

ANALISIS SENTIMEN TERHADAP TEMPAT WISATA BERDASARKAN ULASAN PADA GOOGLE MAPS MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE (Studi Kasus: Kabupaten Brebes)

1st Dimas Teguh Ramadhani
Fakultas Informatika
Telkom University Purwokerto
Purwokerto, Indonesia
dimasteguhramadhani@telkomuniversity.ac.id

2nd Sena Wijayanto, S.Pd., M.T.
Fakultas Rekayasa Industri
Telkom University Purwokerto
Purwokerto, Indonesia
senawijayanto@telkomuniversity.ac.id

3rd Dedy Agung Prabowo, S.Kom., M.Kom
Fakultas Informatika
Telkom University Purwokerto
Purwokerto, Indonesia
dedyaprabowo@telkomuniversity.ac.id

Abstrak — Berwisata adalah salah satu kegiatan rekreasi yang populer di kalangan masyarakat untuk melepas penat dan menikmati keindahan alam atau budaya setempat. Kabupaten Brebes, terbesar ketiga di Jawa Tengah setelah Grobogan, namun menempati urutan terakhir dalam jumlah kunjungan wisatawan pada tahun 2022 menurut data Badan Pusat Statistik (BPS) Jawa Tengah. Kondisi ini mendorong perlunya analisis lebih lanjut mengenai sentimen pengunjung terhadap tempat wisata di Brebes untuk memahami persepsi publik. Penelitian ini bertujuan mengklasifikasikan sentimen pengunjung dari Google Maps dan membandingkan kinerja algoritma Naïve Bayes (NB) dengan Support Vector Machine (SVM). Dari 6734 data ulasan yang dianalisis menggunakan fitur TF-IDF, SVM menunjukkan performa yang lebih unggul dengan akurasi 87%, melampaui Naïve Bayes dengan 86%. Pada pengujian lebih lanjut, SVM menghasilkan nilai precision sebesar 0.69, recall 0.24, dan f1-score 0.36 untuk sentimen negatif, serta precision 0.88, recall 0.98, dan f1-score 0.93 untuk sentimen positif. Sementara itu, Naïve Bayes menghasilkan precision 0.86, recall 0.05, dan f1-score 0.10 untuk sentimen negatif, serta precision 0.86, recall 1.00, dan f1-score 0.92 untuk sentimen positif. Hal ini menunjukkan bahwa SVM tidak hanya lebih akurat secara keseluruhan, tetapi juga lebih seimbang dalam mengenali kedua kelas sentimen, terutama dalam mendeteksi ulasan negatif yang sangat penting sebagai masukan bagi pengembangan sektor pariwisata.

Kata kunci— analisis sentimen, brebes, naïve bayes, support vector machine, wisata

I. PENDAHULUAN

Kabupaten Brebes, yang merupakan kabupaten terluas ketiga di Provinsi Jawa Tengah, memiliki potensi pariwisata yang besar berkat keragaman geografisnya yang mencakup dataran tinggi dan rendah[1]. Namun, data dari Badan Pusat Statistik (BPS) menunjukkan ironi, di mana pada tahun 2022, Brebes menempati urutan terakhir dalam jumlah kunjungan wisatawan di Jawa Tengah dengan hanya 80.938 pengunjung[2]. Angka ini terpaut jauh dari kabupaten lain, yang menandakan adanya tantangan dalam menarik minat wisatawan. Salah satu faktor yang diduga berpengaruh adalah persepsi publik yang tercermin dalam ulasan online, di mana

keluhan mengenai akses, fasilitas, dan pelayanan sering muncul [3].

Di era digital, ulasan di platform seperti Google Maps telah menjadi sumber informasi krusial bagi calon wisatawan dan cerminan langsung dari pengalaman pengunjung [4]. Oleh karena itu, analisis terhadap ulasan ini dapat memberikan wawasan berharga untuk pemangku kepentingan untuk melakukan perbaikan dan evaluasi. Sebagai bagian dari text mining, analisis sentimen, memungkinkan klasifikasi opini dalam teks ke dalam kategori positif atau negatif secara otomatis.

Beberapa penelitian sebelumnya telah berhasil menerapkan metode machine learning untuk analisis sentimen di berbagai domain. Penelitian oleh Ratino, dkk. (2020) [5] dan Pamungkas & Kharisudin (2021) [6] menunjukkan bahwa Support Vector Machine (SVM) sering kali lebih unggul daripada Naïve Bayes (NB) dalam klasifikasi teks. Berdasarkan hal tersebut, penelitian ini bertujuan untuk (1) mengklasifikasikan sentimen masyarakat terhadap tempat wisata di Kabupaten Brebes berdasarkan ulasan di Google Maps, dan (2) membandingkan kinerja algoritma Naïve Bayes Classifier dan Support Vector Machine dengan ekstraksi fitur TF-IDF untuk menemukan model terbaik pada studi kasus ini. Hasil penelitian diharapkan dapat menjadi masukan strategis bagi Dinas Pariwisata Kabupaten Brebes untuk meningkatkan kualitas dan daya tarik destinasi wisatanya.

II. KAJIAN TEORI

A. Analisis Sentimen

Analisis Sentimen adalah Metode komputasi untuk menemukan dan mendapatkan sudut pandang atau emosi yang disampaikan dalam sebuah teks disebut analisis sentimen. Tujuannya adalah untuk mengetahui apakah penulis memiliki sikap positif, negatif, atau netral terhadap suatu subjek, barang, atau layanan.. Dalam konteks penelitian ini, analisis sentimen digunakan untuk mengukur persepsi publik terhadap destinasi wisata di Kabupaten Brebes berdasarkan ulasan yang mereka berikan, sehingga dapat

menjadi acuan untuk evaluasi dan peningkatan kualitas layanan pariwisata [7].

B. Naïve Bayes Classifier

Naïve Bayes Classifier (NBC), yang didasarkan pada Teorema Bayes, adalah teknik klasifikasi probabilistik yang bekerja dengan asumsi "naif" bahwa setiap fitur (dalam kasus ini, setiap kata dalam ulasan) bersifat independen satu sama lain [8]. Meskipun asumsi ini seringkali tidak sepenuhnya benar dalam data dunia nyata, NBC dikenal sangat efisien, cepat, dan seringkali memberikan hasil akurasi yang tinggi, terutama dalam tugas klasifikasi teks. Probabilitas sebuah ulasan masuk ke dalam kelas sentimen tertentu (misalnya, positif) dihitung berdasarkan probabilitas kemunculan kata-kata di dalam ulasan tersebut pada kelas itu.

C. Support Vector Machine

Support Vector Machine (SVM) adalah algoritma supervised learning yang sangat efektif untuk tugas klasifikasi. Tujuan utama SVM adalah menemukan garis pemisah atau hyperplane optimal yang dapat memisahkan titik-titik data dari kelas yang berbeda dengan margin (jarak) semaksimal mungkin [6]. Titik data yang paling dekat dengan hyperplane dan berperan dalam menentukan posisinya disebut support vectors. SVM sangat andal dalam menangani data berdimensi tinggi, seperti data teks yang telah diubah menjadi vektor numerik, dan seringkali menghasilkan akurasi yang sangat baik dalam analisis sentimen [9].

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

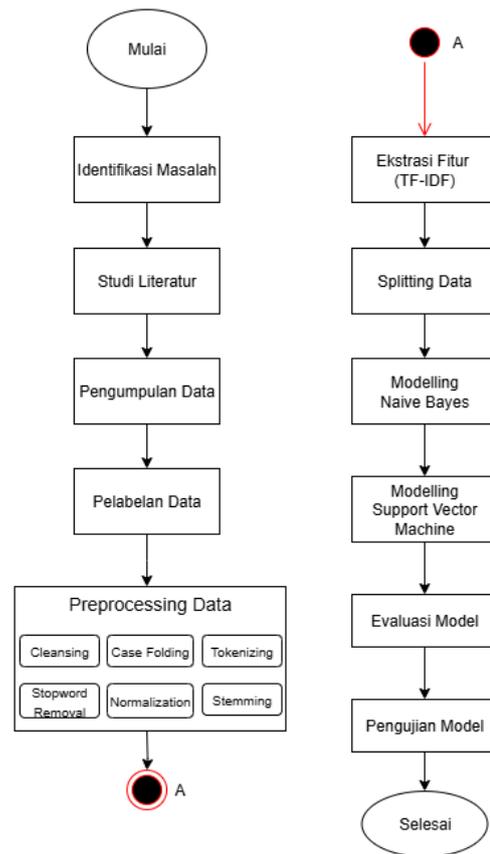
TF-IDF adalah metode pembobotan statistik yang digunakan untuk mengukur tingkat kepentingan sebuah kata dalam suatu dokumen dibandingkan dengan sekumpulan dokumen (korpus) [25]. Metode ini terdiri dari dua komponen: Frekuensi Kata (TF), yang menghitung frekuensi kata muncul dalam satu dokumen, dan Frekuensi Dokumen Inverse (IDF), yang menghitung seberapa unik kata-kata tersebut di seluruh korpus. Kata-kata yang sering muncul dalam satu dokumen tetapi jarang muncul [10].

E. Confusion Matrix

Confusion Matrix digunakan untuk mengevaluasi kinerja model klasifikasi. Ini menyajikan ringkasan hasil prediksi dengan membandingkan kelas aktual dengan kelas yang diprediksi model. Dengan membagi kesalahan yang dibuat oleh model ke dalam empat kategori, confusion matrix memberikan gambaran rinci tentang kesalahan yang dibuat oleh model. True Positive (TP) menunjukkan data positif yang diprediksi benar; True Negative (TN) menunjukkan data negatif yang diprediksi benar; False Positive (FP) menunjukkan data negatif yang salah diprediksi sebagai positif; dan False Negative (FN) menunjukkan data positif yang salah diprediksi sebagai negatif. Untuk mengukur kinerja model secara menyeluruh, metrik evaluasi Accuracy, Precision, Recall, dan F1-Score dapat dihitung dari keempat nilai ini.

III. METODE

Rancangan penelitian ini menggunakan pendekatan kuantitatif dengan beberapa tahapan sistematis yang diuraikan sebagai berikut.



Gambar 1 Diagram Penelitian

A. Pengumpulan & Pelabelan Data

Data penelitian berupa ulasan (review) pengunjung tempat wisata di Kabupaten Brebes yang dikumpulkan dari platform Google Maps menggunakan teknik scraping dengan kata kunci "Wisata Brebes". Sebanyak 6734 ulasan berhasil dikumpulkan. Selanjutnya, dilakukan pelabelan data secara otomatis berdasarkan rating bintang yang diberikan pengguna. Ulasan dengan rating 1-2 bintang dilabeli sebagai sentimen 'Negatif' (nilai 1), dan ulasan dengan rating 3-5 bintang dilabeli sebagai 'Positif' (nilai 0).

B. Preprocessing Data

Untuk menyiapkan data teks agar dapat diolah oleh model, dilakukan serangkaian tahap preprocessing, yaitu:

- Cleansing: Membersihkan teks dari karakter yang tidak relevan seperti URL, mention, tanda baca, angka, dan emoji.
- Case Folding: Mengubah seluruh teks menjadi huruf kecil (lowercase).
- Tokenizing: Memecah kalimat menjadi potongan kata-kata individual (token).
- Normalization: Mengubah kata-kata tidak baku atau singkatan menjadi bentuk bakunya sesuai kamus yang telah dibuat.
- Stopword Removal: Menghapus kata-kata umum yang tidak memiliki makna sentimen (misalnya: "dan", "di", "yang").
- Stemming: Mengubah kata-kata berimbuhan menjadi kata dasarnya menggunakan library Sastrawi.

C. TF-IDF

Setelah data bersih, dilakukan proses ekstraksi fitur untuk mengubah data teks menjadi representasi numerik. Metode yang digunakan adalah Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF bekerja dengan memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam sebuah ulasan (TF) dan frekuensi kemunculannya di seluruh kumpulan ulasan (IDF). Kata yang sering muncul di satu ulasan tetapi jarang muncul di ulasan lain akan mendapatkan bobot yang tinggi, menandakan bahwa kata tersebut signifikan.

D. Pemodelan dan Evaluasi

Data yang telah diubah menjadi vektor TF-IDF dibagi menjadi data latih (80%) dan data uji (20%). penulis juga melakukan serangkaian eksperimen dengan beberapa rasio pembagian data lainnya, yaitu 50:50, 60:40, dan 70:30, meskipun hasil utama yang dilaporkan dalam penelitian ini tetap mengacu pada rasio standar 80:20 yang umum digunakan dalam berbagai penelitian sejenis.

Dua model klasifikasi dibangun dan dibandingkan :

- Naïve Bayes Classifier (NBC) Model probabilistik yang mengasumsikan independensi antar fitur.
- Support Vector Machine (SVM) Model yang mencari hyperplane terbaik untuk memisahkan dua kelas sentimen. Pada penelitian ini, digunakan kernel linear.

Kinerja kedua model dievaluasi menggunakan Confusion Matrix dengan metrik Accuracy, Precision, Recall, dan F1-Score untuk menentukan model mana yang paling efektif dalam mengklasifikasikan sentimen.

IV. HASIL DAN PEMBAHASAN

A. Skenario Pengujian Model

Untuk menguji ketahanan dan konsistensi performa, model dievaluasi menggunakan empat skenario pembagian data yang berbeda: 50:50, 60:40, 70:30, dan 80:20 (data latih:data uji). Hasil dari setiap skenario disajikan pada tabel-tabel berikut.

Tabel 1. Hasil Performa Model Skenario 50:50 NB

Label	Evaluation Matrix					Akurasi
	Precision	Recall	F1 Score	Support	Akurasi	
Negatif	0,96	0,09	0,16	264	0,87	
Positif	0,87	1	0,93	1568		

Tabel 2. Hasil Performa Model Skenario 50:50 SVM

Label	Evaluation Matrix					Akurasi
	Precision	Recall	F1 Score	Support	Akurasi	
Negatif	0,74	0,26	0,39	264	0,88	
Positif	0,89	0,98	0,93	1568		

Tabel 3. Hasil Performa Model Skenario 60:40 NB

Label	Evaluation Matrix					Akurasi
	Precision	Recall	F1 Score	Support	Akurasi	
Negatif	0,95	0,09	0,17	215	0,87	
Positif	0,87	1	0,93	1251		

Tabel 4. Hasil Performa Model Skenario 60:40 SVM

Label	Evaluation Matrix					Akurasi
	Precision	Recall	F1 Score	Support	Akurasi	
Negatif	0,74	0,25	0,37	215	0,88	
Positif	0,88	0,98	0,93	1251		

Tabel 5. Hasil Performa Model Skenario 70:30 NB

Label	Evaluation Matrix					Akurasi
	Precision	Recall	F1 Score	Support	Akurasi	
Negatif	0,93	0,09	0,16	164	0,86	
Positif	0,86	1	0,93	935		

Tabel 6. Hasil Performa Model Skenario 70:30 SVM

Label	Evaluation Matrix					Akurasi
	Precision	Recall	F1 Score	Support	Akurasi	
Negatif	0,65	0,26	0,37	164	0,87	
Positif	0,88	0,98	0,93	935		

Tabel 7. Hasil Performa Model Skenario 80:20 NB

Label	Evaluation Matrix					Akurasi
	Precision	Recall	F1 Score	Support	Akurasi	
Negatif	0,86	0,05	0,10	111	0,86	
Positif	0,86	1	0,92	622		

Tabel 8. Hasil Performa Model Skenario 80:20 SVM

Label	Evaluation Matrix					Akurasi
	Precision	Recall	F1 Score	Support	Akurasi	
Negatif	0,69	0,24	0,36	111	0,87	
Positif	0,88	0,98	0,93	622		

Dari keempat skenario, terlihat bahwa SVM secara konsisten memberikan akurasi yang sedikit lebih tinggi dan metrik yang lebih seimbang. Skenario 80:20, yang merupakan standar umum, menunjukkan bahwa meskipun akurasi SVM (87%) hanya 1% lebih tinggi dari NB (86%), kemampuan SVM dalam mengenali ulasan negatif jauh lebih baik. Hal ini terlihat dari nilai recall (0.24) dan F1-Score (0.36) untuk kelas negatif pada SVM, yang secara signifikan lebih tinggi daripada NB (recall 0.05, F1-Score 0.10). Nilai recall yang sangat rendah pada NB menunjukkan bahwa model ini hampir gagal total dalam mengidentifikasi ulasan negatif, yang merupakan informasi krusial untuk perbaikan.

B. Hasil Uji dengan Data Baru

Untuk menguji kemampuan praktis model, dilakukan prediksi pada beberapa ulasan baru yang tidak ada dalam dataset. Hasilnya disajikan pada Tabel 9.

Tabel 9. Hasil Prediksi Model pada Data Baru

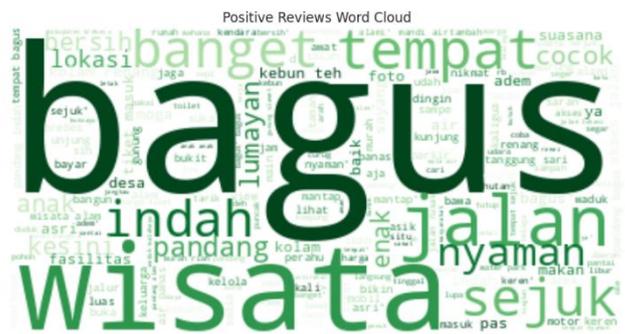
Review	Naive Bayes Prediction	SVM Prediction
Pemandangan sangat indah, saya suka sekali!	Positive	Positive
Tempatnya kotor dan tidak terawat.	Positive	Negative
Pelayanan ramah, tapi fasilitas kurang.	Positive	Negative
Sangat mengecewakan, tidak seperti yang diharapkan.	Negative	Negative
Akses jalan sulit, tapi setelah sampai	Positive	Positive

Review	Naive Bayes Prediction	SVM Prediction
pemandangannya worth it.		
Harga tiket terlalu mahal.	Positive	Negative
Cocok untuk liburan keluarga, banyak spot foto menarik.	Positive	Positive
Tidak ada tempat sampah, banyak sampah berserakan.	Positive	Negative
Pengalaman menyenangkan, pasti akan kembali lagi.	Positive	Positive
Cuaca buruk saat berkunjung, jadi kurang maksimal.	Positive	Positive
Pemandangan di pantai sangat indah, sangat direkomendasikan.	Positive	Positive
Toilet di tempat wisata kotor dan bau, sangat mengecewakan.	Negative	Negative
Makanan di sekitar wisata enak dan murah.	Positive	Positive
Transportasi menuju lokasi sulit ditemukan.	Positive	Negative
Staf di tempat wisata sangat ramah dan membantu.	Positive	Positive
Harga tiket masuk terlalu mahal untuk fasilitas yang ada.	Positive	Negative
Suasana di tempat ini sangat tenang dan cocok untuk relaksasi.	Positive	Positive
Banyak sampah berserakan, kebersihan kurang terjaga.	Positive	Negative
Aktivitas di tempat ini sangat menyenangkan untuk keluarga.	Positive	Positive
Informasi tentang wisata kurang jelas.	Positive	Positive

Tabel 9 mengonfirmasi keunggulan SVM. Pada ulasan yang mengandung kata-kata negatif seperti "kotor", "kurang", dan "mahal", SVM berhasil mengklasifikasikannya sebagai negatif, sementara Naïve Bayes gagal dan salah mengklasifikasikannya sebagai positif. Ini menunjukkan sensitivitas SVM yang lebih baik terhadap kata-kata kunci negatif.

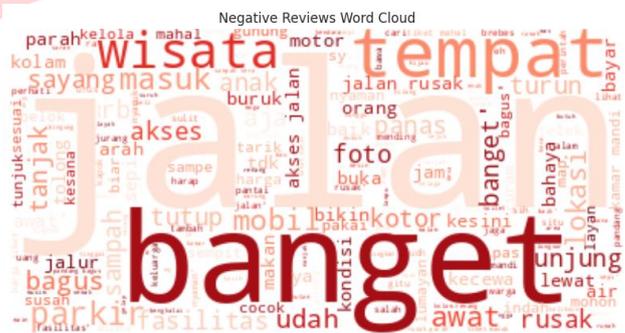
C. Analisis Konten Ulasan

Untuk memahami tema utama dari ulasan positif dan negatif, dibuat visualisasi word cloud.



Gambar 2 Word Cloud Sentimen Positif

Gambar 2 menunjukkan bahwa sentimen positif didominasi oleh kata-kata seperti "bagus", "indah", "tempat", "nyaman", "sejuk", dan "wisata". Hal ini menandakan apresiasi pengunjung terhadap keindahan alam dan suasana yang ditawarkan oleh destinasi wisata di Brebes.



Gambar 3 Word Cloud Sentimen Negatif

Sebaliknya, pada Gambar 3, sentimen negatif didominasi oleh kata-kata seperti "jalan", "rusak", "akses", "mahal", "kotor", dan "banget". Ini mengindikasikan bahwa isu utama yang dikeluhkan pengunjung adalah infrastruktur (akses jalan yang rusak), kebersihan yang kurang terjaga, dan harga tiket yang dianggap tidak sepadan dengan fasilitas yang didapat.

V. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Naïve Bayes dan Support Vector Machine untuk menganalisis sentimen ulasan pengunjung tempat wisata di Kabupaten Brebes dari Google Maps. Dari hasil perbandingan, disimpulkan bahwa algoritma Support Vector Machine (SVM) memiliki kinerja yang lebih unggul dengan akurasi 87% dan kemampuan yang lebih seimbang dalam mengenali sentimen positif maupun negatif dibandingkan Naïve Bayes. Analisis konten menunjukkan bahwa meskipun secara umum sentimen masyarakat cenderung positif, terutama mengapresiasi keindahan alam, terdapat keluhan signifikan terkait infrastruktur akses jalan, kebersihan, dan harga, yang menjadi masukan penting bagi pemerintah daerah dan pengelola wisata untuk perbaikan di masa mendatang.

REFERENSI

- [1] Faraz Sukma Halida, "Analisis Strategi Meningkatkan Kepuasan Pengunjung Objek Wisata Kebun Teh Kaligua Kecamatan Paguyangan Kabupaten Brebes," Institut Agama Islam Negeri Purwokerto, 2019.
- [2] Badan Pusat Statistik Provinsi Jawa Tengah, "Banyaknya Pengunjung Daya Tarik Wisata dan Event Menurut Kabupaten/Kota di Jawa Tengah, 2022." Accessed: Mar. 18, 2024. [Online]. Available: <https://jateng.bps.go.id/statictable/2023/08/02/2714/banyaknya-pengunjung-daya-tarik-wisata-dan-event-menurut-kabupaten-kota-di-jawa-tengah-2022.html>
- [3] Defri Yusach Ansyari, "Peran Dinas Pariwisata Dalam Menggunakan Media Sosial Untuk Promosi Pariwisata Di Kabupaten Brebes Tahun 2017," Universitas Muhammadiyah Yogyakarta, 2019.
- [4] Lutfi Budi Ilmawan and Muhammad Aliyazid Mude, "Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 154–161, 2020.
- [5] Ratino, Noor Hafidz, Sita Anggraeni, and Windu Gata, "Sentimen Analisis Informasi Covid-19 menggunakan Support Vector Machine dan Naïve Bayes," *Jurnal Penelitian Ilmu dan Teknologi Komputer*, vol. 12, no. 2, pp. 1–11, 2020.
- [6] Fajar Sodik Pamungkas and Iqbal Kharisudin, "Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 4, pp. 628–634, 2021.
- [7] Reza Safitri, Nisa Alfira, Dian Tamitiadini, Nufian S Febriana, and Wayan Weda Asmara Dewi, *Analisis Sentimen : Metode Alternatif Penelitian Big Data*. UB Press, 2021.
- [8] Ali Imron, "Analisis Sentimen Terhadap Tempat Wisata Di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier," Universitas Islam Indonesia, 2019.
- [9] Fajrin Septiana Nugraha, "Peningkatan Akurasi Algoritma Support Vector Machine (Svm) Dengan Particle Swarm Optimization (Pso) Pada Analisis Sentimen Review Hotel Di Tasikmalaya (Studi Kasus: Tiket.Com Dan Agoda)," Institut Teknologi Telkom Purwokerto, 2022.
- [10] Eva Rahma Indriyani, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Terhadap Vaksin Astrazeneca Di Twitter," Institut Teknologi Telkom Purwokerto, 2022.