IT-EXPLORE
Jurnal Penerapan Teknologi Informasi dan Komunikasi
ISSN 2828-7940, e-ISSN 2829-1727, DOI: 10.24246/itexplore.v4i3.2025.pp295-306
Volume 04 Nomor 03 Tahun 2025
Journal homepage: https://ejournal.uksw.edu/itexplore

Analisis sentimen ulasan tamu terhadap layanan hotel menggunakan pendekatan *machine learning*

Ricardho Gunawan [1], Hendry Hendry [2]
[1,2] Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Salatiga, Jawa Tengah, Indonesia

Informasi Artikel

Riwayat Artikel:

Dikirim:
22 Februari 2025
Direvisi:
12 Mei 2025
Diterima:
15 Mei 2025

Kata Kunci:

Analisis Sentimen, Grand Diamond Hotel Yogayakarta, *Support Vector Machine* (SVM), *Web Scraping*, Aplikasi Berbasis Web.

Keywords:

Sentiment Analysis, Grand Diamond Hotel Yogyakarta, Support Vector Machine (SVM), Web Scraping, Web-Based Application.

This is an open access article under the CC BY-SA license

ABSTRAK

Analisis sentimen terhadap ulasan tamu merupakan aspek penting dalam meningkatkan kualitas layanan perhotelan. Penelitian ini bertujuan untuk menganalisis sentimen ulasan tamu terhadap layanan Grand Diamond Hotel Yogyakarta menggunakan pendekatan *machine learning* dengan algoritma *Support Vector Machine* (SVM). SVM dipilih karena mampu menangani data berdimensi tinggi seperti teks, serta mampu membentuk batas pemisah yang optimal antara kelas-kelas sentimen. Data penelitian diperoleh melalui *web scraping* dari Traveloka, menghasilkan 1.119 ulasan yang diproses melalui tahap *preprocessing*, translasi, dan pelabelan sentimen menggunakan *library TextBlob*. Setelah pembobotan TF-IDF, data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Model SVM dengan *kernel* linear mencapai akurasi 80% dalam mengklasifikasikan ulasan ke dalam kategori positif, negatif, dan netral. Hasil penelitian ini diimplementasikan dalam aplikasi berbasis web yang dilengkapi fitur visualisasi data dan evaluasi model, sehingga memungkinkan manajemen hotel untuk memantau serta menganalisis sentimen tamu secara efisien dan mendukung peningkatan kualitas layanan berbasis data.

ABSTRACT

Sentiment analysis of guest reviews is a crucial aspect in improving the quality of hotel services. This study aims to analyze the sentiment of guest reviews regarding the services of Grand Diamond Hotel Yogyakarta using a machine learning approach with the Support Vector Machine (SVM) algorithm. SVM was chosen because it can handle high-dimensional data such as text and is capable of forming an optimal separating hyperplane between sentiment classes. The research data was obtained through web scraping from Traveloka, yielding 1,119 reviews, which were processed through preprocessing, translation, and sentiment labeling using the TextBlob library. After TF-IDF weighting, the data was divided into 80% for training and 20% for testing. The linear kernel SVM model achieved 80% accuracy in classifying the reviews into positive, negative, and neutral categories. The results of this study were implemented in a web-based application equipped with data visualization and model evaluation features, allowing hotel management to efficiently monitor and analyze guest sentiment and support data-driven service quality improvement.

Corresponding Author:

Ricardho Gunawan
Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana
Jl. Diponegoro No.52-60, Sidorejo, Salatiga, Jawa Tengah, Indonesia
Email: 672021062@student.uksw.edu

1. PENDAHULUAN

Dalam era digital yang berkembang pesat, teknologi semakin canggih dan memudahkan berbagai aspek kehidupan. Kini, untuk memberikan ulasan tidak lagi memerlukan cara manual seperti menulis di kertas, melainkan dapat dilakukan dengan mudah hanya dengan sentuhan jari melalui ponsel. Ulasan-ulasan tersebut memiliki peran penting, terutama dalam industri perhotelan, karena dapat membantu meningkatkan kualitas layanan dan mendukung kemajuan bisnis perhotelan.

Grand Diamond Hotel Yogyakarta merupakan salah satu hotel Yogyakarta berbintang empat yang menerima banyak ulasan dari para tamu yang menginap. Beberapa aspek yang sering menjadi perhatian dalam ulasan tersebut meliputi layanan, keramahan staf, kebersihan kamar, kenyamanan fasilitas, hingga kualitas makanan. Namun, untuk melihat ulasan-ulasan ini secara manual menjadi tantangan tersendiri karena jumlahnya yang terus bertambah dan beragam[1].

Untuk memperoleh data ulasan tamu, penelitian ini menggunakan sumber dari situs Traveloka. Traveloka merupakan platform yang menyediakan berbagai layanan travel, termasuk pemesanan hotel secara daring, dan menjadi salah satu platform paling populer di Indonesia. Data dari Traveloka dipilih karena ulasan yang tersedia berasal dari tamu yang benar-benar telah menginap, sehingga dinilai valid dan representatif terhadap pengalaman pelanggan. Selain itu, jumlah ulasan yang tersedia cukup banyak, terus diperbarui, dan proses pengambilannya relatif mudah dilakukan, sehingga sangat sesuai untuk keperluan analisis sentimen secara otomatis.

Analisis sentimen mengekstrak informasi dari teks untuk mengidentifikasi kecenderungan sentimen positif, netral, atau negatif. Pengguna internet sering menggunakan teknik ini di media sosial untuk menyampaikan opini atau memberikan evaluasi terhadap suatu topik[2]. Sehingga, untuk melihat ulasan-ulasan tamu secara efektif, diperlukan metode yang tepat untuk mengolah dan mengkategorikan data secara otomatis. Dengan analisis sentimen, ulasan dari tamu dapat dianalisis guna mengidentifikasi sentimen tamu terhadap layanan hotel.

Untuk melakukan klasifikasi sentimen secara akurat, penelitian ini menggunakan algoritma *Support Vector Machine* (SVM). SVM dipilih karena memiliki kemampuan tinggi dalam menangani data berdimensi besar seperti teks, serta mampu membentuk batas pemisah yang optimal antara kelas-kelas sentimen[3].

Berdasarkan permasalahan yang telah dibahas, penelitian ini mengangkat topik dengan judul ”Analisis Sentimen Ulasan Tamu Terhadap Layanan Grand Diamond Hotel Yogyakarta Menggunakan Pendekatan Machine Learning”. Penelitian ini bertujuan untuk membantu manajemen Grand Diamond Hotel Yogyakarta dalam meningkatkan kualitas layanan mereka. Dengan memahami sentimen tamu secara lebih mendalam, manajemen dapat mengidentifikasi aspek-aspek yang perlu diperbaiki serta merancang strategi peningkatan layanan yang tepat sasaran. Penelitian ini juga akan diterapkan dalam bentuk aplikasi berbasis web, yang memudahkan manajemen hotel dalam memantau dan menganalisis sentimen tamu secara efisien.

2. KAJIAN PUSTAKA

Analisis sentimen diterapkan pada ulasan Hotel Gran Mahakam Jakarta yang diperoleh melalui Google *Review* dalam penelitian Prasetyo dan Fahrurozi [3]. Studi ini melihat bagaimana *Support Vector Machine* (SVM) dan Naïve Bayes Classifier (NBC) berfungsi untuk mengkategorikan ulasan ke dalam kategori sentimen positif dan negatif. Hasil menunjukkan bahwa SVM lebih akurat, yaitu 92%, dibandingkan dengan NBC yang hanya mencapai 90%. Selain itu, penelitian ini menemukan bahwa NBC mengalami kesulitan dalam mengklasifikasikan sentimen negatif, karena dataset yang digunakan tidak seimbang (*imbalanced dataset*), akibatnya, nilai precision untuk kelas negatif tidak dapat dihitung. Oleh karena itu, penelitian ini menyimpulkan bahwa SVM memiliki keunggulan dibandingkan NBC dalam analisis sentimen pada ulasan hotel.

Studi tambahan yang dilakukan oleh Irma Surya et al. menganalisis sentimen ulasan pengguna aplikasi Shopee menggunakan algoritma *Support Vector Machine* (SVM). Tujuan dari penelitian ini adalah untuk mengelompokkan ulasan ke dalam kategori positif dan negatif serta mengevaluasi bagaimana algoritma bekerja selama proses klasifikasi. Setelah data penelitian diperoleh melalui metode *web scraping*, 3.000 ulasan diproses melalui tahapan *preprocessing* seperti *case-folding*, penghapusan *stopword*, dan tokenisasi, sebelum metode TF-IDF digunakan untuk pembobotan kata. Setelah proses pelatihan dan pengujian, dari hasil penelitian menunjukkan bahwa algoritma SVM mampu mendapatkan akurasi sebesar sebesar 98%, serta nilai *F1-score* yang juga mencapai 98%. Temuan ini mengindikasikan bahwa SVM merupakan

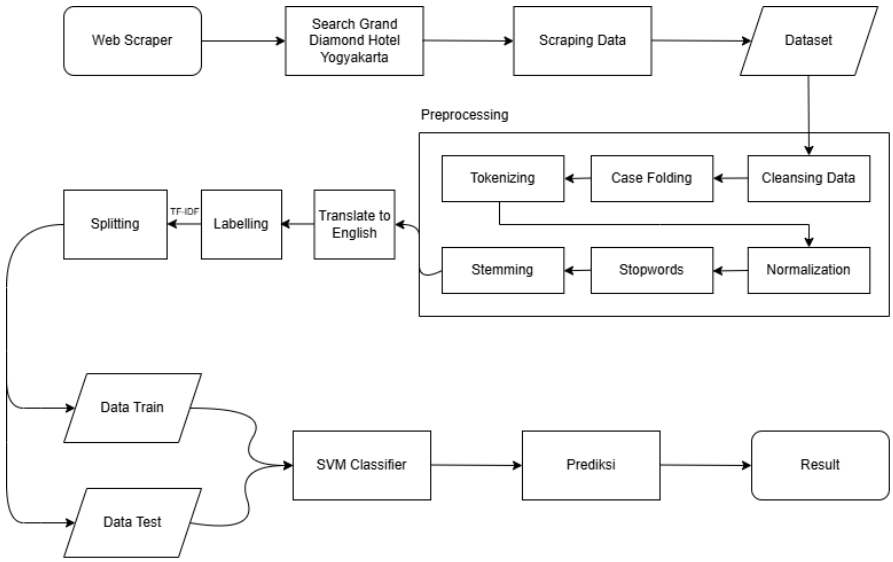
algoritma yang sangat efektif dalam analisis sentimen berbasis teks, khususnya dalam mengklasifikasikan opini pengguna secara akurat[4].

Pada penelitian terdahulu, Arrafiq dan Kurniawan [5] melakukan analisis sentimen terhadap ulasan pengguna layanan perbankan digital SeaBank Indonesia di Instagram menggunakan algoritma *Support Vector Machine* (SVM). Penelitian ini bertujuan untuk mengklasifikasikan komentar pengguna menjadi kategori positif dan negatif serta mengevaluasi kinerja algoritma dalam analisis sentimen. Data diperoleh melalui metode *web scraping* dengan total 1.201 komentar yang diproses melalui tahap *preprocessing*, termasuk *cleaning*, normalisasi, tokenisasi, penghapusan *stopword*, dan *stemming*. Model SVM yang digunakan menunjukkan hasil akurasi sebesar 89%, dengan *precision* 93%, *recall* 83%, dan *F1-score* 88%. Hasil penelitian ini menunjukkan bahwa SVM mampu mengklasifikasikan opini pengguna secara efektif, sehingga dapat digunakan sebagai alat dalam menganalisis kepuasan pelanggan terhadap layanan digital.

Penelitian sebelumnya oleh Pamungkas dan Kharisudin [6] menggunakan *Support Vector Machine* (SVM), Naïve Bayes, dan K-Nearest Neighbor (KNN) algoritma untuk melihat bagaimana masyarakat Indonesia menanggapi pandemi Covid-19 di media sosial Twitter. Tujuan dari penelitian ini adalah untuk membandingkan kinerja ketiga algoritma dalam mengklasifikasikan sentimen positif dan negatif dari 10.000 tanggapan yang dikumpulkan dengan menggunakan metode scraping web. Hasil penelitian menunjukkan bahwa algoritma SVM dengan kernel linear memiliki akurasi sebesar 90,01%; algoritma Naive Bayes memiliki akurasi sebesar 79,20% dan algoritma KNN memiliki akurasi sebesar 62,10%. Oleh karena itu, penelitian ini menyimpulkan bahwa SVM memiliki keunggulan dibandingkan algoritma lainnya dalam menganalisis sentimen berbasis teks di media sosial.

3. METODE PENELITIAN

Untuk menganalisis sentimen ulasan tamu terhadap layanan Grand Diamond Hotel Yogyakarta, terdapat beberapa tahapan penelitian yang harus dilakukan, sebagaimana ditunjukkan dalam diagram aliran pada Gambar 1.



Gambar 1: Diagram aliran penelitian

2.1. Pengumpulan Data

Penelitian ini memanfaatkan teknik *web scraping* dengan ekstensi chrome https://webscraper.io/ untuk mengumpulkan dataset dari situs Traveloka dalam kaitannya dengan ulasan Grand Diamond Hotel Yogyakarta. Metode scraping web digunakan untuk mengekstrak data dari website secara otomatis tanpa menyalinnya secara manual, seperti yang dijelaskan dalam penelitian[6]. Melalui teknik ini, berhasil dikumpulkan sebanyak 1.119 baris data dari situs traveloka, dengan empat atribut utama yang digunakan, yaitu *name*, *rating*, *date*, dan *review*. Keterangan atribut dapat dilihat pada Gambar 2.

	hotel_unit	name	rating	date	review
2	GDHY	F***n	10	2024-11-11	the best pokoke ga nyangka dapet Kamar Segede Itu dengan harga segitu
3	GDHY	D***i	10	2024-11-09	staff sangat sangat ramah, rekomendasi sekali untuk menginap disini
4	GDHY	G***t	6.6	2024-11-08	Bantal tidak empuk dan tipis, bathtub kurang bersih, pas saya wifi tersambung namun akses internet tidak bisa, TV kecil dengan siaran tidak bersih, perlengkapan mandi minim
5	GDHY	Y***i	10	2024-11-01	Ramah, tanggap saat check in maupun check out.
6	GDHY	A***i	10	2024-10-27	Sangat nyaman dan bersih dan rekomendasi untuk disampaikan ke teman dan kerabat.
7	GDHY	HE	10	2024-10-22	Hotel yang cukup lama di jogja. Dulu namanya grand quality & termasuk hotel mewah di jaman nya. Di luar ekspektasi kmarin menginap masih sangat terjaga kualitasnya. Kebersihan, pelayanan ok. Kamar sangat bersih walaupun memang tampilannya ga modern. Sarapannya enak banget banyak jenisnya, suka banget sma vibe restonya. Harga juga termasuk worth it dibanding hotel lainnya di jogja.
8	GDHY	sri d. e.	9.1	2024-10-15	Kamarnya luas bersih, towelnya wangi, sheet no spot. Kita dateng dijmset1 untungnya kamar sudah ready jadi bisa langsung ci, makanan enak, staf staf ramah. Masukan buat tvnya harus diganti deh karena kita nonton layarnya jelek.
9	GDHY	A***a	10	2024-10-07	Salah satu hotel bintang 4 yang menurut saya terkesan mewah, harga juga masih bersahabat. Pelayanan oke dan ramah. Mantap pokoknya. Recommended, kalau ke Jogja balik lagi ke Grand Diamond.
10	GDHY	YB	10	2024-10-01	

Gambar 2: Keterangan atribut dataset

2.2. *Preprocessing* Data

Dataset yang telah dikumpulkan sebelumnya disimpan dalam format .csv, selanjutnya akan melalui tahap *preprocessing* terlebih dahulu sebelum dianalisis lebih lanjut. Berikut beberapa tahapan yang akan dilalui dalam tahap *preprocessing* ini:

2.2.1. *Cleansing Data*

Data yang telah dikumpulkan masih dalam bentuk mentah. Proses pembersihan data dilakukan untuk menghilangkan elemen yang tidak relevan dalam dataset, seperti *URL*, *email*, tanggal, kode HTML, *emoticon*, *hashtag (#)*, *mention (@)*, tanda baca, angka, serta spasi berlebih.

2.2.2. *Case Folding*

Selanjutnya, dataset akan diproses melalui tahap *case folding*, yang berarti mengubah semua huruf dalam setiap kata menjadi huruf kecil guna mempermudah proses analisis lebih lanjut[7].

2.2.3. *Tokenizing*

Proses pemecahan kalimat menjadi bagian yang lebih kecil yang terdiri dari kata per kata dikenal sebagai tahap tokenizing[8]. Penelitian ini, digunakan *library NLTK* untuk melakukan proses tokenisasi.

2.2.4. *Normalization*

Tahap normalisasi bertujuan untuk mengonversi kata-kata tidak baku menjadi bentuk baku sesuai dengan standar KBBI. Proses ini juga mencakup koreksi terhadap kata yang disingkat atau mengalami kesalahan ejaan. Dalam penelitian ini, digunakan kamus normalisasi untuk memastikan standarisasi kata-kata yang tidak baku.

2.2.5. *Stopwords*

Salah satu tahap *preprocessing* teks adalah *stopwords*; tujuan dari tahap ini adalah untuk menghilangkan kata-kata yang tidak penting dari setiap kalimat, yang disusun berdasarkan daftar *stopword* yang telah ditentukan sebelumnya.[9]. Dalam penelitian ini, digunakan *Sastrawi StopWordRemover* untuk menghapus *stopword* pada teks berbahasa Indonesia.

2.2.6. *Stemming*

Stemming adalah tahap terakhir dari *preprocessing*. Ini adalah proses mengubah kata ke bentuk aslinya dengan menghilangkan imbuhan seperti awalan, akhiran, kombinasi awalan dan akhiran, atau sisipan[10]. Penelitian ini menggunakan *Sastrawi Stemmer* untuk melakukan *stemming* pada kata-kata dalam bahasa Indonesia.

2.3. Translasi Teks

Setelah melalui tahapan *preprocessing*, dataset yang awalnya berbahasa Indonesia akan diterjemahkan ke dalam bahasa Inggris pada tahapan ini.

2.3.1. Translasi Teks

Dataset yang akan diterjemahkan adalah dataset yang telah bersih atau data yang telah melalui tahapan *preprocessing*. Dataset dalam bahasa Indonesia diterjemahkan ke dalam bahasa Inggris dengan menggunakan *Google Translator API.*

2.3.2. Validasi Translasi

Untuk mengantisipasi adanya nilai kosong atau terjadinya error, diperlukan pengecekan ulang pada dataset secara manual. Selain itu, perlu juga menangani kasus-kasus khusus yang terjadi dalam proses terjemahan, seperti kesalahan terjemahan, kata yang tidak dapat diterjemahkan, atau kalimat yang ambigu.

2.4. *Labelling*

Setelah dataset ditranslasi ke bahasa Inggris, dataset akan dilakukan tahap *labelling*. Pada tahap *labelling* ini, dataset akan dibagi menjadi tiga kelas, yaitu postif, negatif, dan netral. Penelitian ini menggunakan *library TextBlob* untuk menganalisis sentimen. *Textblob* adalah salah satu *library* yang disediakan oleh *Python* untuk pemrosesan dibidang *Natural Language Processing* yang dapat memberikan tag kata, ekstraksi kata, penerjemahan kata dan *sentiment analysis* [11].

2.5. TF-IDF

Setelah tahap *preprocessing*, translasi teks, dan pelabelan selesai, langkah berikutnya adalah menerapkan pembobotan TF-IDF pada dataset. Dimulai dengan menghitung berapa kali setiap kata muncul dalam dokumen dan seberapa sering kata-kata tersebut muncul di seluruh dokumen. Ini adalah rumus untuk menghitung Term Frequency (TF).

$$TF = \frac{Jumlah\ kemunculan\ kata\ dalam\ dokumen}{Jumlah\ total\ kata\ dalam\ dokumen} \quad (1)$$

Setelah mendapatkan nilai *Term Frequency* (TF), langkah berikutnya adalah menghitung Frekuensi Inverse Dokumen (IDF). Ini dilakukan untuk menentukan seberapa jarang suatu kata muncul di seluruh dokumen. Untuk meningkatkan relevansi analisis teks, kata-kata yang jarang muncul akan diprioritaskan.

$$IDF(t) = \log(\frac{Total\ jumlah\ dokumen}{jumlah\ dokumen\ yang\ mengandung\ kata\ tersebut}) \quad (2)$$

Langkah berikutnya adalah menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) dengan mengalikan kedua nilai tersebut.

$$TF - IDF\ = TF\ \times IDF \quad (3)$$

2.6. *Splitting Data*

Dataset dibagi menjadi dua bagian pada tahap pembagian data, atau *splitting data*. Data pelatihan digunakan untuk melatih model dengan algoritma yang digunakan, sedangkan data pengujian digunakan untuk menguji dan mengevaluasi kinerja model dalam melakukan prediksi[12].

2.7. *Support Vector Machine* (SVM)

Proses klasifikasi dengan algoritma SVM menggunakan kernel linear untuk memisahkan data berdasarkan *hyperplane* terbaik yang memaksimalkan margin antara kelas-kelas yang berbeda. Kemudian model dilatih menggunakan data yang telah melalui tahap *preprocessing*, sehingga dapat mengenali pola dengan lebih baik dan membuat keputusan yang optimal dalam klasifikasi. Setelah proses pelatihan selesai, model yang telah terbentuk disimpan untuk digunakan pada prediksi selanjutnya.

2.8. Evaluasi

Hasil evaluasi model akan dianalisis menggunakan *Confusion matrix* digunakan untuk mengevaluasi tingkat akurasi suatu metode dalam proses pengujian dengan menganalisis jumlah perkiraan yang salah dan

benar untuk masing-masing kelas[2]. Selain itu, evaluasi juga akan dilakukan dengan menggunakan *classification report*, yang mencakup metriks seperti *precision*, *recall*, dan *F1-score* untuk menilai kinerja model secara lebih mendalam.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

Untuk melakukan analisis sentimen, penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) berbasis kernel linear untuk menganalisis ulasan tamu tentang layanan di Grand Diamond Hotel Yogyakarta. Algoritma *Support Vector Machine* (SVM) berfungsi untuk memisahkan data dengan *hyperplane* optimal, sehingga mampu mengklasifikasikan sentimen tamu secara lebih akurat.
Dataset diperoleh dari situs Traveloka Grand Diamond Hotel Yogyakarta menggunakan teknik *web scraping* dengan bantuan ekstensi chrome https://webscraper.io/, data yang berhasil terkumpul sebanyak 1.119 baris data. Tabel 1 menunjukkan contoh hasil *web scraping* dataset.

Tabel 1 Contoh Hasil *Web Scraping* Dataset.

name	*rating*	*date*	*review*
F***n	10	2024-11-11	the best pokoke ga nyangka dapet Kamar Segede Itu dengan harga segitu
D***i	10	2024-11-09	staff sangat sangat ramah, rekomendasi sekali untuk menginap disini
G***t	6.6	2024-11-08	Bantal tidak empuk dan tipis, bathtub kurang bersih, pas saya wifi tersambung namun akses internet tidak bisa, TV kecil dengan siaran tidak bersih, perlengkapan mandi minim

Selanjutnya, dataset yang telah terkumpul akan melalui tahap *preprocessing*, meliputi *cleansing data*, *case folding*, *tokenizing*, *normalization*, *stopwords*, dan *stemming*. Tabel 2 berisi contoh hasil *preprocessing*.

Tabel 2 Contoh Hasil Tahapan *Preprocessing.*

Tahap *Preprocessing*	**Hasil**
Data mentah	the best pokoke ga nyangka dapet Kamar Segede Itu dengan harga segitu
Cleansing Data	the best pokoke ga nyangka dapet Kamar Segede Itu dengan harga segitu
Case Folding	the best pokoke ga nyangka dapet kamar segede itu dengan harga segitu
Tokenizing	[’the’, ’best’, ’pokoke’, ’ga’, ’nyangka’, ’dapet’, ‘kamar’, ‘segede’, ‘itu’, ‘dengan’, ’harga’, ‘segitu’]
Normalization	['the', 'best', 'pokoknya', 'tidak', 'menyangka', 'dapat', 'kamar', 'sebesar', 'itu', 'dengan', 'harga', 'begitu']
Stemming	['the', 'best', 'pokok', 'tidak', 'sangka', 'dapat', 'kamar', 'besar', 'itu', 'dengan', 'harga', 'begitu']

Sebelum tahapan *labelling*, dataset hasil dari tahap *preprocessing* telah menjadi data bersih. Kemudian, tahapan selanjutnya adalah translasi teks, dimana dataset berbahasa Indonesia akan diterjemahkan menggunakan API Google Translator. Tabel 3 menunjukkan contoh hasil translasi teks.

Tabel 3 Contoh Hasil Translasi Teks.

Hasil *Preprocessing*	Translasi Teks
['the', 'best', 'pokok', 'tidak', 'sangka', 'dapat', 'kamar', 'besar', 'itu', 'dengan', 'harga', 'begitu']	['the', 'best', 'core', 'not', 'expect', 'get', 'room', 'big', 'that', 'with', 'price', 'so']

Dataset yang telah melalui tahapan translasi teks selanjutnya akan melalui tahapan *labelling* yang mengkategorikannya menjadi tiga kategori: positif, negatif, dan netral. Pada tahapan *labelling* ini, digunakan *library TextBlob* untuk mempermudah analisis sentimen, karena *library TextBlob* lebih optimal digunakan dengan bahasa Inggris dibandingkan dengan bahasa Indonesia. Tabel 4 menunjukkan contoh hasil dari tahapan labelling.

Tabel 4 Contoh Hasil Tahapan *Labelling.*

Review	***Sentiment***
['the', 'best', 'core', 'not', 'expect', 'get', 'room', 'big', 'that', 'with', 'price', 'so']	Positif
['first', 'enter', 'bed', 'his', 'mattress', 'many', 'hair', 'dust', 'end', 'can', 'move', 'room', 'bathroom', 'his', 'cockroach', 'his', 'bed', 'can', 'dirty', 'ac', 'his', 'yes', 'cold', 'max', 'stay', 'temperature']	Negatif
['if', 'yogyakarta', 'maybe', 'become', 'alternative', 'stay', 'here']	Netral

Hasil dari dataset yang telah di *labelling* akan melalui tahapan TF-IDF, yang digunakan untuk menghitung bobot kata berdasarkan frekuensi kemunculannya dalam dokumen dan seberapa jarang kata tersebut muncul di seluruh kumpulan data, sehingga membantu meningkatkan efektivitas analisis dan klasifikasi teks. Tabel 5 menunjukkan contoh hasil tahapan TF-IDF.

Tabel 5 Contoh Hasil Tahapan TF-IDF.

Kata	D1	D2	D3	D4	D5
acara	0.0	0.0	0.0	0.0	0.0
ada	0.000000	0.0	0.0	0.0	0.0
air	0.000000	0.000000	0.162184	0.0	0.0
akan	0.0	0.0	0.0	0.0	0.0
akhir	0.0	0.0	0.0	0.0	0.0
akses	0.000000	0.000000	0.0	0.295183	0.0
aku	0.0	0.0	0.0	0.0	0.0
alas	0.0	0.0	0.0	0.0	0.0
alergi	0.0	0.0	0.0	0.0	0.0
alhamdulilsudah	0.0	0.0	0.0	0.0	0.0
alir	0.0	0.0	0.0	0.0	0.0

Selanjutnya, dilakukan tahap pembagian data, yang juga dikenal sebagai *splitting data*. Tahap ini membagi dataset yang telah melewati proses *preprocessing* dan translasi menjadi dua puluh persen untuk pengujian data dan delapan puluh persen untuk pelatihan data (80:20). Tabel 6 menunjukkan contoh hasil pembagian data.

Tabel 6 Contoh Hasil *Splitting Data.*

Set Data	Jumlah Data
Training	895
Testing	224
Total	1.119

Sesudah data terbagi menjadi training dan testing, langkah berikutnya adalah melatih model menggunakan algoritma *Support Vector Machine* (SVM) dengan kernel linear. Model ini akan dilatih menggunakan data training (80%) untuk mengidentifikasi pola dan menghasilkan keputusan klasifikasi.

Kode Program 1 Pelatihan Model SVM

```
def train_and_save(self, train_data):
    """Train SVM model and save it"""
    train_data = self.prepare_data(train_data)

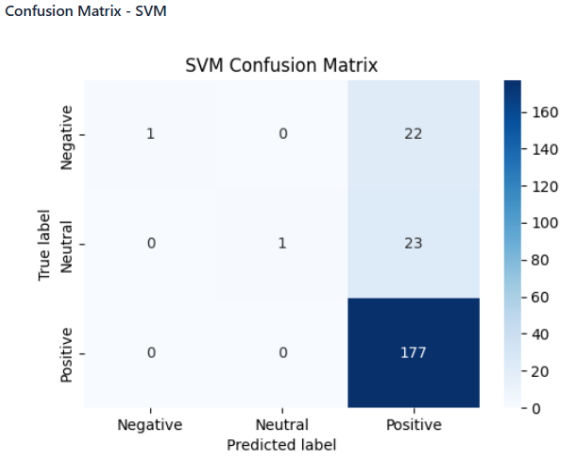
    X_train = train_data['review']
    y_train = train_data['sentiment_label']

    # Preprocessing dan vectorisasi
    X_train_processed = X_train.apply(self.preprocess_text)
    X_train_tfidf = self.vectorizer.fit_transform(X_train_processed)

    # Train SVM model
    logging.info("Training SVM classifier...")
    self.svm_classifier.fit(X_train_tfidf, y_train)
    logging.info("SVM training completed")

    # Save model and vectorizer
    joblib.dump(self.svm_classifier, 'sentiment_model_svm.pkl')
    joblib.dump(self.vectorizer, 'vectorizer.pkl')
```

Pada baris pertama dalam Kode Program 1, fungsi train_and_save() menerima parameter train_data, yang kemudian diproses menggunakan metode prepare_data(). Data pelatihan ini terdiri dari dua kolom utama, yaitu '*review*' sebagai input (X_train) dan 'sentiment_label' sebagai label kelas (y_train). Selanjutnya, tahap *preprocessing* dan vektorisasi dilakukan pada teks ulasan menggunakan metode preprocess_text(). Hasil *preprocessing* ini dikonversi ke dalam representasi numerik menggunakan TF-IDF vectorizer, yang memungkinkan model SVM memahami fitur teks secara matematis. Pada bagian pelatihan model SVM, fungsi fit() digunakan untuk melatih model dengan melihat bagaimana data ulasan berinteraksi yang telah dikonversi (X_train_tfidf) dan label sentimen (y_train). Proses ini dicatat dengan log informasi agar lebih mudah dipantau. Setelah pelatihan selesai, model SVM dan vektorisasi teks yang telah dilatih disimpan ke dalam file menggunakan joblib.dump() dengan nama sentiment_model_svm.pkl dan vectorizer.pkl. Langkah ini memungkinkan model yang telah dilatih untuk digunakan kembali tanpa perlu melakukan pelatihan ulang setiap kali sistem dijalankan. Setelah proses analisis selesai, hasil analisis data berupa label pada data testing akan diperoleh. Evaluasi model tidak hanya bergantung pada nilai akurasi, tetapi juga menggunakan classification report untuk menilai kinerja klasifikasi secara lebih mendalam. Hasil dari Matrix Confusion digambarkan pada Gambar 3.



Gambar 3: Hasil *Confusion Matrix*

Untuk mengevaluasi performa model, *Classification Report* digunakan untuk mengukur nilai akurasi, precision, recall, dan f1-score. Laporan klasifikasi hasil penerapan metode *Support Vector Machine* (SVM)

ditunjukkan pada Gambar 4.

Classification Report - SVM

Accuracy Score: SVM Model Accuracy: 0.80%
Report:

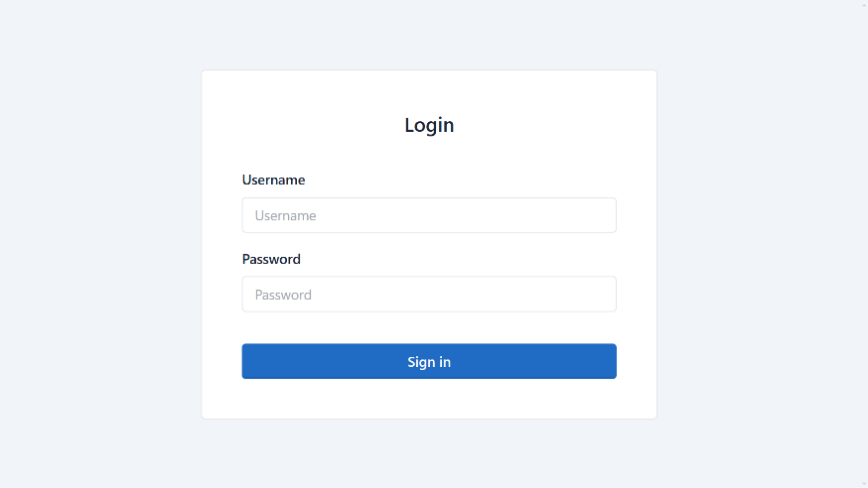
	precision	recall	f1-score	support
Negative	1.00	0.04	0.08	23
Positive	0.80	1.00	0.89	177
Neutral	1.00	0.04	0.08	24
accuracy			**0.80**	**224**
macro avg	0.93	0.36	0.35	224
weighted avg	0.84	0.80	0.72	224

Gambar 4: Hasil *classification report.*

Dari hasil *classification report*, menunjukkan bahwa analisis sentimen ulasan tamu terhadap layanan Grand Diamond Hotel Yogyakarta menghasilkan akurasi sebesar 80%. Model SVM menunjukkan mampu mengidentifikasi sentimen positif dengan sangat tepat sebesar 80% dan recall mencapai 100%, yang mengindikasikan bahwa model dapat mengidentifikasi seluruh ulasan positif secara akurat.

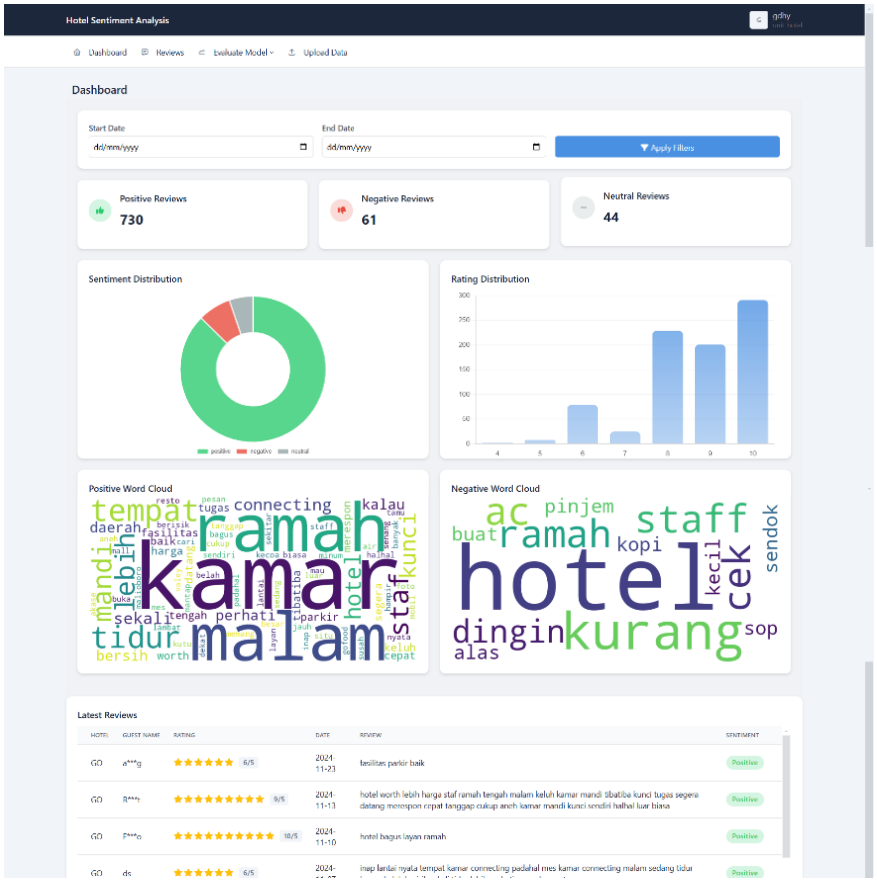
4.2. Implementasi Web

Setelah ulasan tamu Grand Diamond Hotel Yogyakarta dianalisis menggunakan algoritma *Support Vector Machine* (SVM), hasil analisis tersebut diimplementasikan ke dalam bentuk aplikasi berbasis web. Implementasi ini bertujuan untuk mempermudah manajemen hotel dalam memantau dan menganalisis sentimen tamu secara efisien. Seperti pada aplikasi pada umumnya, sebelum mengakses halaman *dashboard*, pengguna harus terlebih dahulu melewati halaman login. Halaman login dapat dilihat pada Gambar 5.



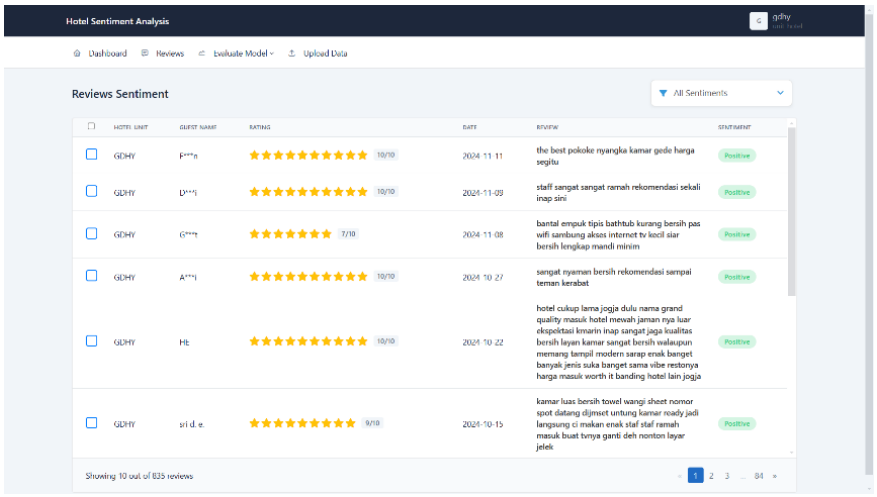
Gambar 5: Halaman Login.

Halaman selanjutnya, halaman *dashboard*, menyajikan hasil analisis dalam tiga kategori sentimen: positif, negatif, dan netral. Selain itu, *dashboard* memiliki visualisasi distribusi sentimen dalam bentuk diagram lingkaran, distribusi rating dalam bentuk diagram batang, dan kata-kata yang paling sering digunakan. Halaman *dashboard* ditunjukkan pada Gambar 6.



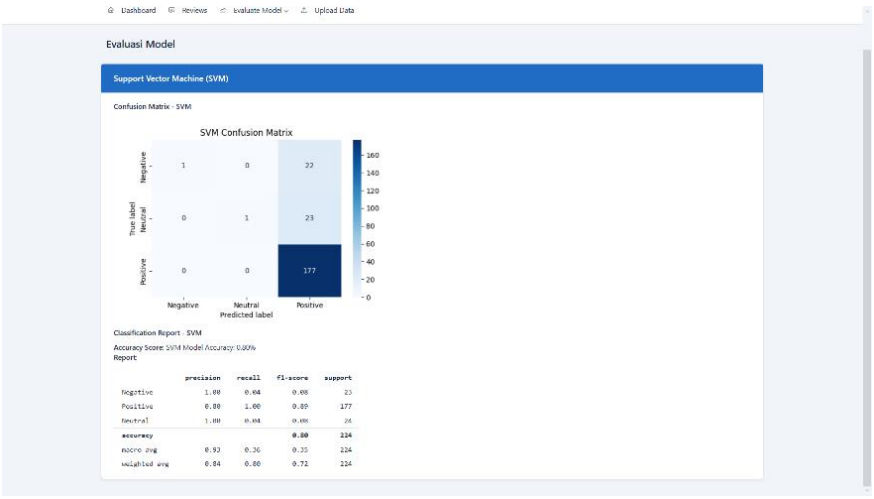
Gambar 6: Halaman *Dashboard*.

Halaman berikutnya adalah halaman *review*, yang menampilkan daftar semua ulasan tamu yang telah di label. Halaman ini juga dilengkapi fitur penyaringan berdasarkan kategori sentimen. Halaman *review* dapat dilihat pada Gambar 7.



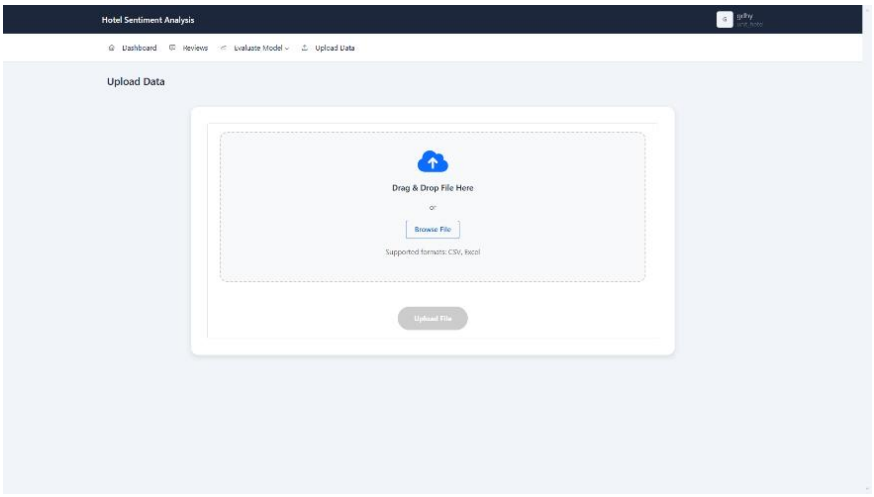
Gambar 7: Tampilan *Review*.

Selanjutnya halaman hasil evaluasi model, halaman ini menampilkan *Confusion Matrix* dan *Classification Report* untuk menilai kinerja model. Halaman evaluasi model ditunjukkan pada Gambar 8.



Gambar 8: Halaman Evaluasi Model.

Halaman terakhir berfungsi sebagai halaman *upload*. untuk menggunggah dataset yang berisikan ulasan-ulasan tamu dengan atribut *name*, *rating*, *date* dan *review*. Halaman *upload* ditunjukkan pada Gambar 9.



Gambar 9: Halaman *Upload.*

V. SIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa algoritma *Support Vector Machine* (SVM) berhasil mengklasifikasikan ulasan tamu tentang layanan Grand Diamond Hotel Yogyakarta ke dalam tiga kelas: positif, negatif, dan netral. Evaluasi dari model menghasilkan akurasi sebesar 80%, yang dimana model SVM memiliki peforma yang sangat baik dalam mengenali sentimen positif, dengan precision 80% dan recall 100%, yang berarti model mampu mengidentifikasi semua ulasan positif dengan akurat. Analisis sentimen ini telah diimplementasikan dalam aplikasi berbasis web untuk memudahkan manajemen hotel dalam memantau dan menganalisis sentimen tamu secara efisien, sehingga dapat mendukung peningkatan layanan dan pengambilan keputusan yang lebih cepat.

Namun, salah satu kekurangan dalam penelitian ini adalah model masih kurang optimal dalam mendeteksi sentimen negatif dan netral, sebagaimana ditunjukkan oleh recall yang sangat rendah sebesar 4% pada kedua kategori tersebut. Model cenderung mengklasifikasikan ulasan sebagai positif, sehingga diperlukan perbaikan lebih lanjut, seperti penyeimbangan data atau penyempurnaan parameter model, agar dapat mengenali ulasan negatif dan netral dengan lebih akurat. Sebagai saran untuk penelitian selanjutnya, disarankan untuk mengeksplorasi teknik *balancing data*, penggunaan model yang lebih kompleks, atau

kombinasi beberapa metode *machine learning* untuk meningkatkan akurasi klasifikasi sentimen secara keseluruhan.

DAFTAR PUSTAKA

[1] H. Ibna Pratama and P. Taqwa Prasetyaningrum, "Penerapan Support Vector Machine untuk Analisis Sentimen pada Google Review Hotel," *Journal of Information System Research*, vol. 6, no. 2, pp. 1246–1254, 2025, doi: 10.47065/josh.v6i2.6645.

[2] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi" *Jurnal SIMETRIS*, vol. 10, no. 2, 2019.

[3] P. Utami, R. Oktariani, N. Pratiwi, and H. Fakhrurroja, "Perbandingan Analisis Sentimen Aplikasi Traveloka dan Tiket.com pada Twitter dengan Metode Support Vector Machine," 2023.

[4] Irma Surya Kumala Idris, Yasin Aril Mustofa*, and Irvan Abraham Salihi, "Analisis Sentimen Terhadap Penggunaaan Aplikasi Shoppe Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura Journal of Electrical and Electronics Engineering*, Dec. 2023, doi: 10.1177/0165551510388123.

[5] M. S. Arrafiq and R. Kurniawan, "Analisis Sentimen Pengguna Terhadap Layanan Aplikasi Seabank Indonesia di Instagram Menggunakan Metode Support Vector Machine," *Journal of Information System Research*, vol. 5, no. 4, 2024, doi: 10.47065/josh.v5i4.5620.

[6] D. Deviacita *et al.*, "Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace," vol. 7, no. 4, 2019.

[7] G. Yunanda, D. Nurjanah, and S. Meliana, "Recommendation System from Microsoft News Data using TF-IDF and Cosine Similarity Methods," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, Jun. 2022, doi: 10.47065/bits.v4i1.1670.

[8] W. Athira Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," 2018. [Online]. Available: http://j-ptiik.ub.ac.id

[9] R. Rinandyaswara, Y. A. Sari, and M. T. Furqon, " Pembentukan daftar stopword menggunakan term based random sampling pada analisis sentimen dengan metode naïve bayes (studi kasus: kuliah daring di masa pandemi)" vol. 9, no. 4, 2022, doi: 10.25126/jtiik.202294707.

[10] Y. Anistyasari and E. Hariadi, " Algoritma baru pembentukan kata dasar pada proses stemming bahasa Indonesia" 2019.

[11] I. Gede, P. Megayasa, P. Pasek, O. Mahawardana, and P. R. Nurbawa, "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap Figure Pemimpin' Menggunakan Python" 2023. [Online]. Available: https://ojs.mahadewa.ac.id/index.php/jmti

[12] B. Wicaksono and N. Cahyono, " Analisis Sentimen Komentar Instagram pada Program Kampus Merdeka dengan Algoritma Naive Bayes dan Decision Tree" 2024. [Online]. Available: https://situs.com