

Implementasi *Lexicon-Based Sentiment Analysis* dan Naive Bayes Untuk Mengetahui Opini Pengunjung Kedai Roti Ibu Saya Berdasarkan Ulasan Aplikasi Google Maps

Arif Wirawan¹, Muhamad Rafi Ramdhani², Muhammad Azkal Azkia³, Musthofa Galih Pradana, M.Kom.⁴

S1 Informatika / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. RS Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450

2210511004@mahasiswa.upnvj.ac.id¹, 2210511019@mahasiswa.upnvj.ac.id²,

2210511038@mahasiswa.upnvj.ac.id³, musthofagalihpradana@upnvj.ac.id⁴

Abstrak. Penelitian ini menganalisis sekitar 500 ulasan Google Maps pada Kedai Roti Ibu Saya, sebuah kedai roti yang terkenal yang terletak di Kota Depok, Jawa Barat Menggunakan analisis sentimen metode *lexicon-based* dan Naive Bayes. Penelitian ini bertujuan untuk mengidentifikasi persepsi dan sentimen publik terhadap Kedai Roti Ibu Saya, serta memberikan rekomendasi strategis untuk meningkatkan kepuasan dan reputasi pelanggan. Sementara penggunaan Naive Bayes digunakan untuk mengklasifikasikan teks menjadi sentimen positif, negatif atau netral. Data ulasan melewati serangkaian tahapan, berupa *pre-processing*, *wordcloud*, pelabelan data, analisis sentimen, *oversampling*, *data training*, hingga evaluasi. Hasil penelitian menunjukkan bahwa 86,5% ulasan bersifat positif, 8,8% netral, dan 4,6% negatif. Hasil *data training* dengan Naive Bayes menunjukkan akurasi mencapai 98%, hal ini menunjukkan kemampuannya yang efektif dalam pemodelan sentimen.

Kata Kunci: Analisis, sentimen, kedai roti ibu saya, naive bayes, *lexicon-based*.

1. Pendahuluan

Kedai Roti Ibu Saya yang terletak di Jalan Limo Raya, Kecamatan Limo, Kota Depok, Jawa Barat, menjadi terkenal karena rotinya yang lezat dan harganya yang terjangkau. Kedai ini menawarkan suasana nyaman dengan area *indoor-outdoor* yang menarik, sehingga menjadi tempat favorit bagi warga lokal dan wisatawan untuk menikmati pengalaman kuliner yang unik.

Perkembangan teknologi informasi, khususnya melalui Google Maps, telah mengubah cara pengunjung berinteraksi dengan Kedai Roti Ibu Saya. Google Maps memungkinkan pengunjung memberikan ulasan dan umpan balik tentang pengalaman mereka, mencakup aspek pelayanan, kualitas makanan, kebersihan, dan fasilitas. Ulasan ini membantu pengelola kedai untuk terus meningkatkan kualitas layanan yang ditawarkan.

Penelitian ini menganalisis sekitar 500 ulasan Google Maps terhadap Kedai Roti Ibu Saya menggunakan kombinasi metode *Lexicon-Based Sentiment Analysis* dan Multinomial Naive Bayes. Analisis sentimen berbasis leksikon berfokus pada polaritas kata dan frasa, sementara Multinomial Naive Bayes digunakan untuk mengklasifikasikan teks kedalam sentimen positif, negatif, atau netral.

Pendekatan gabungan ini menawarkan solusi yang efektif untuk analisis sentimen, menangkap sentimen dasar dan kompleks dengan baik [1]. Tujuan utama penelitian ini adalah mengidentifikasi persepsi dan sentimen publik terhadap Kedai Roti Ibu Saya berdasarkan ulasan Google Maps, mengevaluasi faktor-faktor yang mempengaruhi sentimen, serta memberikan rekomendasi strategis untuk meningkatkan reputasi dan kepuasan pelanggan. Hasil penelitian ini diharapkan memberikan wawasan berharga tentang penggunaan teknik *machine learning* untuk memahami pengalaman pengunjung dan mendukung peningkatan pelayanan di Kedai Roti Ibu Saya.

2. Metode Penelitian

2.1. Lexicon-based Sentiment Analysis

Lexicon-based Sentiment Analysis atau analisis sentimen berbasis leksikon merupakan pendekatan yang bergantung pada asumsi bahwa orientasi semantik teks sangat terkait dengan polaritas kata dan frasa yang muncul di dalamnya. Pendekatan ini berfokus pada kata-kata konten, seperti kata sifat, kata keterangan, kata benda, dan kata kerja, serta frasa dan kalimat yang mengandung kata-kata tersebut.

Meskipun leksikon yang dibuat secara manual umumnya lebih akurat dibandingkan yang dibuat secara otomatis, terutama dalam tugas analisis lintas domain, proses anotasi manual sangat menguras sumber daya manusia dan waktu. Sebaliknya, leksikon yang dibuat secara otomatis cenderung lebih tidak stabil, namun biasanya memiliki ukuran yang lebih besar daripada yang dibuat secara manual. Namun, ukuran yang besar tidak selalu menunjukkan kualitas yang baik. Umumnya, leksikon yang lebih besar memiliki informasi yang kurang rinci, dan sejumlah besar entri dapat menghasilkan lebih sedikit detail dalam deskripsi atau bahkan meningkatkan jumlah *noise* [2].

Pada penelitian ini, kami menggunakan InSet atau *Indonesia Sentiment Lexicon*, yang merupakan leksikon sentimen untuk Bahasa Indonesia. Konstruksi leksikon ini mengacu pada jurnal berjudul "InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs" karya Fajri Koto dan Gemala Y. Rahmaningtyas, yang dipublikasikan oleh IEEE dalam *21st International Conference on Asian Language Processing (IALP)* di Singapura pada bulan Desember 2017. InSet Lexicon ini sendiri terdiri dari 3.609 kata positif dan 6.609 kata negatif dengan bobot berkisar antara -5 hingga +5 [3].

Analisis sentimen berbasis leksikon memiliki beberapa kekurangan, seperti hanya mengambil sentimen dari setiap kata tanpa memperhitungkan konteksnya secara keseluruhan, dan skor sentimen yang dihasilkan sangat bergantung pada pembobotan kata dalam leksikon tersebut [4]. Namun, untuk melakukan analisis dari awal tanpa data yang sudah dilabeli sebelumnya, akan sangat mahal dan rumit untuk melakukan pelabelan sentimen secara manual, terutama bagi mereka yang tidak memiliki spesifikasi tertentu. Oleh karena itu, metode leksikon sangat berguna dalam skenario seperti ini untuk melakukan analisis sentimen secara efisien dan efektif.

2.2. Multinomial Naive Bayes

Algoritma Multinomial Naive Bayes merupakan metode yang populer untuk klasifikasi teks. Pada tahap klasifikasi, algoritma ini menghitung probabilitas suatu teks termasuk dalam kategori sentimen positif, negatif, atau netral, lalu menerapkan aturan pengambilan keputusan untuk mengklasifikasikan teks tersebut.

Multinomial Naive Bayes Classifier atau disingkat MNBC adalah pengembangan dari algoritma Naive Bayes yang digunakan dalam studi ini. Beberapa penelitian terkait telah menggunakan metode ini pada model klasifikasi dengan dokumen teks. Misalnya, dalam penelitian sebelumnya tentang Deteksi Emosi di Komunitas Diabetes di Facebook, metode ini mencapai rata-rata F-score sebesar 75,7% dari 800 dokumen [5]. Dalam penelitian lain, Multinomial Naive Bayes diterapkan untuk mengklasifikasikan basis dokumen teks seperti *email filtering*, dan spesifikasi aktor dalam proses pengembangan perangkat lunak dengan klasifikasi dial atau *multi label*. Hasil-hasil ini menunjukkan bahwa algoritma Multinomial Naive Bayes dapat bekerja dengan baik dalam mengklasifikasikan data dalam bentuk teks.

Dalam formula *Multinomial Naive Bayes Classifier*, kelas teks tidak hanya ditentukan oleh kata-kata yang muncul tetapi juga oleh jumlah kemunculannya. *Multinomial Naive Bayes Classifier* memiliki bentuk umum sebagai berikut (1):

$$C_{MAP} = \arg \max_{c \in \{c_1, c_s\}} P(c|d) \quad (1)$$

atau bisa juga ditulis dalam bentuk (2):

$$C_{MAP} = \arg \max_{c \in \{c_1, \dots, c_s\}} P(c) \prod_{k=1}^m P(t_k | C) \quad (2)$$

Parameter *likelihood* ($P(t_k | c)$) diestimasi dengan menghitung kemunculan t_k dalam semua dokumen teks pelatihan dalam c , menggunakan *Laplace prior* seperti dalam persamaan (3) di bawah ini:

$$P(t_k | c) = \frac{N_k + 1}{n' + |V|} \quad (3)$$

Di mana N_k adalah jumlah kemunculan t_k dalam dokumen pelatihan dalam c , n' adalah jumlah total kemunculan kata dalam c , dan $|V|$ adalah jumlah total kata dalam data pelatihan [6].

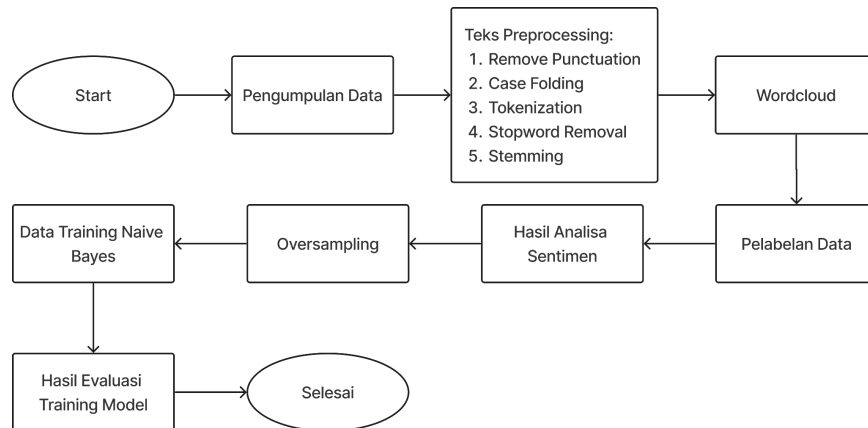
Dengan memanfaatkan pemodelan statistik, Multinomial Naive Bayes dapat menghasilkan klasifikasi yang lebih akurat dibandingkan pendekatan berbasis leksikon, terutama untuk teks yang mengandung bahasa yang lebih kompleks.

Multinomial Naive Bayes juga memiliki keunggulan dalam hal kecepatan dan efisiensi komputasi. Metode ini mampu memproses data dalam jumlah besar dengan waktu yang relatif singkat. Hal ini menjadikannya pilihan yang baik untuk aplikasi *real-time* seperti analisis sentimen pada media sosial, di mana data yang masuk sangat cepat dan memerlukan respons yang hampir seketika.

Selain itu, algoritma ini memiliki sifat *robust* terhadap fitur yang tidak relevan, karena secara otomatis memberikan bobot yang lebih rendah kepada kata-kata yang jarang muncul atau tidak signifikan. Ini membantu mengurangi *noise* dalam data dan meningkatkan kinerja keseluruhan model.

2.3. Tahapan Penelitian

Berikut ini merupakan tahapan-tahapan yang dilakukan dalam penelitian ini, yaitu sebagai berikut:



Gambar 1. Tahapan Penelitian

Setelah memahami gambaran umum tahapan penelitian yang ditunjukkan pada **Gambar 1**, tahapan-tahapan yang dilakukan dalam penelitian ini antara lain, (1) pengumpulan data, dilakukan dengan *crawling data* melalui aplikasi Google Maps terhadap ulasan Kedai Roti Ibu Saya, (2) teks *preprocessing* data, dilakukan dengan *cleansing data* yang meliputi *remove punctuation*, *case folding*, *tokenization*, *stopword removal*, dan *stemming*, (3) *wordcloud*, dilakukan visualisasi untuk melihat kata yang paling sering muncul, (4) pelabelan data, dilakukan dengan *text scoring* berdasarkan kamus kata negatif dan positif (*lexicon*), (5) hasil analisa sentimen, dilakukan untuk menentukan hasil opini pengunjung Kedai Roti Ibu Saya, (6) *oversampling*, dilakukan karena adanya data yang tidak seimbang, (7) data *training* dilakukan untuk membuat model klasifikasi sentimen, (8) hasil evaluasi training model dilakukan untuk mengevaluasi hasil pemodelan klasifikasi *data training*.

Tahapan *remove punctuation* dilakukan untuk menghapus tanda baca, emoji, karakter, *tag html*, dan simbol-simbol yang tidak memberikan manfaat dalam proses analisis sentimen.

Tabel 2. Hasil Remove Punctuation

Before	After
Lokasi strategis, luas dan Asri. Banyak tersedia kursi baik indoor maupun outdoor. Ada 2 kasir yang juga di indoor dan outdoor.\n\nHarga makanan murah dengan porsi kecil. Cocok untuk yang memang porsi kecil	Lokasi strategis luas dan Asri Banyak tersedia kursi baik indoor maupun outdoor Ada 2 kasir yang juga di indoor dan outdoor Harga makanan murah dengan porsi kecil Cocok untuk yang memang porsi kecil

3.2.2. Case Folding

Tahapan *Case Folding* berfungsi untuk mengubah semua teks yang ada menjadi huruf kecil (*lowercase*). Hal ini dilakukan untuk mengurangi variasi huruf besar dan kecil dalam teks sehingga mempermudah analisis sentimen.

Tabel 3. Hasil Case Folding

Before	After
Lokasi strategis luas dan Asri Banyak tersedia kursi baik indoor maupun outdoor Ada 2 kasir yang juga di indoor dan outdoor Harga makanan murah dengan porsi kecil Cocok untuk yang memang porsi kecil	lokasi strategis luas dan asri banyak tersedia kursi baik indoor maupun outdoor ada 2 kasir yang juga di indoor dan outdoor harga makanan murah dengan porsi kecil cocok untuk yang memang porsi kecil

3.2.3. Tokenization

Tahapan *Tokenization* berfungsi untuk memecah teks menjadi potongan-potongan yang lebih kecil yang disebut token. Token tersebut dapat berupa sebuah kata, kata sambung, frasa, maupun imbuhan. Tahapan ini membantu dalam pemrosesan bahasa alami untuk analisis sentimen.

Tabel 4. Hasil Tokenization

Before	After
lokasi strategis luas dan asri banyak tersedia kursi baik indoor maupun outdoor ada 2 kasir yang juga di indoor dan outdoor harga makanan murah dengan porsi kecil cocok untuk yang memang porsi kecil	'lokasi' 'strategis' 'luas' 'dan' 'asri' 'banyak' 'tersedia' 'kursi' 'baik' 'indoor' 'maupun' 'outdoor' 'ada' '2' 'kasir' 'yang' 'juga' 'di' 'indoor' 'dan' 'outdoor' 'harga' 'makanan' 'murah' 'dengan' 'porsi' 'kecil' 'cocok' 'untuk' 'yang' 'memang' 'porsi kecil'

3.2.4. Stopword Removal

Tahapan *Stopword Removal* merupakan metode penghapusan kata-kata umum yang tidak memberikan makna dalam analisis teks. Sehingga, kata umum yang biasanya muncul dan tidak memiliki makna disebut dengan *stopword*. Untuk itu, tahapan ini mengambil kata-kata yang penting dari hasil token sebelumnya.

Tabel 5. Hasil Stopword Removal

Before	After
'lokasi' 'strategis' 'luas' 'dan' 'asri' 'banyak' 'tersedia' 'kursi' 'baik' 'indoor' 'maupun' 'outdoor' 'ada' '2' 'kasir' 'yang' 'juga' 'di' 'indoor' 'dan' 'outdoor' 'harga' 'makanan' 'murah' 'dengan' 'porsi' 'kecil' 'cocok' 'untuk' 'yang' 'memang' 'porsi' 'kecil'	'lokasi' 'strategis' 'luas' 'asri' 'tersedia' 'kursi' 'baik' 'indoor' 'outdoor' 'kasir' 'indoor' 'outdoor' 'harga' 'makanan' 'murah' 'porsi' 'cocok' 'porsi'

3.2.5. Stemming

Tahapan *Stemming* berfungsi untuk mengubah kata-kata dalam teks menjadi bentuk dasarnya dengan cara menghilangkan kata akhiran atau awalnya yang tidak perlu. Tahapan ini membantu untuk memproses bahasa

alami dikarenakan sejumlah kata-kata memiliki makna dasar yang sama tetapi memiliki struktur kata yang berbeda.

Tabel 6. Hasil Stemming

Before					After					
'lokasi'	'strategis'	'luas'	'asri'	'tersedia'	'lokasi'	'strategis'	'luas'	'asri'	'sedia'	'kursi'
'kursi'	'baik'	'indoor'	'outdoor'	'kasir'	'baik'	'indoor'	'outdoor'	'kasir'	'indoor'	
'indoor'	'outdoor'	'harga'	'makanan'	'murah'	'outdoor'	'harga'	'makan'	'murah'	'porsi'	
'porsi'	'cocok'	'porsi'			'cocok'	'porsi'				

3.3. Wordcloud

Teks yang telah melewati *Pre-processing* akan dibuat visualisasi dengan menggunakan Wordcloud. Wordcloud merupakan metode visualisasi teks yang berfungsi menggambarkan kata-kata yang paling sering muncul dan menjadi topik utama dari ulasan Kedai Roti Ibu Saya. Semakin besar ukuran teks pada Wordcloud maka semakin tinggi frekuensi kata tersebut muncul dalam teks ulasan.



Gambar 3. Hasil wordcloud teks ulasan

Hasil visualisasi menggunakan Wordcloud pada **gambar 3.** menunjukkan bahwa kata yang paling sering muncul pada ulasan adalah kata “makan” diikuti dengan kata “enak”, “banget”, “tempat”, “roti”, dan “murah”. Sehingga dapat diartikan bahwa topik utama dalam ulasan tersebut adalah tentang pengalaman makan yang enak di tempat Kedai Roti Ibu Saya yang disertai harga roti yang murah.

Terdapat sejumlah kata lain yang sering disebutkan, seperti “suasana”, “pelayanan”, “menu” dan “harga”. Hal ini mengidentifikasi bahwa terdapat topik lain yang dibahas dalam ulasan. Untuk memastikan apakah opini masyarakat tentang tempat makan tersebut merupakan opini positif, netral, ataupun negatif maka dilakukan pelabelan data emosi menggunakan analisis sentimen dengan metode *lexicon-based*.

3.4. Pelabelan Data

Kumpulan teks yang telah melewati tahapan *preprocessing* perlu diberikan label hasil sentimen. Perhitungan hasil sentimen dilakukan menggunakan perhitungan skor berdasarkan kamus kata *lexicon*. Adapun rumus perhitungan skor sentimen dengan mengurangi skor jumlah kata positif dengan skor jumlah kata negatif (4).

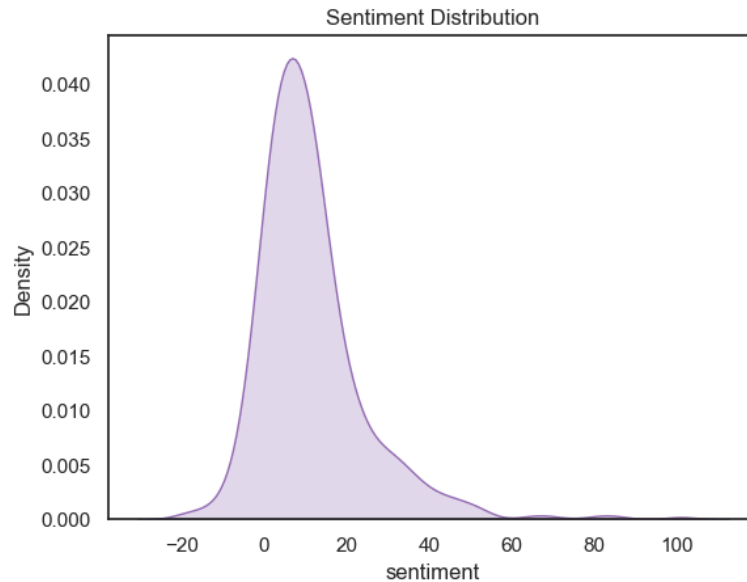
$$Skor = (jumlah\ kata\ positif) - (jumlah\ kata\ negatif) \quad (4)$$

Hasil dari perhitungan skor sentimen ulasan ditunjukkan pada **Tabel 7**.

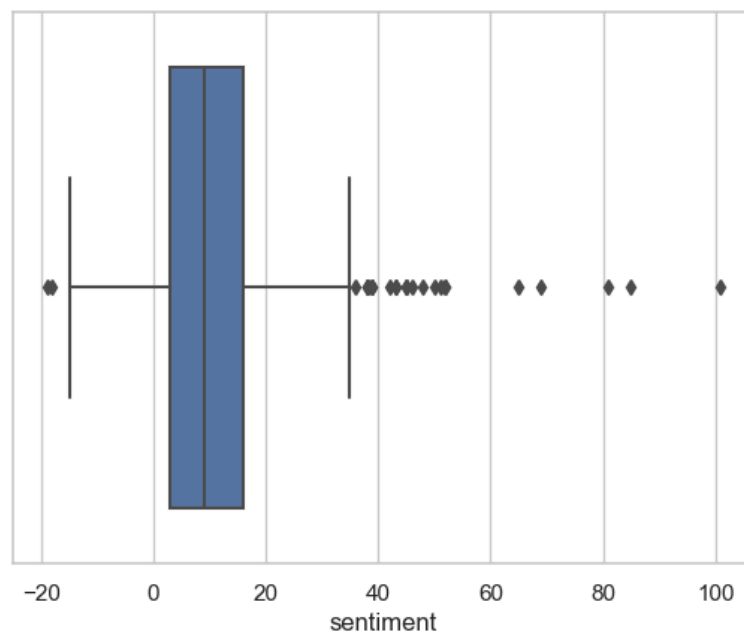
Tabel 7. Hasil skor sentimen

Teks ulasan	Skor Sentimen
lokasi strategis luas asri sedia kursi indoor outdoor kasir indoor outdoor harmakan murah...	42
hargin worth it sih tempat vintage gitu kesini jam an dapet ruang pokok rekomendasi pizzasih wajib garlic bread enak tapi mantap hangat	11

udah x kesini tutup alhamdulillah bukaa konsep klasik tua kalo dateng siang 35
rada panas ruang ac tp krn pintu arah taman bukak ada makan harmurah
banget jam siang sampe jam sore kenyang banget makan sampe ribu
bungkus pizza...



Gambar 4. Hasil KDE skor sentimen



Gambar 5. Hasil Boxplot skor sentimen

Pada visualisasi hasil distribusi sentimen pada **Gambar 4.** dapat dinyatakan bahwa distribusi kepadatan skor sentimen terletak pada skor 0 - 20. Pada hasil *boxplot* pada **Gambar 5.** terdapat skor sentimen yang melebihi skor 100 dan terdapat juga skor yang memiliki nilai negatif hampir menyentuh -20.

3.5. Hasil Analisis Sentimen

Skor sentimen yang telah didapatkan berdasarkan perhitungan skor terhadap kamus *lexicon* akan melewati proses pelabelan sentimen berdasarkan skor yang diperoleh. Pelabelan dibagi menjadi tiga bagian yaitu:

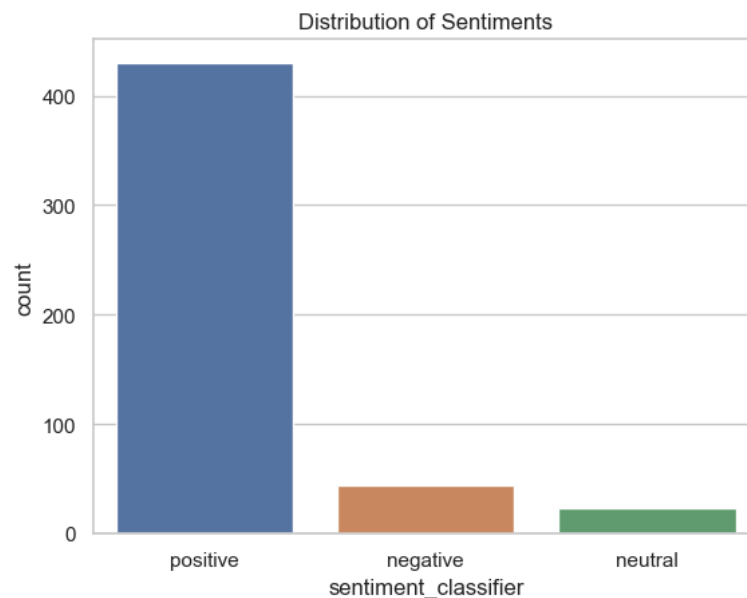
- Sentimen Positif: memiliki skor > 1 .
- Sentimen Netral: memiliki skor $= 0$.
- Sentimen Negatif: memiliki skor < 0 .

Setelah semua tahapan analisis sentimen dilakukan, diperoleh hasil pelabelan kelas sentimen pada 497 data ulasan Kedai Roti Ibu Saya pada aplikasi Google Maps adalah sebagai berikut:

Tabel 8. Hasil pelabelan sentimen pada data

Sentimen	Total	Persentase
Positif	430	86,52%
Netral	44	8,8%
Negatif	23	4,6%

Dari hasil analisis sentimen yang diperoleh pada **Tabel 8.** dapat dinyatakan bahwa data teks memiliki sentimen positif, netral dan negatif. Hasil sentimen didominasi oleh sentimen positif dengan total 430 data teks, lalu disusul dengan sentimen netral dengan total 44 data teks dan sentimen negatif dengan total 23 data teks. Sehingga dapat disimpulkan bahwa opini pengunjung pada ulasan aplikasi Google Maps terhadap Kedai Roti Ibu Saya menunjukkan opini yang positif. Adapun visualisasi yang diperoleh dari hasil pelabelan kelas sentimen ditunjukkan pada **Gambar 6.**



Gambar 6. Visualisasi hasil sentimen

3.6. Oversampling

Dalam menangani masalah ketidakseimbangan data dalam kategori sentimen, maka dilakukan teknik *oversampling* pada data *training* untuk mengurangi adanya bias saat memproses data *training*. *Oversampling* dilakukan untuk menambahkan sampel pada kelas minoritas sehingga distribusi kelas menjadi lebih seimbang.

Tabel 9. Hasil Oversampling

Sentimen	Jumlah Ulasan Sebelum Oversampling	Jumlah Ulasan Sesudah Oversampling
Positif	430	430
Negatif	44	430
Netral	23	430

3.7. Data Training Naive Bayes

Data ulasan yang telah melalui proses *oversampling* kemudian dibagi menjadi data *training* dan data *testing* dengan rasio 80% untuk data *training* dan 20% untuk data *testing*. Algoritma yang digunakan untuk membangun model klasifikasi sentimen merupakan algoritma Naive Bayes.

Tabel 10. Hasil Split Data

Set Data	Jumlah Ulasan
Training	344
Testing	86

3.8. Hasil Evaluasi Training Naive Bayes

Setelah model Naive Bayes dilatih, performa model dievaluasi menggunakan data *testing*. Evaluasi ini mengukur metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengidentifikasi seberapa baik model dalam mengklasifikasikan data.

Tabel 11. Hasil Evaluasi Training Naive Bayes

Metrik	Precision	Recall	F1-Score
Negative	0.98	0.99	0.99
Neutral	0.98	1.00	0.99
Positive	0.99	0.97	0.98
Accuracy			0.98

Berdasarkan hasil di atas, model menunjukkan performa yang sangat baik dengan akurasi keseluruhan sebesar 0.98. *Precision*, *recall*, dan *F1-score* untuk masing-masing kelas juga sangat tinggi, menunjukkan bahwa model mampu mengklasifikasikan data dengan sangat akurat.

Confusion matrix untuk hasil evaluasi model Naive Bayes dapat dilihat pada **Tabel 12** berikut ini:

Tabel 12. Confusion Matrix

	Prediksi Positif	Prediksi Negatif	Prediksi Netral
Aktual Positif	170	3	3
Aktual Negatif	2	169	0
Aktual Netral	0	0	169

Confusion matrix yang disajikan pada **Tabel 12.** menunjukkan kinerja model klasifikasi terhadap dataset uji yang terdiri dari tiga kelas, yaitu positif, negatif, dan netral. Hasil analisis menunjukkan bahwa model memiliki akurasi yang tinggi dengan beberapa kesalahan klasifikasi yang minimal.

Pada kelas positif, model berhasil mengidentifikasi 170 *instance* dengan benar (*True Positives*), sementara terdapat 2 *instance* yang salah diklasifikasikan sebagai negatif (*False Negatives*) dan masing-masing 3 *instance* yang salah diklasifikasikan sebagai negatif dan netral (*False Positives*). Untuk kelas negatif, model mengidentifikasi 169 *instance* dengan benar (*True Negatives*), dengan 3 *instance* yang salah diklasifikasikan sebagai positif (*False Positives*) dan tidak ada yang salah diklasifikasikan sebagai netral (*False Negatives*). Kinerja terbaik terlihat pada kelas netral, di mana model mengidentifikasi semua 169 *instance* dengan benar (*True Neutrals*) tanpa adanya kesalahan klasifikasi (*False Positives* dan *False Negatives*).

Secara keseluruhan, model ini menunjukkan performa yang sangat baik dalam mengklasifikasikan data uji dengan akurasi tinggi, terutama pada kelas netral yang tidak memiliki kesalahan klasifikasi sama sekali. Hasil ini mengindikasikan bahwa model mampu menangani dataset dengan variasi kelas yang berbeda secara efektif.

4. Kesimpulan

Penelitian ini berfokus pada implementasi *Lexicon-Based Sentiment Analysis* dan algoritma Multinomial Naive Bayes untuk menganalisis opini pengunjung Kedai Roti Ibu Saya yang diekstraksi dari ulasan di Google Maps. Dalam rangka mendapatkan hasil yang akurat, penelitian ini melibatkan pengumpulan sekitar 500 ulasan yang diproses melalui beberapa tahap penting.

Tahap pertama dalam penelitian ini adalah *text preprocessing*, yang melibatkan pembersihan data teks dengan menghilangkan tanda baca, angka, dan karakter khusus. Tahap berikutnya adalah pembobotan kata menggunakan

teknik TF-IDF untuk menentukan bobot kata dalam ulasan. Setelah pembobotan, data diklasifikasikan dengan algoritma Multinomial Naive Bayes, yang dipilih karena kemampuannya mengelola data teks dengan banyak fitur secara efisien. Hasil klasifikasi kemudian dievaluasi untuk menentukan apakah setiap ulasan termasuk dalam kategori positif, netral, atau negatif.

Dari hasil analisis, ditemukan bahwa mayoritas ulasan yang dianalisis mengandung sentimen positif, dengan rincian 430 ulasan positif, 44 ulasan netral, dan 23 ulasan negatif. Temuan ini menunjukkan bahwa Kedai Roti Ibu Saya mendapatkan tanggapan yang sangat baik dari pengunjungnya. Lebih lanjut, model klasifikasi sentimen yang dikembangkan menggunakan algoritma Multinomial Naive Bayes menunjukkan performa yang sangat baik dengan akurasi keseluruhan mencapai 98%. Selain itu, metrik evaluasi lainnya seperti *precision*, *recall*, dan *F1-score* juga menunjukkan nilai yang tinggi untuk semua kelas sentimen, mengindikasikan bahwa model ini mampu mengidentifikasi sentimen dengan tingkat kesalahan yang sangat rendah.

Secara keseluruhan, penelitian ini membuktikan efektivitas kombinasi *Lexicon-Based Sentiment Analysis* dan Multinomial Naive Bayes dalam mengolah dan menganalisis data ulasan teks dari platform seperti Google Maps. Implementasi metode ini tidak hanya memberikan gambaran yang akurat tentang opini pengunjung, tetapi juga dapat digunakan sebagai alat yang berguna bagi manajemen Kedai Roti Ibu Saya untuk memahami dan meningkatkan pengalaman pelanggan mereka di masa depan. Dengan demikian, penelitian ini memberikan kontribusi yang signifikan dalam bidang analisis sentimen serta aplikasi praktisnya dalam industri kuliner.

5. Referensi

- [1] Prakoso, Cahyo, and Arief Hermawan. "Perbandingan Model Machine Learning dalam Analisis Sentimen Ulasan Pengunjung Keraton Yogyakarta pada Google Maps." *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 3, 2023, pp. 1292-1302, doi.org/10.30865/klik.v4i3.1419.
- [2] Catelli, Rosario, et al. "Lexicon-Based vs. Bert-Based Sentiment Analysis: A Comparative Study in Italian." *Electronics*, vol. 11, no. 3, 2022, p. 374, doi.org/10.3390/electronics11030374.
- [3] Koto, Fajri, and Gemala Y. Rahmaningtyas. "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs." *2017 International Conference on Asian Language Processing (IALP)*, 2017, pp. 391-394, doi.org/10.1109/IALP.2017.8300625.
- [4] Martua, Evan. "Lexicon-Based Sentiment Analysis of COVID-19 Tweets in Indonesia." *Github*, 2020, <https://github.com/evanmartua34/Twitter-COVID19-Indonesia-Sentiment-Analysis---Lexicon-Based>.
- [5] Balakrishnan, Vimala, and Wandeep Kaur. "String-based Multinomial Naïve Bayes for Emotion Detection among Facebook Diabetes Community." *Procedia Computer Science*, vol. 159, 2019, pp. 30-37, doi.org/10.1016/j.procs.2019.09.157.
- [6] Octaviani, Niken L., et al. "Comparison of Multinomial Naïve Bayes Classifier, Support Vector Machine, and Recurrent Neural Network to Classify Email Spams." *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2020, pp. 17-21, doi.org/10.1109/iSemantic50169.2020.9234296.
- [7] Nainggolan, Hollywrith T., et al. "Klasifikasi Sentimen Menggunakan Algoritma K-Nearest Neighbour (Studi Kasus: Magang Merdeka Belajar)." *Informatik : Jurnal Ilmu Komputer*, vol. 19, no. 1, 2023, pp. 50-57, doi.org/10.52958/iftk.v19i1.4777.