

KLASIFIKASI SENTIMEN ULASAN HOTEL DI ALOR MENGGUNAKAN ALGORITMA NAÏVE BAYES

Putu Devan Satriawan¹, Ryan Peterzon Hadjon², Bonie Empy Giri³, Selus Paru Kelin⁴

^{1,2,3,4} Program Studi Teknologi Informasi Fakultas Teknik Universitas Citra Bangsa

Email : defangaul9@gmail.com¹, ryanphd2@gmail.com², bonieangi28@gmail.com³,
selusparukelin@gmail.com⁴

Abstrak

Perkembangan teknologi informasi telah mendorong pelanggan untuk memberikan ulasan terhadap layanan hotel melalui berbagai platform digital, seperti Google Review dan Traveloka. Ulasan tersebut menjadi sumber informasi yang penting bagi pihak hotel dalam mengevaluasi kualitas pelayanan dan tingkat kepuasan pelanggan. Namun, banyaknya data ulasan menyebabkan proses analisis secara manual menjadi kurang efektif dan membutuhkan waktu yang cukup lama. Oleh karena itu, penelitian ini bertujuan membangun sistem analisis sentimen berbasis web menggunakan algoritma Naïve Bayes untuk mengklasifikasikan ulasan pelanggan Hotel Nusa Kenari ke dalam kategori positif, negatif, dan netral. Dataset yang digunakan dalam penelitian ini terdiri dari 613 data ulasan yang diperoleh dari Google Review, Traveloka, dan hasil wawancara. Tahapan penelitian meliputi preprocessing data berupa *data cleaning*, *case folding*, *tokenizing*, *stopword removal*, *stemming*, pembobotan kata menggunakan TF-IDF, proses oversampling, serta pembagian data menjadi 80% data training dan 20% data testing. Model klasifikasi dibangun menggunakan algoritma Naïve Bayes dan dievaluasi menggunakan Confusion Matrix, Accuracy, Precision, Recall, dan F1-Score. Hasil pengujian menunjukkan bahwa sistem mampu mengklasifikasikan sentimen ulasan pelanggan dengan tingkat akurasi sebesar 80,54%. Hasil tersebut menunjukkan bahwa algoritma Naïve Bayes memiliki performa yang baik dalam analisis sentimen sehingga dapat membantu pengelola Hotel Nusa Kenari memperoleh informasi mengenai persepsi pelanggan secara lebih cepat, objektif, dan efisien sebagai dasar peningkatan kualitas pelayanan.

Kata Kunci: Analisis Sentimen, Naïve Bayes, Hotel, Text Mining, TF-IDF.

I. PENDAHULUAN

Perkembangan teknologi informasi dan internet telah mengubah cara masyarakat menyampaikan pengalaman terhadap suatu produk maupun layanan. Salah satu bentuk penyampaian pendapat yang paling banyak digunakan adalah melalui ulasan pada platform digital, seperti Google Review dan Traveloka. Bagi industri perhotelan, ulasan pelanggan menjadi sumber informasi yang penting karena memuat penilaian mengenai kualitas pelayanan, fasilitas, kebersihan, kenyamanan, serta pengalaman selama menginap. Informasi tersebut dapat dimanfaatkan sebagai bahan evaluasi dalam meningkatkan kualitas pelayanan dan kepuasan pelanggan.

Hotel Nusa Kenari merupakan salah satu penyedia jasa penginapan di Kabupaten Alor yang menerima berbagai ulasan dari pelanggan melalui platform digital. Banyaknya ulasan yang diterima setiap waktu menyebabkan proses analisis secara manual menjadi kurang efektif karena membutuhkan waktu yang lama serta berpotensi menimbulkan penilaian yang subjektif. Akibatnya, informasi penting yang terkandung dalam ulasan pelanggan belum dapat dimanfaatkan secara optimal sebagai dasar pengambilan keputusan untuk meningkatkan kualitas

layanan hotel. Oleh karena itu, diperlukan suatu sistem yang mampu mengolah data ulasan secara otomatis sehingga opini pelanggan dapat dianalisis dengan lebih cepat, akurat, dan objektif.

Salah satu pendekatan yang dapat digunakan adalah analisis sentimen dengan memanfaatkan teknik *text mining*. Analisis sentimen bertujuan mengelompokkan opini pengguna ke dalam kategori positif, negatif, maupun netral berdasarkan isi teks yang diberikan. Salah satu algoritma yang banyak digunakan dalam klasifikasi teks adalah Naïve Bayes karena memiliki proses komputasi yang sederhana, cepat, serta mampu menghasilkan performa klasifikasi yang baik. Untuk meningkatkan kualitas representasi data teks, penelitian ini juga menerapkan metode Term Frequency–Inverse Document Frequency (TF-IDF) sebagai pembobotan kata sebelum proses klasifikasi dilakukan.

Beberapa penelitian sebelumnya menunjukkan bahwa algoritma Naïve Bayes memiliki tingkat akurasi yang baik dalam analisis sentimen pada berbagai bidang, seperti media sosial, e-commerce, maupun layanan publik. Namun, penelitian mengenai analisis sentimen ulasan pelanggan hotel, khususnya pada Hotel Nusa Kenari di Kabupaten Alor, masih relatif terbatas. Oleh karena itu, penelitian ini menerapkan algoritma Naïve

Bayes untuk membangun sistem analisis sentimen berbasis web yang mampu mengklasifikasikan ulasan pelanggan ke dalam kategori positif, negatif, dan netral. Sistem yang dikembangkan diharapkan dapat membantu pengelola hotel memahami persepsi pelanggan secara lebih cepat serta mendukung pengambilan keputusan dalam upaya meningkatkan kualitas pelayanan.

II. DASAR TEORI

2.1 Analisis Sentimen

Analisis sentimen merupakan salah satu teknik dalam text mining yang digunakan untuk mengidentifikasi dan mengelompokkan opini atau pendapat pengguna terhadap suatu produk maupun layanan ke dalam kategori sentimen positif, negatif, atau netral. Analisis sentimen banyak diterapkan pada data ulasan pelanggan karena mampu memberikan gambaran mengenai tingkat kepuasan pengguna terhadap suatu layanan. Pada penelitian ini, analisis sentimen digunakan untuk mengelompokkan ulasan pelanggan Hotel Nusa Kenari sehingga dapat menjadi bahan evaluasi dalam meningkatkan kualitas pelayanan.

2.2 Text Mining

Text mining merupakan proses pengolahan data teks tidak terstruktur menjadi informasi yang dapat dianalisis. Tahapan text mining meliputi preprocessing data, ekstraksi fitur, hingga proses klasifikasi menggunakan algoritma tertentu. Pada penelitian ini, preprocessing dilakukan melalui tahapan data cleaning, case folding, tokenizing, stopword removal, dan stemming agar data ulasan menjadi lebih bersih dan siap diproses pada tahap klasifikasi.

2.3 Algoritma Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi bahwa setiap atribut bersifat independen terhadap atribut lainnya. Algoritma ini memiliki keunggulan berupa proses komputasi yang sederhana, cepat, dan efektif dalam melakukan klasifikasi data teks. Pada penelitian ini, algoritma Naïve Bayes digunakan untuk mengklasifikasikan sentimen ulasan pelanggan Hotel Nusa Kenari ke dalam kategori positif, negatif, dan netral berdasarkan hasil pembobotan kata menggunakan metode TF-IDF. Persamaan dasar algoritma Naïve Bayes adalah:

$$P(H | X) = \frac{P(X | H) \cdot P(H)}{P(X)}$$

dengan:

- $P(H|X)$ = probabilitas hipotesis berdasarkan data,
- $P(X|H)$ = probabilitas data terhadap hipotesis,
- $P(H)$ = probabilitas awal (prior),
- $P(X)$ = probabilitas data.

2.4 TF-IDF (Term Frequency–Inverse Document Frequency)

TF-IDF merupakan metode pembobotan kata yang

digunakan untuk mengubah data teks menjadi representasi numerik. Nilai TF-IDF dihitung berdasarkan frekuensi kemunculan suatu kata dalam dokumen dan tingkat kemunculannya pada seluruh dokumen. Metode ini memberikan bobot yang lebih tinggi pada kata yang dianggap penting sehingga mampu meningkatkan kualitas proses klasifikasi sentimen.

2.5 Confusion Matrix

Confusion Matrix merupakan metode evaluasi yang digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan hasil prediksi terhadap data sebenarnya. Evaluasi dilakukan menggunakan metrik Accuracy, Precision, Recall, dan F1-Score. Pada penelitian ini, confusion matrix digunakan untuk mengetahui tingkat keberhasilan algoritma Naïve Bayes dalam mengklasifikasikan sentimen ulasan pelanggan Hotel Nusa Kenari.

III. METODE PENELITIAN

3.1 Dataset Penelitian

Penelitian ini menggunakan dataset berupa ulasan pelanggan Hotel Nusa Kenari Kabupaten Alor yang diperoleh dari Google Review, Traveloka, dan hasil wawancara secara langsung dengan pelanggan hotel. Seluruh data telah melalui proses pelabelan (labeling) secara manual menjadi tiga kelas sentimen, yaitu positif, negatif, dan netral. Dataset yang digunakan sebanyak 49 data ulasan, kemudian dibagi menjadi data pelatihan (training) dan data pengujian (testing) dengan perbandingan 80:20.

Tabel 1. Distribusi Dataset

Kelas Sentimen	Jumlah Data
Positif	19
Negatif	15
Netral	15
Total	49

3.2 Tahapan Penelitian

Metode penelitian yang digunakan terdiri atas beberapa tahapan, yaitu pengumpulan data, preprocessing teks, pembobotan kata menggunakan TF-IDF, klasifikasi menggunakan algoritma Naïve Bayes, serta evaluasi model menggunakan Confusion Matrix. Alur penelitian ditunjukkan pada Gambar 1.

Tahapan penelitian meliputi:

1. Pengumpulan data ulasan pelanggan.
2. Text preprocessing.
3. Pembobotan fitur menggunakan TF-IDF.
4. Klasifikasi menggunakan algoritma Naïve Bayes.
5. Evaluasi model menggunakan Confusion Matrix.

3.3 Text Preprocessing

Tahapan preprocessing bertujuan membersihkan data sehingga menghasilkan teks yang siap digunakan pada proses klasifikasi. Tahapan preprocessing terdiri atas:

1. Case Folding
2. Tokenizing
3. Stopword Removal
4. Stemming

Sebagai contoh digunakan kalimat berikut.

"Kamar bau dan pelayanan lambat."

Hasil preprocessing ditunjukkan pada Tabel 2.

Tabel 2. Contoh Hasil Preprocessing

Tahapan	Hasil
Kalimat awal	Kamar bau dan pelayanan lambat
Case Folding	kamar bau dan pelayanan lambat
Tokenizing	kamar, bau, dan, pelayanan, lambat
Stopword Removal	kamar, bau, pelayanan, lambat
Stemming	kamar, bau, pelayanan, lambat

3.4 Klasifikasi Menggunakan Algoritma Naïve Bayes
Algoritma Naïve Bayes digunakan untuk menentukan kelas sentimen berdasarkan probabilitas kemunculan kata pada masing-masing kelas.

Sebagai contoh digunakan data uji berikut.

"Kamar bau dan pelayanan lambat."

Setelah preprocessing diperoleh token:

kamar, bau, pelayanan, lambat

3.4.1 Probabilitas Prior

Probabilitas prior dihitung menggunakan Persamaan (4).

$$P(Kelas) = \frac{\text{Jumlah dokumen pada kelas}}{\text{Total dokumen}}$$

Hasil perhitungan prior ditunjukkan pada Tabel 3.

Tabel 3. Probabilitas Prior

Kelas	Jumlah Data	Prior
Positif	19	0,388
Negatif	15	0,306
Netral	15	0,306

3.4.2 Frekuensi Kata

Frekuensi kata diperoleh dari data pelatihan yang telah melalui proses preprocessing.

Tabel 4. Frekuensi Kata

Kata	Positif	Negatif	Netral
kamar	18	19	3
bau	0	15	0
pelayanan	10	9	2
lambat	0	7	1

3.4.3 Probabilitas Likelihood

Likelihood dihitung menggunakan Laplace Smoothing agar tidak terdapat probabilitas bernilai nol.

Persamaan likelihood ditunjukkan pada Persamaan (5).

$$P(w|Kelas) = \frac{f(w, Kelas) + 1}{N_{Kelas} + |V|}$$

dengan:

- (n_i) = frekuensi kata pada kelas
- (N) = jumlah seluruh kata pada kelas
- (V) = jumlah vocabulary.

Pada penelitian ini diketahui:

- Total kata kelas Positif = 110
- Total kata kelas Negatif = 70
- Total kata kelas Netral = 33
- Jumlah Kata Unik = 20

Sebagai contoh, probabilitas likelihood kata bau pada kelas negatif dihitung sebagai berikut.

$$P(bau | Negatif) = \frac{15 + 1}{70 + 20} = \frac{16}{90} = 0,178$$

Perhitungan likelihood seluruh kata ditunjukkan pada Tabel 5.

Tabel 5. Probabilitas Likelihood

Kata	Positif	Negatif	Netral
kamar	0,146	0,222	0,075
bau	0,007	0,177	0,018
pelayanan	0,084	0,111	0,056
lambat	0,007	0,088	0,0189

3.4.4 Probabilitas Posterior

Nilai posterior dihitung menggunakan Persamaan (6).

$$P(H | X) = \frac{P(X | H) \cdot P(H)}{P(X)}$$

Perhitungan untuk setiap kelas adalah sebagai berikut.

Kelas Positif

$$P(U | Positif) = 0,1462 \times 0,0077 \times 0,0846 \times 0,0077 = 7,32 \times 10^{-7}$$

Kelas Negatif

$$P(U | Negatif) = 0,2222 \times 0,1778 \times 0,1111 \times 0,0889 = 3,91 \times 10^{-4}$$

Kelas Netral

$$P(U | Netral) = 0,0755 \times 0,0189 \times 0,0566 \times 0,0189 = 3,91 \times 10^{-4}$$

Hasil perhitungan posterior disajikan pada Tabel 6.

Tabel 6. Hasil Probabilitas Posterior

Kelas	Nilai Posterior
Positif	1.02×10^{-7}
Negatif	9.56×10^{-6}
Netral	3.84×10^{-8}

Berdasarkan hasil perhitungan tersebut, nilai posterior terbesar diperoleh pada kelas Negatif, sehingga kalimat "Kamar bau dan pelayanan lambat" diklasifikasikan sebagai sentimen negatif.

3.5 Evaluasi Model

Evaluasi model dilakukan menggunakan Confusion Matrix untuk mengetahui kemampuan model dalam mengklasifikasikan sentimen ulasan pelanggan.

Confusion Matrix menghasilkan empat metrik evaluasi, yaitu Accuracy, Precision, Recall, dan F1-Score.

Persamaan akurasi ditunjukkan pada Persamaan (7).

$$\text{Akurasi} = \frac{TP + TN + TNN}{\text{Total}} = \frac{17 + 13 + 13}{49} = 0.87 = 87\%$$

Persamaan precision ditunjukkan pada Persamaan (8).

$$\text{Presisi} = \frac{TP}{TP + FP} = \frac{17}{19} = 0.89 = 89\%$$

Persamaan recall ditunjukkan pada Persamaan (9).

$$\text{Recall} = \frac{TP}{TP + FN + FNN} = \frac{17}{19} = 0.89 = 89\%$$

Persamaan F1-Score ditunjukkan pada Persamaan (10).

$$\text{F1-Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 0.89 = 89\%$$

Nilai accuracy menunjukkan tingkat ketepatan model dalam melakukan klasifikasi, sedangkan precision menunjukkan ketepatan prediksi pada suatu kelas. Recall mengukur kemampuan model menemukan seluruh data yang benar pada suatu kelas, sementara F1-Score merupakan rata-rata harmonis antara precision dan recall sehingga memberikan gambaran menyeluruh terhadap performa model.

IV. HASIL DAN PEMBAHASAN

4.1. Deskripsi Umum dan Dataset Penelitian

Penelitian ini berhasil mengimplementasikan sistem klasifikasi sentimen ulasan hotel di Kabupaten Alor dengan menggunakan algoritma Naïve Bayes. Sistem yang dibangun dirancang khusus untuk membantu pihak manajemen hotel dalam mengenali kecenderungan penilaian atau opini pelanggan secara otomatis, cepat, dan objektif berdasarkan ulasan yang tersedia di berbagai platform daring seperti Google Review dan Traveloka, serta hasil wawancara terstruktur pada Hotel Nusa Kenari Indah Kalabahi. Sistem ini berfungsi sebagai alat bantu analisis dalam mengevaluasi kualitas layanan hotel tanpa menggantikan penilaian subjektif dari pelanggan.

Dataset utama yang digunakan dalam penelitian ini berjumlah 613 data ulasan yang telah melalui proses anotasi manual untuk pemberian label sentimen ke dalam tiga kelas, yaitu positif, negatif, dan netral. Sebelum dilakukan tahapan pembersihan (data cleaning), data ulasan tersebut memiliki distribusi awal yang tidak seimbang, dengan rincian kelas sentimen negatif sebanyak 248 ulasan, sentimen positif sebanyak 198 ulasan, dan sentimen netral sebanyak 166 ulasan. Atribut utama dalam dataset ini terdiri dari teks ulasan sebagai fitur input (X) dan label sentimen sebagai target kelas (Y).

4.2. Pra-pemrosesan Data (Text Preprocessing)

Pra-pemrosesan teks merupakan tahapan krusial untuk

mentransformasikan data ulasan mentah yang tidak terstruktur menjadi bentuk yang bersih dan seragam guna meningkatkan kinerja dan akurasi algoritma Naïve Bayes. Implementasi pra-pemrosesan data dalam penelitian ini dilakukan menggunakan bahasa pemrograman Python dengan melalui beberapa tahapan sistematis sebagai berikut:

A. Data Cleaning dan Case Folding

Tahap ini bertujuan untuk membersihkan teks ulasan dari elemen-elemen pengganggu yang tidak memiliki makna dalam analisis sentimen. Proses ini meliputi penghapusan karakter khusus, angka, tanda baca, emoji, serta spasi berlebih. Bersamaan dengan data cleaning, dilakukan pula proses case folding untuk mengubah seluruh huruf dalam teks menjadi huruf kecil (lowercase). Hasil dari tahapan ini memangkas data menjadi 612 ulasan yang konsisten dan siap diproses pada tahapan berikutnya.

B. Tokenizing, Stopword Removal, dan Stemming

Setelah data bersih, kalimat ulasan dipecah menjadi potongan kata tunggal atau token melalui proses tokenizing. Token-token tersebut kemudian disaring menggunakan teknik stopwords removal untuk membuang kata-kata umum yang sering muncul namun tidak memiliki bobot informatif terhadap sentimen, seperti kata "dan", "yang", "di", dan "adalah". Selanjutnya, dilakukan proses stemming untuk mengembalikan kata berimbuhan menjadi kata dasarnya (misalnya kata "pelayanan" diubah menjadi "layanannya").

C. Oversampling Data

Mengingat adanya ketidakseimbangan jumlah data antar kelas sentimen (248 negatif, 198 positif, dan 166 netral) yang dapat menyebabkan model cenderung bias terhadap kelas mayoritas, dilakukan teknik oversampling pada data pelatihan. Melalui penyeimbangan ini, jumlah data pelatihan berhasil diekspansi sehingga total keseluruhan dataset setelah oversampling menjadi 744 data dengan distribusi kelas yang sepenuhnya seimbang. Hal ini sangat membantu model untuk mempelajari karakteristik linguistik setiap kelas dengan lebih adil.

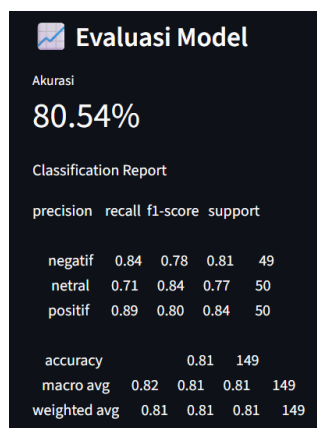
D. Representasi Fitur Menggunakan TF-IDF

Tahap akhir dari pra-pemrosesan adalah mengubah token teks menjadi representasi numerik menggunakan metode Term Frequency - Inverse Document Frequency (TF-IDF). TF-IDF memberikan bobot matematis pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen ulasan relatif terhadap seluruh korpus dataset. Kata-kata bermakna khusus akan mendapatkan bobot yang lebih tinggi, sementara kata umum akan diredam nilainya. Bobot numerik inilah yang menjadi input utama bagi perhitungan probabilitas pada algoritma Naïve Bayes.

4.3. Hasil Pengujian dan Evaluasi Kinerja Model

Setelah data melalui tahapan pra-pemrosesan dan penyeimbangan, dilakukan pembagian data (data splitting) dengan rasio 80% untuk data pelatihan (training data) sebanyak 595 ulasan dan 20% untuk data pengujian (testing data) sebanyak 149 ulasan. Model Naïve Bayes dilatih menggunakan data pelatihan untuk menghitung nilai prior probability dan conditional probability (likelihood) dengan menerapkan Laplace smoothing guna menghindari nilai probabilitas nol.

Pengujian model dilakukan secara ketat menggunakan 149 data uji yang sama sekali belum pernah dipelajari oleh model selama proses pelatihan. Evaluasi kinerja sistem diukur secara objektif menggunakan confusion matrix untuk membandingkan label sentimen aktual dengan hasil prediksi sistem. Berdasarkan pengujian tersebut, diperoleh metrik performa model sebagaimana disajikan pada Gambar 4.1 berikut:



Gambar 4.1 Evaluasi Kinerja Model Naïve Bayes

Berdasarkan laporan hasil evaluasi pada Gambar 4.1, model klasifikasi sentimen berbasis Naïve Bayes menghasilkan tingkat akurasi sebesar 80,54%. Nilai presisi tertinggi dicapai oleh kelas sentimen positif yaitu sebesar 0,89, yang menunjukkan kemampuan sistem dalam meminimalkan kesalahan prediksi positif palsu (false positive). Sementara itu, kelas netral mencatatkan nilai recall tertinggi sebesar 0,84, mengindikasikan bahwa model sangat sensitif dalam mengenali ulasan yang bersifat deskriptif tanpa muatan emosi yang kuat. Secara keseluruhan, nilai F1-score yang seimbang membuktikan bahwa kombinasi pembobotan TF-IDF dan algoritma Naïve Bayes mampu mengelompokkan opini pelanggan hotel di Kabupaten Alor dengan sangat baik.

4.4. Implementasi Antarmuka Sistem Berbasis Web

Untuk mempermudah pemanfaatan model klasifikasi ini oleh pengguna akhir, sistem diimplementasikan ke dalam antarmuka aplikasi berbasis web yang responsif dan terbagi menjadi dua hak akses utama, yaitu halaman admin dan halaman tamu:

A. Antarmuka Halaman Admin (Web Admin)

Halaman admin diawali dengan fitur keamanan Login Admin yang mewajibkan penginputan username dan password valid untuk melindungi privasi data transaksi ulasan. Setelah otentikasi berhasil, admin akan diarahkan ke Dashboard Utama yang menyajikan visualisasi data komprehensif. Dashboard menampilkan ringkasan performa real-time seperti total ulasan masuk (tercatat sebanyak 227 data ulasan pada siklus berjalan dengan proporsi 101 ulasan positif, 55 ulasan netral, dan 71 ulasan negatif). Selain itu, dashboard dilengkapi dengan diagram batang distribusi kelas dan grafik garis tren sentimen per hari untuk memantau fluktuasi kepuasan pelanggan secara berkala.

Admin juga memiliki akses penuh ke Halaman Hasil Klasifikasi Sentimen yang menampilkan tabel data ulasan lengkap beserta penandaan warna visual (warna hijau untuk sentimen positif, kuning untuk netral, dan merah untuk negatif). Pada halaman ini tersedia fitur filter dinamis berbasis kategori sentimen dan rentang tanggal proses untuk mempermudah audit data ulasan, serta tombol aksi hapus untuk membersihkan data yang tidak relevan.

B. Antarmuka Halaman Tamu (Guest System)

Halaman input ulasan tamu didesain dengan antarmuka yang sangat sederhana (minimalis). Melalui halaman ini, tamu hotel dapat mengetikkan pendapat atau menceritakan pengalaman mereka secara langsung selama menginap. Demi menjaga objektivitas, halaman tamu ini bersifat anonim dan mandiri, di mana pelanggan hanya dapat mengirimkan teks ulasan tanpa diberikan hak akses untuk melihat hasil kalkulasi probabilitas maupun label sentimen yang diprediksi oleh sistem. Data yang dikirimkan secara otomatis masuk ke dalam basis data relasional sistem untuk langsung diolah dan diverifikasi oleh pihak admin.

V. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Naïve Bayes untuk mengklasifikasikan sentimen ulasan pelanggan Hotel Nusa Kenari ke dalam kategori positif, negatif, dan netral. Proses klasifikasi dilakukan melalui beberapa tahapan, yaitu preprocessing data, pembobotan kata menggunakan metode TF-IDF, penyeimbangan data dengan oversampling, pembagian dataset menjadi data training dan data testing, serta pembangunan model klasifikasi menggunakan algoritma Naïve Bayes.

Berdasarkan hasil pengujian menggunakan Confusion Matrix, model yang dikembangkan memperoleh tingkat akurasi sebesar 80,54%. Selain itu, hasil evaluasi menggunakan Precision, Recall, dan F1-Score menunjukkan bahwa model mampu

mengklasifikasikan sebagian besar ulasan pelanggan dengan baik. Hasil tersebut membuktikan bahwa algoritma Naïve Bayes memiliki performa yang cukup baik dalam melakukan analisis sentimen terhadap data ulasan hotel.

Sistem analisis sentimen yang dibangun mampu membantu pihak Hotel Nusa Kenari dalam mengidentifikasi opini pelanggan secara otomatis sehingga proses evaluasi kualitas pelayanan dapat dilakukan dengan lebih cepat, objektif, dan efisien. Dengan demikian, hasil penelitian ini diharapkan dapat menjadi salah satu pendukung pengambilan keputusan dalam upaya meningkatkan kualitas pelayanan dan kepuasan pelanggan.

Saepudin, S., Widiastuti, S., & Irawan, C. (2023). *Analisis Sentimen Ulasan Platform Media Sosial Menggunakan Algoritma Naïve Bayes Classifier*. 12, 236–243.

VI. REFERENSI

Arif Bijaksana Putra Negara, Muhandi, H., & Indra Melinda Putri. (2020). *Tampilan Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes dan Seleksi Fitur Information Gain*.

Data, A., Dengan, M., Bayes, N., Klasifikasi, U., Penerima, K., Iuran, B., Kesehatan, J., Desa, P., El, D.-D., Purba, R., Al, E., El, D., & Siboro, Y. O. (2025). *Analisis Kelayakan Penerima Bantuan Iuran Jaminan Kesehatan Menggunakan Teknik Klasifikasi Data Mining dengan Metode Naïve Bayes*. 07(02), 2721–1800.

<https://journal.cattleyadf.org/index.php/jatilima/index>

Hananto, A. D., Erfiana, A. M., Putri, B. L. P., Putri, P. D., & Kurniawan, F. (2023). Algoritma Machine Learning Naïve Bayes pada Analisis Sentimen Kesepakatan Polri dan GNPf-MUI pada Aksi Bela Islam III “212.” *SINTA Journal (Science, Technology, and Agricultural)*, 4(2), 151–160. <https://doi.org/10.37638/sinta.4.2.151-160>

Lukmana, L. O., L, D. R., & Elisya, P. (2025). *APLIKASI THREADS INSTAGRAM DI PLAYSTORE MENGGUNAKAN ALGORITMA NAIVE BAYES*. 13(2).

Puspa, T., Sanjaya, R., Fauzi, A., Fitri, A., & Masruriyah, N. (2023). *Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine*. 4, 16–26. <https://doi.org/10.37373/infotech.v4i1.422>

Rahayu, S., Halawa, H. W., Abdillah, A. F., Mujayanah, A., Informatika, T., Pamulang, U., & Artikel, I. (2025). *PEMANFAATAN BIG DATA ANALYTICS UNTUK ANALISIS POLA PERILAKU KONSUMEN E-COMMERCE STRATEGIS*. 6, 64–73.

Rolando, B., Ariyanto, K., Alexia, K. R., & Hartanti, R. (2022). Peran AI dan Big Data dalam Mengoptimalkan Strategi Pemasaran Digital. *Journal Dinamika Publika*, 1(1), 1–17.