

Analisa Kinerja Metode Support Vector Machine untuk Analisa Sentimen Ulasan Pengguna Google Maps

Andharini Dwi Cahyani

¹Fakultas Teknik, Teknik Informatika, Universitas Trunojoyo Madura, Bangkalan, Indonesia

Email: ^{1,*}andharini.cahyani@trunojoyo.ac.id

Email Penulis Korespondensi: andharini.cahyani@trunojoyo.ac.id

Abstrak—Analisis Sentimen adalah cabang dari penambangan data yang tujuan dasarnya yaitu untuk menganalisis, memahami, memproses, dan mengekstrak data tekstual dalam bentuk pendapat atau ulasan tentang suatu objek tertentu. Dengan adanya perkembangan teknologi salah satunya Google Maps dapat memberi kemudahan dalam mencari informasi mengenai lokasi suatu objek. Google Maps juga memberikan ulasan terhadap suatu objek. Dari ulasan tersebut didapatkan informasi yang bisa dianalisa sentimen apa saja yang diberikan oleh pengunjung. Untuk pengolahan data teks ulasan pelanggan, digunakan pendekatan penambangan teks dengan metode Support Vector Machine. Metode ini merupakan bagian dari teknik klasifikasi yang dapat mengenali dan mengelompokkan data menjadi 2 bagian yang dipisahkan oleh hyperplane. Data yang digunakan dalam studi kasus ini adalah ulasan pengunjung lokasi Api Tak Kunjung Padam, Sumenep, yang tercatat pada Google Maps Review.

Kata Kunci: Support Vector Machine; Text Mining; Klasifikasi; Analisis Sentimen; Google Maps Review

Abstract—Sentiment Analysis is a subset of data mining that analyzes, understands, processes, and extracts textual data in the form of opinions or reviews about a certain object. The advancement of technology, such as Google Maps, can make it easier to find information on the location of an object. Google Maps also gives an object review. These reviews provide information that can be used to examine the feelings expressed by visitors. A text mining methodology with the Support Vector Machine method is utilized to process customer review text data. This method is part of a classification strategy that recognizes and categorizes data into two portions separated by a hyperplane. The data used in this case study are visitors' reviews that is recorded on Google Maps Review of the Eternal Flame, a tourist attraction object located in Sumenep.

Keywords: Support Vector Machine; Text Mining; Classification; Sentiment Analysis; Google Maps Review

1. PENDAHULUAN

Penambangan data atau data mining merupakan salah satu cabang keilmuan dari ilmu komputer yang bertujuan untuk mendapatkan pengetahuan ataupun informasi dari database. Penambangan data menggunakan teknik statistik, matematika, kecerdasan buatan dan Machine Learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dari pengetahuan yang terkait dari berbagai database besar. Penambangan data ini banyak diterapkan dalam sektor industri, bisnis, wisata, edukasi maupun pengambilan keputusan strategis [1][2]. Pada data mining ada beberapa pendekatan, diantaranya adalah asosiasi [3][4], clustering [5][6] dan klasifikasi [7][8]. Klasifikasi merupakan teknik yang umumnya digunakan untuk menggolongkan atau menentukan suatu kriteria dari suatu data yang didasarkan pada data sebelumnya yang sudah dipelajari [9]. Pada penambangan data, objek yang diolah tidak terbatas pada data numerik dan kategorikal saja, melainkan juga data lainnya seperti teks. Penambangan teks atau text mining memiliki arti secara luas sebagai proses penggalian informasi di mana pengguna berinteraksi dengan sekumpulan dokumen menggunakan alat analisis yang merupakan komponen dalam penambangan data [10]. Penambangan teks atau text mining adalah teknik yang digunakan untuk menyelesaikan suatu kasus seperti information retrieval classification [11] [12], dan juga information extraction [13]. Konsep text mining yang dipakai dalam klasifikasi adalah melakukan klasifikasi terhadap dokumen teks untuk pengelompokan dokumen sesuai dengan pembahasan topik [14].

Salah satu permasalahan yang sering diangkat dalam text mining adalah analisis sentimen. Analisa sentimen merupakan cabang dari penambangan data yang tujuan dasarnya yaitu untuk menganalisis, memahami, memproses, dan mengekstrak data tekstual dalam bentuk pendapat atau ulasan tentang entitas seperti produk, layanan, organisasi, individu, dan topik tertentu. Metode klasifikasi seperti Support Vector Machine (SVM), Naive Bayes, dan Decision Tree sering digunakan untuk membantu dalam proses analisis sentimen [15].

Pada penelitian Pratama, dkk [16] tentang analisis sentimen pengguna terhadap wisata pantai yang ada dikota Malang. Penelitian ini menggunakan metode TF-IDF dan Support Vector Machine untuk mengelompokkan ulasan pengunjung wisata menjadi kelompok negatif dan positif. Pada penelitian ini, data yang digunakan dalam opini pengunjung dalam bahasa Indonesia dengan jumlah 674 dari tahun 2013-2018 yang diperoleh dari disitus TripAdvisor dan tercatat sejumlah 43 objek pariwisata pantai Malang Selatan. Penelitian ini mendapatkan hasil klasifikasi yang cukup baik dengan rata-rata Accuracy yang dihasilkan sebesar 85%.

Syaifudin dan Irawan [17] melakukan teknik gabungan antara clustering dan classification opini wisata pantai. Metode yang digunakan untuk melakukan clustering adalah K-Means, sedangkan metode sentimen analisis yang digunakan yaitu metode Support Vector Machine. Data yang digunakan adalah data tweet sejumlah 500 tweet dari 10 pantai yang ada di Indonesia. Dalam penelitian ini hasil yang didapatkan akurasi sebesar 74,39% dengan menerapkan metode Support Vector Machine untuk menganalisis sentimen terhadap pendapat wisata pantai.

Selanjutnya, Widodo dkk [18] meneliti tentang analisis sentimen untuk wisata gunung Bromo dan Semeru dengan menggunakan data dari Tripadvisor. Penelitian ini diterapkan pada data opini terhadap wisata gunung Bromo dan juga Semeru dengan menerapkan metode klasifikasi Support Vector Machine (SVM). Data yang sudah ada tersebut dikelompokkan menjadi tiga kategori yaitu positif dan negatif. dari hasil penelitian ini yang menerapkan metode SVM diperoleh hasil akurasi untuk gunung Bromo dan Semeru masing-masing 71% dan 62%.

Penelitian berikutnya dengan memanfaatkan situs Tripadvisor untuk memperoleh data sentimen dari wisata Candi Borobudur untuk dilakukann klasifikasi sentimen. Data yang digunakan adalah data review mengenai wisata Candi Borobudur dalam bentuk bahasa Inggris yang berjumlah 3591 data. Pada penelitian ini dilakukan perbandingan untuk menganalisa sentimen dengan menerapkan Support Vector Machine dan K-Nearest Neighbor. Hasil penelitian ini menyimpulkan bahwa metode SVM lebih optimal dibandingkan dengan metode KNN [19].

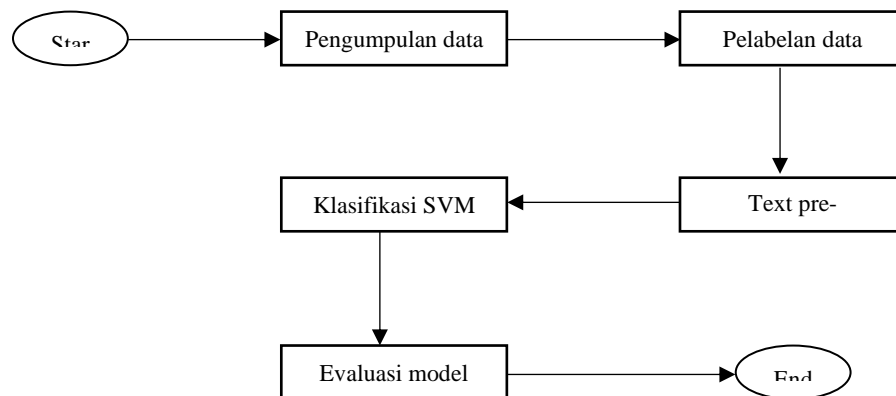
Penelitan berikutnya dengan memanfaatkan ulasan di google review untuk memperoleh data ulasan untuk dilakukan analisa sentimen. Data yang berhasil dikumpulkan berjumlah 2.143 yang merupakan data ulasan dari Summarecon Mal Bekasi yang didapatkan dengan cara web scrapping. Penelitian ini menerpakan dua metode untuk didapatkan hasil perbandingan antara Support Vector Machine dan Naive Bayes. Dalam penelitian ini diperoleh hasil perbandingan akurasi dari metode Naive Bayes dan Support Vector Machine masing-masing 80,95% dan 100% [20]. Dari penelitian tersebut metode SVM memiliki kinerja yang cukup baik dibandingkan dengan metode klasifikasi lainnya dalam pengklasifikasian data teks.

Pada penelitian ini akan dilakukan sentimen analisis terhadap ulasan masyarakat mengenai tempat wisata Api Tak Kunjung Padam. Data ulasan disini diambil dari ulasan Google Maps Review dimana tujuannya untuk mengetahui penerapan metode klasifikasi Support Vector Machine dalam pengklasifikasikan sentimen positif ataupun sentimen negatif pada data ulasan pengunjung objek wisata tersebut dengan harapan metode tersebut bekerja cukup baik dan menghasilkan akurasi diatas 80% dan juga hasil analisis ulasan tersebut dapat memberikan informasi yang berguna bagi pemerintah khususnya Dinas Kebudayaan dan Pariwisata Kota Pamekasan, pengelola objek wisata Api Tak Kunjung Padam maupun pihak lain yang membutuhkan dalam mengelola tempat wisata menjadi lebih terarah dan baik.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Ada beberapa tahapan pada penelitian ini meliputi pengumpulan data, pelabelan, text pre-processing, klasifikasi SVM, dan dilanjutkan dengan evaluasi model. Gambar 1 menunjukkan bagaimana proses dalam melakukan analisis sentimen dimulai dengan pengumpulan data hingga evaluasi sistem. Tahap pertama adalah pengumpulan data ulasan pengguna yang diperoleh dari Google Maps. Tahap selanjutnya adalah pelabelan data menjadi data bernilai ulasan positif dan ulasan negatif. Pada proses pelabelan ini dibantu oleh 3 orang ahli Bahasa Indonesia.



Gambar 1. Rancangan Proses

Setelah semua data diberi label, maka selanjutnya proses pemodelan dimulai dengan adanya text-preprocessing. Pada tahap ini dilakukan case folding, cleansing, tokenize, slang words, stopwords removal, dan stemming. Kemudian dilanjutkan dengan proses pemodelan klasifikasi dengan Support Vector Machine. Setelah hasil klasifikasi didapat proses selanjutnya adalah dilakukan evaluasi dengan menghitung nilai akurasi, recall, precision, dan f-1 score untuk mengetahui kinerja dari metode yang diusulkan.

2.2. Dataset

Data yang akan digunakan dalam penelitian ini adalah data ulasan yang diambil dari ulasan google maps dari objek wisata Api Tak Kunjung Padam (<https://maps.app.goo.gl/UdSfkssYfs82HhHW9>). Proses pengambilan data

dilakukan dengan menggunakan tools Instant Data Scraping dan menghasilkan total 1745 data. Pada penelitian ini, data yang diperoleh diseleksi berdasarkan kriteria berikut ini:

1. Data yang diolah adalah data yang memiliki ulasan pendapat pengguna, bukan hanya rating saja. Pada data penelitian ini, ditemukan sebanyak 814 data rating saja tanpa review. Hal ini menyebabkan data tersebut tidak dapat diolah lebih lanjut untuk analisa sentimen.
2. Data ulasan yang diolah adalah yang menggunakan bahasa Indonesia, bukan bahasa daerah ataupun bahasa Inggris. Hal ini terkait dengan library yang digunakan dalam penelitian ini

Setelah data diseleksi menggunakan kriteria diatas, total data yang diperoleh adalah sebanyak 861 data. Tabel dibawah ini menunjukkan beberapa data yang digunakan dalam penelitian ini.

Tabel 1. Potongan Dataset

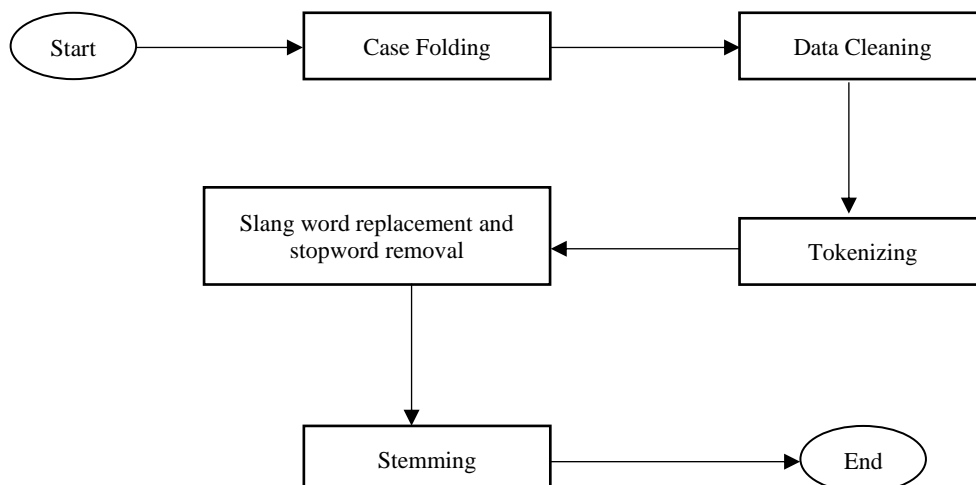
No	Ulasan
1	Wisata yg sangat natural banget...
2	Sangat bagus
3	Ya saya suka dengan pelayanan rest areanya
4	Parkir Rp. 10.000, tanpa menunggu apa apa bisa langsung ke lokasi yang penuh dengan pilihan makanan, oleh oleh di sekitar wisata Api Abadi itu. Paling asyik kalau malam karena api bisa tampak lebih jelas.
5	Panas sekali cuacanya
....
857	Api alami cocok buat bakar jagung
858	Ada begitu banyak kucing berkeliaran
859	lebih baik kamu pergi di malam hari
860	Bersyukurlah atas kebaikan yang diberikan oleh Yang Mahakuasa ...
861	Keren

2.3. Pelabelan Data

Setelah data terkumpul, proses selanjutnya yaitu melakukan pelabelan pada setiap data secara manual. Pelabelan data dilakukan oleh 3 orang ahli Bahasa Indonesia. Untuk menghindari ketidakintegritasian data sehingga jika terjadi perbedaan pendapat, pihak ketiga dapat mengatasi permasalahan tersebut dengan menggunakan metode majority voting yakni label dengan suara terbanyak yang akan digunakan sebagai kesimpulan label. Pemilihan kandidat pelabelan tersebut berasal dari tingkat SMP/MTs merupakan tenaga pendidik bahasa Indonesia dan sudah menempuh pendidikan Indonesia atau sastra Indonesia.

2.4. Text Preprocessing

Tahap selanjutnya yaitu preprocessing, tahap ini memiliki tujuan agar data ulasan tersebut menjadi lebih terstruktur untuk memudahkan dalam proses selanjutnya. Berikut adalah alur dari tahap preprocessing. Pada Gambar 2 dapat dilihat alur dari text preprocessing. Diawali dari proses case folding untuk mengubah dokumen ke huruf kecil. Kemudian masuk ke data cleaning untuk menghapus tanda baca dan karakter lainnya yang tidak penting. Selanjutnya proses tokenizing untuk memecah kalimat dalam bentuk token. Setelah proses tokenizing selesai masuk ke proses slang word replacement yaitu untuk mengubah kata tidak baku kedalam bahasa baku dan proses stopword removal untuk menghapus kata yang tidak penting dan yang terakhir adalah proses stemming yaitu untuk mengubah kata kedalam bentuk kata dasar. Setelah proses selesai hasil berupa dokumen yang sudah siap di proses ke tahap selanjutnya.



Gambar 2. Tahapan Text Preprocessing

2.5. Proses Klasifikasi Support Vector Machine (SVM)

Pada tahap proses klasifikasi yang akan dilakukan dalam penelitian ini menerapkan model klasifikasi yaitu metode Support Vector Machine. SVM merupakan suatu model yang umum digunakan untuk pengklasifikasian atau support vector classification dan juga Support Vector Regression. SVM terkenal karena memiliki konsep yang lebih jelas secara sistematis dibandingkan dengan pendekatan klasifikasi lainnya. Tidak hanya itu SVM tidak hanya mengklasifikasikan dalam bentuk linear akan tetapi dalam bentuk non linear juga [17].

Support Vector Machine (SVM) adalah metode pengklasifikasian dengan cara kerja menganalisa data dan mengenali pola dari data yang ada. Biasanya SVM digunakan untuk penelitian yang berpusat pada klasifikasi biner yang memiliki dua kelas yakni positif dan negatif. SVM adalah model klasifikasi yang membentuk hyperplane dalam ruang input multidimensi untuk memisahkan data terbaik. Dengan mengukur batas-batas keanggotaan kelas, kita dapat menentukan hyperplane dengan paling akurat. SVM awalnya digunakan untuk mengklasifikasikan data numerik, tetapi beberapa penelitian yang telah disampaikan diatas menunjukkan bahwa SVM juga sangat efisien dan cepat dalam menangani masalah dengan data teks. Karena sifat bahasa yang cenderung memiliki data berdimensi tinggi tetapi memiliki kecenderungan untuk saling berkorelasi, maka metode SVM dapat bekerja dengan baik untuk mengelompokkan data menjadi 2 bagian dengan menggunakan hyperplane [14]. Langkah-langkah proses metode klasifikasi dengan menggunakan Support Vector Machine bisa diperhatikan berikut ini:

1. Menginisialisasi nilai awal dari α , c , epsilon, lambda dan juga gamma.
2. Menginputkan data latih (training) berdasarkan kemunculan keyword dalam satu kalimat.
3. Mencari dot product untuk setiap data dengan menggunakan fungsi kernel [K]. Formula fungsi kernel linear adalah sebagai berikut:

$$K(x, x') = \exp(-\gamma |x - x'|^2) \quad (1)$$

Fungsi linier digunakan sebagai fungsi kernel. Sebelum itu data harus ditranspose terlebih dahulu untuk dilakukan perkalian matrik $Ax A^T$.

4. Menghitung matriks dengan rumus sebagai berikut:

$$D_{ij} = Y_i Y_j (K(X_i X_j) + \lambda^2) \quad (2)$$

Dengan ketentuan:

D_{ij} = Elemen matriks ke i j

Y_i = Kelas data ke i

Y_j = Kelas data ke j

λ^2 = Batas teoritis

5. Mencari nilai error dengan rumus:

$$E_i = \sum_{j=1}^i \alpha_j D_{ij} \quad (3)$$

6. Mencari nilai dari delta alpha dengan formula berikut:

$$\delta \alpha_i = \min\{\max[\gamma(1 - E_i) - \alpha_i] C - \alpha_i\} \quad (4)$$

mencari nilai alpha baru dengan formula sebagai berikut:

$$\alpha_i = \alpha_i + \delta \alpha_i \quad (5)$$

7. Mencari nilai bias dengan formula sebagai berikut:

$$b = -\frac{1}{2}(w \cdot x^+ + w \cdot x^-) \quad (6)$$

8. Setelah nilai α , w , dan b diketahui, selanjutnya masuk ke tahap pengujian. untuk melakukan pengujian pertama dilakukan perhitungan dot product antara data testing dengan semua data training dengan fungsi kernel:

$$K(x, x_i) = x \cdot x_i \quad (7)$$

Setelah itu pengujian dilakukan dengan fungsi keputusan:

$$f(x) = w \cdot x + b \text{ atau } f(x) = \sum_{i=1}^m \alpha_i x_i K(x, x_i) + b \quad (8)$$

2.6. Evaluasi Model

Pada tahapan ini dilakukan perhitungan akurasi untuk melihat seberapa baik pemodelan menggunakan metode klasifikasi yang digunakan untuk mengklasifikasikan ulasan dari wisata Api Tak Kunjung padam. Dalam proses evaluasi penelitian ini memanfaatkan confusion matrix untuk menghitung nilai akurasi, recall, precision dan f1-score.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini terdapat beberapa uji coba yang dilakukan dengan menginputkan beberapa parameter C dan gamma yang masing-masing berbeda untuk mengetahui nilai c dan Gamma terbaik sehingga menghasilkan nilai akurasi yang optimal. Nilai c yang diinputkan dinataranya 0,01, 0,1, 1, 10, 100 dan nilai gamma 0,01, 0,1, 0,5, 1, 100. Pengujian ini juga dilakukan dengan beberapa pembagian data latih dan data uji dengan presentase 80%:20%, 70%:30%, 60%:40% dan 50%:50%. Berikut adalah hasil dari skenario uji coba tersebut:

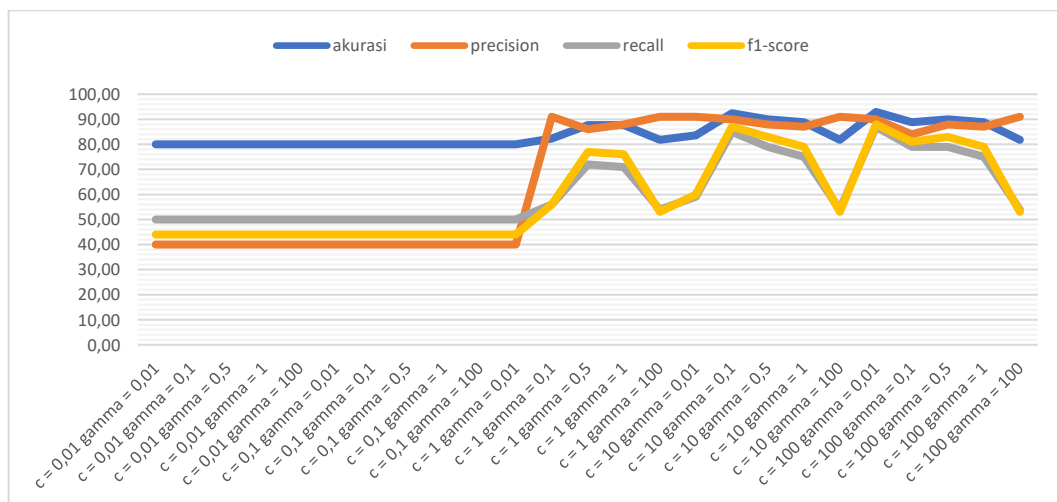
3.1. Pembagian data 80%:20%

Pengujian pertama pada penelitian ini untuk masing-masing nilai c dan gamma dilakukan dengan pembagian data latih dan data uji dengan presentase 80%:20%. Dari pengujian tersebut didapatkan hasil akurasi, precision, recall, dan F1-score dalam tabel berikut:

Tabel 2. Hasil pengujian data 80%:20%

Parameter	Akurasi	Precision	Recall	F1-score
c = 0,01 gamma = 0,01	80,00%	40%	50%	44%
c = 0,01 gamma = 0,1	80,00%	40%	50%	44%
c = 0,01 gamma = 0,5	80,00%	40%	50%	44%
c = 0,01 gamma = 1	80,00%	40%	50%	44%
c = 0,01 gamma = 100	80,00%	40%	50%	44%
c = 0,1 gamma = 0,01	80,00%	40%	50%	44%
c = 0,1 gamma = 0,1	80,00%	40%	50%	44%
c = 0,1 gamma = 0,5	80,00%	40%	50%	44%
c = 0,1 gamma = 1	80,00%	40%	50%	44%
c = 0,1 gamma = 100	80,00%	40%	50%	44%
c = 1 gamma = 0,01	80,00%	40%	50%	44%
c = 1 gamma = 0,1	82,35%	91%	56%	56%
c = 1 gamma = 0,5	87,65%	86%	72%	77%
c = 1 gamma = 1	87,65%	88%	71%	76%
c = 1 gamma = 100	81,76%	91%	54%	53%
c = 10 gamma = 0,01	83,53%	91%	59%	60%
c = 10 gamma = 0,1	92,35%	90%	85%	87%
c = 10 gamma = 0,5	90,00%	88%	79%	83%
c = 10 gamma = 1	88,82%	87%	75%	79%
c = 10 gamma = 100	81,76%	91%	54%	53%
c = 100 gamma = 0,01	92,94%	90%	87%	88%
c = 100 gamma = 0,1	88,82%	84%	79%	81%
c = 100 gamma = 0,5	90,00%	88%	79%	83%
c = 100 gamma = 1	88,82%	87%	75%	79%
c = 100 gamma = 100	81,76%	91%	54%	53%

Pada Tabel 2 dapat dilihat hasil pengujian dengan akurasi paling tinggi didapatkan pada pengujian dengan nilai c = 100 dan gamma 0,01 dengan hasil akurasi 92,94%, precision = 90%, recall = 87%, dan F1-score = 88%. Berikut adalah grafik dari pengujian tersebut:



Gambar 3. Grafik pembagian data 80%:20%

Dari Gambar 3 terlihat statistik dimana pada parameter c yang rendah menyebabkan akurasi yang rendah. Ini terjadi karena semakin rendah parameter c maka batas keputusan atau decision boundary semakin besar atau melebar. Hal ini memungkinkan kesalahan klasifikasi karena batas kesalahan yang luas. Sebaliknya jika parameter c semakin tinggi maka batas kesalahan semakin sempit dan membuat akurasi semakin tinggi. Selain itu nilai gamma cenderung semakin kecil membuat akurasi semakin tinggi. Sehingga pada percobaan ini akurasi tertinggi ada pada parameter c = 100 dan gamma = 0,01.

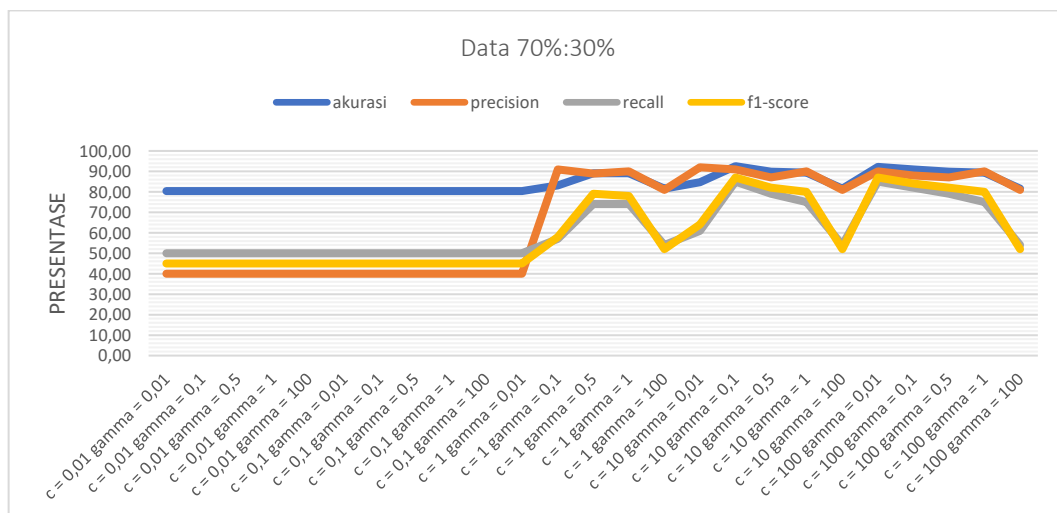
3.2. Pembagian data 70%:30%

Pengujian pertama pada penelitian ini untuk masing-masing nilai c dan gamma dilakukan dengan pembagian data latih dan data uji dengan presentase 70%:30%. Dari pengujian tersebut didapatkan hasil akurasi, precision, recall, dan F1-score dalam tabel berikut:

Tabel 3. Hasil pengujian data 70%:30%

Parameter	Akurasi	Precision	Recall	F1-score
c = 0,01 gamma = 0,01	80,39%	40%	50%	45%
c = 0,01 gamma = 0,1	80,39%	40%	50%	45%
c = 0,01 gamma = 0,5	80,39%	40%	50%	45%
c = 0,01 gamma = 1	80,39%	40%	50%	45%
c = 0,01 gamma = 100	80,39%	40%	50%	45%
c = 0,1 gamma = 0,01	80,39%	40%	50%	45%
c = 0,1 gamma = 0,1	80,39%	40%	50%	45%
c = 0,1 gamma = 0,5	80,39%	40%	50%	45%
c = 0,1 gamma = 1	80,39%	40%	50%	45%
c = 0,1 gamma = 100	80,39%	40%	50%	45%
c = 1 gamma = 0,01	80,39%	40%	50%	45%
c = 1 gamma = 0,1	83,14%	91%	57%	58%
c = 1 gamma = 0,5	89,02%	89%	74%	79%
c = 1 gamma = 1	89,02%	90%	74%	78%
c = 1 gamma = 100	81,57%	81%	54%	52%
c = 10 gamma = 0,01	84,71%	92%	61%	64%
c = 10 gamma = 0,1	92,55%	91%	85%	87%
c = 10 gamma = 0,5	89,80%	87%	79%	82%
c = 10 gamma = 1	89,41%	90%	75%	80%
c = 10 gamma = 100	81,57%	81%	54%	52%
c = 100 gamma = 0,01	92,16%	90%	85%	87%
c = 100 gamma = 0,1	90,98%	88%	82%	84%
c = 100 gamma = 0,5	89,80%	87%	79%	82%
c = 100 gamma = 1	89,41%	90%	75%	80%
c = 100 gamma = 100	81,57%	81%	54%	52%

Pada Tabel 3 dapat dilihat hasil pengujian dengan akurasi paling tinggi didapatkan pada pengujian dengan nilai c = 10 dan gamma 0,1 dengan hasil akurasi 92,55%, precision = 91%, recall = 85%, dan F1-score = 87%. Berikut adalah grafik dari pengujian tersebut:



Gambar 4. Grafik pembagian data 70%:30%

Pada Gambar 4 terlihat pada saat parameter c rendah menyebabkan akurasi rendah. Hal ini karena pengaruh dari parameter c yang rendah menyebabkan batas keputusan semakin besar. Dari grafik tersebut hyperplane terbaik ada pada parameter c = 10 dan gamma 0,1. Meskipun pada penelitian ini parameter c paling tinggi 100 dan gamma paling rendah 0,01 namun parameter paling optimal pada pengujian ini ada pada parameter c =10 dan gamma = 0,1. Hal ini bisa dipengaruhi oleh faktor presentase pembagian data.

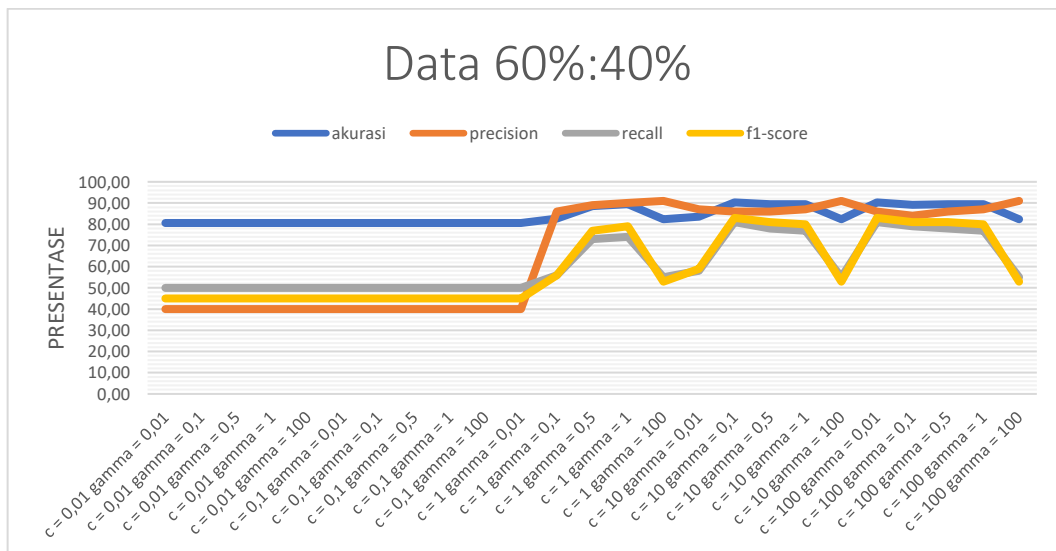
3.3. Pembagian data 60%:40%

Pengujian pertama pada penelitian ini untuk masing-masing nilai c dan gamma dilakukan dengan pembagian data latih dan data uji dengan presentase 60%40%, Dari pengujian tersebut didapatkan hasil akurasi, precision, recall, dan F1-score dalam tabel berikut:

Tabel 4. Hasil pengujian data 60%:40%

Parameter	Akurasi	Precision	Recall	F1-score
c = 0,01 gamma = 0,01	80,59%	40%	50%	45%
c = 0,01 gamma = 0,1	80,59%	40%	50%	45%
c = 0,01 gamma = 0,5	80,59%	40%	50%	45%
c = 0,01 gamma = 1	80,59%	40%	50%	45%
c = 0,01 gamma = 100	80,59%	40%	50%	45%
c = 0,1 gamma = 0,01	80,59%	40%	50%	45%
c = 0,1 gamma = 0,1	80,59%	40%	50%	45%
c = 0,1 gamma = 0,5	80,59%	40%	50%	45%
c = 0,1 gamma = 1	80,59%	40%	50%	45%
c = 0,1 gamma = 100	80,59%	40%	50%	45%
c = 1 gamma = 0,01	80,59%	40%	50%	45%
c = 1 gamma = 0,1	82,65%	86%	56%	56%
c = 1 gamma = 0,5	88,53%	89%	73%	77%
c = 1 gamma = 1	89,41%	90%	74%	79%
c = 1 gamma = 100	82,35%	91%	55%	53%
c = 10 gamma = 0,01	83,53%	87%	58%	59%
c = 10 gamma = 0,1	90,29%	86%	81%	83%
c = 10 gamma = 0,5	89,41%	86%	78%	81%
c = 10 gamma = 1	89,41%	87%	77%	80%
c = 10 gamma = 100	82,35%	91%	55%	53%
c = 100 gamma = 0,01	90,29%	86%	81%	83%
c = 100 gamma = 0,1	89,12%	84%	79%	81%
c = 100 gamma = 0,5	89,41%	86%	78%	81%
c = 100 gamma = 1	89,41%	87%	77%	80%
c = 100 gamma = 100	82,35%	91%	55%	53%

Pada Tabel 4 dapat dilihat hasil pengujian dengan akurasi paling tinggi didapatkan pada pengujian dengan nilai c = 10, gamma 0,1 dan c = 100, gamma = 0,01 dengan hasil akurasi 90,29%, precision = 86%, recall = 81%, dan F1-score = 83%. Berikut adalah grafik dari pengujian tersebut:



Gambar 5. Grafik pembagian data 60%:40%

Dari Gambar 5 terlihat statistik dimana pada parameter c yang rendah menyebabkan akurasi yang rendah. Ini terjadi karena semakin rendah parameter c maka batas keputusan atau decision boundary semakin besar atau melebar. Hal ini memungkinkan kesalahan klasifikasi karena batas kesalahan yang luas. Sebaliknya jika parameter c semakin tinggi maka batas kesalahan semakin sempit dan membuat akurasi semakin tinggi. Selain itu nilai γ cenderung semakin kecil membuat akurasi semakin tinggi. Sehingga pada percobaan ini akurasi tertinggi ada pada parameter $c = 10$, $\gamma = 0,1$ dan $C = 100$, $\gamma = 0,01$.

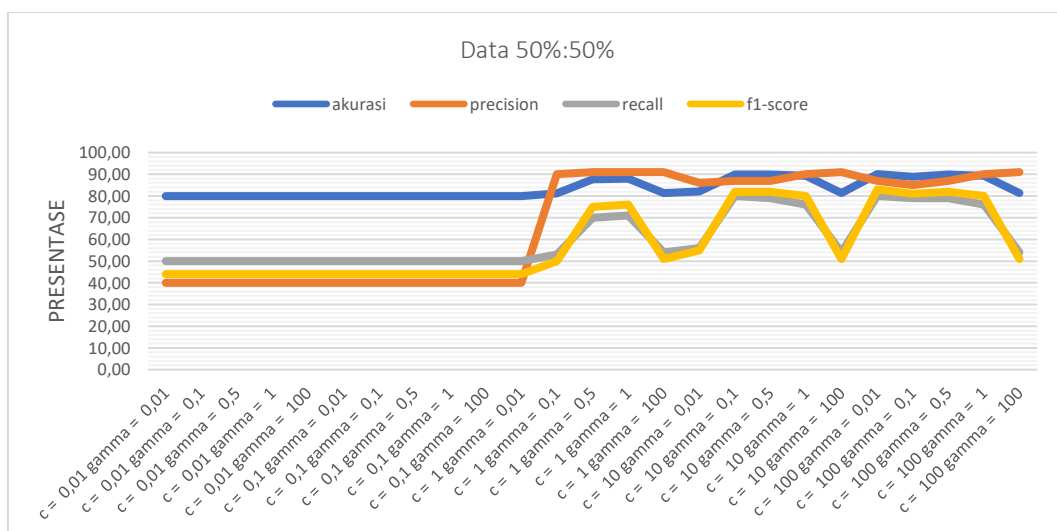
3.4. Pembagian data 50%:50%

Pengujian pertama pada penelitian ini untuk masing-masing nilai c dan γ dilakukan dengan pembagian data latih dan data uji dengan presentase 60%:40%. Dari pengujian tersebut didapatkan hasil akurasi, precision, recall, dan F1-score dalam tabel berikut:

Tabel 5. Hasil pengujian data 50%:50%

Parameter	Akurasi	Precision	Recall	F1-score
$c = 0,01 \gamma = 0,01$	79,95%	40%	50%	44%
$c = 0,01 \gamma = 0,1$	79,95%	40%	50%	44%
$c = 0,01 \gamma = 0,5$	79,95%	40%	50%	44%
$c = 0,01 \gamma = 1$	79,95%	40%	50%	44%
$c = 0,01 \gamma = 100$	79,95%	40%	50%	44%
$c = 0,1 \gamma = 0,01$	79,95%	40%	50%	44%
$c = 0,1 \gamma = 0,1$	79,95%	40%	50%	44%
$c = 0,1 \gamma = 0,5$	79,95%	40%	50%	44%
$c = 0,1 \gamma = 1$	79,95%	40%	50%	44%
$c = 0,1 \gamma = 100$	79,95%	40%	50%	44%
$c = 1 \gamma = 0,01$	79,95%	40%	50%	44%
$c = 1 \gamma = 0,1$	81,13%	90%	53%	50%
$c = 1 \gamma = 0,5$	87,74%	91%	70%	75%
$c = 1 \gamma = 1$	87,97%	91%	71%	76%
$c = 1 \gamma = 100$	81,37%	91%	54%	51%
$c = 10 \gamma = 0,01$	82,08%	86%	56%	55%
$c = 10 \gamma = 0,1$	89,86%	87%	80%	82%
$c = 10 \gamma = 0,5$	89,86%	87%	79%	82%
$c = 10 \gamma = 1$	89,39%	90%	76%	80%
$c = 10 \gamma = 100$	81,37%	91%	54%	51%
$c = 100 \gamma = 0,01$	90,09%	87%	80%	83%
$c = 100 \gamma = 0,1$	88,92%	85%	79%	81%
$c = 100 \gamma = 0,5$	89,86%	87%	79%	82%
$c = 100 \gamma = 1$	89,39%	90%	76%	80%
$c = 100 \gamma = 100$	81,37%	91%	54%	51%

Pada Tabel 5 dapat dilihat hasil pengujian dengan akurasi paling tinggi didapatkan pada pengujian dengan nilai $c = 100$ dan $\gamma = 0,01$ dengan hasil akurasi 90.09%, precision = 87%, recall = 80%, dan F1-score = 83%. Berikut adalah grafik dari pengujian tersebut:



Gambar 6. Grafik pembagian data 60%:40%

Dari Gambar 6. terlihat statistik dimana pada parameter c yang rendah menyebabkan akurasi yang rendah. Ini terjadi karena semakin rendah parameter c maka batas keputusan atau decision boundary semakin besar atau melebar. Hal ini memungkinkan kesalahan klasifikasi karena batas kesalahan yang luas. Sebaliknya jika parameter c semakin tinggi maka batas kesalahan semakin sempit dan membuat akurasi semakin tinggi. Selain itu pada grafik tersebut, parameter c yang kecil menyebabkan parameter γ tidak terlalu berpengaruh terhadap akurasi. Selain itu nilai γ cenderung semakin kecil membuat akurasi semakin tinggi. Sehingga pada percobaan ini akurasi tertinggi ada pada parameter $c = 100$ dan $\gamma = 0,01$.

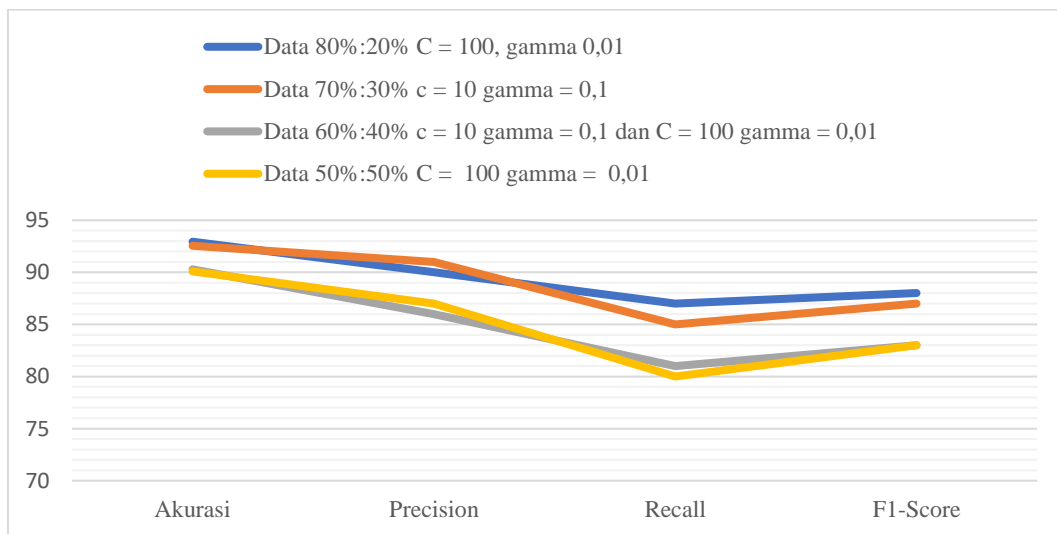
3.4. Hasil akurasi terbaik

Setelah dilakukan uji coba dari beberapa pembagian data uji dan data latih terhadap masing-masing nilai c dan γ didapatkan hasil akurasi terbaik dari masing-masing pengujian. Berikut tabel memuat hasil pengujian tersebut:

Tabel 6. Hasil akurasi terbaik

Pengujian	Parameter	Akurasi	Precision	Recall	F1-Score
Data 80%:20%	$c = 100, \gamma = 0,01$	92,94%	90%	87%	88%
Data 70%:30%	$c = 10, \gamma = 0,1$	92,55%	91%	85%	87%
Data 60%:40%	$c = 10, \gamma = 0,1$ dan $c = 100, \gamma = 0,01$	90,29%	86%	81%	83%
Data 50%:50%	$c = 100, \gamma = 0,01$	90,09%	87%	80%	83%

Pada Tabel 6 merupakan hasil akurasi terbaik dari masing-masing pengujian dengan beberapa pembagian data latih dan data uji terhadap masing-masing nilai c dan γ . Dari pengujian tersebut hasil akurasi paling optimal ada pada pengujian dengan data 80%:20% dengan nilai $c = 100$ dan $\gamma = 0.01$ dengan hasil akurasi 92,94%.



Gambar 7. Grafik pembagian data 60%:40%

Pada Gambar 7 terlihat akurasi terbaik ada pada percobaan pembagian data dengan presentase 80%:20%. Pada penelitian ini semakin besar data latih menghasilkan akurasi yang semakin baik. Hal ini disebabkan dengan banyaknya data latih, maka kemampuan sistem untuk mengenali model dan karakteristik data menjadi semakin lebih baik. Pada data diatas juga dapat disimpulkan bahwa parameter c yang besar dan juga γ yang kecil juga memberikan kontribusi dalam peningkatan nilai akurasi.

4. KESIMPULAN

Penelitian ini menggunakan data teks berupa ulasan pengunjung lokasi wisata Api Tak Kunjung Padam yang diperoleh dari Google Maps Review. Pengumpulan data dilakukan dengan menggunakan tools yang tersedia dengan memasukkan URL target. Dari 1745 data yang diperoleh, total ada 861 data yang diperoleh yang siap untuk diolah lebih lanjut untuk analisa sentimen. Ujicoba model SVM yang dibangun pada penelitian ini dilakukan dengan membagi data input menjadi data latih dan data uji dengan presentase 80%:20%, 70%:30%, 60%:40%, dan 50%:50%. Untuk masing – masing skenario uji coba, dilakukan perbedaan paramater yang digunakan secara sistematis untuk mengetahui peran masing – masing parameter terkait dengan accuracy, recall, precision dan F1-Score. Nilai c yang diteliti pada studi ini adalah = 0.01; 0.01; 1; 10; 100. Sedangkan nilai γ yang digunakan

adalah: 0.01; 0.01; 0.5; 1; 100. Dari pengujian tersebut didapatkan hasil akurasi terbaik pada pengujian dengan pembagian data 80%:20% dengan nilai $c = 100$ dan $\gamma = 0,01$ dengan akurasi sebesar 92,94%, precision = 90%, recall = 87% dan F1-score = 88%. Hasil tersebut menunjukkan semakin besar presentase data latih yang digunakan cenderung semakin besar akurasi yang dihasilkan. Selain itu faktor yang mempengaruhi akurasi tersebut adalah parameter c dan γ . Semakin besar nilai parameter c maka cenderung semakin tinggi nilai akurasi karena batas kesalahan semakin sempit sehingga terjadinya kesalahan prediksi semakin menurun. Sedangkan semakin kecil nilai parameter γ cenderung memiliki nilai akurasi yang semakin tinggi.

REFERENCES

- [1] A. Sucipto, and H. Kusumodestoni, 2022. Implementasi Data Mining Algoritma Apriori Terhadap Penjualan Spare Part Motor di Yamaha Arista Subang. AMRI (Analisa Metode Rekayasa Informatika), 1(1), pp.52-58.
- [2] A. Wahyudi, S.O. Tampubolon, N. Afrilia Putri, A. Ghassa, E. Rasywir, and D. Kisbianty, 2022. Penerapan Data Mining Algoritma Naive Bayes Klasifier untuk Mengetahui Minat Beli Pelanggan terhadap INDIHOME. Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM), 2(2), pp.240-247.
- [3] A. Nofianti, M.Y. Yawan, and M.A. Nazar, M.A., 2023. Implementasi Data Mining dalam Pengolahan Data Transaksi Toko Sembako Menggunakan Algoritma Apriori (Studi Kasus: Toko Devan Mart). G-Tech: Jurnal Teknologi Terapan, 7(1), pp.165-173.
- [4] S.P. Tamba, 2022. Penerapan Data Mining Algoritma Apriori Dalam Menentukan Stok Bahan Baku Pada Restoran Nelayan Menggunakan Metode Association Rule. Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA), 5(2), pp.97-102.
- [5] R.D. Romadhona, 2022, September. Pemanfaatan Teknik Clustering untuk Mencapai Kecerdasan Bisnis Menggunakan Algoritma K-Means. In Seminar Nasional Sistem Informasi (SENASIF) (Vol. 6, pp. 3015-3023).
- [6] Hasanah, N.N. and Purnomo, A.S., 2022. Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus: Perpustakaan Politeknik LPP Yogyakarta). Jurnal Teknologi Dan Sistem Informasi Bisnis, 4(2), pp.300-311.
- [7] J.I. Pratiwi, D.R. Utami, and U. Enri, U., 2022. Klasifikasi Data Mining Dalam Menentukan Pemberian Pinjaman Berbasis Arisan Online dengan Algoritma C4. 5. InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan, 6(2), pp.262-266.
- [8] U.P. Sanjaya, T. Pribadi, and I.W.D. Prastya, I2022. Klasifikasi Dana Hibah Usaha Mikro Kecil dan Menengah dengan Metode Naive Bayes. Indonesian Journal of Computer Science, 11(3).
- [9] A. H. Yunial, "Analisa Perbandingan Algoritma Klasifikasi Support Vector Machine, Decision Tree Dan Naive Bayes," Pros. Semin. Nas. Inform. dan Sist. Inf. , vol. 5, no. 2, 2020.
- [10] Z. Alhaq, A. Mustopa, S. Mulyatun, and J. D. Santoso, "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter," J. Inf. Syst. Manag., vol. 3, no. 2, pp. 44–49, 2021, doi: 10.24076/joism.2021v3i2.558.
- [11] Kumar, A. and Starly, B., 2022. "FabNER": information extraction from manufacturing process science domain literature using named entity recognition. Journal of Intelligent Manufacturing, 33(8), pp.2393-2407.
- [12] D.K. Jain, P. Boyapati, J. Venkatesh, and M. Prakash, 2022. An intelligent cognitive-inspired computing with big data analytics framework for sentiment analysis and classification. Information Processing & Management, 59(1), p.102758.
- [13] W. Fu, E.K. Choi, and H.S. Kim, 2022. Text mining with network analysis of online reviews and consumers' satisfaction: A case study in Busan wine bars. Information, 13(3), p.127.
- [14] Lestari, T.P., 2022. Analisis Text Mining pada Sosial Media Twitter Menggunakan Metode Support Vector Machine (SVM) dan Social Network Analysis (SNA). Jurnal Informatika Ekonomi Bisnis, pp.65-71.
- [15] M. I. Fikri, T. S. Sabrila, and Y. Azhar, 2020. Perbandingan metode naive bayes dan support vector machine pada analisis sentimen twitter. SMATIKA Jurnal: STIKI Informatika Jurnal, 10(02), pp.71-76.
- [16] Y.T. Pratama, 2018. Analisis Sentimen Opini Pelanggan Terhadap Aspek Pariwisata Pantai Malang Selatan Menggunakan TF-IDF dan Support Vector Machine (Doctoral dissertation, Universitas Brawijaya).
- [17] Syaifudin, Y.W. and Irawan, R.A., 2018. Implementasi Analisis Clustering Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode K-Means. Jurnal Informatika Polinema, 4(3), pp.189-189.
- [18] E. Widodo, I. Al Maksur, K.D. Larasati, and N.P. Saraswati, 2019, November. Analisis Sentimen Tripadvisor Terhadap Pariwisata Gunung Bromo dan Gunung Semeru. In Seminar Multimedia & Artificial Intelligence (Vol. 2, pp. 43-48).
- [19] R.P. Saputri, W.S. Winahju, dan K. Fithriasari, 2020. Klasifikasi Sentimen Wisatawan Candi Borobudur pada Situs TripAdvisor Menggunakan Support Vector Machine dan K-Nearest Neighbor. Jurnal Sains dan Seni ITS, 8(2), pp.D349-D356.
- [20] H. Herlawati, R. T. Handayanto, P. D. Atika, F. N. Khasanah, A. Y. P. Yusuf, and D. Y. Septia, 2021. Analisis Sentimen Pada Situs Google Review dengan Naive Bayes dan Support Vector Machine. Jurnal Komtika (Komputasi dan Informatika), 5(2), pp.153-163.