

Analisis Sentimen Ulasan Pelanggan pada Kafe Gamat Bendungan Hilir Menggunakan Algoritma *Naive bayes*

Ariyo Sheva Adhityas¹, Shafar Satya Nugraha², Muhammad Antar Syuhada*³

^{1,2,3}program Studi Sistem Informasi, Telkom University Purwokerto
Jl. DI Panjaitan No.128, Karangreja, Purwokerto Kidul, Kec. Purwokerto Sel., Kabupaten Banyumas, Jawa Tengah Indonesia

¹ariyosheva@student.telkomuniversity.ac.id

²shafarsatyaa@student.telkomuniversity.ac.id

³antarsyuhada@student.telkomuniversity.ac.id

Dikirim pada 19-11-20224, Direvisi pada 24-11-2024, Diterima pada 30-11-2024

Abstrak

Perkembangan bisnis kuliner di Indonesia yang pesat turut dipengaruhi oleh kemajuan teknologi internet, termasuk penggunaan platform ulasan seperti *Google Maps*. Kafe Gamat Bendungan Hilir, yang strategis dan memiliki menu bervariasi, memanfaatkan ulasan pelanggan untuk meningkatkan loyalitas. Penelitian ini bertujuan menganalisis sentimen ulasan pelanggan kafe tersebut menggunakan algoritma *Naive bayes*. Data diperoleh melalui scraping ulasan *Google Maps* sebanyak 997 data. Tahapan analisis meliputi *preprocessing* data, seperti *case folding*, *cleaning*, *tokenization*, *stopword removal*, dan *stemming*, serta *feature extraction* menggunakan metode TF-IDF. Setelah itu, dilakukan klasifikasi menggunakan algoritma *Naive bayes*, diikuti evaluasi dengan metrik *precision*, *recall*, *f1-score*, dan akurasi. Metode SMOTE digunakan untuk menangani ketidakseimbangan data. Hasil menunjukkan akurasi model sebesar 72%. Penelitian ini diharapkan dapat membantu Kafe Gamat memahami opini pelanggan dan meningkatkan strategi layanan serta kepuasan pelanggan.

Kata Kunci: Analisis, Sentimen, *Google Maps*, *Naive bayes*

Ini adalah artikel akses terbuka di bawah lisensi [CC BY-SA](#).



Penulis Koresponden:

Muhammad Antar Syuhada

Program Studi Sistem Informasi, Universitas Telkom Kampus Purwokerto, Jl. D.I Panjaitan No.128 Purwokerto, 53147 Kabupaten Banyumas, Jawa Tengah, Indonesia Email: antarsyuhada@student.telkomuniversity.ac.id

I. PENDAHULUAN

A. Pendahuluan

Istilah kafe berasal dari kata *coffee* yang berarti kopi. Istilah ini awalnya umum dijumpai di Negara Perancis yang kemudian diadopsi oleh kota-kota di Inggris pada abad ke-19. Kafe secara harfiah memiliki arti yakni kedai kopi atau tempat untuk menikmati makanan dan minuman sambil menikmati hiburan. Perkembangan bisnis kuliner di Indonesia semakin pesat, selain produk makanan olahan, minuman olahan juga mendominasi, salah satunya saat ini menjadi produk olahan paling populer bagi para penikmat, yaitu kopi yang menjadi top konsumen *purchase* di tahun 2019[1]. Saat ini, sektor kopi Indonesia tumbuh lebih makmur dan menunjukkan tanda-tanda pertumbuhan yang menggembirakan. Dengan bertambahnya banyaknya kedai kopi di Indonesia, jumlah pelanggan yang memilih untuk membeli barang kopi meningkat setiap tahunnya, dan konsumsi kopi meningkat hingga 8% setiap tahunnya[2].

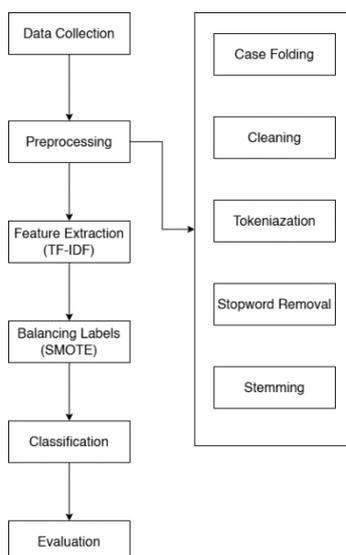
Perkembangan kafe yang pesat saat ini dipengaruhi oleh kemajuan teknologi internet[3]. Berdasarkan hasil wawancara dengan manajer Kafe Gamat Bendungan Hilir, didapat hasil bahwa saat ini, Kafe Gamat Bendungan Hilir terlatak di wilayah Jakarta Pusat. Kafe ini menjadi salah satu tempat yang sering dikunjungi oleh masyarakat karena tempatnya yang strategis dan menu yang bervariasi. Penjualan dan kelangsungan hidup perusahaan itu sendiri dapat dipengaruhi oleh loyalitas pelanggan. Pihak kafe harus memperhatikan ide dan pendapat pelanggan mereka untuk meningkatkan loyalitas pelanggan.

Google Review adalah salah satu fitur yang tersedia di *Google Maps*. Fitur ini memungkinkan pengguna memberikan penilaian berupa angka, ulasan teks, dan foto terkait suatu lokasi. Pengguna dapat dengan mudah memberikan penilaian berdasarkan berbagai faktor yang memengaruhi kepuasan pelanggan[4]. Namun, manajemen kafe akan merasa sangat sulit untuk mendapatkan dan memeriksa setiap evaluasi pelanggan secara individual. Untuk mendukung keputusan tentang operasi, layanan, dan pemasaran berdasarkan evaluasi pelanggan, diperlukan sebuah sistem yang dapat mengevaluasi sentimen ulasan pelanggan. Salah satunya dengan melakukan analisis sentimen. Analisis sentimen dapat diaplikasikan sebagai sarana untuk memahami pandangan masyarakat terhadap suatu persoalan, terdapat platform *Google review* yang dapat digunakan untuk mengetahui sekaligus menganalisis opini dari para pengunjung, dengan menggunakan teknik crawling dan diolah menggunakan metode *Naive bayes*.

Naive Bayes (NB) adalah algoritma klasifikasi probabilistik yang populer. Meskipun sederhana, algoritma ini sangat efisien dan memiliki banyak aplikasi di dunia nyata, mulai dari rekomendasi produk hingga diagnosis medis, serta pengendalian kendaraan otonom[5]. Model *Naive bayes* cocok untuk digunakan pada dataset yang sangat besar, baik dalam pembangunan maupun analisis lanjutan. Sebagai metode klasifikasi yang sederhana namun canggih, model ini tetap efektif bahkan dalam skenario yang kompleks[6]. Terdapat penelitian sebelumnya dari Prasetyo dkk yang menggunakan metode *Naive bayes* pada analisis sentimen pelanggan pada Kafe Kopi Paste, algoritma *Naive bayes* menghasilkan rata-rata akurasi yang tinggi, yaitu sebesar 93,33% [7]. Berdasarkan penelitian tersebut, penelitian ini bertujuan untuk mengetahui sentimen pelanggan Kafe Gamat Bendungan Hilir menggunakan algoritma *Naive bayes*.

II. METODE PENELITIAN

Metode penelitian yang dilakukan pada penelitian mencakup beberapa tahapan, yaitu pengumpulan data, *preprocessing data*, *extraction feature*, *classification*, serta *evaluation*. Proses pengumpulan data dilakukan dengan scrapping data *Google Maps* menggunakan *Google Colab*. Selanjutnya, dilakukan *preprocessing data* untuk membersihkan data. *Feature extraction* dilakukan dengan metode TF-IDF untuk melakukan pembobotan pada setiap kata, kemudian dilakukan klasifikasi menggunakan algoritma *Naive bayes*, dan hasil dari klasifikasi kemudian dievaluasi. Tahapan penelitian dapat dilihat pada gambar 1:



Gambar 1. Alir Penelitian

A. Data collection

Pengambilan data dilakukan dengan cara scraping dengan jumlah 997 ulasan dan diambil pada Oktober 2024, melalui review *Google Maps*.

B. Preprocessing

Text preprocessing adalah langkah awal dalam pengolahan data teks, di mana data yang mengandung noise akan dibersihkan untuk mempermudah analisis lebih lanjut. Proses ini mencakup berbagai rutin dan metode yang bertujuan untuk mempersiapkan data teks agar siap digunakan dalam operasi knowledge discovery pada sistem *text mining*[8]. Tahapan pada *preprocessing data* adalah *case folding*, *cleaning*, *tokenization*, *stopword removal*, dan *stemming*. Bahasa pemrograman yang digunakan adalah *Python* dan menggunakan library *pandas*, *nlTK*, *numpy*, *re*, dan *string*.

1. Case folding

Proses *case folding* bertujuan untuk menyamakan semua huruf kapital menjadi huruf kecil.

2. Cleaning

Cleaning adalah proses membersihkan dokumen dari kata-kata yang tidak relevan guna mengurangi noise. Kata-kata yang dihapus meliputi karakter HTML, kata kunci, emotikon, tagar (#), mention username (@username), angka, dan URL.

3. Tokenization

Tokenisasi adalah proses menyeleksi, memecah, dan memotong kata-kata dalam dokumen menjadi istilah-istilah (*term*) berdasarkan pemisah seperti spasi.

4. Stopword Removal

Stopword adalah proses menyaring kata-kata dalam dokumen yang tidak relevan atau tidak memiliki hubungan dengan analisis sentimen[9].

5. Stemming

Stemming adalah tahap yang bertujuan untuk memetakan atau menguraikan berbagai bentuk kata dalam teks menjadi kata dasar[10].

C. Feature extraction (TF-IDF)

Feature extraction adalah proses mengubah data dalam bentuk teks menjadi vektor angka sehingga dapat diproses oleh komputer menggunakan *Python*, disertai dengan pembuktian perhitungan secara manual[11]. *Term Frequency* (TF) adalah teknik untuk menentukan bobot sebuah dokumen dengan menghitung frekuensi kemunculan sebuah *term* dalam dokumen tersebut. *Inverse Document Frequency* (IDF) adalah metode yang digunakan untuk mengukur seberapa luas *term* tersebut tersebar di seluruh dokumen. Rumus TF-IDF adalah sebagai berikut:

$$tf_{t,d} = \frac{\text{Jumlah kemunculan term } t \text{ dalam dokumen } d}{\text{Total jumlah kemunculan term dalam dokumen } d}$$

$$idf_t = \log\left(\frac{N}{df_t}\right)$$

N = Total jumlah dokumen dalam kumpulan dokumen

Setelah nilai TF dan IDF diperoleh, langkah berikutnya adalah menghitung nilai TF-IDF dengan cara mengalikan nilai TF dengan nilai IDF[12].

$$tfidf_{t,d} = tf_{t,d} * idf_t$$

D. Balancing Labels (SMOTE)

SMOTE adalah metode yang digunakan untuk menyeimbangkan data kelas yang tidak seimbang melalui teknik oversampling. Pendekatan ini bekerja dengan menghasilkan duplikasi data sintetis pada kelas minoritas sehingga jumlahnya seimbang dengan kelas mayoritas[13].

E. Naive bayes Algorithm

Naive bayes Classifier adalah metode klasifikasi yang didasarkan pada Teorema Bayes. Metode ini menggunakan pendekatan probabilitas dan statistik yang dikembangkan oleh ilmuwan asal Inggris, Thomas Bayes. Metode ini memprediksi peluang di masa depan berdasarkan data atau pengalaman dari masa lalu, sehingga dikenal dengan istilah Teorema Bayes. Rumus algoritma *Naive bayes* adalah sebagai berikut[14]:

$$P\left(\frac{H}{X}\right) = \frac{\left(P\left(\frac{X}{H}\right)P(H)\right)}{P(X)}$$

$P\left(\frac{H}{X}\right)$ = Probabilitas posterior

$P\left(\frac{X}{H}\right)$ = *Likelihood* yang merupakan probabilitas prediktor kelas yang diberikan

$P(H)$ = Probabilitas prior dari kelas a

$P(X)$ = Probabilitas prior dari kelas y

F. Evaluation

Pada tahap evaluasi, kinerja sistem dianalisis untuk menilai hasil klasifikasi yang telah dilakukan. Proses ini mencakup perhitungan nilai *accuracy*, *precision*, *recall*, dan *f1-score*. *Precision* mengukur tingkat akurasi model dalam memprediksi data positif, sementara *recall* menghitung seberapa efektif model dalam menemukan data positif. *F1-score* digunakan untuk menilai keseimbangan antara *precision* dan *recall*.

III. HASIL DAN PEMBAHASAN

A. Data collection

Proses crawling data ulasan Kafe Gamat Bendungan Hilir pada *Google Maps* menggunakan *Google Colab*, berhasil diperoleh sebanyak 997 data. Data ulasan diambil pada rentang waktu Januari 2022 sampai Oktober 2024.

	name	rating	date	snippet
0	Geraldi Ryan Wibinata	3.0	sebulan lalu	Tempatnya enak dan cozy buat kerja ataupun bua...
1	Ema Lihdya	5.0	3 bulan lalu	Salah satu cafe yg nyaman di daerah benhil. Te...
2	putri dianita	5.0	seminggu lalu	Lokasi dan akses: Agak masuk2 tapi G Maps sesu...
3	putu nelly	5.0	seminggu lalu	Tempatnya nyaman, keliatan kecil dr luar tp di...
4	Krisna Vidya	5.0	3 bulan lalu	Senang sekali nostalgia masa masa dulu suka ja...
...
993	Lydia Lia Lucyawaty	5.0	6 tahun lalu	NaN
994	faiha Anezka	4.0	5 tahun lalu	NaN
995	Tresvel Nazwil	4.0	4 tahun lalu	NaN
996	Karin Sentosa	4.0	4 tahun lalu	NaN
997	Abie Algiffary	4.0	5 tahun lalu	NaN

998 rows x 4 columns

Gambar 2. Hasil Crawling Data

B. Preprocessing

Tahap *preprocessing data* dilakukan dengan beberapa proses, yaitu *case folding*, *cleaning*, *tokenization*, *stopword removal*, dan *stemming*. Hasil data setelah *preprocessing* berisi 997 data yang sudah dibersihkan dari elemen yang tidak diperlukan dan siap digunakan untuk analisis dengan algoritma *Naive bayes*.

a. Case folding

Proses *case folding* dilakukan untuk mengubah semua huruf dalam teks menjadi huruf kecil untuk menyamakan format dan menghilangkan perbedaan akibat kapitalisasi. Proses *case folding* penting untuk menjaga konsistensi data saat analisis teks dilakukan.

	name	rating	date	snippet	cleaning	case_folding
0	Geraldi Ryan Wibinata	3.0	sebulan lalu	Tempatnya enak dan cozy buat kerja ataupun bua...	tempatnya enak dan cozy buat kerja ataupun bua...	tempatnya enak dan cozy buat kerja ataupun bua...
1	Emah Lihdy	5.0	3 bulan lalu	Salah satu cafe yg nyaman di daerah benhil. Te...	salah satu cafe yg nyaman di daerah benhil tem...	salah satu cafe yg nyaman di daerah benhil tem...
2	putri dianita	5.0	seminggu lalu	Lokasi dan akses: Agak masuk2 tapi G Maps sesu...	lokasi dan akses agak masuk2 tapi g maps sesua...	lokasi dan akses agak masuk2 tapi g maps sesua...
3	putu nelly	5.0	seminggu lalu	Tempatnya nyaman, keliatan kecil dr luar tp di...	tempatnya nyaman keliatan kecil dr luar tp di ...	tempatnya nyaman keliatan kecil dr luar tp di ...
4	Krisna Vidya	5.0	3 bulan lalu	Senang sekali nostalgia masa masa dulu suka ja...	senang sekali nostalgia masa masa dulu suka ja...	senang sekali nostalgia masa masa dulu suka ja...
...
993	Lydia Lia Lucyawaty	5.0	6 tahun lalu	NaN	NaN	NaN
994	faiha Anezka	4.0	5 tahun lalu	NaN	NaN	NaN
995	Tresvel Nazwil	4.0	4 tahun lalu	NaN	NaN	NaN
996	Karin Sentosa	4.0	4 tahun lalu	NaN	NaN	NaN
997	Abie Algiffary	4.0	5 tahun lalu	NaN	NaN	NaN

Gambar 3. Hasil Case Folding

b. *Cleaning*

Proses *cleaning* adalah proses menghapus elemen-elemen yang tidak relevan seperti angka, simbol, tanda baca, URL, atau emoji dari teks. *Cleaning* dilakukan untuk mendapatkan data yang lebih bersih dan fokus pada informasi yang penting untuk analisis.

	name	rating	date	snippet	cleaning
0	Geraldi Ryan Wibinata	3.0	sebulan lalu	Tempatnya enak dan cozy buat kerja ataupun bua...	tempatnya enak dan cozy buat kerja ataupun bua...
1	Emah Lihdy	5.0	3 bulan lalu	Salah satu cafe yg nyaman di daerah benhil. Te...	salah satu cafe yg nyaman di daerah benhil tem...
2	putri dianita	5.0	seminggu lalu	Lokasi dan akses: Agak masuk2 tapi G Maps sesu...	lokasi dan akses agak masuk2 tapi g maps sesua...
3	putu nelly	5.0	seminggu lalu	Tempatnya nyaman, keliatan kecil dr luar tp di...	tempatnya nyaman keliatan kecil dr luar tp di ...
4	Krisna Vidya	5.0	3 bulan lalu	Senang sekali nostalgia masa masa dulu suka ja...	senang sekali nostalgia masa masa dulu suka ja...
...
993	Lydia Lia Lucyawaty	5.0	6 tahun lalu	NaN	NaN
994	faiha Anezka	4.0	5 tahun lalu	NaN	NaN
995	Tresvel Nazwil	4.0	4 tahun lalu	NaN	NaN
996	Karin Sentosa	4.0	4 tahun lalu	NaN	NaN
997	Abie Algiffary	4.0	5 tahun lalu	NaN	NaN

Gambar 4. Hasil Cleaning

c. *Tokenization*

Proses *tokenization* dilakukan untuk memecah teks menjadi unit-unit kecil seperti kata, frasa, atau kalimat yang disebut token untuk mempermudah analisis.

	name	rating	date	snippet	cleaning	tokenize
0	Geraldi Ryan Wibinata	3.0	sebulan lalu	Tempatnya enak dan cozy buat kerja ataupun bua...	tempatnya enak dan cozy buat kerja ataupun bua...	tempatnya enak dan cozy buat kerja ataupun bua...
1	Emah Lihdy	5.0	3 bulan lalu	Salah satu cafe yg nyaman di daerah benhil. Te...	salah satu cafe yg nyaman di daerah benhil tem...	salah satu cafe yg nyaman di daerah benhil tem...
2	putri dianita	5.0	seminggu lalu	Lokasi dan akses: Agak masuk2 tapi G Maps sesu...	lokasi dan akses agak masuk2 tapi g maps sesua...	lokasi dan akses agak masuk2 tapi g maps sesua...
3	putu nelly	5.0	seminggu lalu	Tempatnya nyaman, keliatan kecil dr luar tp di...	tempatnya nyaman keliatan kecil dr luar tp di ...	tempatnya nyaman keliatan kecil dr luar tp di ...
4	Krisna Vidya	5.0	3 bulan lalu	Senang sekali nostalgia masa masa dulu suka ja...	senang sekali nostalgia masa masa dulu suka ja...	senang sekali nostalgia masa masa dulu suka ja...
...
993	Lydia Lia Lucyawaty	5.0	6 tahun lalu	NaN	NaN	NaN
994	faiha Anezka	4.0	5 tahun lalu	NaN	NaN	NaN
995	Tresvel Nazwil	4.0	4 tahun lalu	NaN	NaN	NaN
996	Karin Sentosa	4.0	4 tahun lalu	NaN	NaN	NaN
997	Abie Algiffary	4.0	5 tahun lalu	NaN	NaN	NaN

Gambar 5. Hasil *Tokenization*

d. *Stopword Removal*

Proses *stopword removal* dilakukan untuk menghapus kata-kata umum seperti "yang," "dan," atau "di" yang dianggap tidak memiliki makna signifikan dalam analisis teks.

	name	rating	date	snippet	cleaning	case_folding	normalized_text	stopwords
0	Geraldi Ryan Wibinata	3.0	sebulan lalu	Tempatnya enak dan cozy buat kerja ataupun bua...	tempatnya enak dan cozy buat kerja ataupun bua...	tempatnya enak dan cozy buat kerja ataupun bua...	tempatnya enak dan cozy buat kerja ataupun bua...	[tempatnya, enak, cozy, kerja, nongki, outdoor...
1	Ema Lihdya	5.0	3 bulan lalu	Salah satu cafe yg nyaman di daerah berhili tem...	salah satu cafe yg nyaman di daerah berhili tem...	salah satu cafe yg nyaman di daerah berhili tem...	salah satu cafe yg nyaman di daerah berhili tem...	[salah, cafe, yg, nyaman, daerah, berhili, temp...
2	putri dianita	5.0	seminggu lalu	Lokasi dan akses: Agak masuk2 tapi G Maps sesua...	lokasi dan akses agak masuk2 tapi g maps sesua...	lokasi dan akses agak masuk2 tapi g maps sesua...	lokasi dan akses agak masuk2 tapi g maps sesua...	[lokasi, akses, masuk2, g, maps, sesuai, seati...
3	putu nelly	5.0	seminggu lalu	Tempatnya nyaman, keliatan kecil di luar tp di...	tempatnya nyaman keliatan kecil di luar tp di...	tempatnya nyaman keliatan kecil di luar tp di...	tempatnya nyaman keliatan kecil di luar tp di...	[tempatnya, nyaman, keliatan, di, tp, ruangan...
4	Krisna Vidya	5.0	3 bulan lalu	Senang sekali nostalgia masa masa dulu suka ja...	senang sekali nostalgia masa masa dulu suka ja...	senang sekali nostalgia masa masa dulu suka ja...	senang sekali nostalgia masa masa dulu suka ja...	[senang, nostalgia, suka, jajan, mie, ayam, ga...
993	Lydia Lia Lucyawaty	5.0	6 tahun lalu	NaN	NaN	NaN	NaN	[]
994	faiha Anezka	4.0	5 tahun lalu	NaN	NaN	NaN	NaN	[]
995	Tresvel Nazwil	4.0	4 tahun lalu	NaN	NaN	NaN	NaN	[]
996	Karin Sentosa	4.0	4 tahun lalu	NaN	NaN	NaN	NaN	[]
997	Abie Aligiffary	4.0	5 tahun lalu	NaN	NaN	NaN	NaN	[]

Gambar 6. Hasil *Stopword Removal*

e. Stemming

Stemming adalah proses mengubah kata ke bentuk dasarnya dengan menghilangkan imbuhan seperti awalan, akhiran, atau sisipan untuk menyederhanakan analisis teks.

	name	rating	date	snippet	cleaning	case_folding	normalized_text	stopwords	normalized	stemming_data
0	Geraldi Ryan Wibinata	3.0	sebulan lalu	Tempatnya enak dan cozy buat kerja ataupun bua...	tempatnya enak dan cozy buat kerja ataupun bua...	tempatnya enak dan cozy buat kerja ataupun bua...	tempatnya enak dan cozy buat kerja ataupun bua...	[tempatnya, enak, cozy, kerja, nongki, outdoor...	tempatnya enak dan cozy buat kerja ataupun bua...	tempat enak cozy kerja nongki outdoor indoor d...
1	Ema Lihdya	5.0	3 bulan lalu	Salah satu cafe yg nyaman di daerah berhili tem...	salah satu cafe yg nyaman di daerah berhili tem...	salah satu cafe yg nyaman di daerah berhili tem...	salah satu cafe yg nyaman di daerah berhili tem...	[salah, cafe, yg, nyaman, daerah, berhili, temp...	salah satu cafe yg nyaman di daerah berhili tem...	salah cafe yg nyaman daerah berhili tempat luas...
2	putri dianita	5.0	seminggu lalu	Lokasi dan akses: Agak masuk2 tapi G Maps sesua...	lokasi dan akses agak masuk2 tapi g maps sesua...	lokasi dan akses agak masuk2 tapi g maps sesua...	lokasi dan akses agak masuk2 tapi g maps sesua...	[lokasi, akses, masuk2, g, maps, sesuai, seati...	lokasi dan akses agak masuk2 tapi g maps sesua...	lokasi akses masuk2 g maps sesuai seating 2 la...
3	putu nelly	5.0	seminggu lalu	Tempatnya nyaman, keliatan kecil di luar tp di...	tempatnya nyaman keliatan kecil di luar tp di...	tempatnya nyaman keliatan kecil di luar tp di...	tempatnya nyaman keliatan kecil di luar tp di...	[tempatnya, nyaman, keliatan, di, tp, ruangan...	tempatnya nyaman keliatan kecil di luar tp di...	tempat nyaman liat di tp ruang luas indoor out...
4	Krisna Vidya	5.0	3 bulan lalu	Senang sekali nostalgia masa masa dulu suka ja...	senang sekali nostalgia masa masa dulu suka ja...	senang sekali nostalgia masa masa dulu suka ja...	senang sekali nostalgia masa masa dulu suka ja...	[senang, nostalgia, suka, jajan, mie, ayam, ga...	senang sekali nostalgia masa masa dulu suka ja...	senang nostalgia suka jajan mie ayam gamat ked...
993	Lydia Lia Lucyawaty	5.0	6 tahun lalu	NaN	NaN	NaN	NaN	[]	NaN	
994	faiha Anezka	4.0	5 tahun lalu	NaN	NaN	NaN	NaN	[]	NaN	
995	Tresvel Nazwil	4.0	4 tahun lalu	NaN	NaN	NaN	NaN	[]	NaN	
996	Karin Sentosa	4.0	4 tahun lalu	NaN	NaN	NaN	NaN	[]	NaN	
997	Abie Aligiffary	4.0	5 tahun lalu	NaN	NaN	NaN	NaN	[]	NaN	

Gambar 7. Hasil *Stemming*

C. Feature extraction (TF-IDF)

Tahap *feature extraction* menggunakan rumus TF-IDF digunakan untuk menilai pentingnya kata dalam dokumen relatif terhadap kumpulan dokumen.

0	0.0	10	40	50	abang	about	ac	ada	adalah	adem	...	\
0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.065061	0.0	0.0	...	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.123146	0.0	0.0	...	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.225455	0.000000	0.0	0.0	...	
3	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.380190	0.0	0.0	...	
4	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0	...	
	worth	ya	yaa	yamin	yang	yg	you	your	yummy	zonk		
0	0.0	0.0	0.0	0.0	0.071105	0.148428	0.0	0.0	0.0	0.0		
1	0.0	0.0	0.0	0.0	0.000000	0.140471	0.0	0.0	0.0	0.0		
2	0.0	0.0	0.0	0.0	0.0	0.223051	0.000000	0.0	0.0	0.0		
3	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0		
4	0.0	0.0	0.0	0.0	0.000000	0.114702	0.0	0.0	0.0	0.0		

Gambar 8. Hasil *Feature extraction* dengan Rumus TF-IDF

Data yang telah dilakukan pembobotan selanjutnya dilakukan *labelling* dengan package `textblob`. *Labelling* dilakukan untuk memberikan label pada tiap data, berupa label negatif, netral, atau positif.



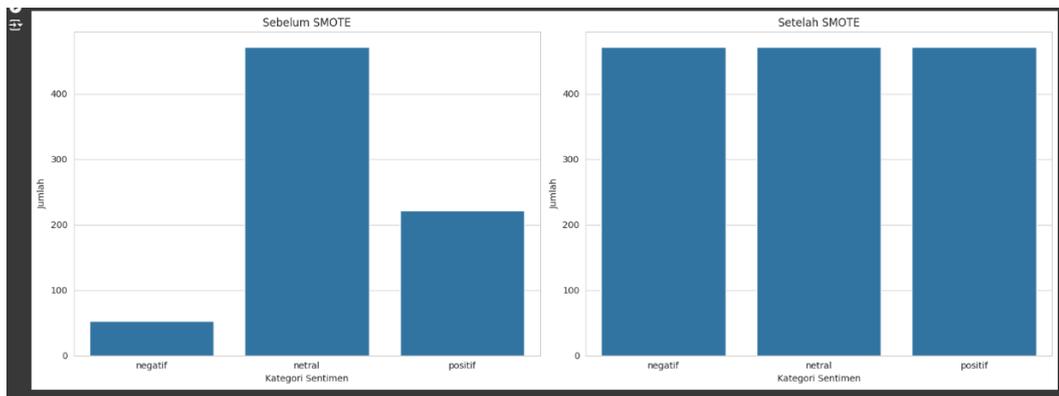
	stemming_data	score	sentiment
0	tempat enak cozy kerja nongki outdoor indoor d...	-0.075000	Negative
1	salah cafe yg nyaman daerah benhil tempat luas...	0.000000	Neutral
2	lokasi akses masuk2 g maps sesuai seating 2 la...	0.158333	Positive
3	tempat nyaman liat dr tp ruang luas indoor out...	0.000000	Neutral
4	senang nostalgia suka jajan mie ayam gamat ked...	1.000000	Positive
...
993		NaN	Negative
994		NaN	Negative
995		NaN	Negative
996		NaN	Negative
997		NaN	Negative

998 rows x 3 columns

Gambar 9. Labelling Data menggunakan Textblob

D. SMOTE

Data yang telah dilakukan *feature extraction* menggunakan rumus TF-IDF kemudian diseimbangkan menggunakan metode SMOTE.



Gambar 10. Hasil *Balancing data* menggunakan SMOTE

E. Evaluation

Tahapan selanjutnya yaitu *evaluation*. Data yang telah melalui proses *preprocessing* dilakukan klasifikasi menggunakan algoritma *Naive bayes*. Hasil klasifikasi adalah sebagai berikut:

```

x_train = 1723
x_test = 431
y_train = 1723
y_test = 431

```

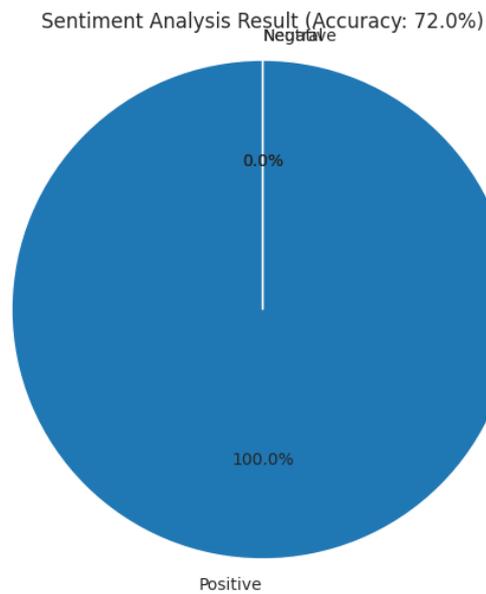
Gambar 11. Hasil *Train and Test Data*

Tahap pertama proses klasifikasi yaitu melakukan *training* dan *testing* data. Data dibagi menjadi dua bagian, yaitu data *train* dan data *test*. Data *train* digunakan untuk melatih model mempelajari pola, hubungan, atau fitur dari data yang diberikan, sedangkan data *test* digunakan untuk menilai kinerja model setelah proses pelatihan, menggunakan data baru yang belum pernah digunakan oleh model sebelumnya.

Tahap selanjutnya pada tahap klasifikasi yaitu proses klasifikasi *Naive bayes*. Gambar X merupakan hasil evaluasi model klasifikasi berdasarkan metrik seperti precision, recall, f1-score, dan support untuk setiap kelas (0, 1, 2). Accuracy model mencapai 73%, sedangkan metrik rata-rata seperti macro avg yang menghitung rata-rata antar kelas secara seimbang dan weighted avg yang menghitung rata-rata dengan mempertimbangkan jumlah data di setiap kelas menunjukkan performa keseluruhan model. Performa berbeda antar kelas terlihat dari metrik, seperti kelas 0 yang memiliki recall tinggi (1.00) namun precision lebih rendah (0.60), sementara kelas 1 memiliki precision sempurna (1.00) tetapi recall rendah (0.26).

	precision	recall	f1-score	support
0	0.60	1.00	0.75	154
1	1.00	0.26	0.41	143
2	0.90	0.91	0.90	134
accuracy			0.73	431
macro avg	0.83	0.72	0.69	431
weighted avg	0.82	0.73	0.68	431

Gambar 12. Hasil Klasifikasi menggunakan Algoritma *Naive bayes*



Gambar 13. Hasil Akurasi *Naive bayes* pada Analisis Sentimen

Confusion Matrix

Actual	0	154	0	0
	1	92	37	14
	2	12	0	122
		0	1	2
		Predicted		

Gambar 14. *Confusion matrix*

Gambar X menunjukkan hasil analisis sentimen dengan akurasi 72%, di mana semua data digolongkan ke dalam sentimen positif (100%), sedangkan sentimen negatif tidak terdeteksi (0%). Gambar X menunjukkan *confusion matrix* untuk model klasifikasi dengan tiga kelas. Matriks menunjukkan bahwa kelas 0 diklasifikasikan dengan benar sebanyak 154 kali, namun kelas 1 dan 2 memiliki sejumlah besar kesalahan klasifikasi, terutama kelas 1 yang sering salah diprediksi sebagai kelas 0.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma *Naive bayes* efektif untuk menganalisis sentimen ulasan pelanggan Kafe Gamat Bendungan Hilir, dengan akurasi sebesar 72%. Meskipun terdapat kesalahan klasifikasi pada data minoritas, metode ini dapat memberikan gambaran yang jelas tentang opini pelanggan dan menjadi dasar pengambilan keputusan untuk meningkatkan strategi bisnis dan pelayanan.

DAFTAR PUSTAKA

- [1] T. A. Sari, E. Sinduningrum, Dan F. Noor Hasan, "Klik: Kajian Ilmiah Informatika Dan Komputer Analisis Sentimen Ulasan Pelanggan Pada Aplikasi Fore Coffee Menggunakan Metode Naive Bayes," *Media Online*), Vol. 3, No. 6, Hlm. 773–779, 2023.
- [2] P. Hartini *Dkk.*, "Jurnal Informatika Dan Rekayasa Perangkat Lunak *Naive Bayes Classifier* Untuk Analisis Sentimen Ulasan Pelanggan Pada Domo Coffee And Resto".
- [3] D. A. Hamidah, R. Salkiawati, Dan R. Sari, "Analisis Sentimen Ulasan Customer Kopi Tmlst Menggunakan Algoritma Naive Bayes," *Journal Of Students' Research In Computer Science*, Vol. 5, No. 1, Hlm. 27–40, Mei 2024.
- [4] F. U. Haq Dan H. Rachmat, "Penggunaan *Google Review* Sebagai Penilaian Kepuasan Pengunjung Dalam Pariwisata," *Tornare-Journal Of Sustainable Tourism Research*, 2020.
- [5] V. Jackins, S. Vimal, M. Kaliappan, Dan M. Y. Lee, "Ai-Based Smart Prediction Of Clinical Disease Using Random Forest Classifier And *Naive Bayes*," *Journal Of Supercomputing*, Vol. 77, No. 5, Hlm. 5198–5219, Mei 2021.
- [6] I. Wickramasinghe Dan H. Kalutarage, "*Naive Bayes: Applications, Variations And Vulnerabilities: A Review Of Literature With Code Snippets For Implementation*," *Soft Comput*, Vol. 25, No. 3, Hlm. 2277–2293, Feb 2021.
- [7] Prasetyo, V. Riandaru, Ryanda, I. Akbar, Prima, Dan D. Ardy, "Analisis Sentimen Dan Kategorisasi Review Pelanggan Pada Cafe Kopi Paste Dengan Metode *Naive Bayes* Dan K-Nearest Neighbor," Jul 2023, Diakses: 30 November 2024.
- [8] E. Yuniar, D. S. Utsalinah, Dan D. Wahyuningsih, "Implementasi Scrapping Data Untuk Sentiment Analysis Pengguna Dompot Digital Dengan Menggunakan Algoritma Machine Learning," *Jurnal Janitra Informatika Dan Sistem Informasi*, Vol. 2, No. 1, Hlm. 35–42, Apr 2022.
- [9] D. Darwis, E. Shintya Pratiwi, A. Ferico, Dan O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," 2020.

- [10] A. Muzaki Dan A. Witanti, "Sentiment Analysis Of The Community In The Twitter To The 2020 Election In Pandemic Covid-19 By Method *Naive Bayes* Classifier," *Jurnal Teknik Informatika (Jutif)*, Vol. 2, No. 2, Hlm. 101–107, Mar 2021.
- [11] K. Tri Putra, M. Amin Hariyadi, Dan C. Crysdiyan, "Perbandingan *Feature Extraction* Tf-Idf Dan Bow Untuk Analisis Sentimen Berbasis Svm."
- [12] I. Anbi Fahrezi, N. Azmi Verdikha, Dan U. Muhammadiyah Kalimantan Timur, "Analisis Sentimen Twitter Atas Isu Hak Angket Menggunakan Pembobotan Tf-Idf Dan Algoritma Svm," 2024.
- [13] N. N. Sholihah Dan A. Hermawan, "Implementation Of Random Forest And Smote Methods For Economic Status Classification In Cirebon City," *Jurnal Teknik Informatika (Jutif)*, Vol. 4, No. 6, Hlm. 1387–1397, Des 2023.
- [14] A. Felicia Watratan, A. B. Puspita, D. Moeis, S. Informasi, Dan S. Profesional Makassar, "Implementasi Algoritma *Naive Bayes* Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," 2020