ANALISIS SENTIMEN ULASAN TEMPAT WISATA WADUK GAJAH MUNGKUR MENGGUNAKAN METODE NAÏVE BAYES

Sayid Muchammad Rosyid Aridho, Ridwan Dwi Irawan, Pramono
Teknik Informatika, Universitas Duta Bangsa Surakarta
Jl. Bhayangkara No.55,Tipes.Kec. Serengan, Kota Surakarta, Jawa Tengah
220103213@mhs.udb.ac.id

ABSTRAK

Perkembangan teknologi informasi mendorong masyarakat membagikan opini melalui platform digital seperti Google Maps. Ulasan pengunjung pada objek wisata dapat dimanfaatkan untuk mengetahui kecenderungan opini publik, namun data teks yang tidak terstruktur menyulitkan proses analisis manual. Selain itu, pelabelan berdasarkan rating bintang berpotensi menimbulkan bias karena tidak selalu sesuai dengan isi ulasan. Penelitian ini bertujuan mengklasifikasikan sentimen ulasan pengunjung Waduk Gajah Mungkur menggunakan metode Multinomial Naïve Bayes. Data sebanyak 2.785 ulasan diperoleh melalui *web scraping* dan diproses menggunakan *preprocessing* serta pembobotan TF-IDF. Ketidakseimbangan kelas ditangani menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE), sedangkan optimasi parameter dilakukan menggunakan GridSearchCV. Evaluasi model menggunakan *Stratified 5-Fold Cross Validation* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa model terbaik memperoleh *accuracy* sebesar 66,03%, F1 *Macro* 44,95%, dan *F1 Weighted* 70,25%. Model memiliki performa terbaik pada kelas positif dengan *F1-score* 80,81%, sementara kelas netral (23,20%) dan negatif (30,83%) masih rendah akibat ketidakseimbangan data dan potensi bias pelabelan berbasis rating.

Kata kunci : *Analisis Sentimen, Naïve Bayes, TF-IDF, SMOTE, Google Maps*

1. PENDAHULUAN

Perkembangan teknologi saat ini membuat masyarakat semakin mudah menyampaikan pendapat melalui platform digital seperti Google Maps. Selain digunakan sebagai layanan navigasi, Google Maps juga menyediakan fitur ulasan dan rating yang memungkinkan pengguna memberikan penilaian terhadap suatu tempat. Ulasan tersebut menjadi sumber informasi penting, terutama pada bidang pariwisata, karena sering dijadikan acuan oleh calon pengunjung sebelum mendatangi suatu lokasi.

Waduk Gajah Mungkur merupakan salah satu destinasi wisata di Kabupaten Wonogiri yang memiliki banyak ulasan pengunjung di Google Maps. Banyaknya data ulasan dalam bentuk teks tidak terstruktur membuat proses analisis secara manual menjadi tidak efisien dan sulit dilakukan secara menyeluruh. Selain itu, rating bintang yang diberikan pengguna tidak selalu sesuai dengan isi ulasan sehingga berpotensi menimbulkan bias dalam penentuan sentimen [1], [2].

Distribusi data yang tidak seimbang juga menjadi permasalahan tersendiri karena model klasifikasi cenderung lebih fokus pada kelas mayoritas sehingga kemampuan mengenali kelas minoritas menjadi kurang optimal [3].

Penelitian ini bertujuan mengklasifikasikan sentimen ulasan pengunjung Waduk Gajah Mungkur pada platform Google Maps menggunakan metode Multinomial Naïve Bayes dengan pembobotan TF-IDF. Untuk memperoleh model yang optimal, diterapkan SMOTE dalam menangani ketidakseimbangan distribusi kelas, *hyperparameter tuning* menggunakan *GridSearchCV*, serta evaluasi menggunakan *Stratified 5-Fold Cross Validation* dengan metrik *accuracy*, *precision*, *recall*, dan F1-*score.*

2. TINJAUAN PUSTAKA

Beberapa penelitian terdahulu yang relevan telah menerapkan metode Naïve Bayes untuk analisis sentimen pada ulasan Google Maps. Analisis sentimen ulasan wisata di Kalimantan Barat menggunakan Naïve Bayes dengan TF-IDF memperoleh akurasi 0,76 dengan mayoritas ulasan bersentimen positif [4].

Penerapan Naïve Bayes pada ulasan Google Maps layanan rumah sakit menghasilkan performa cukup tinggi dengan *accuracy* 84% dan F1-*score* 90% [5].

Analisis sentimen pada ulasan wisata pantai di Karawang menemukan variasi sentimen dengan dominasi negatif pada beberapa objek wisata [6].

Naïve Bayes masih banyak digunakan dalam analisis sentimen ulasan Google Maps karena kesederhanaan dan efisiensi komputasinya. Namun, penelitian-penelitian tersebut belum menangani ketidakseimbangan distribusi data, belum menerapkan *stratified cross validation*, serta menggunakan pelabelan berbasis rating yang berpotensi tidak konsisten dengan isi teks ulasan. Oleh karena itu, penelitian ini menambahkan SMOTE, Stratified 5-Fold *Cross Validation*, dan *hyperparameter tuning* pada model Naïve Bayes untuk menghasilkan klasifikasi sentimen yang lebih representatif.

2.1. Analisis Sentimen

Dalam bidang *Natural Language Processing* (NLP), analisis sentimen digunakan untuk

mengidentifikasi kecenderungan opini yang terkandung dalam suatu teks. Secara umum, hasil klasifikasi sentimen dibedakan menjadi tiga kategori utama, yaitu positif, negatif, dan netral [2].

Analisis sentimen dapat dilakukan pada tiga level: *document-level*, *sentence-level*, dan *aspect-level* [2].

Penelitian ini menggunakan pendekatan *document-level* karena setiap ulasan pengunjung diperlakukan sebagai satu dokumen utuh yang merepresentasikan opini secara keseluruhan.

2.2. Text Mining

Text mining digunakan untuk mengekstraksi informasi dari data berbentuk teks yang sebelumnya tidak terstruktur agar dapat diproses oleh algoritma *machine learning*. Dalam analisis sentimen, *text mining* mencakup tiga tahapan yang saling terhubung, yaitu *text preprocessing* yang meliputi pembersihan data seperti *tokenizing*, *stopword removal*, dan *stemming*; *feature extraction* yang mengubah teks menjadi representasi numerik menggunakan metode seperti TF-IDF; serta *classification* yang mengelompokkan teks ke dalam kelas sentimen menggunakan algoritma seperti Naïve Bayes. Ketiga tahapan ini memungkinkan komputer mengenali pola opini dalam teks secara otomatis dan terstruktur [7].

2.3. Naïve Bayes

Algoritma Naïve Bayes bekerja dengan pendekatan berbasis probabilitas berdasarkan Teorema Bayes dan mengasumsikan setiap fitur bersifat independen. Pada klasifikasi data teks, Naïve Bayes sering digunakan karena memiliki proses komputasi yang sederhana dan efisien [8].

Secara matematis, *Teorema Bayes* dinyatakan sebagai berikut:

$$P(C \mid X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (1)$$

Keterangan:

- P(C|X): probabilitas posterior, yaitu peluang data X termasuk ke dalam kelas C
- P(X|C): probabilitas *likelihood*, yaitu peluang munculnya data X pada kelas C
- P(C): probabilitas *prior*, yaitu peluang awal dari kelas C
- P(X): probabilitas *evidence*, yaitu peluang munculnya data X secara keseluruhan.

Dalam proses klasifikasi teks, suatu dokumen direpresentasikan sebagai kumpulan fitur atau kata X=(x_1,x_2,...,x_n). Dengan asumsi independensi antar fitur, maka persamaan Naïve Bayes dapat dikembangkan menjadi:

$$P(C \mid X) = P(C) \times \prod_{i=1}^{n} P(x_i \mid C) \quad (2)$$

Keterangan:

- x_i: fitur ke-i (kata dalam dokumen)
- n: jumlah seluruh fitur (kata) dalam dokumen
- ∏: simbol perkalian untuk seluruh probabilitas fitur.

2.4. TF-IDF

TF-IDF (Term Frequency – Inverse Document Frequency) merupakan metode pembobotan kata yang digunakan untuk menentukan tingkat kepentingan suatu kata terhadap dokumen tertentu dalam kumpulan dokumen. Dalam analisis teks, TF-IDF membantu menghasilkan representasi fitur yang lebih informatif sebelum proses klasifikasi dilakukan [2].

TF-IDF terdiri dari dua komponen utama, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF). Term Frequency menunjukkan seberapa sering suatu kata muncul dalam dokumen, sedangkan Inverse Document Frequency menunjukkan seberapa penting kata tersebut dalam keseluruhan dokumen.

Secara matematis dirumuskan sebagai berikut:

$$TF(t,d) = \frac{f(t,d)}{\sum_{t'} f(t',d)} \quad (3)$$

$$IDF(t) = \log\left(\frac{N}{df(t)}\right) \quad (4)$$

$$TF\text{-}IDF(t,d) = TF(t,d) \times IDF(t) \quad (5)$$

Keterangan:

f(t,d) : jumlah kemunculan kata t pada dokumen d,
N : jumlah total dokumen, dan
df(t) : jumlah dokumen yang mengandung kata t.

Dengan metode ini, kata yang sering muncul dalam suatu dokumen namun jarang muncul pada dokumen lain akan memiliki bobot lebih tinggi, sehingga membantu model dalam mengidentifikasi fitur yang lebih informatif.

2.5. SMOTE

SMOTE (Synthetic Minority Over-sampling Technique) merupakan teknik yang digunakan untuk mengatasi permasalahan ketidakseimbangan data dengan cara menghasilkan data sintetis pada kelas minoritas. Metode ini bekerja dengan membuat sampel baru berdasarkan kedekatan antar data dalam ruang fitur, sehingga distribusi kelas menjadi lebih seimbang. Penggunaan SMOTE terbukti dapat meningkatkan performa model klasifikasi pada dataset yang tidak seimbang [3].

2.6. Cross Validation

Cross Validation merupakan teknik evaluasi model dalam *machine learning* yang bekerja dengan membagi dataset menjadi beberapa bagian (*fold*), kemudian model dilatih dan diuji secara bergantian sehingga setiap data mendapat kesempatan menjadi data uji. Teknik ini lebih andal dibandingkan pembagian data satu kali (*train-test split*) karena hasil evaluasinya tidak bergantung pada satu partisi data tertentu.

Performa model diukur menggunakan empat metrik evaluasi. *Accuracy* mengukur proporsi prediksi yang benar terhadap keseluruhan data uji. *Precision* mengukur ketepatan model dalam memprediksi suatu kelas, yaitu seberapa besar proporsi prediksi yang memang benar-benar sesuai kelasnya. *Recall* mengukur kemampuan model dalam mendeteksi seluruh data yang sebenarnya termasuk dalam suatu

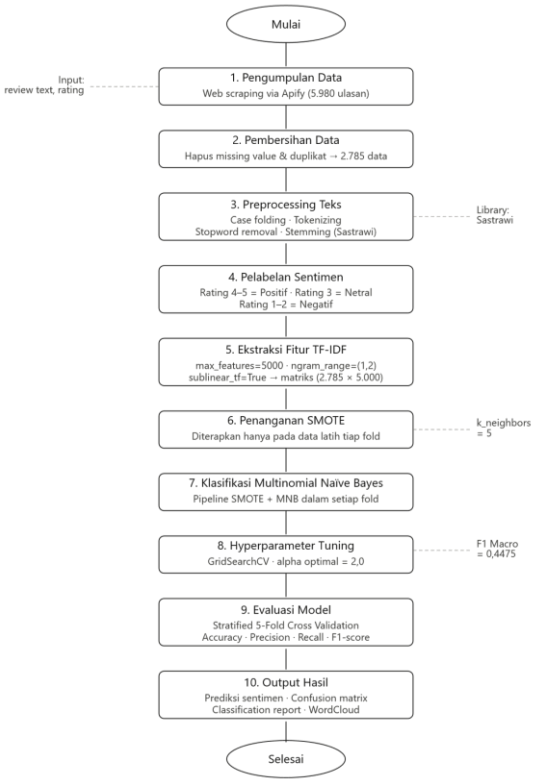
kelas. *F1-score* merupakan rata-rata harmonik antara *precision* dan *recall* yang memberikan gambaran keseimbangan antara keduanya, terutama berguna ketika distribusi kelas tidak seimbang [9].

2.7. Google Maps

Google Maps adalah layanan peta online yang dapat diakses melalui web maupun perangkat mobile. Aplikasi ini tidak hanya menyediakan petunjuk arah, tetapi juga berbagai informasi seperti tempat wisata, lokasi kuliner, kondisi lalu lintas, serta rute alternatif. Selain itu, pengguna juga dapat memberikan ulasan dan penilaian terhadap tempat yang pernah dikunjungi, sehingga menjadi sumber informasi penting bagi pengguna lain [10].

3. METODE PENELITIAN

Metode eksperimen diterapkan dalam penelitian ini untuk menguji performa klasifikasi sentimen pada ulasan Google Maps. Seluruh implementasi dilakukan menggunakan *Python* dengan *library* Pandas, *Scikit-learn*, Sastrawi, dan Matplotlib di lingkungan Visual Studio Code. Gambaran umum alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur penelitian analisis sentimen ulasan Waduk Gajah Mungkur menggunakan metode Naïve Bayes

Berdasarkan Gambar 1, alur penelitian dimulai dengan pengumpulan data, pembersihan data, *preprocessing* teks, dan pelabelan sentimen berbasis rating. Selanjutnya dilakukan ekstraksi fitur TF-IDF, penanganan ketidakseimbangan kelas menggunakan SMOTE, klasifikasi menggunakan Multinomial Naïve Bayes, serta *hyperparameter tuning* untuk menentukan konfigurasi model optimal. Tahap akhir adalah evaluasi model menggunakan Stratified 5-Fold *Cross Validation* hingga menghasilkan output berupa prediksi sentimen, *confusion matrix*, dan *word cloud*. Penjelasan rinci setiap tahapan diuraikan pada subbab berikut.

3.1. Pengumpulan dan Pembersihan Data

Data berupa ulasan pengunjung Waduk Gajah Mungkur dikumpulkan dari platform Google Maps melalui teknik *web scraping* menggunakan layanan Apify. Atribut yang diambil meliputi teks ulasan dan rating bintang (skala 1–5), karena kedua atribut tersebut secara langsung merepresentasikan opini pengguna dan umum digunakan dalam penelitian analisis sentimen [1].

Data awal yang berhasil dikumpulkan berjumlah 5.980 ulasan.

Selanjutnya dilakukan *data cleaning* untuk memastikan kualitas dataset sebelum memasuki tahap analisis. Proses ini mencakup penghapusan data dengan nilai kosong (*missing value*) sebanyak 3.005 data dan data duplikat sebanyak 3.054 data. Setelah proses pembersihan selesai, diperoleh dataset bersih sebanyak 2.785 ulasan yang valid dan siap digunakan pada tahap selanjutnya [9].

3.2. *Preprocessing* Text

Tahap *preprocessing* bertujuan membersihkan data teks agar dapat diproses secara optimal oleh algoritma klasifikasi. Kualitas data sangat menentukan performa model yang dihasilkan, karena teks mentah dari ulasan pengguna umumnya mengandung banyak *noise* seperti karakter khusus, huruf kapital tidak konsisten, dan kata-kata tidak bermakna [9].

Tahapan *preprocessing* yang diterapkan meliputi *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. *Case folding* dilakukan dengan mengubah seluruh huruf menjadi huruf kecil, sedangkan *tokenizing* digunakan untuk memecah kalimat menjadi kata-kata. Selanjutnya dilakukan *stopword removal* untuk menghapus kata umum yang tidak memiliki pengaruh signifikan terhadap analisis sentimen menggunakan library Sastrawi. Tahap terakhir adalah *stemming*, yaitu mengubah kata menjadi bentuk dasar dengan menghilangkan imbuhan menggunakan algoritma ECS Indo pada library Sastrawi [9]. Hasil dari tahap ini berupa teks yang lebih bersih dan siap digunakan pada proses ekstraksi fitur.

3.3. Pelabelan Data

Setiap ulasan diberikan label sentimen secara otomatis berdasarkan nilai rating yang diberikan pengguna. Pendekatan pelabelan berbasis rating ini dipilih karena rating dianggap mencerminkan kecenderungan opini pengguna secara umum dan telah banyak digunakan dalam penelitian analisis sentimen sebelumnya. Ketentuan pelabelan yang diterapkan

adalah rating 4–5 dikategorikan sebagai sentimen positif, rating 3 sebagai sentimen netral, dan rating 1–2 sebagai sentimen negatif.

Pembagian label tersebut mengacu pada tiga kategori utama dalam analisis sentimen, yaitu positif, netral, dan negatif. Pendekatan ini termasuk dalam klasifikasi *document-level sentiment analysis*, karena setiap ulasan diperlakukan sebagai satu kesatuan opini terhadap suatu objek atau layanan [1].

3.4. Ekstraksi Fitur dengan TF-IDF

Setelah melalui tahap *preprocessing*, teks ulasan dikonversi menjadi representasi numerik menggunakan metode TF-IDF dengan parameter max_features=5000, ngram_range=(1,2), dan sublinear_tf=True menggunakan TfidfVectorizer dari library Scikit-learn [2].

Proses ini menghasilkan matriks fitur berukuran (2.785 × 5.000) yang selanjutnya digunakan sebagai input pada tahap penanganan ketidakseimbangan kelas menggunakan SMOTE dan proses klasifikasi Multinomial Naïve Bayes.

3.5. Penanganan SMOTE

Dataset pada penelitian ini memiliki distribusi kelas yang tidak seimbang, di mana jumlah data sentimen positif jauh lebih banyak dibandingkan kelas netral dan negatif. Kondisi tersebut dapat menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas sehingga kemampuan dalam mengenali kelas minoritas menjadi kurang optimal.

Untuk mengatasi permasalahan tersebut, diterapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) setelah tahap ekstraksi fitur TF-IDF. SMOTE digunakan untuk menyeimbangkan distribusi data dengan menghasilkan sampel sintetis pada kelas minoritas berdasarkan kedekatan antar data dalam ruang fitur.

Dengan distribusi data yang lebih seimbang, model diharapkan dapat belajar secara lebih optimal dan meningkatkan kemampuan klasifikasi pada seluruh kelas sentimen, terutama pada kelas minoritas [3].

3.6. Klasifikasi dengan Multinomial Naïve Bayes

Model klasifikasi yang digunakan dalam penelitian ini adalah Multinomial Naïve Bayes, yaitu varian Naïve Bayes yang dirancang khusus untuk data dengan distribusi frekuensi diskret seperti jumlah kemunculan kata dalam dokumen teks [8].

Algoritma ini bekerja berdasarkan Teorema Bayes dengan menghitung probabilitas suatu dokumen termasuk ke dalam kelas tertentu berdasarkan distribusi kata yang dikandungnya. Multinomial Naïve Bayes dipilih karena kesederhanaannya, kecepatan komputasi yang tinggi, serta kemampuannya bekerja secara efektif pada dataset teks berskala besar. Model dilatih menggunakan data yang telah melewati proses TF-IDF, kemudian digunakan untuk memprediksi label sentimen pada data uji di setiap iterasi *cross validation*.

3.7. Hyperparameter Tuning

Untuk memperoleh performa model yang lebih optimal, dilakukan proses *hyperparameter tuning* pada parameter alpha (α) algoritma Multinomial Naïve Bayes menggunakan metode *GridSearchCV*. Parameter alpha digunakan sebagai *additive smoothing* untuk menghindari probabilitas nol pada kata yang tidak muncul dalam data latih serta membantu mengontrol generalisasi model [8].

Proses tuning dilakukan dengan menguji beberapa kandidat nilai alpha menggunakan *Stratified 5-Fold Cross Validation* dan metrik evaluasi F1 Macro sebagai acuan pemilihan parameter terbaik. Penggunaan F1 Macro dipilih karena mampu memberikan evaluasi yang lebih seimbang pada dataset tidak seimbang [3].

Nilai alpha terbaik yang diperoleh kemudian digunakan sebagai parameter final pada tahap evaluasi model.

3.8. Evaluasi dengan 5-Fold *Cross Validation*

Evaluasi model penelitian ini menggunakan *Stratified 5-Fold Cross Validation*. Dataset dibagi menjadi 5-*fold* dengan proporsi kelas yang tetap seimbang pada setiap iterasi. Proses pelatihan dan pengujian dilakukan sebanyak 5 kali secara bergantian, di mana 1-*fold* digunakan sebagai data uji dan 4-*fold* lainnya digunakan sebagai data latih. Nilai performa akhir diperoleh dari rata-rata hasil evaluasi seluruh iterasi. Pendekatan ini dipilih karena mampu menghasilkan evaluasi yang lebih stabil dan mengurangi bias akibat ketidakseimbangan distribusi kelas pada dataset [11].

Performa model diukur menggunakan empat metrik evaluasi yang umum digunakan dalam penelitian klasifikasi teks, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* dengan *F1 Macro* sebagai metrik utama karena memberikan bobot yang sama pada seluruh kelas tanpa memandang ketidakseimbangan distribusinya. Keempat metrik ini dihitung pada setiap *fold* kemudian dirata-ratakan untuk memperoleh nilai performa akhir model.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

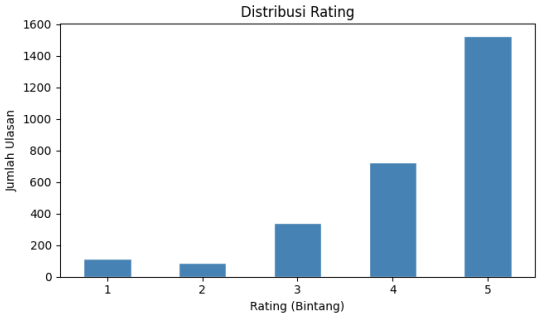
Pengumpulan data dilakukan menggunakan teknik *web scraping* pada platform Google Maps menggunakan layanan Apify. Data yang diperoleh berupa teks ulasan dan rating bintang (1–5) dari pengunjung Waduk Gajah Mungkur. Hasil *scraping* kemudian diekspor dalam format CSV untuk diproses menggunakan Python.

Selanjutnya dilakukan tahap data *cleaning* dengan menghapus data duplikat, *missing value*, dan ulasan yang tidak memiliki teks bermakna. Setelah proses pembersihan, diperoleh sebanyak 2.785 ulasan

valid yang siap digunakan pada tahap analisis berikutnya.

4.2. Deskripsi dan Distribusi Dataset

Dataset hasil *scraping* yang telah dibersihkan selanjutnya dianalisis distribusinya berdasarkan nilai rating. Distribusi rating pada dataset ditampilkan pada Gambar 2.



Gambar 2. Distribusi rating ulasan pengunjung Waduk Gajah Mungkur

Berdasarkan Gambar 2, distribusi rating pada dataset menunjukkan pola yang sangat tidak merata antar nilai bintang. Secara rinci, rating 5 bintang mendominasi dengan jumlah 1.525 ulasan, diikuti rating 4 bintang sebanyak 723 ulasan, rating 3 bintang sebanyak 338 ulasan, rating 1 bintang sebanyak 113 ulasan, dan rating 2 bintang sebanyak 86 ulasan. Distribusi rating secara lengkap ditampilkan pada Tabel 1.

Tabel 1. Distribusi rating ulasan pengunjung Waduk Gajah Mungkur

Rating	Jumlah Ulasan	Persentase
5 Bintang	1.525	54,8%
4 Bintang	723	26,0%
3 Bintang	338	12,1%
2 Bintang	86	3,1%
1 Bintang	113	4,1%
Total	2.785	100%

Berdasarkan Tabel 1, rating 5 bintang mendominasi dataset dengan persentase 54,8%, diikuti rating 4 bintang sebesar 26,0%. Sementara itu, rating 1 dan 2 bintang hanya berjumlah 7,2% dari total data, sedangkan rating 3 bintang sebesar 12,1%. Distribusi ini menunjukkan bahwa dataset bersifat tidak seimbang dan didominasi oleh ulasan positif. Kondisi tersebut menjadi pertimbangan dalam proses pelabelan sentimen serta penerapan metode penanganan ketidakseimbangan kelas pada tahap pemodelan [12].

4.3. *Preprocessing* Teks

Tahap preprocessing menghasilkan teks yang lebih bersih dan terstruktur sehingga dapat meningkatkan kualitas data sebelum digunakan dalam proses klasifikasi. Proses ini dilakukan untuk mengurangi noise pada data teks dan memastikan konsistensi representasi kata. Contoh perbandingan teks sebelum dan sesudah *preprocessing* ditampilkan pada Tabel 2.

Tabel 2. Contoh hasil *preprocessing* teks ulasan

No	Teks Asli	Hasil *Preprocessing*
1	Istimewa	istimewa
2	Tempat Wisata Asyik di Wonogiri	Tempat Wisata Asyik Wonogiri
3	Bagus waduknya, kalo sabtu minggu ramai...	bagus waduk kalo sabtu minggu ramai...
4	Sangat menyenangkan	sangat senang
5	Mantappp	mantap

Dari Tabel 2 dapat dilihat bahwa *preprocessing* berhasil menyederhanakan teks secara signifikan. Kata depan "di" berhasil dihapus sebagai *stopword*, kata "menyenangkan" berhasil distem menjadi bentuk dasarnya "senang", serta seluruh tanda baca dan karakter tidak relevan berhasil dibersihkan. Hasil *preprocessing* ini menghasilkan representasi teks yang lebih bersih dan konsisten sehingga siap untuk diproses pada tahap ekstraksi fitur.

4.4. Pelabelan Sentimen

Setelah tahap *preprocessing*, setiap ulasan diberikan label sentimen berdasarkan *rating* bintang pada Google Maps menggunakan fungsi label_sentiment(). Pendekatan pelabelan berbasis *rating* dipilih karena banyak digunakan dalam penelitian analisis sentimen dan dianggap mampu merepresentasikan opini pengguna secara umum. Distribusi kelas sentimen hasil pelabelan ditampilkan pada Tabel 3.

Tabel 3. Distribusi kelas sentimen hasil pelabelan

Kelas Sentimen	Rentang Rating	Jumlah Ulasan	Persentase
Positif	4-5 Bintang	2.248	80,72%
Netral	3 Bintang	338	12,14%
Negatif	1-2 Bintang	199	7,15%
Total		2.785	100%

Dari Tabel 3 terlihat bahwa distribusi kelas sentimen tidak seimbang, dengan kelas positif mendominasi sebanyak 2.248 ulasan (80,72%), diikuti kelas netral 338 ulasan (12,14%), dan kelas negatif 199 ulasan (7,15%). Ketidakseimbangan ini berpotensi menyebabkan model lebih fokus mempelajari pola pada kelas mayoritas sehingga performa klasifikasi pada kelas netral dan negatif menjadi kurang optimal. Oleh karena itu, setelah tahap ekstraksi fitur *TF-IDF*, diterapkan metode *Synthetic Minority Over-sampling Technique* (*SMOTE*) untuk menyeimbangkan distribusi kelas sebelum proses pelatihan model.

4.5. Ekstraksi Fitur TF-IDF dan Pembagian Data

Setelah proses pelabelan sentimen, teks pada kolom *processed_text* dikonversi menjadi representasi

numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Proses ini diperlukan karena algoritma Multinomial Naïve Bayes hanya dapat memproses data dalam bentuk numerik.

Parameter TF-IDF yang digunakan meliputi max_features=5000, ngram_range=(1,2), dan sublinear_tf=True. Proses vektorisasi menggunakan TfidfVectorizer dari *library Scikit-learn* menghasilkan matriks fitur berukuran (2.785 × 5.000), yang menunjukkan bahwa seluruh ulasan berhasil direpresentasikan menjadi fitur numerik secara konsisten.

Untuk mengetahui kata yang paling informatif, dilakukan analisis terhadap 10 *term* dengan rata-rata bobot TF-IDF tertinggi. Hasil analisis ditampilkan pada Tabel 4.

Tabel 4. 10 *term* dengan rata-rata bobot TF-IDF tertinggi

Peringkat	*Term*	Rata-rata bobot
1	tempat	0,0351
2	bagus	0,0274
3	wisata	0,0268
4	banyak	0,0200
5	yg	0,0190
6	waduk	0,0188
7	keluarga	0,0187
8	kurang	0,0184
9	buat	0,0162
10	luas	0,0138

Dari Tabel 4 dapat dilihat bahwa kata "tempat" memiliki bobot TF-IDF tertinggi sebesar 0,0351, diikuti kata "bagus" dan "wisata". Kemunculan kata-kata tersebut menunjukkan bahwa sebagian besar ulasan membahas kondisi dan pengalaman pengunjung terhadap tempat wisata. Selain itu, kata "waduk" juga muncul sebagai salah satu *term* dominan, yang menunjukkan bahwa ulasan secara konsisten merujuk pada objek penelitian.

Secara keseluruhan, kemunculan kata "bagus" (konotasi positif) dan "kurang" (konotasi negatif) dalam 10 besar menunjukkan bahwa representasi fitur cukup seimbang untuk membedakan polaritas sentimen. Matriks TF-IDF berukuran (2.785 × 5.000) selanjutnya digunakan sebagai input pada tahap SMOTE dan klasifikasi Multinomial Naïve Bayes.

4.6. Klasifikasi Naïve Bayes dengan *Cross Validation*

Pada tahap ini dilakukan proses klasifikasi menggunakan algoritma Multinomial Naïve Bayes. Evaluasi model dilakukan dengan metode *Stratified 5-Fold Cross Validation*, yang bertujuan untuk mengukur performa model secara lebih stabil dengan mempertahankan proporsi kelas pada setiap f*old*. Hasil evaluasi model baseline ditampilkan pada Gambar 3.

```
=== Hasil Baseline — 5-Fold Stratified CV ===
Akurasi CV (per fold) : [0.8061 0.8079 0.8079 0.8061 0.8079]
Rata-rata Akurasi     : 0.8072 ± 0.0009
Rata-rata F1 Macro    : 0.2997 ± 0.0039
Rata-rata F1 Weighted : 0.7219 ± 0.0014
Rata-rata Precision   : 0.3359
Rata-rata Recall      : 0.3342
```

Gambar 3. Hasil evaluasi model Naïve Bayes menggunakan Stratified *5-Fold* Cross Validation

Berdasarkan Gambar 3, model *Multinomial Naïve Bayes* menghasilkan rata-rata akurasi sebesar 0,8072 dengan standar deviasi ±0,0009, yang menunjukkan performa cukup stabil pada setiap *fold* dalam *Stratified 5-Fold Cross Validation*. Namun, nilai *F1-score macro* sebesar 0,2997 menunjukkan bahwa kemampuan model dalam mengklasifikasikan seluruh kelas secara seimbang masih rendah, terutama pada kelas minoritas seperti sentimen netral dan negatif. Sementara itu, nilai *F1-score weighted* sebesar 0,7219 dipengaruhi oleh dominasi kelas positif dalam dataset. Nilai *precision* sebesar 0,3359 dan *recall* sebesar 0,3342 juga menunjukkan bahwa kemampuan model dalam mengenali seluruh kelas sentimen masih terbatas. Oleh karena itu, pada tahap selanjutnya diterapkan metode *SMOTE* untuk meningkatkan performa klasifikasi pada kelas minoritas.

4.7. Penanganan Data Menggunakan SMOTE

Berdasarkan hasil evaluasi model *baseline*, meskipun akurasi model cukup tinggi, performa klasifikasi pada kelas minoritas masih rendah. Hal ini terlihat dari nilai *F1-score macro* yang rendah, sehingga model belum mampu mengenali sentimen netral dan negatif secara optimal akibat distribusi dataset yang tidak seimbang dan didominasi kelas positif.

Untuk mengatasi permasalahan tersebut, diterapkan metode *Synthetic Minority Over-sampling Technique* (*SMOTE*) setelah tahap ekstraksi fitur *TF-IDF*. Penerapan *SMOTE* dilakukan hanya pada data latih di setiap iterasi *cross validation* untuk menghindari *data leakage* dan menjaga validitas evaluasi model.

Hasil evaluasi model setelah penerapan SMOTE menggunakan metode *Stratified 5-Fold Cross Validation* ditampilkan pada Gambar 4.

```
=== Hasil Setelah SMOTE — 5-Fold Stratified CV =
Rata-rata Akurasi     : 0.6528 ± 0.0191
Rata-rata F1 Macro    : 0.4431 ± 0.0056
Rata-rata F1 Weighted : 0.6969 ± 0.0124
Rata-rata Precision   : 0.4362
Rata-rata Recall      : 0.5070
```

Gambar 4. Hasil evaluasi model setelah penerapan SMOTE

Berdasarkan Gambar 4, rata-rata akurasi model menurun menjadi 0,6528 dengan standar deviasi ±0,0191. Namun, nilai *F1-score macro* meningkat menjadi 0,4431, yang menunjukkan perbaikan

kemampuan model dalam mengenali kelas minoritas. Nilai *precision* dan *recall* juga meningkat masing-masing menjadi 0,4362 dan 0,5070. Sementara itu, *F1-score weighted* sedikit menurun menjadi 0,6969 akibat berkurangnya dominasi kelas mayoritas selama proses pelatihan.

Perbandingan performa model sebelum dan sesudah SMOTE disajikan pada Tabel 5.

Tabel 5. Perbandingan performa model sebelum dan sesudah SMOTE

Metrik	Baseline	SMOTE	Delta
Akurasi	0,8072	0,6528	-0,1544
F1 Macro	0,2997	0,4431	+0,1434
F1 Weighted	0,7219	0,6969	-0,0250
Precision	0,3359	0,4362	+0,1003
Recall	0,3342	0,5070	+0,1728

Berdasarkan Tabel 5, penerapan *SMOTE* terbukti meningkatkan performa model dalam menangani ketidakseimbangan data, terutama pada metrik *F1-score macro*, *precision*, dan *recall*. Meskipun akurasi menurun, model menjadi lebih representatif dalam mengenali seluruh kelas sentimen, khususnya kelas minoritas.

4.8. Hyperparameter Tuning

Setelah penerapan SMOTE menunjukkan peningkatan performa model, selanjutnya dilakukan proses *hyperparameter tuning* untuk menentukan nilai alpha optimal pada algoritma Multinomial Naïve Bayes. Hasil tuning menggunakan GridSearchCV dengan *Stratified 5-Fold Cross Validation* pada tujuh kandidat nilai alpha disajikan pada Tabel 6.

Tabel 6. Hasil *hyperparameter tuning* nilai alpha

Alpha (α)	F1 Macro CV (mean)	F1 Macro CV (std)	F1 Train (mean)
0,01	0,4117	0,0100	0,8120
0,10	0,4381	0,0125	0,7868
0,30	0,4443	0,0101	0,7617
0,50	0,4429	0,0091	0,7494
0,70	0,4435	0,0088	0,7396
1,00	0,4431	0,0056	0,7290
2,00	**0,4475**	**0,0078**	**0,7107**

Berdasarkan Tabel 6, nilai alpha=2,0 menghasilkan F1 Macro CV tertinggi sebesar 0,4475 dengan standar deviasi terkecil 0,0078, sehingga dipilih sebagai nilai alpha optimal. Pola pada tabel menunjukkan bahwa nilai alpha yang terlalu kecil (0,01) menyebabkan *overfitting*, yang ditandai oleh gap besar antara F1 Train (0,8120) dan F1 Macro CV (0,4117). Seiring meningkatnya nilai alpha, gap tersebut menyempit karena regularisasi yang semakin kuat, hingga mencapai nilai terkecil sebesar 0,2632 pada alpha=2,0. Nilai alpha optimal ini selanjutnya digunakan sebagai konfigurasi model pada tahap evaluasi akhir.

4.9. Evaluasi Akhir Model

Setelah penerapan SMOTE dan penentuan nilai alpha optimal sebesar 2,0 melalui *hyperparameter tuning*, model akhir dievaluasi melalui *classification report* yang mencakup *precision*, *recall*, dan F1-*score* untuk masing-masing kelas sentimen. Hasil evaluasi akhir model ditampilkan pada Gambar 5.

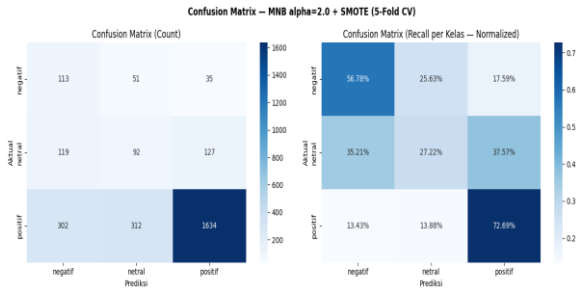
```
Best Alpha : 2.0

=== Classification Report Final ===

              precision    recall  f1-score   support

     negatif     0.2116    0.5678    0.3083       199
      netral     0.2022    0.2722    0.2320       338
     positif     0.9098    0.7269    0.8081      2248

    accuracy                         0.6603      2785
   macro avg     0.4412    0.5223    0.4495      2785
weighted avg     0.7740    0.6603    0.7025      2785

=== HASIL EVALUASI FINAL ===
Akurasi         : 66.03%
F1 Macro        : 44.95%
F1 Weighted     : 70.25%
Precision Macro : 44.12%
Recall Macro    : 52.23%
```

Gambar 5. Hasil evaluasi akhir model

Berdasarkan Gambar 5, model *Multinomial* Naïve Bayes dengan penerapan SMOTE menghasilkan akurasi sebesar 66,03%, nilai F1 *Macro* sebesar 44,95%, dan F1 Weighted sebesar 70,25%. Nilai *recall macro* sebesar 52,23% menunjukkan bahwa model cukup mampu mendeteksi seluruh kelas sentimen, sedangkan *precision macro* sebesar 44,12% menunjukkan bahwa masih terdapat kesalahan prediksi pada beberapa kelas, terutama kelas minoritas.

Kelas positif mencatat performa terbaik dengan *precision* 0,9098, *recall* 0,7269, dan F1-*score* 0,8081, didukung oleh dominasi jumlah data sehingga model memiliki lebih banyak pola pembelajaran. Sebaliknya, kelas negatif memperoleh F1-*score* 0,3083 dan kelas netral hanya 0,2320, menunjukkan bahwa model masih kesulitan membedakan karakteristik kedua kelas minoritas secara konsisten meskipun SMOTE telah diterapkan. Secara keseluruhan, Multinomial Naïve Bayes cukup efektif untuk kelas mayoritas, namun performanya pada kelas minoritas masih perlu ditingkatkan.

4.10. Confusion Matrix



Gambar 6. Confusion Matrix

Berdasarkan Gambar 6, *confusion matrix* digunakan untuk menganalisis perbandingan antara label aktual dan hasil prediksi model pada setiap kelas sentimen. Matriks ini menunjukkan perbedaan performa model yang signifikan antar kelas. Kelas positif mencatat prediksi benar tertinggi (1.634 data), meskipun masih terdapat kesalahan ke kelas negatif (302 data) dan netral (312 data). Kelas negatif memprediksi dengan benar 113 data namun salah ke kelas netral (51 data) dan positif (35 data). Kelas netral mencatat performa terendah dengan hanya 92 prediksi benar, sementara sebagian besar data salah diklasifikasikan ke kelas negatif (119 data) dan positif (127 data), menjadikannya kelas paling sulit dikenali. Hasil ini menegaskan bahwa model masih kesulitan mengenali kelas minoritas, sejalan dengan nilai F1 Macro yang relatif rendah.

4.11. Visualisasi *word cloud*

Pada tahap ini dilakukan visualisasi data menggunakan *word cloud* untuk mengetahui kata-kata yang paling sering muncul pada masing-masing kelas sentimen. Visualisasi ini bertujuan untuk memberikan gambaran umum mengenai kecenderungan opini pengunjung berdasarkan kata-kata dominan yang digunakan dalam ulasan. Hasil visualisasi ditampilkan pada Gambar 7.

Gambar 7. *Word cloud* per kelas sentimen ulasan pengunjung Waduk Gajah Mungkur

Berdasarkan Gambar 7, sentimen positif didominasi kata "tempat", "wisata", "bagus", dan "waduk", mencerminkan kepuasan pengunjung terhadap daya tarik dan suasana wisata. Sentimen netral didominasi kata "tempat", "kurang", "parkir", dan "masuk", mengindikasikan ulasan yang lebih bersifat deskriptif terhadap kondisi fasilitas tanpa kecenderungan opini yang kuat. Sentimen negatif didominasi kata "kurang", "kotor", "mahal", dan "fasilitas", mencerminkan ketidakpuasan pengunjung terutama pada aspek kebersihan, harga, dan pengelolaan tempat wisata. Ketiga visualisasi ini memperkuat hasil analisis sentimen sebelumnya dan dapat dijadikan acuan bagi pengelola dalam memprioritaskan perbaikan layanan.

5. KESIMPULAN DAN SARAN

Penelitian ini menerapkan metode Multinomial Naïve Bayes dengan pembobotan TF-IDF untuk mengklasifikasikan sentimen ulasan pengunjung Waduk Gajah Mungkur pada Google Maps menggunakan teks ulasan sebagai input dan rating bintang sebagai acuan pelabelan otomatis. Dataset yang digunakan berjumlah 2.785 ulasan dengan distribusi sentimen positif sebesar 80,72%, netral 12,14%, dan negatif 7,15%. Hasil evaluasi *baseline* menunjukkan akurasi sebesar 80,72%, namun nilai F1 *Macro* yang rendah (0,2997) mengindikasikan bahwa model belum mampu mengenali kelas minoritas secara optimal. Setelah penerapan SMOTE, nilai F1 *Macro* meningkat menjadi 0,4431 dengan akurasi turun menjadi 0,6528 sebagai konsekuensi berkurangnya dominasi kelas mayoritas. Proses *tuning* menggunakan *GridSearchCV* menghasilkan parameter terbaik alpha = 2,0 dengan F1 *Macro* CV sebesar 0,4475. Evaluasi akhir model menghasilkan *accuracy* 66,03%, F1 *Macro* 44,95%, dan F1 *Weighted* 70,25%, di mana model telah menghasilkan klasifikasi sentimen yang berfungsi dengan baik pada kelas positif (F1-*Score* 80,81%), meskipun masih terbatas pada kelas negatif (30,83%) dan netral (23,20%) akibat ketidakseimbangan distribusi data dan potensi bias pada pelabelan berbasis rating. Hasil *word cloud* menunjukkan ulasan positif didominasi kata "bagus", "tempat", dan "wisata", sedangkan ulasan negatif didominasi kata "kurang", "kotor", dan "mahal", yang dapat dijadikan acuan bagi pengelola dalam memprioritaskan perbaikan layanan. Untuk penelitian selanjutnya disarankan menggunakan pelabelan manual, membandingkan dengan algoritma lain seperti SVM atau *Random Forest*, serta menambahkan matriks ROC-AUC dan mengeksplorasi *deep learning* seperti IndoBERT untuk meningkatkan performa klasifikasi khususnya pada kelas minoritas.

DAFTAR PUSTAKA

[1] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, 2nd ed. Cambridge: Cambridge University Press, 2020. doi: 10.1017/9781108639286.

[2] S. Minaee, N. Kalchbrenner, E. Cambria Erik and Nikzad, M. Chenaghlu, and J. Gao, "Deep learning-based text classification: A comprehensive review," *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–40, 2021, doi: 10.1145/3439726.

[3] A. Santosa, R. Puspita, and D. Nugraha, "Analisis sentimen calon presiden 2024 di media sosial X menggunakan Naive Bayes dan SMOTE," *Jurnal Media Informatika Budidarma*, vol. 8, no. 3, pp. 1854–1862, 2024, doi: 10.30865/mib.v8i3.7708.

[4] M. Rifa'i, H. Sujaini, and I. Prawira, "Sentiment analysis objek wisata Kalimantan Barat pada Google Maps menggunakan metode Naïve Bayes," *Jurnal Sistem dan Teknologi Informasi*, vol. 9, no. 3, pp. 247–253, 2021.

[5] A. Rizki, B. Prihartono, and M. Fathurrohman, "Analisis sentimen ulasan Google Maps Rumah Sakit Khalishah di Cirebon dengan algoritma Naïve Bayes," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 7, no. 1, pp. 45–53, 2025.

[6] I. Khofifah, R. Rahayu, and M. Yusuf, "Analisis sentimen menggunakan Naïve Bayes untuk melihat review masyarakat terhadap tempat wisata pantai di Kabupaten Karawang pada ulasan Google Maps," *Jurnal Sistem Informasi*, vol. 8, no. 2, pp. 112–120, 2022.

[7] F. A. Ramadhan, S. H. Sitorus, and T. Rismawan, "Penerapan metode Multinomial Naïve Bayes untuk klasifikasi judul berita clickbait dengan Term Frequency–Inverse Document Frequency," *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, vol. 11, no. 1, pp. 70–78, 2023, doi: 10.26418/justin.v11i1.57452.

[8] S. Raschka, Y. Liu, and V. Mirjalili, *Machine Learning with PyTorch and Scikit-Learn*. Birmingham: Packt Publishing, 2022.

[9] C. C. Aggarwal, *Machine Learning for Text*. Cham: Springer, 2023. doi: 10.1007/978-3-031-36150-8.

[10] A. H. Dalimunthe, A. L. Sipahutar, and S. P. Sipayung, "Analisis sentimen pengunjung wisata heritage Kota Semarang menggunakan Naive Bayes pada ulasan Google Maps," *Jurnal Elektronika dan Teknologi Informasi*, vol. 4, no. 2, pp. 21–28, 2023, doi: 10.5201/jet.v4i2.412.

[11] M. Kuhn and K. Johnson, *Feature Engineering and Selection: A Practical Approach for Predictive Models*. Boca Raton: CRC Press, 2021. doi: 10.1201/9781315108230.

[12] R. Situmorang and L. Tamyis Usrin and Rajagukguk, "Analisis sentimen destinasi wisata di Jawa Barat pada Twitter menggunakan algoritma Naive Bayes Classifier," *Jurnal Sistem Informasi dan Teknik Komputer*, vol. 8, pp. 339–342, 2023, [Online]. Available: http://ojs.fikom-methodist.net/index.php/methotika