

Analisis Sentimen terhadap RSUD Salatiga Menggunakan SVM dan TF-IDF

Windy Livia Azzahra ^{1*}, Evangs Mailoa ²

^{1*,2}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Kota Salatiga, Provinsi Jawa Tengah, Indonesia.

Email: 672021211@student.uksw.edu ^{1*}, evangs.mailoa@uksw.edu ²

Histori Artikel:

Dikirim 1 November 2024; *Diterima dalam bentuk revisi* 15 November 2024; *Diterima* 30 Desember 2024; *Diterbitkan* 10 Januari 2025. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Rumah Sakit Umum Daerah (RSUD) Salatiga memiliki peran penting dalam memberikan layanan kesehatan. Penelitian ini menganalisis sentimen masyarakat terhadap RSUD Salatiga menggunakan metode SVM dengan kernel linear untuk analisis sentimen dan TF-IDF digunakan untuk ekstraksi fitur. Dataset terdiri dari 414 ulasan yang telah diproses, termasuk case folding, pembersihan data, tokenisasi, normalisasi, penghapusan kata, dan stemming. Evaluasi menunjukkan model mencapai akurasi 84,00%; presisi 84,00%; recall 83,25%; dan F1-score 83,53%. Sebanyak 55,8% ulasan menunjukkan sentimen positif dan 44,2% sentimen negatif, menyoroti perlunya peningkatan pada sistem antrean, waktu tunggu, dan fasilitas parkir. Metode SVM dan TF-IDF dipilih karena kemampuannya menangani data teks besar dengan akurasi tinggi. Penelitian ini memberikan kontribusi praktis berupa rekomendasi seperti implementasi sistem antrean berbasis teknologi. Keterbatasan meliputi jumlah data yang terbatas dan bias platform, sehingga disarankan eksplorasi algoritma lain, seperti Naive Bayes dan Random Forest.

Kata Kunci: Analisis Sentimen; Rumah Sakit; Support Vector Machine (SVM); TF-IDF.

Abstract

The Salatiga Regional General Hospital (RSUD) plays an important role in providing healthcare services. This research analyzes public sentiment towards RSUD Salatiga using the SVM method with a linear kernel for sentiment analysis and TF-IDF for feature extraction. The dataset consists of 414 processed reviews, including case folding, data cleaning, tokenization, normalization, stopword removal, and stemming. Evaluation shows that the model achieved an accuracy of 84.00%; precision of 84.00%; recall of 83.25%; and an F1-score of 83.53%. A total of 55.8% of reviews indicated positive sentiment and 44.2% negative sentiment, highlighting the need for improvements in the queue system, waiting times, and parking facilities. The SVM and TF-IDF methods were chosen for their ability to handle large text data with high accuracy. This research provides practical contributions in the form of recommendations such as the implementation of a technology-based queue system. Limitations include the limited amount of data and platform bias, so exploring other algorithms, such as Naive Bayes and Random Forest, is recommended.

Keyword: Sentiment Analysis; Hospital; Support Vector Machine (SVM); TF-IDF.

1. Pendahuluan

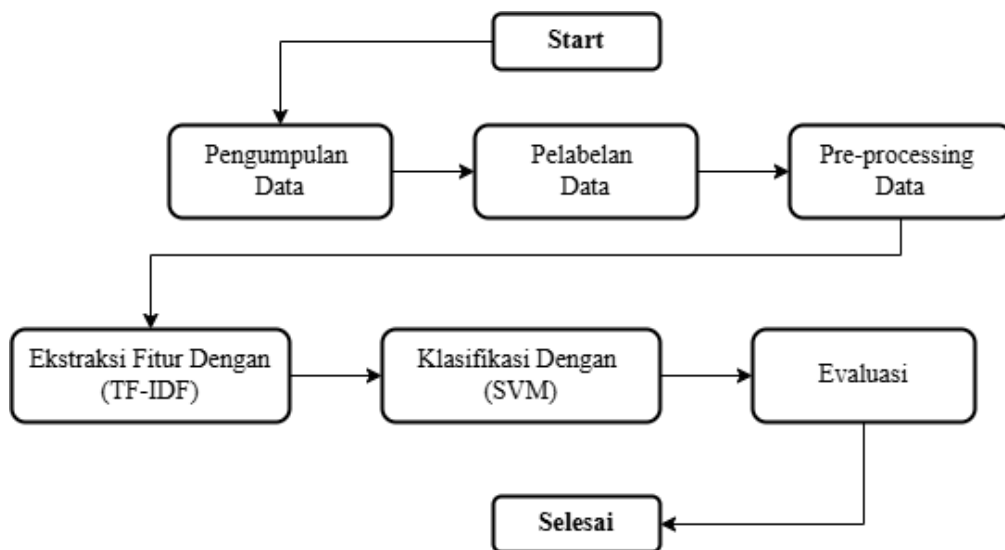
Dalam era digital saat ini, media sosial dan platform *online* telah menjadi sarana utama bagi masyarakat untuk berbagi pengalaman dan opini terkait produk atau layanan yang digunakan. Salah satu sektor yang sering menjadi perhatian adalah layanan kesehatan, khususnya rumah sakit. Rumah Sakit Umum Daerah (RSUD) Salatiga, sebagai fasilitas pelayanan kesehatan milik pemerintah daerah, mendapat banyak umpan balik dan penilaian dari masyarakat terkait layanan yang diberikan (Gifari *et al.*, 2022). Pada tahun 2021, RSUD Salatiga tercatat sebagai rumah sakit dengan jumlah pasien tertinggi di Kota Salatiga, mencapai 10.722 pasien, melampaui enam rumah sakit lainnya di wilayah tersebut (Data Statistik Sektoral Kota Salatiga, 2021). Analisis sentimen, yang merupakan bagian dari pemrosesan bahasa alami (*Natural Language Processing*), digunakan untuk mengekstraksi dan mengklasifikasikan opini atau sentimen dari data teks secara luas, seperti ulasan atau komentar di media sosial (Liu, 2022). Analisis sentimen terhadap layanan RSUD Salatiga membantu pihak rumah sakit mengidentifikasi aspek pelayanan yang memerlukan peningkatan berdasarkan data masyarakat (Muttaqien *et al.*, 2022). Penelitian ini menggunakan metode *Support Vector Machine (SVM)* dan *TF-IDF (Term Frequency-Inverse Document Frequency)* untuk menganalisis sentimen masyarakat terhadap layanan RSUD Salatiga. *SVM* adalah algoritma pembelajaran mesin yang efektif untuk klasifikasi teks, sedangkan *TF-IDF* merupakan metode pembobotan kata yang sering diterapkan dalam pemrosesan bahasa alami (Pranata *et al.*, 2024). Klasifikasi dilakukan untuk membagi ulasan ke dalam dua kategori, yaitu sentimen positif dan negatif, sehingga hasil analisis dapat menggambarkan persepsi masyarakat secara lebih terarah (Dzukaidah & Prasvita, 2022).

SVM memiliki sejumlah keunggulan, termasuk kemampuannya menangani data berdimensi tinggi, kemampuan bekerja dengan jumlah fitur yang lebih besar dari jumlah sampel, serta efektivitasnya dalam menangani data yang tidak seimbang (Ditami dkk., 2022). Namun, kelemahan *SVM* meliputi sensitivitas terhadap pemilihan parameter serta biaya komputasi yang tinggi untuk melatih model pada *dataset* besar. *TF-IDF (Term Frequency-Inverse Document Frequency)* adalah metode yang digunakan untuk mengevaluasi relevansi sebuah kata dalam dokumen tertentu dibandingkan dengan koleksi dokumen lainnya (Septiani & Isabela, 2022). Teknik ini digunakan untuk mengidentifikasi fitur penting dari teks. Penelitian ini memiliki karakteristik unik dibandingkan studi sebelumnya, yaitu pada penggunaan *dataset* yang secara khusus berisi ulasan terkait layanan RSUD Salatiga (Khan & Srivastava, 2024). *Dataset* diperoleh dari berbagai sumber *online*, seperti media sosial, situs ulasan, dan forum diskusi, untuk memastikan hasil analisis mencerminkan persepsi masyarakat secara representatif. Selain itu, penelitian ini juga mengevaluasi pengaruh teknik pra-pemrosesan data teks, seperti *case folding*, *data cleaning*, tokenisasi, normalisasi, penghapusan kata, dan *stemming*, terhadap kinerja model *SVM* dan *TF-IDF*. Dengan penerapan langkah-langkah tersebut, akurasi serta efisiensi analisis diharapkan dapat meningkat. Penelitian bertujuan untuk memberikan informasi terkait sentimen masyarakat terhadap layanan RSUD Salatiga dan menjadi rujukan bagi rumah sakit serta fasilitas kesehatan lainnya dalam memanfaatkan analisis sentimen untuk meningkatkan kualitas pelayanan (Astuti *et al.*, 2023).

2. Metode Penelitian

2.1 Alur Metode Penelitian

Alur metode penelitian ini dibangun untuk analisis sentimen terhadap RSUD Salatiga menggunakan metode *Support Vector Machine* dan *TF-IDF* terdiri dari beberapa tahapan. Tahapan tersebut antara lain pengumpulan data, pelabelan data, *pre-processing* data, ekstraksi fitur menggunakan *TF-IDF*, menggunakan *Support Vector Machine (SVM)* untuk pemodelan klasifikasi, dan evaluasi. Desain alur metode penelitian yang dirancang dapat dilihat pada gambar 1 berikut ini:



Gambar 1. Alur Metode Penelitian

2.2 Pengumpulan Data

Data dikumpulkan dari ulasan dan komentar masyarakat terkait layanan RSUD Salatiga. Sebagian besar data berasal dari komentar di Google Maps, serta beberapa sumber *online* lainnya, termasuk media sosial resmi RSUD Salatiga dan situs ulasan kesehatan. Proses pengumpulan data dilakukan dengan teknik *scraping* pada platform yang mendukung *API* media sosial untuk mengakses ulasan dan komentar, serta secara manual untuk situs yang tidak dapat diproses secara otomatis. Data yang diperoleh disimpan dalam format *CSV* untuk menjaga struktur data tetap terorganisir (Gifari dkk., 2022). Pengumpulan data mencakup periode tahun 2019 hingga Oktober 2024, menghasilkan *dataset* sebanyak 624 entri. Validitas data dipastikan dengan hanya menggunakan ulasan dari akun yang aktif dan relevan dengan layanan RSUD Salatiga, sementara keberagaman data dijaga dengan mencakup berbagai jenis ulasan, baik positif maupun negatif.

2.3 Pelabelan Data

Proses pelabelan data dilakukan dengan mengklasifikasikan sentimen ke dalam dua kategori: positif dan negatif. Pelabelan dilakukan secara manual dengan membaca setiap ulasan yang telah dikumpulkan (Karimah, 2024). Ulasan yang mencerminkan kepuasan terhadap pelayanan atau fasilitas diberikan label sebagai sentimen positif. Sebaliknya, ulasan yang mengandung kritik atau menunjukkan ketidakpuasan dikategorikan sebagai sentimen negatif. Contoh proses pelabelan disajikan pada tabel 1 berikut:

Tabel 1. Pelabelan Dataset

Ulasan	Label
RSUD Salatiga selalu memberikan pelayanan yg terbaik. Semoga kedepan terus berkembang, sukses terus RSUD Salatiga 😊 ...	Positif
Tidak efektif dalam sistem antrean, dokter datang terlambat, kemudian sudah 3 jam antre belum di periksa.	Negatif
Ruangan IGD yang cukup cepat. Ruang poli dan bangsal yang cukup luas, bersih, rapi dan tertata. Keunggulan RS terdapat poli umum dan poli eksklusif.	Positif
Perlu dibangun Gedung Parkir. Karena pasien atau pengunjung RS banyak yang kesulitan menemukan tempat parkir yang layak dan memadai	Negatif

2.4 Pre-Processing Data

Pada tahap *pre-processing*, data mentah diubah ke dalam format *CSV* yang sesuai dengan kebutuhan analisis. Atribut yang tidak relevan dihapus untuk meningkatkan efisiensi. Tahapan utama dalam proses *pre-processing* meliputi:

- a) *Case Folding*
Semua huruf besar diubah menjadi huruf kecil untuk menyeragamkan data menggunakan pustaka *NLTK*. Sebagai contoh, kata "BAGUS" diubah menjadi "bagus".
- b) Pembersihan Data
Data yang tidak relevan, seperti emotikon, angka, dan tanda baca, dihapus menggunakan pustaka *regular expression (re)* untuk menghilangkan karakter yang tidak diperlukan.
- c) Tokenisasi
Kalimat dipecah menjadi kata-kata individu atau token. Seluruh simbol, seperti tanda seru dan titik koma, dihapus menggunakan modul *nlk.word_tokenize*.
- d) Normalisasi
Kata-kata tidak standar atau informal diubah menjadi bentuk baku. Proses ini bertujuan untuk menyeragamkan variasi penulisan agar teks lebih mudah diproses oleh algoritma. Normalisasi dilakukan menggunakan daftar kata yang telah disiapkan secara manual.
- e) Penghapusan Kata
Stopword yang tidak relevan dalam kalimat dihapus menggunakan pustaka *NLTK*, *corpus*, dan Sastrawi. Proses ini bertujuan untuk menghilangkan kata-kata yang tidak signifikan, sehingga meningkatkan efisiensi analisis.
- f) *Stemming*
Kata-kata dinormalisasi menjadi bentuk dasar menggunakan pustaka Sastrawi untuk memastikan konsistensi dalam teks.

2.5 Ekstraksi Fitur Dengan TF-IDF

Tahap ekstraksi fitur bertujuan untuk mengubah data menjadi format yang sesuai untuk algoritma *machine learning*, yang dapat digunakan untuk menyelesaikan tugas yang diinginkan (Batubara & Maharani, 2020). Dalam penelitian ini, ekstraksi fitur berfokus pada identifikasi emosi terhadap layanan RSUD. Proses ini dilakukan dengan menggunakan algoritma TF-IDF, di mana teks diekstraksi melalui vektorisasi TF-IDF. Formula yang digunakan adalah sebagai berikut:

- 1) Menghitung TF (*Term Frequency*)

$$TF(t, d) = \frac{\text{jumlah kemunculan } t \text{ dalam } d}{\text{jumlah dokumen yang mengandung } t}$$

- 2) Menghitung IDF (*Inverse Document Frequency*)

$$IDF(t) = \log\left(\frac{\text{jumlah total dokumen}}{\text{jumlah dokumen yang mengandung } t}\right)$$

- 3) Menghitung TF-IDF

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

Hasilnya terdiri dari set data latihan yang didasarkan pada parameter fitur maksimum, yang menunjukkan jumlah kata unik yang dapat dikenali dalam ulasan secara keseluruhan. Proses ini terdiri dari sebuah kalimat berisikan kumpulan kata yang dipecah menjadi kata per kata dan kemudian mengubahnya menjadi bilangan unik.

2.6 Klasifikasi Dengan SVM

Support Vector Machine (SVM) adalah metode untuk melakukan prediksi dalam kasus klasifikasi maupun regresi (Raehanun, 2021). Implementasi SVM melibatkan tahap pelatihan diikuti dengan tahap pengujian. Metode ini bekerja berdasarkan prinsip supervised learning, yang memerlukan data berlabel untuk melatih model. Implementasi SVM melibatkan pelatihan model dengan dataset yang ada dan pengujian model dengan data baru. SVM dikembangkan untuk menyelesaikan masalah klasifikasi karena memiliki kemampuan generalisasi yang lebih baik dibandingkan dengan teknik sebelumnya. SVM adalah salah satu metode *machine learning* yang melatih model menggunakan dataset pelatihan, kemudian membuat generalisasi dan prediksi dari data baru (Prabowo & Azizah, 2020). Dalam pemodelan klasifikasi, SVM memiliki konsep matematis yang lebih jelas dan matang daripada metode klasifikasi lainnya. SVM juga mampu menangani masalah regresi dan klasifikasi, baik linear maupun *non-linear*. SVM menggunakan *kernel linear*, yang dipilih karena sesuai untuk dataset dengan karakteristik *linear separable* dan lebih efisien dibandingkan kernel lainnya. *Hyperplane* optimal ditentukan dengan memaksimalkan margin antar kelas, menghasilkan klasifikasi yang lebih akurat. *Hyperplane* ini dibangun dengan mengukur jarak terdekat antara *hyperplane* dan vektor pendukung dari masing-masing kelas, memastikan margin maksimal (Fahlevvi, 2022). Keunggulan metode SVM adalah konsep structural risk minimization (SRM), yang membantu mengatasi masalah *overfitting*. Metode *Support Vector Machine* memiliki beberapa model pendekatan atau biasa disebut *kernel* seperti *linear*, *polynomial* dan *gaussian/ radial basis function* (RBF). Persamaan *kernel* tersebut dapat dilihat sebagai berikut:

a) *Linear*

$$K(x_i, x_j) = x_i^T x_j + C$$

Dengan x_i dan x_j merupakan vektor dari dataset dan C merupakan constant.

b) *Polynomial*

$$K(x_i, x_j) = (\gamma \cdot x_i^T x_j + c)^d, \gamma > 0$$

Dengan x_i dan x_j merupakan vektor dari dataset, γ adalah parameter untuk mengontrol kecepatan proses learning, C merupakan constant dan d merupakan pangkat *polynomial* yang digunakan.

c) *Gaussian/Radial basis Function* (RBF)

$$K(x_i, x_j) = \exp(-\gamma \|X_i - X_j\|^2), \gamma > 0$$

Dengan x_i dan x_j merupakan vektor dari dataset, γ adalah parameter untuk mengontrol kecepatan proses learning dan *exp* merupakan basis dari logaritma alami.

Tujuan penggunaan *kernel* adalah untuk menerapkan model dalam ruang berdimensi lebih tinggi (*feature space*) tanpa perlu mendefinisikan fungsi pemetaan dari ruang input ke *feature space*. Dengan demikian, kasus yang tidak dapat dipisahkan secara linear dalam ruang input diharapkan menjadi dapat dipisahkan secara linear di *feature space*. *Hyperplane* kemudian dapat digunakan sebagai decision boundary dengan efisien (Cristianini, 2000). Dalam penelitian ini, hanya *kernel linear* yang digunakan karena kernel *linear* memiliki tingkat akurasi yang lebih tinggi dibandingkan *kernel* lainnya (Aruan dkk., 2022). Sebagai salah satu jenis kernel paling mendasar, *kernel linear* sangat tepat diaplikasikan pada dataset yang memiliki karakteristik pemisahan linear. Penerapannya efektif ketika klasifikasi data dapat dilakukan menggunakan garis pemisah atau *hyperplane*. Dalam formulasi matematisnya, dengan $x \in \mathbb{R}^n$, kernel *linear* dinyatakan sebagai $k(x, x') = \{x, x'\} = \sum x_i x_i'$. Pada vektor padat, waktu komputasi kernel membutuhkan $O(n)$, sedangkan untuk vektor sparse dapat direduksi menjadi $O(|\text{nnz}(x) \cap \text{nnz}(x')|)$, dengan $\text{nnz}(\cdot)$ merepresentasikan kumpulan indeks non-zero dari vektor dan $|\cdot|$ menandakan dimensi himpunan. Aplikasi *kernel linear* sering ditemui dalam pengolahan data berbentuk vektor, khususnya dalam analisis teks, di mana dokumen dikonversi menjadi vektor frekuensi kata.

2.7 Evaluasi

Pada penelitian ini, *Confusion matrix* digunakan untuk mengukur evaluasi performa. *Confusion matrix* digunakan karena memberikan informasi tentang bagaimana model membuat prediksi, termasuk jenis dan jumlah kesalahan yang dibuat. Setiap kolom pada *confusion matrix* mewakili instance dari kelas prediksi, sementara setiap baris mewakili instance dari kelas actual (Atika dkk., 2022). *Confusion matrix* dalam penelitian ini dibuat menggunakan library scikit-learn. Akurasi, presisi, recall, dan pengukuran F1 adalah metrik kinerja yang umum digunakan dalam *Confusion matrix*, dan penjelasannya dapat ditemukan di tabel 2 *Confusion matrix* berikut:

Tabel 2. *Confusion Matrix*

<i>Actual</i>	<i>Predicted</i>	
	<i>Positive</i>	<i>True Positive (TP)</i>
<i>Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

True Positive (TP) adalah jumlah data positif yang diklasifikasikan dengan benar oleh system; *True Negative (TN)* adalah jumlah data negatif yang diklasifikasikan dengan benar oleh sistem; *False Positive (FP)* adalah jumlah data negatif yang salah diklasifikasikan sebagai positif oleh system; dan, *False Negative (FN)* adalah jumlah data positif yang salah diklasifikasikan sebagai negatif oleh system (Kristiawan & Widjaja, 2021).

- 1) *Accuracy* (Akurasi) adalah tingkat kedekatan antara nilai prediksi dengan nilai aktual. Rumus untuk menghitung akurasi adalah:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

- 2) *Precision* (Presisi) adalah tingkat ketepatan antara informasi yang diminta dan jawaban yang diberikan oleh sistem. Rumus untuk menghitung presisi adalah:

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

- 3) *Recall* adalah tingkat keberhasilan sistem dalam mengenali data yang relevan dari keseluruhan data positif. Rumus untuk menghitung *recall* adalah:

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

- 4) *F1-Measure (F1 Score)* adalah penyetaraan nilai *precision* dan *recall* karena kedua nilai tersebut biasanya memiliki perbedaan yang cukup tinggi. Rumus untuk menghitung *F1-Measure (F1 Score)* adalah:

$$F1 - Measure = 2 \times \frac{(recall \times precision)}{(recall + precision)} \times 100\%$$

3. Hasil dan Pembahasan

3.1 Hasil

Dataset penelitian ini terdiri dari 624 ulasan masyarakat terhadap RSUD Salatiga, yang diperoleh dari Google Maps dan media sosial resmi rumah sakit. Setelah melalui proses pembersihan data untuk menghapus data duplikat, tidak relevan, atau kosong, jumlah data valid yang digunakan adalah 414

ulasan. Selain itu data yang sudah terkumpul kemudian diberikan label Sentimen Positif atau Negatif secara manual. Proses pra-pemrosesan meliputi case folding, pembersihan data, tokenisasi, normalisasi, penghapusan kata berhenti, dan stemming, menghasilkan data yang siap untuk ekstraksi fitur menggunakan TF-IDF. Hasil perubahan data sebelum pemrosesan dan setelah pemrosesan ditunjukkan pada tabel 3.

Tabel 3. Hasil *Pre-processing*

Ulasan	Pre-processing	Hasil
<p>Pelayanan poli saraf sangat baik, dokter dan perawatnya ramah... Bapak saya periksa di layani dengan baik, dan sekarang alhamdulillah sudah membaik... Terima kasih RSUD</p>	<i>Case Folding</i>	pelayanan poli saraf sangat baik, dokter dan perawatnya ramah.... bapak saya periksa di layani dengan baik, dan sekarang alhamdulillah sudah membaik... terima kasih rsud
	Pembersihan Data	pelayanan poli saraf sangat baik dokter dan perawatnya ramah bapak saya periksa di layani dengan baik dan sekarang alhamdulillah sudah membaik terima kasih rsud
	Tokenisasi	“pelayanan,” “poli,” “saraf,” “sangat,” “baik,” “dokter,” “dan,” “perawatnya,” “ramah,” “bapak,” “saya,” “periksa,” “di,” “layani,” “dengan,” “baik,” “dan,” “sekarang,” “alhamdulillah,” “sudah,” “membaik,” “terima,” “kasih,” “rsud”
	Normalisasi	“pelayanan,” “poliklinik,” “saraf,” “sangat,” “baik,” “dokter,” “dan,” “perawatnya,” “ramah,” “bapak,” “saya,” “periksa,” “di,” “layani,” “dengan,” “baik,” “dan,” “sekarang,” “alhamdulillah,” “sudah,” “membaik,” “terima,” “kasih,” “rsud”
	Penghapusan Kata	“pelayanan,” “poliklinik,” “saraf,” “dokter,” “perawatnya,” “ramah,” “periksa,” “layani,” “alhamdulillah,” “membaik,” “terima,” “kasih,” “rsud”
	<i>Stemming</i>	layan poliklinik saraf dokter awat ramah periksa layan alhamdulillah baik terima kasih rsud

Tahap ekstraksi fitur bertujuan untuk mengubah data teks yang telah di pre-processing menjadi format numerik yang dapat digunakan oleh algoritma pembelajaran mesin. Pada penelitian ini, metode TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan untuk memberi nilai setiap kata dalam dokumen berdasarkan frekuensi kemunculannya dan seberapa jarang kata tersebut muncul dalam seluruh dokumen. Berikut adalah proses ekstraksi fitur menggunakan TF-IDF dari contoh yang telah dipreprocessing sebelumnya.

- 1) Menghitung TF (*Term Frequency*)
Misalkan ada dua dokumen (d1, d2, dan d3):
 - a) d1: ['alhamdulillah,' 'layan,' 'rsud,' 'puas,' 'layan,' 'ruang,' 'rawat,' 'inap']
 - b) d2: ['layan,' 'rsud,' 'puas,' 'informasi']
 - c) d3: ['layan,' 'ramah,' 'cepat,' 'tanggap']

$$TF(puas, d1) = \frac{1}{8} = 0,125$$

- 2) Menghitung IDF (*Inverse Document Frequency*)

$$IDF(puas) = \log\left(\frac{3}{2}\right) = 0,176$$

- 3) Menghitung TF-IDF

$$TF - IDF(puas, d1) = TF(puas, d1) \times IDF(puas) = 0,125 \times 0,176 = 0,022$$

Tabel 4. Pembobotan Kata Dengan TF-IDF

Ulasan	layan	ramah	fasilitas	perawat	cepat	sehat
layan vip bagus perawat ramah	0,207	0,330	0	0,388	0	0
fasilitas layan sehat lengkap	0,201	0	0,552	0	0	0,511
ramah cepat tanggap	0	0,404	0	0	0,565	0



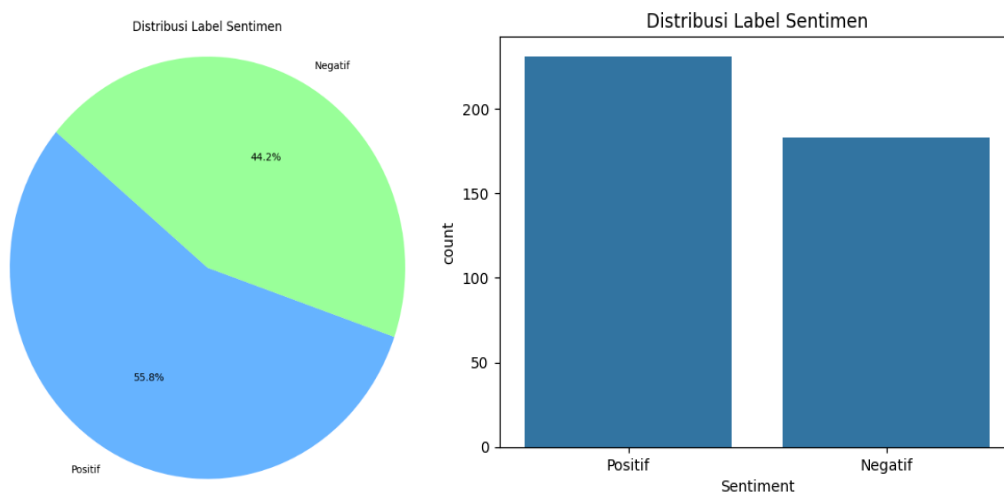
Gambar 2. Visualisasi *Wordcloud*

Gambar 2 menyajikan *wordcloud* (frekuensi kata) untuk memvisualisasikan kata-kata dominan dalam ulasan positif dan negatif. Beberapa kata dominan dalam ulasan positif meliputi "ramah," "layan," "pasien," "dokter," "terima kasih," dan "baik." Kata-kata ini mencerminkan persepsi positif masyarakat, terutama terkait keramahan staf dan kualitas pelayanan yang banyak diapresiasi melalui ungkapan terima kasih. Sebaliknya, kata-kata seperti "antre," dan "waktu," yang muncul dalam ulasan negatif menunjukkan ketidakpuasan terkait waktu tunggu dan sistem antrean. Selain itu, kata-kata seperti "ruang," "poliklinik," "parkir," dan "BPJS" menunjukkan perhatian masyarakat terhadap fasilitas rumah sakit. Temuan ini memberikan gambaran menyeluruh mengenai aspek layanan dan fasilitas yang menjadi perhatian utama pasien dan pengunjung RSUD Salatiga, yang akan dianalisis lebih lanjut menggunakan metode *Support Vector Machine* dan TF-IDF untuk memahami sentimen masyarakat secara mendalam.

Tabel 5. Hasil Evaluasi *Confusion Matrix*

<i>Actual</i>	<i>Predicted</i>	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	63	12
<i>Negative</i>	8	42
<i>Accuracy</i>	84,00%	
<i>Precision</i>	84,00%	
<i>Recall</i>	83,25%	
<i>F1-Score</i>	83,53%	

Berdasarkan hasil evaluasi menggunakan *Confusion Matrix* pada Tabel 5. Model *Support Vector Machine* yang digunakan dalam analisis sentimen terhadap RSUD Salatiga menunjukkan performa yang cukup baik. Model berhasil mencapai *Accuracy* sebesar 84,00%; yang berarti model dapat mengklasifikasikan sentimen dengan tepat pada 84% dari keseluruhan data, nilai *precision* sebesar 84,00% menunjukkan tingkat ketepatan model dalam mengklasifikasikan sentimen positif dari seluruh prediksi positif yang dibuat, sedangkan *recall* sebesar 83,25% menggambarkan kemampuan model dalam mengenali seluruh sentimen positif yang sebenarnya ada dalam dataset. *F1-Score* sebesar 83,53% mengindikasikan keseimbangan yang baik antara *precision* dan *recall*. Secara keseluruhan, hasil ini menunjukkan bahwa model memiliki performa yang reliabel dalam menganalisis sentimen terhadap RSUD Salatiga dari dataset yang ada.



Gambar 3. Diagram Hasil Distribusi Label Sentimen

Gambar 3. menunjukkan hasil pelabelan manual terhadap 414 data yang valid, diperoleh distribusi sentimen sebagai berikut:

- 1) Sentimen Positif: 231 data (55,8%)
- 2) Sentimen Negatif: 183 data (44,2%)

Distribusi sentimen yang menunjukkan dominasi sentimen positif (55,8%) menunjukkan persepsi masyarakat yang cukup baik terhadap layanan RSUD Salatiga. Namun, proporsi sentimen negatif yang signifikan (44,2%) mengindikasikan bahwa masih ada aspek layanan yang perlu diperbaiki. Ulasan negatif memberikan informasi spesifik tentang perlunya memperbaiki efisiensi sistem antrean, mengurangi waktu tunggu pasien, serta meningkatkan kapasitas fasilitas parkir.

3.2 Pembahasan

Penelitian ini menganalisis sentimen masyarakat terhadap layanan RSUD Salatiga menggunakan algoritma *Support Vector Machine (SVM)* dan metode *Term Frequency-Inverse Document Frequency (TF-IDF)*. Hasil evaluasi menunjukkan model mencapai tingkat akurasi sebesar 84,00%, dengan presisi 84,00%, *recall* 83,25%, dan *F1-score* 83,53%. Hasil ini mengindikasikan bahwa kombinasi metode *SVM* dan *TF-IDF* mampu menghasilkan performa klasifikasi yang andal dalam konteks analisis sentimen, sejalan dengan temuan serupa pada penelitian lain (Aruan *et al.*, 2022; Ditami *et al.*, 2022). Sentimen dari dataset mengungkapkan bahwa 55,8% ulasan masyarakat masuk dalam kategori positif, sedangkan 44,2% termasuk kategori negatif. Distribusi ini menunjukkan persepsi masyarakat yang umumnya baik terhadap layanan RSUD Salatiga, meskipun proporsi ulasan negatif tetap signifikan. Sentimen positif terutama mencerminkan penghargaan terhadap keramahan staf dan kualitas pelayanan, sementara sentimen negatif lebih sering terkait dengan waktu tunggu, sistem antrean, dan fasilitas parkir. Temuan ini didukung oleh penelitian lain yang menunjukkan bahwa keluhan terkait layanan operasional dan infrastruktur sering kali menjadi perhatian utama dalam analisis sentimen layanan publik (Astuti *et al.*, 2023; Dzukaidah & Prasvita, 2022). Penggunaan *SVM* dalam penelitian ini didasarkan pada kemampuan algoritma ini untuk menangani data berdimensi tinggi serta menghasilkan klasifikasi yang konsisten meskipun dataset tidak seimbang (Cristianini, 2000; Fahlevvi, 2022). Keunggulan ini juga diakui dalam penelitian lain yang menggunakan *SVM* untuk klasifikasi teks, termasuk pada domain layanan publik dan media sosial (Gifari *et al.*, 2022; Atika *et al.*, 2022). Sementara itu, *TF-IDF* dipilih karena kemampuannya mengekstraksi fitur yang relevan dari data teks, yang juga sesuai dengan temuan Batubara dan Maharani (2020) serta Septiani dan Isabela (2022). Penelitian ini memberikan wawasan penting tentang persepsi masyarakat terhadap layanan RSUD Salatiga. Hasilnya juga memperkuat relevansi analisis sentimen sebagai alat evaluasi kinerja layanan publik, sebagaimana dibuktikan dalam penelitian lain (Aruan *et al.*, 2022; Kristiawan & Widjaja, 2021).

4. Kesimpulan

Penelitian ini berhasil menganalisis sentimen masyarakat terhadap RSUD Salatiga menggunakan metode *Support Vector Machine (SVM)* dan *TF-IDF* dengan akurasi 84,00%. Model menunjukkan performa seimbang dengan presisi 84,00%, *recall* 83,25%, dan *F1-score* 83,53%. Distribusi sentimen mengungkapkan 55,8% sentimen positif dan 44,2% sentimen negatif, menunjukkan persepsi masyarakat yang umumnya baik terhadap layanan rumah sakit ini. Analisis lebih lanjut terhadap sentimen negatif menyoroti perlunya perbaikan pada waktu tunggu, sistem antrean, dan fasilitas parkir. Hasil analisis *wordcloud* juga menunjukkan bahwa pelayanan, keramahan staf, dan fasilitas adalah fokus utama perhatian masyarakat. Penelitian ini memberikan kontribusi praktis bagi RSUD Salatiga dengan merekomendasikan langkah konkret, seperti implementasi sistem antrean berbasis teknologi seperti QR Code untuk mengurangi waktu tunggu dan meningkatkan pengalaman pasien, serta peningkatan fasilitas parkir untuk mengatasi keluhan aksesibilitas. Namun, penelitian ini memiliki keterbatasan, seperti jumlah data yang belum mencakup seluruh segmen populasi dan potensi bias dalam proses pengumpulan data. Untuk penelitian selanjutnya, disarankan menggunakan algoritma tambahan untuk membandingkan kinerja dengan metode saat ini seperti *Naive Bayes* untuk pendekatan probabilistik sederhana atau *Random Forest* untuk analisis fitur yang lebih kompleks. Penambahan kelas netral dalam pelabelan data juga diharapkan memberikan wawasan yang lebih konkrit.

5. Ucapan Terima Kasih

Terima kasih kepada dosen Universitas Kristen Satya Wacana atas bimbingan dan masukan berharga selama penelitian ini. Ucapan terima kasih juga ditujukan kepada keluarga, teman, dan semua pihak yang telah memberikan dukungan, baik secara moral maupun material. Semua bantuan yang diberikan sangat berarti dalam menyelesaikan penelitian ini.

6. Daftar Pustaka

- Aruan, J. D. C., Rahayudi, B., & Ridok, A. (2022). Analisis Sentimen Opini Masyarakat terhadap Pelayanan Rumah Sakit Umum Daerah menggunakan Metode Support Vector Machine dan Term Frequency-Inverse Document Frequency. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(5), 2072-2078.
- Astuti, Y., Ruldeviyani, Y., Salbari, F., & Prayogi, A. (2023). Sentiment Analysis of Electricity Company Service Quality Using Naïve Bayes. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 7(2), 389-396. <https://doi.org/10.29207/resti.v7i2.4627>.
- Atika, D., Styawati, S., & Aldino, A. A. (2022). Term Frequency-Inverse Document Frequency Support Vector Machine untuk Analisis Sentimen Opini Masyarakat terhadap Tekanan Mental pada Media Sosial Twitter. *Jurnal Teknologi dan Sistem Informasi*, 3(4), 86-97. <https://doi.org/10.33365/jtsi.v3i4.2054>.
- Batubara, A. S. R. T., & Maharani, W. (2020). Kecendrungan Sentimen Dengan Pendekatan Support Vector Machine Pada Komunitas Yang Berpengaruh Di Twitter. *eProceedings of Engineering*, 7(3).
- Cristianini, N. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press.
- Ditami, G. R., Ripanti, E. F., & Sujaini, H. (2022). Implementasi Support Vector Machine untuk Analisis Sentimen Terhadap Pengaruh Program Promosi Event Belanja pada Marketplace. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 8(3), 508-516. <https://doi.org/10.26418/jp.v8i3.56478>.
- Dzukaidah, M. F., & Prasvita, D. S. (2022, August). ANALISIS SENTIMEN PROGRAM BANTUAN SOSIAL TUNAI PADA SOSIAL MEDIA TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. In *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya* (Vol. 3, No. 2, pp. 569-580).
- Fahlevvi, M. R. (2022). Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia di Google Playstore Menggunakan Metode Support Vector Machine. *Jurnal Teknologi dan Komunikasi Pemerintahan*, 4(1), 1-13. <https://doi.org/10.33701/jtkp.v4i1.2701>.
- Gifari, O. I., Adha, M., Hendrawan, I. R., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, 2(1), 36-40. <https://doi.org/10.46229/jifotech.v2i1.330>.

- Karimah, N. (2024). Multi-Aspect Sentiment Analysis Pada Review Film Menggunakan Metode Bidirectional Encoder Representations From Transformers (BERT). *Komputika: Jurnal Sistem Komputer*, 13(1), 63-72. <https://doi.org/10.34010/komputika.v13i1.11098>.
- Khan, M., & Srivastava, A. (2024). Sentiment analysis of Twitter data using machine learning techniques. *International Journal of Engineering and Management Research*, 14(1), 196-203. <https://doi.org/10.5281/zenodo.10791485>.
- Kristiawan, K., & Widjaja, A. (2021). Perbandingan algoritma machine learning dalam menilai sebuah lokasi toko ritel. *Jurnal Teknik Informatika dan Sistem Informasi*, 7(1). <https://doi.org/10.28932/jutisi.v7i1.3182>.
- Liu, B. (2022). *Sentiment analysis and opinion mining*. Springer Nature.
- Muttaqien, D. D., Tibyani, T., & Hartono, P. P. (2022). Implementasi Support Vector Machine pada Analisis Sentimen mengenai Bantuan Sosial di Era Pandemi Covid-19 pada Pengguna Twitter. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(1), 163-171.
- Prabowo, W. A., & Azizah, F. (2020). Sentiment analysis for detecting cyberbullying using TF-IDF and SVM. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 4(6), 1142–1148. <https://doi.org/10.29207/resti.v4i6.2753>.
- Pranata, R. A., Rudiman, R., & Verdikha, N. A. (2024). Metode pembobotan TF-IDF untuk klasifikasi teks quick count pemilihan wakil presiden Indonesia 2024 pada X Twitter dengan metode SVM. *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, 18(2), 127–138. <https://doi.org/10.54914/jit.v10i2.1279>.
- Raehanun, M. (2021). Analisis Support Vector Machine (SVM) Dalam Prediksi Permintaan Emas Perhiasan (Studi Kasus: Permintaan Emas Perhiasan dari Beberapa Negara Tertentu Periode Tahun 2000-2021).
- Septiani, D., & Isabela, I. (2022). Analisis term frequency inverse document frequency (tf-idf) dalam temu kembali informasi pada dokumen teks. *Sistem dan Teknologi Informasi Indonesia (SINTESIA)*, 1(2), 81-88.