Available *online* at: https://bit.ly/SISFOTEKJAR

Jurnal Sistem Informasi dan Teknologi Jaringan

(SISFOTEKJAR)

ISSN (Print) 2808-7917 | ISSN (Online) 2807-9256

Analisis Sentimen Ulasan Wisata Sumenep dengan SVM dan TF-IDF

M. Yahya Ubaidillah[1], Saiful Bahri[2]

[1,2] Program Studi Informatika, Fakultas Sains dan Kesehatan, Universitas PGRI Sumenep, Jl. Trunojoyo, Gedungan Barat, Gedungan, Kecamatan Batuan, Kabupaten Sumenep, Jawa Timur 69451, Indonesia
[1]ubay.connect@stkippgrisumenep.ac.id, [2]saifulbahrirois94@gmail.com

ARTICLE INFORMATION

Received: February 25, 2026
Revised: March 20, 2026
Available online: March 31, 2026

KATA KUNCI

analisis sentimen, ulasan wisata, Google Maps, Support Vector Machine, TF-IDF

CORRESPONDENCE

Phone: +62 857-3534-5201

E-mail: ubay.connect@stkippgrisumenep.ac.id

ABSTRAK

Penelitian ini bertujuan mengklasifikasikan sentimen ulasan destinasi wisata di Kabupaten Sumenep menggunakan pembobotan *Term Frequency–Inverse Document Frequency* (TF-IDF) dan algoritma *Support Vector Machine* (SVM), serta menilai kinerja model pada distribusi kelas yang tidak seimbang. Data diperoleh melalui *web scraping* ulasan enam destinasi wisata pada Google Maps. Dari 797 ulasan yang terkumpul, sebanyak 725 ulasan digunakan setelah data tanpa teks dan ulasan dengan peringkat tiga dikeluarkan. Peringkat empat dan lima diberi label positif, sedangkan peringkat satu dan dua diberi label negatif. Praproses teks meliputi *case folding*, pembersihan teks, normalisasi kata, penghapusan *stopword*, dan *stemming*. Teks kemudian direpresentasikan menggunakan TF-IDF dengan maksimum 3.000 fitur berupa *unigram* dan *bigram*. Klasifikasi dilakukan menggunakan SVM dengan kernel linear, parameter $C = 1{,}0$, dan bobot kelas balanced. Dataset akhir terdiri atas 643 ulasan positif dan 82 ulasan negatif. Model menghasilkan akurasi 87,50%, nilai F1 tertimbang 85,61%, nilai F1 makro 59,09%, dan *balanced accuracy* 57,42%. *Recall* kelas positif mencapai 96,09%, sedangkan *recall* kelas negatif hanya 18,75%. Model menunjukkan performa yang tinggi dalam mengenali ulasan positif, tetapi belum memadai dalam mendeteksi ulasan negatif. Ketimpangan kelas dan keterbatasan variasi ulasan negatif masih memengaruhi hasil klasifikasi, sehingga evaluasi model perlu mempertimbangkan metrik per kelas, bukan hanya akurasi keseluruhan.

1. PENDAHULUAN

Kabupaten Sumenep memiliki potensi pariwisata yang beragam, meliputi wisata bahari, alam, sejarah, dan budaya. Pantai Lombang, Pantai Slopeng, Gili Labak, Gili Iyang, Pantai Sembilan, dan Museum Keraton Sumenep merupakan beberapa destinasi yang banyak dikenal oleh wisatawan. Perkembangan pariwisata daerah terlihat dari peningkatan jumlah kunjungan wisatawan. Data Dinas Kebudayaan, Pemuda, Olahraga, dan Pariwisata Kabupaten Sumenep menunjukkan bahwa jumlah kunjungan meningkat dari 1.057.622 orang pada 2022 menjadi 1.523.102 orang pada 2023 [1]. Peningkatan tersebut perlu diikuti dengan evaluasi terhadap pengalaman pengunjung agar pengembangan destinasi tidak hanya berorientasi pada jumlah kunjungan, tetapi juga memperhatikan kualitas pelayanan dan pengelolaan wisata.

Pengalaman wisatawan kini banyak disampaikan melalui platform daring. Google Maps tidak hanya menyediakan informasi mengenai lokasi destinasi, tetapi juga memberikan ruang bagi pengunjung untuk menyampaikan rating dan ulasan tertulis. Informasi tersebut dapat digunakan oleh calon wisatawan

sebagai bahan pertimbangan sebelum berkunjung, sedangkan pengelola destinasi dapat memanfaatkannya untuk memahami penilaian pengunjung. Sejumlah penelitian telah menggunakan ulasan Google Maps untuk mengidentifikasi kecenderungan sentimen terhadap tempat wisata [2], [3]. Namun, ketika jumlah ulasan semakin banyak, pembacaan secara manual menjadi tidak efisien dan menyulitkan proses penarikan kesimpulan secara konsisten.

Analisis sentimen dapat digunakan untuk mengidentifikasi opini, evaluasi, dan sikap yang terkandung dalam teks, terutama untuk menentukan kecenderungan sentimen positif atau negatif [4]. Dalam proses klasifikasi, teks tidak dapat langsung diolah oleh algoritma sehingga perlu diubah menjadi representasi numerik. Salah satu pendekatan yang banyak digunakan adalah *Term Frequency–Inverse Document Frequency* (TF-IDF). Pembobotan TF-IDF memberikan nilai pada suatu kata berdasarkan frekuensi kemunculannya dalam sebuah dokumen dan tingkat kelangkaannya pada seluruh kumpulan dokumen [5]. Dengan pembobotan tersebut, kata yang dianggap lebih representatif terhadap isi suatu ulasan dapat memperoleh bobot yang lebih besar.

Algoritma yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM). SVM membentuk bidang pemisah atau *hyperplane* dengan margin maksimum untuk memisahkan data dari kelas yang berbeda [6]. Pada klasifikasi teks, data umumnya memiliki jumlah fitur yang tinggi karena setiap kata dapat menjadi fitur, sedangkan sebagian besar nilai fiturnya bersifat nol. Joachims menunjukkan bahwa karakteristik tersebut membuat SVM sesuai digunakan untuk mempelajari pengklasifikasi teks dengan banyak fitur yang relevan [7]. Kombinasi representasi TF-IDF dan SVM kemudian banyak diterapkan dalam penelitian klasifikasi sentimen.

Penerapan SVM pada ulasan wisata telah dilakukan pada beberapa objek penelitian. Suryawan, Utami, dan Fredlina menganalisis 669 ulasan wisata di Ubud yang diperoleh dari TripAdvisor. Penelitian tersebut menghasilkan akurasi 84,01%, presisi 90,40%, recall 89,83%, dan nilai F1 sebesar 90,11% [8]. Syahlan, Irmayanti, dan Alam menggunakan komentar Google Maps untuk mengklasifikasikan sentimen pengunjung Taman Air Mancur Sri Baduga Purwakarta ke dalam kelas positif, negatif, dan netral. Model yang digunakan menghasilkan akurasi 81%, presisi 94%, dan recall 99% [2]. Ipmawati, Saifulloh, dan Kusnawi juga menggunakan ulasan Google Maps terhadap tempat wisata di Daerah Istimewa Yogyakarta dan memperoleh rata-rata akurasi sebesar 83,8% [3].

Kajian yang lingkup wilayahnya dekat dengan penelitian ini dilakukan oleh Fatah, Rochman, Kamil, dan Su'ud terhadap opini wisata Madura yang diperoleh dari Twitter [9]. Penelitian tersebut menggunakan 1.814 data dan menerapkan SVM dengan validasi silang lima bagian. Hasil terbaiknya mencapai akurasi 92,592%, tetapi nilai F1 yang dilaporkan hanya 32,051%. Perbedaan yang besar antara akurasi dan nilai F1 memperlihatkan bahwa nilai akurasi saja belum tentu cukup menggambarkan kemampuan model pada setiap kelas. Penelitian yang lebih baru oleh Ramdani dan Cahyana menggunakan 2.469 ulasan Google Maps dan Twitter mengenai Kampung Naga. SVM dengan representasi TF-IDF menghasilkan akurasi 83%, sedangkan LSTM menghasilkan akurasi 81% [10].

Kombinasi TF-IDF dan SVM juga digunakan pada objek ulasan selain destinasi wisata. Nirmala dan Dwidasmara menggunakan 3.000 ulasan aplikasi M-Paspor dengan komposisi 1.500 data positif dan 1.500 data negatif. SVM kernel RBF yang dievaluasi menggunakan *k-fold cross-validation* menghasilkan rata-rata akurasi 83,62%, presisi 84,70%, recall 83,65%, dan nilai F1 sebesar 83,51% [11]. Zebua dan Dwidasmara menggunakan TF-IDF dan SVM pada ulasan aplikasi Citilink dan memperoleh akurasi 88%, presisi 90%, recall 83%, serta nilai F1 sebesar 85% [12]. Kedua penelitian tersebut menunjukkan bahwa hasil klasifikasi dipengaruhi oleh karakteristik dataset, distribusi kelas, konfigurasi model, dan skenario evaluasi yang digunakan.

Distribusi kelas menjadi salah satu persoalan penting pada klasifikasi sentimen. Apabila jumlah data pada satu kelas jauh lebih besar, algoritma dapat lebih banyak mempelajari karakteristik kelas mayoritas dan kurang mampu mengenali kelas minoritas [13]. Dalam kondisi tersebut, akurasi yang tinggi belum tentu diikuti oleh kemampuan yang seimbang dalam mendeteksi setiap kelas. Sumantiawan, Suseno, dan Syafei menunjukkan bahwa penerapan SMOTE dan Tomek Links pada klasifikasi ulasan pelanggan menggunakan SVM meningkatkan akurasi dari 68% menjadi 92% dan nilai F1 dari 71% menjadi 89% [14]. Temuan tersebut menunjukkan bahwa distribusi kelas serta pemilihan metrik evaluasi perlu diperhatikan dalam menafsirkan performa model.

Penelitian terdahulu telah memperlihatkan bahwa SVM dapat digunakan untuk menganalisis sentimen wisata di Ubud, Purwakarta, Yogyakarta, Kampung Naga, dan wilayah Madura. Namun, penelitian-penelitian tersebut memiliki perbedaan dalam sumber data, objek destinasi, pembentukan label, jumlah kelas, dan distribusi data. Penelitian wisata Madura sebelumnya menggunakan opini dari Twitter dan tidak secara khusus membahas ulasan Google Maps dari destinasi wisata di Kabupaten Sumenep [9]. Sementara itu, penelitian berbasis Google Maps pada wilayah lain belum menggambarkan karakteristik ulasan Pantai Sembilan, Gili Labak, Pantai Slopeng, Museum Keraton Sumenep, Pantai Lombang, dan Gili Iyang. Kesenjangan tersebut menjadi dasar penggunaan data ulasan dari enam destinasi tersebut dalam penelitian ini.

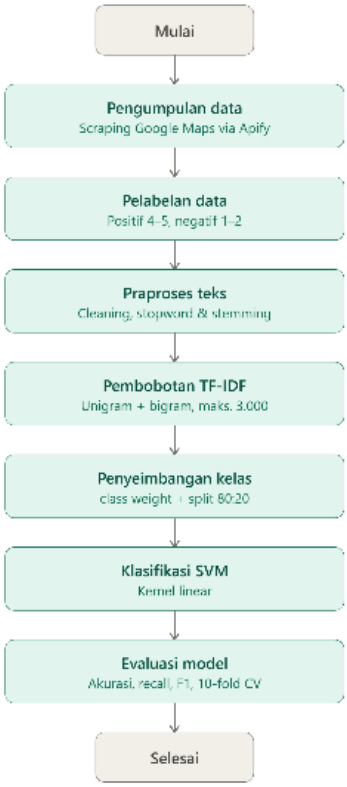
Penelitian ini menggunakan TF-IDF untuk merepresentasikan teks dan SVM kernel linear untuk mengklasifikasikan ulasan ke dalam sentimen positif dan negatif. Karena distribusi data didominasi oleh kelas positif, parameter bobot kelas seimbang diterapkan selama proses pelatihan. Kontribusi penelitian terletak pada penggunaan data ulasan enam destinasi wisata di Kabupaten Sumenep serta penyajian evaluasi performa setiap kelas, bukan hanya akurasi keseluruhan. Evaluasi per kelas diperlukan untuk mengetahui sejauh mana model dapat mengenali ulasan negatif sebagai kelas dengan jumlah data yang lebih sedikit.
Berdasarkan uraian tersebut, penelitian ini bertujuan menerapkan pembobotan TF-IDF dan algoritma SVM untuk mengklasifikasikan sentimen ulasan destinasi wisata di

Kabupaten Sumenep. Penelitian ini juga bertujuan mengevaluasi kemampuan model dalam mengenali sentimen positif dan negatif menggunakan *confusion matrix*, akurasi, presisi, recall, dan nilai F1. Hasil penelitian diharapkan dapat memberikan gambaran yang lebih proporsional mengenai kemampuan dan keterbatasan model dalam mengolah ulasan wisata dengan distribusi kelas yang tidak seimbang.

2. METODE PENELITIAN

1. Rancangan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan rancangan eksperimen klasifikasi teks. Unit analisis penelitian berupa ulasan pengunjung terhadap destinasi wisata di Kabupaten Sumenep yang tersedia secara publik pada Google Maps. Proses penelitian terdiri atas pengumpulan data, seleksi data, pelabelan sentimen, praproses teks, pembobotan *Term Frequency–Inverse Document Frequency* (TF-IDF), pembagian data, pelatihan model *Support Vector Machine* (SVM), dan evaluasi model. Alur penelitian disajikan pada Gambar 1.



Gambar 1. Alur Penelitian

2. Pengumpulan dan Seleksi Data

Data penelitian diperoleh dari ulasan publik pada Google Maps terhadap enam destinasi wisata di Kabupaten Sumenep, yaitu Pantai Sembilan, Gili Labak, Pantai Slopeng, Museum Keraton Sumenep, Pantai Lombang, dan Gili Iyang. Pengumpulan data dilakukan menggunakan teknik web scraping melalui platform Apify dengan perangkat Google Maps Reviews Scraper.

Atribut yang dikumpulkan meliputi nama destinasi, teks ulasan, *rating* bintang, dan nama pengulas. Namun, identitas pengulas tidak digunakan dalam pemodelan maupun ditampilkan pada hasil penelitian. Atribut yang digunakan dalam proses analisis dibatasi pada nama destinasi, teks ulasan, dan *rating* bintang. Data yang diperoleh kemudian diperiksa untuk menghapus ulasan duplikat dan ulasan yang tidak memiliki teks.

Seleksi data dilakukan sebelum tahap pelabelan untuk memastikan bahwa setiap ulasan memiliki teks yang dapat diolah. Ulasan tanpa teks dikeluarkan karena tidak dapat direpresentasikan menjadi fitur TF-IDF, sedangkan data duplikat dihapus agar satu ulasan tidak memberikan pengaruh berulang dalam proses pelatihan. Selanjutnya, ulasan dengan *rating* tiga tidak disertakan karena penelitian ini menggunakan klasifikasi biner yang terdiri atas sentimen positif dan negatif. Tahapan tersebut mengurangi jumlah data dari 797 menjadi 725 ulasan yang kemudian digunakan pada proses praproses dan klasifikasi. Prosedur seleksi ini memastikan bahwa data yang dianalisis sesuai dengan kriteria dan rancangan klasifikasi yang telah ditetapkan.

Proses pengumpulan menghasilkan 797 ulasan unik. Sebaran ulasan berdasarkan destinasi disajikan pada Tabel 1.

Tabel 1. Sebaran Ulasan Berdasarkan Destinasi

Lokasi Destinasi	Jumlah Ulasan
Pantai Sembilan	217
Gili Labak	152
Pantai Slopeng	136
Museum Keraton Sumenep	126
Pantai Lombang	116
Gili Iyang	50
Total	797

3. Pelabelan Sentimen

Pelabelan sentimen dilakukan berdasarkan *rating* bintang yang diberikan oleh pengunjung. Ulasan dengan *rating* empat dan lima diberi label positif, sedangkan ulasan dengan *rating* satu dan dua diberi label negatif. Ulasan dengan *rating* tiga tidak digunakan karena penelitian ini menerapkan klasifikasi biner dan menempatkan *rating* tersebut sebagai kategori netral. Ketentuan pelabelan disajikan pada Tabel 2.

Tabel 2. Ketentuan Pelabelan Sentimen

Rating	Label Sentimen	Status Data
1–2	Negatif	Digunakan
3	Netral	Tidak digunakan
4–5	Positif	Digunakan

Setelah ulasan tanpa teks dan ulasan dengan rating tiga dikeluarkan, diperoleh 725 ulasan yang terdiri atas 643 ulasan positif dan 82 ulasan negatif. Ulasan positif mencakup 88,69% data, sedangkan ulasan negatif mencakup 11,31% data.

Pelabelan berdasarkan rating digunakan sebagai label operasional atau label proksi. Rating bintang tidak selalu menggambarkan keseluruhan isi teks karena suatu ulasan dapat mengandung pujian dan keluhan dalam satu kalimat. Oleh karena itu, hasil klasifikasi dalam penelitian ini ditafsirkan sebagai kemampuan model mempelajari kecenderungan label berbasis rating dari teks ulasan.

Perbedaan jumlah data pada kedua kelas menunjukkan adanya ketidakseimbangan kelas. Dalam kondisi tersebut, model cenderung lebih banyak mempelajari pola kelas mayoritas sehingga kemampuan mengenali kelas minoritas dapat menurun [13].

4. *Praproses Teks*

Praproses dilakukan untuk menyeragamkan bentuk teks dan mengurangi unsur yang tidak digunakan sebagai fitur klasifikasi. Tahapan praproses terdiri atas *case folding*, pembersihan teks, normalisasi kata, penghapusan *stopword*, dan *stemming*. Urutan praproses ditunjukkan pada Gambar 2.

Gambar 2. Alur Praproses Teks

Tahap *case folding* mengubah seluruh huruf menjadi huruf kecil. Proses ini dilakukan agar kata yang memiliki perbedaan kapitalisasi diperlakukan sebagai fitur yang sama. Sebagai contoh, kata "Pantai", "pantai", dan "PANTAI" diubah menjadi "pantai".

Pembersihan teks dilakukan dengan menghapus URL, angka, tanda baca, simbol, emoji, dan karakter nonalfabet. Spasi berlebih yang terbentuk setelah penghapusan karakter juga diseragamkan. Tahap ini bertujuan mengurangi unsur yang tidak digunakan dalam pembentukan fitur.

Normalisasi kata dilakukan untuk mengubah bentuk tidak baku, singkatan, dan penulisan informal menjadi bentuk yang lebih seragam. Sebagai contoh, kata "yg" diubah menjadi "yang", "ga" menjadi "tidak", dan "mantab" menjadi "mantap". Proses normalisasi menggunakan kamus yang berisi pasangan kata tidak baku dan kata penggantinya.

Penghapusan *stopword* dilakukan terhadap kata-kata umum yang tidak memberikan informasi pembeda menggunakan daftar *stopword* bahasa Indonesia dari pustaka Sastrawi. Daftar tersebut perlu diperiksa agar kata negasi, seperti "tidak", "bukan", "belum", "kurang", dan "jangan", tidak dihapus apabila masih diperlukan untuk mempertahankan makna sentimen.

Tahap *stemming* dilakukan untuk mengubah kata berimbuhan menjadi kata dasar menggunakan pustaka Sastrawi. Sebagai contoh, kata "mengunjungi" diubah menjadi "kunjung", sedangkan kata "pemandangan" diubah menjadi "pandang". Proses tersebut menerapkan pendekatan *confix-stripping* yang dikembangkan untuk menangani pembentukan kata berimbuhan dalam bahasa Indonesia [15].

Untuk memperlihatkan perubahan teks pada setiap tahap, beberapa ulasan digunakan sebagai contoh. Hasil praproses menunjukkan bahwa karakter yang tidak diperlukan dapat dihilangkan, kata tidak baku dapat diseragamkan, dan kata berimbuhan dapat diubah ke bentuk dasar. Contoh hasil praproses disajikan pada Tabel 3.

Tabel 3. Contoh Hasil Praproses Teks

No.	Teks Ulasan Awal	Teks Hasil Praproses
1	Pantainya bagus banget, cocok untuk liburan keluarga.	pantai bagus cocok libur keluarga
2	Tempatnya indah, tetapi fasilitas kurang terawat.	tempat indah fasilitas kurang rawat
3	Akses jalan tidak bagus dan cukup sulit dilalui.	akses jalan tidak bagus sulit lalu
4	Pemandangannya sangat indah dan menenangkan.	pandang sangat indah tenang

5. *Pembobotan TF-IDF*

Teks hasil praproses diubah menjadi representasi numerik menggunakan TF-IDF. Pembobotan ini memperhitungkan frekuensi kemunculan suatu kata dalam sebuah dokumen dan tingkat kelangkaan kata tersebut pada seluruh kumpulan dokumen [5].

Nilai *term frequency* menunjukkan frekuensi kemunculan kata $t$dalam dokumen $d$dan dirumuskan sebagai:

$$TF(t,d) = f_{t,d} \qquad (1)$$

dengan $f_{t,d}$ merupakan jumlah kemunculan kata $t$dalam dokumen d.

Nilai *inverse document frequency* mengukur tingkat kelangkaan kata pada seluruh dokumen dan dirumuskan sebagai:

$$IDF(t) = \log\left(\frac{N}{df_t}\right) \qquad (2)$$

dengan N merupakan jumlah seluruh dokumen dan df_t merupakan jumlah dokumen yang mengandung kata t.

Bobot TF-IDF diperoleh melalui perkalian nilai TF dan IDF:

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

Setiap baris pada matriks TF-IDF mewakili satu ulasan, sedangkan setiap kolom mewakili kata yang terbentuk dalam kosakata. Matriks tersebut digunakan sebagai data masukan dalam proses pelatihan dan pengujian SVM.

6. Klasifikasi Support Vector Machine

Klasifikasi sentimen dilakukan menggunakan algoritma *Support Vector Machine* (SVM). Algoritma ini bekerja dengan membentuk bidang pemisah atau *hyperplane* yang memiliki margin maksimum antara kelas positif dan negatif [6]. Kernel linear digunakan karena fitur yang dihasilkan melalui pembobotan TF-IDF memiliki dimensi tinggi dan sebagian besar nilainya bersifat nol. Karakteristik tersebut sesuai dengan penerapan SVM pada klasifikasi teks [7].

Fungsi keputusan SVM linear dapat dinyatakan sebagai:

$$f(\boldsymbol{x}) = \text{sign}(\boldsymbol{w}^T \boldsymbol{x} + b) \qquad (4)$$

dengan $\boldsymbol{x}$merupakan vektor fitur TF-IDF, $\boldsymbol{w}$merupakan vektor bobot, dan $b$merupakan nilai bias. Hasil fungsi keputusan menentukan kelas sentimen dari setiap ulasan.

Model dibangun menggunakan kelas SVC dari pustaka Scikit-learn dengan kernel linear dan parameter regularisasi $C = 1{,}0$. Parameter C mengatur keseimbangan antara upaya memaksimalkan margin pemisah dan meminimalkan kesalahan klasifikasi. Nilai $C = 1{,}0$ digunakan untuk memberikan keseimbangan antara kedua tujuan tersebut.

Dataset penelitian memiliki distribusi yang tidak seimbang, yaitu 643 ulasan positif dan 82 ulasan negatif. Untuk mengurangi kecenderungan model terhadap kelas positif, parameter class_weight="balanced" diterapkan pada proses pelatihan. Parameter tersebut memberikan penalti yang lebih besar terhadap kesalahan klasifikasi pada kelas yang memiliki jumlah data lebih sedikit. Strategi ini tidak menambah jumlah ulasan negatif, tetapi menyesuaikan kontribusi kesalahan masing-masing kelas dalam proses pembentukan model [13].

Parameter probability=True diaktifkan agar model dapat menghasilkan estimasi probabilitas untuk setiap kelas. Konfigurasi model yang digunakan dapat dirangkum sebagai berikut:

Tabel 4. Konfigurasi Model Support Vector Machine

Parameter	Nilai
Model	*Support Vector Classification*
Kernel	Linear
Parameter regularisasi (C)	1
Bobot kelas	balanced
Estimasi probabilitas	TRUE

Model dilatih menggunakan data latih yang telah direpresentasikan dalam bentuk matriks TF-IDF. Setelah proses pelatihan selesai, model digunakan untuk memprediksi kelas sentimen pada data uji. Kinerja model selanjutnya dianalisis menggunakan *confusion matrix*, akurasi, presisi, *recall*, dan nilai F1 untuk setiap kelas.

7. Skenario Pengujian

Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan fungsi train_test_split. Pembagian dilakukan secara terstratifikasi melalui parameter stratify=y agar proporsi kelas positif dan negatif pada data latih maupun data uji tetap mendekati distribusi dataset awal. Parameter random_state=42 digunakan untuk menjaga konsistensi pembagian data ketika eksperimen dijalankan kembali.

Teks ulasan direpresentasikan menggunakan TF-IDF dengan jumlah fitur maksimum sebanyak 3.000. Konfigurasi ngram_range=(1,2) digunakan untuk membentuk fitur berupa *unigram* dan *bigram*. *Unigram* merepresentasikan kemunculan satu kata, sedangkan *bigram* merepresentasikan dua kata yang muncul secara berurutan. Penggunaan kedua jenis fitur tersebut memungkinkan model mengenali kata tunggal sekaligus frasa pendek yang memiliki makna sentimen, seperti “sangat indah”, “tidak bersih”, dan “kurang terawat”.

Pada pengujian menggunakan pembagian 80:20, model dilatih menggunakan data latih dan selanjutnya digunakan untuk memprediksi label sentimen pada data uji. Hasil prediksi dibandingkan dengan label aktual untuk memperoleh *confusion matrix*, akurasi, presisi, *recall*, dan nilai F1. Presisi dan *recall* pada keluaran utama dihitung dengan menetapkan kelas positif sebagai kelas acuan, sedangkan nilai F1 dihitung menggunakan pendekatan tertimbang agar kontribusi setiap kelas disesuaikan dengan jumlah datanya.

Selain pengujian menggunakan data uji, kestabilan model dianalisis melalui validasi silang terstratifikasi sebanyak sepuluh bagian. Pada setiap putaran, sembilan bagian digunakan untuk melatih model dan satu bagian digunakan sebagai data validasi. Proses tersebut diulang sampai seluruh bagian pernah menjadi data validasi. Stratifikasi digunakan untuk mempertahankan proporsi kelas pada setiap bagian data sehingga evaluasi tidak hanya bergantung pada satu pembagian dataset [16].

Metrik yang digunakan pada validasi silang adalah nilai F1 tertimbang. Nilai tersebut dipilih karena dataset memiliki jumlah ulasan positif dan negatif yang tidak seimbang. Hasil validasi silang dilaporkan dalam bentuk nilai rata-rata dan simpangan baku. Nilai rata-rata menggambarkan performa umum model, sedangkan simpangan baku menunjukkan tingkat perubahan performa pada pembagian data yang berbeda.

Tabel 5. Konfigurasi Skenario Pengujian.

Komponen	Konfigurasi
Rasio data latih dan data uji	80:20
Metode pembagian	Terstratifikasi
random_state	42
Jumlah maksimum fitur	3.000
Rentang fitur	*Unigram* dan *bigram*
Model klasifikasi	SVC
Kernel	Linear
Parameter (C)	1
Bobot kelas	balanced
Validasi silang	10 bagian terstratifikasi
Metrik validasi	Nilai F1 tertimbang

8. Evaluasi Model

Evaluasi dilakukan menggunakan *confusion matrix*, akurasi, presisi, *recall*, dan nilai F1. Pada klasifikasi biner, *confusion matrix* terdiri atas *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN).

Akurasi menunjukkan proporsi keseluruhan data yang diprediksi dengan benar dan dihitung menggunakan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Presisi menunjukkan proporsi data yang benar di antara seluruh data yang diprediksi sebagai suatu kelas:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

Recall menunjukkan proporsi data aktual suatu kelas yang dapat dikenali oleh model:

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

Nilai F1 merupakan rata-rata harmonik antara presisi dan *recall*:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

Presisi, *recall*, dan nilai F1 dihitung untuk kelas positif dan negatif. Hasil evaluasi juga disajikan dalam bentuk *macro average* dan *weighted average*. *Macro average* memberikan bobot yang sama pada setiap kelas, sedangkan *weighted average* memperhitungkan jumlah data pada setiap kelas. Perbedaan kedua jenis agregasi tersebut perlu diperhatikan karena dataset penelitian memiliki distribusi kelas yang tidak seimbang [1], [17]. Metrik berbasis *confusion matrix* memang dapat menghasilkan interpretasi berbeda bergantung pada distribusi dan jenis tugas klasifikasi.

Akurasi tidak digunakan sebagai satu-satunya dasar untuk menilai model. Kemampuan mendeteksi sentimen negatif dianalisis secara khusus menggunakan presisi, *recall*, dan nilai F1 kelas negatif.

9. Perangkat Penelitian

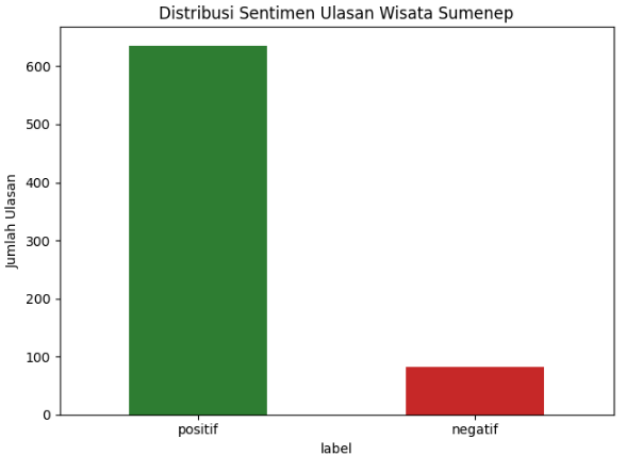
Pengolahan data dan pembangunan model dilakukan menggunakan bahasa pemrograman Python pada lingkungan Google Colaboratory. Pustaka Pandas digunakan untuk membaca, membersihkan, dan mengelola dataset, sedangkan Sastrawi digunakan untuk penghapusan *stopword* dan *stemming* bahasa Indonesia. Pustaka Scikit-learn digunakan untuk pembentukan fitur TF-IDF, pembagian dataset, pelatihan model SVM, validasi silang, dan evaluasi model [18].

10. Pertimbangan Etis

Data penelitian diperoleh dari ulasan yang tersedia secara publik pada Google Maps. Identitas pengulas tidak digunakan sebagai fitur klasifikasi dan tidak ditampilkan pada tabel, gambar, maupun pembahasan hasil. Atribut yang digunakan hanya meliputi nama destinasi, teks ulasan, dan rating bintang yang berkaitan langsung dengan tujuan penelitian. Contoh ulasan yang ditampilkan dalam artikel juga tidak disertai nama akun untuk mengurangi risiko pengungkapan identitas pengguna. Penggunaan data dibatasi untuk kepentingan analisis akademik dan tidak diarahkan untuk menilai individu yang memberikan ulasan.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil pengolahan data ulasan wisata Sumenep yang diperoleh dari Google Maps menggunakan pembobotan TF-IDF dan klasifikasi Support Vector Machine. Hasil yang ditampilkan meliputi distribusi data sentimen, kinerja model berdasarkan confusion matrix dan metrik evaluasi, serta pembahasan terhadap kemampuan model dalam mengklasifikasikan ulasan positif dan negatif. Pembahasan difokuskan pada interpretasi performa model dan pengaruh ketidakseimbangan data terhadap hasil klasifikasi. Distribusi Data Sentimen Setelah proses seleksi dan pelabelan, diperoleh 725 ulasan yang terdiri atas 643 ulasan positif dan 82 ulasan negatif. Kelas positif mencakup 88,69% dari keseluruhan data, sedangkan kelas negatif hanya mencakup 11,31%. Distribusi tersebut menunjukkan bahwa jumlah ulasan positif hampir delapan kali lebih banyak dibandingkan ulasan negatif.



Gambar 3. Distribusi Sentimen Ulasan Wisata Sumenep

Dominasi kelas positif menunjukkan bahwa sebagian besar pengunjung memberikan *rating* empat atau lima terhadap destinasi yang diulas. Namun, distribusi yang tidak seimbang dapat menyebabkan model lebih banyak mempelajari pola bahasa kelas positif dan kurang sensitif dalam mengenali kelas negatif [13]. Oleh karena itu, hasil model tidak hanya dinilai berdasarkan akurasi, tetapi juga berdasarkan presisi, *recall*, dan nilai F1 pada setiap kelas.

1. Hasil Evaluasi Model

Model SVM diuji menggunakan data uji yang berasal dari pembagian dataset dengan rasio 80:20. Hasil evaluasi menunjukkan bahwa model memperoleh akurasi sebesar 87,50%. Presisi kelas positif mencapai 90,44%, sedangkan *recall* kelas positif mencapai 96,09%. Nilai F1 kelas positif sebesar 93,18%, menunjukkan bahwa model memiliki kemampuan yang tinggi dalam mengenali ulasan positif.

Pada evaluasi keseluruhan, model memperoleh nilai F1 tertimbang sebesar 85,61%. Namun, nilai F1 makro hanya mencapai 59,09%. Perbedaan tersebut menunjukkan bahwa performa model belum merata pada kedua kelas. Nilai F1 tertimbang lebih banyak dipengaruhi oleh kelas positif karena jumlah datanya lebih besar, sedangkan nilai F1 makro memberikan bobot yang sama terhadap kelas positif dan negatif.

Tabel 6. Hasil Evaluasi Model pada Data Uji

Metrik Evaluasi	Nilai
Akurasi	87,50%
Presisi kelas positif	90,44%
Recall kelas positif	96,09%
F1 kelas positif	93,18%
Presisi kelas negatif	37,50%
Recall kelas negatif	18,75%
F1 kelas negatif	25,00%
F1 makro	59,09%
F1 tertimbang	85,61%
Balanced accuracy	57,42%

Hasil evaluasi setiap kelas disajikan pada Tabel 7.

Tabel 7. Hasil Evaluasi Model pada Setiap Kelas

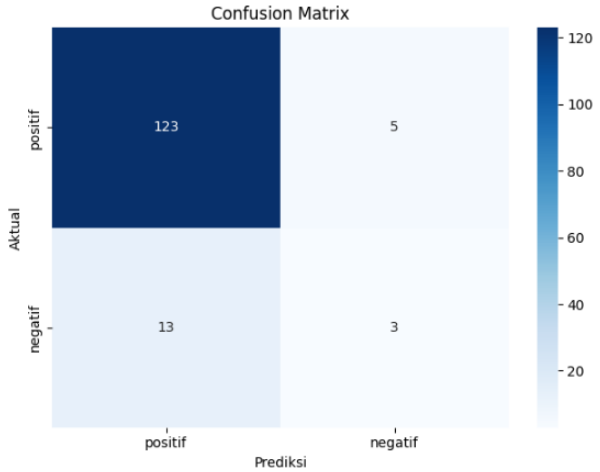
Kelas	Presisi	*Recall*	Nilai F1	*Support*
Negatif	0,38	0,19	0,25	16
Positif	0,9	0,96	0,93	128
Rata-rata makro	0,64	0,57	0,59	144
Rata-rata tertimbang	0,85	0,88	0,86	144

Tabel 7 memperlihatkan adanya perbedaan performa yang cukup besar antara kedua kelas. Model memperoleh nilai F1 sebesar 0,93 pada kelas positif, tetapi hanya 0,25 pada kelas negatif. *Recall* kelas negatif sebesar 0,19 menunjukkan bahwa sebagian besar ulasan negatif belum dapat dikenali dengan benar.

Selain pengujian menggunakan data uji, model dievaluasi menggunakan validasi silang 10 bagian dengan metrik *weighted F1-score*. Hasil pengujian awal menghasilkan rata-rata nilai F1 tertimbang sebesar 86,38% dengan simpangan baku 2,98%. Nilai tersebut menunjukkan bahwa performa tertimbang model relatif tidak banyak berubah pada pembagian data yang berbeda.

2. *Analisis Confusion Matrix*

Hasil prediksi model terhadap data uji divisualisasikan menggunakan *confusion matrix* pada Gambar 4.



Gambar 4. Confusion Matrix Hasil Klasifikasi

Berdasarkan *confusion matrix*, sebanyak 123 dari 128 ulasan positif berhasil diklasifikasikan sebagai positif, sedangkan lima ulasan positif diprediksi sebagai negatif. Pada kelas negatif, hanya tiga dari 16 ulasan yang berhasil dikenali sebagai negatif. Sebanyak 13 ulasan negatif lainnya salah diklasifikasikan sebagai positif.

Kesalahan prediksi lebih banyak terjadi ketika model menghadapi ulasan negatif. Temuan tersebut sejalan dengan rendahnya *recall* kelas negatif sebesar 18,75%. Meskipun parameter class_weight="balanced" telah digunakan, jumlah data negatif yang jauh lebih sedikit menyebabkan pola bahasa pada kelas tersebut belum terwakili secara memadai selama proses pelatihan.

3. *Analisis Kesalahan Klasifikasi*

Matriks konfusi memperlihatkan 18 kesalahan prediksi, yang terdiri atas 13 ulasan negatif yang diprediksi sebagai positif dan lima ulasan positif yang diprediksi sebagai negatif. Kesalahan paling banyak terjadi pada kelas negatif. Dari 16 ulasan negatif pada data uji, hanya tiga ulasan yang berhasil dikenali sesuai label aktual.

Kesalahan tersebut dapat dipengaruhi oleh keterbatasan jumlah ulasan negatif, penggunaan kalimat dengan opini campuran, dan pelabelan berdasarkan rating. Sebuah ulasan dapat memberikan rating tinggi tetapi tetap memuat kritik terhadap fasilitas atau pelayanan. Sebaliknya, kata yang tampak negatif tidak selalu menunjukkan sentimen negatif apabila digunakan dalam bentuk negasi atau ungkapan tertentu.

Penggunaan *unigram* dan *bigram* membantu model mengenali kata tunggal serta pasangan kata berurutan, tetapi belum sepenuhnya menangkap hubungan makna dalam kalimat yang panjang. Kondisi ini dapat menyebabkan ulasan yang mengandung pujian dan keluhan secara bersamaan lebih sulit diklasifikasikan. Analisis kesalahan tersebut memperlihatkan bahwa kualitas representasi kelas negatif masih perlu ditingkatkan.

Tabel 8. Kategori Kesalahan Klasifikasi

Jenis Kesalahan	Jumlah
Ulasan positif diprediksi negatif	5
Ulasan negatif diprediksi positif	13
Total kesalahan	**18**

4. *Pembahasan*

Hasil pengujian menunjukkan bahwa SVM dengan pembobotan TF-IDF mampu mengenali ulasan positif dengan baik. Hal ini ditunjukkan oleh *recall* kelas positif sebesar 96,09% dan nilai F1 sebesar 93,18%. Akurasi model sebesar 87,50% berada pada kisaran hasil beberapa penelitian sentimen wisata sebelumnya yang melaporkan akurasi antara 81% dan 84,01% [2],[3],[8]. Perbandingan langsung perlu dilakukan secara hati-hati karena masing-masing penelitian menggunakan sumber data, distribusi kelas, metode pelabelan, dan skenario evaluasi yang berbeda.

Kemampuan model dalam mengenali ulasan negatif masih rendah. Dari 16 ulasan negatif pada data uji, hanya tiga ulasan yang berhasil diklasifikasikan dengan benar. Kondisi tersebut menghasilkan *recall* sebesar 18,75% dan nilai F1 sebesar 25% pada kelas negatif. Temuan ini menunjukkan bahwa akurasi keseluruhan belum cukup untuk menggambarkan kemampuan model pada dataset yang tidak seimbang.

Perbedaan antara nilai F1 tertimbang sebesar 85,61% dan nilai F1 makro sebesar 59,09% memperlihatkan bahwa kinerja model

masih didominasi oleh kelas positif. Nilai tertimbang dipengaruhi oleh jumlah data pada setiap kelas, sedangkan nilai makro memberikan bobot yang sama kepada kelas positif dan negatif. Oleh karena itu, evaluasi per kelas lebih tepat digunakan untuk menilai kemampuan model dalam mendeteksi kelas minoritas [13].

Penggunaan class_weight="balanced" belum sepenuhnya mengatasi ketidakseimbangan data. Parameter tersebut meningkatkan penalti terhadap kesalahan pada kelas negatif, tetapi tidak menambah jumlah dan variasi ulasan negatif yang dipelajari model. Penelitian terdahulu menunjukkan bahwa teknik penyeimbangan seperti SMOTE dan Tomek Links dapat meningkatkan kinerja SVM pada data tidak seimbang [14]. Namun, penerapannya pada data teks perlu dilakukan hanya terhadap data latih untuk mencegah kebocoran data.

Rendahnya performa kelas negatif juga dapat dipengaruhi oleh pelabelan berbasis rating. Rating bintang tidak selalu mewakili seluruh isi ulasan karena satu teks dapat memuat pujian dan keluhan sekaligus. Dengan demikian, model dapat digunakan sebagai alat penyaring awal untuk mengelompokkan ulasan positif dalam jumlah besar. Namun, prediksi terhadap kelas negatif masih memerlukan pemeriksaan manual karena sebagian besar ulasan negatif belum dikenali dengan benar.

Dari sisi penerapan, model dapat digunakan sebagai alat penyaring awal untuk mengelompokkan ulasan dalam jumlah besar. Namun, hasil prediksi tidak sebaiknya digunakan sebagai satu-satunya dasar penilaian terhadap suatu destinasi. Ulasan yang diprediksi negatif perlu diperiksa kembali karena model masih memiliki keterbatasan dalam mengenali kelas tersebut.

Informasi sentimen juga dapat membantu pengelola destinasi mengidentifikasi ulasan yang memerlukan perhatian lebih lanjut. Meskipun demikian, penelitian ini baru mengklasifikasikan polaritas umum dan belum memisahkan aspek kebersihan, fasilitas, aksesibilitas, harga, pelayanan, dan kondisi lingkungan. Pengembangan analisis berbasis aspek akan memberikan informasi yang lebih spesifik bagi pengelola wisata.

5. Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan. Data hanya diperoleh dari Google Maps dan mencakup enam destinasi wisata di Kabupaten Sumenep sehingga hasil penelitian belum dapat digeneralisasikan untuk seluruh destinasi wisata. Pelabelan berdasarkan rating juga berpotensi menimbulkan ketidaksesuaian antara label dan isi teks.

Distribusi data yang didominasi oleh ulasan positif menyebabkan kemampuan model dalam mengenali sentimen negatif masih rendah. Penelitian ini juga belum melakukan analisis berbasis aspek sehingga belum dapat mengidentifikasi secara khusus sentimen wisatawan terhadap fasilitas, kebersihan, aksesibilitas, harga, pelayanan, atau kondisi lingkungan destinasi.

4. KESIMPULAN

Penelitian ini menerapkan pembobotan *Term Frequency–Inverse Document Frequency* dan algoritma *Support Vector Machine* dengan kernel linear untuk mengklasifikasikan sentimen ulasan enam destinasi wisata di Kabupaten Sumenep. Dari 725 ulasan yang digunakan, sebanyak 643 ulasan termasuk kelas positif dan 82 ulasan termasuk kelas negatif. Model dengan parameter regularisasi $C = 1{,}0$dan bobot kelas balanced menghasilkan akurasi sebesar 87,50%, nilai F1 tertimbang 85,61%, dan nilai F1 makro 59,09%. Pada evaluasi per kelas, model memperoleh *recall* 96,09% dan nilai F1 93,18% untuk kelas positif, sedangkan *recall* kelas negatif hanya mencapai 18,75% dengan nilai F1 sebesar 25%.

Hasil tersebut menunjukkan bahwa model lebih mampu mengenali ulasan positif daripada ulasan negatif. Penggunaan bobot kelas seimbang belum sepenuhnya mengatasi pengaruh ketimpangan jumlah data karena sebagian besar ulasan negatif masih diprediksi sebagai positif. Dengan demikian, akurasi keseluruhan tidak dapat digunakan sebagai satu-satunya dasar untuk menilai kinerja model. Pengembangan penelitian berikutnya perlu diarahkan pada penambahan dan pemeriksaan data kelas negatif, validasi kesesuaian label rating dengan isi ulasan, serta pengujian metode penanganan ketidakseimbangan kelas yang diterapkan hanya pada data latih.

5. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Informatika, Fakultas Sains dan Kesehatan, Universitas PGRI Sumenep, yang telah memberikan dukungan fasilitas dan sumber daya selama pelaksanaan penelitian ini. Terima kasih juga kepada seluruh rekan sejawat di lingkungan Universitas PGRI Sumenep atas masukan dan diskusi yang bermanfaat dalam proses penyelesaian artikel ini, Penelitian ini tidak menerima pendanaan eksternal.

REFERENCES

[1] Pemerintah Kabupaten Sumenep, “Target PAD Disbudporapar 2024 Naik Menjadi Rp847 juta | Kabupaten Sumenep,” Media Center Kabupaten Sumenep. Accessed: Aug. 05, 2026. [Online]. Available: https://www.sumenepkab.go.id/berita/baca/target-pad-disbudporapar-2024-naik-menjadi-rp847-juta

[2] M. S. Syahlan, D. Irmayanti, and S. Alam, “Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Support Vector Machine (Svm),” *Simtek: jurnal sistem informasi dan teknik komputer*, vol. 8, no. 2, pp. 315–319, 2023.

[3] J. Ipmawati, S. Saifulloh, and K. Kusnawi, “Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine,” *MALCOM: Indonesian Journal of Machine*

Learning and Computer Science, vol. 4, no. 1, pp. 247–256, Jan. 2024, doi: 10.57152/malcom.v4i1.1066.

[4] B. Liu, "Sentiment analysis and opinion mining," 2012.

[5] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Inf. Process. Manag.*, vol. 24, no. 5, pp. 513–523, 1988.

[6] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[7] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *European conference on machine learning*, 1998, pp. 137–142.

[8] I. W. B. Suryawan, N. W. Utami, and K. Q. Fredlina, "Analisis sentimen review wisatawan pada objek wisata ubud menggunakan algoritma support vector machine," *Jurnal Informatika Teknologi dan Sains (Jinteks)*, vol. 5, no. 1, pp. 133–140, 2023.

[9] D. A. Fatah, "Sentiment Analysis of Madura Tourism Opinion Using Support Vector Machine (SVM)," *Technium*, vol. 16, no. 1, pp. 243–249, 2023.

[10] R. Ramdani and R. Cahyana, "Analisis Sentimen Ulasan Wisata Budaya Menggunakan Metode Support Vector Machine dan Long Short-Term Memory," *Jurnal Algoritma*, vol. 22, no. 2, pp. 1188–1199, 2025.

[11] N. L. P. H. Nirmala and I. B. G. Dwidasmara, "Analisis Sentimen Ulasan Aplikasi M-Paspor Menggunakan TF-IDF dan Support Vector Machine," *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, vol. 3, no. 2, pp. 431–438, 2025.

[12] D. B. M. Zebua and I. B. G. Dwidasmara, "Analisis Sentimen Ulasan Aplikasi Citilink Menggunakan Metode Support Vector Machine dengan TF-IDF," *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, vol. 3, no. 2, pp. 439–446, 2025.

[13] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009.

[14] D. I. Sumantiawan, J. E. Suseno, and W. A. Syafei, "Sentiment analysis of customer reviews using support vector machine and smote-tomek links for identify customer satisfaction," *Jurnal Sistem Informasi Bisnis*, vol. 13, no. 1, pp. 1–9, 2023.

[15] M. Adriani, J. Asian, B. Nazief, S. M. M. Tahaghoghi, and H. E. Williams, "Stemming Indonesian: A confix-stripping approach," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 6, no. 4, pp. 1–33, 2007.

[16] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.

[17] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.