



Penerapan Algoritma Naïve Bayes Dalam Analisis Sentimen Masyarakat Terhadap STMIK Widya Cipta Dharma

Helmelya Putri Jelita*, Muhammad Ibnu Saad, Wahyuni

Teknik Informatika, STMIK Widya Cipta Dharma, Samarinda, Indonesia

Email: 2143018@wicida.ac.id, saad@wicida.ac.id, wahyuni@wicida.ac.id

Email Korespondensi penulis: 2143018@wicida.ac.id

Abstrak- Penelitian ini menerapkan algoritma Naïve Bayes untuk menganalisis sentimen masyarakat terhadap STMIK Widya Cipta Dharma dengan menggunakan ulasan Google Maps sebagai sumber data utama. Penelitian ini bertujuan untuk mengklasifikasikan persepsi masyarakat ke dalam tiga kategori: positif, netral, dan negatif. Metodologi yang digunakan mengikuti kerangka kerja CRISP-DM, yang menggabungkan tahapan-tahapan seperti preprocessing data (pembersihan teks, penghilangan stopword, dan stemming), TF-IDF untuk ekstraksi fitur, dan SMOTE untuk mengatasi ketidakseimbangan kelas. Label sentimen diperoleh dari kombinasi peringkat ulasan (1-5 bintang) dan konten tekstual. Hasilnya menunjukkan bahwa Naïve Bayes mencapai akurasi 91% dalam mengklasifikasikan kelas mayoritas (positif), tetapi kesulitan dengan kelas minoritas (netral dan negatif), menghasilkan presisi dan recall 0% untuk kategori-kategori ini. Setelah menerapkan SMOTE, recall untuk kelas negatif meningkat menjadi 100%, meskipun akurasi keseluruhan turun menjadi 38%, yang mencerminkan pertukaran antara pengenalan kelas yang seimbang dan kinerja model. Studi ini menyoroti efektivitas algoritme dalam menangani data teks berskala besar, tetapi menggarisbawahi tantangan dalam mengelola set data yang tidak seimbang. Temuan ini memberikan wawasan yang dapat ditindaklanjuti bagi STMIK Widya Cipta Dharma untuk meningkatkan kualitas layanan dan citra institusi dengan memanfaatkan umpan balik dari masyarakat. Penelitian di masa depan dapat mengeksplorasi algoritma hibrida atau teknik preprocessing tingkat lanjut untuk mengoptimalkan akurasi analisis sentimen di semua kelas.

Kata Kunci: Analisis sentimen, Naïve Bayes, Google Maps, ulasan, Rating, SMOTE.

Abstract- This study applies the Naïve Bayes algorithm to analyze public sentiment toward STMIK Widya Cipta Dharma using Google Maps reviews as the primary data source. The research aims to classify community perceptions into three categories: positive, neutral, and negative. The methodology follows the CRISP-DM framework, incorporating stages such as data preprocessing (text cleaning, stopword removal, and stemming), TF-IDF for feature extraction, and SMOTE to address class imbalance. Sentiment labels were derived from a combination of review ratings (1–5 stars) and textual content. Results indicate that Naïve Bayes achieved 91% accuracy in classifying the majority (positive) class but struggled with minority classes (neutral and negative), yielding 0% precision and recall for these categories. After applying SMOTE, recall for the negative class improved to 100%, although overall accuracy dropped to 38%, reflecting a trade-off between balanced class recognition and model performance. The study highlights the algorithm's effectiveness in handling large-scale text data but underscores challenges in managing imbalanced datasets. These findings provide actionable insights for STMIK Widya Cipta Dharma to enhance service quality and institutional image by leveraging public feedback. Future research could explore hybrid algorithms or advanced preprocessing techniques to optimize sentiment analysis accuracy across all classes.

Keywords: Sentiment analysis, Naïve Bayes, Google Maps, reviews, rating, SMOTE.

1. PENDAHULUAN

Di era digital saat ini, media sosial dan platform daring menjadi tempat utama bagi masyarakat untuk menyampaikan opini, pengalaman, dan persepsi mereka terhadap berbagai institusi, termasuk lembaga pendidikan tinggi. Salah satu platform yang banyak digunakan masyarakat untuk memberikan rating dan ulasan adalah Google Maps. Melalui fitur ulasan pada Google Maps, masyarakat dapat memberikan komentar dan rating terhadap kualitas layanan suatu institusi secara terbuka dan real-time. STMIK Widya Cipta Dharma, sebagai salah satu institusi pendidikan tinggi di Kalimantan Timur, juga menerima banyak ulasan dari masyarakat melalui platform ini.

Analisis terhadap data ulasan tersebut sangat penting untuk memahami sentimen masyarakat, baik yang bersifat positif, netral, maupun negatif. Pemanfaatan teknik analisis sentimen menjadi salah satu cara efektif untuk mengeksperisikan diri di era digital saat ini, media sosial dan platform daring menjadi tempat utama bagi masyarakat untuk menyampaikan opini, pengalaman, dan persepsi mereka terhadap berbagai institusi, termasuk lembaga pendidikan tinggi. Salah satu platform yang banyak digunakan masyarakat untuk memberikan penilaian dan ulasan adalah Google Maps. Melalui fitur ulasan pada Google Maps, masyarakat dapat memberikan komentar dan rating terhadap kualitas layanan suatu institusi secara terbuka dan real-time, [1] Google Maps merupakan salah satu alamat website yang dapat menulis ulasan untuk tempat yang pernah dikunjungi.” Ulasan tersebut dapat dianalisis secara otomatis menggunakan algoritma Naive Bayes





untuk mengetahui sentimen masyarakat, apakah bernada positif atau negatif. [2] Analisis sentimen dilakukan untuk mengukur dan memahami opini publik terhadap STMIK Widya Cipta Dharma yang diharapkan dapat menjadi evaluasi instansi tersebut dan pendukung keputusan strategis yang lebih baik. STMIK Widya Cipta Dharma, sebagai salah satu institusi pendidikan tinggi di Kalimantan Timur, juga menerima banyak ulasan dari masyarakat melalui platform ini. Ulasan yang diberikan oleh masyarakat pada platform digital seperti Google Maps mencerminkan pandangan dan pengalaman mereka terhadap layanan yang diterima. Data ini bersifat terbuka dan dinamis, sehingga dapat menjadi sumber informasi yang kaya untuk dieksplorasi. Namun, mengingat jumlah ulasan yang terus bertambah, proses analisis secara manual menjadi tidak efisien. Oleh karena itu, diperlukan metode otomatis yang mampu mengolah data teks dalam skala besar untuk mengidentifikasi pola dan tren opini publik.

Salah satu algoritma yang sering digunakan dalam analisis sentimen adalah Naïve Bayes. [3] Agar analisis terhadap sentimen yang terdapat dalam komentar mahasiswa dapat dilakukan dengan lebih baik, maka digunakan algoritma Naïve Bayes untuk membantu dalam mengidentifikasi atau mengklasifikasikan sentimen mahasiswa. Algoritma ini dikenal memiliki performa baik dalam klasifikasi teks, terutama dalam kasus-kasus dengan data besar dan bersifat tidak terstruktur, [4] Naïve Bayes merupakan metode klasifikasi yang sederhana dan cepat, serta mampu menghasilkan akurasi tinggi pada data berjumlah besar, Naïve Bayes bekerja berdasarkan prinsip probabilitas dengan mengasumsikan bahwa setiap fitur atau kata dalam dokumen bersifat independen. Meskipun asumsi ini terkesan sederhana, dalam praktiknya algoritma ini terbukti efektif dalam berbagai studi klasifikasi teks dan analisis sentimen.

Penelitian sebelumnya telah menunjukkan efektivitas Naïve Bayes dalam menganalisis sentimen pengguna terhadap berbagai layanan. [5] Menerapkannya pada data Twitter BMKG dan mencapai akurasi 69,97%. [6] Menggunakannya untuk analisis produk dan layanan perusahaan melalui Twitter dengan akurasi 89%. [7] Mengkombinasikan metode ini dengan pendekatan berbasis leksikon pada layanan UB Press dan mencapai akurasi 87,83%.

Kelebihan lain dari algoritma Naïve Bayes adalah kecepatannya dan efisiensinya dalam proses pelatihan dan prediksi, menjadikannya cocok untuk diterapkan pada data ulasan dalam jumlah besar seperti di Google Maps. Selain itu, metode ini dapat diintegrasikan dengan teknik pra-pemrosesan teks seperti stemming, stopword removal, dan tokenisasi untuk meningkatkan akurasi klasifikasi.

Dengan mempertimbangkan kompleksitas bahasa dan konteks pada ulasan masyarakat umum, pendekatan analisis harus mampu menangani variasi bahasa, termasuk bahasa Indonesia informal dan campuran bahasa asing. Oleh karena itu, dalam penerapan algoritma Naïve Bayes pada penelitian ini, dilakukan pula tahapan preprocessing data yang mencakup normalisasi teks, stemming menggunakan algoritma bahasa Indonesia, dan penghapusan stopwords.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma Naïve Bayes dalam menganalisis sentimen masyarakat terhadap STMIK Widya Cipta Dharma, dengan data diambil dari ulasan Google Maps. Analisis ini diharapkan dapat memberikan wawasan berharga bagi institusi dalam meningkatkan mutu layanan, citra institusi, serta pemahaman yang lebih dalam terhadap kebutuhan dan harapan masyarakat. [8] Analisis sentimen merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen dalam opini. [9] Menyatakan bahwa teknik analisis asosiasi juga digunakan dalam data mining untuk mendukung strategi bisnis.

Web scraping merupakan proses pengumpulan data secara otomatis dari halaman web. [10] Menjelaskan bahwa teknik ini umum digunakan dalam studi berbasis media sosial. [11] Menyatakan bahwa web scraping dapat mengumpulkan teks, gambar, dan informasi dari berbagai situs, serta meningkatkan efisiensi operasional, [12] Penggunaan program web scraping terbukti menghasilkan data yang jauh lebih lengkap, akurat, dan konsisten dibandingkan pengumpulan data secara manual.

Naïve Bayes sendiri menggunakan teori probabilitas. [13] Menyatakan bahwa metode ini melihat frekuensi kemunculan setiap klasifikasi pada data pelatihan. [14] Menambahkan bahwa teorema Bayes digunakan untuk memprediksi fakta berdasarkan kejadian sebelumnya. [15] Ketidakseimbangan data pada klasifikasi sentimen dapat menyebabkan bias terhadap kelas mayoritas dan mengabaikan kelas minoritas, yang berdampak buruk terhadap akurasi prediksi model. Untuk menangani data tidak seimbang, digunakan metode SMOTE (Synthetic Minority Oversampling Technique). Ketidakseimbangan data pada klasifikasi sentimen dapat menyebabkan bias terhadap kelas mayoritas dan mengabaikan kelas minoritas, yang berdampak buruk terhadap akurasi prediksi model [16] Menjelaskan bahwa SMOTE membuat data sintesis antara titik-titik minoritas. Dalam penelitian ini, SMOTE diterapkan untuk menyetarakan distribusi sentimen positif, netral, dan negatif, yang terbukti meningkatkan kinerja model klasifikasi.

Evaluasi kinerja model dilakukan menggunakan confusion matrix. [17] confusion matrix mengukur akurasi klasifikasi melalui komponen TP, TN, FP, dan FN. Sebagai kerangka analisis, digunakan metode CRISP-DM yang terdiri dari enam tahapan, seperti dijelaskan oleh [18] dan [19], [20] CRISP-DM tidak hanya terbatas pada proyek data mining tradisional, tetapi juga sangat sesuai diterapkan pada proyek analisis dan data sains modern selain itu, digunakan metode TF-IDF (Term Frequency-Inverse Document Frequency) sebagai teknik representasi teks. [21] Term Frequency (tf(w,d)) dianggap memiliki proporsi kepentingan sesuai total kemunculannya dalam teks atau dokumen. Inverse Document Frequency (IDF) merupakan metode pembobotan token yang berfungsi untuk memonitor kemunculan token dalam



himpunan teks. [22] Representasi teks dalam penelitian ini menggunakan metode TF-IDF karena mampu mengurangi pengaruh kata-kata umum dan memberikan bobot lebih tinggi pada kata-kata yang penting dan unik dalam korpus.

2. METODOLOGI PENELITIAN

2.1 Cross Industry Standard Process for Data Mining (CRISP-DM)

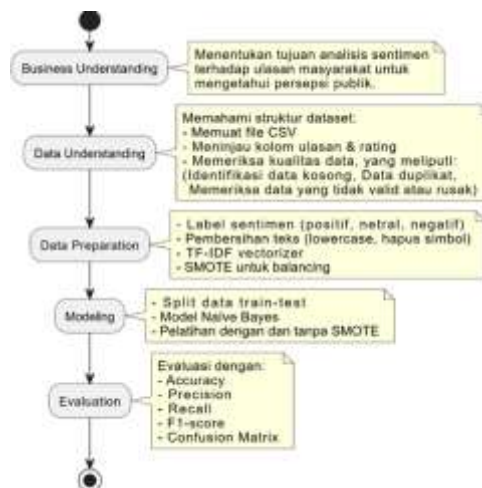
Gambar 1 menggambarkan alur penelitian yang mengikuti tahapan metode CRISP-DM (Cross Industry Standard Process for Data Mining), yang terdiri dari enam tahap utama, yaitu: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Dalam konteks penelitian ini, proses dilakukan secara menyeluruh hingga tahap Evaluation, yaitu tahap evaluasi performa model klasifikasi sentimen. Namun, penelitian ini tidak melanjutkan ke tahap Deployment, yaitu proses penerapan model ke dalam sistem atau website. Hal ini disebabkan fokus utama penelitian adalah pada pengujian akurasi dan efektivitas model Naïve Bayes dalam analisis sentimen.



Gambar 1: Metode CRISP-DM

2.2 Alur Penelitian

Gambar 1 menunjukkan alur penelitian yang mengacu pada metode CRISP-DM, yang terdiri dari beberapa tahapan mulai dari pemahaman bisnis hingga evaluasi model. Dalam penelitian ini, proses dilaksanakan hingga tahap evaluasi saja, tanpa dilanjutkan ke tahap deployment atau penerapan hasil model ke dalam Website untuk dipergunakan.



Gambar 2. Alur penelitian

2.2.1 Business Understanding

Tahapan pertama adalah memahami tujuan bisnis, yaitu melakukan analisis sentimen terhadap ulasan masyarakat untuk mengetahui persepsi publik. Penelitian ini bertujuan untuk mengetahui bagaimana tanggapan atau opini

masyarakat, khususnya yang disampaikan melalui media sosial, terhadap suatu topik tertentu. Pemahaman ini penting agar proses selanjutnya dapat terarah dengan jelas sesuai tujuan akhir yang ingin dicapai.

2.2.2 Data Understanding

Pada tahap ini, peneliti mulai memahami struktur dari dataset yang akan digunakan. Proses ini meliputi :

1. Memuat data dalam format CSV ke dalam program,
2. Meninjau isi kolom yang tersedia, terutama kolom ulasan (teks) dan rating, yang menjadi komponen utama dalam analisis sentimen.
3. Memeriksa kualitas data, yang meliputi:
 - a. Identifikasi data kosong, yaitu memeriksa nilai yang hilang pada kolom-kolom penting untuk menghindari gangguan dalam proses analisis.
 - b. Deteksi data duplikat, untuk memastikan tidak ada entri yang berulang yang dapat mempengaruhi hasil pelatihan model.
 - c. Memeriksa data yang tidak valid atau rusak, seperti baris data dengan format yang salah, karakter yang tidak dikenali, atau nilai rating yang berada di luar batas yang diharapkan.

Tahap ini bertujuan untuk memastikan bahwa data yang akan digunakan memiliki kualitas yang memadai dan sesuai dengan kebutuhan klasifikasi sentimen. Dengan memahami dan memeriksa data secara menyeluruh, potensi kesalahan dalam proses analisis lebih lanjut dapat diminimalisir.

2.2.3 Data Preperation

Setelah memahami data, tahap berikutnya adalah mempersiapkan data agar siap digunakan dalam pemodelan. Proses ini meliputi :

1. Pemberian label sentimen: yaitu mengelompokkan data menjadi tiga kelas positif, netral, dan negatif.
2. Pembersihan teks: seperti mengubah huruf menjadi lowercase, menghapus simbol, dan karakter tidak penting lainnya.
3. Ekstraksi fitur dengan TF-IDF: metode ini digunakan untuk mengubah teks menjadi bentuk numerik yang dapat diproses oleh algoritma, dari confusion matrix dapat diketahui metrik atau pengukuran tolak ukur kinerja model klasifikasi dengan menggunakan persamaan sebagai berikut:
 - a. Precision Precision adalah rasio perbandingan jumlah data prediksi benar positif dengan keseluruhan data prediksi positif atau dapat dituliskan pada persamaan berikut :

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

- b. Recall Recall adalah rasio perbandingan jumlah data prediksi benar positif dengan jumlah data benar positif dan data salah negatif atau dapat dituliskan pada persamaan berikut :

$$\text{Precision} = \frac{TP}{TP + FN} \quad (2)$$

- c. F1-Score F1-Score adalah parameter ukuran keberhasilan retrieval yang menggabungkan precision dan recall. Perhitungan metrik F1-Score melibatkan informasi FP dan FN sehingga membuat metrik ini cocok digunakan untuk kasus imbalanced data. Nilai dari F1-Score diperoleh dari persamaan berikut.

$$\text{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

- d. Accuracy Accuracy adalah rasio prediksi benar dengan keseluruhan data. Perhitungan accuracy diperoleh dari persamaan berikut.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

4. SMOTE (Synthetic Minority Oversampling Technique): digunakan untuk mengatasi ketidakseimbangan kelas dalam data, sehingga model bisa belajar secara adil dari semua kelas sentimen.

2.2.4 Modeling

Tahapan modeling dilakukan setelah data siap. Langkah-langkah utamanya :

1. Membagi dataset menjadi data latih (train) dan data uji (test)
2. Membangun model klasifikasi menggunakan algoritma Naïve Bayes dengan persamaan sebagai berikut :

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (5)$$

Keterangan:

$P(H|X)$ = probabilitas hipotesis H pada kondisi X

$P(H)$ = probabilitas hipotesis H

$P(X|H)$ = probabilitas X pada kondisi H

$P(X)$ = probabilitas X.

Pada tahap ini dilakukan pembangunan model klasifikasi untuk menganalisis sentimen dari data ulasan masyarakat terhadap STMIK Widya Cipta Dharma. Algoritma yang digunakan dalam penelitian ini adalah Naïve Bayes, yaitu salah satu algoritma klasifikasi berbasis probabilistik yang umum digunakan untuk analisis teks. Naïve Bayes bekerja berdasarkan Teorema Bayes dengan asumsi bahwa fitur-fitur (kata) bersifat saling independen. Meskipun asumsi ini jarang sepenuhnya terpenuhi dalam data teks, algoritma ini telah terbukti efektif dalam banyak studi klasifikasi dokumen dan analisis sentimen karena kemudahannya, efisiensinya dalam pelatihan, serta kemampuannya mengatasi data dalam jumlah besar.

3. Melakukan pelatihan model baik dengan data asli maupun data yang sudah di-balancing menggunakan SMOTE, untuk melihat perbandingan hasil dan performa dari kedua pendekatan tersebut.

Model dibangun dengan dua pendekatan :

- a. Tanpa SMOTE, yaitu menggunakan data asli yang telah melalui proses preprocessing dan vectorization (TF-IDF).
- b. Dengan SMOTE, yaitu data yang telah di-balance untuk mengatasi ketidakseimbangan kelas (positif, netral, negatif).

Kedua model tersebut kemudian dibandingkan kinerjanya untuk menilai dampak penggunaan SMOTE terhadap performa klasifikasi. Melakukan pelatihan model baik dengan data asli maupun data yang sudah di-balancing menggunakan SMOTE, untuk melihat perbandingan hasil dan performa dari kedua pendekatan tersebut.

2.2.5 Evaluation

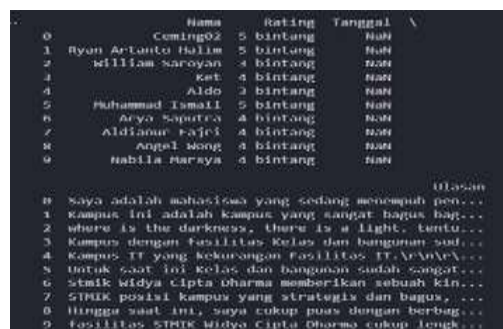
Tahap terakhir adalah evaluasi, yang bertujuan untuk mengukur seberapa baik model bekerja. Evaluasi dilakukan dengan menggunakan matrix - matrix berikut:

1. Accuracy : mengukur tingkat ketepatan klasifikasi.
2. Precision : sejauh mana prediksi positif benar-benar relevan.
3. Recall : seberapa banyak data positif yang berhasil dikenali.
4. F1-score : kombinasi dari precision dan recall untuk keseimbangan performa.
5. Confusion Matrix : memberikan gambaran detail jumlah prediksi benar dan salah pada setiap kelas sentimen.

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Proses pengambilan data menggunakan webscarping melalui platform ulasan dan rating yang diberikan langsung oleh masyarakat dalam google maps, ulasan ini menghasilkan 3 sentiment positif, netral, dan negatif jika dianalisis dengan tepat, dapat memberikan gambaran tentang kepuasan dan harapan masyarakat. Oleh karena itu, diperlukan metode analisis yang mampu mengolah data teks tersebut secara sistematis untuk mengungkap persepsi publik secara objektif. Analisis sentiment ini menggunakan algoritma naive bayes dengan metode CRISP-DM hingga menampilkan hasil akurasi



	Name	Rating	Tanggal
0	Cemilng02	5 bintang	NaN
1	Ryon Arianto Halim	5 bintang	NaN
2	William Sarayan	4 bintang	NaN
3	Ket	4 bintang	NaN
4	Aldo	3 bintang	NaN
5	Muhammad Ismail	5 bintang	NaN
6	Arya Smitra	4 bintang	NaN
7	Aldianur Fajri	4 bintang	NaN
8	Angel Wong	4 bintang	NaN
9	Nabila Nuraya	4 bintang	NaN

Ulasan

```
0 Saya adalah mahasiswa yang sedang mencoba pen...
1 Kampus ini adalah kampus yang sangat bagus bag...
2 where is the darkness, there is a light. tentu...
3 Kampus dengan fasilitas Kelas dan bangunan sud...
4 Kampus IT yang kekurangan Fasilitas IT. AAAAA...
5 Untuk saat ini kelas dan bangunan sudah sangat...
6 STMIK Widya Cipta Dharma memberikan sebuah kin...
7 STMIK postal kampus yang strategis dan bagus. ...
8 Hingga saat ini saya cukup puas dengan berbang...
9 Fasilitas STMIK Widya Cipta Dharma cukup legh...
```

Gambar 3. 10 data teratas

3.2 Data Understanding

Berdasarkan hasil eksplorasi awal, dataset ini terdiri dari 275 entri dengan 4 kolom yaitu Nama, Rating, Tanggal, dan Ulasan. Kolom Nama dan Rating memiliki data lengkap tanpa nilai kosong, sementara kolom Ulasan hanya memiliki 164 data non-null, menunjukkan adanya 111 entri yang tidak memuat isi ulasan. Kolom Tanggal sepenuhnya kosong, sehingga tidak dapat digunakan dalam analisis. Secara tipe data, ketiga kolom (Nama, Rating, dan Ulasan) bertipe objek (teks), sedangkan kolom Tanggal terdeteksi sebagai float64 akibat kosong seluruhnya. Gambar ini menunjukkan bahwa fokus analisis hanya akan efektif dilakukan pada kolom Rating dan Ulasan, sementara kolom Tanggal sebaiknya diabaikan atau dihapus karena tidak memberikan informasi berarti.

```
Informasi Dataset:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 275 entries, 0 to 274
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Nama        275 non-null   object
1   Rating      275 non-null   object
2   Tanggal     0 non-null     float64
3   Ulasan      164 non-null   object
dtypes: float64(1), object(3)
memory usage: 8.7+ KB
```

Gambar 4. Informasi Dataset

3.3 Data Preperation

Dari hasil output tersebut terlihat bahwa proses persiapan data telah dilakukan dengan mengelompokkan nilai rating ke dalam kategori sentimen. Rating yang awalnya ditulis dalam format teks seperti "5 bintang" diubah menjadi label sentimen seperti positif atau netral, lalu disimpan dalam kolom baru bernama Sentimen. Misalnya, rating 4 dan 5 bintang dianggap sebagai ulasan positif, sementara rating 3 bintang dikategorikan sebagai netral. Meski belum terlihat ulasan dengan rating rendah, kemungkinan rating 1 atau 2 akan masuk ke kategori negatif di bagian data lainnya. Langkah ini penting karena sentimen yang sudah dikelompokkan nantinya akan digunakan sebagai acuan atau target dalam proses pelatihan model klasifikasi. Dengan begitu, data sudah berada dalam kondisi yang siap untuk tahap analisis lebih lanjut seperti pembersihan teks dan pelatihan algoritma.

```
...      Rating Sentimen
0   5 bintang  positif
1   5 bintang  positif
2   3 bintang  netral
3   4 bintang  positif
4   3 bintang  netral
5   5 bintang  positif
6   4 bintang  positif
7   4 bintang  positif
8   4 bintang  positif
9   4 bintang  positif
```

Gambar 5. Labeling berdasarkan rating

3.3.1 Pembersihan Teks

Bagian ini merupakan proses awal dalam pembersihan data teks, yang bertujuan untuk menghilangkan karakter-karakter yang tidak diperlukan sebelum dilakukan analisis lebih lanjut seperti analisis sentimen atau klasifikasi teks. Fungsi `clean_text()` yang didefinisikan dalam kode ini berfungsi untuk membersihkan input teks dengan terlebih dahulu memastikan bahwa input berupa string. Kemudian, seluruh huruf dalam teks diubah menjadi huruf kecil agar tidak terjadi perbedaan makna antara huruf kapital dan huruf kecil. Setelah itu, menggunakan modul `re` (regular expression), fungsi ini menghapus semua karakter yang bukan huruf alfabet atau spasi, termasuk angka, simbol, dan tanda baca. Hasil akhir

dari fungsi ini adalah teks yang terdiri hanya dari huruf kecil dan spasi, sehingga lebih konsisten dan siap digunakan dalam proses analisis berikutnya.

```
import re

# Fungsi membersihkan teks (tanpa stopwords removal)
def clean_text(text):
    # Pastikan 'text' adalah string
    if not isinstance(text, str):
        text = str(text)
    # Menghapus karakter non-alfabet dan mengubah ke huruf kecil
    text = text.lower()
    text = re.sub(r'[^\a-zA-Z\s]', '', text)
    return text

# Contoh input
contoh_teks = "Halo! Ini adalah contoh #teks, dengan angka 123 dan simbol @!$."

# Tampilkan hasil pembersihan
print("Teks sebelum dibersihkan:")
print(contoh_teks)

print("\nTeks setelah dibersihkan:")
print(clean_text(contoh_teks))
```

Gambar 7. Proses pembersihan Data

```
Teks sebelum dibersihkan:
Halo! Ini adalah contoh #teks, dengan angka 123 dan simbol @!$.

Teks setelah dibersihkan:
halo ini adalah contoh teks dengan angka dan simbol
```

Gambar 8. Contoh hasil data sebelum dan sesudah dibersihkan

```
Ulasan \
0 Saya adalah mahasiswa yang sedang menempuh pen...
1 Kampus ini adalah kampus yang sangat bagus bag...
2 where is the darkness, there is a light. tentu...
3 kampus dengan fasilitas kelas dan bangunan sud...
4 kampus IT yang kekurangan fasilitas IT.\nkam...
5 Untuk saat ini kelas dan bangunan sudah sangat...
6 stmik widya cipta dharna memberikan sebuah kin...
7 STMIK posisi kampus yang strategis dan bagus, ...
8 hingga saat ini, saya cukup puas dengan berbag...
9 fasilitas STMIK Widya Cipta Dharna cukup lengk...

Ulasan_Bersih
0 saya adalah mahasiswa yang sedang menempuh pen...
1 kampus ini adalah kampus yang sangat bagus bag...
2 where is the darkness there is a light tenta...
3 kampus dengan fasilitas kelas dan bangunan sud...
4 kampus it yang kekurangan fasilitas it.\nkamp...
5 untuk saat ini kelas dan bangunan sudah sangat...
6 stmik widya cipta dharna memberikan sebuah kin...
7 stmik posisi kampus yang strategis dan bagus t...
8 hingga saat ini saya cukup puas dengan berbag...
9 fasilitas stmik widya cipta dharna cukup lengk...
```

Gambar 9. Menampilkan 10 data sebelum dan sesudah di bersihkan

3.3.2 TF-IDF

Proses ini merupakan bagian dari ekstraksi fitur teks menggunakan metode TF-IDF (Term Frequency–Inverse Document Frequency). Dalam konteks ini, teks yang telah dibersihkan dan disimpan di kolom Ulasan_Bersih pada DataFrame df diubah menjadi representasi numerik yang dapat digunakan dalam model machine learning. TfidfVectorizer() dari pustaka scikit-learn digunakan untuk melakukan transformasi ini, di mana setiap kata dalam dokumen diberi bobot berdasarkan seberapa sering kata tersebut muncul dalam satu dokumen (term frequency) dan seberapa jarang muncul di seluruh dokumen (inverse document frequency). Parameter max_features=5000 membatasi jumlah fitur (kata unik) maksimal menjadi 5.000 kata paling penting berdasarkan skor TF-IDF tertinggi. Setelah dilakukan fit_transform(), hasilnya disimpan dalam variabel X_text dalam bentuk matriks sparse, di mana setiap baris mewakili satu dokumen ulasan, dan setiap kolom mewakili bobot TF-IDF dari sebuah kata. Proses ini penting untuk mengubah teks menjadi format numerik agar bisa digunakan oleh algoritma klasifikasi seperti Naïve Bayes.

```
# Ekstraksi fitur menggunakan TF-IDF
vectorizer = TfidfVectorizer(max_features=5000)
X_text = vectorizer.fit_transform(df['Ulasan_Bersih'])

ac acnya ada adalah adem agak agar agya ahli aj \
0 0.0 0.0 0.000000 0.206543 0.000000 0.0 0.000000 0.0 0.0 0.0
1 0.0 0.0 0.134129 0.160471 0.000000 0.0 0.000000 0.0 0.0 0.0
2 0.0 0.0 0.000000 0.000000 0.000000 0.0 0.000000 0.0 0.0 0.0
3 0.0 0.0 0.000000 0.000000 0.000000 0.0 0.215313 0.0 0.0 0.0
4 0.0 0.0 0.000000 0.000000 0.000000 0.0 0.000000 0.0 0.0 0.0
5 0.0 0.0 0.000000 0.000000 0.188192 0.0 0.144177 0.0 0.0 0.0
6 0.0 0.0 0.000000 0.000000 0.000000 0.0 0.000000 0.0 0.0 0.0
7 0.0 0.0 0.247231 0.000000 0.000000 0.0 0.000000 0.0 0.0 0.0
8 0.0 0.0 0.135987 0.000000 0.000000 0.0 0.000000 0.0 0.0 0.0
9 0.0 0.0 0.128964 0.000000 0.000000 0.0 0.000000 0.0 0.0 0.0

... where wicida widya wifi wifisignal wow yaitu yaminmaju \
0 ... 0.000000 0.0 0.198372 0.0 0.0 0.0 0.0 0.0 0.0
1 ... 0.000000 0.0 0.000000 0.0 0.0 0.0 0.0 0.0 0.0
2 ... 0.177553 0.0 0.000000 0.0 0.0 0.0 0.0 0.0 0.0
3 ... 0.000000 0.0 0.000000 0.0 0.0 0.0 0.0 0.0 0.0
4 ... 0.000000 0.0 0.000000 0.0 0.0 0.0 0.0 0.0 0.0
5 ... 0.000000 0.0 0.000000 0.0 0.0 0.0 0.0 0.0 0.0
6 ... 0.000000 0.0 0.153557 0.0 0.0 0.0 0.0 0.0 0.0
7 ... 0.000000 0.0 0.000000 0.0 0.0 0.0 0.0 0.0 0.0
8 ... 0.000000 0.0 0.000000 0.0 0.0 0.0 0.0 0.0 0.0
9 ... 0.000000 0.0 0.148188 0.0 0.0 0.0 0.0 0.0 0.0

... yang yg
8 0.255882 0.000000
9 0.000000 0.000000

[10 rows x 742 columns]
```

Gambar 9. Ekstraksi fitur TF-IDF

3.3.3 SMOTE Untuk Balancing

Proses ini menjelaskan pelatihan model Naive Bayes dengan memanfaatkan teknik SMOTE untuk menangani ketidakseimbangan data. Pertama, data pelatihan yang awalnya berbentuk sparse matrix diubah menjadi array berdimensi penuh menggunakan `toarray()`, karena SMOTE hanya dapat bekerja pada data dalam format dense. Kemudian, SMOTE digunakan untuk menyeimbangkan distribusi label dengan membuat data sintesis sehingga jumlah sampel dari setiap kelas menjadi seimbang. Setelah itu, model Naive Bayes dilatih menggunakan data yang telah diseimbangkan. Model ini kemudian diuji pada data uji, dan hasil prediksinya dievaluasi menggunakan `classification_report` untuk melihat metrik seperti akurasi, presisi, recall, dan F1-score. Akhirnya, confusion matrix divisualisasikan untuk menunjukkan seberapa baik model mengklasifikasikan masing-masing kelas sentimen setelah penerapan SMOTE.

```
X_train_dense = X_train.toarray() # SMOTE butuh dense array
smote = SMOTE(random_state=42)
X_train_res, y_train_res = smote.fit_resample(X_train_dense, y_train)

model_nb_smote = MultinomialNB()
model_nb_smote.fit(X_train_res, y_train_res)
y_pred_nb_smote = model_nb_smote.predict(X_test)

print("Evaluasi Model Naive Bayes dengan SMOTE:")
print(classification_report(y_test, y_pred_nb_smote))

# Confusion Matrix
cm_nb_smote = confusion_matrix(y_test, y_pred_nb_smote, labels=model_nb.classes_)
disp_nb_smote = ConfusionMatrixDisplay(confusion_matrix=cm_nb_smote, display_labels=model_nb.classes_)
disp_nb_smote.plot(cmap='Greens')
plt.title("Confusion Matrix - Dengan SMOTE")
plt.show()
```

Gambar 10. SMOTE Untuk Balancing

3.4 Modeling

3.4.1 Data Train – Test

Proses ini digunakan untuk membagi data ulasan yang sudah diolah menjadi dua bagian, yaitu data untuk melatih model dan data untuk menguji hasilnya. Variabel X berisi fitur teks ulasan yang telah diubah menjadi angka menggunakan metode TF-IDF, sedangkan y berisi label sentimen seperti positif, netral, atau negatif. Fungsi `train_test_split` membagi data tersebut secara acak, di mana 80% data digunakan untuk pelatihan dan 20% sisanya untuk pengujian. Pemakaian `random_state=42` bertujuan agar hasil pembagian datanya tetap sama setiap kali dijalankan, sehingga hasil evaluasi model menjadi konsisten. Proses ini penting agar kita bisa menilai seberapa baik model mampu memprediksi data yang belum pernah dilihat sebelumnya.

```
X = X_text
y = df['Sentimen']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Gambar 11. Data Train - Test

3.4.2 Naive Bayes Tanpa SMOTE

Tahapan ini digunakan untuk melatih dan mengevaluasi model klasifikasi menggunakan algoritma Naive Bayes tanpa penyeimbangan data. Model dilatih dengan data pelatihan, lalu digunakan untuk memprediksi sentimen pada data uji. Hasil prediksi dibandingkan dengan data asli menggunakan `classification_report` untuk melihat nilai akurasi, presisi, recall, dan F1-score.

```
model_nb = MultinomialNB()
model_nb.fit(X_train, y_train)
y_pred_nb = model_nb.predict(X_test)

print("Evaluasi Model Naive Bayes Tanpa SMOTE:")
print(classification_report(y_test, y_pred_nb))

# Confusion Matrix
cm_nb = confusion_matrix(y_test, y_pred_nb, labels=model_nb.classes_)
disp_nb = ConfusionMatrixDisplay(confusion_matrix=cm_nb, display_labels=model_nb.classes_)
disp_nb.plot(cmap='Blues')
plt.title("Confusion Matrix - Tanpa SMOTE")
plt.show()
```

Gambar 12 Naive Bayes Tanpa SMOTE

3.4.3 Naive Bayes Dengan SMOTE

Kode ini digunakan untuk melatih model Naive Bayes dengan data yang telah diseimbangkan menggunakan SMOTE. Karena SMOTE hanya bisa digunakan pada data dalam bentuk array padat, data pelatihan `X_train` terlebih dahulu dikonversi ke bentuk dense menggunakan `toarray()`. Setelah itu, SMOTE diterapkan untuk menyeimbangkan jumlah data di setiap kelas sentimen. Data hasil SMOTE (`X_train_res` dan `y_train_res`) digunakan untuk melatih ulang model `MultinomialNB`.

```
X_train_dense = X_train.toarray() # SMOTE butuh dense array
smote = SMOTE(random_state=42)
X_train_res, y_train_res = smote.fit_resample(X_train_dense, y_train)

model_nb_smote = MultinomialNB()
model_nb_smote.fit(X_train_res, y_train_res)
y_pred_nb_smote = model_nb_smote.predict(X_test)

print("Evaluasi Model Naive Bayes dengan SMOTE:")
print(classification_report(y_test, y_pred_nb_smote))

# Confusion Matrix
cm_nb_smote = confusion_matrix(y_test, y_pred_nb_smote, labels=model_nb.classes_)
disp_nb_smote = ConfusionMatrixDisplay(confusion_matrix=cm_nb_smote, display_labels=model_nb.classes_)
disp_nb_smote.plot(cmap='Greens')
plt.title("Confusion Matrix - Dengan SMOTE")
plt.show()
```

Gambar 13 Naive Bayes Dengan SMOTE

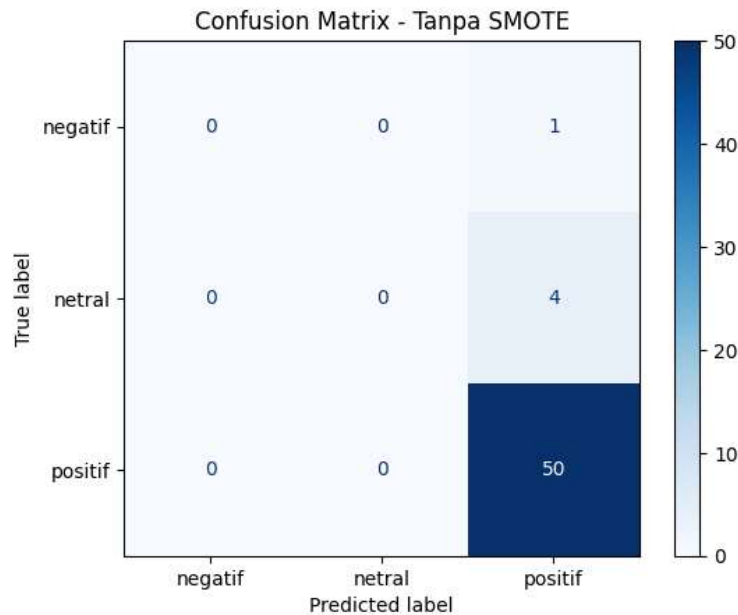
3.5 Evaluation

3.5.1 Confusion Matrix Naive Bayes Tanpa SMOTE

Hasil evaluasi model Naive Bayes tanpa penerapan SMOTE menunjukkan bahwa model terlalu fokus pada kelas "positif". Hal ini terlihat dari nilai precision dan recall untuk kelas "positif" yang sangat tinggi, sementara dua kelas lainnya, yaitu "negatif" dan "netral", sama sekali tidak terdeteksi oleh model (semuanya bernilai nol). Situasi ini mengindikasikan adanya ketidakseimbangan jumlah data antar kelas, di mana kelas "positif" mendominasi dataset. Meskipun akurasi terlihat tinggi, yaitu 91%, hasil tersebut sebenarnya menyesatkan karena model hanya berhasil mengklasifikasikan satu jenis kelas. Oleh karena itu, diperlukan pendekatan seperti SMOTE untuk menyeimbangkan distribusi data dan meningkatkan kemampuan model dalam mengenali seluruh kelas secara adil.

Evaluasi Model Naive Bayes Tanpa SMOTE:				
	precision	recall	f1-score	support
negatif	0.00	0.00	0.00	1
netral	0.00	0.00	0.00	4
positif	0.91	1.00	0.95	50
accuracy			0.91	55
macro avg	0.30	0.33	0.32	55
weighted avg	0.83	0.91	0.87	55

Gambar 15 Evaluasi Model Naive Bayes Tanpa SMOTE



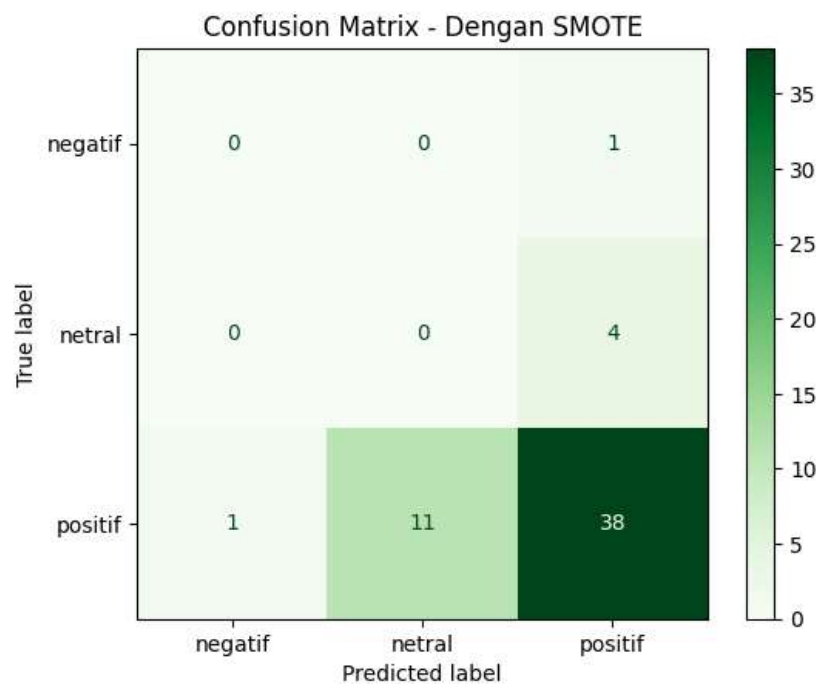
Gambar 16 Hasil Visualisasi Naive Bayes Tanpa SMOTE

3.5.2 Confusion Matrix Naive Bayes Dengan SMOTE

Setelah diterapkannya metode SMOTE, performa model Naive Bayes mengalami penurunan akurasi dari sebelumnya 91% menjadi 69%. Meskipun demikian, model mulai menunjukkan sedikit kemampuan untuk menangani data yang lebih seimbang. Kelas "positif" masih mendominasi prediksi dengan precision sebesar 0.88 dan recall 0.76, sementara kelas "negatif" dan "netral" belum berhasil dikenali sama sekali (precision dan recall bernilai nol). Hal ini mengindikasikan bahwa penerapan SMOTE belum cukup efektif untuk mengatasi ketimpangan distribusi kelas. Nilai macro average yang masih rendah menandakan bahwa model belum dapat memprediksi semua kelas secara seimbang. Dengan demikian, meskipun ada sedikit peningkatan dalam keragaman prediksi, model masih memerlukan pendekatan tambahan seperti penambahan data atau penyempurnaan parameter agar performanya menjadi lebih merata.

Evaluasi Model Naive Bayes dengan SMOTE:				
	precision	recall	f1-score	support
negatif	0.00	0.00	0.00	1
netral	0.00	0.00	0.00	4
positif	0.88	0.76	0.82	50
accuracy			0.69	55
macro avg	0.29	0.25	0.27	55
weighted avg	0.80	0.69	0.74	55

Gambar 17. Evaluasi Model Naive Bayes Dengan SMOTE



Gambar 18. Hasil Visualisasi Naive Bayes Dengan SMOTE

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma Naive Bayes dalam menganalisis sentimen masyarakat terhadap STMIK Widya Cipta Dharma berdasarkan data ulasan yang diperoleh dari Google Maps. Pendekatan yang digunakan mengikuti metode CRISP-DM, dimulai dari tahap pemahaman bisnis, pemahaman data, persiapan data, pemodelan, hingga evaluasi. Beberapa proses penting yang dilakukan dalam penelitian ini meliputi pembersihan data teks (text cleaning), penghapusan kata-kata umum yang tidak bermakna (stopword removal), proses stemming untuk menyederhanakan kata, serta ekstraksi fitur menggunakan metode TF-IDF. Selain itu, untuk mengatasi permasalahan ketidakseimbangan kelas sentimen, penelitian ini menerapkan teknik Synthetic Minority Oversampling Technique (SMOTE). Hasil dari proses klasifikasi menunjukkan bahwa algoritma Naive Bayes memiliki performa yang sangat baik dalam mengenali sentimen mayoritas, yaitu sentimen positif, dengan tingkat akurasi mencapai 91%. Namun demikian, model mengalami kegagalan dalam mengenali kelas minoritas, yakni sentimen netral dan negatif, yang ditunjukkan oleh nilai presisi dan recall sebesar 0% pada kedua kelas tersebut. Ketika data diseimbangkan menggunakan SMOTE, performa klasifikasi terhadap kelas negatif meningkat secara signifikan dengan recall mencapai 100%. Akan tetapi, hal ini diiringi dengan penurunan akurasi keseluruhan model menjadi 38%, yang menunjukkan adanya trade-off antara akurasi total dan kemampuan model dalam mengenali semua kelas sentimen secara seimbang. Temuan ini memberikan dua implikasi



utama. Pertama, Naïve Bayes tetap menjadi salah satu algoritma yang efisien dan efektif untuk klasifikasi teks dalam jumlah besar. Kedua, penggunaan metode balancing seperti SMOTE penting untuk meningkatkan representasi kelas minoritas, namun perlu diiringi dengan pemilihan strategi evaluasi yang tepat agar performa model tetap optimal secara menyeluruh. Dari sisi praktis, hasil analisis ini memberikan gambaran yang relevan bagi STMIK Widya Cipta Dharma dalam memahami persepsi publik terhadap layanan institusi. Informasi yang diperoleh dari analisis sentimen ini dapat dijadikan dasar dalam menyusun strategi perbaikan mutu layanan, peningkatan kepuasan mahasiswa, serta penguatan citra institusi. Untuk penelitian selanjutnya, disarankan agar pendekatan hybrid seperti penggabungan Naïve Bayes dengan algoritma lain (misalnya SVM atau Random Forest) atau penerapan teknik feature selection dan embedding yang lebih kompleks turut dieksplorasi. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi terhadap pengembangan analisis sentimen di bidang pendidikan tinggi, tetapi juga membuka jalan bagi penelitian lanjutan yang lebih adaptif terhadap tantangan data tidak seimbang.

REFERENCES

- [1] W. Khofifah, D. N. Rahayu, and A. M. Yusuf, "Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps," *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 16, no. 4, pp. 28–38, Jan. 2022, doi: 10.35969/interkom.v16i4.192.
- [2] M. S. Assyifa, "ANALISIS SENTIMEN TERHADAP STMIK WIDYA CIPTA DHARMA MENGGUNAKAN PENDEKATAN LEXICON," SKRIPSI, STMIK WIDYA CIPTA DHARMA, SAMARINDA, 2025.
- [3] E. Eunike, W. Wahyuni, and P. Adytia, "Analisis Sentimen Kepuasan Mahasiswa Terhadap Laboratorium Komputer STMIK Widya Cipta Dharma Menggunakan Algoritma Naïve Bayes," *Wicida Repository*, pp. 1–7, 2024.
- [4] S. Styawati, A. Rahman Isnain, N. Hendrastuty, and L. Andraini, "Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis," *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 6, no. 1, Jan. 2021.
- [5] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *Jurnal TEKNO KOMPAK*, vol. 15, no. 1, pp. 131–145, 2018.
- [6] Affandy and O. Nandiyati, "Sentiment Analysis Berbasis Algoritma Naïve Bayes Classifier untuk Identifikasi Persepsi Masyarakat Terhadap Produk / Layanan Perusahaan," *JOINS Journal of Information System*, vol. 5, no. 1, pp. 126–135, May 2020, doi: 10.33633/joins.v5i1.3608.
- [7] H. P. A. Sormin, D. E. Ratnawat, and N. Y. Setiawan, "ANALISIS SENTIMEN MASYARAKAT TERHADAP LAYANAN UB PRESS DENGAN MENGGUNAKAN METODE NAÏVE BAYES DAN LEXICON-BASED FEATURES," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 1, pp. x–x, Jan. 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] E. Hokijuliandy, H. Napitupulu, and F. Firdaniza, "Analisis Sentimen Menggunakan Metode Klasifikasi Support Vector Machine (SVM) dan Seleksi Fitur Chi-Square," *SisInfo – Jurnal Sistem Informasi dan Informatika*, vol. 5, no. 2, Aug. 2023.
- [9] S. Purnomo, H. Pratiwi, and M. I. Sa'ad, "Penerapan Data Mining Dalam Menganalisis Pola Belanja Konsumen Menggunakan Market Basket Analysis," *METIK JURNAL*, vol. 7, no. 2, pp. 111–120, Dec. 2023, doi: 10.47002/metik.v7i2.678.
- [10] M. A. Brown, A. Gruen, G. Maldoff, S. Messing, Z. Sanderson, and M. Zimmer, "Web Scraping for Research: Legal, Ethical, Institutional, and Scientific Considerations," *Web Scraping for Research*, Dec. 2024, [Online]. Available: <http://arxiv.org/abs/2410.23432>
- [11] D. Chrisinta and J. E. Simarmata, "Eksplorasi Teknik Web Scraping pada Data Mining: Pendekatan Pencarian Data Berbasis Python," *Faktor Exacta*, vol. 17, no. 1, pp. 58–68, Mar. 2024, doi: 10.30998/faktorexacta.v17i1.22393.
- [12] M. A. Khder, "Web scraping or web crawling: State of art, techniques, approaches and application," *International Journal of Advances in Soft Computing and its Applications*, vol. 13, no. 3, pp. 144–168, Nov. 2021, doi: 10.15849/ijasca.211128.11.
- [13] M. K. Insan, U. Hayati, and O. Nurdiawan, "ANALISIS SENTIMEN APLIKASI BRIMO PADA ULASAN PENGGUNA DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAIVE BAYES," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, Feb. 2023.
- [14] T. I. Rais, "ANALISIS SENTIMEN TERHADAP KOMENTAR VIDEO YOUTUBE RAIDEN SHOGUN - JUDGMENT OF EUTHYMIA MENGGUNAKAN METODE MAJORITY VOTING," UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH, Jakarta, 2022.
- [15] D. Dablain, B. Krawczyk, and N. V. Chawla, "DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data," *IEEE Trans Neural Netw Learn Syst*, vol. 34, no. 9, pp. 6390–6404, Sep. 2023, doi: 10.1109/TNNLS.2021.3136503.
- [16] G. Douzas and F. Bacao, "Geometric SMOTE: Effective oversampling for imbalanced learning through a geometric extension of SMOTE," *Geometric SMOTE*, Sep. 2017, [Online]. Available: <http://arxiv.org/abs/1709.07377>
- [17] E. W. H. C. Candana, I. G. A. Gunaidi, and D. G. H. Divayana, "PERBANDINGAN FUZZY TSUKAMOTO, MAMDANI DAN SUGENO DALAM PENENTUAN HARI BAIK PERNIKAHAN BERDASARKAN WARIGA MENGGUNAKAN CONFUSION MATRIX," *Jurnal Ilmu Komputer Indonesia (JIK)*, vol. 6, no. 2, Nov. 2021.





- [18] F. Salsabila, I. Fitrianti, Y. Umaidah, and N. Heryana, "PENERAPAN METODE CRISP-DM UNTUK ANALISA PENDAPATAN BERSIH BULANAN PEKERJA INFORMAL DI PROVINSI JAWA BARAT DENGAN ALGORITMA K-MEANS," *DINAMIK*, vol. 28, no. 2, 2023.
- [19] N. Sakti, F. Natsir, and S. Istianah, "Penentuan Penjualan Barang Berdasarkan Pengelompokan Produk dengan K-Means Clustering Metode CRISP-DM Pada CV.Sembako Dina," *Zetroem*, vol. 5, no. 2, 2023.
- [20] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput Sci*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [21] O. I. Gifari, M. Adha, I. Rifky Hendrawan, F. Freddy, and S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY)*, vol. 2, no. 1, Mar. 2022.
- [22] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Information*, vol. 10, no. 4, pp. 1–68, 2019, doi: 10.3390/info10040150.

