

Analisis Sentimen Komentar Universitas di Indonesia Menggunakan Metode Naive Bayes dan SVM

Benediktus Silaban^{a1}, Ida Ayu Gde Suwiprabayanti^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹silaban.2308561139@student.unud.ac.id
²iagsuwiprabayantiputra@unud.ac.id

Abstract

This research aims to analyze public sentiment towards Universitas Indonesia based on user reviews collected from Google Maps. In the era of digital information, online reviews serve as invaluable feedback channels, significantly influencing an institution's reputation and prospective student choices. This study employs a sentiment analysis approach to automatically classify reviews into positive, negative, and neutral categories. The methodology involves several key stages: data collection from Google Maps, comprehensive text preprocessing (including cleaning, tokenization, stopword removal, and stemming), and feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF). For classification, two prominent machine learning algorithms, Support Vector Machine (SVM) and Multinomial Naive Bayes, are utilized. Both models are trained and evaluated on the processed dataset to assess their performance in accurately classifying sentiment. A comparative analysis will be conducted to highlight the strengths and weaknesses of each algorithm in this specific context. The findings are expected to provide Universitas Indonesia with actionable insights into public perception, identify areas for improvement, and contribute to the understanding of sentiment analysis applications in educational contexts.

Keywords: Sentiment Analysis, Google Maps, Support Vector Machine (SVM), Naive Bayes, Text Mining, Natural Language Processing.

1. Pendahuluan

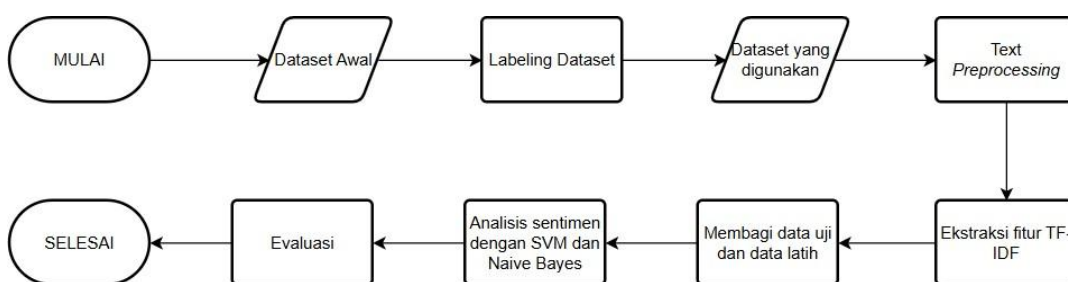
Di era digital saat ini, ulasan daring telah menjadi sumber informasi yang sangat berharga bagi individu maupun institusi. Masyarakat modern semakin mengandalkan pengalaman dan opini yang dibagikan secara daring untuk membuat keputusan, mulai dari pembelian produk hingga pemilihan tempat pendidikan. Bagi institusi pendidikan tinggi seperti Universitas Indonesia, umpan balik dari mahasiswa, alumni, dan pengunjung yang tersebar di berbagai platform digital, termasuk Google Maps, merupakan cerminan persepsi publik yang krusial. Reputasi universitas, kualitas fasilitas, lingkungan akademik, hingga pengalaman layanan dapat terekam dalam ulasan-ulasan ini, yang pada gilirannya sangat mempengaruhi calon mahasiswa dan citra institusi secara keseluruhan.

Volume ulasan yang sangat besar di platform seperti Google Maps seringkali membuat analisis manual menjadi tidak praktis dan memakan waktu. Keterbatasan ini menghambat kemampuan universitas untuk secara cepat dan efisien memahami sentimen yang mendasari ulasan tersebut. Oleh karena itu, diperlukan sebuah pendekatan otomatis yang mampu mengekstraksi informasi sentimen dari data teks yang tidak terstruktur ini. Analisis sentimen, sebuah cabang dari pengolahan bahasa alami (Natural Language Processing) dan penambangan teks (text mining), menawarkan solusi untuk mengidentifikasi, mengekstrak, dan mengkuantifikasi polaritas emosional dari data tekstual.

Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna Universitas Indonesia di Google Maps. Data ulasan akan dikumpulkan dan melalui serangkaian tahapan pra-pemrosesan teks yang komprehensif, meliputi pembersihan data, tokenisasi, penghapusan *stopwords*, dan *stemming*, untuk mempersiapkan data bagi model *machine learning*. Ekstraksi fitur akan dilakukan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF), yang akan mengubah teks menjadi representasi numerik yang dapat dipahami oleh algoritma. Dua algoritma klasifikasi *machine learning* yang populer dan terbukti efektif dalam analisis sentimen, yaitu Support Vector Machine (SVM) dan Naive Bayes, akan digunakan untuk mengklasifikasikan ulasan ke dalam kategori sentimen positif, negatif, dan netral. Melalui perbandingan kinerja kedua algoritma ini, penelitian ini diharapkan dapat memberikan wawasan mendalam mengenai persepsi publik terhadap Universitas Indonesia, mengidentifikasi kekuatan dan kelemahan yang dipersepsikan, serta memberikan kontribusi metodologis dalam penerapan analisis sentimen pada konteks pendidikan tinggi di Indonesia.

2. Metode Penelitian

2.1 Alur Penelitian



Gambar 1. Alur Penelitian

Penelitian ini mengadopsi pendekatan kuantitatif dengan tahapan yang sistematis untuk menganalisis sentimen ulasan pengguna. Proses dimulai dengan pengumpulan data berupa ulasan daring dari Google Maps terkait Universitas Telkom Bandung dari platform kaggle. Data mentah tersebut kemudian melalui serangkaian pra-pemrosesan data yang komprehensif, meliputi pembersihan teks dari karakter tidak relevan, normalisasi huruf, tokenisasi untuk memecah teks menjadi kata-kata individual, penghapusan *stopwords* guna menghilangkan kata-kata umum yang tidak berkontribusi pada sentimen, dan *stemming* untuk mereduksi kata ke bentuk dasarnya.

Tahap krusial selanjutnya adalah pelabelan sentimen secara manual untuk mengklasifikasikan setiap ulasan ke dalam kategori positif, negatif, atau netral, yang menjadi *ground truth* bagi pelatihan model. Setelah data bersih dan terlabeli, dilakukan ekstraksi fitur menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk mengubah representasi teks menjadi vektor numerik. Data yang telah menjadi fitur numerik kemudian dipecah menjadi data pelatihan dan data pengujian dengan rasio 80:20 menggunakan *stratified splitting* untuk menjaga proporsi kelas sentimen.

Selanjutnya, pelatihan model dilakukan dengan mengimplementasikan dua algoritma klasifikasi: Support Vector Machine (SVM) dan Multinomial Naive Bayes, menggunakan data pelatihan. Terakhir, kedua model yang telah dilatih dievaluasi secara menyeluruh menggunakan metrik evaluasi seperti akurasi, presisi, *recall*, dan F1-Score, serta *confusion matrix*, untuk kemudian dilakukan perbandingan dan analisis hasil guna mengidentifikasi kinerja terbaik dan memberikan implikasi dari temuan yang ada.

2.2 Pengumpulan Data

Tahap pengumpulan data merupakan fondasi utama dalam penelitian analisis sentimen ini.

Data yang digunakan adalah ulasan pengguna yang berasal dari platform daring Google Maps, dengan fokus objek penelitian pada Telkom University Bandung. Ulasan-ulasan ini dipilih karena merefleksikan opini dan pengalaman langsung dari berbagai pihak, termasuk mahasiswa, alumni, maupun pengunjung, terhadap universitas.

Proses pengumpulan data dilakukan dengan pendekatan yang sistematis untuk memastikan cakupan ulasan yang relevan. Meskipun dalam kasus ini data ulasan telah tersedia di Kaggle dengan nama *TelU reviews* dan berisi sekitar 931 ulasan. Data yang diekstraksi meliputi teks ulasan itu sendiri, rating bintang yang diberikan pengguna, serta informasi lain seperti nama pengguna dan tanggal ulasan. Kualitas dan kuantitas data yang terkumpul menjadi krusial dalam menjamin representasi sentimen yang akurat dan relevansi analisis lebih lanjut.

Selanjutnya, dilakukan proses labeling data berdasarkan skor ulasan. Ulasan dengan skor 1, 2 dikategorikan sebagai sentimen negatif, ulasan dengan skor 3 dikategorikan sebagai sentimen netral, sedangkan ulasan dengan skor 4 dan 5 dikategorikan sebagai sentimen positif. Pendekatan klasifikasi biner ini dipilih untuk menyederhanakan analisis dan fokus pada perbedaan persepsi pengguna secara tegas antara positif, negatif, dan netral.

Setelah dilakukan proses pelabelan berdasarkan nilai skor ulasan, dataset terbagi ke dalam tiga kategori final sentimen, yaitu positif, negatif, dan netral. Ulasan dengan skor 4 dan 5 diklasifikasikan sebagai sentimen positif, sedangkan skor 1, 2 diklasifikasikan sebagai sentimen negatif, dan skor 3 diklasifikasikan sebagai sentimen netral. Berdasarkan hasil pelabelan, terdapat sebanyak 810 ulasan positif, 102 ulasan negatif, dan 19 ulasan netral. Distribusi ini menunjukkan bahwa mayoritas pengguna memberikan ulasan yang bersifat positif terhadap Universitas Telkom Bandung, sementara ulasan negatif jumlahnya lebih sedikit. Ketidakseimbangan jumlah data antar kelas ini menjadi perhatian dalam proses pemodelan, karena berpotensi mempengaruhi performa model klasifikasi sentimen yang digunakan.

2.3 Pra-pemrosesan Data

Tahap pra-pemrosesan data merupakan langkah krusial untuk mengubah data ulasan mentah yang tidak terstruktur menjadi format yang bersih, relevan, dan siap untuk analisis oleh model *machine learning*. Proses ini diawali dengan pembersihan data, di mana teks ulasan dinormalisasi dengan mengubah semua huruf menjadi huruf kecil, menghapus URL, angka, tanda baca, simbol, serta karakter yang tidak relevan lainnya, dan juga menyingkirkan spasi berlebih serta baris duplikat atau kosong. Selanjutnya, dilakukan pelabelan sentimen, yaitu proses pemberian label manual pada setiap ulasan ke dalam kategori 'Positif', 'Negatif', atau 'Netral', yang menjadi dasar 'ground truth' untuk pelatihan model. Setelah bersih dan terlabeli, teks ulasan kemudian melalui tokenisasi, di mana setiap kalimat dipecah menjadi unit-unit kata atau token. Token-token ini selanjutnya melewati penghapusan *stopwords*, yaitu proses eliminasi kata-kata umum dalam Bahasa Indonesia yang tidak memiliki kontribusi signifikan terhadap makna sentimen (misalnya, "yang", "dan", "di"). Tahap akhir pra-pemrosesan adalah *stemming*, menggunakan pustaka sastrawi untuk mereduksi setiap kata ke bentuk dasarnya (misalnya, "mengajar" menjadi "ajar", "terhormat" menjadi "hormat"). Seluruh proses ini memastikan bahwa data teks yang akan diekstraksi fiturnya adalah data yang berkualitas tinggi dan bebas dari 'noise', sehingga dapat meningkatkan akurasi dan efisiensi model klasifikasi sentimen.

2.4 TF-IDF (Term Frequency-Inverse Document Frequency)

Setelah melalui tahapan pra-pemrosesan data, langkah selanjutnya adalah mengubah data teks yang terstruktur menjadi representasi numerik yang dapat dipahami oleh algoritma *machine learning*. Metode yang digunakan untuk tujuan ini adalah TF-IDF. Teknik ini digunakan untuk menentukan tingkat kepentingan suatu kata dalam sebuah dokumen dibandingkan dengan keseluruhan kumpulan dokumen. Bobot TF-IDF diperoleh dengan mengalikan frekuensi kemunculan kata dalam dokumen tertentu (TF) dengan nilai kebalikan dari jumlah dokumen yang mengandung kata tersebut (IDF), yang mencerminkan

keunikannya[1]. Sebuah kata akan dianggap penting apabila memiliki skor TF-IDF yang tinggi. Perhitungan nilai TF-IDF untuk sebuah kata t dalam dokumen d dirumuskan sebagai berikut:

$$TF(t, d) = \frac{\text{Jumlah frekuensi kata 't' dalam dokumen 'd'}}{\text{Jumlah seluruh kata dalam dokumen 'd'}} \quad (1)$$

$$IDF(t) = \log \left(\frac{\text{Jumlah seluruh dokumen}}{\text{Jumlah dokumen yang mengandung kata 't'}} \right) \quad (2)$$

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

2.5 Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan algoritma pembelajaran *supervised* yang sangat efektif untuk masalah klasifikasi dan regresi[2]. Prinsip utama SVM adalah menemukan *hyperplane* optimal yang paling baik memisahkan kelas-kelas data dalam ruang berdimensi tinggi[3]. *Hyperplane* ini dipilih sedemikian rupa sehingga memaksimalkan *margin* (jarak terdekat) antara titik-titik data dari kelas yang berbeda, yang dikenal sebagai *support vectors*. Untuk data yang tidak dapat dipisahkan secara linier, SVM menggunakan *kernel trick* untuk memetakan data ke ruang berdimensi lebih tinggi di mana pemisahan linier menjadi mungkin[4]. Dalam konteks klasifikasi teks dengan fitur TF-IDF, *kernel* linear seringkali merupakan pilihan yang efektif dan efisien.

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

2.6 Naïve Bayes

Naive Bayes digunakan karena memiliki algoritma klasifikasi probabilistik sederhana yang didasarkan pada Teorema Bayes dengan asumsi kuat tentang independensi antar fitur. Meskipun asumsi ini seringkali tidak sepenuhnya benar di dunia nyata. Naive Bayes menunjukkan kinerja yang mengejutkan baiknya dalam tugas klasifikasi teks, terutama dengan dataset yang besar dan fitur yang banyak seperti dalam kasus TF-IDF[5]. Keunggulan utamanya adalah kesederhanaan, kecepatan pelatihan, dan efisiensi komputasi. Varian yang paling umum digunakan untuk klasifikasi dokumen adalah Multinomial Naive Bayes, yang cocok untuk fitur yang merepresentasikan hitungan frekuensi (seperti bobot TF-IDF). Model ini menghitung probabilitas bahwa suatu dokumen termasuk dalam kelas tertentu berdasarkan probabilitas kemunculan setiap kata dalam dokumen tersebut di setiap kelas[1], [2], [3], [4], [5], [6]. Rumus dasar teorema bayes:

$$P(C_k|x) \propto P(x|C_k)P(C_k) \quad (5)$$

2.7 Evaluasi

Tahap evaluasi model merupakan fase krusial dalam penelitian ini yang bertujuan untuk mengukur dan menilai kinerja model klasifikasi SVM dan Naive Bayes yang telah dilatih. Evaluasi dilakukan pada data pengujian (testing set) yang belum pernah dilihat oleh model selama proses pelatihan, sehingga memberikan indikasi yang objektif mengenai kemampuan generalisasi model terhadap data baru. Metrik evaluasi yang digunakan mencakup akurasi, presisi (*precision*), *recall*, dan F1-score, yang masing-masing memberikan perspektif berbeda tentang kinerja model. Dalam konteks klasifikasi multi-kelas, metrik-metrik ini dihitung untuk setiap kelas (Positif, Negatif, Netral). Definisi dasarnya adalah sebagai berikut:

- True Positive (TP): Jumlah data positif yang diprediksi dengan benar sebagai positif.

- True Negative (TN): Jumlah data negatif yang diprediksi dengan benar sebagai negatif.
- False Positive (FP): Jumlah data negatif yang salah diprediksi sebagai positif.
- False Negative (FN): Jumlah data positif yang salah diprediksi sebagai negatif.

Dari komponen-komponen *Matrix* tersebut, performa model diukur menggunakan empat metrik utama, akurasi, presisi, *recall*, dan *F1-score*. Perhitungan setiap metrik dilakukan dengan rumus sebagai berikut:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (6)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (7)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

$$\text{F1-Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (9)$$

Dalam implementasinya, semua nilai metrik ini dihitung secara otomatis menggunakan fungsi *accuracy_score* dan *classification_report* dari *library Scikit-learn*.

3. Hasil dan Diskusi

3.1. Pra-pemrosesan Data

Dataset yang digunakan dalam penelitian ini terdiri dari 931 ulasan Google Maps mengenai Universitas Telkom Bandung. Setelah dilakukan proses pembersihan data dan pelabelan sentimen, diperoleh distribusi kelas sentimen sebagaimana ditunjukkan pada Tabel 1.

Tabel 1. Distribusi Sentimen Dataset

Sentimen	Jumlah
Positif	810
Negatif	102
Netral	19

Distribusi ini mengindikasikan adanya ketidakseimbangan kelas yang signifikan, di mana kelas Positif merupakan mayoritas dominan. Meskipun demikian, pelabelan manual memastikan 'ground truth' yang akurat untuk pelatihan model. Selanjutnya, teks ulasan yang telah dibersihkan melalui proses tokenisasi, penghapusan stopwords menggunakan kamus stopwords Bahasa Indonesia, dan stemming dengan pustaka Sastrawi. Hasil dari ketiga proses ini adalah representasi teks yang lebih standar dan ringkas, mengurangi variasi kata dan menghilangkan "noise" yang tidak perlu, sehingga mempersiapkan data untuk ekstraksi fitur yang lebih efektif.

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

Tahap ekstraksi fitur bertanggung jawab untuk mengubah teks yang sudah diproses menjadi format numerik yang dapat diinterpretasikan oleh algoritma machine learning. Dalam penelitian ini, metode TF-IDF diaplikasikan pada kolom *processed_text*. Hasil dari proses TF-IDF adalah sebuah matriks fitur berukuran 931, 362 yang berarti 931 ulasan telah berhasil diubah menjadi

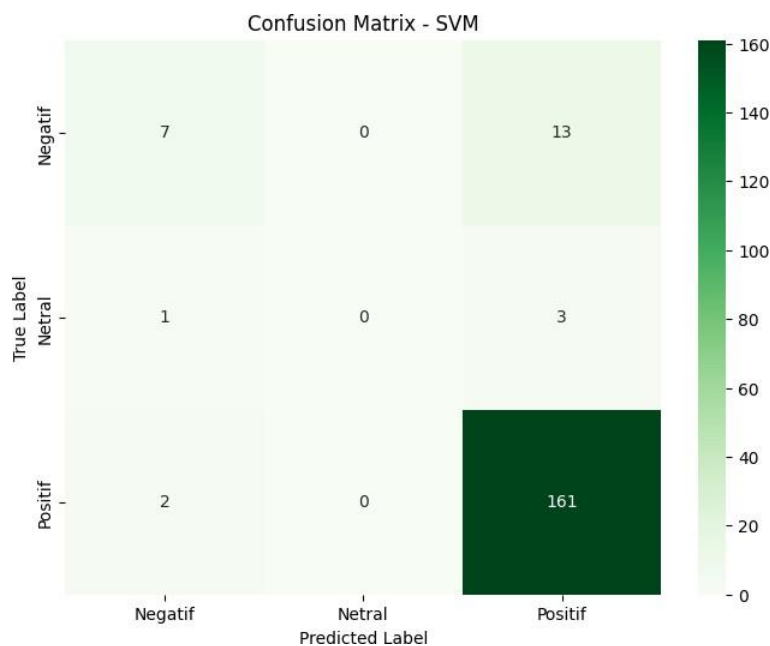
362 fitur numerik. Penggunaan parameter *max_features=5000* dan *min_df=5* pada *TfidfVectorizer* secara efektif membatasi kosakata yang dipertimbangkan, sehingga hanya kata-kata dan bigram yang paling relevan dan sering muncul (minimal di 5 dokumen) yang dipertahankan. Parameter *ngram_range=(1,2)* juga turut berkontribusi dengan menghasilkan fitur baik dari kata tunggal (unigram) maupun pasangan kata (bigram), yang terbukti penting dalam menangkap konteks sentimen, seperti frasa "tidak baik" yang memiliki makna berbeda dari sekadar kata "baik". Proses ekstraksi fitur ini berhasil mengonversi data tekstual menjadi representasi vektor padat yang siap untuk tahap pembagian data dan pelatihan model.

3.3. Support Vector Machine (SVM)

Model SVM diuji menggunakan data testing dengan rasio pembagian data 80:20. Akurasi keseluruhan model SVM sebesar 89.84%. Hasil evaluasi model ditunjukkan pada Tabel 2.

Tabel 2. Hasil Evaluasi SVM

Kelas	Precision	Recall	F1-Score
Negatif	0.70	0.35	0.47
Netral	0.00	0.00	0.00
Positif	0.91	0.99	0.95



Gambar 2. Confusion Matrix SVM

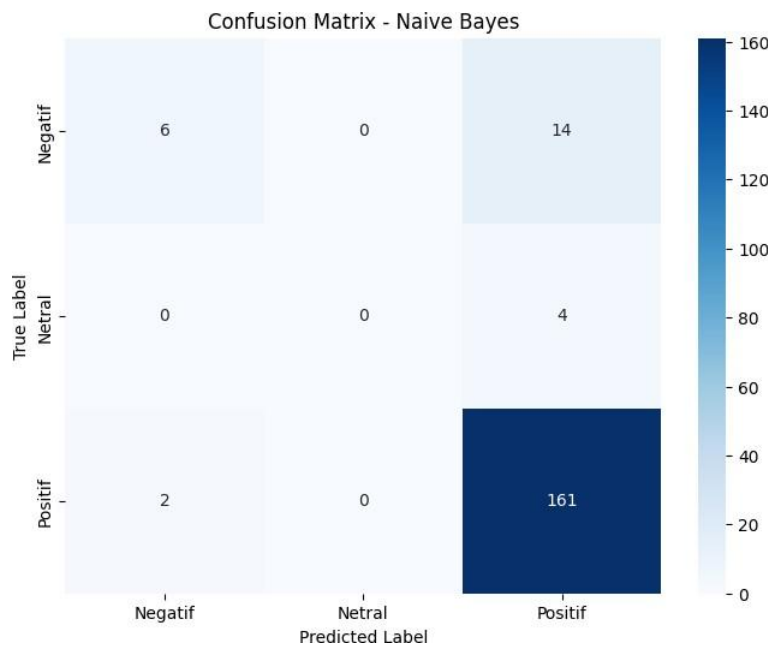
3.4. Pemodelan Naïve Bayes

Algoritma Multinomial Naive Bayes juga dilatih dan dievaluasi pada data yang sama. Model

Naive Bayes menghasilkan akurasi keseluruhan sebesar 89.30%, sedikit di bawah performa SVM. Serupa dengan SVM, Naive Bayes menunjukkan kinerja yang sangat kuat dalam mengklasifikasikan sentimen Positif, dengan *precision* 0.90, *recall* 0.99, dan *F1-score* 0.94. Ini kembali menegaskan kemudahan model dalam mengidentifikasi kelas mayoritas. Namun, pada kelas minoritas, Naive Bayes juga menunjukkan keterbatasan yang serupa dengan SVM. Untuk sentimen negatif, model ini mencapai *precision* 0.75 tetapi dengan *recall* yang lebih rendah yaitu 0.30, menghasilkan *F1-Score* 0.43.

Tabel 3. Hasil Evaluasi Naive Bayes

Kelas	Precision	Recall	F1-Score
Negatif	0.75	0.30	0.43
Netral	0.00	0.00	0.00
Positif	0.90	0.99	0.94



Gambar 3. Confusion Matrix Naïve Baye

4. Pembahasan

4.1. Analisis Kinerja Model

Berdasarkan hasil pengujian, model SVM memperoleh akurasi lebih tinggi dibandingkan Naive Bayes, yaitu sebesar 89.84%. Hal ini menunjukkan bahwa SVM memiliki kemampuan yang lebih baik dalam memisahkan data sentimen berdasarkan fitur TF-IDF. Kedua model menunjukkan performa yang sangat baik pada kelas positif. Tingginya nilai *recall* pada kelas positif menunjukkan bahwa sebagian besar ulasan positif berhasil dikenali dengan benar oleh model.

4.2. Pengaruh Ketidakseimbangan Data

Distribusi data yang tidak seimbang menjadi faktor utama yang mempengaruhi performa klasifikasi. Jumlah data positif yang jauh lebih banyak dibandingkan data negatif dan netral menyebabkan model cenderung memprediksi kelas mayoritas. Kondisi ini terlihat pada kelas netral yang tidak berhasil diprediksi oleh kedua model, ditandai dengan nilai *precision*, *recall*,

dan F1-score sebesar 0.00. Jumlah data netral yang sangat sedikit menyebabkan model kesulitan mempelajari pola karakteristik sentimen netral.

4.3. Perbandingan SVM dan Naïve Bayes

Model SVM menunjukkan performa yang lebih baik dibandingkan Naive Bayes terutama pada recall kelas negatif. Hal ini menunjukkan bahwa SVM lebih mampu mengenali ulasan negatif dibandingkan Naive Bayes. Sementara itu, Naive Bayes memiliki precision yang sedikit lebih tinggi pada kelas negatif. Namun recall yang rendah menunjukkan bahwa masih banyak ulasan negatif yang gagal dikenali. Secara umum, SVM lebih cocok digunakan pada penelitian ini karena mampu memberikan performa klasifikasi yang lebih stabil pada data teks berbasis TF-IDF.

5. Kesimpulan

Penelitian ini telah berhasil mengimplementasikan sistem analisis sentimen untuk mengklasifikasikan ulasan pengguna dari Google Maps ke dalam kategori sentimen positif, negatif, dan netral, menggunakan algoritma *Support Vector Machine* (SVM) dan Multinomial Naive Bayes. Melalui serangkaian tahapan mulai dari pra-pemrosesan data yang komprehensif (pembersihan teks, tokenisasi, penghapusan stopwords, dan stemming), ekstraksi fitur menggunakan TF-IDF, hingga pelatihan dan evaluasi model, kedua algoritma menunjukkan kemampuan yang signifikan dalam klasifikasi sentimen. Secara keseluruhan, kedua model mencapai akurasi yang tinggi (SVM sekitar 89.84% dan Naive Bayes sekitar 89.30%). Keduanya sangat efektif dalam mengklasifikasikan sentimen Positif, yang merupakan kelas mayoritas, dengan metrik presisi, recall, dan F1-Score yang sangat tinggi. Namun, evaluasi mendalam menunjukkan tantangan signifikan pada klasifikasi kelas minoritas. Model SVM dan Naive Bayes sama-sama menunjukkan kinerja yang sangat rendah pada kelas Negatif (terutama dari sisi recall) dan gagal total dalam memprediksi kelas Netral, yang terindikasi dari nilai presisi, recall, dan F1-Score sebesar 0.00 untuk kelas ini. Disparitas kinerja yang mencolok antara kelas mayoritas dan minoritas ini sebagian besar disebabkan oleh ketidakseimbangan kelas (*class imbalance*) dalam dataset ulasan. Meskipun demikian, penelitian ini berhasil menunjukkan kapabilitas algoritma machine learning dalam memahami sentimen dari data teks tidak terstruktur dan menyoroti perlunya strategi penanganan *class imbalance* yang lebih canggih untuk tugas klasifikasi multi-kelas dengan distribusi data yang tidak merata. Hasil ini memberikan wawasan awal mengenai persepsi pengguna dan dapat menjadi dasar bagi pengembangan sistem analisis sentimen yang lebih robust di masa mendatang.

Daftar Pustaka

- [1] M. Hafizh Mahendra, D. Triantoro Murdiansyah, and K. Muslim Lhaksmana, "Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer," Aug. 2023.
- [2] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.
- [3] A. S. Nugroho, A. B. Witarto, and D. Handoko, "Support Vector Machine-Teori dan Aplikasinya dalam Bioinformatika 1," 2003. [Online]. Available: <http://asnugroho.net>
- [4] T. Meisya, N. Arifin, and R. Mayasari, "Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinasi Covid-19," 2021, [Online]. Available: <https://doi.org/10.31598>
- [5] J. Jefriyanto, N. Ainun, and M. Arif, "Application of Naïve Bayes Classification to Analyze Performance Using Stopwords," *Journal of Information System*, vol. 1, no. 2, pp. 49–53, 2023, [Online]. Available: <http://gemapublisher.com/index.php/jiste>
- [6] Astari Ni Made Ayu Juli, Dewa Gede Hendra Divayana, and Gede Indrawan, "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier," *Jurnal Sistem dan Informatika (JSI)*, vol. 15, no. 1, pp. 27–29, Nov. 2020, doi: 10.30864/jsi.v15i1.332.