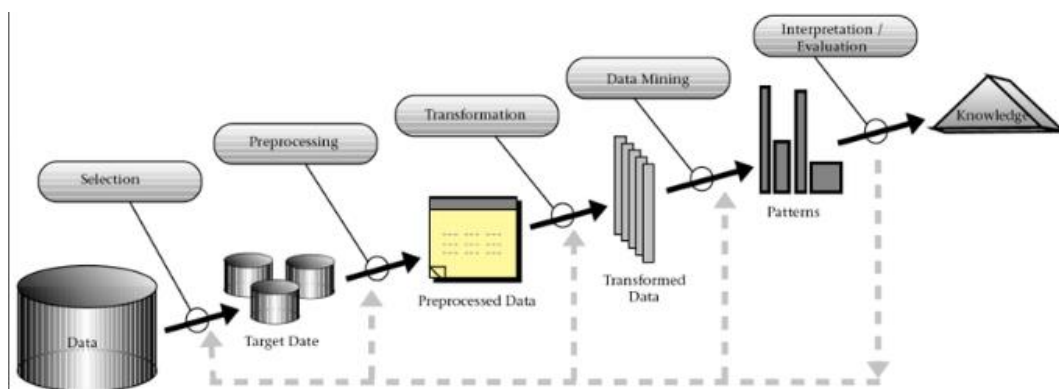


BAB II

LANDASAN TEORI

2.1. *Knowledge Discovery in Database (KDD)*

Knowledge Discovery in Database (KDD) merupakan proses sistematis untuk menemukan dan mengubah informasi bernilai menjadi pengetahuan yang berguna dari kumpulan data berukuran besar. Data mining menjadi salah satu tahap penting yang berfungsi menggali pola menggunakan metode atau algoritma tertentu, sehingga data mining merupakan bagian dari keseluruhan proses KDD dan tidak berdiri sendiri (Pengembangan et al., 2025). Proses KDD meliputi pemilihan data, pembersihan, transformasi, serta penerapan algoritma klasifikasi seperti SVM untuk menentukan polaritas sentimen pelanggan guna mendukung pengambilan keputusan berbasis data dalam peningkatan layanan kedai kopi. Pendekatan ini telah banyak digunakan dalam analisis sentimen pelanggan dan terbukti efektif dalam mengevaluasi kualitas pelayanan secara otomatis (Dwilestari et al., 2025).



Gambar 2.1 Tahapan Dalam Knowledge Discovery in Database

Sumber : https://www.researchgate.net/figure/Metode-Knowledge-Discovery-and-Data-Mining-KDD_fig1_361105655

Tahapan *Knowledge Discovery in Database* (KDD) adalah sebagai berikut (Ubaid et al., 2025):

1. Pemilihan Data (Data Selection)

Tahap awal untuk memilih data yang relevan dengan tujuan penelitian dari sumber data yang tersedia, sehingga hanya data yang memiliki keterkaitan langsung dengan permasalahan yang diproses lebih lanjut.

2. Preprocessing

Tahap pembersihan data dari nilai kosong (*missing values*), duplikasi, dan inkonsistensi untuk meningkatkan kualitas data agar hasil analisis menjadi lebih akurat.

3. Transformasi

Tahap mengubah data ke dalam format yang sesuai dengan metode atau algoritma yang digunakan, seperti normalisasi data atau konversi data teks menjadi representasi numerik.

4. Data Mining

Tahap inti dalam KDD yang menerapkan algoritma tertentu untuk menemukan pola, hubungan, atau model dari data yang telah diproses, misalnya menggunakan algoritma klasifikasi.

5. Evaluasi (*Interpretation/Evaluation*)

Tahap penilaian hasil data mining untuk memastikan pola atau model yang dihasilkan valid, akurat, dan dapat digunakan sebagai dasar pengambilan keputusan.

2.2. Data Mining

Data mining merupakan rangkaian proses untuk mengekstraksi pola, hubungan, atau informasi yang bernilai dari kumpulan data dalam jumlah besar, yang selanjutnya dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan yang lebih akurat dan efektif. Salah satu pendekatan utama dalam data mining adalah klasifikasi, yaitu proses pengelompokan data ke dalam kelas atau kategori tertentu berdasarkan karakteristik yang dimiliki, seperti pengelompokan data berdasarkan kualitas pelayanan. Dalam penerapannya, terdapat berbagai algoritma klasifikasi yang sering digunakan, antara lain *Naive Bayes*, *Decision Tree*, *Support Vector Machine (SVM)*, dan *K-Nearest Neighbors (KNN)*, yang masing-masing memiliki keunggulan dan karakteristik tersendiri dalam mengolah serta menganalisis data (Knn et al., 2025).

Data mining merupakan bagian dari machine learning yang berfokus pada proses penemuan pola, hubungan, atau informasi penting dari data yang telah dipilih dengan menerapkan teknik atau metode tertentu. Proses ini bertujuan untuk mengekstraksi pengetahuan yang tersembunyi dalam data sehingga dapat memberikan wawasan yang bernilai. Informasi yang dihasilkan dari data mining selanjutnya dapat dimanfaatkan untuk mendukung pengambilan keputusan yang lebih tepat dan berbasis data (Informasi et al., 2025)

2.3. Evaluasi

Evaluasi adalah proses memperkirakan atau meramalkan nilai, hasil, atau perilaku suatu variabel di masa depan atau kondisi yang belum terjadi berdasarkan pola atau informasi yang ditemukan dari data historis. Dalam konteks data mining

dan machine learning, evaluasi dilakukan dengan membangun model dari data yang ada sehingga dapat digunakan untuk menentukan kemungkinan hasil pada data baru, seperti mengevaluasi kualitas pelayanan dari ulasan teks mereka menggunakan algoritma klasifikasi(Endra & Rizal, n.d.).

2.4. Kualitas Pelayanan

Kualitas pelayanan dalam industri kedai kopi dipengaruhi oleh berbagai aspek seperti kualitas produk, harga, pelayanan, serta kenyamanan tempat, yang mana faktor-faktor ini secara signifikan memengaruhi persepsi konsumen terhadap kualitas pelayanan. Untuk memahami hal tersebut, banyak penelitian menggunakan survei pelanggan dan ulasan sebagai sumber data utama karena dapat menggambarkan persepsi dan pengalaman konsumen secara langsung dalam konteks bisnis kedai kopi(Atmosphere et al., 2023).

Faktor-Faktor yang Mempengaruhi Kualitas Pelayanan Adalah sebagai berikut (A. A. Maulana, 2024):

1. **Kualitas Produk dan Pelayanan:** Evaluasi kualitas pelayanan sangat dipengaruhi oleh penilaian pelanggan terhadap mutu produk dan kualitas layanan yang mereka terima.
2. **Kewajaran Harga:** Harga yang dirasakan wajar dan sebanding dengan manfaat yang diperoleh turut memberikan kontribusi terhadap penilaian kualitas pelayanan yang diterima pelanggan.
3. **Komunikasi yang Efektif:** Penyampaian informasi yang jelas dan interaksi yang baik antara perusahaan dan pelanggan dapat meningkatkan

pemahaman serta membentuk persepsi pelanggan terhadap kualitas pelayanan.

2.5. Analisis Sentimen

Analisis sentimen adalah proses untuk menggali, memahami, serta mengelompokkan opini atau ekspresi perasaan yang terkandung dalam suatu teks. Teknik ini digunakan untuk menentukan kecenderungan sentimen suatu teks, apakah bersifat positif atau negatif (Pebrianti & Karawang, 2025).

Analisis sentimen merupakan teknik untuk mengidentifikasi dan mengkategorikan opini atau emosi yang terkandung dalam teks ulasan atau opini publik, membantu dalam memahami feedback pengguna terhadap suatu topik atau produk. Teknik ini banyak dipakai dalam penelitian berbasis teks untuk menilai persepsi dan sentimen masyarakat terhadap layanan, produk, atau kebijakan tertentu (Sulistiowati & Santoso, 2023).

Analisis sentimen bertujuan untuk mengidentifikasi, memahami, dan mengklasifikasikan opini atau emosi yang terkandung dalam data teks sehingga dapat menggambarkan persepsi dan sikap masyarakat terhadap suatu topik atau objek tertentu secara otomatis. Dengan teknik ini, bisnis dapat mengukur sentimen pelanggan secara objektif, memetakan kecenderungan positif, negatif, atau netral dalam ulasan, serta membantu pengambilan keputusan yang lebih tepat berdasarkan data nyata dari pelanggan (Fitriono et al., 2025). Dalam konteks penelitian seperti *Analisis Sentimen untuk Evaluasi Kualitas Pelayanan Kedai Kopi SAA*, tujuan analisis sentimen adalah menggali wawasan ulasan pelanggan

untuk menilai tingkat kualitas layanan terhadap produk dan pelayanan, yang kemudian dapat digunakan untuk strategi peningkatan kualitas bisnis.

Jenis-Jenis Sentimen Adalah sebagai berikut (Sosial et al., 2024):

1. Sentimen Positif

Merupakan kategori sentimen di mana teks menunjukkan opini, perasaan, atau evaluasi yang mendukung, pujian, atau kepuasan pelanggan terhadap suatu produk, layanan, atau pengalaman. Sentimen positif mencerminkan persepsi yang baik dari pelanggan terhadap objek yang dikomentari.

2. Sentimen Negatif

Merupakan kategori sentimen yang menunjukkan ketidakpuasan, kritik, kekecewaan, atau evaluasi buruk terhadap suatu produk atau layanan. Sentimen ini sering muncul dalam bentuk keluhan atau kritik pelanggan terhadap aspek yang dianggap kurang memuaskan.

3. Sentimen Netral

Sentimen netral mencakup teks yang tidak secara eksplisit bersifat positif maupun negatif. Ulasan netral biasanya bersifat informatif atau deskriptif tanpa ekspresi emosi yang kuat terhadap bagian tertentu dari produk atau layanan.

Untuk memberikan gambaran yang lebih jelas mengenai dasar penentuan kategori sentimen sebagaimana digunakan pada pengolahan data ulasan pelanggan di Bab III (Tabel 3.1, Tabel 3.9, dan Tabel 3.10), berikut diuraikan kriteria umum yang menjadi acuan dalam proses pelabelan sentimen positif, netral, dan negatif pada penelitian ini:

- a. Kriteria Sentimen Positif.** Ulasan dikategorikan positif apabila mengandung kata atau frasa yang menunjukkan pujian, kepuasan, atau

penilaian baik, seperti “enak”, “ramah”, “cepat”, “nyaman”, “memuaskan”, dan “nikmat”. Contohnya ulasan “Pelayanan ramah dan cepat” dan “Kopinya sangat nikmat” pada Tabel 3.1 dikategorikan positif karena secara eksplisit mengandung kata-kata yang menunjukkan apresiasi dan kepuasan pelanggan.

b. Kriteria Sentimen Negatif. Ulasan dikategorikan negatif apabila mengandung kata atau frasa yang menunjukkan keluhan, ketidakpuasan, atau kritik, seperti “kurang”, “kecewa”, “pahit”, “kotor”, “mahal”, dan “tidak ramah”. Contohnya ulasan “Kopinya pahit dan kurang enak” dan “Pelayanan lama dan tidak ramah” pada Tabel 3.1 dikategorikan negatif karena mengandung unsur kritik terhadap rasa dan pelayanan.

c. Kriteria Sentimen Netral. Ulasan dikategorikan netral apabila bersifat deskriptif atau informatif tanpa penilaian tegas ke arah positif maupun negatif, misalnya hanya menyatakan fakta atau kondisi apa adanya. Contohnya ulasan “Kopinya biasa saja” dan “Menu tidak terlalu banyak” pada Tabel 3.1 dikategorikan netral karena tidak mengandung kata sifat yang secara jelas mengarah pada pujian maupun kritik.

Kriteria di atas juga menjadi dasar dalam mengaitkan hasil klasifikasi sentimen dengan penilaian kualitas pelayanan (Baik/Kurang Baik) sebagaimana disajikan pada Bab III. Ulasan dengan sentimen positif maupun netral cenderung dikategorikan sebagai kualitas pelayanan baik, karena tidak menunjukkan adanya keluhan atau kekecewaan pelanggan yang signifikan. Sebaliknya, ulasan dengan sentimen negatif dikategorikan sebagai kualitas pelayanan kurang baik, karena mencerminkan adanya ketidakpuasan pelanggan terhadap produk atau layanan yang diterima.

Contoh penerapan analisis sentimen pada bisnis/UMKM :

Penerapan analisis sentimen pada bisnis/UMKM banyak dilakukan untuk menilai opini pelanggan secara otomatis dari data ulasan pengguna. Misalnya, analisis sentimen pada aplikasi *MPStore Super App UMKM* menggunakan SVM untuk mengevaluasi persepsi pengguna terhadap aplikasi tersebut, serta studi pada penjualan produk UMKM di Kabupaten Nias yang menggunakan algoritma klasifikasi untuk mengetahui respons konsumen terhadap produk UMKM(Prayugah et al., 2024).

2.6. Text Preprocessing

Tahap text preprocessing merupakan proses awal yang bertujuan untuk menyiapkan data teks agar sesuai dengan kebutuhan sistem yang akan dibangun. Pada tahap ini dilakukan pengurangan kosakata yang tidak relevan, penghilangan *noise* seperti simbol, kata tidak bermakna, dan karakter khusus, serta penyeragaman bentuk teks. Melalui proses tersebut, data teks yang semula tidak terstruktur dapat diubah menjadi data yang lebih terorganisasi, sehingga mampu mempermudah dan mempercepat proses pengolahan serta meningkatkan efektivitas analisis pada tahap selanjutnya(B. A. Maulana & Fahmi, 2024).

Berikut adalah tahapan preprocessing data yaitu(Ariaji et al., 2025):

1. Cleaning

Proses penghapusan elemen yang tidak relevan atau tidak dibutuhkan, sehingga hanya karakter alfabet yang dipertahankan.

2. Remove Duplicates

Tahap penghapusan data teks yang bersifat duplikat agar tidak memengaruhi hasil analisis.

3. Case Folding

Proses mengubah seluruh teks menjadi huruf kecil (*lowercase*) untuk menyeragamkan bentuk kata.

4. Tokenizing

Tahap pemecahan kalimat menjadi unit kata atau token yang akan digunakan dalam analisis lebih lanjut.

5. Filtering

Proses pemilihan kata-kata penting dengan menghapus kata penghubung atau *stopword* yang tidak memiliki makna signifikan.

6. Stemming

Tahap mengubah kata berimbuhan menjadi bentuk kata dasar untuk menyederhanakan variasi kata dalam teks.

7. Labeling

Proses pemberian label sentimen menggunakan metode *Valence Aware Dictionary and sEntiment Reasoner* (VADER), yang mengklasifikasikan sentimen teks ke dalam kategori positif dan negatif, serta dirancang khusus untuk menganalisis teks media sosial yang bersifat informal.

2.7. Pembobotan Kata (*Feature Extraction*)

Proses pembobotan kata (*feature extraction*) merupakan tahapan yang bertujuan untuk mengubah data berbentuk teks menjadi representasi numerik berupa vektor angka agar dapat diproses oleh komputer. Pada tahap ini, setiap kata atau token hasil preprocessing direpresentasikan dalam bentuk nilai numerik berdasarkan tingkat kemunculan dan kepentingannya dalam dokumen. Representasi vektor ini sangat penting karena algoritma *machine learning* seperti

Support Vector Machine (SVM) hanya dapat bekerja pada data numerik(Lestari & Hutagalung, n.d.).

2.7.1. TF-IDF (*Term Frequency–Inverse Document Frequency*)

TF-IDF (*Term Frequency–Inverse Document Frequency*) merupakan metode pembobotan kata yang terdiri dari dua komponen utama, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Masing-masing komponen memiliki rumus dan mekanisme perhitungan yang berbeda. Nilai TF digunakan untuk mengukur frekuensi kemunculan suatu kata dalam sebuah dokumen, sedangkan IDF berfungsi untuk menilai tingkat kepentingan kata tersebut berdasarkan penyebarannya dalam keseluruhan dokumen. Pada tahap akhir, nilai TF dan IDF dikombinasikan untuk menghasilkan bobot TF-IDF, yang merepresentasikan tingkat relevansi suatu kata dalam dokumen tertentu(Ibrahim, n.d.).

Untuk memperoleh nilai TF-IDF digunakan persamaan (1). Namun, sebelum nilai TF-IDF dihitung, terlebih dahulu ditentukan nilai TF (*Term Frequency*) menggunakan persamaan (2) dan nilai IDF (*Inverse Document Frequency*) menggunakan persamaan (3) (Pradana et al., 2025).

$$w_{t,d} = tf_{t,d} \times idf_t \quad (1)$$

$$tf_{t,d} = 1 + \log_{10}(f_{t,d}), \quad \text{jika } f_{t,d} > 0 \quad (2)$$

$$idf_t = \log_{10} \left(1 + \frac{N}{df_t} \right) \quad (3)$$

Keterangan :

$w_{t,d}$: bobot kata (*term*) terhadap dokumen d

$tf_{t,d}$: nilai *Term Frequency* dari *term* pada dokumen d

idf_t : nilai *Inverse Document Frequency* dari *term*

N : jumlah seluruh dokumen

df_t : jumlah dokumen yang mengandung *term*

$f_{t,d}$: frekuensi kemunculan *term* dalam dokumen d

2.8. *Support Vector Machine (SVM)*

Support Vector Machine (SVM) merupakan salah satu metode klasifikasi yang banyak digunakan untuk pengolahan data teks karena mampu menghasilkan tingkat akurasi yang baik. SVM termasuk ke dalam metode *supervised learning* yang bekerja dengan memanfaatkan data berlabel untuk membangun model klasifikasi. Secara matematis, SVM memiliki dasar teori yang kuat dan jelas, sehingga sering digunakan dalam berbagai permasalahan klasifikasi (Svm et al., 2021).

Prinsip kerja SVM adalah menentukan hyperplane optimal yang mampu memisahkan data ke dalam kelas-kelas yang berbeda dengan memaksimalkan jarak antar kelas tersebut. Hyperplane merupakan fungsi pemisah antar kelas, di mana pada ruang dua dimensi fungsi ini berupa garis (*line*), pada ruang tiga dimensi berupa bidang (*plane*), dan pada ruang berdimensi lebih tinggi dikenal sebagai *hyperplane* (Utami, 2025).

Proses klasifikasi dilakukan dengan menggunakan garis pemisah (*hyperplane*) yang berfungsi untuk membedakan antara kelas opini 1+ dan opini 1-. *Hyperplane* yang optimal adalah *hyperplane* yang memiliki jarak maksimum terhadap titik data pelatihan terdekat dari masing-masing kelas. Semakin besar nilai margin yang dihasilkan, maka semakin kecil kemungkinan terjadinya kesalahan generalisasi pada proses klasifikasi. Margin didefinisikan sebagai jarak antara suatu vektor data dalam kelas tertentu terhadap *hyperplane* (Aisah et al., 2023).

Beberapa jenis rumus kernel yang umum digunakan dalam metode Support Vector Machine (SVM) antara lain sebagai berikut:

1. Kernel Linear

Kernel ini digunakan apabila data dapat dipisahkan secara linier. Kernel linear memanfaatkan vektor fitur dari dua titik data untuk membentuk fungsi pemisah yang sederhana.

$$K(x, x') = x \cdot x'$$

Di sini, x dan x' adalah vektor fitur dari dua titik data.

2. Kernel Polinomial

Kernel polinomial diterapkan pada data yang memiliki hubungan berbentuk polinomial, sehingga mampu merepresentasikan pola yang lebih kompleks dibandingkan kernel linear. Kernel ini melibatkan konstanta dan derajat polinomial sebagai parameter pengendali kompleksitas model.

$$K(x, x') = (x \cdot x' + c)^d$$

Di mana c adalah konstanta dan d adalah derajat polinomial.

3. *Kernel Radial Basis Function* (RBF)

Kernel RBF efektif dalam menangani data dengan pola nonlinier yang kompleks. Kernel ini menggunakan parameter tertentu untuk mengatur tingkat pengaruh masing-masing titik data terhadap pembentukan hyperplane.

$$K(x, x') = \exp\left(-\frac{|x - x'|^2}{2\sigma^2}\right)$$

Di mana σ adalah parameter yang mengontrol seberapa besar pengaruh data titik tersebut.

Penggunaan kernel bertujuan untuk mentransformasikan ruang fitur data ke dimensi yang lebih tinggi, sehingga permasalahan klasifikasi dapat diselesaikan secara lebih optimal, khususnya pada data yang tidak dapat dipisahkan secara linier (Svm et al., 2021).

2.9. Evaluasi Model *Support Vector Machine* (SVM)

Pada tahap evaluasi, dilakukan penilaian terhadap kinerja model *Support Vector Machine* (SVM) yang telah dibangun. Evaluasi kinerja model dilakukan dengan menggunakan confusion matrix untuk menghitung nilai accuracy, precision, recall, dan F1-score. Hasil evaluasi ini digunakan sebagai dasar dalam menilai tingkat ketepatan, sensitivitas, serta efektivitas model dalam melakukan proses klasifikasi. Selain itu, tahap evaluasi juga bertujuan untuk mengukur kemampuan model dalam menangani kesalahan serta mengidentifikasi faktor-faktor yang menyebabkan terjadinya kesalahan klasifikasi. Pengukuran performa model dilakukan menggunakan beberapa metrik klasifikasi dengan rumus perhitungan sebagaimana ditunjukkan pada persamaan berikut (Dewi et al., 2025): Adapun persamaan dari akurasi dapat dilihat pada rumus persamaan berikut:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Adapun persamaan dari presisi dapat dilihat pada rumus persamaan berikut :

$$\text{Precision} = \frac{TP}{TP + FP}$$

Adapun persamaan dari recall dapat dilihat pada rumus persamaan berikut:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Adapun persamaan dari f1-score dapat dilihat pada rumus persamaan berikut:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

2.10. Penelitian Terdahulu

Tabel 2.1 Tabel Penelitian Terdahulu

No	Peneliti	Judul	Thn	Metode	Hasil
1	Firdaus et al.	Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Menggunakan Algoritma Support Vector Machine: Studi Kasus Depot Bamara	2024	Data ulasan pelanggan Depot Bamara dari Google Review tahun 2021–2022. Metode yang digunakan meliputi web scraping, pelabelan sentimen positif dan negatif, pelabelan aspek makanan, pelayanan, harga, dan suasana, preprocessing, pembobotan TF-IDF, klasifikasi SVM, K-Fold Cross Validation, serta Root Cause Analysis.	Model SVM menghasilkan rata-rata accuracy 92%, precision 94%, recall 95%, dan f-measure 95%. Aspek harga memiliki accuracy tertinggi sebesar 96%. Hasil RCA menunjukkan masalah utama pada opsi kepedasan, SOP pelayanan, pelatihan staf, penataan tempat duduk, dan parkir.

2	Fadhilah et al.	Support Vector Machine (SVM) untuk Analisis Sentimen Ulasan Pelanggan dalam Penyusunan SWOT Restoran Kebuli Tarim	2025	Data ulasan pelanggan dari Google Review dan TikTok. Setelah proses pembersihan, data valid yang digunakan sebanyak 1.451 ulasan. Metode yang digunakan adalah Aspect-Based Sentiment Analysis, DINESERV, kamus aspek, KeyBERT, InSet Lexicon, TF-IDF, SVM kernel linear, GridSearchCV, evaluasi model, dan analisis SWOT.	Model terbaik menggunakan SVM kernel linear dengan parameter $C = 2$ pada rasio data 90:10. Hasil evaluasi menunjukkan accuracy 89,73%, balanced accuracy 86,66%, dan AUC 0,96. Analisis SWOT menemukan kekuatan pada cita rasa, pelayanan ramah, kenyamanan, dan nilai harga; sedangkan kelemahan terdapat pada inkonsistensi rasa, pelayanan, kenyamanan ruang makan, dan persepsi nilai.
3	Putri et al.	Analisis Kualitas Layanan Fasilitas Malang Smart Arena melalui Analisis Sentimen Menggunakan Support Vector Machine	2025	Data berupa ulasan pengunjung Malang Smart Arena dari Google Review. Metode yang digunakan meliputi web scraping, text preprocessing, topic modeling dengan LDA, pelabelan sentimen menggunakan InSet Lexicon, pembobotan TF-IDF, klasifikasi	Model SVM menghasilkan accuracy 95,39%, macro avg precision 93%, recall 72%, dan F1-score 79%. Aspek tangibles memperoleh skor service quality terendah sehingga menjadi prioritas perbaikan, terutama pada fasilitas bermain, kebersihan, dan kondisi fisik layanan.

				SVM, service quality score, confusion matrix, dan Root Cause Analysis.	
4	Khotimah	Analisis Sentimen terhadap Kualitas Pelayanan: Tinjauan Literatur	2024	Data berupa 200 artikel dari Google Scholar melalui Publish or Perish, kemudian diseleksi menjadi 22 artikel sesuai kriteria tahun 2020–2024 dan topik analisis sentimen terhadap pelayanan. Metode yang digunakan adalah Systematic Literature Review dengan analisis approach, tools, dan platform.	Hasil kajian menunjukkan bahwa platform yang paling banyak digunakan dalam penelitian analisis sentimen pelayanan adalah Twitter, sedangkan algoritma yang paling banyak digunakan adalah Naïve Bayes Classifier. SVM juga ditemukan sebagai salah satu algoritma yang digunakan dalam beberapa penelitian terkait pelayanan.

5	Fardeen et al.	Perbandingan Analisis Sentimen untuk Prediksi Kepuasan Pelanggan Kedai Kopi di Kofind Menggunakan Algoritma SVM dan Naive Bayes	2025	Data berupa 486 ulasan pelanggan Kedai Kopi Kofind dari Tokopedia periode Februari–Juni 2023. Metode yang digunakan meliputi preprocessing, lowercasing, data cleaning, tokenisasi, stopword removal, pelabelan positif, netral, dan negatif, TF-IDF, SMOTE, SVM, Naïve Bayes, confusion matrix, dan cross validation.	SVM memperoleh akurasi 99,19%, sedangkan Naïve Bayes memperoleh akurasi 89,11%. Hasil ini menunjukkan bahwa SVM lebih unggul dalam mengklasifikasikan sentimen pelanggan kedai kopi dibandingkan Naïve Bayes.
6	Penelitian ini	Analisis Sentimen untuk Evaluasi Kualitas Pelayanan Kedai Kopi SAA Menggunakan Algoritma Support Vector Machine (SVM)	2026	Data penelitian berupa 50 ulasan pelanggan Kedai Kopi SAA yang diklasifikasikan ke dalam sentimen positif, netral, dan negatif serta kualitas pelayanan baik dan kurang baik. Metode yang digunakan adalah Knowledge Discovery in	Penelitian ini diarahkan untuk mengetahui proses penerapan SVM dalam mengklasifikasikan sentimen ulasan pelanggan dan mengevaluasi kualitas pelayanan Kedai Kopi SAA. Dalam rancangan proposal, hasil evaluasi model ditampilkan sangat tinggi, tetapi nilai akhir tetap perlu divalidasi berdasarkan data testing final agar

				Database, data selection, preprocessing, TF-IDF, klasifikasi Support Vector Machine, dan evaluasi model menggunakan accuracy, precision, recall, F1-score, serta confusion matrix.	hasil penelitian lebih kuat secara akademik.
--	--	--	--	---	---