

ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN HOTEL XYZ DI KOTA TANGERANG DENGAN ALGORITMA SVM

¹⁾Yamka Sudirman, ²⁾Dwi Sartika Simatupang

^{1,2)}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Esa Unggul

^{1,2)}yamkasudirman@gmail.com, Dwi.sartika@esaunggul.ac.id

INFO ARTIKEL

Riwayat Artikel :

Diterima : 16 Oktober 2025

Disetujui : 20 Oktober 2025

Kata Kunci : analisis sentimen, support vector machine, web scraping, tf-idf

ABSTRAK

Analisis sentimen berbasis aspek dilakukan untuk mengklasifikasikan opini pelanggan terhadap Hotel XYZ di Kota Tangerang pada aspek fasilitas, kamar, kebersihan, dan pelayanan. Data ulasan dikumpulkan melalui web scraping, di labeli menggunakan metode lexicon-based, dan di representasikan secara numerik dengan TF-IDF. Ketidakseimbangan data di atasi menggunakan SMOTE, dan klasifikasi dilakukan dengan algoritma Support Vector Machine dengan pengujian parameter $C = 1 – 5$. Hasil eksperimen menunjukkan F1-score tertinggi pada aspek fasilitas 90%, diikuti kebersihan 83%, kamar 79%, dan pelayanan 67%, dengan parameter $C = 3$ memberikan performa optimal pada sebagian besar aspek.

ARTICLE INFO

Article History :

Received : Oct 16, 2025

Accepted : Oct 20, 2025

Keywords:

sentiment analysis, support vector machine, web scraping, tf-idf

ABSTRACT

Aspect-based sentiment analysis was conducted to classify customer opinions towards Hotel XYZ in Tangerang City on the aspects of facilities, rooms, cleanliness, and service. Review data were collected through web scraping, labeled using a lexicon-based method, and represented numerically with TF-IDF. Data imbalance was addressed using SMOTE, and classification was performed using the Support Vector Machine algorithm with testing parameters $C = 1 – 5$. The experimental results showed the highest F1-score in the facility aspect of 90%, followed by cleanliness 83%, rooms 79%, and service 67%, with parameter $C = 3$ providing optimal performance in most aspects.

1. PENDAHULUAN

Industri perhotelan di era digital menghadapi tantangan besar dalam mengelola reputasi *online*, terutama melalui ulasan pelanggan di platform seperti *Google Maps*, *Tiket.com*, dan *Traveloka*. Ulasan pelanggan dapat memengaruhi persepsi calon tamu dan berdampak pada keputusan mereka dalam memilih hotel. Jika pelanggan memberikan ulasan positif, maka hotel berpotensi meningkatkan jumlah reservasi dan loyalitas pelanggan (Gabbard, 2023). Sebaliknya, ulasan negatif dapat menurunkan reputasi hotel dan berdampak pada pendapatan jangka panjang. Fenomena ini semakin relevan bagi hotel - hotel di kota tangerang, termasuk hotel XYZ kota tangerang, yang beroperasi di lingkungan dengan persaingan ketat.

Dalam memahami persepsi pelanggan, analisis sentimen berbasis aspek menjadi pendekatan yang lebih efektif dibandingkan analisis sentimen biasa (Morama et al., 2022). Pendekatan ini tidak hanya menentukan apakah ulasan bersifat positif atau negatif, tetapi juga mengidentifikasi aspek spesifik yang dikomentari pelanggan, seperti fasilitas, kebersihan, kamar, dan pelayanan (Dewia et al., 2023). Dengan demikian, manajemen hotel dapat memahami kekuatan dan kelemahan mereka secara lebih mendetail dan mengambil keputusan strategis yang lebih tepat.

Untuk melakukan analisis ini, data ulasan hotel dikumpulkan melalui *web scraping* menggunakan *SerpAPI* dan *Web Scraper*, yang memungkinkan ekstraksi ulasan secara otomatis dari website *Google Maps*, *Traveloka*, dan *Tiket.com* (Prasetyo et al., 2023). Data yang diperoleh akan dikategorikan berdasarkan aspek yang dibahas dalam ulasan, kemudian di lakukan tahap *preprocessing* untuk membersihkan data sebelum di analisis lebih lanjut.

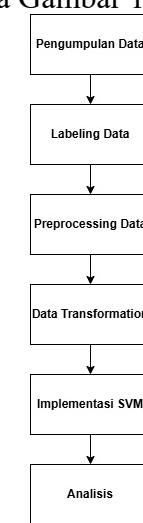
Dalam penelitian ini, algoritma *Support Vector Machine* digunakan untuk mengklasifikasikan sentimen ulasan berdasarkan aspek fasilitas, kebersihan, kamar, dan pelayanan. Pemilihan SVM didasarkan pada penelitian sebelumnya yang berjudul "*Sentiment Analysis of Hotel Reviews Using Machine Learning Techniques*" (Anis et al., 2021). Penelitian tersebut membandingkan performa lima algoritma *machine learning*, yaitu *Naïve*

Bayes, *K- Nearest Neighbor*, *SVM*, *Logistic Regression*, dan *Random Forest*, dalam menganalisis sentimen ulasan hotel. Hasil penelitian menunjukkan bahwa SVM memiliki performa tertinggi dengan akurasi sebesar 0.863, sehingga menjadikannya algoritma yang lebih andal dalam tugas klasifikasi sentimen ulasan hotel (Yustihan et al., 2021). Berdasarkan temuan tersebut, SVM dianggap sebagai pilihan terbaik untuk menganalisis sentimen berbasis aspek dalam penelitian ini.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengidentifikasi sentimen ulasan pelanggan terhadap hotel XYZ kota tangerang berdasarkan aspek fasilitas, kebersihan, kamar, dan pelayanan menggunakan algoritma SVM (Maulana et al., 2023). Hasil klasifikasi sentimen berbasis aspek dalam penelitian ini di harapkan dapat membantu manajemen hotel dalam memahami opini pelanggan secara lebih mendetail dan meningkatkan kualitas layanan mereka berdasarkan aspek- aspek yang paling berpengaruh terhadap kepuasan pelanggan.

2. METODE

Penelitian ini menggunakan algoritma *Support Vector Machine* untuk melakukan klasifikasi sentimen terhadap ulasan hotel (Radiena & Nugroho., 2023). Metodologi penelitian terdiri dari enam tahapan, yaitu: pengumpulan data, *labeling data*, *preprocessing data*, *data transformation*, implementasi SVM, dan analisis. Diagram alur metodologi penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Data ulasan dikumpulkan dari website *GoogleMaps*, *Tiket.com*, dan *Traveloka* menggunakan metode *SerpAPI* dan *Web Scraper* (Mitchell., 2024). Teknik ini digunakan untuk mengotomatisasi proses ekstraksi ulasan pengguna yang berkaitan dengan layanan hotel dari berbagai sumber daring. Sebagai ilustrasi, beberapa contoh ulasan yang berhasil dikumpulkan ditampilkan pada Tabel 1 berikut.

Tabel 1. Contoh Data Ulasan

Contoh Data Ulasan

Biaya tambahan breakfast untuk anak di bawah 12 tahun sebaiknya tidak terlalu mahal.
Kamar bersih , bantal empuk . Dan staff yg ramah
Hotel nyaman dan bersih, gak menyesal deh staycation di sini. Selalu langganan di hotel ini. Kamar bersih, toilet bersih, staf nya ramah. Lokasi nya ok.
Hotel ya nyaman, dekat dengan tol. Wangi, bersih.

2.2. Labeling Data

Proses *labeling* dilakukan menggunakan metode *lexicon based* (Muhammad et al., 2022). Metode ini melibatkan penggunaan kamus InSet (*Indonesia Sentiment Lexicon*) yang berisi 3.609 kata positif dan 6.609 kata negatif, dengan tujuan mengelompokkan ulasan ke dalam kelas positif atau negatif. Ulasan juga diklasifikasikan ke dalam empat aspek, yaitu fasilitas, kamar, kebersihan, dan pelayanan, berdasarkan kemunculan kata kunci tertentu yang mewakili masing-masing aspek tersebut. Daftar aspek beserta kata kunci yang digunakan untuk mengelompokkan ulasan ditampilkan pada Tabel 2.

Tabel 2. Daftar Aspek

Aspek	Kata Kunci	Deskripsi
Fasilitas	wifi, tv, kolam renang, gym, lift, sarapan, breakfast, parkir, lounge, air panas, shuttle, bar	Aspek fasilitas mencakup layanan penunjang hotel seperti wifi, sarapan, gym, shuttle, dan parkir.
	Kamar, ranjang, kasur, tempat tidur, toilet, ac, bathroom, shower, lemari, jendela, lampu, tv	Aspek kamar mencakup kenyamanan, luas, kelengkapan, dan kondisi kamar.
Kebersihan	bersih, kebersihan, higienis, rapih, debu, bau, kotor, sampah	Aspek kebersihan mencakup kondisi kamar, kamar mandi, linen, serta area umum hotel yang bersih.
	pelayanan, staf, service, resepsiionis, ramah , helpful, sopan, cepat tanggap, slow response, senyum, staff	Aspek pelayanan mencakup sikap staf, kecepatan layanan, keramahan, dan respons terhadap kebutuhan tamu.

2.3. Preprocessing Data

Data ulasan hotel yang telah dikumpulkan akan melalui tahapan *preprocessing* yang mencakup lima langkah utama (Jafari., 2022). Diantaranya adalah pembersihan teks dari angka, tanda baca, atau simbol (*clean text*), perubahan huruf besar ke huruf kecil (*casefolding*), membagi teks menjadi token (*tokenization*), menghapus kata-kata umum yang tidak memiliki nilai signifikan dalam analisis (*removing stopwords*), menghapus imbuhan seperti awalan, akhiran, atau sisipan dari sebuah kata (*stemming*) (Fauziah et al., 2021).

2.4. Data Transformation

Setelah data ulasan melalui tahap *preprocessing*, langkah berikutnya adalah melakukan *data transformation* sebagai *feature extraction* menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF). TF-IDF mengubah setiap kata dalam ulasan menjadi bobot numerik, sehingga kata-kata yang signifikan dalam satu ulasan namun jarang muncul di ulasan lain diberi bobot lebih tinggi, meningkatkan relevansi fitur yang dianalisis (Zhou & Chen., 2021). Dalam penerapannya, digunakan parameter $max_df = 0.5$ untuk mengecualikan kata yang terlalu sering muncul di banyak ulasan, dan $min_df = 2$ untuk mengabaikan kata yang muncul sangat jarang (Jain et al., 2024). Dengan representasi numerik yang lebih informatif ini, model klasifikasi dapat fokus pada kata-kata kunci yang membedakan kategori, sehingga prediksi sentimen menjadi lebih akurat dan hasil analisis lebih dapat direplikasi.

2.5. Implementasi SVM

Setelah data melalui proses *data transformation* menggunakan TF-IDF, langkah berikutnya adalah membangun model menggunakan algoritma *Support Vector Machine* (SVM). Data yang telah di transformasi dibagi menjadi 80% data latih dan 20% data uji, dengan *random_state*=16 untuk memastikan hasil pembagian yang konsisten. Model SVM dilatih menggunakan kernel linear, dengan pengujian beberapa nilai parameter C (1–5) untuk menemukan konfigurasi yang memberikan performa terbaik (Makthum & Muhamid., 2023). Evaluasi model dilakukan menggunakan *confusion matrix* serta metrik

klasifikasi *accuracy*, *precision*, *recall*, dan *F1-score*.

2.6. Analisis

Tahap analisis dilakukan terhadap model *Support Vector Machine* dalam mengklasifikasikan sentimen ulasan hotel berdasarkan aspek seperti fasilitas, kebersihan, kamar, dan pelayanan. Selain itu, dilakukan juga analisis hasil pengujian model berdasarkan *hyperparameter* dan metrik evaluasi untuk menilai kualitas model yang dihasilkan.

3. HASIL DAN PEMBAHASAN

3.1. Pembagian Aspek

Hasil scraping menghasilkan sekitar 3.930 data ulasan hotel dengan aspek yang berbeda - beda. Setelah dikategorikan berdasarkan aspek, jumlah data berkembang menjadi 4.173. Rincian distribusinya meliputi 1.029 data pada aspek kamar, 1.213 data pada aspek pelayanan, 986 data pada aspek kebersihan, dan 945 data pada aspek fasilitas. Tabel 3 menunjukkan jumlah ulasan yang dilakukan untuk setiap aspek, serta rasio sentimen positif dan negatif.

Tabel 3. Distribusi Sentimen

Aspek	Positif	Negatif	Total
Kamar	904	125	1029
Pelayanan	1146	67	1213
Kebersihan	890	96	986
Fasilitas	631	314	945
Total	3571	602	4173

Berdasarkan distribusi sentimen pada Tabel 3, terlihat bahwa aspek fasilitas merupakan satu - satunya aspek yang memiliki distribusi sentimen yang relatif mendekati seimbang, yaitu terdiri dari 631 ulasan positif dan 314 ulasan negatif. Sementara itu, pada aspek lainnya seperti kamar, pelayanan, dan kebersihan, terjadi ketimpangan distribusi yang cukup ekstrem, di mana jumlah ulasan positif jauh lebih dominan dibandingkan ulasan negatif. Ketidakseimbangan kelas seperti ini berpotensi mengurangi performa algoritma *Support Vector Machine*, karena model dapat menjadi bias terhadap kelas mayoritas dan mengalami penurunan akurasi dalam mengklasifikasikan data dari kelas minoritas.

3.2. Synthetic Minority Oversampling Technique (SMOTE)

Pada tahap ini dikarenakan pelabelan data menggunakan pendekatan *lexicon based* menghasilkan ketidakseimbangan kelas, di mana jumlah data berlabel positif jauh lebih tinggi

dibandingkan negatif yang terlihat pada Tabel 4. Ketidakseimbangan ini dapat memengaruhi performa dan akurasi model, karena model cenderung bias terhadap kelas mayoritas. Oleh karena itu, dilakukan teknik *oversampling* untuk meningkatkan proporsi data pada kelas minoritas. Salah satu metode *oversampling* yang digunakan adalah SMOTE (*Synthetic Minority Over-sampling Technique*), yang bertujuan untuk menyeimbangkan jumlah data antar kelas dengan cara menghasilkan data sintetis pada kelas minoritas. SMOTE ini hanya diterapkan pada data pelatihan, bukan pada data pengujian.

3.3. Uji Parameter

Dilakukan skenario pengujian terhadap algoritma *Support Vector Machine* dengan menggunakan *kernel linear* dan variasi nilai parameter regulasi C dari 1 hingga 5. Pengujian ini diterapkan pada masing-masing aspek ulasan, yaitu aspek kamar, pelayanan, kebersihan, dan fasilitas. Aspek fasilitas dipilih sebagai representasi data dengan distribusi sentimen yang mendekati seimbang, sementara aspek kamar, pelayanan, dan kebersihan dipilih karena memiliki ketidakseimbangan kelas yang cukup signifikan. Tujuan dari pengujian ini adalah untuk mengevaluasi pengaruh nilai parameter C terhadap performa klasifikasi SVM, serta melihat sejauh mana ketidakseimbangan data memengaruhi hasil klasifikasi pada tiap aspek. Hasil dari uji coba ini akan di rata - ratakan dari masing - masing aspek.

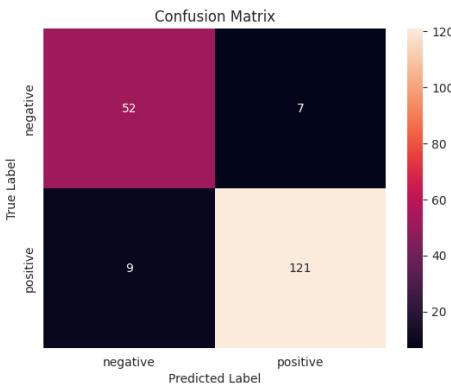
3.3.1. Aspek Fasilitas

Aspek fasilitas hotel diuji menggunakan ulasan pengunjung yang mencakup tanggapan tentang kelengkapan, ketersediaan, kenyamanan, fungsi, dan aksesibilitas fasilitas. Kinerja klasifikasi sentimen pada aspek ini dinilai menggunakan hasil *confusion matrix* dari algoritma SVM, dengan memanfaatkan parameter nilai C dan *kernel linear*. Detail pengujian tertera pada Tabel 4.

Tabel 4. Hasil Pengujian Aspek Fasilitas

C	Parameter Nilai C			
	Accuracy	Precision	Recall	F1-Score
1	92%	90%	90%	90%
2	92%	90%	90%	90%
3	92%	90%	91%	90%
4	91%	89%	90%	90%
5	91%	90%	90%	90%

Berdasarkan Tabel 4, dapat diamati bahwa nilai $C=3$ memberikan hasil performa terbaik untuk aspek fasilitas. Untuk melihat distribusi hasil klasifikasi secara lebih *detail*, berikut disajikan *confusion matrix* yang dihasilkan oleh algoritma SVM dengan parameter optimal tersebut pada Gambar 2.



Gambar 2. *Confusion Matrix* Fasilitas

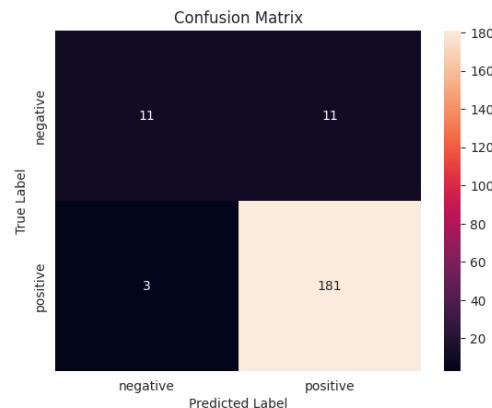
3.3.2. Aspek Kamar

Pengujian aspek kamar hotel melibatkan analisis ulasan pengunjung yang membahas kondisi, kenyamanan, kelengkapan fasilitas, dan kesesuaian informasi kamar. Kinerja klasifikasi sentimen untuk aspek ini dievaluasi menggunakan *confusion matrix* dari algoritma SVM, yang dikonfigurasi dengan parameter nilai C dan *kernel linear*. Hasil pengujian ini disajikan lengkap dalam Tabel 5.

Tabel 5. Hasil Pengujian Aspek Kamar

Parameter Nilai C				
C	Accuracy	Precision	Recall	F1-Score
1	91%	78%	71%	74%
2	93%	84%	74%	78%
3	93%	86%	74%	79%
4	93%	84%	74%	78%
5	93%	86%	74%	79%

Berdasarkan Tabel 5, dapat diamati bahwa nilai $C=3$ memberikan hasil performa terbaik untuk aspek kamar. Untuk melihat distribusi hasil klasifikasi secara lebih *detail*, berikut disajikan *confusion matrix* yang dihasilkan oleh algoritma SVM dengan parameter optimal tersebut pada Gambar 3.



Gambar 3. *Confusion Matrix* Kamar

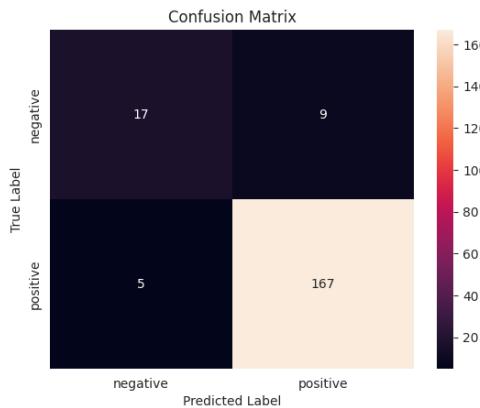
3.3.3. Aspek Kebersihan

Pengujian aspek kebersihan hotel melibatkan analisis ulasan pengunjung yang membahas tingkat kebersihan di berbagai area, seperti kamar, kamar mandi, lobi, dan koridor. Perlu diketahui bahwa persepsi kebersihan seringkali berkorelasi dengan kesan pertama pengunjung terhadap kualitas pemeliharaan hotel. Guna mengevaluasi performa klasifikasi sentimen pada aspek ini, digunakan hasil *confusion matrix* yang diperoleh dari algoritma SVM, dengan memanfaatkan parameter nilai C dan *kernel linear*. Detail pengujian dapat dilihat pada Tabel 6.

Tabel 6. Hasil Pengujian Aspek Kebersihan

C	Parameter Nilai C			
	Accuracy	Precision	Recall	F1-Score
1	93%	86%	81%	83%
2	91%	86%	72%	77%
3	91%	86%	72%	77%
4	92%	87%	74%	79%
5	91%	85%	70%	75%

Berdasarkan Tabel 6, dapat diamati bahwa nilai $C=1$ memberikan hasil performa terbaik untuk aspek kebersihan. Untuk melihat distribusi hasil klasifikasi secara lebih *detail*, berikut disajikan *confusion matrix* yang dihasilkan oleh algoritma SVM dengan parameter optimal tersebut pada Gambar 4.



Gambar 4. *Confusion Matrix* Kebersihan

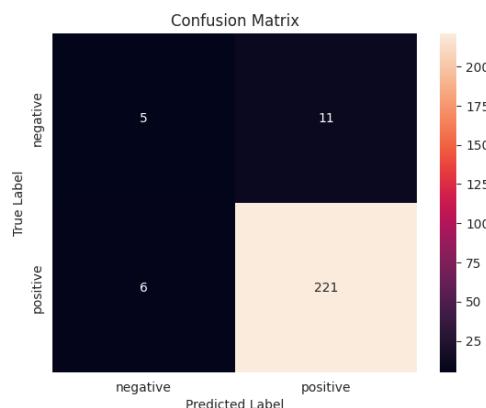
3.3.4. Aspek Pelayanan

Pengujian performa klasifikasi sentimen pada aspek pelayanan hotel dilakukan dengan menganalisis ulasan pengunjung yang membahas sikap, responsivitas, dan profesionalitas staf dalam melayani tamu. Pelayanan memegang peranan penting dalam kepuasan tamu secara keseluruhan, sebab melibatkan interaksi personal yang signifikan. Evaluasi performa klasifikasi ini memanfaatkan hasil *confusion matrix* yang diperoleh dari algoritma SVM, menggunakan parameter nilai C dan *kernel linear*. Rincian hasil pengujian dipaparkan pada Tabel 7.

Tabel 7. Hasil Pengujian Aspek Pelayanan

C	Parameter Nilai C			
	Accuracy	Precision	Recall	F1-Score
1	93%	67%	61%	63%
2	93%	70%	61%	64%
3	93%	70%	64%	67%
4	93%	73%	65%	67%
5	93%	73%	65%	67%

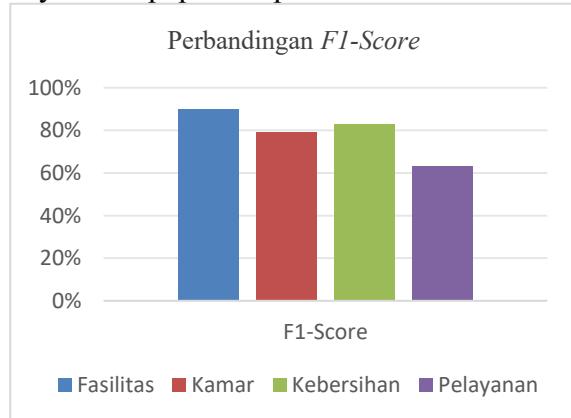
Berdasarkan Tabel 7, dapat diamati bahwa nilai C=3 memberikan hasil performa terbaik untuk aspek pelayanan. Untuk melihat distribusi hasil klasifikasi secara lebih *detail*, berikut disajikan *confusion matrix* yang dihasilkan oleh algoritma SVM dengan parameter optimal tersebut pada Gambar 5.



Gambar 5. *Confusion Matrix* Pelayanan

3.3.5. Perbandingan *F1-Score* terhadap Proporsi sentimen

Dilakukan perbandingan antara nilai *F1-Score* yang dihasilkan model pada masing-masing aspek terhadap proporsi sentimen dalam data ulasan. Pemilihan *F1-Score* sebagai metrik utama didasarkan pada kondisi distribusi data yang tidak seimbang, dimana akurasi cenderung tidak merepresentasikan kemampuan model dalam mengklasifikasikan kedua kelas secara adil. *F1-Score* dinilai lebih relevan karena mencerminkan keseimbangan antara *precision* dan *recall*, khususnya dalam menangani kelas minoritas seperti sentimen negatif. *Chart* hasil pengujian nilai *F1-Score* pada aspek fasilitas, aspek kamar, aspek kebersihan, dan aspek pelayanan dipaparkan pada Gambar 6.



Gambar 6. Perbandingan *F1-Score*

3.4. Analisis Hasil Pengujian

Analisis hasil pengujian dilakukan terhadap keempat aspek utama, yaitu fasilitas, kamar, kebersihan, dan pelayanan. Hasil pada beberapa aspek menunjukkan performa model yang cenderung lebih rendah, yang kemungkinan dipengaruhi oleh ketimpangan distribusi sentimen dalam data ulasan. Evaluasi dalam pengujian ini hanya berfokus pada satu metrik utama, yakni *F1-Score*, yang digunakan untuk

menilai kemampuan model dalam mengklasifikasikan sentimen secara seimbang pada data yang tidak merata antar kelas. Pada aspek fasilitas, performa model cenderung stabil dengan *F1-Score* konsisten di angka 90%, menunjukkan bahwa perubahan nilai C tidak memberikan pengaruh signifikan terhadap peningkatan kinerja model.

Pada aspek kamar, peningkatan performa terjadi pada $C = 2$ dan $C = 3$ dengan *F1-Score* mencapai 78%–79%, dibandingkan nilai terendah pada $C = 1$ yaitu 74%. Hal ini menunjukkan bahwa penyesuaian parameter C berdampak pada peningkatan hasil klasifikasi pada aspek ini. Untuk aspek kebersihan, performa terbaik terjadi pada $C = 1$ dan $C = 4$, masing-masing menghasilkan *F1-Score* 83% dan 79%, sedangkan nilai terendah berada pada $C = 5$ dengan *F1-Score* sebesar 75%, yang mengindikasikan adanya fluktuasi sensitivitas model terhadap variasi parameter C.

Sementara itu, pada aspek pelayanan, terlihat adanya peningkatan *F1-Score* dari 63% pada $C = 1$ menjadi 67% pada $C = 3$ hingga $C = 5$, yang menunjukkan bahwa parameter C memberikan pengaruh terhadap performa klasifikasi. Meskipun demikian, performa klasifikasi pada aspek ini tetap lebih rendah dibanding aspek lainnya. Hal ini dapat dijelaskan secara teoritis melalui distribusi data ulasan yang digunakan. Berdasarkan Tabel 3, rasio sentimen positif dan negatif pada aspek pelayanan sangat timpang, dengan 1.146 ulasan positif dibandingkan hanya 67 ulasan negatif dari total 1.213 ulasan. Ketimpangan jumlah ini menyebabkan model cenderung bias terhadap kelas mayoritas, sehingga mengurangi kemampuan klasifikasi untuk mendeteksi sentimen negatif secara akurat, yang berimplikasi pada *F1-Score* yang lebih rendah.

4. PENUTUP

4.1. Kesimpulan

Berdasarkan hasil pengujian dan analisis yangtelah dilakukan terhadap empat aspek ulasanhotel, yaitu fasilitas, kamar, kebersihan, dan pelayanan, dapat disimpulkan bahwa performa model *Support Vector Machine* (SVM) dipengaruhi oleh nilai parameter regulasi C yang digunakan. Pengujian menggunakan nilai C dari 1 hingga 5 menunjukkan bahwa nilai C = 3 secara

umum memberikan hasil *F1-Score* terbaik, terutama pada aspek kamar dan pelayanan yang memiliki distribusi sentimen tidak seimbang.

Pada aspek fasilitas, variasi nilai C tidak memberikan pengaruh signifikan karena nilai *F1-Score* cenderung stabil di semua parameter. Sebaliknya, aspek kebersihan justru menunjukkan hasil terbaik pada C yang lebih rendah, yaitu $C = 1$, yang mengindikasikan bahwa model dengan regulasi ringan lebih sesuai.

Secara keseluruhan, penggunaan metrik *F1-Score* dalam evaluasi terbukti efektif dalam mengukur performa model pada data ulasan yang tidak seimbang, dan nilai parameter $C = 3$ dapat dijadikan sebagai parameter optimal untuk membangun model klasifikasi sentimen yang seimbang dan akurat pada sebagian besar aspek yang diuji.

4.2. Saran

Sebagai saran untuk penelitian selanjutnya, disarankan untuk menguji algoritma klasifikasi lain selain *Support Vector Machine*, seperti *Random Forest*, *Naive Bayes*, atau *Logistic Regression*, agar dapat dibandingkan performanya terhadap data ulasan dengan karakteristik serupa. Perbandingan ini bertujuan untuk mengetahui sejauh mana efektivitas SVM dalam menangani data dengan distribusi sentimen yang tidak seimbang, serta memberikan gambaran yang lebih komprehensif terhadap pemilihan model terbaik dalam konteks analisis sentimen pada ulasan hotel.

5. DAFTAR PUSTAKA

- Gabbard, D. (2023). The impact of online reviews on hotel performance. *Journal of Modern Hospitality*, 2(1), 26-36.
- Morama, H. C., Ratnawati, D. E., & Arwani, I. (2022). Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(4), 1702-1708.
- Dewia, N. P. A., ERa, N. A. S., Karyawatia, A. E., Made, I. B., Mahendraa, I. B. G. D., & Wibawaa, I. G. A. (2023). Analisis Sentimen Berbasis Aspek Ulasan Pelanggan Hotel di Bali Menggunakan Metode

- Decision Tree. *Jurnal Elektronik Ilmu Komputer Udayana p-ISSN*, 2301, 5373.
- Prasetyo, M. R., & Fahrurrozi, A. (2023). Analisa Sentimen Pada Ulasan Google Untuk Hotel Gran Mahakam Jakarta Menggunakan Pendekatan Machine Learning. *Jurnal Ilmiah Informatika Komputer*, 28(3), 203-217.
- Anis, S., Saad, S., & Aref, M. (2020, September). Sentiment analysis of hotel reviews using machine learning techniques. In *International Conference on Advanced Intelligent Systems and Informatics* (pp. 227-234). Cham: Springer International Publishing.
- Rahma Yustihan, S., & Pandu Adikara, P. (2021). Analisis Sentimen berbasis Aspek terhadap Data Ulasan Rumah Makan menggunakan Metode Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(3), 1017-1023.
- Maulana, I., Apriandari, W., & Pambudi, A. (2023). Analisis Sentimen Berbasis Aspek Terhadap Ulasan Aplikasi Mypertamina Menggunakan Support Vector Machine. *Idealis: Indonesia Journal Information System*, 6(2), 172-181.
- Radiena, G., & Nugroho, A. (2023). Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Kai Access Menggunakan Metode Support Vector Machine. *Jurnal Pendidikan Teknologi Informasi (JUKANTI)*, 6(1), 1-10.
- Mitchell, R. (2024). *Web scraping with python*. " O'Reilly Media, Inc.".
- Jafari, R. (2022). *Hands-On Data Preprocessing in Python: Learn how to effectively prepare data for successful data analytics*. Packt Publishing Ltd.
- Muhammad, R. H., Laksana, T. G., & Arifa, A. B. (2022). Combination of support vector machine and lexicon-based algorithm in twitter sentiment analysis. *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, 8(1), 59-71.
- Zhou, Z., & Chen, Y. (2021). Svm barrage classification algorithm based on TF-IDF weighting and Its improvement. *World Scientific Research Journal*, 7(7), 142-151.
- Jain, S., Jain, S. K., & Vasal, S. (2024, April). An effective TF-IDF model to improve the text classification performance. In *2024 IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 1-4). IEEE.
- Fauziah, Y., Yuwono, B., & Aribowo, A. S. (2021, December). Lexicon based sentiment analysis in Indonesia languages: A systematic literature review. In *RSF Conference Series: Engineering and Technology* (Vol. 1, No. 1, pp. 363-367).
- Makhtum, A. R., & Muhajir, M. (2023). Sentiment analysis of omnibus law using support vector machine (SVM) with linear kernel. *BAREKENG J. Ilmu Mat. dan Terap*, 17(4), 2197-2206.