



Implementasi *Naive Bayes Classifier* Dalam Menganalisis Sentimen Pelanggan Mie Gacoan Pada Instagram

Audina Tazkia*¹, Yuni Arkhiansyah²

*^{1,2} Jurusan Teknik Informatika, Institut Informatika dan Bisnis Darmajaya, Jl.Z.A Pagar Alam No.93, Labuhan Ratu, Bandar Lampung 35142, Indonesia

Email Penulis Korespondensi: ¹ audina.2011010090@mail.darmajaya.ac.id

Abstrak

Media sosial telah menjadi bagian umum dalam kehidupan masyarakat pada era digital saat ini. Salah satu platform media sosial yang sangat populer adalah Instagram, di mana pengguna dapat memberikan komentar terhadap foto atau video yang dibagikan. Komentar-komentar ini mengandung opini dan permasalahan yang dapat diolah menjadi analisis sentimen. Mie Gacoan, sebagai restoran mie pedas yang populer di Indonesia, menerima penilaian baik dan buruk dari pelanggan. Dalam penelitian ini, penulis bertujuan untuk memberikan solusi dengan menghasilkan analisis sentimen terhadap Mie Gacoan di Instagram menggunakan metode *Naive Bayes Classifier*. Metode ini digunakan untuk mengklasifikasikan komentar-komentar pengguna menjadi kategori positif atau negatif. Dengan menganalisis sentimen masyarakat terhadap Mie Gacoan di Instagram, diharapkan hasil penelitian ini dapat memberikan wawasan yang berguna bagi restoran tersebut untuk meningkatkan kualitas produk dan pelayanan mereka. Hasilnya juga dapat menjadi referensi bagi pemilik bisnis lainnya dalam mengoptimalkan media sosial sebagai platform interaksi dengan pelanggan. Pada penelitian ini ekstraksi fitur dilakukan dengan menerapkan metode TF-IDF. Kemudian data dibagi menjadi 80% data latih dan 20% data uji. Hasil penelitian ini menunjukkan tingkat akurasi sebesar 90,59%, precision 87,50%, recall 95,45%, dan f1-score 91,30%.

Kata Kunci—Analisis sentiment, *Naive Bayes Classifier*, Mie Gacoan, Instagram.

Abstract

Social media has become a common part of people's lives in today's digital age. One of the most popular social media platforms is Instagram, where users can comment on shared photos or videos. These comments contain opinions and issues that can be processed into sentiment analysis. Mie Gacoan, as a popular spicy noodle restaurant in Indonesia, receives both good and bad ratings from customers. In this research, the author aims to provide a solution by generating a sentiment analysis of Mie Gacoan on Instagram using the *Naive Bayes Classifier* method. This method is used to classify user comments into positive or negative categories. By analyzing people's sentiments towards Mie Gacoan on Instagram, it is hoped that the results of this study can provide useful insights for the restaurant to improve the quality of their products and services. The results can also be a reference for other business owners in optimizing social media as a platform for interaction with customers. In this research, feature extraction was carried out using the TF-IDF method. Then the data is divided into 80% training data and 20% test data. The

results of this research show an accuracy level of 90.59%, precision 87.50%, recall 95.45%, and f1-score 91.30%.

Keywords—*Sentiment analysis, Naïve Bayes Classifier, Mie Gacoan, Instagram.*

1. PENDAHULUAN

Pada era digital saat ini, media sosial telah menjadi sesuatu yang umum dan telah banyak digunakan untuk memenuhi kebutuhan masyarakat. Salah satu platform media sosial yang banyak digunakan adalah Instagram. Instagram telah menjadi salah satu media sosial yang saat ini mengubah menjadi sebuah tren gaya hidup dikalangan masyarakat, khususnya di kalangan generasi muda yang secara aktif menggunakan media sosial dalam kehidupan sehari-hari mereka. Instagram merupakan sebuah aplikasi yang memungkinkan pengguna untuk mengambil foto dan video, menerapkan filter digital, dan membagikannya. Selain itu, pengguna lain juga dapat memberikan komentar sebagai bentuk pengakuan atau penilaian terhadap foto atau video yang dibagikan.

Mie Gacoan pertama kali didirikan pada tahun 2016 di Malang, oleh Anton Kurniawan. Mie Gacoan merupakan salah satu restoran mie pedas yang populer di Indonesia selama beberapa tahun belakangan ini. Restoran yang menjual mie sebagai menu utamanya ini sangat diminati oleh banyak orang dan menjadi restoran yang paling populer jika dibandingkan dengan restoran-restoran lainnya. Pada saat ini makanan dengan tingkat kepedasan yang tinggi telah menjadi salah satu trend yang sangat populer akhir-akhir ini. Hal inilah yang Mie Gacoan coba tawarkan melalui berbagai variasi mie yang mereka sediakan. Pelanggan utama restoran ini adalah orang-orang yang menyukai mie dan makanan pedas. Dapat dibayangkan bahwa nantinya restoran ini akan menjadi salah satu tempat favorit anak muda, khususnya penyuka makanan pedas. Dibalik kepopuleran Mie Gacoan pada saat ini, tentunya tidak dapat dipungkiri bahwasannya ada penilaian positif dan penilaian negatif dari pelanggan. Banyak dari pelanggan Mie Gacoan mengutarakan komentar mereka mengenai kualitas pelayanan Mie Gacoan di Instagram. Komentar tersebut dimanfaatkan untuk melakukan analisis sentiment guna mengetahui kecenderungan komentar pelanggan terhadap pelayanan Mie Gacoan, apakah komentar pada Instagram cenderung positif atau negatif.

Kemudian opini dan permasalahan pada komentar tersebut diolah menjadi sebuah informasi melalui suatu cara yaitu analisis sentimen. Analisis sentimen adalah bidang studi yang menganalisa pendapat, sentimen, penilaian, evaluasi, emosi, serta sikap seseorang terhadap produk, peristiwa, isu, topik, individu dan organisasi [1]. Dalam melakukan analisis sentimen menggunakan data komentar Instagram terdapat permasalahan yang dihadapi yaitu penulisan kata yang disingkat dan penggunaan bahasa gaul dalam berkomentar. Oleh karena itu, dilakukan *preprocessing* terhadap data Instagram untuk menghasilkan data yang lebih baik. Penelitian ini bertujuan untuk mengukur analisis sentimen pelanggan pengguna Instagram terhadap Mie Gacoan di Indonesia.

Peneliti ingin memberikan solusi dengan cara menganalisis sentimen pelanggan terhadap Mie Gacoan di Instagram dengan memanfaatkan metode *Naive Bayes Classifier*. *Naive Bayes Classifier* digunakan untuk mengklasifikasikan sentimen pelanggan dan juga untuk mendapatkan hasil pengujian dengan nilai akurasi yang tinggi. *Naive Bayes* adalah algoritma yang digunakan untuk mengklasifikasikan suatu variable tertentu dengan memanfaatkan metode probabilitas dan statistic [2]. Studi literatur mengenai pembahasan analisis sentimen yang terkait dengan topik dalam penelitian ini yaitu penelitian dengan judul “Penerapan Analisis Sentimen Melalui Data Instagram Untuk Mengetahui Reputasi Wisata Kuliner Di Kota Bandung [3]”. Berdasarkan penelitian dengan 864 dataset komentar dan postingan, Algoritma *Naive Bayes* menunjukkan hasil akurasi yang lebih tinggi daripada algoritma *Decision Tree* dan K-NN, dengan tingkat akurasi mencapai 86.87% yang telah diuji menggunakan perangkat RapidMiner.

Penelitian yang dilakukan dengan judul “Perbandingan Metode *Naive Bayes* dan *Support Vector Machine* pada Analisis Sentimen Twitter [4]”. Berdasarkan penelitian tersebut Algoritma

Naïve Bayes memiliki tingkat akurasi yang lebih tinggi, mencapai 73,65%, dibandingkan dengan SVM yang memiliki tingkat akurasi sebesar 70,20%.

Penelitian yang dilakukan dengan judul “*Comparison Naïve Bayes Classifier, K-Nearest Neighbor And Support Vector Machine In The Classification Of Individual On Twitter Account* [2]”. Berdasarkan penelitian yang menggunakan dataset 130 akun *twitter*, didapatkan bahwa hasil evaluasi perhitungan dengan menggunakan *confusion matrix* menunjukkan bahwa persentase akurasi dari perbandingan antara hasil klasifikasi psikolog secara manual dan hasil klasifikasi sistem adalah 31.5% untuk algoritma *Naïve Bayes Classifier*, 23.8% untuk algoritma *K-Nearest Neighbor*, dan 28.5% untuk algoritma *Support Vector Machine*.

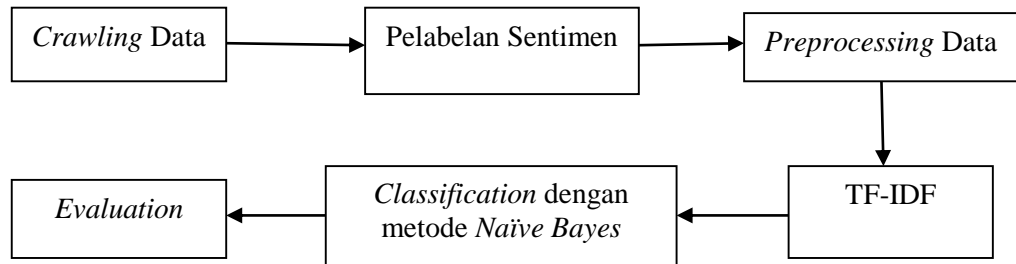
Penelitian yang dilakukan dengan judul “*Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN* [5]”. Berdasarkan penelitian tersebut algoritma *k-nearest neighbors*, dengan pembagian data pada rasio 60:40 terhadap dataset ulasan pengguna di aplikasi Bibit dan Bareksa, menghasilkan nilai akurasi, presisi, dan recall sebesar 85,14%, 91,91%, dan 76,44% untuk Bibit. Sedangkan untuk Bareksa, nilai tersebut adalah 81,70%, 87,15%, dan 75,73%.

Penelitian yang dilakukan dengan judul “*Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19* [6]”. Berdasarkan penelitian yang menggunakan data *tweet* yang berkaitan dengan COVID-19, algoritma Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19).

2. METODE PENELITIAN

2.1. Kerangka Pemikiran

Kerangka pemikiran merupakan konsep dasar penelitian yang didapatkan dari fakta-fakta, observasi, dan tinjauan kepustakaan [7]. Rancangan Penelitian yang akan dilaksanakan disajikan ke dalam Gambar 2.1.



Gambar 2.1. Kerangka Pemikiran

Keterangan:

1. *Crawling Data*

Pada tahap ini akan dilakukan penarikan dataset komentar dengan cara *crawling* menggunakan URL postingan Instagram @mie.gacoan untuk menarik data.

2. Pelabelan Sentimen

Pada tahap ini dataset yang telah diambil akan dilakukan proses pelabelan secara manual dan akan diberi label positif dan negatif.

3. *Preprocessing Data*

Pada tahap ini data akan diolah dengan *Preprocessing* berupa *Cleaning*, *Case Folding*, *Tokenizing*, *Normalisasi*, *Stopword Removal*, dan *Stemming*. Hal tersebut bertujuan untuk memperbaiki kualitas data tersebut, menghilangkan permasalahan yang bisa saja terjadi pada saat pemrosesan data, dan membuat dataset menjadi lebih efektif dan efisien.

4. TF-IDF

Pada tahap TF-IDF dilakukan pembobotan terhadap kata-kata dalam komentar untuk mengetahui bobot dari setiap kata tersebut.

5. Classification

Pada tahap *Classification* akan dilakukan pengklasifikasian data dengan menggunakan algoritma *Naïve Bayes*.

6. Evaluation

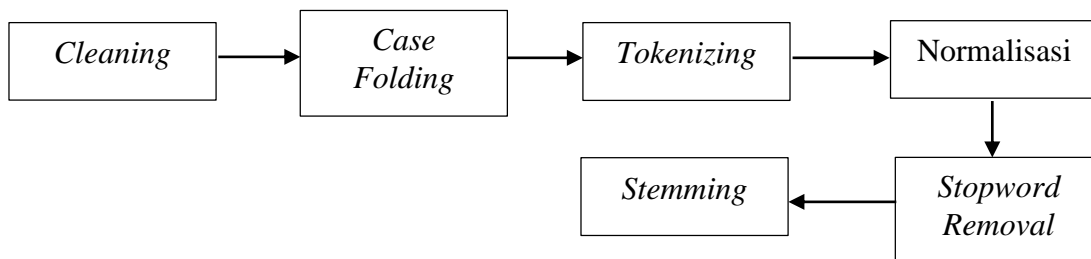
Pada tahap *Evaluation* dilakukan perhitungan *accuracy*, *precision*, *recall*, dan *F1-Score* dengan *confusion matrix* pada komentar yang telah diklasifikasikan dengan algoritma *Naïve Bayes*. Hal ini bertujuan untuk mengetahui keakuratan dokumen setelah diklasifikasi menjadi sentiment positif dan sentiment negatif.

2.2. Metode Pengumpulan Data

Pada penelitian ini pengumpulan data mengenai sentiment pelanggan terhadap pelayanan Mie Gacoan di Instagram yang diperoleh dengan cara *crawling* menggunakan tools *Export Comment for IG*. Data yang diambil yaitu sebanyak 1156 data dengan rentang waktu November 2023 – Januari 2024.

2.3. Praproses Data

Praproses data adalah langkah dimana teks diseragamkan dalam bentuk dan formatnya sehingga dapat dipersiapkan untuk diproses pada langkah berikutnya. Tahapan praproses yang digunakan diilustrasikan dalam Gambar 2.2.



Gambar 2.2. Tahapan *Text Preprocessing*

Keterangan:

1. *Cleaning*

Cleaning adalah tahapan menghilangkan atribut yang ada pada data seperti simbol, tanda baca, *hashtag*, URL, angka dan *noise* lainnya. Sehingga data menjadi lebih efisien dan dapat diolah dengan baik.

2. *Case Folding*

Case Folding adalah tahapan dimana semua huruf kapital pada teks komentar akan diubah ke dalam huruf kecil. Hal ini bertujuan agar dokumen pada teks komentar memiliki bentuk yang standar.

3. *Tokenizing*

Tokenizing adalah tahapan dimana kalimat akan dipecah menjadi unit-unit terkecil atau token. Tahap ini bertujuan untuk memudahkan pemrosesan dan analisis data.

4. Normalisasi

Normalisasi adalah tahapan penormalan bentuk kata yang kurang normal seperti singkatan, bahasa asing, dan kata yang kurang baku menjadi baku.

5. *Stopword Removal*

Stopword Removal adalah tahapan menghapus atau menghilangkan kata-kata yang yang tidak bermakna dalam dataset dan tidak memiliki pengaruh terhadap analisis sentimen.

6. *Stemming*

Stemming adalah tahapan mengembalikan suatu kata ke dalam bentuk kata dasar atau menghilangkan imbuhan (awalan dan akhiran) pada kata tersebut.

2.4. TF-IDF

Pada tahap ini akan dilakukan pemberian bobot pada setiap kata dengan menggunakan metode TF-IDF. TF-IDF merupakan algoritma yang memberikan nilai pada bobot pada kata-kata dalam teks. Hal itu dapat dilihat pada persamaan 1:

$$TF - IDF(d,t) = TD(d,t) \times IDF(t) \quad (1)$$

Pada persamaan (1), TF merupakan frekuensi suatu kata dalam dokumen, IDF merupakan nilai invers dari dokumen yang mengandung kata tersebut, t merupakan kata, dan d merupakan dokumen.

2.5. Classification dengan metode Naïve Bayes

Naïve Bayes Classifier adalah suatu algoritma klasifikasi yang diterapkan dalam *machine learning*. Secara umum teorema Bayes dinotasikan pada persamaan 2:

$$P(A|B) = \frac{P(A) P(B|A)}{P(B)} \quad (2)$$

Keterangan:

B = *Class* data yang dimasukkan.

A = Data hipotesis.

P(A|B) = Probabilitas hipotesis A yang mengacu pada kondisi B.

P(A) = Probabilitas hipotesis A.

P(B|A) = Probabilitas B berdasarkan kondisi pada hipotesis A.

P(B) = Probabilitas B

2.6. Rancangan Pengujian Model

Pada penelitian ini menggunakan proses evaluasi dengan *confusion matrix*. *Confusion matrix* merupakan alat yang berguna untuk mengukur kinerja suatu metode klasifikasi. Berikut ini merupakan contoh *confusion matrix* pada Tabel 2.1.

Tabel 2.1. *Confusion Matrix*

| Fakta | Prediksi | |
|---------|-----------------------|-----------------------|
| | Negatif | Positif |
| Negatif | <i>True Negative</i> | <i>False Positive</i> |
| Positif | <i>False Negative</i> | <i>True Positive</i> |

Untuk mengukur performa klasifikasi *Naïve Bayes*, metode yang diterapkan melibatkan perhitungan akurasi (*accuracy*), presisi (*precision*), *recall*, dan *f1-score* dengan persamaan sebagai berikut ini:

- a. Akurasi (*accuracy*), yaitu jumlah prediksi yang benar pada dataset yang diperoleh dari tabel *confusion matrix*. Sehingga ditentukan pada persamaan (3):

$$\text{Akurasi (accuracy)} = \frac{TN+TP}{TP+TN+FN+FP} \times 100\% \quad (3)$$

- b. Presisi (*precision*), yaitu rasio prediksi antara jumlah positif yang benar dan salah. Sehingga dapat dihitung menggunakan persamaan (4):

$$\text{Presisi (precision)} = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

- c. *Recall*, yaitu rasio prediksi dari jumlah positif benar dan jumlah positif yang benar ditambah dengan jumlah negatif yang salah. Sehingga dapat dihitung melalui persamaan (5):

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

- d. *F1 Score*, yaitu usaha untuk menyeimbangkan nilai presisi dan *recall*. Sehingga dapat dihitung melalui persamaan (6):

$$F1 \text{ Score} = 2 \times \frac{(\text{Recall} \times \text{Presisi})}{(\text{Recall} + \text{Presisi})} \quad (6)$$

3. HASIL DAN PEMBAHASAN

3.1. Dataset

Setelah melakukan *crawling* data pada Instagram @mie.gacoan dalam periode November 2023 – Januari 2024 menggunakan *tools Export Comment for IG*. Data yang terkumpul terdiri dari komentar-komentar pelanggan yang kemudian disusun menjadi sebuah dataset, dengan total data sebanyak 1156. Beberapa *comment* yang dikumpulkan ditampilkan dalam Gambar 3.1. Kemudian setelah proses pelabelan data, jumlah data yang digunakan dalam penelitian adalah sebanyak 424 data. Hal tersebut disebabkan karena penelitian ini hanya menggunakan *comment* dengan nilai sentimen positif dan negatif.

| User Id | Username | Comment Id | Comment Text | Profile URL | Avatar URL | Date |
|----------|----------------|-------------|---|------------------------------|---|------|
| 2,94E+09 | rattnaadaw | 1,81112E+16 | @kulo_fhitria ayo too | https://www.instagram.com/ra | https://scontent-cgk1-1.12/28/2023, 10:46:12 PM | |
| 8,3E+09 | kulo_fhitria | 1,81516E+16 | @rattnaadaw hayok kapan | https://www.instagram.com/ki | https://scontent-cgk1-1.12/28/2023, 10:47:03 PM | |
| 5,61E+10 | dusk_r95 | 1,79819E+16 | Info web untuk lokernya | https://www.instagram.com/di | https://scontent-cgk1-1.12/29/2023, 8:25:47 AM | |
| 5,73E+10 | _regaindraa | 1,82795E+16 | Finally mie gacoan sudah buka di tasik | https://www.instagram.com/_1 | https://instagram.fsrg6-12/29/2023, 3:01:39 PM | |
| 4,51E+10 | dwkrmdani | 1,80106E+16 | Minnn gratis oh min 😊😊 | https://www.instagram.com/di | https://scontent-cgk1-1.12/29/2023, 3:33:16 PM | |
| 6,22E+10 | ulfah_hasanah | 1,7898E+16 | Daerah mana kak | https://www.instagram.com/ul | https://instagram.fsrg6-12/29/2023, 5:20:20 PM | |
| 5,69E+10 | salummeliyan | 1,83061E+16 | hihi aku udh ke sono enak tau | https://www.instagram.com/sa | https://scontent-cgk1-1.12/29/2023, 6:28:34 PM | |
| 3,56E+10 | masmunari60 | 1,81086E+16 | Ngantri panjang ditasik mah | https://www.instagram.com/m | https://scontent-cgk1-1.12/29/2023, 9:50:26 PM | |
| 6,35E+10 | inss4042 | 1,80083E+16 | Buka cabang di Banjarbaru, Kalimantan selatan | https://www.instagram.com/in | https://instagram.fsrg6-12/29/2023, 9:54:52 PM | |
| 5,3E+10 | kapten_ekadia | 1,79822E+16 | Siapa yang pernah makan saya | https://www.instagram.com/ki | https://scontent-cgk1-1.12/30/2023, 7:05:08 AM | |
| 5,62E+09 | anisitirohanii | 1,8089E+16 | ME GACOOAN Bantar gebang - Cimuning | https://www.instagram.com/ai | https://scontent-cgk1-1.12/30/2023, 11:42:30 AM | |
| 5,39E+10 | dstiinrainiii | 1,8275E+16 | Knpaaa sii gabuka cabang d garutt | https://www.instagram.com/di | https://instagram.fsrg6-12/30/2023, 2:58:31 PM | |
| 5,36E+10 | miraclegracia2 | 1,83212E+16 | B | https://www.instagram.com/m | https://scontent-cgk1-1.12/30/2023, 6:40:06 PM | |
| 1,11E+10 | eka.srwt | 1,79151E+16 | @ekaaoktariani | https://www.instagram.com/el | https://instagram.fcgk1(12/30/2023, 6:53:12 PM | |
| 5,2E+10 | vitry.indriaa | 1,80178E+16 | @fasidol99 mie gacoan yng d pondok gede juga | https://www.instagram.com/vi | https://scontent-cgk1-1.12/31/2023, 12:11:08 AM | |
| 5,57E+10 | heo_hyun273 | 1,80361E+16 | @pitelii_ cepet pulang buk, kita harus serbuu.. | https://www.instagram.com/h | https://instagram.fcgk1(12/31/2023, 12:40:41 AM | |
| 5,75E+10 | rismaayuristia | 1,82612E+16 | Enk nya g main 😊 | https://www.instagram.com/ri | https://scontent-cgk1-1.12/31/2023, 9:57:25 AM | |
| 5,46E+10 | calon_flighta | 1,80517E+16 | Ini yang di bali itu kk | https://www.instagram.com/c | https://instagram.fcgk1(12/31/2023, 10:16:08 AM | |
| 5,46E+10 | calon_flighta | 1,80189E+16 | Aku tau kesitu satu keluarga pada bingung karn: | https://www.instagram.com/c | https://instagram.fcgk1(12/31/2023, 10:18:55 AM | |
| 5,45E+10 | flourecentx | 1,79079E+16 | @salmarqss cobainn bubb | https://www.instagram.com/fl | https://scontent-cgk1-1.12/31/2023, 10:55:36 AM | |

Gambar 3.1. Dataset Comment tentang Mie Gacoan

3.2. Pelabelan

Pada penelitian ini dilakukan pelabelan secara manual oleh pelabel. Pelabel memberikan label positif untuk sentiment positif dan negatif untuk sentiment negatif. Beberapa comment tentang pelayanan Mie Gacoan yang telah diberi label ditampilkan dalam Tabel 3.1 dan Tabel 3.2.

Tabel 3.1. Sentimen Positif

| Comment | Label |
|--|---------|
| hihi aku udh ke sono enak tau | Positif |
| Malahan aku suka bangetttt sama mie gacoan | Positif |
| Favorit aku udang keju @megacoangatsu | Positif |
| mie hompimpa, kenapa jarang di pesen padahal enak banget?? | Positif |
| Udang keju krna itu enak banget tp aku selalu pesan | Positif |

Tabel 3.2. Sentimen Negatif

| Comment | Label |
|--|---------|
| Ngantri panjang ditasik mah | Negatif |
| @fasidol99 mie gacoan yng d pondok gede juga karyawannya kurang stroong...mie'y udh mau hrs d tambah nahan pedas minumnya bru dtng?? | Negatif |
| pelayanan ny g bngt | Negatif |
| Mie ini bikin ngantri yg gw GK suka perut dh laper duluan | Negatif |

| | |
|--|---------|
| Gacoan cabang tugu yoga super ancurrrrr!!!! Makanan dingin, pangsit dingin lembek kaya pangsit rebus jdnya | Negatif |
|--|---------|

3.3. Preprocessing Data

Preprocessing Data adalah tahap dimana teks yang akan diklasifikasi akan dibersihkan dan dipersiapkan sebelum dilakukan analisis dokumen [8]. Pada penelitian ini tahapan *preprocessing* data menggunakan fitur *text preprocessing* yang ada pada *library* NLTK (*Natural Language Toolkit*). Tahapan *preprocessing* yang dilakukan yaitu:

1. Cleaning

Tahapan ini bertujuan untuk menghilangkan atribut yang ada pada data seperti simbol, tanda baca, *hashtag*, URL, angka, dan noise. Hasil pemrosesan *cleaning* data dapat dilihat pada Tabel 3.3.

Tabel 3.3. Contoh Data Hasil *Cleaning*

| | |
|---------------|---------------------------------------|
| Input | Favorit aku udang keju @megacoangatsu |
| Output | Favorit aku udang keju |

2. Case Folding

Tahap ini merupakan proses untuk mengubah semua data huruf kapital yang ada pada dataset menjadi huruf kecil (non-kapital). Hasil pemrosesan *case folding* dapat dilihat pada Tabel 3.4.

Tabel 3.4. Contoh Data Hasil *Case Folding*

| | |
|---------------|---|
| input | GACOAN GADA NIATAN MAU BUKA CABANG DI SERANG APA???????? |
| output | gacoan gada niatan mau buka cabang di serang apa |

3. Tokenizing

Tahap ini merupakan proses memecahkan kalimat menjadi satuan kata yang menyusunnya. Hasil pemrosesan *tokenizing* dapat dilihat pada Tabel 3.5.

Tabel 3.5. Contoh Data Hasil *Tokenizing*

| | |
|---------------|-------------------------------------|
| input | favorit aku udang keju |
| output | ['favorit', 'aku', 'udang', 'keju'] |

4. Normalisasi

Tahap normalisasi adalah tahap penormalan bentuk kata. Pada tahap normalisasi yaitu mengubah bentuk kata yang kurang normal seperti singkatan, bahasa asing dan kata yang kurang baku dalam *dataset* menjadi kata baku. Hasil pemrosesan normalisasi dapat dilihat pada Tabel 3.6.

Tabel 3.6. Contoh Normalisasi

| | |
|--------------|--|
| Input | gak banget apa lagi pelayanan banget |
| Ouput | ['tidak ', 'banget ', 'layan ', 'banget '] |

5. Stopword Removal

Tahap ini bertujuan untuk menghapus kata-kata yang terkandung dalam daftar kata *stopword*, yang dalam hal ini peneliti menggunakan *stopword* yang terdapat dalam *library* NLTK (*Natural Language Toolkit*) dengan dataset bahasa Indonesia. Hasil pemrosesan *stopword removal* dapat dilihat pada Tabel 3.7.

Tabel 3.7. Contoh Hasil *Stopword Removal*

| | |
|--------------|---|
| Input | sayang nya yg punya alergi udang hanya bisa menatapi sambil ngiler bikin dimsum ayam nya donk min |
| Ouput | ['sayang', 'alergi', 'udang', 'bisa', 'menatapi', 'ngiler', 'dimsum', 'ayam', 'donk', 'min'] |

6. *Stemming*

Stemming merupakan tahapan yang bertujuan untuk mengembalikan suatu kata ke bentuk akar katanya atau menghilangkan imbuhan (awalan dan akhiran) yang terdapat pada kata tersebut. Hasil pemrosesan *Stemming* dapat dilihat pada tabel 3.8.

Tabel 3.8. Contoh Data Hasil *Stemming*

| <i>Comment</i> |
|---|
| Suka |
| Minum favorit green thai tea |
| Kangen gaco buka batam |
| Enak banget |
| nyari sakit pakai tunggu superwin akun menang |
| Mie suit udang keju best combo |

7. *Wordcloud*

Wordcloud merupakan suatu gambar yang berisi kumpulan kata yang besarnya suatu kata merepresentasikan kemunculan atau tingkat kepentingan. Hal tersebut dapat dilihat Pada Gambar 3.2. dan Gambar 3.3. yang merupakan *wordcloud* sentimen negatif dan positif pada dataset.

Gambar 3.2. *Wordcloud* Sentimen Negatif dalam Dataset.



Gambar 3.3. Wordcloud Sentimen Positif dalam Dataset.

3.4. Pembobotan TF-IDF

Hasil dari dataset yang telah dilatih menggunakan TF-IDF didapatkan kosakata yang berjumlah 1513 kata. Selanjutnya, pada Tabel 3.9 ditampilkan perbandingan perhitungan manual dan otomatis.

Tabel 3.9. Perbandingan TF-IDF

| Term | Manual | | | | Otomatis | |
|----------|--------|----|-------|--------|-------------|-------------|
| | TF | DF | IDF | e | TF-IDF | TF-IDF |
| Dimsum | 1 | 11 | 4,567 | 20,857 | 0,536410618 | 0,536410618 |
| Enak | 1 | 32 | 3,556 | 12,645 | 0,417665022 | 0,417665022 |
| rambutan | 1 | 5 | 5,26 | 27,668 | 0,617805967 | 0,617805967 |
| Udang | 1 | 39 | 3,363 | 11,31 | 0,394996476 | 0,394996476 |

3.5. Evaluation

Evaluasi performa model menggunakan Multinomial Naive Bayes. Setelah dataset melalui tahap *preprocessing* dan pengklasifikasian menggunakan algoritma *Naive Bayes Classifier* didapatkan nilai akurasi, presisi, *recall*, dan *F1-score* dengan memanfaatkan *Confusion Matrix* seperti pada Gambar 3.4

```

Accuracy: 90.59%
Precision: 87.50%
Recall: 95.45%
F1-Score: 91.30%
Confusion Matrix:
[[35  6]
 [ 2 42]]
TN: 35, FP: 6, FN: 2, TP: 42
Classification Report:

```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negatif | 0.95 | 0.85 | 0.90 | 41 |
| positif | 0.88 | 0.95 | 0.91 | 44 |
| accuracy | | | 0.91 | 85 |
| macro avg | 0.91 | 0.90 | 0.91 | 85 |
| weighted avg | 0.91 | 0.91 | 0.91 | 85 |

Gambar 3.4. Hasil Pengujian Evaluasi Performa

Dari Gambar 3.4 dapat diketahui bahwa hasil evaluasi menggunakan Multinomial Naive Bayes yaitu akurasi sebesar 90,59%, *precision* 87,50%, *recall* 95,45%, dan *f1-score* 91,30%.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilaksanakan, dapat disimpulkan bahwa jumlah data yang digunakan dalam penelitian ini mencapai 424 data yang diperoleh melalui pengumpulan informasi dari platform Instagram. Analisis terhadap opini pelanggan terkait pelayanan mie gacoan menunjukkan dominasi sentimen positif, di mana sebanyak 218 dari total responden menyatakan sentimen positif, menyumbang sebanyak 51,4%, sementara 206 responden mengungkapkan sentimen negatif, sekitar 48,6%. Proses ekstraksi fitur dilakukan dengan menerapkan metode TF-IDF untuk mendapatkan representasi yang lebih baik dari teks. Pembagian data dilakukan secara proporsional, dengan 80% data digunakan sebagai data latih dan 20% sebagai data uji. Hasil evaluasi menunjukkan tingkat akurasi sebesar 90,59% ketika menerapkan algoritma Naïve Bayes dalam pengolahan data. Algoritma *Naïve Bayes Classifier* cocok digunakan dalam pengklasifikasian karena menghasilkan tingkat akurasi yang tinggi dibandingkan dengan metode klasifikasi lainnya.

5. SARAN

Setelah dilakukan penelitian ini terdapat saran untuk penelitian selanjutnya adalah sebagai berikut:

1. Pengembangan pada analisis sentimen selanjutnya, dapat menambahkan optimasi jenis klasifikasi lainnya seperti *Deep Learning*, yang berpotensi dapat meningkatkan performa klasifikasi secara signifikan.
2. Pada penelitian selanjutnya, dapat melakukan filterisasi (penyaringan) data instagram untuk mendeteksi instagram yang berulang dengan menggunakan bot pendeteksi otomatis instagram duplikat. Hal ini bertujuan untuk mengurangi noise atau gangguan dalam data.
3. Pengembangan pada penyeimbangan data dapat dilakukan dengan menggunakan metode seperti SMOTE, sehingga mendapatkan hasil resampling data untuk meningkatkan akurasi hasil analisis.

UCAPAN TERIMA KASIH

Penulis menyampaikan rasa terima kasih kepada Tim Redaksi Jurnal Teknik Politeknik Negeri Sriwijaya atas kesempatan yang telah diberikan, sehingga artikel ilmiah ini dapat dipublikasikan.

DAFTAR PUSTAKA

- [1] A. Faesal, et al. "Sentimen Analisis Pada Data Tweet Pengguna Twitter Terhadap Produk Penjualan Toko ONLINE Menggunakan Metode K-Means," *Jurnal MATRIK*, pp. 207-213, 2020.
- [2] A. C. Khotimah, E. Utami. "COMPARISON NAÏVE BAYES CLASSIFIER, K-NEAREST NEIGHBOR AND SUPPORT VECTOR MACHINE IN THE CLASSIFICATION OF INDIVIDUAL ON TWITTER ACCOUNT," *Jurnal Teknik Informatika (JUTIF)*, pp. 673-680, 2022.
- [3] Finandra S, et al "PENERAPAN ANALISIS SENTIMEN MELALUI DATA INSTAGRAM UNTUK MENGETAHUI REPUTASI WISATA KULINER DI KOTA BANDUNG," *Journal Engineering*, p. 2, 2021.
- [4] M. I. Fikri, et al. "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *Jurnal SMATIKA*, vol. II, pp. 71-76, 2020.
- [5] Putra A. D. & Juanita., S. "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN," *Jurnal Teknik Informatika dan Sistem Informasi*, pp. 634-646, 2021.

-
- [6] Khairunnisa, S., Adiwijaya, Al Faraby, S. "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *Jurnal Media Informatika Budidarma*, vol. V, pp. 406-414, 2021.
- [7] Z. e. a. Syahputri, "Kerangka Berfikir Penelitian Kuantitatif," *Jurnal Ilmu Pendidikan dan Pengajaran*, p. 2, 2023.
- [8] C. P. Habibi M, "Clustering User Characteristics Based on the influence of," *Indonesian Journal of Computing and Cybernetics Systems (IJCCS)*, p. 400, 2019.
- [9] Cindo M, et al. "Literatur Review: Metode Klasifikasi Pada Sentimen Analisis," pp. 66-70, 2019.
- [10] N., Munasatya & S., Novianto. "Natural Language Processing untuk Analisis Sentimen Presiden Jokowi Menggunakan Multi Layer Perceptron," *Journal Techno.com*, vol. 19, no. XIX, pp. 237-244, 2020.
- [11] Septiani Riska Kurnia, et al. "Klasifikasi Sentimen Terhadap Ibu Kota Nusantara (IKN) pada Media Sosial Menggunakan Naive Bayes," *Jurnal Teknik*, vol. 16, no. 02, pp. 245-254, 2022.
- [12] D. Alita & A. Rahman Isnain, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," *Jurnal Komputasi*, vol. 8, no. 02, pp. 50-58, 2020.
- [13] S., Sari, et al, "Sentiment Analysis Against Beauty Shaming Comments on Twitter Social Media Using SentiStrength Algorithm," *Indonesian Journal of Informatic Research and Software Engineering Journal*, vol. 1, pp. 71-78, 2021.
- [14] Yudhistira, et al. "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clusterin," *Journal of Artificial Intelligence and Technology Information*, vol. 1, pp. 20-28, 2023.
- [15] A. Primandani., R. Waluyo, "ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," *Jurnal Teknologi Information dan Ilmu Komputer (JTIK)*, vol. 8, no. 1, pp. 147-156, 2019.