

# Analisis Sentimen Terhadap Hotel Keprabon Menggunakan Algoritma Naïve Bayes

Rajendra Nandana Hernanto,  
Sistem Informasi,  
Telkom University Purwokerto,  
Purwokerto, Indonesia  
rajendranandana@student.telkomuniver  
sity.ac.id

Sena Wijayanto,  
Sistem Informasi,  
Telkom University Purwokerto,  
Purwokerto, Indonesia  
senawijayanto@telkomuniversity.ac.id

**Abstrak** — Hotel Keprabon sebagai salah satu hotel non-bintang di Surakarta menghadapi tantangan dalam memahami persepsi pelanggan terhadap kualitas layanan mereka. Banyaknya ulasan pelanggan yang tersebar di platform digital menyebabkan kesulitan dalam melakukan analisis secara manual. Sistem ini dirancang untuk menerima masukan berupa data ulasan pelanggan dan menghasilkan keluaran berupa klasifikasi sentimen positif atau negatif. Topik ini penting karena ulasan pelanggan di platform seperti Traveloka dan Tiket.com sangat memengaruhi citra dan keputusan calon pelanggan. Saat ini, Hotel Keprabon belum memiliki sistem terstruktur untuk mengolah opini pelanggan secara otomatis, sehingga diperlukan pendekatan yang efisien untuk menggali insight dari data yang tidak terstruktur. Solusi yang dikembangkan berupa sistem analisis sentimen berbasis klasifikasi. Data ulasan dikumpulkan dari platform digital, lalu diproses melalui tahap pembersihan, normalisasi, dan konversi menjadi format terstruktur. Sentimen ditentukan menggunakan pendekatan klasifikasi berbasis model statistik, dilatih pada data yang telah diseimbangkan dan diberi bobot kata. Sistem ini dibangun menggunakan pendekatan pembelajaran mesin yang sederhana namun efektif untuk teks Bahasa Indonesia. Sistem ini berhasil mengklasifikasikan ulasan dengan tingkat akurasi mencapai 87%. Kontribusi utama dari penelitian ini adalah sistem otomatis yang mampu membantu pengelola hotel dalam mengevaluasi layanan berdasarkan opini pelanggan secara efisien dan objektif.

**Kata kunci**— analisis sentimen, ulasan, naïve bayes, TF-IDF, hotel, SMOTE

## I. PENDAHULUAN

Pariwisata merupakan salah satu sektor perekonomian yang dapat meningkatkan pendapatan daerah dan kesejahteraan masyarakat setempat [1]. Salah satu kota dengan potensi pariwisata yang tinggi merupakan Kota Surakarta, Jawa Tengah[2]. Adanya potensi pariwisata yang tinggi, memiliki hubungan yang erat dengan industri perhotelan, terutama bisnis perhotelan[3]. Berdasarkan data dari Badan Pusat Statistik (BPS) Kota Surakarta, jumlah hotel di Kota Surakarta pada 2024 mencapai 164 [4].

Melihat banyaknya hotel di Kota Surakarta, persaingan bisnis perhotelan menjadi ketat. Hotel merupakan fasilitas

yang penting dalam industri pariwisata, karena berperan sebagai tempat menginap para wisatawan. Hotel Keprabon merupakan salah satu hotel non bintang yang berada pada kecamatan Banjarsari. Hotel ini terletak pada jantung Kota Surakarta yang dekat dengan berbagai destinasi wisata di Kota Surakarta seperti salah satunya adalah Pura Mangkunegaran. Sektor perhotelan dianggap penting karena menyediakan fasilitas dan layanan penginapan, makanan, minuman, serta berbagai layanan lain yang dikelola secara profesional dan bertahap[5]. Faktor-faktor seperti fasilitas, kenyamanan, harga, layanan, dan ulasan pengunjung sebelumnya berpengaruh keputusan saat memesan hotel. Salah satu bentuk evaluasi Hotel Keprabon adalah melihat umpan balik yang diberikan dari pelanggan berupa ulasan, baik dari platform digital atau secara langsung yang dapat dimanfaatkan hotel untuk meningkatkan mutu layanan mereka. Ulasan ini memberikan perspektif subjektif dari tamu tentang pengalaman mereka selama menginap di hotel dan dapat mempengaruhi pendapat tamu lainnya. Dengan memahami perasaan pelanggan, hotel dapat menentukan bidang yang perlu ditingkatkan untuk meningkatkan kepuasan pelanggan dan reputasi hotel. Ulasan online sangat mempengaruhi cara pelanggan dalam melihat barang atau layanan dan membuat keputusan pembelian mereka(Farrel dkk., 2023). Dengan adanya persaingan bisnis ini, Hotel Keprabon ingin memahami pandangan pelanggan terhadap Hotel Keprabon apakah puas atau kecewa sebelum meningkatkan kualitas mereka. Salah satu langkah yang diambil adalah dengan mengetahui respon pelanggan hotel melalui ulasan pelanggan di aplikasi atau platform yang terhubung dengan hotel.

Dalam dunia digital ini, ulasan online telah menjadi sumber informasi penting bagi wisatawan dan pengelola hotel. *Traveloka* adalah salah satu aplikasi yang berfokus pada layanan bepergian seperti pemesanan hotel atau tiket. *Traveloka* didirikan pada tahun 2012 dengan tujuan utama untuk memudahkan para pelancong dalam memilih akomodasi yang cocok [7]. Selain *Traveloka*, salah satu

layanan pemesanan hotel secara *online* adalah *Tiket.com*. *Tiket.com* menyediakan pemesanan tiket hotel dan hal seputar akomodasi lainnya secara *online*[8]. Ulasan yang ditinggalkan pelanggan hotel pada berbagai platform tersebut dapat memberikan tanggapan berharga tentang pengalaman mereka. Penilaian dari pelanggan umumnya berupa rating dan pengalaman yang sering kali disampaikan dalam bentuk komentar atau saran. Kualitas layanan hotel umumnya dinilai menggunakan *rating* dan masukan dari para tamu. Walaupun metode *rating* sering dipakai dalam mengukur layanan hotel, *rating* sendiri tidak dapat sepenuhnya menangkap keluhan atau kekurangan yang dirasakan pelanggan[9]. Maka dari itu, ulasan adalah sesuatu berguna dalam menentukan ekspresi atau pengalaman pelanggan berdasarkan yang mereka ungkapkan. Data ulasan yang digunakan untuk analisis sentimen pada penelitian ini diambil dari ulasan *Traveloka* dan ulasan *Tiket.com*. Platform tersebut dipilih karena berdasarkan pelanggan yang menginap dari Hotel Keprabon.

Analisis sentimen adalah suatu metode yang digunakan untuk mengenali, menganalisis, dan memproses teks untuk menangkap konten emosional dalam representasi bahasa dari pendapat atau pemikiran. [10]. Melalui analisis sentimen, wawasan berharga dapat diperoleh dari data yang tidak terstruktur dengan mengklasifikasikan pendapat yang dinyatakan sesuai dengan emosi yang terkandung di dalamnya sebagai sentimen positif atau negatif[11]. Dengan menerapkan analisis sentimen Hotel Keprabon pada berbagai ulasan dari beberapa platform, manajemen Hotel Keprabon dapat memperoleh seluruh gambaran dari perasaan dan pendapat para pengunjung hotel. Hotel Keprabon ingin melakukan peningkatan infrastruktur hotel, tetapi sebelum itu perlu melihat pandangan pelanggan terhadap Hotel Keprabon terlebih dahulu karena dari pihak Hotel Keprabon juga kesulitan terhadap fasilitas mana yang akan mereka perbaiki, maka dari itu diperlukan pendapat atau pandangan pelanggan terhadap Hotel Keprabon.

Tahapan analisis sentimen terdapat metode klasifikasi yang umum digunakan yaitu, Algoritma *Naïve Bayes* yang banyak digunakan pada data *mining* atau *text mining*[12]. Algoritma ini membuat pernyataan probabilitas tentang kejadian di masa depan berdasarkan pengalaman sebelumnya dan oleh karena itu disebut sebagai teorema Bayes. Algoritma *Naïve Bayes* ditandai oleh efisiensinya dan sangat baik untuk dataset yang memiliki banyak fitur (atribut) dan sejumlah besar data[13]. Selain *Naïve Bayes*, terdapat beberapa metode lain yang digunakan untuk analisis sentimen, seperti *Decision Tree*, K-NN, dan *Support Vector Machine*. Dalam penelitian ini, algoritma *Naïve Bayes* dipilih karena dinilai efisien dalam mengklasifikasikan opini, terutama pada data berbasis teks seperti ulasan dari pengunjung[14]. Menurut penelitian yang dilakukan oleh Fatikhah, Muhtarom, dan Oktaviani (2024) dalam jurnal ZONasi: Jurnal Sistem Informasi, algoritma *Naïve Bayes* terbukti efektif dalam menganalisis sentimen ulasan pelanggan terhadap Mie Gacoan Solo Veteran di

platform Google Maps. Penelitian ini menggunakan algoritma *Naïve Bayes* yang dipadukan dengan teknik peningkatan data menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE), dan berhasil meraih akurasi sebesar 91% dalam proses klasifikasi sentimen ulasan. Hal tersebut menunjukkan bahwa metode *Naïve Bayes* dapat digunakan secara optimal dalam mengelompokkan opini berbasis teks, khususnya untuk ulasan pengguna di media digital[15].

Berdasarkan permasalahan yang disebutkan, studi ini bertujuan untuk mengetahui bagaimana pelanggan mengungkapkan pendapat dan penilaian mereka dalam ulasan dan bagaimana Hotel Keprabon dapat memanfaatkan informasi ini untuk meningkatkan kualitas layanan. Dengan menganalisis sentimen pelanggan, area yang perlu diperbaiki dapat diidentifikasi untuk meningkatkan kepuasan pelanggan serta reputasi hotel. Selain itu, studi ini juga bertujuan untuk menentukan sejauh mana suasana positif dan negatif mendominasi dan seberapa baik algoritma *Naïve Bayes* mampu mengklasifikasikan dan mengevaluasi perasaan pelanggan berdasarkan ulasan dari *Traveloka* dan *Tiket.com*.

## II. KAJIAN TEORI

### A. Analisis Sentimen

Analisis sentimen adalah bidang yang luas dalam pengolahan bahasa alami, linguistik komputer, dan analisis teks. Tujuannya adalah untuk mengenali dan menginterpretasikan opini, suasana hati, penilaian, sikap, evaluasi, atau emosi seseorang terhadap topik, produk, layanan, organisasi, orang, atau aktivitas tertentu.

### B. Hotel

Hotel adalah usaha komersial yang menawarkan kamar untuk menginap serta makanan, hiburan, dan layanan lainnya dengan pembayaran setiap hari. Ada perbedaan antara hotel tanpa klasifikasi bintang dan hotel bintang. Yang terakhir umumnya menawarkan standar yang lebih tinggi dengan fasilitas dan layanan yang lebih lengkap, seperti kolam renang atau ruang kebugaran. Kualitas sebuah hotel bervariasi tergantung pada klasifikasinya.

### C. Naïve Bayes

*Naïve Bayes* adalah metode klasifikasi dari bidang pembelajaran mesin yang didasarkan pada teorema Bayes. Dalam hal ini, diasumsikan bahwa semua fitur dalam data pelatihan saling independen.[17]. Kekuatan *Naïve Bayes* termasuk implementasinya yang sederhana, kecepatan pemrosesan yang tinggi, serta kebutuhan yang relatif sedikit akan data pelatihan untuk klasifikasi yang efektif[18].

### D. Klasifikasi

Klasifikasi adalah suatu prosedur untuk mengembangkan model atau fungsi yang menggambarkan dan membedakan berbagai konsep atau kategori data[19]. Tujuannya adalah untuk memprediksi keanggotaan objek yang sebelumnya tidak dikenal ke dalam kelas tertentu[20]. Proses klasifikasi dibagi menjadi dua fase utama: fase pelatihan dan fase pengujian. Pada fase pelatihan, model dibangun

menggunakan data yang ada, sedangkan pada fase pengujian, model diuji dengan data baru yang belum pernah digunakan sebelumnya untuk menilai akurasi.

#### E. Crawling

*Crawling* merujuk pada proses di mana salinan data yang relevan diambil dari *World Wide Web*[21]. Selain itu, *crawling* juga digunakan untuk mengumpulkan informasi dari halaman web. Metode pengumpulan informasi ini dapat dilakukan secara menyeluruh atau secara terfokus pada konten tertentu.[22].

#### F. Text Preprocessing

*Text Preprocessing* adalah proses di mana ditentukan data mana dari setiap dokumen yang akan diproses lebih lanjut.[23]. Tujuan dari langkah ini adalah untuk mengubah data teks menjadi bentuk yang lebih terstruktur. Proses ini memainkan peran penting dalam analisis sentimen, karena struktur data yang lebih tinggi menghasilkan hasil prediksi yang lebih tepat[24].

#### G. TF-IDF

*Term Frequency-Inverse Document Frequency* (TF-IDF) adalah metode untuk memberi bobot pada kata-kata, di mana frekuensi suatu istilah di dalam dokumen tertentu (*Term Frequency/TF*) digabungkan dengan kelangkaannya dalam keseluruhan kumpulan dokumen (*Inverse Document Frequency/IDF*). Dengan cara ini, setiap kata dalam teks mendapatkan bobot tertentu.[25]. Singkatnya, tujuan dari metode TF-IDF adalah untuk menentukan total bobot istilah dalam sebuah dokumen[26].

#### H. Confusion Matrix

Matriks kebingungan adalah alat untuk menilai kinerja model prediksi, dengan mengukur tingkat akurasi dalam proses klasifikasi[27]. Ini digunakan untuk mengevaluasi kualitas klasifikasi dan memberikan metrik seperti akurasi, presisi, *recall*, dan *F1-score*.

#### I. Traveloka

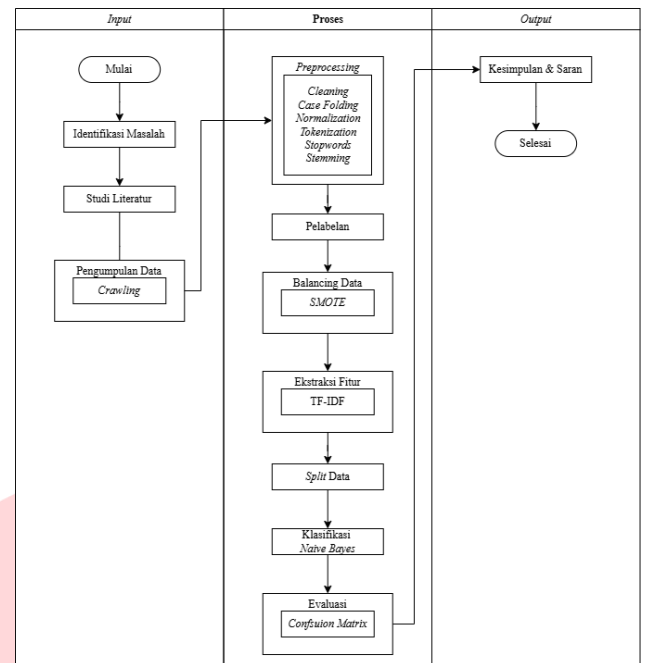
*Traveloka* adalah perusahaan yang pertama kali menyediakan layanan pemesanan online untuk hotel dan tiket pesawat, yang didirikan oleh Ferry Unardi, Derianto Kusuma, dan Albert[28].

#### J. Tiket.com

*Tiket.com* adalah portal perjalanan daring (Online Travel Agent/OTA) yang menawarkan layanan pemesanan untuk berbagai jenis tiket, termasuk tiket pesawat, reservasi hotel, tiket kereta api, serta tiket masuk untuk konser dan acara hiburan lainnya[29]. Platform ini memiliki sistem pemesanan yang cepat, harga yang bersaing, dan beragam penawaran layanan.

### III. METODE

Beberapa tahapan yang dilakukan untuk melakukan penelitian ini diilustrasikan pada Gambar 1.



GAMBAR 1.  
(DIAGRAM ALIR PENELITIAN)

Tahap pertama dilakukan dengan identifikasi masalah pada Hotel Keprabon. Kondisi yang bertujuan untuk mencari gambaran terhadap permasalahan. Kegiatan tersebut di ambil dari sumber yang terpercaya secara langsung dengan pihak Hotel Keprabon. Selanjutnya ditentukan tujuan dan manfaat dari penelitian ini.

Pada tahap studi literatur, mencari informasi yang berhubungan dari bermacam-macam sumber seperti jurnal atau artikel ilmiah yang berhubungan dengan penelitian ini serta memperoleh literatur yang sesuai dengan objek kajian maka dilakukan studi literatur terlebih dahulu. Studi literatur dalam hal ini terkait dengan penelitian analisis sentimen hotel menggunakan algoritma *Naïve Bayes*. Jurnal atau artikel ilmiah yang digunakan diperoleh dari sumber yang terpercaya seperti IEEE atau *Google Scholar* dengan rentang waktu 5 tahun terakhir. Tahap ini dilakukan bertujuan agar validitas penelitian ini terpenuhi.

Proses pengumpulan data penelitian ini dijalankan dengan tahap *crawling* pada *Traveloka* dan *Tiket.com* menggunakan *Instant Data Scraper* dan *Web Scraper*. Data diambil dari Januari 2017 hingga November 2024. Data yang diambil berupa teks ulasan pelanggan Hotel Keprabon. Data yang diambil menggunakan Bahasa Indonesia. Data disimpan dalam format *.xlsx*.

Setelah data dikumpulkan, dilakukan *text preprocessing*. *Text Preprocessing* adalah proses yang bertujuan untuk pembersihan kata yang tidak diperlukan sebelum data diolah untuk dilanjutkan ke tahap berikutnya[30]. *Text Preprocessing* mencakup beberapa tahap, yaitu: *cleaning*, *case folding*, *tokenizing*, *stopwords removal*, *normalization*, *stemming*.

Setelah melakukan proses *preprocessing* data, selanjutnya dilakukan *labelling* pada data yang sudah

diproses menjadi tiga kategori yaitu positif, negatif, dan netral. *Lexicon* adalah *tools* yang digunakan dalam proses *labelling*. Proses *labelling* dilakukan dengan menganalisis teks yang telah diterapkan proses *preprocessing* lalu ditentukan nilai *polarity*nya terlebih dahulu. Setelah itu, ditentukan label sentimennya berdasarkan nilai *polarity* nya.

Kemudian dilakukan *balancing* data. *Balancing* data digunakan ketika data yang digunakan tidak seimbang, di mana jumlah data pada satu kelompok jauh lebih banyak daripada kelompok lainnya. Masalah ini dapat menyebabkan model yang dibuat terlalu fokus pada kelompok data mayoritas dan gagal memprediksi kelompok minoritas[31]. Untuk mengatasi masalah ini, digunakan metode SMOTE. SMOTE dapat mengumpulkan data tambahan pada kelompok minoritas sehingga jumlah data di kedua kelompok seimbang. Ini membantu model yang dikembangkan untuk mencegah *overfitting* dan belajar secara adil.

Setelah data terlebih dahulu diberi label dan kemudian diseimbangkan, ekstraksi fitur dilakukan dengan menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*). Teknik ini digunakan untuk mengubah data teks menjadi bentuk numerik yang dapat diproses oleh algoritma *NaïveBayes*. Dengan TF-IDF, bobot untuk kata-kata dalam konten dokumen ditetapkan.[1].

Sebelum melakukan klasifikasi, langkah berikutnya adalah memisahkan data menjadi dua bagian, yaitu data latih dan data uji. Data latih dimanfaatkan untuk melatih algoritma klasifikasi untuk mengenali pola pada data. Sementara itu, data uji digunakan untuk menilai sejauh mana model dapat melakukan klasifikasi dengan baik dan tepat. Pengujian ini bertujuan untuk memberikan gambaran pada model mengenai kapabilitas dalam mengelompokkan data yang belum pernah dijumpai sebelumnya.

Tahap berikutnya dilakukan klasifikasi dengan menerapkan algoritma *Naïve Bayes*. Klasifikasi ini mengakar pada *Teorema Bayes* dengan menggunakan metode statistik dan probabilitas. Menurut *Teorema Bayes* sendiri, setiap variabel bersifat independen.

Tahap terakhir dalam penelitian ini adalah melakukan evaluasi dengan *confusion matrix*. *Confusion matrix* merupakan salah satu teknik untuk menentukan akurasi pada ide penambangan data atau *data mining* [32]. *Confusion matrix* digunakan untuk melihat tingkat akurasi atau performa dari klarifikasi dan menghasilkan nilai *accuracy*, *precision*, *recall*, dan *f1 score*.

#### IV. HASIL DAN PEMBAHASAN

##### A. Pengumpulan Data

Data diambil dari halaman Hotel Keprabon pada *Traveloka*, dan *Tiket.com*. Proses pengambilan data dilakukan dengan menggunakan fitur *extensions* dalam *Google Chrome*, seperti *Web Scraper*. *Web Scraper* digunakan untuk mengambil data pada *Traveloka* dan

*Tiket.com*. Total data yang diperoleh dari proses ini sebanyak 1339 data.

| nama       | rating | tanggal   | ulasan                                                                                                                                                                                                                                                                                                                                                                                                                             |
|------------|--------|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Arif A.    | 10     | 01-Sep-24 | Hotel murah tapi bersih, meskipun terlihat tua, terawat. Don't expect to much ya, properti memang kusam dan tua, kamar mandi pake timba dan gayung, closet sudah kusam. Tapi lumayan bersih, AC meskipun uda udzur juga tetep dingin mengigil. Staff ramah-ramah, dan grtis kopi dan teh bikin sendiri all day. Harga on the spot saya lihat 100 K saja, dan bisa sewa room buat mingguan dan sebulan juga, murah banget pokoknya. |
| Surya F.   | 10     | 26-Jun-24 | Untuk harga segitu, cukup lumayan. Lokasi strategis. Rekomendasi menginap backpacker di Solo.                                                                                                                                                                                                                                                                                                                                      |
| B***a      | 10     | 17-Jun-24 | Bapak yang melayani sangat ramah. Hotelnya nyaman. Next pingin ngeinap lagi d sini.                                                                                                                                                                                                                                                                                                                                                |
| Sukesih M. | 10     | 25-Mei-24 | Puas deh ngeinap di sini, murah tapi ga murahan.                                                                                                                                                                                                                                                                                                                                                                                   |

(a)

| nama           | rating | ulasan                                                                                                                                                                                                                  |
|----------------|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Oktavia Novita | 5,0    | worth it syekali harga segitu dgn fasilitas yg maksimal                                                                                                                                                                 |
| SA             | 4,0    | worth it dengan harga dibawah 100k, tpat cukup bersih dan strategis                                                                                                                                                     |
| BR             | 4,0    | Untuk harga segitu udah cukuplah minusnya cuma dibagian handuk sama selimut yang usang/lama Staf ramah banget.                                                                                                          |
| RM             | 4,0    | Ukuran kamarnya luas .. cocok untuk yg bawa keluarga .. air panasnya kadang berfungsi kadang tidak , TVnya masih jadul , bebas masak atau refill air galon, kmrn minta checkout telat jam 1 ga bisa . Tru...Selengkapny |

(b)

GAMBAR 2.

(HASIL SCRAPING DATA: TIKET.COM (A), TRAVELOKA (B))

##### B. Text Preprocessing

*Text preprocessing* dilakukan dengan memanfaatkan *Google Colab* dengan melalui enam langkah, yaitu *cleaning*, *case folding*, *tokenizing*, *stopwords removal*, *normalization*, *stemming*. Proses *cleaning* dilakukan untuk menyingkirkan komponen yang tidak ada relevansi pada data ulasan, seperti tanda baca, angka, URL, dan komponen lainnya yang tidak diperlukan.

TABEL 1.  
(CLEANING)

| Data sebelum <i>cleaning</i>                                    | Data setelah <i>cleaning</i>                               |
|-----------------------------------------------------------------|------------------------------------------------------------|
| Yang jelas kotor banyak kecoa dikamar nomor 24. Sudah itu saja. | Yang jelas kotor banyak kecoa dikamar nomor sudah itu saja |
| Nyaman. Wifinya cepet 😊😁 ...                                    | Nyaman Wifinya cepet                                       |
| Murah letak strategis pelayanan ramah https://vwborobudur. Icu. | Murah letak strategis pelayanan ramah vwborobudur Icu      |

Tahap selanjutnya adalah *case folding* yang dilakukan untuk merubah semua karakter dalam dokumen menjadi bentuk *lowercase* atau mengubah seluruh kata menjadi huruf kecil. Proses ini bertujuan agar sistem tidak membaca data

berulang ketika ada dua kata yang sama tetapi memiliki *case* huruf yang beda.

TABEL 2.  
(CASE FOLDING)

| Data sebelum <i>case folding</i>                                | Data setelah <i>case folding</i>                           |
|-----------------------------------------------------------------|------------------------------------------------------------|
| Yang jelas kotor banyak kecoa dikamar nomor 24. Sudah itu saja. | yang jelas kotor banyak kecoa dikamar nomor sudah itu saja |
| Nyaman. Wifinya cepet 😊😌 ...                                    | nyaman wifinya cepet                                       |
| Murah letak strategis pelayanan ramah https://vwborobudur. Icu. | murah letak strategis pelayanan ramah vwborobudur icu      |

Tahap normalization berfungsi untuk mentransformasi kata yang sebelumnya tidak baku menjadi kata baku yang sesuai dengan kaidah ejaan yang benar. Normalisasi penting karena banyak pengguna yang menggunakan kata-kata tidak formal atau *slang* seperti: ok, nggak, ngga, ga, dan sebagainya.

TABEL 3.  
(NORMALIZATION)

| Data sebelum <i>normalization</i>                          | Data setelah <i>normalization</i>                     |
|------------------------------------------------------------|-------------------------------------------------------|
| yang jelas kotor banyak kecoa dikamar nomor sudah itu saja | yang jelas kotor banyak kecoa dikamar sudah itu saja  |
| nyaman wifinya cepet                                       | nyaman wifinya cepat                                  |
| murah letak strategis pelayanan ramah vwborobudur icu      | murah letak strategis pelayanan ramah vwborobudur icu |

*Tokenization* atau tokenisasi adalah proses fragmentasi teks menjadi unit lebih kecil atau token, yang bisa berupa leksikon, karakter khusus, atau komponen lainnya. Tahapan ini berfungsi untuk mengenali setiap segmen teks sehingga dapat diproses lebih mendalam.

TABEL 4.  
(TOKENIZATION)

| Data sebelum <i>tokenization</i>                      | Data setelah <i>tokenization</i>                                                 |
|-------------------------------------------------------|----------------------------------------------------------------------------------|
| yang jelas kotor banyak kecoa dikamar sudah itu saja  | ['yang', 'jelas', 'kotor', 'banyak', 'kecoa', 'dikamar', 'sudah', 'itu', 'saja'] |
| nyaman wifinya cepat                                  | ['nyaman', 'wifinya', 'cepat']                                                   |
| murah letak strategis pelayanan ramah vwborobudur icu | ['murah', 'letak', 'strategis', 'pelayanan', 'ramah', 'vwborobudur', 'icu']      |

Kemudian, dilakukan proses *stopwords removal*. Proses *stopwords removal* merupakan tahap eliminasi kata yang tidak memiliki arti signifikan terhadap data teks ulasan. *Stopwords removal* dilakukan agar membantu model fokus ke kata-kata penting yang berpengaruh dalam menentukan sentiment ketika proses pelabelan. Tahap *stopwords removal* dilakukan dengan teliti dalam memilih kata yang akan

dieliminasi agar tidak mengubah makna dari ulasan dan menghasilkan sentimen yang berbeda dari yang seharusnya.

TABEL 5.  
(STOPWORDS REMOVAL)

| Data sebelum <i>stopwords removal</i>                                            | Data setelah <i>stopwords removal</i> |
|----------------------------------------------------------------------------------|---------------------------------------|
| ['yang', 'jelas', 'kotor', 'banyak', 'kecoa', 'dikamar', 'sudah', 'itu', 'saja'] | kotor banyak kecoa dikamar            |
| ['nyaman', 'wifinya', 'cepat']                                                   | nyaman wifinya cepat                  |
| ['murah', 'letak', 'strategis', 'pelayanan', 'ramah', 'vwborobudur', 'icu']      | murah letak strategis pelayanan ramah |

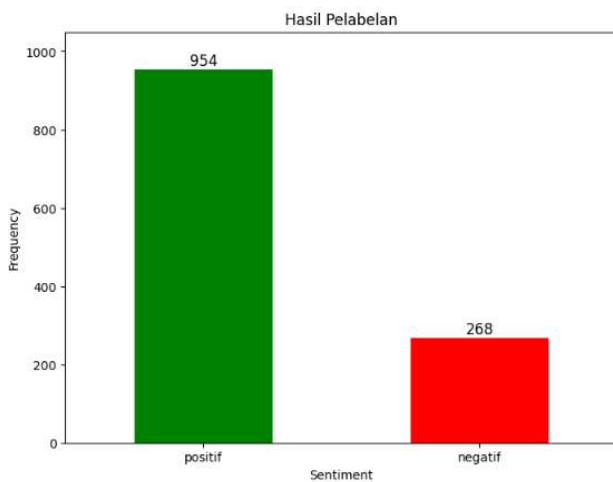
Proses *stemming* merupakan tahap mengubah kata dengan imbuhan menjadi kata dasar. Tahap ini dilakukan dengan menghapus imbuhan kata depan, sisipan, belakang atau gabungan dari tiap kata. *Stemming* membantu model agar menganggap satu kata memiliki makna yang sama.

TABEL 6.  
(STEMMING)

| Data sebelum <i>stemming</i>          | Data setelah <i>stemming</i>      |
|---------------------------------------|-----------------------------------|
| kotor banyak kecoa dikamar            | kotor banyak kecoa kamar          |
| nyaman wifinya cepat                  | nyaman wifi cepat                 |
| murah letak strategis pelayanan ramah | Murah letak strategis layan ramah |

### C. Pelabelan

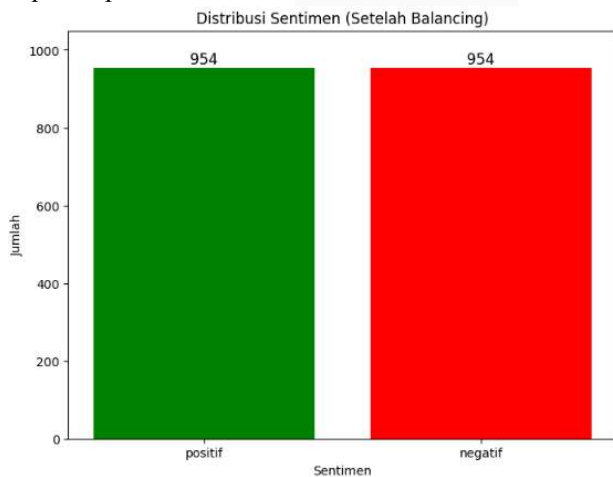
Setelah proses *preprocessing* selesai, diterapkan proses pelabelan yang bertujuan untuk mengklasifikasikan ulasan berdasarkan kelas sentimen yang telah ditentukan, apakah positif atau negatif. Proses pelabelan dilakukan secara otomatis dengan menggunakan *library* dari InSet (Indonesia Lexicon). Proses pelabelan ditentukan berdasarkan nilai polaritas yang dimiliki tiap kata, yang akan menentukan label sentimennya. Jika total skor lebih besar dari nol akan dilabelkan positif dan sebaliknya untuk yang negatif, jika nilainya sama akan dilabelkan netral. Proses pelabelan menghasilkan jumlah data positif sebanyak 954 dan negatif 268 data yang telah divisualisasikan pada Gambar 3.



GAMBAR 3.  
(VISUALISASI PELABELAN)

#### D. Balancing Data

Berdasarkan visualisasi data yang telah dilakukan pelabelan, dihasilkan jumlah data yang tidak seimbang pada tiap kategori, tahap selanjutnya adalah melakukan proses *balancing* data. Proses ini mengatasi ketidakseimbangan jumlah data antar kelas yang bertujuan untuk memastikan jumlah data dalam tiap kategori bersifat proporsional. Proses *balancing* data dilakukan dengan menggunakan metode SMOTE untuk menyeimbangkan data dengan membuat sampel sintesis dari kelas minoritas. Hasil *balancing* data ditampilkan pada Gambar 4.



GAMBAR 4  
(HASIL PROSES BALANCING DATA)

#### E. Ekstraksi Fitur

Tahap berikutnya adalah ekstraksi fitur. Fase ekstraksi fitur, tahapan yang dilakukan setelah dilakukan *stemming* data, mentransformasikan data menjadi bentuk vektor. Proses ekstraksi fitur dilakukan dengan menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*) yang memanfaatkan *library Count Vectorizer*. TF-IDF bertujuan untuk mengonversi teks dalam dokumen ke dalam representasi numerik, di mana setiap kata diberikan bobot yang mencerminkan tingkat signifikansinya dalam suatu dokumen serta keseluruhan kumpulan dokumen. Hasil ekstraksi ditampilkan pada Gambar 5.

|   | term  | idf      | tf0      | tf1      | tf2 | tf3 | tf4 | tf5      |
|---|-------|----------|----------|----------|-----|-----|-----|----------|
| 0 | abon  | 6.800393 | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.0 | 0.000000 |
| 1 | abu   | 6.463921 | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.0 | 0.000000 |
| 2 | ac    | 1.954632 | 0.157869 | 0.228938 | 0.0 | 0.0 | 0.0 | 0.115604 |
| 3 | acara | 6.800393 | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.0 | 0.000000 |
| 4 | acnya | 6.618071 | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.0 | 0.000000 |

GAMBAR 5.  
(HASIL EKSTRAKSI FITUR TF-IDF)

#### F. Split Data

Tahap selanjutnya adalah *split* data atau pembagian data. Dataset dibagi menjadi data *training* dan data *testing*. Tiap data memiliki perbandingan persentase data 80% untuk data *training*, dan 20% untuk data *testing*.

#### G. Klasifikasi Naïve Bayes

Setelah data dibagi menjadi dua bagian, dilakukan proses klasifikasi menggunakan *Naïve Bayes*. Dalam proses klasifikasi menggunakan *Naïve Bayes*, setiap langkahnya dikombinasikan menjadi satu alur terintegrasi yang terdiri dari *count vectorizer*, *transformer*, dan *classifier*. Model klasifikasi *Naïve Bayes* yang digunakan adalah *Multinomial NB* untuk proses klasifikasinya. Setelah semua tahapan selesai, dilakukan evaluasi kinerja model *Naïve Bayes* dengan memanfaatkan *confusion matrix*.

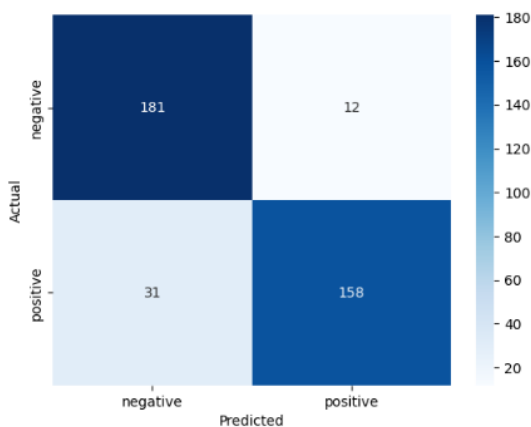
#### H. Confusion Matrix

Tahap selanjutnya melakukan evaluasi performa model *Naïve Bayes* yang telah dilaksanakan sebelumnya menggunakan metode *Confusion Matrix*. Untuk mengevaluasi performa model dan mengukur seberapa baik kinerja *Naïve Bayes*, perlu diterapkan uji model, di mana hasil kinerja ditampilkan dalam bentuk *Confusion Matrix*. *Confusion Matrix* akan menghasilkan tingkat akurasi model yang diperoleh dengan membagi total data yang benar diklasifikasikan dengan total data yang ada.

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.85      | 0.94   | 0.89     | 193     |
| 2            | 0.93      | 0.84   | 0.88     | 189     |
| accuracy     |           |        | 0.89     | 382     |
| macro avg    | 0.89      | 0.89   | 0.89     | 382     |
| weighted avg | 0.89      | 0.89   | 0.89     | 382     |

GAMBAR 6.  
(HASIL EVALUASI CONFUSION MATRIX)

Berdasarkan hasil evaluasi dengan *Confusion Matrix* pada Gambar 6, akurasi yang diperoleh dari pemodelan *Naïve Bayes* sebesar 87%. *Accuracy* menunjukkan persentase prediksi yang benar dari pemodelan *Naïve Bayes*, *precision* mengukur ketepatan prediksi terhadap kelas tertentu, *recall* mengukur seberapa besar model dapat mengenali data dari suatu kelas, *f1-score* merujuk pada rata-rata antara *precision* dan *recall*, *support* menggambarkan jumlah sampel untuk kelas. Visualisasi *Confusion Matrix* dapat dilihat pada Gambar 7.



GAMBAR 7.  
(VISUALISASI CONFUSION MATRIX)

### I. Implikasi

Implikasi penelitian menjabarkan efek hasil analisis sentimen dengan ulasan *Traveloka*, dan *Tiket.com* pada Hotel Keprabon secara nyata dengan memperhatikan berdasarkan hasil visualisasi data mengenai kata-kata yang sering muncul pada ulasan yang telah melalui proses *preprocessing* menggunakan *WorldCloud*.



GAMBAR 8.

(*WORLD CLOUD SENTIMEN: POSITIF(A), NEGATIF(B)*)

Berdasarkan Gambar 8(a), kata “hotel” sering muncul pada sentimen positif, mengindikasikan kata yang paling banyak dibahas. Dari kata tersebut, dapat diartikan banyak pelanggan yang memiliki penilaian positif tentang hotel. Gambar 8(b) menampilkan kata yang sering muncul pada sentimen negatif ialah kata “kamar”. Pada kata-kata ini diartikan pada bagian kamar hotel belum seluruh pelanggan puas dengan layanan kamar hotel. Kata-kata yang sering muncul pada sentimen negatif ini dapat dimanfaatkan hotel dalam memudahkan

dalam mencari celah kekurangan mereka. Kata-kata ini dapat digunakan sebagai pertimbangan dan untuk mengoreksi kekurangan yang lain.

Berdasarkan hasil visualisasi *worldcloud* kata-kata yang sering muncul, hotel dapat menentukan bidang yang perlu ditingkatkan atau perbaiki pada tiap sentimen tergantung dari konteks ulasan yang dimaksud.

### V. KESIMPULAN

Berdasarkan penelitian yang telah dilaksanakn mengenai analisis sentimen terhadap ulasan pelanggan terkait Hotel Keprabon menggunakan algoritma *Naïve Bayes*, penelitian ini berhasil mengklasifikasikan ulasan pelanggan dan diperoleh ulasan positif sebanyak 954 data dan negatif sebanyak 268 data yang mana diperoleh dari *platform Traveloka* dan *Tiket.com*. Hal ini menunjukkan pandangan sebagian besar pelanggan yang pernah menginap di Hotel Keprabon merasa puas.

Hasil kinerja algoritma *Naïve Bayes* yang diterapkan untuk melakukan analisis sentimen ini menunjukkan bahwa algoritma ini memperoleh nilai *accuracy* 88%, *precision* 89%, *recall* 89%, dan *f1-score* 89% dengan proporsi pembagian data training dan data testing 80:20.

### REFERENSI

- [1] M. Afrad, C. Febrianto, S. Wijayanto, and Y. Fathoni, “Sentiment Analysis of Visitor Reviews on Baturaden Tourist Attraction Using Machine Learning Methods,” *Edu Komputika*, vol. 11, no. 1, 2024, [Online]. Available: <https://journal.unnes.ac.id/journals/edukom/>
- [2] S. Maulidia and E. T. Astuti, “Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Potensi Pariwisata Tahun 2020,” *Seminar Nasional Official Statistics*, pp. 405–414, 2022.
- [3] M. P. V. Langi, B. Lagarensen, and O. Lintong, “ANALISIS STRATEGI PELAKU BISNIS PERHOTELAN DALAM Mendukung Pariwisata Kota Manado,” *Jurnal Hospitaliti dan Pariwisata*, vol. 2, 2024.
- [4] Badan Pusat Statistik Kota Surakarta, *Direktori Hotel dan Akomodasi Lainnya Di Kota Surakarta 2024*, vol. 8, 2024.
- [5] I. Suwarno and E. Martini, “Peran Customer Relationship Management dalam Meningkatkan Loyalitas Pelanggan: Studi Kasus di Hotel Zaen Syariah Solo,” 2021.
- [6] D. Farrel Putra Rachmawan, M. Hanindia Prami Swari, and C. Aji Putra, “Analisis Sentimen Review Hotel Favehotel Kusumanegara Menggunakan Algoritma Multinomial Naive Bayes,” vol. 3, pp. 155–159, 2023, [Online]. Available: [www.traveloka.com](http://www.traveloka.com)
- [7] Z. A. Diekson, M. R. B. Prakoso, M. S. Q. Putra, M. S. A. F. Syaputra, S. Achmad, and R. Sutoyo, “Sentiment Analysis for Customer Review: Case Study Of Traveloka,”

- in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 682–690. doi: 10.1016/j.procs.2022.12.184.
- [8] B. Budiman, Z. Silvana Anggraeni, C. Habibi, and N. Alamsyah, “Analisis Sentimen Publik pada Media Sosial Twitter Terhadap Tiket.com Menggunakan Algoritma Klasifikasi,” *Jurnal Informatika*, vol. 11, no. 1, pp. 1–10, Apr. 2024, doi: 10.31294/inf.v11i1.17988.
- [9] B. Bayu Baskoro *et al.*, “Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR),” vol. 3, no. 2, pp. 21–029, 2021, doi: 10.20895/INISTA.V3I2.
- [10] A. Safira, A. S. Masyarakat...v, and F. N. Hasan, “Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier,” *Jurnal Sistem Informasi*, vol. 5, no. 1, 2023.
- [11] B. Ramadhani and R. R. Suryono, “Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 2, p. 714, Apr. 2024, doi: 10.30865/mib.v8i2.7458.
- [12] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional,” vol. 15, no. 1, 2021.
- [13] P. Sofyan Zakaria, R. Julianto, and R. Surya Bernada, “Implementasi Naive Bayes Menggunakan Python Dalam Klasifikasi Data,” 2023. [Online]. Available: <https://jurnalmahasiswa.com/index.php/biikma>
- [14] D. Arya Utama, “Analisis Sentimen Tempat Wisata Kabupaten Gunungkidul Menggunakan Metode Naive Bayes Berdasarkan Kritik Saran Wisatawan,” *3rd Seminar Nasional Mahasiswa Fakultas Teknologi Informasi*, vol. 2, no. 2, 2023.
- [15] T. Fatikhah, M. Muhtarom, and I. Oktaviani, “Analisis Sentimen Ulasan Mie Gacoan Solo Veteran Di Google Maps Menggunakan Algoritma Naive Bayes,” *ZONasi: Jurnal Sistem Informasi*, vol. 6, no. 3, 2024.
- [16] A. Z. Amrullah, A. Sofyan Anas, M. Adrian, and J. Hidayat, “Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square,” *Jurnal*, vol. 2, no. 1, 2020, doi: 10.30812/bite.v2i1.804.
- [17] A. Dwi, “Metode Naive Bayes Untuk Klasifikasi Masa Studi Sarjana,” *Teknologipintar.org*, vol. 3, no. 4, pp. 2023–2024, 2023.
- [18] I. Arfanda, W. Ramdhan, R. A. Yusda, and H. Artikel, “Naive Bayes Dalam Menentukan Penerima Bantuan Langsung Tunai,” vol. 1, no. 1, 2021, doi: 10.47709/briliance.vxix.xxxx.
- [19] S. Febriani and H. Sulistiani, “Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5,” vol. 2, pp. 89–95, 2021.
- [20] D. Azzahra Nasution, H. H. Khotimah, and N. Chamidah, “Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-Nn,” 2019.
- [21] I. N. Husada, E. H. Fernando, H. Sagala, A. E. Budiman, and H. Toba, “Ekstraksi dan Analisis Produk di Marketplace Secara Otomatis dengan Memanfaatkan Teknologi Web Crawling,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 5, no. 3, Jan. 2020, doi: 10.28932/jutisi.v5i3.1977.
- [22] J. Budiarto, “Identifikasi Kebutuhan Masyarakat Nusa Tenggara Barat pada Pandemi Covid-19 di Media Sosial dengan Metode Crawling,” vol. 2, no. 4, pp. 244–250, 2021.
- [23] L. Hermawan, M. B. Ismiati, J. Bangau, N. 60, and M. Charitas, “Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval,” *TRANSFORMATIKA*, vol. 17, no. 2, pp. 188–199, 2020.
- [24] K. Sriwenda Putri, R. Setiawan, and A. Pambudi, “Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan Naïve Bayes Classifier,” 2023.
- [25] A. N. Assidyk, E. B. Setiawan, S. Si, I. Kurniawan, S. Pd, and M. Si, “Analisis Perbandingan Pembobotan TF-IDF dan TF-RF pada Trending Topic di Twitter dengan Menggunakan Klasifikasi K-Nearest Neighbor,” 2020.
- [26] M. Afrad, “Utilization Of Principal Component Analysis To Improve Emotion Classification Performance In Text Using Artificial Neural Networks,” 2024. [Online]. Available: <https://www.kaggle.com/Datasets/ishantjuyal/emotions-in-text>
- [27] E. W. Hary Candana, I. Gede, A. Gunadi, and D. G. H. Divayana, “Perbandingan Fuzzy Tsukamoto, Mamdani dan Sugeno Dalam Penentuan Hari Baik Pernikahan Berdasarkan Wariga Menggunakan Confusion Matrix,” *Jurnal Ilmu Komputer Indonesia (JIK)*, vol. 6, no. 2, 2021.
- [28] A. Yuni Widowati and C. Budihartanti, “Analisis Kepuasan Pengguna Terhadap Aplikasi Traveloka Dengan Menerapkan Metode Tam (Technology Acceptance Model),” vol. 6, no. 2, 2019.
- [29] M. Raka Alfajri, A. Wardhana, and M. Si, “Pengaruh Customer Relationship Management (Crm) Terhadap Loyalitas Pelanggan Tiket.Com (Survey Pada Pengguna Tiket.Com Di Kota Bandung),” 2020. [Online]. Available: [www.similarweb.com](http://www.similarweb.com),
- [30] A. Malik Zuhdi, E. Utami, and S. Raharjo, “Analisis Sentiment Twitter Terhadap Capres Indonesia 2019 Dengan Metode K-Nn,” 2019.
- [31] H. Hidayatullah, Purwantoro, and Y. Umaidah, “Penerapan Naïve Bayes Dengan Optimasi Information Gain dan Smote untuk Analisis Sentimen Pengguna Aplikasi Chatgpt,” *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 3, pp. 1546–1553, Jun. 2023.
- [32] B. Pratiwi, A. Handayani, and Sarjana, “Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi WSN Menggunakan Confusion Matrix,” vol. 6, 2020.

