Puskesmas di Kota Denpasar.

3.5.1. Denpasar Utara

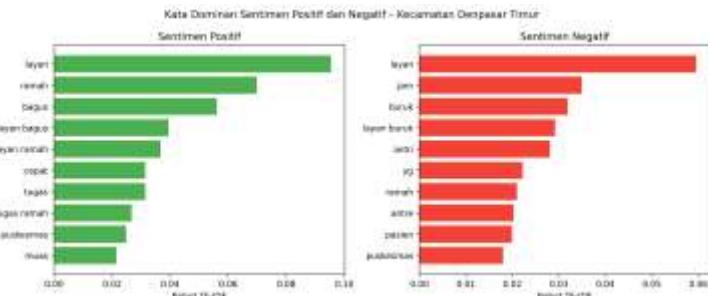
Berdasarkan visualisasi kata dominan sentimen positif dan negatif Kecamatan Denpasar Utara pada Gambar 2, terlihat pola persepsi masyarakat yang cukup kontras namun saling melengkapi. Pada sentimen positif, kata-kata seperti “*layan*”, “*ramah*”, “*bagus*”, dan “*cepat*” muncul dengan bobot TF-IDF tertinggi, yang menunjukkan bahwa aspek pelayanan, sikap petugas, dan kecepatan layanan menjadi faktor utama yang diapresiasi oleh masyarakat terhadap Puskesmas di wilayah ini. Kemunculan frasa seperti “*layan bagus*” dan “*layan ramah*” menguatkan bahwa kualitas interaksi antara petugas dan pasien merupakan sumber kepuasan utama. Sebaliknya, pada sentimen negatif, kata “*layan*” juga tetap dominan, namun disertai dengan kata-kata seperti “*jam*”, “*buruk*”, “*lambat*”, dan “*antri*”, yang mengindikasikan bahwa permasalahan utama yang dikeluhkan masyarakat berkaitan dengan waktu pelayanan, antrean, dan ketidaksesuaian jam layanan. Temuan ini menunjukkan bahwa meskipun kualitas sikap dan pelayanan petugas dinilai positif, masih terdapat aspek operasional, khususnya manajemen waktu dan alur layanan, yang perlu menjadi fokus perbaikan untuk meningkatkan kepuasan masyarakat secara menyeluruh.



Gambar 2. Kata Dominan pada Sentimen di Denpasar Utara

3.5.2. Denpasar Timur

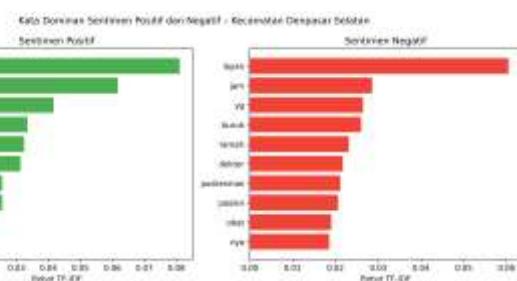
Berdasarkan visualisasi kata dominan sentimen positif dan negatif pada Kecamatan Denpasar Timur pada Gambar 3, terlihat bahwa persepsi kepuasan masyarakat terhadap layanan Puskesmas didominasi oleh aspek kualitas pelayanan dan sikap petugas. Pada sentimen positif, kata “layan”, “ramah”, dan “bagus” memiliki bobot TF-IDF tertinggi, yang mengindikasikan bahwa keramahan petugas, kualitas layanan, serta pengalaman pelayanan yang baik menjadi faktor utama pembentuk kepuasan masyarakat. Kemunculan frasa seperti “layan bagus” dan “layan ramah” menunjukkan bahwa penilaian positif sering kali muncul dalam konteks deskriptif yang mengaitkan langsung kualitas layanan dengan sikap petugas. Sebaliknya, pada sentimen negatif, kata “layan”, “jam”, “buruk”, dan “antri” mendominasi, yang merefleksikan keluhan masyarakat terkait lamanya waktu pelayanan, antrean, serta kualitas layanan yang dirasakan kurang memuaskan. Temuan ini menunjukkan bahwa meskipun layanan Puskesmas di Kecamatan Denpasar Timur secara umum dipersepsikan positif, aspek efisiensi waktu dan manajemen antrean masih menjadi isu utama yang perlu mendapat perhatian untuk peningkatan kualitas layanan secara berkelanjutan.



Gambar 3. Kata Dominan pada Sentimen di Denpasar Timur

3.5.3. Denpasar Selatan

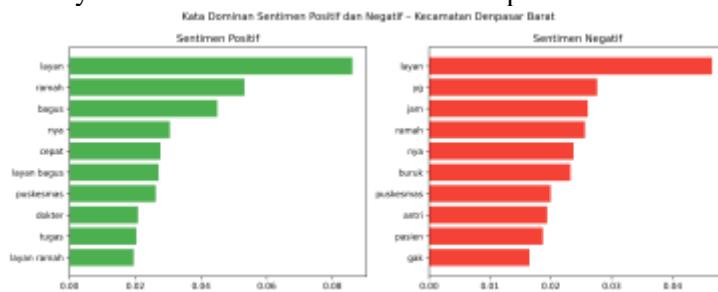
Gambar 4 menampilkan visualisasi kata dominan sentimen positif dan negatif pada Kecamatan Denpasar Selatan, terlihat bahwa persepsi positif masyarakat masih didominasi oleh aspek pelayanan, yang tercermin dari tingginya bobot kata “layan”, “ramah”, dan “bagus”, serta kemunculan kata “cepat” dan “puskesmas” yang menunjukkan apresiasi terhadap proses layanan dan institusi. Hal ini mengindikasikan bahwa interaksi petugas dengan pasien serta kualitas layanan secara umum menjadi faktor utama pembentuk kepuasan. Namun demikian, pada sisi sentimen negatif, kata “layan” juga muncul dengan bobot tertinggi, diikuti oleh “jam”, “buruk”, dan “ramah”, yang mengisyaratkan adanya ketidakpuasan terkait jam pelayanan, kualitas layanan tertentu, serta inkonsistensi sikap petugas. Selain itu, kemunculan kata “dokter”, “obat”, dan “pasien” pada sentimen negatif menunjukkan bahwa aspek medis dan ketersediaan layanan kesehatan masih menjadi sumber keluhan di wilayah ini. Secara keseluruhan, hasil ini menunjukkan bahwa meskipun Kecamatan Denpasar Selatan memiliki persepsi layanan yang relatif positif, masih terdapat isu spesifik yang perlu mendapat perhatian untuk meningkatkan kepuasan masyarakat secara berkelanjutan.



Gambar 4. Kata Dominan pada Sentimen di Denpasar Selatan

3.5.4. Denpasar Barat

Berdasarkan visualisasi kata dominan sentimen positif dan negatif di Kecamatan Denpasar Barat yang ditampilkan pada gambar 5, terlihat bahwa persepsi positif masyarakat terutama didominasi oleh kata “layan”, “ramah”, dan “bagus”, yang mengindikasikan bahwa kualitas pelayanan dan sikap petugas menjadi aspek utama yang diapresiasi oleh pengguna layanan Puskesmas di wilayah ini. Kemunculan kata seperti “cepat”, “dokter”, dan “tugas” pada sentimen positif juga menunjukkan bahwa kecepatan pelayanan serta peran tenaga kesehatan turut berkontribusi terhadap kepuasan masyarakat. Di sisi lain, sentimen negatif masih didominasi oleh kata “layan”, yang menandakan bahwa isu pelayanan juga menjadi sumber utama keluhan. Kata-kata seperti “jam”, “antri”, “buruk”, dan “pasien” mengindikasikan permasalahan yang berkaitan dengan waktu pelayanan, antrean, serta pengalaman pasien saat menerima layanan. Temuan ini menunjukkan bahwa meskipun kualitas pelayanan secara umum dinilai baik, aspek pengelolaan waktu dan antrean masih perlu mendapatkan perhatian lebih untuk meningkatkan kepuasan layanan Puskesmas di Kecamatan Denpasar Barat.



Gambar 5. Kata Dominan pada Sentimen di Denpasar Barat

4. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada seluruh pihak yang secara langsung maupun tidak langsung telah memberikan dukungan dalam pelaksanaan penelitian ini, khususnya penyedia platform Google Maps yang memungkinkan ketersediaan data ulasan sebagai sumber data penelitian. Apresiasi juga disampaikan kepada rekan sejawat di lingkungan akademik atas diskusi, masukan, dan saran konstruktif selama proses perancangan metodologi, analisis data, hingga penyusunan naskah.

5. Referensi

- [1] Aswitha Sari Suprihathin, “Analisis Tingkat Kepuasan Pasien Terhadap Pelayanan Kefarmasian di Puskesmas Malawili Kabupaten Sorong,” *JPM*, vol. 3, no. 3, pp. 191–199, Apr. 2025, doi: 10.61132/obat.v3i3.1305.
- [2] S. D. Beni Azhar, “ANALISIS KUALITAS PELAYANAN PUBLIK DALAM MEMBERIKAN KEPUASAN KEPADA MASYARAKAT PADA PUSKESMAS LEMPAKE KOTA SAMARINDA,” *DDK*, vol. 21, no. 2, p. 28, Dec. 2020, doi: 10.31293/ddk.v21i2.5002.
- [3] I. A. M. C. Dewi, I. K. Dharmendra, and N. W. Setiasih, “ANALISIS SENTIMEN REVIEW APLIKASI SATU SEHAT MOBILE MENGGUNAKAN MODEL SAMPLING TOMEK LINKS,” *Jurnal Teknologi Informasi dan Komputer*, vol. 9, no. 5, Oct. 2023, doi: 10.36002/jutik.v9i5.2644.

- [4] F. N. Ziyadah and E. Alisah, "Analisis Fuzzy Servqual Dalam Mengukur Tingkat Kepuasan Pasien BPJS Kesehatan Terhadap Kualitas Layanan Pada UPTD Puskesmas," *JRMM*, vol. 3, no. 6, pp. 258–273, Nov. 2024, doi: 10.18860/jrmm.v3i6.27847.
- [5] I. K. Dharmendra, I. M. A. W. Putra, and Y. P. Atmojo, "Evaluation of the Effectiveness of SMOTE and Random Under Sampling in Emotion Classification of Tweets," *Journal of Informatics*, vol. 9, no. 2, p. 182, Dec. 2024, doi: 10.51211/itbi.v9i2.3183.
- [6] I. K. Dharmendra, I. G. N. A. Kusuma, I. A. M. C. Dewi, and Edwar, "IMPLEMENTASI TEXT MINING UNTUK KLASIFIKASI OPINI ALUMNI PADA PERGURUAN TINGGI," *Jurnal Teknologi Informasi dan Komputer*, vol. 9, no. 3, Art. no. 3, Jul. 2023, Accessed: Aug. 18, 2023. [Online]. Available: <https://jurnal.undhirabali.ac.id/index.php/jutik/article/view/2504>
- [7] E. Edwar, I. G. A. N. R. Semadi, M. Samsudin, and I. K. Dharmendra, "Perbandingan Metode Seleksi Fitur Pada Analisis Sentimen (Studi Kasus Opini PILKADA DKI 2017)," *INFORMATICS FOR EDUCATORS AND PROFESSIONAL : Journal of Informatics*, vol. 8, no. 1, pp. 11–18, Jul. 2023, doi: 10.51211/itbi.v8i1.2408.
- [8] Y. Karaman, F. Akdeniz, B. K. Savaş, and Y. Becerikli, "A Comparative Analysis of SVM, LSTM and CNN-RNN Models for the BBC News Classification," in *Innovations in Smart Cities Applications Volume 6*, M. Ben Ahmed, A. A. Boudhir, D. Santos, R. Dionisio, and N. Benaya, Eds., Cham: Springer International Publishing, 2023, pp. 473–483. doi: 10.1007/978-3-031-26852-6_44.
- [9] B. B. Baskoro, I. Susanto, and S. Khomsah, "Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR)," *INISTA (Journal of Informatics Information System Software Engineering and Applications)*, vol. 3, no. 2, Art. no. 2, Jun. 2021, doi: 10.20895/inista.v3i2.218.
- [10] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *mib*, vol. 5, no. 2, p. 640, Apr. 2021, doi: 10.30865/mib.v5i2.2937.
- [11] B. Obeidat and S. Alourd, "Healthcare equity in focus: bridging gaps through a spatial analysis of healthcare facilities in Irbid, Jordan," *Int J Equity Health*, vol. 23, no. 1, p. 52, Mar. 2024, doi: 10.1186/s12939-024-02120-8.
- [12] S. Lin, S. Qin, L. Peng, X. Sun, and X. Dou, "Assessing Geographic Barriers to Access Long-Term Services and Supports in Chengdu, China: A Spatial Accessibility Analysis," *Sustainability*, vol. 17, no. 7, p. 3222, Apr. 2025, doi: 10.3390/su17073222.
- [13] H.-Q. Huang-fu *et al.*, "Spatial profiling of geographical accessibility to maternal healthcare and coverage of maternal health service utilisation in Nepal: a geospatial analysis based on demographic and health survey," *BMJ Glob Health*, vol. 10, no. 2, p. e017229, Feb. 2025, doi: 10.1136/bmjgh-2024-017229.
- [14] "Aplikasi Sistem Antrian Provinsi Bali | Satu Data Indonesia Provinsi Bali." Accessed: Jan. 05, 2026. [Online]. Available: <https://balisatudata.baliprov.go.id/application/sistem-antrian-provinsi-bali>
- [15] Y. Liu, Y. Wan, X. Shen, Z. Ye, and J. Wen, "Product Customer Satisfaction Measurement Based on Multiple Online Consumer Review Features," *Information*, vol. 12, no. 6, p. 234, May 2021, doi: 10.3390/info12060234.
- [16] "Stemming Bahasa Indonesia · sastrawi/sastrawi Wiki · GitHub." Accessed: Jun. 26, 2019. [Online]. Available: <https://github.com/sastrawi/sastrawi/wiki/Stemming-Bahasa-Indonesia>
- [17] I. K. Dharmendra, N. N. U. Januhari, I. P. Ramayasa, and I. M. A. W. Putra, "Uji Komparasi Sentiment Analysis Pada Opini Alumni Terhadap Perguruan Tinggi," *Jurnal Teknik Informatika UNIKA Santo Thomas*, pp. 1–6, May 2022, doi: 10.54367/jtiust.v7i1.1748.
- [18] M. Das, S. K., and P. J. A. Alphonse, "A Comparative Study on TF-IDF feature Weighting Method and its Analysis using Unstructured Dataset," 2023, *arXiv*. doi: 10.48550/ARXIV.2308.04037.
- [19] E. M. O. N. Haryanto, A. K. A. Estetikha, and R. A. Setiawan, "IMPLEMENTASI SMOTE UNTUK MENGATASI IMBALANCED DATA PADA SENTIMENT ANALYSIS SENTIMENT HOTEL DI NUSA TENGGARA BARAT DENGAN MENGGUNAKAN ALGORITMA SVM," *Informasi Interaktif*, vol. 7, no. 1, Art. no. 1, Jan. 2022.