


Sentiment Analysis System Of Bali Tourism Using Naive Bayes Algorithm And Web Framework

Nengah Widya Utami^{1*}, I Nyoman Purnama², I Made Ari Prayoga³

^{1*,2,3}Fakultas Teknologi Informasi dan Desain, Universitas Primakara

Article Info	ABSTRACT
Keywords: sentiment analysis, tourist, Bali, Naive Bayes, web framework	To obtain trends and impacts that may occur in the Bali tourism industry after the pandemic requires tourism actors to maintain the existence of the tourist beauty and culture they have. This research aims to develop a sentiment analysis system in the Bali tourism sector using the Naive Bayes algorithm and the Web Framework. This research stages carried out include Data Collection (Scraping), Data Cleaning, Feature Extraction, Modeling, and Web Platform Development. The data used was 2779 review data. The results show that most of the visitor reviews are in the "Very Positive" category, namely 1244. Next, 776 reviews are in the "Positive" category, 328 "Neutral". The words that appeared most frequently included "place", "walk", "beautiful", "nice". The evaluation results show that the Bayes algorithm shows an accuracy value of 71%, which means Naive Bayes produces sufficient accuracy for sentiment analysis. In this research, we succeeded in developing a website with a web framework to predict the sentiment of a review in real time and it is hoped that it can help related parties understand and respond to reviews more effectively, improve the tourist experience, and advance the tourism sector in Bali.
This is an open access article under the CC BY-NC license 	Corresponding Author: Nengah Widya Utami Fakultas Teknologi Informasi dan Desain, Universitas Primakara widya@primakara.ac.id

INTRODUCTION

Bali is a leading tourism destination in Indonesia with a high number of tourists visiting each year. The development of tourism in Bali post COVID-19 pandemic may vary greatly depending on government policies, the global situation related to the pandemic, and the level of tourist confidence. However, there are some trends and impacts that may occur in Bali's tourism industry after the pandemic including: 1) Economic Recovery, where Bali relies heavily on tourism as one of the main sources of income; 2) Adaptation to Health Protocols, where Bali and other tourism destinations will most likely continue to implement strict health protocols to ensure the safety of tourists and locals; 3) Tourism Diversification i.e. The pandemic has shown the risks associated with depending on one particular tourism sector. Therefore, Bali may try to diversify its tourism sector by offering diverse tourism experiences, such as ecotourism, culture, and sports; 4) Technology and Digitalization, where the tourism industry in Bali can adopt more technology and digital solutions to enhance the tourist experience and simplify destination management. This may include the use of apps for

bookings, cashless payments, and real-time tourist information; 5) Tourism Promotion, to recover the tourism industry, Bali may increase its destination promotion and marketing efforts. These campaigns may focus on safety, cleanliness, and Bali's natural and cultural beauty; 6) Infrastructure Development, where infrastructure development such as transportation, accommodation, and other public facilities may be undertaken to improve Bali's competitiveness as a tourist destination; and 7) Environmental Awareness, where Bali can take steps to engage local communities and tourists in sustainability initiatives.

Currently, Bali tourism is slowly growing and developing again. To obtain trends and impacts that may occur in the Bali tourism industry after the pandemic requires tourism actors to maintain the existence of the beauty of tourism and culture that is owned. Visitors or tourists to Bali have started to increase in several tourist sites in Bali. Travelers' reviews and opinions are widespread on online platforms such as travel sites, social media, and discussion forums have a huge impact on the image and growth of the tourism industry. Therefore, collecting, analyzing, and understanding review sentiment is crucial for decision-making and management of tourism destinations. With the development of information technology, searching for information about tourist attractions can now be easily searched.

One of the online booking sites that has been recognized by the global community, TripAdvisor provides services and information about tourism which has a role as a consideration for tourists who want to travel. (Suryawan et al., 2023). In addition, the review feature on Google Maps is one of the things from the big data era, where everyone can leave a review after they visit a place. (Irvandi et al., 2023). To generate information from opinion data on online platforms, it is necessary to analyze text mining with sentiment analysis for villas in Ubud by separating opinions into positive, negative and neutral sentiments. Text mining is a technique that can be used for document classification, clustering, information extraction, sentiment analysis, and information retrieval. Text mining is a type of data mining that seeks to find interesting patterns in large amounts of text data. (Ginantra et al., 2022).

Sentiment analysis is used to identify and analyze user perceptions of digital applications, such as web-based information systems and social media, to optimize the marketing of business unit products and services according to market demand. In the context of traditional marketing of tourist destinations, it is necessary to know consumer perceptions of tourism products to analyze the effect of product promotion on the decision to visit a tourist destination. (Singgalen, 2021). Sentiment analysis is also known as opinion mining, which combines text mining and natural language processing. Text mining means adding data in the form of text obtained from documents. The purpose of text mining is to find words that describe documents so that it can perform analysis that builds relationships between these words. (Sari & Hayuningtyas, 2019).

The Naïve Bayes algorithm is one of the algorithms included in the classification. (Sari & Hayuningtyas, 2019). Naïve Bayes is a simple classification method based on Bayes' theorem with strong naive assumptions as the name suggests. Naïve Bayes makes the assumption that the presence or absence of a particular feature of a class is not related to other features. (Widyarto et al., 2023). The purpose of classification is to determine the class of a review whether it is in the positive class or the negative class. An important step of the

classification process is the representation of a sentence or document into numerical form so that it can be understood by a computer. (Putu et al., 2021).

The research that conducted sentiment analysis of Bali tourism reviews is research conducted by the researcher himself with the title "Sentiment Analysis of Tourist Reviews on Ubud Tourism Objects Using the Support Vector Machine Algorithm". The results obtained in this study produced positive sentiment as much as 551 data and negative sentiment as much as 118 data based on test data as much as 669 data, these results resulted in a positive value on tourist objects in Ubud with the results of accuracy of 84.01%, recall of 89.83%, precision of 90.40% and F1-Score of 90.11%. (Suryawan et al., 2023). In addition, Edgarsa, et al (2023) also conducted a study entitled "Implementation of the Naïve Bayes Classifier Method for Sentiment Analysis of Tourist Attractions in West Nusa Tenggara". The results of sentiment analysis using naïve bayes produce an F1-score macro worth 0.76 or 76%. (Widyarto et al., 2023).

Thus, the researcher conducted a study entitled "Web-based Bali Tourism Sentiment Analysis System Using Naive Bayes Algorithm and Web Framework". This system will process tourist reviews from various online platforms to extract tourist visitor sentiments towards tourism destinations in Bali. The use of this technology is expected to help related parties understand and respond to reviews more effectively, improve the tourist experience, and advance the tourism sector in Bali.

METHODS

Literature Review

The steps in this research are as follows.

- Data Collection: Review data from various online platforms (travel sites, social media, discussion forums) will be collected and stored for analysis.
- Data Cleaning and Processing: This stage includes data cleaning, tokenization, stop words removal, and text normalization.
- Feature Extraction: Important features of the text will be extracted using methods such as TF-IDF or word embeddings.
- Modeling with Naive Bayes Algorithm: Implementation of Naive Bayes algorithm for sentiment classification of the processed reviews.
- Web Platform Development using Flask: Building a web platform that displays sentiment analysis results with informative and user-friendly visualizations.
- Evaluation and Validation: Testing the system using test data to measure sentiment classification accuracy and overall system performance.

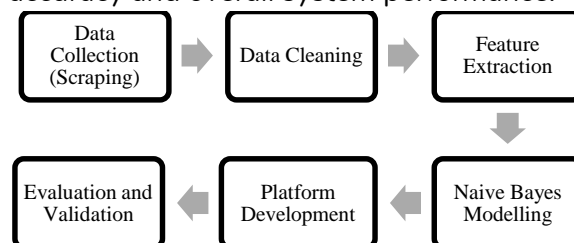


Figure 1: Research steps

RESULTS AND DISCUSSION

Data Collection (Scraping)

Data collection is intended to obtain data that will be used in research. The data used for this sentiment analysis research are reviews of some of the most frequently visited tourist attractions in Ubud through the Google Map platform, namely Monkey Forest, Goa Gajah, Ubud Art Market, Ceking Rice Terrace, and Campuhan Ridge. In the scraping process using Instant Data Scraper. Data obtained in the last 1 (one) year amounted to 2779 review data. Here is an example of data that was successfully scraped.

Table 1. Review Dataset

No.	Review
1	I was a little nervous at first but it was a fun experience. It's very cool to see the monkeys running around, climbing, and eating. The baby monkeys are especially adorable. As long as you follow the rules you should be fine! Some of them were a little mischievous where as others were very chill resting.
2	Very good place to have a walk and see cute monkeys. Be careful though they are still wild animals so just follow the rules of the park and you'll be fine! During the weekend there's a traditional dance representation as well so that's good :)
3	The forest is so beautiful! We loved all the trees, vines, plants, the creek and the Koi
4	Beautiful and the monkeys are friendly because they feed them all the time. Staff is excellent, helpful and place has temple. Don't touch the monkeys nor feed them, they might however jump on you to say hi did you bring me something?
5	An amazing experience that has to be on your list when in Ubud. The reviews of aggressive monkeys is something we did not see and the monkeys
...	
2780	Nice calm walk when you want to relax from Ubud daily rush.

Data Cleaning

The cleansing process in the data preprocessing stage is done by eliminating duplicate data and missing values. This process uses the Google Colab platform. Data cleansing by identifying and removing missing or invalid values, normalizing data if needed, removing duplicate data if any, and adjusting data types if needed.

```
data.dropna(inplace=True)
data.drop_duplicates(inplace=True)
data.to_excel('dataset_clean.xlsx',
index=False)
print(data.head())
```

The results of data cleaning show that from a dataset of 2779 to 2704 are shown in the following figure.



Figure 2. Data cleaning results

Feature Extraction

To perform feature extraction on clean datasets in Google Colab using various feature extraction techniques provided by libraries such as scikit-learn or pandas.

```
Mean Features:
Series([], dtype: float64)

Median Features:
Series([], dtype: float64)

Standard Deviation Features:
Series([], dtype: float64)
```

Feature 3 Extraction Results

Naive Bayes Modeling

In the application of naive bayes, the labeling process is carried out first. The labeling process is done automatically with the BERT algorithm. The results of the labeling process are as in the following table.

```
tokenizer = AutoTokenizer.from_pretrained('nlptown/bert-base-multilingual-uncased-sentiment')
model =
AutoModelForSequenceClassification.from_pretrained('nlptown/bert-base-multilingual-uncased-sentiment')
```

Dataset after Labeling

Review	score	category
I was a little nervous at first but it was a fun experience. It's very cool to see the monkeys running around, climbing, and eating. The baby monkeys are especially adorable. As long as you follow the rules you should be fine! Some of them were a little mischievous where as others were very chill resting.	4	Positive
The forest is so beautiful! We loved all the trees, vines, plants, the creek and the Koi pond. ...	5	Very Positive
Beautiful and the monkeys are friendly because they feed them all the time. Staff is excellent, helpful and place has temple. Don't touch the monkeys nor feed them, they might however jump on you to say hi did you bring me something? But don't...	4	Positive
An amazing experience that has to be on your list when in Ubud.	5	Very Positive

From the results of the labeling process, it is obtained that most visitor reviews are in the "Very Positive" category, namely 1244. Next 776 reviews fall into the "Positive" category, "Neutral" as many as 328, "Negative" as many as 173 and "Very Negative" as many as 259 reviews.

```
sentiment_category
Very Positive    1244
Positive         776
Neutral          328
Very Negative    259
Negative         173
Name: count, dtype: int64
```

Figure 4. Sentiment Category Results

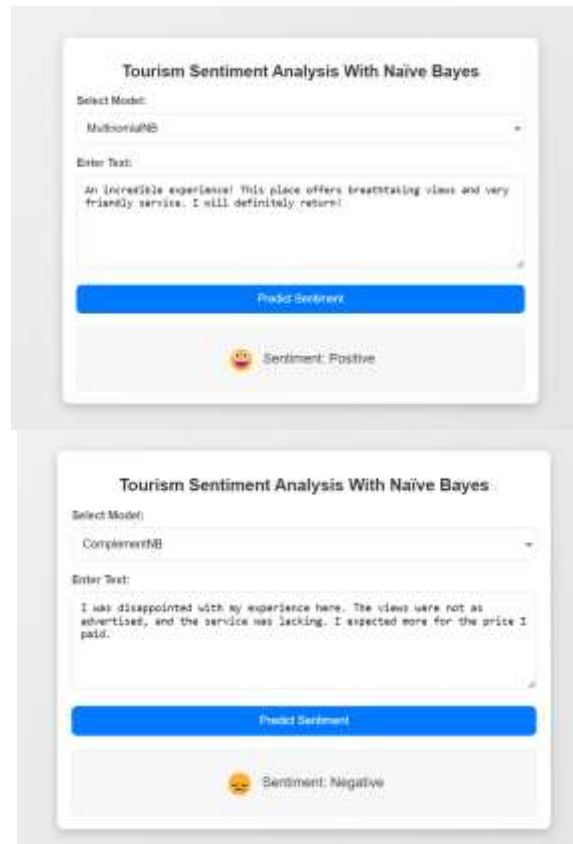


Figure 7. Web Framework

Evaluation and Validation

The evaluation includes evaluation in terms of web functionality. In terms of the web successfully predicting sentiment on sentences inputted by users.

CONCLUSION

Based on the research results of the Web-Based Bali Tourism Sentiment Analysis System Using Naive Bayes Algorithm and Web Framework, it is expected to make it easier for users to see the visualization results of the application of the naive bayes model on the dataset of tourist attractions reviews in Ubud. The research stages include Data Collection (Scraping), Data Cleaning, Feature Extraction, Naive Bayes Modeling, and Web Platform Development. The data used amounted to 2779 review data. The results show that visitor reviews are mostly in the "Very Positive" category. The most common words include "place", "walk", "rice", "beautiful", "nice". The evaluation results show that the bayes algorithm with shows an accuracy value of 0.71 which means naive bayes produces sufficient accuracy for sentiment analysis. This research successfully developed a website with a web framework to predict the sentiment of a review in real time.

REFERENCE

Ginantra, N. L. W. S. R., Yanti, C. P., Prasetya, G. D., Sarasvananda, I. B. G., & Wiguna, I. K. A.

- G. (2022). Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-NN. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 11(3), 205–215. <https://doi.org/10.23887/janapati.v11i3.49450>
- I. W. B. Suryawan, N. W. Utami, and K. Q. Fredlina, “Analisis Sentimen Review Wisatawan pada Objek Wisata Ubud Menggunakan Algoritma Support Vector Machine,” *J. Inform. Teknol. dan Sains*, vol. 5, no. 1, pp. 133–140, 2023.
- Irvandi, Irawan, B., & Nurdiawan, O. (2023). Naive Bayes Dan Wordcloud Untuk Analisis Sentimen Wisata Halal Pulau Lombok. *INFOTECH Journal*, 9(1), 236–242. <https://doi.org/10.31949/infotech.v9i1.5322>
- Putu, N. L. P. M., Ahmad Zuli Amrullah, & Ismarmiaty. (2021). Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 123–131. <https://doi.org/10.29207/resti.v5i1.2587>
- Sari, R., & Hayuningtyas, R. Y. (2019). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website. *Indonesian Journal on Software Engineering (IJSE)*, 5(2), 51–60. <https://doi.org/10.31294/ijse.v5i2.6957>
- Singgalen, Y. A. (2021). Analisis Sentimen dan Pemodelan Topik dalam Optimalisasi Pemasaran Destinasi Pariwisata Prioritas di Indonesia. *Journal of Information Systems and Informatics*, 3(3), 459–470. <https://doi.org/10.51519/journalisi.v3i3.171>
- Suryawan, I. W. B., Utami, N. W., & Fredlina, K. Q. (2023). Analisis Sentimen Review Wisatawan pada Objek Wisata Ubud Menggunakan Algoritma Support Vector Machine. *Jurnal Informatika Teknologi Dan Sains*, 5(1), 133–140.
- Widyarto, E. B., Informatika, F., Telkom, U., Informatika, F., Telkom, U., Lhaksmana, K. M., Informatika, F., Telkom, U., & Belakang, A. L. (2023). *Implementasi Metode Naive Bayes Classifier Terhadap Analisis Sentimen Tempat Wisata di Nusa Tenggara Barat*. 10(5), 4934–4941.
- W. Khofifah, D. N. Rahayu, and A. M. Yusuf, “Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps,” *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 16, no. 4, pp. 28–38, 2022, doi: 10.35969/interkom.v16i4.192.