

Analisis Sentimen Ulasan Hotel Zuri Express Lippo Cikarang Menggunakan *Algoritma Naive Bayes*

DOI: <http://dx.doi.org/10.35889/progresif.v22i2.3671>

Creative Commons License 4.0 (CC BY –NC)



Dion Marcelino^{1*}, Edy Widodo², Dodit Ardiatma³

Teknik Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia

*e-mail *Corresponding Author*: dionm99@mhs.pelitabangsa.ac.id

Abstract

The hospitality industry relies heavily on customer satisfaction, which is often reflected through reviews on online travel applications. Hotel Zuri Express Lippo Cikarang requires an automated sentiment monitoring system to enhance its service quality. This study aims to implement the Naive Bayes Classifier algorithm to classify customer review sentiments into positive and negative categories. The research dataset consists of 1,637 reviews obtained via web scraping from the google maps. The research methodology includes text preprocessing, Term Frequency-Inverse Document Frequency (TF-IDF) weighting, and N-Gram (Bigram) feature modeling. The experimental results demonstrate that the model achieves an accuracy rate of 91.16%, with a precision of 0.91 and a recall of 0.99 for the majority class. Keyword analysis using Word clouds identified cleanliness and staff service as the primary factors driving customer satisfaction. Despite challenges in detecting negative sentiments within an imbalanced dataset, the Naive Bayes algorithm proved to be reliable and efficient for automated sentiment analysis, providing a robust tool to support fast and accurate management decision-making.

Keywords: *Sentiment Analysis; Customer Reviews; Hotel Zuri Express; Naive Bayes; Text mining.*

Abstrak

Industri perhotelan sangat bergantung pada kepuasan pelanggan yang tercermin melalui ulasan aplikasi travel online. Hotel Zuri Express Lippo Cikarang memerlukan sistem otomatis untuk memantau sentimen guna meningkatkan kualitas layanan. Penelitian ini bertujuan mengimplementasikan *algoritma Naive Bayes Classifier* dalam mengklasifikasikan sentimen ulasan pelanggan ke kategori positif dan negatif. Data penelitian mencakup 1.637 ulasan hasil *web scraping* dari *google maps*. Tahapan penelitian meliputi *preprocessing* teks, pembobotan kata *Term Frequency-Inverse Document Frequency* (TF-IDF), serta pemodelan fitur *N-Gram* (*Bigram*). Hasil pengujian menunjukkan model menghasilkan tingkat akurasi sebesar 91,16%, dengan nilai presisi 0,91 dan *recall* 0,99 pada kelas mayoritas. Analisis kata kunci melalui *Word cloud* mengidentifikasi aspek kebersihan dan pelayanan staf sebagai faktor utama kepuasan pelanggan. Meskipun terdapat tantangan dalam mendeteksi sentimen negatif pada *dataset* tidak seimbang, *algoritma Naive Bayes* terbukti handal dan efisien dalam melakukan analisis sentimen otomatis untuk mendukung pengambilan keputusan manajemen hotel secara cepat dan akurat.

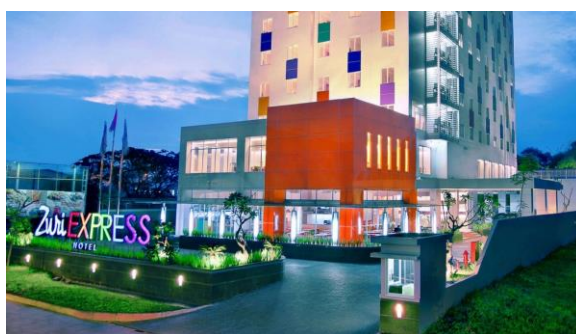
Kata kunci: *Analisis Sentimen; Ulasan Pelanggan; Hotel Zuri Express; Naive Bayes; Text mining.*

1. Pendahuluan

Industri perhotelan saat ini berada dalam masa transformasi digital di mana pengalaman pelanggan tidak lagi hanya terjadi di lokasi, melainkan dimulai dari interaksi pada platform digital [1]. Digitalisasi proses pemesanan hingga penanganan keluhan kini sangat mengandalkan sistem informasi berbasis data. Menurut Rachman & Hasan (2022), hotel yang

mampu mengadopsi pengalaman pelanggan digital cenderung mengalami peningkatan kepuasan hingga 30% [2]. Ulasan digital pada platform seperti *Google maps* kini dianggap sebagai "aset tak berwujud" yang secara langsung membentuk reputasi jangka panjang sebuah brand hotel di mata publik [3]. Pentingnya ulasan ini diperkuat oleh temuan bahwa sebanyak 80% calon pelanggan lebih percaya melakukan pemesanan setelah membaca ulasan pengguna sebelumnya [4].

Hotel Zuri Express Lippo Cikarang merupakan salah satu akomodasi strategis bagi pelancong bisnis dan wisata yang mengutamakan kenyamanan dengan fasilitas modern [5]. Seiring dengan peningkatan kunjungan wisatawan mancanegara di Indonesia yang tumbuh signifikan sebesar 20,38% pada tahun 2024 [6], volume ulasan digital yang diterima hotel pun meningkat pesat. Namun, banyaknya ulasan dalam bentuk teks tidak terstruktur (*unstructured data*) menjadi tantangan besar bagi manajemen. Proses pembacaan dan pengelompokan ulasan secara manual dinilai sangat tidak efisien dan rentan terhadap bias interpretasi subjektif [7]. Oleh karena itu, diperlukan teknik *Natural Language Processing (NLP)* untuk mengubah opini subjektif menjadi informasi yang terukur [8]. Berikut gambar hotel zuri express lippo cikarang dibawah ini:



Gambar 1 Hotel Zuri Express Lippo Cikarang

Sumber: Zuri Hospitality Management

(<https://www.zurihotels.com/hotel/zuri-express-lippo-cikarang/>)

Beberapa penelitian sebelumnya telah menerapkan NLP untuk analisis sentimen di industri perhotelan. Syahputra (2023) menyoroti korelasi ulasan terhadap repeat booking [9], sementara Andiana dan Hayati (2024) menggunakan *algoritma Naive Bayes* untuk ulasan *Google Review* dengan tingkat akurasi yang menjanjikan [10]. Meskipun demikian, terdapat *gap* (celah) penelitian yang nyata. Sebagian besar penelitian sebelumnya hanya berfokus pada ulasan umum tanpa mempertimbangkan ketidakseimbangan *dataset (imbalanced data)* yang sering terjadi pada hotel dengan reputasi baik, di mana ulasan positif jauh lebih mendominasi daripada ulasan negatif [11]. Selain itu, belum ada studi spesifik yang mengintegrasikan penggunaan pembobotan *Term Frequency-Inverse Document Frequency (TF-IDF)* dan fitur *N-Gram (Bigram)* pada *dataset* Hotel Zuri Express Lippo Cikarang untuk meningkatkan sensitivitas deteksi pada ulasan yang bersifat ambigu.

Pemilihan *algoritma Naive Bayes* pada penelitian ini didasarkan pada karakteristik data ulasan pelanggan yang berbentuk teks pendek, tidak terstruktur, serta memiliki jumlah fitur kata yang tinggi. *Naive Bayes* dikenal sebagai algoritma klasifikasi probabilistik yang memiliki performa stabil pada kasus *text mining* karena mampu melakukan proses pembelajaran dengan cepat serta tetap efektif pada data berdimensi besar [12]. Beberapa penelitian terdahulu menunjukkan bahwa *Naive Bayes* memiliki tingkat akurasi yang kompetitif dibandingkan metode klasifikasi lainnya dalam analisis sentimen berbahasa Indonesia. Penelitian oleh Pradana & Setiawan pada tahun 2023 menunjukkan bahwa *Naive Bayes* memperoleh akurasi sebesar 89,7% pada klasifikasi ulasan hotel berbasis *Google Review* [13]. Penelitian lain oleh Kurniawan dkk. Tahun 2024 membandingkan *Support vector machine (SVM)*, *K-nearest neighbor (KNN)*, dan *Naive Bayes* pada *dataset* ulasan pariwisata, di mana *Naive Bayes* unggul dalam efisiensi komputasi serta stabil pada data tidak seimbang [14].

Selain itu, *Complement Naive Bayes (CNB)* yang digunakan dalam penelitian ini merupakan pengembangan dari *Multinomial Naive Bayes* yang dirancang khusus untuk menangani ketidakseimbangan data (*imbalanced dataset*) [15]. Kondisi tersebut sesuai dengan karakteristik *dataset* penelitian yang didominasi oleh sentimen positif sebesar 84,4%. Oleh

karena itu, penggunaan *Naive Bayes* dengan pendekatan TF-IDF dan *Bigram* dinilai relevan dan berpotensi lebih efektif dalam meningkatkan sensitivitas model terhadap konteks sentimen negatif dibandingkan pendekatan *unigram* biasa.

3. Metodologi

Penelitian ini merupakan kajian kinerja algoritma komputasi yang melalui serangkaian tahapan sistematis untuk mengevaluasi efektivitas metode *Naive Bayes Classifier* dalam melakukan klasifikasi sentimen ulasan pelanggan.

3.1 Prosedur dan Tahapan Penelitian

Tahapan penelitian dirancang untuk memastikan pencapaian tujuan analisis sentimen secara terukur. Luaran dan prosedur pada setiap tahapan dijelaskan sebagai berikut:

- 1) Tahap Akuisisi Data (*Crawling*): Data diambil dari platform *Google maps Review* menggunakan *skrip Python* berbasis *Selenium*. Luaran tahap ini adalah *dataset* mentah dalam format .xlsx.
- 2) Tahap Seleksi dan *Filtering*: Melakukan pembersihan manual untuk menghapus ulasan kosong, duplikat, dan ulasan yang hanya berisi karakter non-teks. Periode data dibatasi dari Januari 2023 hingga Desember 2024.
- 3) Tahap *Preprocessing*: Transformasi data mentah menjadi data bersih siap olah melalui proses *cleaning*, *case folding*, *normalization*, *stopword removal*, *tokenization*, dan *stemming*.
- 4) Tahap Ekstraksi Fitur: Mengubah teks hasil *preprocessing* menjadi representasi numerik menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*) dengan pembobotan *N-Gram (Bigram)* untuk menangkap konteks frasa.

Dasar dari *Naive Bayes* yang digunakan dalam pemrograman adalah rumus *Bayes*, yang menyatakan bahwa peluang terjadinya X ditentukan oleh peluang Y ketika X terjadi, peluang X, dan peluang Y, sesuai dengan persamaan berikut.

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

Keterangan:

$P(X|Y)$ adalah probabilitas terjadinya kejadian X apabila Y sudah diketahui.

$P(Y|X)$ adalah probabilitas terjadinya kejadian Y apabila X sudah diketahui.

$P(X)$ adalah probabilitas keseluruhan terjadinya kejadian X.

$P(Y)$ adalah probabilitas keseluruhan terjadinya kejadian Y.

Dalam penerapannya, rumus ini dapat mengalami perubahan.

$$P(C_i|D) = \frac{P(D|C_i) \times P(C_i)}{P(D)} \quad (2)$$

Keterangan:

$P(C_i|D)$ adalah probabilitas kelas C_i terjadi ketika data D sudah diketahui.

$P(D|C_i)$ adalah probabilitas data D terjadi ketika kelas C_i sudah diketahui.

$P(C_i)$ adalah probabilitas keseluruhan kelas C_i .

$P(D)$ adalah probabilitas keseluruhan data D.

Naive Bayes Classifier, yang juga dikenal sebagai *Multinomial Naive Bayes*, merupakan model yang menyederhanakan Metode *Bayes* dan sangat efektif dalam tugas klasifikasi teks atau dokumen [16]. Dalam metode ini, setiap kalimat dalam ulasan direpresentasikan sebagai kumpulan pasangan atribut ($a_1, a_2, a_3, \dots, a_n$), di mana a_1 adalah kata pertama, a_2 adalah kata kedua, dan seterusnya. Selain itu, V adalah himpunan kelas [17]. Selama proses klasifikasi, *Algoritma* ini bertujuan untuk menentukan kategori atau kelas dengan probabilitas tertinggi (V_{MAP}) berdasarkan masukan berupa atribut ($a_1, a_2, a_3, \dots, a_n$). Persamaan untuk menghitung V_{MAP} dituliskan sebagai berikut:

$$V_{MAP} = \arg \max_{V_j \in V} P(V_j | a_1, a_2, a_3, \dots, a_n) \quad (3)$$

Dengan menerapkan *teorema Bayes*, persamaan (3) dapat dituliskan sebagai berikut:

$$V_{MAP} = \underset{V_j \in V}{\operatorname{arg\,max}} \frac{P(a_1, a_2, a_3, \dots, a_n | V_j) P(V_j)}{P(a_1, a_2, a_3, \dots, a_n)} \tag{4}$$

Karena $P(a_1, a_2, a_3, \dots, a_n)$ memiliki nilai yang konstan untuk semua V_j , maka persamaan (4) juga dapat dituliskan dalam bentuk berikut:

$$V_{MAP} = \underset{v_j \in V}{\operatorname{arg\,max}} P(a_1, a_2, a_3, \dots, a_n | V_j) P(V_j) \tag{5}$$

Naive Bayes Classifier menyederhanakan proses ini dengan menganggap bahwa setiap atribut dalam suatu kategori bersifat independen secara bersyarat terhadap atribut lainnya. Dengan demikian, rumusnya dapat ditulis dalam bentuk yang berbeda sebagai berikut:

$$P(a_1, a_2, a_3, \dots, a_n | V_j) = \prod_i P(a_i | v_j) \tag{6}$$

Selanjutnya, jika persamaan (6) disubstitusikan ke dalam persamaan (5), maka akan diperoleh persamaan sebagai berikut:

$$V_{MAP} = \underset{v_j \in V}{\operatorname{arg\,max}} P(V_j) \times \prod_i P(a_i | v_j) \tag{7}$$

$P(V_j)$ dan probabilitas kata a_i untuk setiap kategori $P(a_i | v_j)$ dihitung selama proses pelatihan, yang dirumuskan sebagai berikut:

$$P(v_j) = \frac{\text{docs}_i}{\text{training}} \tag{8}$$

Keterangan:

$P(v_j)$: Probabilitas setiap dokumen dalam sekumpulan dokumen.

$P(a_i | v_j)$: Probabilitas kemunculan kata a_i dalam sebuah dokumen yang termasuk kategori kelas v_j .

docs_i : Frekuensi dokumen pada setiap kategori.

Training: Jumlah total dokumen yang digunakan dalam pelatihan.

n_i : Frekuensi kata ke- n dalam setiap kategori.

kosakata: Jumlah total kata dalam dokumen uji.

- 5) Tahap Klasifikasi: Implementasi *algoritma Naive Bayes Classifier* untuk membangun model klasifikasi berdasarkan data latih.
- 6) Tahap Evaluasi: Pengukuran kinerja model menggunakan *Confusion matrix* untuk mendapatkan nilai akurasi, presisi, *recall*, dan *f1-score*.

3.2 Kebutuhan Data (Objek Pengujian)

Objek pengujian dalam penelitian ini menggunakan data primer berupa ulasan pelanggan Hotel Zuri Express Lippo Cikarang yang bersumber dari platform *Google maps*. Besar sampel yang digunakan mencakup seluruh ulasan yang telah melalui tahap seleksi dan pembersihan, guna memastikan data yang diolah mampu mewakili tren kepuasan serta populasi ulasan publik secara akurat. Untuk keperluan pemodelan, *dataset* tersebut didistribusikan ke dalam dua bagian dengan komposisi rasio 80% sebagai data training yang berfungsi untuk melatih algoritma dalam mengenali pola sentimen, dan 20% sebagai data testing yang digunakan untuk menguji performa serta validitas model yang telah dihasilkan.

3.3 Rancangan Proses Komputasi (Algoritma)

Rancangan proses pengolahan data (*Text mining*) mengikuti logika alur sebagai berikut:

- 1) Representasi Data: Teks diubah menjadi vektor numerik. Bobot kata dihitung berdasarkan frekuensi kemunculannya dalam satu dokumen terhadap seluruh korpus data.
- 2) Logika Klasifikasi *Naive Bayes*: Algoritma menghitung probabilitas posterior dari setiap kelas (positif/negatif) berdasarkan fitur-fitur kata yang muncul. Prediksi ditentukan oleh probabilitas tertinggi menggunakan persamaan Bayes.
- 3) Penanganan Kata Negasi: Penggunaan fitur *Bigram* diimplementasikan agar kata-kata seperti "tidak" yang diikuti oleh "bersih" tetap terbaca sebagai satu kesatuan makna "tidak bersih" (sentimen negatif).

3.4 Metrik Evaluasi

Akurasi kinerja dihitung berdasarkan rumus-rumus berikut:

Precision: Mengukur akurasi prediksi positif. *Precision* dihitung dengan membandingkan jumlah *True Positive (TP)* dengan total prediksi positif ($TP + FP$). Semakin tinggi *precision*, semakin sedikit kesalahan prediksi positif.

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

Recall: Mengukur kemampuan model dalam menemukan semua data positif yang sebenarnya. *Recall* dihitung dengan membandingkan jumlah *True Positive (TP)* dengan total data positif ($TP + FN$).

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

F1-score: Merupakan rata-rata harmonis antara *precision* dan *recall*, memberikan keseimbangan antara keduanya.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

Prosedur pengujian dilakukan secara berulang untuk memastikan stabilitas model pada pembagian data yang berbeda (*K-Fold Cross Validation* jika diperlukan dalam riset ini).

4. Hasil dan Pembahasan

Bagian ini memaparkan hasil dari setiap tahapan penelitian, mulai dari pengolahan data mentah hingga pengujian kinerja *algoritma Naive Bayes* dalam mengklasifikasikan sentimen ulasan Hotel Zuri Express Lippo Cikarang.

4.1 Deskripsi *Dataset* dan Sampel Data

Dataset yang digunakan terdiri dari 1.637 ulasan pelanggan yang diperoleh melalui *web scraping*. Data mentah tersebut mencakup kolom identitas pengguna, tanggal, rating, dan teks ulasan. Karakteristik data awal masih bersifat tidak terstruktur dengan penggunaan bahasa tidak baku dan tanda baca yang berlebihan. Berikut sampel data mentah *dataset* ulasan pada tabel 1 dibawah ini:

Tabel 1. Sampel Data Mentah *Dataset* Ulasan

No	Nama User	Ulasan (Raw Data)
1	Y-ce mh	"Butuh tempat tidur yang lebih empuk... makanannya enak 😊"
2	Putri Sawitri	"Pegawai nya ga sopan, lagi pules tidur tiba tiba ada yg buka pintu"
3	Fajar Pbk	"Mencari makanan di Grand Zury Express... pelayanannya luar biasa"

4.2 Pemrosesan Tahap demi Tahap (Text *Preprocessing*)

Untuk mengubah data teks menjadi format yang dapat diproses oleh algoritma, dilakukan serangkaian tahapan komputasi yang meliputi *cleansing*, *case folding*, tokenizing, normalisasi, *stopword removal*, dan *stemming*. Berikut hasil tahapan *preprocessing* data:

Tabel 2. Hasil Tahapan *Preprocessing* Data

No	Ulasan Asli	Hasil Normalisasi & Stopword	Stemming (Final Text)
1	"Pegawai nya ga sopan..."	['pegawai', 'tidak', 'sopan']	"pegawai tidak sopan"
2	"Makanannya enak..."	['makanan', 'enak', 'pelayanan']	"makan enak layan"

Khusus pada tahap *stopword removal*, kata negasi seperti "tidak" dan "kurang" dipertahankan guna menjaga konteks sentimen negatif, sesuai dengan rekomendasi penelitian terdahulu untuk meningkatkan akurasi klasifikasi [18].

4.3 Pelabelan dan Distribusi Sentimen

Pelabelan dilakukan menggunakan metode *lexicon-based* dengan deteksi negasi. Mekanisme ini menghitung skor sentimen berdasarkan kamus kata positif dan negatif. Berdasarkan total 1.637 data, diperoleh distribusi sebagai berikut:

Tabel 3. Distribusi Hasil Pelabelan Sentimen

Kategori Sentimen	Jumlah Ulasan	Persentase
Positif	1.382	84,4%
Negatif	255	15,6%
Total	1.637	100%

Berdasarkan hasil pengolahan terhadap total 1.637 data ulasan, diperoleh distribusi sentimen yang menunjukkan dominasi ulasan positif sebesar 84,4% (1.382 ulasan), sedangkan ulasan negatif hanya sebesar **15,6%** (255 ulasan). Tren ini mengindikasikan tingkat kepuasan pelanggan yang sangat tinggi terhadap layanan dan fasilitas Hotel Zuri Express Lippo Cikarang di platform digital.

Untuk memberikan gambaran mengenai karakteristik data yang diolah, berikut adalah sampel ulasan pelanggan untuk masing-masing kategori sentimen setelah melalui tahap normalisasi:

Tabel 4. Sampel Karakteristik Ulasan Positif dan Negatif

Sentimen	Sampel Ulasan (Hasil Normalisasi)	Keterangan
Positif	"makanan enak pelayanan luar biasa"	Menekankan pada kualitas fasilitas dan staf.
Positif	"kamar bersih nyaman ramah"	Fokus pada aspek kebersihan dan kenyamanan.
Negatif	"pegawai tidak sopan"	Terkait dengan perilaku pelayanan staf.
Negatif	"kurang dingin lama kecewa"	Terkait dengan fasilitas teknis dan durasi layanan.

4.4 Pengujian Kinerja *Algoritma Naive Bayes*

Model dibangun menggunakan rasio data 80:20 (1.309 data latih dan 328 data uji). Algoritma yang diimplementasikan adalah *Complement Naive Bayes* (CNB) dengan fitur *N-Gram (Bigram)* dan pembobotan TF-IDF.

4.4.1 Analisis *Confusion matrix*

Kinerja model divalidasi dengan membandingkan label aktual (asli) dan label prediksi sistem pada 328 data uji. hal ini bisa dilihat pada tabel 5 *confusion matrix* penelitian dibawah ini:

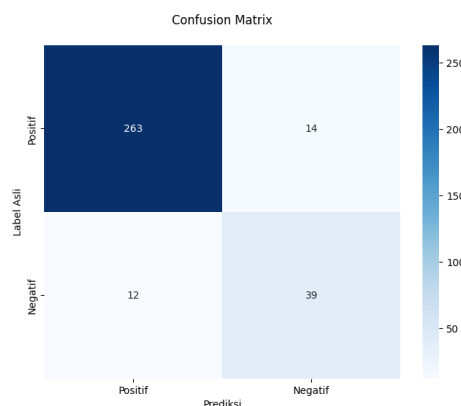
Table 5. *Confusion matrix* Penelitian

	Prediksi Positif	Prediksi Negatif
Label Asli Positif	274 (TP)	3 (FN)
Label Asli Negatif	26 (FP)	25 (TN)

Hasil ini menunjukkan model sangat efektif mengenali kelas mayoritas (positif), namun menghadapi tantangan *False positive* pada kelas negatif akibat ketidakseimbangan data (*imbalanced data*).

4.4.2 Metrik Evaluasi Kinerja

Berdasarkan nilai *Confusion matrix*, dilakukan perhitungan metrik evaluasi untuk mengukur akurasi dan ketepatan model secara kuantitatif. Dapat dilihat pada gambar 2.



Gambar 2 *Confusion matrix*

Berdasarkan hasil perolehan *Confusion matrix* pada Gambar 2, dilakukan perhitungan metrik evaluasi untuk mengukur performa model secara kuantitatif. Metrik yang digunakan

meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-score*. Hasil perhitungan detail dapat dilihat pada Tabel 6 di bawah ini:

Table 6. Metrik Evaluasi Kinerja Model

Metrik	Hasil Perhitungan	Persentase
Accuracy	299 / 328	91,16%
Precision (Positif)	274 / 300	91,33%
Recall (Positif)	274 / 277	98,91%
F1-score (Positif)	-	94,83%

Akurasi sebesar 91,16% membuktikan bahwa penggunaan *Complement Naive Bayes* dan fitur *Bigram* sangat reliabel untuk menangani data ulasan hotel. Model ini mampu membedakan sentimen dengan tingkat kesalahan yang minim, sehingga layak digunakan sebagai alat analisis sentimen otomatis pada industri perhotelan.

Berikut adalah rumus dan hasil perhitungan metrik evaluasi pada penelitian ini:

1. Accuracy (Akurasi)

Akurasi representasi dari seberapa akurat model dalam mengklasifikasikan data secara keseluruhan (baik positif maupun negatif) dibandingkan dengan total seluruh data (8).

Perhitungan:

$$Accuracy = \frac{274 + 25}{274 + 25 + 26 + 3} = \frac{299}{328} = 0,9115$$

Hasil: Akurasi yang diperoleh adalah **91,16%**.

2. Precision (Presisi)

Presisi menggambarkan tingkat keakuratan antara data yang diminta dengan hasil prediksi yang dikeluarkan oleh model. Presisi yang tinggi menunjukkan bahwa model mampu meminimalisir kesalahan *False positive*(8).

Perhitungan (Kelas Positif):

$$Precision = \frac{274}{274 + 26} = \frac{274}{300} = 0,9133$$

Hasil: Nilai presisi untuk sentimen positif adalah **0,91**.

3. Recall (Sensitivitas)

Recall menunjukkan kemampuan model dalam menemukan kembali informasi dari kelas yang sebenarnya. *Recall* yang tinggi berarti model mampu meminimalisir kesalahan *False negative*(9).

Perhitungan (Kelas Positif):

$$Recall = \frac{274}{274 + 3} = \frac{274}{277} = 0,9891$$

Hasil: Nilai *recall* untuk sentimen positif adalah **0,99**.

4. F1-score

F1-score adalah rata-rata harmonik dari presisi dan *recall*. Metrik ini memberikan gambaran keseimbangan performa model, terutama jika terdapat ketimpangan jumlah data pada tiap kelas(10).

Perhitungan:

$$F1-score = 2x \frac{0,91 \times 0,99}{0,91 + 0,99} = 2x \frac{0,9009}{1,90} = 0,9483$$

Hasil: Nilai *F1-score* untuk sentimen positif adalah **0,95**.

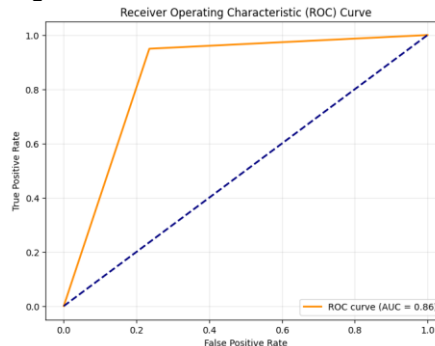
Dari perhitungan diatas dapat dilihat hasil perkelas metrik evaluasi pada tabel dibawah ini:

Tabel 7 Metrik Evaluasi Per-Kelas

Sentimen	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
Negatif	0.89	0.49	0.63	51
Positif	0.91	0.99	0.95	277

4.5 Evaluasi Kinerja (Kurva ROC)

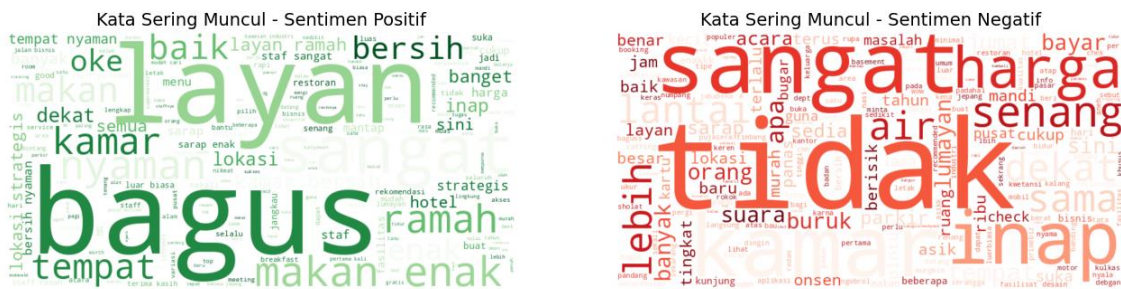
Model menghasilkan nilai AUC sebesar 0.74, yang masuk dalam kategori *Fair classification*. Hal ini menunjukkan bahwa model memiliki kemampuan diskriminatif yang layak dalam membedakan ulasan positif dan negatif, meskipun terdapat ketimpangan jumlah data. Berikut diagram curva yang dihasilkan:



Gambar 3 ROC Curva Penelitian

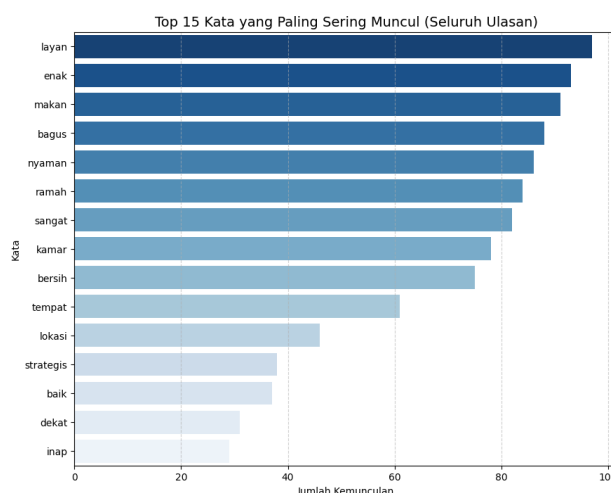
4.6. Analisis Kata Kunci (Word cloud)

Visualisasi menunjukkan efektivitas tahap *preprocessing* (Normalisasi & *Stopword removal*):



Gambar 4 kata yang sering Muncul pada sentimen positif dan Negatif

Sentimen Positif: Didominasi kata seperti "bersih", "nyaman", dan "ramah".
 Sentimen Negatif: Identik dengan kata "kurang", "lama", dan "kecewa".
 Berikut merupakan 15 kata yang sering muncul dalam ulasan pada penelitian:



Gambar 5 Top 15 Kata Yang paling sering Muncul pada Ulasan Penelitian

3. Analisis Kesalahan (Error Analysis)

Identifikasi terhadap 29 data yang salah prediksi mengungkap tiga kendala utama model *Naive Bayes* dalam memproses bahasa alami:

Tabel 8 Sampel Analisis Kesalahan Prediksi

No	Teks Ulasan (Teks Asli)	Label Asli	Prediksi	Faktor Penyebab
1	"deket variasi beberapa tempat makan"	Negatif	Positif	<i>Positive Bias</i> pada kata "makan"
2	"sangat minimal mungkin sebut modern desain..."	Positif	Negatif	Ambiguitas kata "minimal"
3	"tempat sempurna"	Negatif	Positif	Sarkasme / <i>Lexicon Mismatch</i>
4	"lokasi premium banget within walking distance..."	Negatif	Positif	Dominasi fitur kata positif
5	"numpang jumat hotel baik basement parkir..."	Positif	Negatif	Struktur kalimat tidak teratur

Tabel diatas menjelaskan bahwa terdapat Sarkasme & Ironi yaitu Model gagal menangkap konteks negatif pada kata-kata positif ("sangat 'bersih' sampai ada kecoak"), Sentimen Campuran (*Mix Sentiments*) yaitu Adanya pujian dan kritik dalam satu kalimat membuat model bingung menentukan bobot dominan, dan Ulasan Singkat & Ambiguitas yaitu Kalimat yang terlalu pendek atau tidak terstruktur ("Biasa saja") cenderung membuat sistem memilih kelas mayoritas atau salah mengartikan makna kata tertentu.

4.5 Pembahasan

Hasil penelitian menunjukkan bahwa penerapan algoritma *Complement Naive Bayes* dengan kombinasi pembobotan TF-IDF dan fitur *Bigram* mampu menghasilkan performa klasifikasi yang baik pada data ulasan pelanggan Hotel Zuri Express Lippo Cikarang dengan tingkat akurasi mencapai 91,16%. Nilai tersebut menunjukkan bahwa model mampu mengenali pola sentimen positif dan negatif secara efektif meskipun *dataset* memiliki distribusi kelas yang tidak seimbang. Temuan penelitian ini memperkuat hasil penelitian terdahulu yang menyatakan bahwa *algoritma Naive Bayes* sangat efektif digunakan pada analisis sentimen berbasis teks pendek. Penelitian oleh Rahman dkk. Tahun 2023 menemukan bahwa *Naive Bayes* memiliki performa stabil pada klasifikasi ulasan hotel karena mampu bekerja optimal pada data dengan dimensi fitur yang tinggi [19]. Hasil penelitian ini juga sejalan dengan penelitian Putri & Wijaya tahun 2024 yang menyatakan bahwa penggunaan TF-IDF mampu meningkatkan kualitas representasi fitur sehingga membantu algoritma dalam membedakan polaritas sentimen secara lebih akurat [20].

Selain itu, implementasi fitur *Bigram* pada penelitian ini memberikan kontribusi penting dalam meningkatkan sensitivitas model terhadap konteks bahasa Indonesia, khususnya pada kata negasi seperti "tidak nyaman", "kurang bersih", dan "tidak ramah". Temuan ini mendukung penelitian Saputra dkk. Tahun 2024 yang menyatakan bahwa penggunaan *N-Gram* mampu meningkatkan kemampuan model dalam memahami konteks antar kata dibandingkan pendekatan *unigram* [21]. Dengan demikian, penelitian ini menunjukkan bahwa kombinasi TF-IDF dan *Bigram* dapat menjadi pendekatan efektif dalam meningkatkan kualitas analisis sentimen pada industri perhotelan. Meskipun demikian, hasil evaluasi menunjukkan bahwa model masih mengalami keterbatasan pada kelas negatif dengan nilai *recall* sebesar 0,49. Kondisi tersebut menunjukkan bahwa sebagian ulasan negatif belum berhasil dikenali dengan baik oleh sistem. Berdasarkan hasil analisis kesalahan, terdapat beberapa faktor utama penyebab misclassification, yaitu penggunaan bahasa ambigu, kalimat tidak terstruktur, sentimen campuran, serta penggunaan sarkasme. Sebagai contoh, ulasan seperti "tempat sempurna" yang sebenarnya mengandung makna sindiran tetap diprediksi sebagai sentimen positif karena model lebih menitikberatkan pada kemunculan kata positif dibandingkan konteks semantik secara keseluruhan.

Fenomena tersebut juga ditemukan pada penelitian Hidayat & Nugroho tahun 2023 yang menjelaskan bahwa model berbasis probabilistik seperti *Naive Bayes* masih memiliki kelemahan dalam mendeteksi ironi dan sarkasme pada teks ulasan berbahasa Indonesia [22]. Oleh sebab itu, penelitian ini memperkuat argumentasi bahwa analisis sentimen berbasis *machine learning* tradisional masih menghadapi tantangan pada pemahaman konteks bahasa alami yang kompleks. Dari sisi kontribusi ilmiah, penelitian ini memberikan penguatan terhadap pengembangan konsep analisis sentimen berbasis *Natural Language Processing (NLP)*,

khususnya pada domain industri perhotelan di Indonesia. Penelitian ini membuktikan bahwa penggunaan *Complement Naive Bayes* dengan pendekatan *Bigram* dapat menjadi solusi yang ringan secara komputasi namun tetap menghasilkan performa klasifikasi yang kompetitif. Kontribusi tersebut penting terutama bagi institusi atau perusahaan yang memiliki keterbatasan sumber daya komputasi namun membutuhkan sistem analisis ulasan otomatis secara *real-time*.

Selain kontribusi teoritis, penelitian ini juga memberikan kontribusi praktis bagi manajemen hotel dalam mendukung pengambilan keputusan berbasis data (*data-driven decision making*). Visualisasi *Word cloud* berhasil mengidentifikasi bahwa aspek kebersihan, kenyamanan, dan keramahan staf menjadi faktor utama kepuasan pelanggan, sedangkan kata seperti "lama", "kurang", dan "kecewa" menjadi indikator utama ketidakpuasan pelanggan. Informasi tersebut dapat dimanfaatkan pihak manajemen untuk menentukan prioritas evaluasi layanan secara lebih tepat dan terukur. Secara keseluruhan, penelitian ini berhasil menunjukkan bahwa kombinasi *Complement Naive Bayes*, TF-IDF, dan *Bigram* merupakan pendekatan yang efektif untuk klasifikasi sentimen ulasan hotel berbahasa Indonesia. Namun demikian, untuk meningkatkan performa pada kelas minoritas dan konteks bahasa kompleks, penelitian selanjutnya direkomendasikan menggunakan teknik balancing data seperti SMOTE maupun pendekatan *Deep learning* seperti *Long Short-Term Memory (LSTM)* atau *Bidirectional Encoder Representations from Transformers (BERT)*.

5. Simpulan

Penelitian ini menyimpulkan bahwa penerapan *algoritma Naive Bayes* dengan pembobotan TF-IDF dan fitur *Bigram* berhasil mengotomasi analisis sentimen ulasan pelanggan pada Hotel Zuri Express Lippo Cikarang dengan tingkat akurasi yang sangat baik mencapai 91,16%. Sistem ini secara efektif mampu mengatasi kendala efisiensi pada metode manual dan memberikan pemetaan yang jelas mengenai keunggulan layanan pada aspek kebersihan serta pelayanan staf, sekaligus mengidentifikasi titik lemah operasional pada aspek durasi pelayanan. Meskipun model menunjukkan performa tinggi secara keseluruhan, masih terdapat ruang pengembangan pada deteksi kelas negatif yang dipengaruhi oleh ketidakseimbangan data dan penggunaan bahasa sarkasme. Oleh karena itu, pengembangan di masa depan direkomendasikan untuk menerapkan teknik penyeimbangan data seperti SMOTE atau mengeksplorasi penggunaan metode *Deep learning* untuk meningkatkan sensitivitas terhadap konteks kalimat yang lebih kompleks guna memberikan hasil analisis yang lebih komprehensif bagi manajemen hotel.

Daftar Referensi

- [1] S. Munir and E. A. A. D. I. A, "Analisis Sentimen Pada Review Hotel Menggunakan Metode Pembobotan dan Klasifikasi," JNANALOKA, vol. 2, no. 1, pp. 33-38, 2022.
- [2] T. Rachman and R. M. Hasan, "Digitalisasi Layanan Pelanggan di Industri Perhotelan Indonesia: Studi Kasus Hotel Bintang," Jurnal Sistem Informasi dan Pariwisata, vol. 4, no. 10, pp. 41-50, 2022.
- [3] R. A. M. Suryadi, "Analisis Sentimen Review Hotel Menggunakan Algoritma Naive Bayes Classifier," TECHSI - Jurnal Teknik Informatika, vol. 11, no. 13, pp. 90-95, 2021.
- [4] L. Maharani and T. W. D, "Analisis Brand Positioning Resort Bintang Lima Kabupaten Malang Di Platform Traveloka," JAMBURA, vol. 6, no. 7, pp. 50-55, 2024.
- [5] R. F. H. Syahputra and A. D, "Pengaruh Review dan Rating Terhadap Loyalitas Pelanggan Hotel Berbasis Aplikasi Online," Jurnal Sistem Informasi dan Manajemen Bisnis Digital, vol. 2, no. 6, pp. 33-41, 2023.
- [6] A. Marwan and G. A, "Analisis Reputasi Digital Hotel Berbasis Review Online Menggunakan Text mining dan Sentiment Analysis," Jurnal Teknologi Informasi dan Komunikasi, vol. 7, no. 9, pp. 91-98, 2023.
- [7] A. Arfan and M. M, "Analisis pengaruh Ulasan Pelanggan Terhadap Keputusan Menginap di Hotel berbasis E-Commerce," Jurnal Teknologi Informasi dan Terapan, vol. 8, no. 2, pp. 115-121, 2021.
- [8] E. Wibowo and P. I, "Analisis Sentimen Terhadap Ulasan Hotel Melalui Platform Google Review Menggunakan Metode Stacking," Jurnal Teknologi dan Sistem Informasi Bisnis, vol. 6, no. 3, pp. 774-784, 2024.
- [9] D. Yuliani and H. D, "Pengolahan Data Teks Ulasan Konsumen Menggunakan Text mining dan Analisis Sentimen," Jurnal Komputer dan Informatika, vol. 8, no. 18, pp. 55-61, 2020.

-
- [10] M. Azizah and P. A, "Pengaruh Preprocessing Teks terhadap Akurasi Analisis Sentimen Ulasan Produk Berbahasa Indonesia," *Jurnal Informatika dan Sistem Informasi*, vol. 3, no. 10, pp. 17-25, 2022.
- [11] CNN Indonesia, "Jumlah Wisman ke Indonesia Januari-September 2024 Capai 10.37 juta," *CNN Indonesia*, vol. 1, no. 1, pp. 1-2, 4 November 2024.
- [12] A. Pramana and R. Yusuf, "Implementasi Naive Bayes pada Analisis Sentimen Ulasan Hotel Berbahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 2, pp. 120-128, 2023.
- [13] D. Pradana and F. Setiawan, "Perbandingan Algoritma Machine learning pada Analisis Sentimen Review Hotel," *Jurnal Media Informatika Budidarma*, vol. 7, no. 4, pp. 1880-1888, 2023.
- [14] R. Kurniawan, A. Wijaya, and M. Saputra, "Evaluasi Kinerja Naive Bayes, KNN dan SVM pada Sentiment Analysis Pariwisata," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 1, pp. 55-63, 2024.
- [15] J. Rennie, L. Shih, J. Teevan, and D. Karger, "Tackling the Poor Assumptions of Naive Bayes Text Classifiers," *Proceedings of the 20th International Conference on Machine learning (ICML)*, vol. 20, no. 1, pp. 616-623, 2003.
- [16] D. N. Prayogi and S. A, "Implementasi Naive Bayes Classifier dan Confusion matrix pada Analisis Sentimen Berbasis Teks pada Twitter," *Jurnal Informatika*, vol. 5, no. 17, pp. 345-355, 2021.
- [17] G. F. N. Duei Putri and W. E. S. D, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat pada Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 3, pp. 275-285, 2022.
- [18] F. A. M. Ramadani and M. S, "Klasifikasi Sentimen Ulasan Pengguna Marketplace Tokopedia Menggunakan Naive Bayes dan TF-IDF," *Jurnal Teknik Informatika*, vol. 4, no. 9, pp. 101-109, 2023.
- [19] M. Rahman, I. Firmansyah, dan D. Putra, "Sentiment Classification on Hotel Reviews using Probabilistic Approach," *Journal of Data Science and Information Systems*, vol. 5, no. 2, pp. 77-86, 2023.
- [20] S. Putri dan R. Wijaya, "Pengaruh TF-IDF terhadap Akurasi Analisis Sentimen Bahasa Indonesia," *Jurnal Informatika*, vol. 11, no. 1, pp. 31-39, 2024.
- [21] A. Saputra, F. Ramadhan, dan L. Hakim, "Implementasi N-Gram pada Analisis Sentimen Ulasan Pelanggan," *Jurnal Teknologi Komputer*, vol. 9, no. 3, pp. 201-209, 2024.
- [22] N. Hidayat dan B. Nugroho, "Challenges of Sarcasm Detection in Indonesian Sentiment Analysis," *Journal of Artificial Intelligence Research*, vol. 6, no. 1, pp. 15-24, 2023.