

$$TF(t, d) = \frac{\text{Jumlah kemunculan term } t \text{ dalam dokumen } d}{\text{Total jumlah term dalam dokumen } d} \quad (2.13)$$

Kemudian untuk menghitung nilai *invers document frequency* menggunakan persamaan berikut.

$$idf_i = \log\left(\frac{D}{df_i}\right) \quad (2.14)$$

Dengan demikian rumus perhitungan *TF-IDF* adalah penggabungan dari rumus TF dan IDF dengan melakukan operasi perkalian nilai seperti persamaan berikut.

$$W_{ij} = tf_{ij} \times idf_i \quad (2.15)$$

Keterangan:

W_{ij} : bobot *TF-IDF*

tf_i : frekuensi suatu kata

idf_i : *inverse document frequency*

D : jumlah dokumen

df_i : jumlah dokumen yang mengandung kata i

2.8 Evaluasi

Evaluasi model dilakukan dengan tujuan untuk mengukur performa metode yang digunakan dalam penelitian. digunakan dua pendekatan evaluasi yang berbeda yaitu *confusion marix* untuk model klasifikasi sentimen.

2.8.1 Confusion Matrix

Confusion matrix adalah tabel yang digunakan untuk mengevaluasi performa model klasifikasi dengan membandingkan hasil prediksi model terhadap label. Tabel ini terdiri dari empat komponen utama.

True Positive (TP): Prediksi positif yang benar.

True Negative (TN): Prediksi Negatif yang benar.

False Positive (FP): Prediksi positif yang salah.

False Negative (FN): Prediksi Negatif yang salah.

Dari keempat komponen tersebut, dapat diturunkan metrik-metrik evaluasi sebagai berikut [6].

Akurasi (*Accuracy*)

Akurasi mengukur proporsi total prediksi yang benar dari keseluruhan data. Metrik ini memberikan gambaran umum tentang performa model.

$$Accuracy = \frac{Juml \text{ Prediksi Benar}}{Total \text{ Jumlah Prediksi}} = \frac{TP+TN}{TP+FN+TN+FP} \quad (2.16)$$

Presisi (*Precision*)

Presisi mengukur berapa banyak dari data yang diprediksi sebagai positif yang benar-benar positif. Metrik ini relevan ketika nilai dari False Positive tinggi.

$$Precision = \frac{TP}{TP+FP} \quad (2.17)$$

Recall

Recall mengukur kemampuan model untuk mengidentifikasi kembali semua data yang sebenarnya positif. Metrik ini relevan ketika nilai dari False Negative tinggi.

$$Recall = \frac{TP}{TP+FN} \quad (2.18)$$

F1-Score

F1-Score adalah rata-rata harmonis dari precision dan recall. Metrik ini memberikan skor tunggal yang menyeimbangkan keduanya, sangat berguna pada dataset yang tidak seimbang.

$$F1\text{-Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.19)$$

2.9 Penelitian Terdahulu

Tabel 2.1 Penelitian Terdahulu

Penelitian terkait	Permasalahan	Solusi	Hasil
Sidabutar & Juardi (2025)[2]	Analisis sentimen masyarakat terhadap aplikasi halodoc	Melakukan analisis sentiment berbasis <i>text mining</i> terhadap ulasan pengguna Halodoc	Sentiment pengguna cenderung positif, tetapi belum ada pengujian dengan
Sowjanya et al. (2023)[10]	Ketidakseimbangan kelas pada dataset kesehatan menyebabkan penurunan akurasi model.	Mengembangkan metode Modified SMOTE untuk menyeimbangkan dataset kesehatan sebelum klasifikasi.	Akurasi dan F1-Score meningkat signifikan dibandingkan SMOTE standar pada data tidak seimbang.
Budaya & Suniantara (2024)[8]	Perbandingan algoritma analisis sentimen pada Google Reviews untuk puskesmas	Menggunakan kombinasi SMOTE + TF-IDF dengan beberapa algoritma machine learning	Kombinasi SMOTE + TF-IDF terbukti meningkatkan akurasi dibandingkan tanpa SMOTE.
P. Indobert (2025)[14]	Perbedaan performa model berdasarkan metode pelabelan (manual, leksikon, transformer)	Membandingkan pelabelan data manual, leksikon, dan transformer untuk performa IndoBERT	Model dengan pelabelan transformer cenderung unggul tetapi pelabelan manual dan leksikon memberikan

			distribusi kelas lebih stabil
M. I. Adrian & E. A. Laksana (2025)[15]	Penentuan parameter SVM yang kurang optimal pada klasifikasi sentimen	Menggunakan algoritma PSO untuk optimasi nilai parameter SVM	Optimasi parameter, khususnya parameter C, meningkatkan akurasi dan F1-score dibanding SVM tanpa optimasi
F. Dewi et al. (2025)[16]	Sensitivitas performa model terhadap pilihan hyperparameter	Studi tuning hyperparameter SVM menggunakan grid/eksperimen	Diperoleh bahwa nilai C dan parameter lain berdampak signifikan terhadap metrik seperti presisi dan recall
S. N. Khan et al. (2023)[17]	Perbedaan generalisasi SVM linear vs non-linear di berbagai domain	Tinjauan karakteristik SVM linear dan non-linear	SVM linear cenderung lebih stabil pada data teks berdimensi tinggi dibanding non-linear

Pada Tabel 2.1 diperlihatkan penelitian-penelitian terdahulu yang berkaitan dengan penelitian ini berikut penjelasan.

Penelitian oleh Sidabutar dan Juardi (2025)[2] membahas analisis sentimen terhadap ulasan pengguna aplikasi Halodoc dengan tujuan untuk mengetahui kecenderungan opini masyarakat terhadap layanan kesehatan digital yang

disediakan oleh aplikasi tersebut. Permasalahan utama yang diangkat dalam penelitian ini adalah besarnya jumlah ulasan pengguna yang masuk ke platform aplikasi, sehingga analisis secara manual menjadi tidak efektif dan memerlukan waktu yang lama. Selain itu, variasi bahasa yang digunakan oleh pengguna, seperti penggunaan kata tidak baku, singkatan, serta ekspresi emosional, turut menyulitkan proses analisis sentimen secara konvensional. Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan pendekatan text mining dalam proses klasifikasi sentimen, sehingga ulasan pengguna dapat dikelompokkan secara otomatis ke dalam kategori sentimen tertentu. Hasil penelitian menunjukkan bahwa sentimen pengguna terhadap aplikasi Halodoc cenderung positif. Meskipun demikian, penelitian ini belum mengkaji lebih lanjut pengaruh variasi parameter pada model klasifikasi yang digunakan, serta belum mempertimbangkan kondisi ketidakseimbangan data sentimen. Oleh karena itu, masih terdapat peluang untuk melakukan pengembangan, khususnya dalam hal evaluasi parameter model dan penerapan teknik penanganan data tidak seimbang.

Penelitian selanjutnya dilakukan oleh Sowjanya et al. (2023)[10] yang secara khusus menyoroti permasalahan ketidakseimbangan kelas (class imbalance) pada dataset di bidang kesehatan. Permasalahan ini muncul ketika jumlah data pada kelas tertentu, terutama kelas minoritas, jauh lebih sedikit dibandingkan kelas lainnya, sehingga model klasifikasi cenderung bias terhadap kelas mayoritas. Kondisi tersebut berdampak pada rendahnya kemampuan model dalam mengenali kelas minoritas, yang tercermin dari nilai akurasi dan F1-score yang kurang optimal. Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan penggunaan metode Modified Synthetic Minority Over-sampling Technique (Modified SMOTE) sebagai upaya penyeimbangan data sebelum proses klasifikasi dilakukan. Hasil eksperimen menunjukkan bahwa metode Modified SMOTE mampu meningkatkan performa model secara signifikan jika dibandingkan dengan penggunaan SMOTE standar. Namun demikian, penelitian ini tidak diterapkan pada data teks ulasan berbahasa Indonesia dan tidak secara spesifik mengombinasikan metode tersebut dengan algoritma Support Vector Machine (SVM) dalam konteks analisis sentimen.

Penelitian oleh Budaya dan Suniantara (2024)[8] membahas perbandingan beberapa algoritma analisis sentimen pada data ulasan Google Reviews puskesmas. Permasalahan utama yang diangkat dalam penelitian ini adalah ketidakseimbangan distribusi data sentimen yang berdampak langsung pada performa model klasifikasi. Untuk mengatasi permasalahan tersebut, penelitian ini mengombinasikan teknik SMOTE dan pembobotan fitur menggunakan TF-IDF, kemudian diterapkan pada beberapa algoritma machine learning. Hasil penelitian menunjukkan bahwa penerapan SMOTE mampu meningkatkan akurasi model dibandingkan dengan kondisi tanpa SMOTE. Meskipun demikian, penelitian ini belum membahas secara mendalam pengaruh variasi parameter C pada algoritma SVM, serta tidak melakukan perbandingan terhadap metode pelabelan data yang digunakan. Dengan demikian, aspek pengaruh parameter dan kualitas pelabelan data masih belum menjadi fokus utama dalam penelitian ini.

Penelitian yang dilakukan oleh P. Indobert (2025)[14] berfokus pada pengaruh metode pelabelan data terhadap performa model analisis sentimen. Permasalahan yang diangkat berkaitan dengan perbedaan kualitas hasil pelabelan dan distribusi kelas yang dihasilkan dari metode pelabelan manual, leksikon, dan transformer-based. Penelitian ini membandingkan ketiga metode pelabelan tersebut terhadap performa model IndoBERT. Hasil penelitian menunjukkan bahwa pelabelan berbasis transformer menghasilkan performa terbaik dari sisi akurasi dan F1-score, sementara pelabelan manual dan leksikon cenderung menghasilkan distribusi kelas yang lebih stabil. Meskipun memberikan kontribusi penting terkait kualitas pelabelan data, penelitian ini tidak menggunakan algoritma SVM dan belum membahas penanganan data tidak seimbang menggunakan teknik SMOTE.

Penelitian oleh Adrian dan Laksana (2025)[15] mengkaji permasalahan penentuan parameter SVM yang kurang optimal dalam klasifikasi sentimen ulasan IMDb. Dalam penelitian ini, parameter C pada SVM dianggap memiliki peran penting dalam menentukan keseimbangan antara margin pemisah dan kesalahan klasifikasi. Permasalahan yang diangkat adalah penggunaan nilai parameter C yang tidak optimal dapat menyebabkan menurunnya performa model. Untuk mengatasi hal tersebut, penelitian ini menerapkan algoritma Particle Swarm Optimization