

## ANALISIS SENTIMEN ULASAN MIE GACOAN SOLO VETERAN DI GOOGLE MAPS MENGGUNAKAN ALGORITMA NAIVE BAYES

Tariska Zidny Fatikhah<sup>1</sup>, Moh Muhtarom<sup>2</sup>, Intan Oktaviani<sup>3</sup>

<sup>1,2,3</sup>Universitas Duta Bangsa Surakarta

(Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Duta Bangsa Surakarta)

(Jl. Bhayangkara No.55, Tipes, Serengan, Surakarta, Jawa Tengah, telp. (0271) 719552)

e-mail: <sup>1</sup>zidnytariska@gmail.com, <sup>2</sup>muhtarom@udb.ac.id, <sup>3</sup>intan\_oktaviani@udb.ac.id

### Abstrak

*Mie Gacoan Solo Veteran merupakan salah satu cabang retoran cepat saji yang sangat diminati berbagai kalangan, sehingga tidak sedikit masyarakat yang meninggalkan penilaian terhadap restoran tersebut. Penelitian ini bertujuan untuk membantu perusahaan Mie Gacoan Solo Veteran dalam upaya memahami tanggapan konsumen terhadap merek dagang mereka dan bagaimana masyarakat menilai produk yang disajikan berdasarkan opini dan ulasan yang terdapat pada platform Google Maps. Data dikumpulkan melalui proses crawling, kemudian diberi label dan dilakukan preprocessing sebelum diekstraksi menggunakan teknik Term Frequency-Inverse Document Frequency (TF-IDF) serta dimodelkan dengan algoritma Naive Bayes. Hasil pengujian menunjukkan bahwa pada data yang tidak seimbang, model memperoleh akurasi sebesar 86%. Namun, setelah mengatasi ketidakseimbangan data melalui metode oversampling menggunakan Synthetic Minority Over-sampling Technique (SMOTE), akurasi model meningkat menjadi 91%.*

**Kata kunci:** Analisis Sentimen, Python, TF-IDF, SMOTE, Mie Gacoan

### Abstract

*Mie Gacoan Solo Veteran is one of the fastfood restaurant branches that is very popular with various groups, so that quite a few people have left an assessment of it. This research aims to help the Mie Gacoan Solo Veteran company in its efforts to understand consumer responses to their trademark and how people assess the products presented based on opinions and reviews on the Google Maps platform. Data was collected through a crawling process, then labeled and preprocessed before being extracted using the Term Frequency-Inverse Document Frequency (TF-IDF) technique and modeled with the Naive Bayes algorithm. The test results show that on unbalanced data, the model obtains an accuracy of 86%. However, after addressing data imbalance through an oversampling method using Synthetic Minority Over-sampling Technique (SMOTE), the model accuracy increased to 91%.*

**Keywords:** Sentiment Analysis, Python, TF-IDF, SMOTE, Mie Gacoan

### 1. PENDAHULUAN

Pesatnya kemajuan industri kuliner telah mendorong pertumbuhan yang signifikan dalam sektor restoran cepat saji[1]. Maraknya kehadiran usaha perkulininan melalui bermacam ciri khas maupun keunikannya. Bisnis kuliner dengan konsep makanan pedas telah menjadi tren yang sangat diminati oleh banyak orang dan dikunjungi di Indonesia, salah satunya adalah Mie Gacoan. Di antara banyak restoran yang ada, Mie Gacoan telah hadir sebagai salah satu restoran cepat saji berkonsep kekinian dan juga harganya yang sangat terjangkau membuat Mie Gacoan dengan mudah diterima dan sangat diminati di berbagai kalangan, terutama pada kalangan remaja. Merek dagang Mie Gacoan dimiliki oleh PT Pesta Pora Abadi, perusahaan induk dari bisnis mie pedas paling populer di Indonesia. Mie Gacoan telah ada sejak awal tahun 2016, dan dalam waktu tersebut telah berkembang menjadi pemain dominan di pasar, terutama di Pulau Bali, Jawa Barat, Jawa Tengah, dan Jawa Timur. Saat ini, Mie Gacoan benar-benar berkembang untuk menjadi merek teratas di negara ini[2]. Keberhasilan Mie Gacoan dalam menarik minat konsumen membuat restoran ini mendapatkan banyak perhatian dan ulasan dari

masyarakat, salah satunya yaitu cabang Mie Gacoan Solo Veteran, yang berlokasi di Jl. Veteran No.309, Tipes, Kecamatan Serengan, Kota Surakarta. Banyak konsumen yang meninggalkan ulasan pada platform Google Maps, memberikan penilaian berdasarkan pengalaman mereka secara langsung.

Google Maps merupakan aplikasi peta yang dimaksudkan untuk membantu orang menempuh perjalanan dengan memberikan petunjuk ke tempat tujuan mereka. Pada aplikasi Google Maps terdapat fitur Google Review yang dikembangkan oleh Google untuk memberikan penilaian atau ulasan terhadap suatu tempat dalam bentuk rating maupun kalimat[3]. Pada ulasan tersebut, tentunya opini yang muncul akan sangat beragam dan bervariasi. Setiap konsumen memiliki pandangan dan pengalaman yang berbeda, sehingga mengungkapkan opini mereka dengan cara yang unik. Hal ini mencerminkan beragam perspektif dan penilaian terhadap produk atau layanan yang diberikan, membuat analisis sentimen menjadi penting untuk memahami kecenderungan umum serta perasaan individu dari setiap ulasan yang ada.

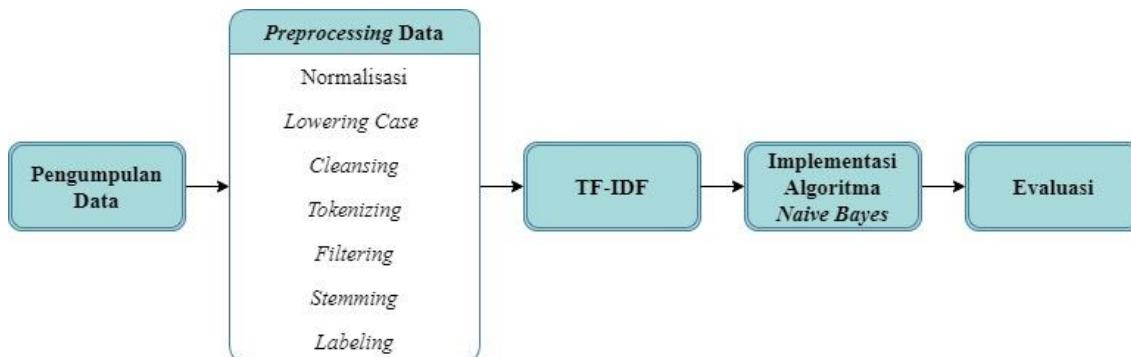
Analisis sentimen merupakan teknik yang secara otomatis mengkaji, mengekstrak, dan memproses data teksual guna menemukan data emosional dari ekspresi pemikiran(Gifari et al., 2022). Selain itu, analisis sentimen dapat memberikan wawasan yang lebih mendalam bagi perusahaan dalam memahami bagaimana produk dan layanan mereka diterima oleh masyarakat. Dengan mengkategorikan opini pelanggan menjadi sentimen positif, negatif, atau netral, perusahaan dapat memperoleh penggambaran dengan semakin mendalam terkait kelebihan serta kekurangannya. Perihal tersebut sangat penting dalam industri kuliner yang kompetitif, di mana ulasan dan penilaian pelanggan dapat mempengaruhi reputasi dan daya tarik restoran secara signifikan.

Mengacu pada latar belakang sebelumnya, penulis turut meneliti terkait sentimen dalam ulasan Mie Gacoan Solo Veteran di Google Maps. Penelitian ini bertujuan untuk membantu perusahaan Mie Gacoan Solo Veteran dalam upaya memahami tanggapan terhadap merek dagangnya juga bagaimana masyarakat dalam menilai suatu produk melalui opini juga ulasan terdapat pada Google Maps. Fokus utama penelitian ini adalah untuk memahami persepsi konsumen mengenai berbagai aspek dari restoran tersebut yang diulas oleh pelanggan di platform Google Maps. Penelitian ini tidak hanya fokus pada satu aspek, tetapi menggabungkan berbagai dimensi yang membentuk pengalaman keseluruhan pelanggan di Mie Gacoan Solo Veteran.

Berikut adalah beberapa penelitian terkait mengenai analisis sentimen yaitu Analisis Sentimen Ulasan Pelanggan Online Ubi Madu Cilembu Abah Nana Menggunakan Algoritma Naïve Bayes, hasil penelitian ini menunjukkan bahwa algoritma Naïve Bayes dapat digunakan untuk mengklasifikasikan sentimen ulasan pelanggan terhadap pelayanan dan produk Ubi Madu Cilembu Abah Nana dengan tingkat akurasi yang cukup tinggi, diperoleh hasil akurasi 86.29% pada eksperimen 1 yang menggunakan operator Split Data, dan akurasi 86.12% pada eksperimen 2 yang menggunakan operator Cross Validation dengan bantuan ahli bahasa(Al et al., 2023). Analisis Sentimen Opini Masyarakat Terhadap Kuliner Dki Jakarta Dengan Metode Naïve Baiyes Dan Support Vector Machine, nilai akurasi Naïve Bayes lebih tinggi daripada nilai akurasi Support Vector Machine untuk akurasi Naïve Bayes mendapatkan skor 76.00% dan akurasi dari support vector machine mendapatkan skor 74.00%(Sasongko et al., 2023).

## 2. METODE PENELITIAN

Berikut adalah gambaran mengenai tahapannya perihal diteliti dari penulis



Gambar 1. Tahapan Metode Penelitian

## 2.1 Pengumpulan Data

Dalam tahapan mengumpulkan data yang bersumber di Google Maps didapatkan dengan cara *crawling* data menggunakan web *scraping* melalui website Apify. Dataset yang digunakan yaitu berupa data rating dan teks ulasan pengunjung tentang pengalaman mereka setelah mengunjungi restoran Mie Gacoan Solo Veteran. Data yang awalnya berjumlah 1335 telah melalui proses normalisasi, yang akhirnya menghasilkan 582 data yang siap untuk dianalisis lebih lanjut.

## 2.2 Preprocessing Data

Untuk memudahkan proses analisis, kumpulan data ulasan harus melalui tahap persiapan sebelum memasuki proses klasifikasi. Tahap ini dirancang untuk mengubah data yang awalnya tidak terstruktur atau mengandung banyak noise menjadi data yang terstruktur(Husada & Paramita, 2021). Tahapan *preprocessing* ada beragam tergantung pada kebutuhan, namun umumnya meliputi, normalisasi, *lowering case*, *cleansing*, *tokenizing*, *filtering*, dan *stemming*(Hermawan & Bellaniar Ismiati, 2020).

Untuk membuat analisis lebih mudah, kata-kata yang mengandung ejaan dan slang diubah melalui proses normalisasi(Darwis et al., 2020). Kata slang dengan dimaknai termasuk frasa sulitnya dipahami, kata-kata yang tidak baku, atau frasa maknanya berbeda dari kata asli ataupun kata gaul yang digunakan anak remaja saat ini(Anggini et al., 2022). *Lowering case* merupakan proses pengubahan bentuk kalimat atau data menjadikan huruf kecil secara keseluruhan(Sumitro et al., 2021). *Cleansing* adalah proses pembersihan data teks yaitu dengan menghapus data yang tidak konsisten atau tidak relevan, *cleansing* dilakukan agar karakter yang bersifat noise dapat dihilangkan (Fathonah & Herliana, 2021). *Tokenizing* ialah suatu tahap pemotongan string kata yang didasarkan pada penyusunan kata tersebut(Syah & Witanti, 2022). *Filtering* juga dikenal sebagai *stopword*, adalah proses mengambil kata penting dari *tokenizing*. Terdapat algoritma *stoplist* yang digunakan untuk menghilangkan kata yang tidak penting, serta *wordlist* melalui digunakan untuk menyimpan kata yang penting. *Stemming* ialah proses pemetaan dan penguraian bentuk dari suatu kata menjadi bentuk dasarnya(Wardani & Cipta Nugraha, 2020).

## 2.3 TF-IDF

Selanjutnya, data yang sudah melalui Text *Preprocessing* akan dibagi kedalam data pelatihan dan data pengujian lalu dilakukan pembobotan atau *term weighting*, *term weighting* merupakan bentuk proses pembobotan tiap kata guna mengoptimalkan kemampuan analisis sentimen(Gifari et al., 2022). Penelitian ini memanfaatkan *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF diaplikasikan guna menghitung jumlah kemunculan kata pada dokumen lalu mengubah kata tersebut ke dalam bentuk angka(Aziz & Fauziah, 2022). Setelah proses pembobotan selesai, tahap pertama pengujian akan dilaksanakan menggunakan data yang tidak seimbang untuk mengevaluasi kinerja model pada kondisi asli. Pada tahap pengujian kedua, dilakukan penyeimbangan data dengan menerapkan berteknik oversampling menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE) guna meningkatkan representasi kelas minoritas dalam dataset.

## 2.4 Metode Naive Bayes

Dalam tahapan pengklasifikasian data menerapkan algoritma Naive Bayes. Naive Bayes adalah teknik pemrosesan teks yang digunakan untuk klasifikasi sederhana yang memiliki tingkat akurasi yang tinggi(Fauzan & Hikmah, 2022).

## 2.5 Evaluasi

Pada tahap evaluasi, perhitungan akurasi (*accuracy*), presisi (*precision*), *recall*, dan *F1-Score* dilakukan melalui menggunakan *confusion matrix* dalam ulasan yang telah diklasifikasikan menggunakan algoritma Naive Bayes. Tujuan dalam evaluasi guna menentukan kedalaman model yang telah dibangun dapat secara akurat mengklasifikasikan dokumen menjadi sentimen positif dan sentimen negatif. Dengan demikian, evaluasi ini memberikan gambaran menyeluruh mengenai performa dan efektivitas model dalam mengidentifikasi dan memisahkan berbagai jenis sentimen yang terdapat dalam ulasan.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengumpulan Data

Data ulasan Mie Gacoan Solo Veteran dikumpulkan menggunakan website Apify dari Google Maps. Setiap ulasan terdapat atribut *Stars* dan *Text* ulasan yang berjumlah 589 data setelah dilakukan normalisasi.

**Tabel 1.** Contoh Data Hasil *Scraping*

| Stars | Text                                                |
|-------|-----------------------------------------------------|
| 5     | suka dimsumnya                                      |
| 5     | Pelayanannya mantep dan efisien                     |
| 5     | Kasirnya baik ramah pelayanannya cepat              |
| 3     | Resto dg menu utama mie.tingkatkan LG pelayanan nya |
| 5     | Suasana nya enak buat nongkrong                     |
| 5     | tempatnya nyaman, makananya enak                    |

#### 3.2 Preprocessing Data

Tahap mendasar dari analisis sentimen adalah preprocessing, yang merupakan proses penting dalam penggalian data dan bermanfaat untuk meningkatkan kinerja. Preprocessing yang digunakan dalam penelitian ini meliputi hal-hal berikut Normalisasi, *Lowering Case*, *Cleansing*, *Tokenizing*, *Filtering* atau *Stopwords*, *Stemming*. Proses preprocessing ini penting karena membantu dalam menyiapkan data secara optimal untuk tahap analisis sentimen selanjutnya, yaitu ekstraksi fitur menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency) dan pemodelan menggunakan algoritma Naive Bayes.

| stars | text                                                       |
|-------|------------------------------------------------------------|
| 0     | 5 suka dimsumnya                                           |
| 1     | 5 Enak kalii looo miee gacoan ini juara lah pok...<br>...  |
| 2     | 5 Adakan free udang keju                                   |
| 3     | 5 NaN                                                      |
| 4     | 5 mie nya enak dan juga murah,tempat nya juga as...<br>... |
| 1331  | 5 NaN                                                      |
| 1332  | 5 tempatnya luas dan nyaman, pelayanan cepat               |
| 1333  | 5 NaN                                                      |
| 1334  | 5 NaN                                                      |
| 1335  | 5 Oke mie nya enak pangsit nya gurih                       |

1336 rows x 2 columns

**Gambar 2.** Sebelum Normalisasi

| text                                                       |
|------------------------------------------------------------|
| 0 suka dimsumnya                                           |
| 1 Enak kali lo mie gacoan ini juara lah pokonya ...<br>... |
| 2 Adakan gratis udang keju                                 |
| 4 mie nya enak dan juga murah,tempat nya juga as...<br>... |
| 5 Pelayanan bagus ramah cepat<br>...                       |
| 1327 Tempat nyaman, pelayanan ramah                        |
| 1328 Sangat bagus pelayanan ramah pokoknya the best        |
| 1330 Pelayanan cepat, rasa mantap sesuai SOC\nSayan...     |
| 1332 tempatnya luas dan nyaman, pelayanan cepat            |
| 1335 Oke mie nya enak pangsit nya gurih                    |

582 rows x 1 columns

**Gambar 3.** Normalisasi

Pada gambar 3 merupakan hasil dari normalisasi, dilakukan penghapusan data yang tidak lengkap untuk membantu menjaga konsistensi dan standar dalam proses analisis data. Gambar 4 menunjukkan hasil dari proses *lowering case* dan *cleansing*, pada proses tersebut mengubah kalimat menjadikannya huruf kecil serta menghilangkan karakter yang tidak diperlukan dalam analisis sentimen. Gambar 5 merupakan hasil dari proses *tokenizing*, Gambar 6 hasil dari proses *filtering*, dan gambar 7 adalah hasil dari proses *stemming*.

|      | clean_text                                        |
|------|---------------------------------------------------|
| 0    | suka dimsumnya                                    |
| 1    | enak kali lo mie ini juara lah pokonya tidak ...  |
| 2    | adakan gratis udang keju                          |
| 4    | mie nya enak dan juga murah tempat nya juga as... |
| 5    | pelayanan bagus ramah cepat                       |
| ...  | ...                                               |
| 1327 | tempat nyaman pelayanan ramah                     |
| 1328 | sangat bagus pelayanan ramah pokoknya the best    |
| 1330 | pelayanan cepat rasa mantap sesuai soc sayang...  |
| 1332 | tempatnya luas dan nyaman pelayanan cepat         |
| 1335 | oke mie nya enak pangsit nya gurih                |

582 rows × 1 columns

**Gambar 4. Lowering Case & Cleansing**

|      | tokens                                             |
|------|----------------------------------------------------|
| 0    | [suka, dimsumnya]                                  |
| 1    | [enak, kali, lo, mie, ini, juara, lah, pokonya...] |
| 2    | [adakan, gratis, udang, keju]                      |
| 4    | [mie, nya, enak, dan, juga, murah, tempat, nya...] |
| 5    | [pelayanan, bagus, ramah, cepat]                   |
| ...  | ...                                                |
| 1327 | [tempat, nyaman, pelayanan, ramah]                 |
| 1328 | [sangat, bagus, pelayanan, ramah, pokoknya, th...] |
| 1330 | [pelayanan, cepat, rasa, mantap, sesuai, soc, ...] |
| 1332 | [tempatnya, luas, dan, nyaman, pelayanan, cepat]   |
| 1335 | [oke, mie, nya, enak, pangsit, nya, gurih]         |

582 rows × 1 columns

**Gambar 5. Tokenizing**

|      | filtered_tokens                                    |
|------|----------------------------------------------------|
| 0    | [suka, dimsumnya]                                  |
| 1    | [enak, kali, lo, mie, juara, pokonya, bosan, k...] |
| 2    | [adakan, gratis, udang, keju]                      |
| 4    | [mie, nya, enak, murah, nya, asik, nyaman]         |
| 5    | [pelayanan, bagus, ramah, cepat]                   |
| ...  | ...                                                |
| 1327 | [nyaman, pelayanan, ramah]                         |
| 1328 | [bagus, pelayanan, ramah, pokoknya, the, best]     |
| 1330 | [pelayanan, cepat, mantap, sesuai, soc, sayang...] |
| 1332 | [tempatnya, luas, nyaman, pelayanan, cepat]        |
| 1335 | [oke, mie, nya, enak, pangsit, nya, gurih]         |

582 rows × 1 columns

**Gambar 6. Filtering**

|      | stemmed_tokens                                     |
|------|----------------------------------------------------|
| 0    | [suka, dimsumnya]                                  |
| 1    | [enak, kali, lo, mie, juara, poko, bosan, kalo...] |
| 2    | [adakan, gratis, udang, keju]                      |
| 4    | [mie, nya, enak, murah, nya, asik, nyaman]         |
| 5    | [layan, bagus, ramah, cepat]                       |
| ...  | ...                                                |
| 1327 | [nyaman, layan, ramah]                             |
| 1328 | [bagus, layan, ramah, pokok, the, best]            |
| 1330 | [layan, cepat, mantap, sesuai, soc, sayang, te...] |
| 1332 | [tempat, luas, nyaman, layan, cepat]               |
| 1335 | [oke, mie, nya, enak, pangsit, nya, gurih]         |

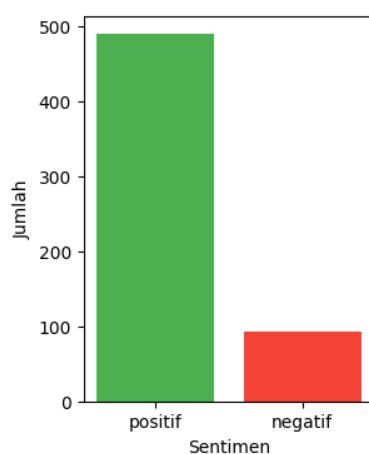
582 rows × 1 columns

**Gambar 7. Stemming**

Setelah proses *stemming* selesai, tahap selanjutnya adalah melakukan pelabelan pada data. Dalam proses pelabelan secara otomatis ini dilakukan dengan menggunakan bahasa pemrograman python, data *rating* akan digunakan untuk menentukan hasil sentimen positif, dan negatif. Pembagian kelas sentimen berdasarkan rating dibagi menjadi 2 yaitu untuk ulasan rating 4 dan 5 termasuk dalam kelas positif, untuk rating 1 dan 2 termasuk kelas negatif.

|      |                                                  | stemmed_text                     | sentimen | stars |
|------|--------------------------------------------------|----------------------------------|----------|-------|
| 0    |                                                  | suka dimsumnya                   | positif  | 5     |
| 1    | enak kali lo mie juara poko bosan kalo makan mie |                                  | positif  | 5     |
| 2    |                                                  | adakan fre udang keju            | positif  | 5     |
| 4    | mie nya enak murah nya asik nyaman               |                                  | positif  | 5     |
| 5    |                                                  | layan bagus ramah cepat          | positif  | 5     |
| ...  |                                                  | ...                              | ...      | ...   |
| 1327 |                                                  | nyaman layan ramah               | positif  | 5     |
| 1328 |                                                  | bagus layan ramah pokok the best | positif  | 5     |
| 1330 | layan cepat mantap sesuai soc sayang tempat      |                                  | positif  | 5     |
| 1332 |                                                  | tempat luas nyaman layan cepat   | positif  | 5     |
| 1335 | oke mie nya enak pangsit nya gurih               |                                  | positif  | 5     |

582 rows × 3 columns

**Gambar 8.** Hasil Labeling**Gambar 9.** Diagram Hasil Labeling

Gambar 9 menunjukkan diagram hasil labelling bersentimen positif semakin dominan berjumlah 84.0 % ataupun 489 data. Disusulkan dari bersentimen negatif dengan berjumlah 16.0% atau 93 data.

### 3.3 TF-IDF

Proses TF-IDF menggunakan *Tf idf Vectorizer* dari scikit-learn untuk mengubah teks menjadi representasi numerik yang didasarkan pada bobot kata yang dihitung menggunakan formula TF-IDF(Br Sinulingga & Sitorus, 2024). Dalam pembagian data penulis menggunakan data dengan perbandingan 70:30 sebagai data *training* dan *testing*. Berikut adalah *Source Code* yang digunakan.

```
from sklearn.model_selection import train_test_split # Import train_test_split untuk membagi data menjadi set pelatihan dan set pengujian
from sklearn.feature_extraction.text import TfidfVectorizer # Import TfidfVectorizer untuk mengubah teks menjadi vektor fitur TF-IDF

# Pisahkan fitur dan label
X = data['stemmed_text'] # sebagai fitur
y = data['sentimen'] # sebagai label

# Bagi data menjadi set pelatihan dan set pengujian
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Mengubah teks menjadi vektor fitur menggunakan TF-IDF
vectorizer = TfidfVectorizer()
X_train_vect = vectorizer.fit_transform(X_train)
X_test_vect = vectorizer.transform(X_test)

# Menampilkan jumlah ulasan dan jumlah kata unik (fitur)
print('{} Number of reviews has {} words'.format(text_tf_idf.shape[0], text_tf_idf.shape[1]))
```

582 Number of reviews has 1331 words

**Gambar 10.** Source Code TF-IDF

### 3.4 Implementasi Algoritma Naive Bayes

```
[58] # Membangun model Naive Bayes
      clf = MultinomialNB()
      clf.fit(X_train_vect, y_train)

      # Prediksi label untuk data pengujian
      predicted = clf.predict(X_test_vect)
```

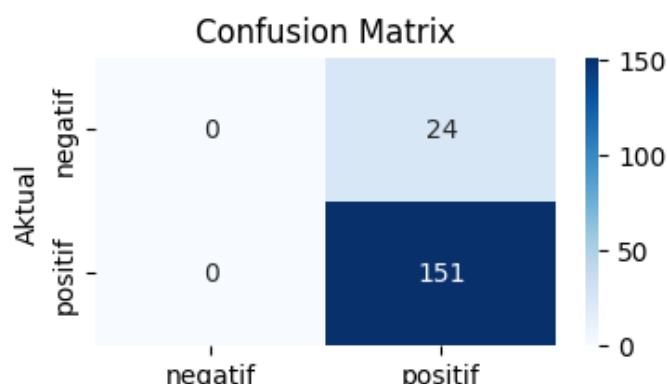
Gambar 11. Source Code Naive Bayes

### 3.5 Evaluasi

Setelah dilakukan analisis sentimen, akan dilakukan evaluasi menggunakan metode confusion matrix untuk melihat apakah model yang diaplikasikan mempunyai akurasi yang optimal. Selama proses evaluasi, Naive Bayes mengukur *Accuracy*, *precision*, *recall*, dan *F1-score*.

```
- Accuracy: 0.8628571428571429
Precision: 0.7445224489795919
Recall: 0.8628571428571429
F1 Score: 0.7993339176161262
Confusion Matrix:
[[ 0 24]
 [ 0 151]]
Classification Report:
              precision    recall   f1-score   support
negatif       0.00     0.00     0.00      24
positif       0.86     1.00     0.93     151
accuracy          -         -        -      175
macro avg       0.43     0.50     0.46     175
weighted avg    0.74     0.86     0.80     175
```

Gambar 12. Akurasi Algoritma Naive Bayes Tanpa SMOTE



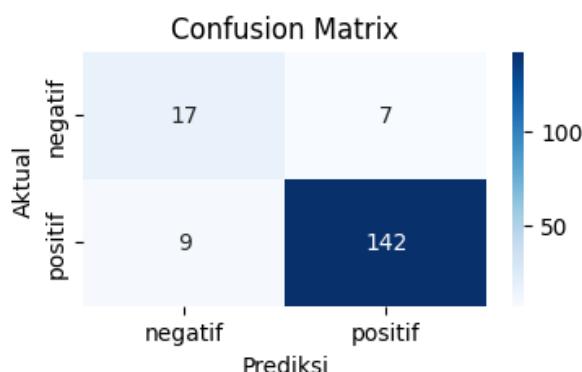
Gambar 13. Evaluasi Confusion Matrix tanpa SMOTE

Pada gambar 12 juga gambar 13 merupakan gambar dari akurasi data *imbalance* menggunakan Naive Bayes dan pengujian model Naive Bayes pada data *imbalance* menggunakan *confusion matrix*. Nilai akurasi pada Naive Bayes memberikan nilai sebesar 86%.

```

    ➔ Accuracy: 0.9085714285714286
    Precision: 0.9119905597757946
    Recall: 0.9085714285714286
    F1 Score: 0.910095238095238
    Confusion Matrix:
    [[ 17  7]
     [ 9 142]]
    Classification Report:
               precision    recall   f1-score  support
    negatif        0.65      0.71      0.68      24
    positif        0.95      0.94      0.95     151
    accuracy       0.89      0.91      0.90     175
    macro avg       0.80      0.82      0.81     175
    weighted avg    0.91      0.91      0.91     175

```

**Gambar 14.** Akurasi Algoritma Naive Bayes dengan SMOTE**Gambar 15.** Evaluasi Confusion Matrix dengan SMOTE

Pada gambar 14 merupakan gambar dari akurasi data yang sudah dilakukan *oversampling* menggunakan metode SMOTE menggunakan Naive Bayes. Pada gambar 15 merupakan pengujian model Naive Bayes menggunakan *confusion matrix* pada data data yang sudah dilakukan *oversampling* menggunakan metode SMOTE. Nilai akurasi pada Naive Bayes memberikan nilai sebesar 91%. Berikut hasil perbandingan nilai akurasi dari data *imbalance* dan data yang dilakukan *oversampling*.

**Tabel 2.** Tabel Nilai Akurasi Tanpa SMOTE

| No | Label   | Accuracy | Precision | Recall |
|----|---------|----------|-----------|--------|
| 1  | Positif | 86%      | 86%       | 100%   |
| 2  | Negatif |          | 0%        | 0%     |

**Tabel 3.** Tabel Nilai Akurasi dengan SMOTE

| No | Label   | Accuracy | Precision | Recall |
|----|---------|----------|-----------|--------|
| 1  | Positif | 91%      | 95%       | 94%    |
| 2  | Negatif |          | 65%       | 71%    |

Berdasarkan kedua proses yang telah dilakukan sebelumnya, pada proses kedua, yaitu data yang dilakukan *oversampling* menghasilkan *accuracy* yang paling besar yaitu 91% dibandingkan dengan data *imbalance*. Berikut *word cloud* yang dihasilkan dari analisis sentimen ini.



**Gambar 16. Word Cloud**

### **3.6 Pembahasan**

Pada penelitian ini, untuk menganalisis sentimen data ulasan Mie Gacoan Solo Veteran yang diambil dari Google Maps, menggunakan algoritma Naive Bayes. Data ulasan yang dikumpulkan awalnya menunjukkan ketidakseimbangan antara jumlah ulasan positif dan negatif, yang dapat mempengaruhi performa model dalam memprediksi sentimen dengan akurat. Untuk mengatasi masalah ketidakseimbangan data ini, diterapkan metode oversampling SMOTE (*Synthetic Minority Over-sampling Technique*). Hasil menunjukkan bahwa algoritma Naive Bayes yang menggunakan metode oversampling SMOTE memberikan akurasi yang lebih optimal. Hal ini karena Metode SMOTE dapat digunakan untuk menyeimbangkan data yang sebelumnya *imbalance*(Arifiyanti & Wahyuni, 2020). Metode SMOTE digunakan dalam penelitian ini menunjukkan bahwa penyeimbangan data sangat penting untuk analisis sentimen, terutama dalam kasus di mana ada ketidakseimbangan yang signifikan dalam data asli. Dengan data yang lebih seimbang, model dapat menemukan pola dengan lebih mudah dan menghasilkan hasil klasifikasi yang lebih akurat.

#### **4. KESIMPULAN**

Mengacu temuan pengujian dan perihal penelitian ini maka analisis sentimen dalam ulasan Mie Gacoan Solo Veteran di Google Maps dengan melakukan pengujian sebanyak dua kali yaitu pada data *imbalance* dan pada metode *oversampling* SMOTE, mendapatkan sebuah konklusi jika data yang tidak seimbang dapat mempengaruhi akurasi yang di dapatkan. Data *imbalance* pada pengujian pertama mendapatkan nilai akurasi sebesar 86% dengan *Precession* positif 86%, negatif 0%, *Recall* positif 86%, negatif 0%. Pada pengujian kedua dengan menggunakan data *imbalance* yang sama namun digunakan metode *oversampling* SMOTE dihasilkan rata-rata akurasi yang didapatkan lebih tinggi yaitu sebesar 91% *Precession* positif 95%, negatif 65%, *Recall* positif 94%, negatif 71%.

## **Daftar Pustaka**

- [1] N. G. Woen and S. Santoso, "Pengaruh Kualitas Layanan, Kualitas Produk, Promosi, dan Harga Normal terhadap Kepuasan dan Loyalitas Konsumen," *Jurnal Maksipreneur: Manajemen, Koperasi, dan Entrepreneurship*, vol. 10, no. 2, p. 146, Feb. 2021, doi: 10.30588/jmp.v10i2.712.
  - [2] G. A. Prasasti and P. Maisara, "PENGARUH FASILITAS, HARGA DAN CITA RASA TERHADAP KEPUASAN KONSUMEN MIE GACOAN DI SOLO RAYA," *Publik: Jurnal Manajemen Sumber Daya Manusia, Administrasi dan Pelayanan Publik*, vol. 9, no. 2, pp. 276–288, May 2022, doi: 10.37606/publik.v9i2.327.
  - [3] F. U. Haq, " PENGGUNAAN GOOGLE REVIEW SEBAGAI PENILAIAN KEPUASAN PENGUNJUNG DALAM PARIWISATA," *Tornare*, vol. 2, no. 1, p. 10, Jan. 2020, doi: 10.24198/tornare.v2i1.25826.
  - [4] O. I. Gifari, Muh. Adha, F. Freddy, and F. F. S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *Journal of Information Technology*, vol. 2, no. 1, pp. 36–40, Mar. 2022, doi: 10.46229/jifotech.v2i1.330.
  - [5] M. R. Al, F. Zain, and M. Kamayani, "Analisis Sentimen Ulasan Pelanggan Online Ubi Madu Cilembu Abah Nana Menggunakan Algoritma Naïve Bayes," *Jurnal Sistem Komputer dan Informatika (JSON) Hal: 11-*, vol. 21, no. 1, 2023, doi: 10.30865/json.v5i1.6646.

- [6] M. F. Sasongko, F. Fauziah, and N. Hayati, "Analisis Sentimen Opini Masyarakat Terhadap Kuliner DKI Jakarta dengan Metode Naïve Bayes dan Support Vector Machine," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 7, no. 3, p. 241, Apr. 2023, doi: 10.30998/string.v7i3.13931.
- [7] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, pp. 18–26, Feb. 2021, doi: 10.34148/teknika.v10i1.311.
- [8] L. Hermawan and M. Bellaniar Ismiati, "Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval," *Jurnal Transformatika*, vol. 17, no. 2, p. 188, Jan. 2020, doi: 10.26623/transformatika.v17i2.1705.
- [9] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA," *Edutic - Scientific Journal of Informatics Education*, vol. 7, no. 1, Nov. 2020, doi: 10.21107/edutic.v7i1.8779.
- [10] N. Anggini, N. Y. Afifah, and E. Syaputra, "Pengaruh Bahasa Gaul (SLANG) Terhadap Bahasa Indonesia Pada Generasi Muda," *Jurnal Multidisiplin Dehasen (MUDE)*, vol. 1, no. 3, Jun. 2022, doi: 10.37676/mude.v1i3.2477.
- [11] P. A. Sumitro, Rasiban, D. I. Mulyana, and W. Saputro, "Analisis Sentimen Terhadap Vaksin Covid-19 di Indonesia pada Twitter Menggunakan Metode Lexicon Based," *J-ICOM - Jurnal Informatika dan Teknologi Komputer*, vol. 2, no. 2, pp. 50–56, Oct. 2021, doi: 10.33059/jicom.v2i2.4009.
- [12] F. Fathonah and A. Herliana, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes," *Jurnal Sains dan Informatika*, vol. 7, no. 2, pp. 155–164, Dec. 2021, doi: 10.34128/jsi.v7i2.331.
- [13] H. Syah and A. Witanti, "ANALISIS SENTIMEN MASYARAKAT TERHADAP VAKSINASI COVID-19 PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 5, no. 1, pp. 59–67, Feb. 2022, doi: 10.47080/simika.v5i1.1411.
- [14] N. W. Wardani and P. G. S. Cipta Nugraha, "STEMMING DOKUMEN TEKS BAHASA BALI DENGAN METODE RULE BASE APPROACH," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 7, no. 3, pp. 510–521, Dec. 2020, doi: 10.35957/jatisi.v7i3.538.
- [15] A. Aziz and Fauziah, "Analisis Sentimen Identifikasi Opini Terhadap Produk, Layanan dan Kebijakan Perusahaan Menggunakan Algoritma TF-IDF dan SentiStrength," Abdul Aziz, 2022.
- [16] Abd. C. Fauzan and K. Hikmah, "IMPLEMENTASI ALGORITMA NAIVE BAYES DALAM ANALISIS POLARISASI OPINI MASYARAKAT TERKAIT VAKSIN COVID-19," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 7, no. 2, pp. 122–128, Jul. 2022, doi: 10.36341/rabit.v7i2.2403.
- [17] J. E. Br Sinulingga and H. C. K. Sitorus, "Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 14, no. 1, pp. 42–53, Feb. 2024, doi: 10.34010/jamika.v14i1.11946.
- [18] A. A. Arifiyanti and E. D. Wahyuni, "SMOTE: METODE PENYEIMBANG KELAS PADA KLASIFIKASI DATA MINING," *SCAN - Jurnal Teknologi Informasi dan Komunikasi*, vol. 15, no. 1, Feb. 2020, doi: 10.33005/scan.v15i1.1850.

