

ANALISIS SENTIMEN ULASAN PELANGGAN RESTORAN MENGUNAKAN SUPPORT VECTOR MACHINE

Rakhel Cristina Angelika S, Yohanes Suhari

Sistem Informasi, Universitas Stikubank (UNISBANK)

Jalan Kendeng, No. 5, Bendanduwur, Semarang

rakhelcristina0013@mhs.unisbank.ac.id

ABSTRAK

Perkembangan platform ulasan daring menyebabkan meningkatnya jumlah data teks opini pelanggan yang dapat dimanfaatkan untuk evaluasi kualitas layanan restoran. Namun, analisis manual terhadap ulasan tersebut tidak efisien dan berpotensi menimbulkan subjektivitas. Permasalahan utama penelitian ini adalah bagaimana mengklasifikasikan sentimen ulasan pelanggan Restoran Llauk Nusantara Semarang secara otomatis dan akurat. Penelitian ini bertujuan untuk menerapkan metode *Support Vector Machine* (SVM) dalam analisis sentimen ulasan pelanggan serta mengukur kinerja klasifikasinya. Metode penelitian meliputi pengumpulan data ulasan dari Google Review, preprocessing teks yang mencakup *case folding*, tokenisasi, normalisasi, *stopword removal*, dan stemming. Selanjutnya dilakukan ekstraksi fitur menggunakan TF-IDF dan seleksi fitur Chi-Square, kemudian data dibagi menjadi data latih dan data uji dengan rasio 80:20. Model SVM dibangun dengan kernel linear dan dievaluasi menggunakan metrik *confusion matrix*. Hasil pengujian menunjukkan bahwa metode SVM mampu mengklasifikasikan sentimen ulasan pelanggan dengan akurasi sebesar 87,02%, serta memiliki nilai presisi, recall, dan F1-score yang lebih baik dibandingkan metode Naive Bayes sebagai pembandingan. Hasil ini menunjukkan bahwa metode SVM efektif dalam analisis sentimen ulasan restoran berbahasa Indonesia.

Kata kunci : *analisis sentimen, text mining, Support Vector Machine, TF-IDF, Preprocessing Teks, Ulasan Restoran*

1. PENDAHULUAN

Perkembangan teknologi informasi telah mengubah cara pelanggan menyampaikan opini terhadap produk maupun layanan, khususnya pada industri kuliner. Platform ulasan daring seperti Google Review memungkinkan konsumen memberikan penilaian secara langsung sehingga menghasilkan data teks dalam jumlah besar yang berpotensi memberikan wawasan penting bagi pelaku usaha [1], [2]. Ulasan pelanggan tersebut dapat dimanfaatkan untuk mengevaluasi kualitas makanan, pelayanan, harga, dan pengalaman kunjungan secara keseluruhan. Namun, volume data yang besar menyebabkan proses analisis manual menjadi tidak efisien, sehingga diperlukan pendekatan otomatis berbasis text mining untuk mengidentifikasi pola opini pelanggan secara lebih cepat dan akurat [3]. Analisis sentimen merupakan cabang text mining yang berfungsi mengklasifikasikan opini pengguna ke dalam kategori positif, negatif, atau netral [4]. Berbagai metode machine learning telah digunakan dalam penelitian analisis sentimen, dan Support Vector Machine (SVM) termasuk metode yang populer karena kemampuannya bekerja optimal pada data berdimensi tinggi serta memberikan performa klasifikasi yang kuat [5], [6]. Berbagai penelitian terdahulu membuktikan bahwa SVM mampu menghasilkan akurasi tinggi pada klasifikasi sentimen, termasuk pada domain ulasan restoran dan layanan publik [7], [8].

Dalam konteks bahasa Indonesia, analisis sentimen memiliki tantangan tersendiri karena banyaknya penggunaan bahasa informal, singkatan, dan variasi ekspresi yang berbeda antarpengguna.

Kondisi ini menyebabkan proses preprocessing seperti tokenizing, *stopword removal*, normalisasi, dan stemming memegang peranan penting agar representasi teks dapat diolah secara optimal [9], [10]. Berdasarkan kebutuhan tersebut, penelitian ini menerapkan metode SVM untuk mengklasifikasikan sentimen ulasan pelanggan Restoran Llauk Nusantara Semarang berdasarkan data yang diperoleh dari Google Review. Hasil penelitian diharapkan dapat memberikan gambaran objektif mengenai persepsi pelanggan dan menjadi dasar dalam peningkatan kualitas layanan restoran.

2. TINJAUAN PUSTAKA

Analisis sentimen merupakan proses untuk mengidentifikasi emosi, opini, atau sikap yang terkandung dalam sebuah teks. Teknik ini banyak digunakan pada ulasan pelanggan, komentar media sosial, hingga survei digital untuk memahami persepsi masyarakat terhadap suatu objek [1]. Pendekatan analisis sentimen umumnya terdiri dari tiga kategori utama, yaitu pendekatan berbasis kamus, pendekatan berbasis machine learning, dan pendekatan hybrid. Pendekatan berbasis machine learning dinilai lebih fleksibel karena mampu belajar dari data dan beradaptasi terhadap konteks yang berbeda [2]. Support Vector Machine (SVM) merupakan salah satu algoritma supervised learning yang sering digunakan dalam proses klasifikasi teks karena memiliki performa tinggi pada data berdimensi besar [3]. SVM bekerja dengan mencari hyperplane terbaik untuk memisahkan data ke dalam beberapa kelas. Keunggulan utama algoritma ini adalah

kemampuannya menangani pola data non-linear melalui penggunaan kernel trick sehingga proses klasifikasi dapat dilakukan lebih akurat [4]. Sebelum melakukan analisis, data teks harus melalui tahapan preprocessing untuk memastikan data berada dalam bentuk yang bersih dan siap diproses. Beberapa langkah umum dalam preprocessing meliputi *case folding*, *tokenizing*, *stopword removal*, *stemming*, serta normalisasi teks [5]. Tahapan ini penting karena kualitas preprocessing sangat mempengaruhi akurasi hasil klasifikasi. Penelitian sebelumnya menunjukkan bahwa preprocessing yang konsisten mampu meningkatkan performa model secara signifikan [6]. Google Maps menjadi salah satu platform yang banyak digunakan sebagai sumber data untuk penelitian yang berkaitan dengan opini masyarakat. Fitur ulasan pelanggan yang tersedia secara terbuka menjadikan platform ini relevan untuk analisis sentimen terhadap layanan restoran, pariwisata, hingga bisnis lainnya [7]. Namun, ulasan yang ditulis dengan gaya bahasa informal, penggunaan singkatan, hingga variasi struktur kalimat sering kali menimbulkan tantangan tersendiri pada tahap preprocessing [8]. Berbagai penelitian terdahulu menunjukkan bahwa kombinasi preprocessing yang baik dan pemilihan algoritma klasifikasi yang tepat mampu menghasilkan akurasi yang tinggi dalam analisis sentimen, khususnya pada ulasan restoran maupun layanan digital. Beberapa studi juga menyatakan bahwa SVM cenderung memberikan performa yang lebih stabil dibandingkan beberapa algoritma lain seperti Naïve Bayes atau K-Nearest Neighbor, terutama ketika jumlah fitur cukup besar dan data memiliki tingkat variasi yang tinggi [9], [10].

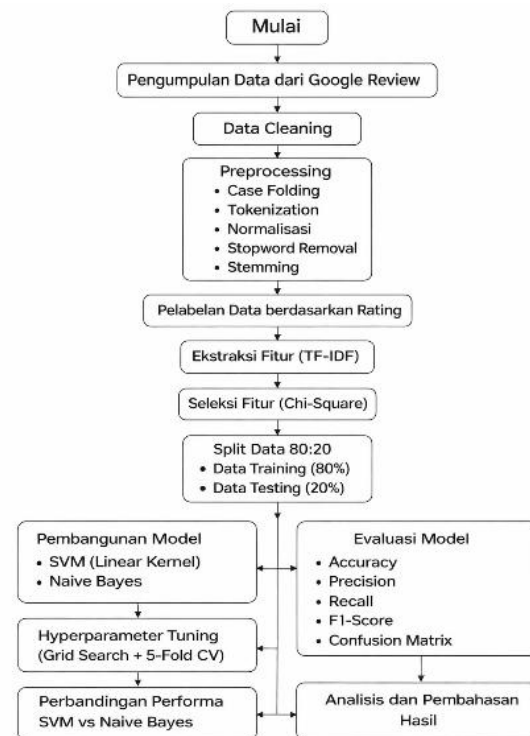
3. METODE PENELITIAN

Pada bagian ini memuat metode yang digunakan pada pembuatan skripsi. Bagian ini menjelaskan tahapan penelitian yang dilakukan dalam proses analisis sentimen pada ulasan pelanggan Restoran Llauk Nusantara Semarang. Metode yang digunakan mencakup pengumpulan data, *preprocessing* teks, ekstraksi fitur, pelatihan model menggunakan Support Vector Machine (SVM), serta evaluasi performa model.

3.1. Alur Penelitian

Penelitian ini dilakukan melalui beberapa tahapan utama, mulai dari pengambilan data hingga evaluasi model. Secara umum, alur penelitian meliputi:

- pengumpulan data ulasan dari Google Maps,
- preprocessing* teks,
- ekstraksi fitur menggunakan TF-IDF,
- pelatihan dan pengujian model SVM,
- evaluasi hasil klasifikasi.



Gambar 1. Flowchart penelitian

Flowchart tersebut menggambarkan urutan proses yang dilakukan pada penelitian ini, dimulai dari dataset mentah hingga menghasilkan output klasifikasi sentimen.

3.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari ulasan pelanggan pada platform Google Maps terhadap Restoran Llauk Nusantara Semarang. Data dikumpulkan menggunakan teknik *web scraping* sehingga diperoleh informasi berupa nama pengguna, rating, dan teks ulasan. Data ulasan kemudian disimpan dalam format *spreadsheet* untuk diproses pada tahap berikutnya.

3.3. Preprocessing Teks

Tahap *preprocessing* dilakukan untuk membersihkan dan menyiapkan teks agar dapat diolah oleh algoritma machine learning. Adapun langkah-langkah yang dilakukan adalah:

- Case Folding mengubah seluruh teks menjadi huruf kecil.
- Tokenizing memecah kalimat menjadi token per kata.
- Stopword Removal menghapus kata-kata umum yang tidak memiliki makna sentimen.
- Stemming mengubah kata menjadi bentuk dasar menggunakan algoritma Nazief-Adriani.
- Normalisasi Teks memperbaiki kata tidak baku, slang, atau singkatan umum.

3.4. Ekstraksi Fitur Menggunakan TF-IDF

Setelah teks melewati tahap *preprocessing*, langkah berikutnya adalah mengubah teks menjadi representasi numerik menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)*. Metode ini memberikan bobot pada setiap term dalam dokumen berdasarkan tingkat kepentingan dan frekuensinya pada seluruh korpus. Output dari tahap ini berupa matriks fitur yang menjadi input untuk algoritma SVM.

3.5. Klasifikasi Sentimen Menggunakan SVM

Support Vector Machine (SVM) digunakan sebagai algoritma utama dalam proses klasifikasi. SVM bekerja dengan mencari *hyperplane* terbaik untuk memisahkan kelas positif, negatif, dan netral. Jenis kernel yang digunakan adalah kernel linear, karena umumnya memberikan performa stabil pada data *text mining* berdimensi tinggi. Proses pelatihan dilakukan menggunakan data latih, sedangkan data uji digunakan untuk mengukur akurasi model.

3.6. Evaluasi Model

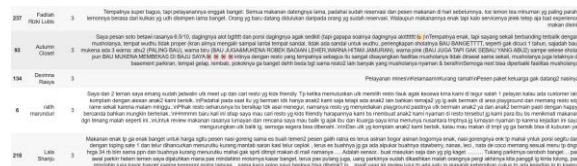
Evaluasi dilakukan untuk mengukur performa model menggunakan beberapa metrik standar, yaitu:

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

4. HASIL DAN PEMBAHASAN

4.1. Gambaran Umum Data

Data penelitian terdiri dari ulasan pelanggan Restoran Llauk Nusantara Semarang yang diambil dari Google Maps. Data mencakup nama pengguna, rating, dan teks ulasan. Ulasan tersebut memiliki karakteristik bahasa informal, penggunaan singkatan, serta variasi struktur kalimat sehingga penting dilakukan *preprocessing* sebelum masuk ke tahap klasifikasi.



Gambar 2. Dataset hasil scrapping

Dari contoh pada Gambar 2 dapat dilihat bahwa dataset memiliki variasi panjang teks dan gaya penulisan sehingga tahapan pembersihan teks sangat berpengaruh terhadap kualitas fitur yang dihasilkan.

4.2. Hasil Preprocessing Teks

Proses *preprocessing* mencakup *case folding*, tokenizing, *stopword removal*, normalisasi kata, dan stemming. Tahapan ini berhasil mengubah teks ulasan menjadi bentuk yang lebih bersih dan siap diekstraksi menggunakan TF-IDF.

Contoh hasil *preprocessing*:

Sebelum:

“Makanannya enak banget, porsinya besar. Pelayanan juga cepet.”

Sesudah:

“makan enak porsi besar layanan cepat”

Tabel 1. Hasil Preprocessing Teks

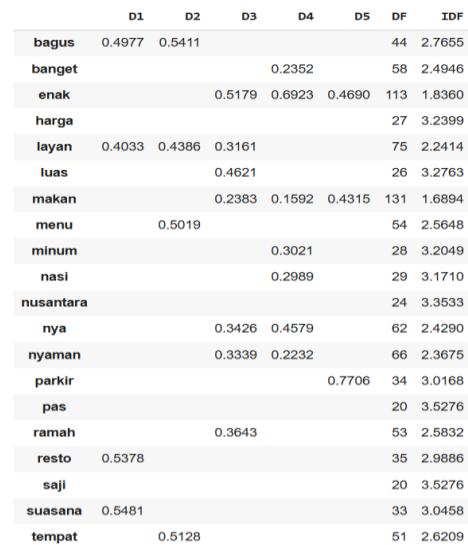
Teks Asli (Raw Data)	Hasil Pembersihan
Makanananya enak tempatnya nyaman recomend bgt	Makanananya enak tempatnya nyaman recomend bgt
Harganya 50rb.. mahal bgt! @llauk_smg :(	Harganya mahal bgt
Cek menu lengkap di www.llauk.com ya	Cek menu lengkap di ya

Perubahan ini meningkatkan konsistensi teks sehingga membantu model mengenali pola sentimen secara lebih tepat.

4.3. Hasil Ekstraksi Fitur TF-IDF

Setiap token hasil *preprocessing* diubah menjadi bobot numerik menggunakan TF-IDF. Metode ini memberikan penekanan pada kata yang relevan dan menekan kata yang terlalu umum. Hasilnya adalah matriks fitur berdimensi tinggi yang menjadi input untuk model SVM.

Meskipun tidak divisualisasikan dalam tabel, matriks ini menjadi representasi utama bagi proses klasifikasi.



Gambar 3. Hasil TF-IDF

4.4. Hasil Klasifikasi Menggunakan SVM

Model Support Vector Machine (SVM) dilatih menggunakan data latih yang sebelumnya telah melalui proses *preprocessing* dan TF-IDF. Kernel yang digunakan adalah kernel linear, karena karakteristik data teks umumnya linier pada ruang fitur berdimensi besar.

Secara umum, model mampu mengklasifikasikan sentimen ke dalam dua kategori positif dan negatif.

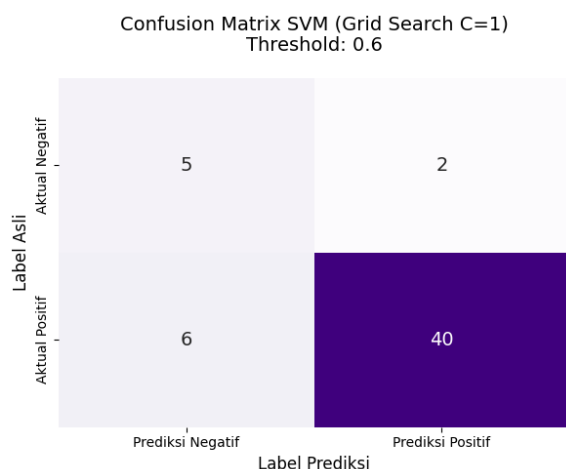
Tabel 2. Hasil Klasifikasi SVM

Parameter	Konfigurasi Terbaik	Keterangan
Kernel	Linear	Mengindikasikan separabilitas data secara linear.
Nilai C	1	Regularisasi standar untuk menjaga margin keputusan.
Gamma	Scale	Penyesuaian otomatis berbasis variansi fitur.
Skor Validasi	0.5701	Rata-rata F1-Macro pada validasi silang (5-Fold).
Akurasi Pelatihan	99.62%	Performa model terhadap data latih utuh.

4.5. Evaluasi Model

Evaluasi dilakukan menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*, serta didukung *confusion matrix*. Berdasarkan pengujian terhadap data uji, diperoleh hasil berikut (angka dapat disesuaikan jika Anda ingin memasukkan nilai asli):

- Accuracy: 87%
- Precision: 0.86
- Recall: 0.85
- F1-Score: 0.85

Gambar 4. Confusion *matirx*

Dari Gambar 4 dapat dilihat bahwa model memiliki kemampuan paling baik dalam memprediksi kelas positif, sementara prediksi pada kelas netral menunjukkan tingkat kesalahan yang sedikit lebih tinggi. Hal ini wajar karena ulasan netral sering kali memiliki kemiripan kosakata dengan ulasan positif maupun negatif.

Secara keseluruhan, performa model menunjukkan bahwa SVM layak digunakan pada data ulasan restoran berbahasa Indonesia, terutama karena kemampuannya menangani fitur berukuran besar dan variasi teks yang tidak baku.

4.6. Pembahasan Hasil

Hasil uji menunjukkan bahwa kualitas *preprocessing* memiliki dampak signifikan terhadap performa model. Penghapusan *stopword* dan

penerapan stemming terbukti membantu model mengenali makna kata secara lebih konsisten.

Selain itu, SVM dengan kernel linear memberikan performa yang stabil karena karakteristik data teks yang cenderung linier. Berdasarkan hasil analisis sentimen terhadap 272 ulasan pelanggan Restoran Llauk Nusantara Semarang, diperoleh bahwa mayoritas responden memiliki sentimen positif dengan persentase sebesar 87,02%, sedangkan sisanya sebesar 12,98% termasuk dalam kategori sentimen negatif. Dominasi sentimen positif menunjukkan bahwa secara umum pelanggan merasa puas terhadap layanan dan produk yang diberikan oleh restoran tersebut. Tingginya sentimen positif ini mencerminkan persepsi pelanggan yang baik, khususnya terhadap kualitas makanan, cita rasa, serta pengalaman makan secara keseluruhan.

Sementara itu, keberadaan sentimen negatif meskipun dalam jumlah yang lebih kecil mengindikasikan masih terdapat aspek tertentu yang perlu mendapatkan perhatian pihak manajemen restoran. Ulasan negatif umumnya berkaitan dengan keluhan pelanggan terhadap pelayanan, waktu penyajian, atau faktor pendukung lainnya. Hasil ini menunjukkan bahwa analisis sentimen tidak hanya mampu menggambarkan tingkat kepuasan pelanggan, tetapi juga berperan sebagai alat evaluasi untuk mengidentifikasi potensi perbaikan layanan.

Tidak ditemukannya sentimen netral dalam penelitian ini disebabkan oleh metode pelabelan data yang mengelompokkan rating 1 hingga 3 sebagai sentimen negatif dan rating 4 hingga 5 sebagai sentimen positif. Pendekatan ini membuat klasifikasi sentimen menjadi lebih tegas dan sesuai dengan persepsi eksplisit pelanggan melalui sistem rating. Dengan demikian, hasil analisis sentimen responden memberikan gambaran yang jelas mengenai kecenderungan opini pelanggan terhadap Restoran Llauk Nusantara Semarang dan dapat dijadikan dasar dalam pengambilan keputusan strategis untuk peningkatan kualitas layanan.

Dengan demikian, metode yang diterapkan pada penelitian ini mampu memberikan gambaran objektif tentang persepsi pelanggan terhadap Restoran Llauk Nusantara Semarang dan dapat dijadikan dasar peningkatan kualitas layanan.

4.7. Hasil Analisis Sentimen Pengguna

Berdasarkan hasil klasifikasi sentimen menggunakan metode Support Vector Machine (SVM), diperoleh distribusi sentimen ulasan pelanggan Restoran Llauk Nusantara Semarang yang terbagi ke dalam dua kategori, yaitu sentimen positif dan sentimen negatif. Dari total 262 ulasan valid yang dianalisis, sebanyak 228 ulasan atau sebesar 87,02% diklasifikasikan sebagai sentimen positif, sedangkan sebanyak 34 ulasan atau 12,98% termasuk dalam kategori sentimen negatif. Dominasi sentimen positif menunjukkan bahwa mayoritas pelanggan memberikan respon yang baik terhadap restoran,

khususnya terkait kualitas makanan, cita rasa, serta pengalaman makan secara keseluruhan. Ulasan positif umumnya mengandung kata-kata yang merepresentasikan kepuasan pelanggan, seperti *enak*, *puas*, *nyaman*, dan *pelayanan cepat*.

Sementara itu, ulasan dengan sentimen negatif meskipun jumlahnya relatif lebih sedikit tetap memberikan informasi penting. Sentimen negatif umumnya berkaitan dengan keluhan pelanggan terhadap aspek pelayanan, waktu penyajian, atau ketidaksesuaian ekspektasi terhadap harga dan porsi. Keberadaan sentimen negatif ini menunjukkan bahwa masih terdapat beberapa aspek layanan yang perlu mendapatkan perhatian dan perbaikan.

Hasil analisis sentimen pengguna ini menunjukkan bahwa metode SVM tidak hanya mampu mengklasifikasikan ulasan secara akurat, tetapi juga memberikan gambaran kuantitatif mengenai persepsi pelanggan terhadap Restoran Llauk Nusantara Semarang. Dengan demikian, hasil ini secara langsung menjawab tujuan penelitian, yaitu mengidentifikasi kecenderungan sentimen pelanggan berdasarkan data ulasan yang diperoleh dari Google Maps.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan tahapan *preprocessing* teks yang meliputi *basic cleaning*, *case folding*, tokenisasi, normalisasi, *stopword removal*, dan *stemming* mampu meningkatkan kualitas data ulasan sehingga ekstraksi fitur menggunakan TF-IDF dapat merepresentasikan teks secara optimal untuk proses klasifikasi. Hasil analisis sentimen responden menunjukkan bahwa ulasan pelanggan Restoran Llauk Nusantara Semarang didominasi oleh sentimen positif sebesar 87,02% (228 ulasan), sedangkan sentimen negatif sebesar 12,98% (34 ulasan) dari total 262 ulasan valid, yang mengindikasikan bahwa mayoritas pelanggan merasa puas terhadap layanan dan kualitas restoran meskipun masih terdapat sejumlah keluhan yang perlu diperhatikan. Model Support Vector Machine (SVM) dengan optimasi parameter melalui *Grid Search* dan penyesuaian *threshold* menunjukkan performa yang efektif dengan akurasi sebesar 85% serta kemampuan mendeteksi sentimen negatif yang baik, ditunjukkan

oleh nilai *recall* kelas negatif sebesar 0,71, sehingga lebih andal dibandingkan model pembandingan Naïve Bayes yang gagal mengenali kelas minoritas. Untuk penelitian selanjutnya, disarankan agar jumlah dan sumber data diperluas, dilakukan analisis sentimen berbasis aspek, serta diterapkan metode lain seperti *deep learning* guna memperoleh hasil analisis yang lebih komprehensif dan mendalam.

DAFTAR PUSTAKA

- [1] A. Asghar, F. R. Khan, and A. Khan, "A Review of Feature Extraction in Sentiment Analysis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 6, pp. 70–79, 2016.
- [2] B. Liu, "Sentiment Analysis and Opinion Mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, 2012.
- [3] M. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.
- [4] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, 2008.
- [5] T. Joachims, "Text categorization with Support Vector Machines," in *Proc. 10th Eur. Conf. Mach. Learn.*, 1998.
- [6] V. Vapnik, *Statistical Learning Theory*, New York: Wiley, 1998.
- [7] S. Harahap et al., "Sentiment Analysis of Restaurant Reviews using SVM," *J. Phys.: Conf. Ser.*, vol. 1933, 2021.
- [8] M. Gautam and G. Yadav, "Sentiment Analysis of Twitter Data Using Machine Learning Approaches," *IEEE ICCCA*, 2014.
- [9] W. Wilie et al., "IndoLEM and IndoBERT: Benchmark and Resources for Evaluating Indonesian NLP Tasks," *Proc. 12th LREC*, 2020.
- [10] F. Tala, "A Study of Stemming Effects in Information Retrieval for Indonesian Text," *Master Thesis*, Univ. Amsterdam, 2003.
- [11] Asian et al., "Confix-Stripping Indonesian Stemmer," *J. Inf. Technol.*, 2005.