BIJMT : Brilliant International Journal Of Management And Tourism

E- ISSN: 2827-8380
P- ISSN: 2810-076X

Research Article

Guest Sentiment Analysis Based on Online Reviews to Optimize Guest Satisfaction at Hotel X

I Dewa Gde Tri Jaya Cahyadi [1], Putu Diah Sastri Pitanatri [2], Prastha Adyatma [3*]

[1] Politeknik Pariwisata Bali; email : prasthaadyatma8@gmail.com
[2] Politeknik Pariwisata Bali; email : diahsastri@ppb.ac.id
[3] Politeknik Pariwisata Bali; email : prasthaadyatma8@gmail.com
* Corresponding Author: Prastha Adyatma

Abstract: Tourism has become one of the largest and fastest-growing industries globally. Advances in technology and communication have brought significant changes in various aspects, especially in the hospitality industry. Using a dataset from November 2024 to January 2025, this sentiment analysis was conducted using the Naive Bayes classification method (Gaussian, Multinomial, and Bernoulli). The results show that 72.99% of reviews are positive, while 19.08% are negative and 7.93% are neutral. The Naive Bayes model demonstrates high accuracy in classifying positive sentiment but exhibits differences in classification accuracy for the negative and neutral categories due to class imbalance. Occupancy data reveals a peak in 2023 and a significant decline in 2024. This study reveals the importance of ongoing sentiment analysis to establish management strategies, address service gaps, and improve guest satisfaction, which aims to improve guest satisfaction in the competitive hospitality market.

Keywords: Sentiment Analysis, TripAdvisor, Guest Satisfaction, Naïve Bayes, Hotel Manajement

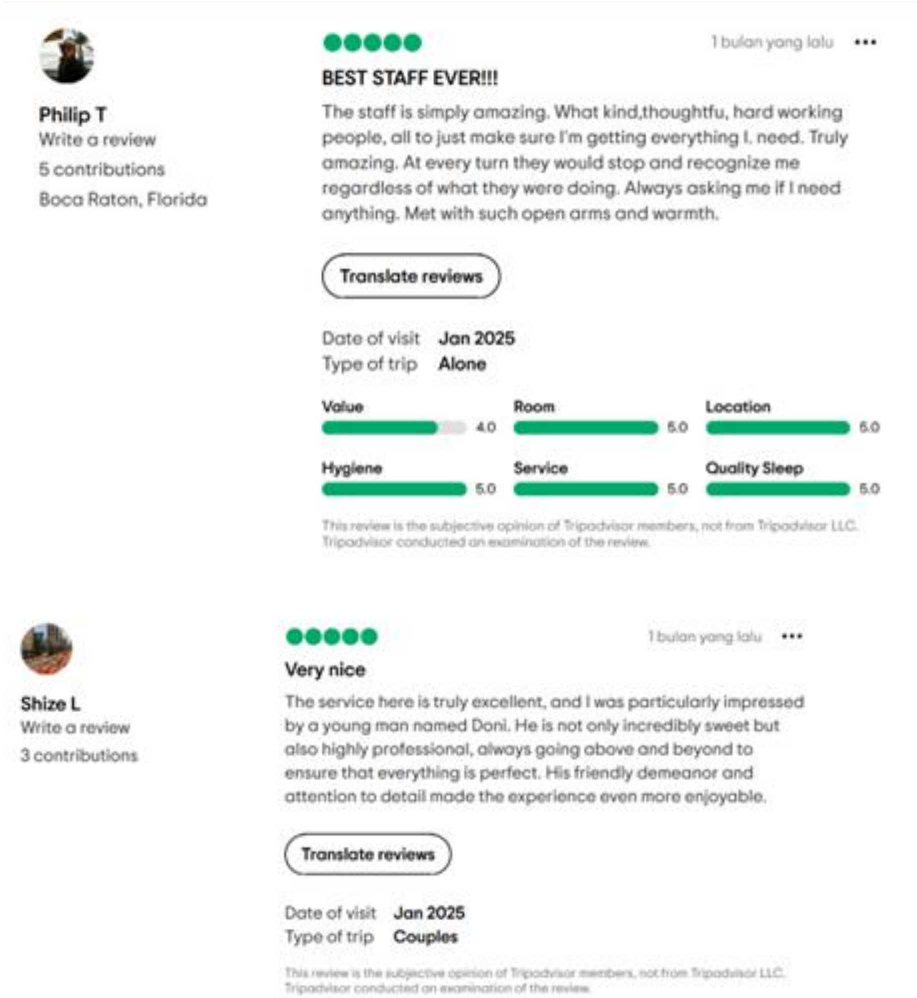
1. Introduction

Advances in technology and communication have brought about significant changes in various aspects, particularly in the hospitality industry. The success of a hotel business heavily relies on guest satisfaction. According to Herlambang & Komara (2022), customer satisfaction is the process of customer evaluation where the assessment exceeds customer expectations. However, measuring customer satisfaction traditionally is often quite challenging. Sentiment analysis, which utilizes text data from online reviews, provides a more in-depth approach to understanding guest perceptions. In this context, online reviews significantly influence guests' decisions when choosing accommodation. Online reviews themselves are shared by consumers regarding the products or services they have used. Through online reviews, prospective guests have the opportunity to understand the perspectives of other guests in assessing product quality and service satisfaction levels based on previous user experiences.

TripAdvisor has emerged as a major and influential online review site in the hospitality sector, where users can share their experiences with various travel-related services (Xiang et al., 2017). Important aspects of online reviews include review quality, review consistency, number of reviews, and updated information provided.

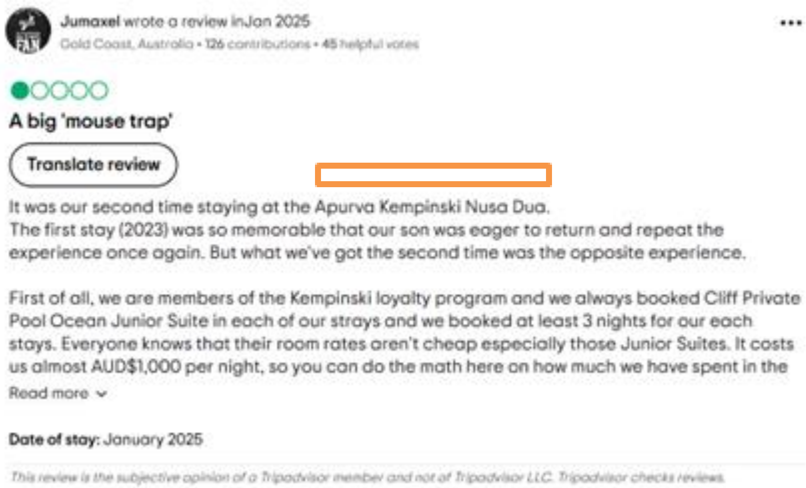
Received: May 19, 2025
Revised: June 01, 2025
Accepted: June 16, 2025
Published: June 20, 2025
Curr. Ver.: June 20, 2025

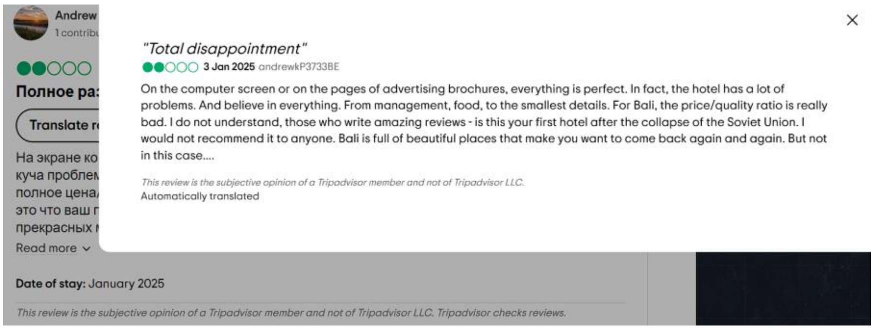
Figure 1. Sample Review from a Satisfied Customer of Hotel X



Source : TripAdvisor

Figure 2. Sample Review From a Dissatisfied Customer Regarding Hotel X's Performance





Source : TripAdvisor

Figure 1 displays a guest review expressing satisfaction with the service provided. Meanwhile, Figure 2 shows an example of a guest leaving a dissatisfied review about Hotel X's service on TripAdvisor. It can be concluded that every guest has their own expectations, and sometimes these expectations aren't met, leading to consumer dissatisfaction. This highlights the hotel management's task to always set targets to ensure consumers using their accommodation services feel their expectations are fulfilled.

Tabel 1. Occupancy X Hotel From 2022 – 2024

NO.	TAHUN	TOTAL ROOM OCCUPIED	TOTAL RATA-RATA OCCUPANCY (%)
1	**2022**	**108,089**	**63.21%**
2	**2023**	**136,475**	**79.81%**
3	**2024**	**138,695**	**81.05%**

Source: Data processed from Hotel X, (2024) Based on the table above, in 2024, Hotel X experienced an insignificant increase in its total average occupancy, growing by only 1.24% compared to 2023. This figure might be influenced by several external and internal factors related to customer satisfaction. This can make consumers more cautious in their purchasing decisions, especially for more exclusive or higher-priced destinations like Hotel X, which holds the status of a luxury hotel with a rather fantastic price.

According to BPS (Statistics Indonesia) data from 2023, an increase in occupancy rates below 3% per year is considered a poor indicator for the hotel business, as it signifies slow revenue growth and can point to unfavorable market conditions. This could indicate a decline in the hotel's profitability and business sustainability. Furthermore, intense competition from alternative accommodation options and a decreasing interest in luxury hotels add to the challenges for this hotel.

The analysis of occupancy growth from 2022 to 2023 shows a significant increase. However, the increase in 2024 indicates stagnant growth, a condition in which growth is slow, with a product's growth being less than 2-3% per year.

Based on sentiment analysis from online reviews on platforms like TripAdvisor, it can be seen whether there are issues with service quality or certain aspects of facilities that customers deem inadequate. Recurring negative reviews can provide clues about strategies that need improvement.

The objective of this research is to gain an understanding of customer perceptions and sentiments regarding various aspects of services provided by Hotel X, and to identify service aspects that need improvement to enhance customer satisfaction.

2. Literature Riview

- Sentimen Analysis

Technology enables customers to provide product reviews, which significantly impact potential customer purchasing decisions more than a company's marketing activities. This is because online reviews left by experienced customers are recognized as objective and reliable information (Alnsour, 2018).

Sentiment analysis focuses on an individual's perspective expressing or implying positive or negative sentiment. Most sentiment analyses are related to people on social media (Supriyanto et al., 2023).

The application of sentiment analysis can be used to assess and classify guest reviews based on the positive, negative, or neutral feelings contained within the text. With sentiment analysis, hotels can discover what their guests like. Once the data is collected, the next step is to analyze sentiment patterns to identify aspects that receive appreciation or negative comments from guests, such as staff behavior and service, room cleanliness, facility comfort, or even the hotel's location.

The information obtained from such analysis can help hotels understand customer preferences and provide insights for improving service quality. For instance, if reviews indicate that guests frequently complain about room cleanliness or slow staff responses, management can immediately take action to rectify these issues. However, if reviews highlight positive aspects such as excellent facilities or a strategic location with a good view, the hotel can use this information as promotional material to attract more customers. Research by Supriyanto et al. (2023) also emphasizes that sentiment analysis can assist hotels in designing more effective marketing strategies by highlighting their competitive advantages based on customer comments.

- Online Reviews

TripAdvisor is one of the largest and most popular online review platforms in the world, allowing travelers to share their experiences regarding hotels, restaurants, tourist destinations, and various other attractions. According to a 2021 study by Travel Weekly, approximately 83% of travelers stated that online reviews influenced their decisions when choosing accommodation. This platform enables travelers to openly share their experiences, which then becomes an important source of information for prospective customers in their decision-making process.

Reviews on TripAdvisor are considered more trustworthy than the service providers' own promotional content, as the information provided is current and based on direct experience. The influence of TripAdvisor reviews on customer decisions is very significant, as consumers are more likely to choose businesses with positive ratings and reviews.

Online reviews on the TripAdvisor platform are divided into 3 categories: positive reviews, which are further divided into two parts, "Excellent" with a rating of 5 and "Very good" with a rating of 4; neutral reviews, which are "Average" with a rating of 3; and negative reviews, which are divided into two parts, "Poor" with a rating of 2 and "Terrible" with a rating of 1. These ratings provide an overview of a guest's experience with a hotel or tourist destination.

Research by Ariansyah (2025) indicates that 87.5% of travelers read online reviews before deciding to book a room, with 84.4% of them stating that these reviews influenced their purchasing decision for a service product. Additionally, the reputation and ranking of a hotel or tourist destination heavily depend on guest reviews. High rankings on the TripAdvisor platform often increase a property's attractiveness, while negative reviews can damage an established reputation.

- Customer Satisfaction

Customer satisfaction is a feeling of pleasure or disappointment that arises after comparing a product's perceived performance against expectations (Maulidiah et al., 2023). According to Kelly (2020), customer satisfaction is a measurement or indicator of the extent to which customers or users of a company's product or service are very pleased with the products or services received. Customer satisfaction is a comparison between expectations and perceived experience.

Customer satisfaction is key to the success of a business. Satisfied customers tend to be loyal and recommend their experiences to others, which helps enhance reputation and attract new customers. Customer satisfaction itself is influenced by the quality of the product or service, appropriate pricing, good service, and an overall pleasant experience.

According to Indrasari in Mabrur et al. (2022), in determining the level of customer satisfaction, there are five main factors that companies must consider:

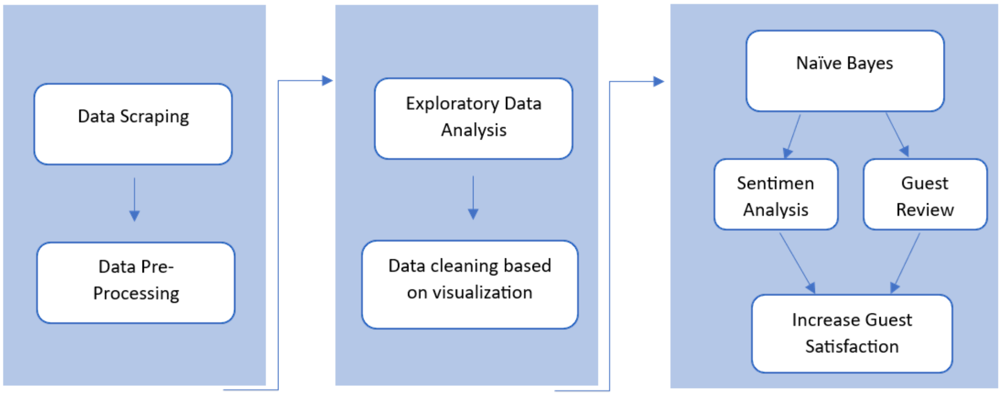
- Product Quality. Customers will feel satisfied if their evaluation shows that the product they used is indeed of high quality.

- Service Quality. Customers will feel satisfied if they receive good service or service that meets their expectations. This is closely related to their comfort with the product they chose.

- Emotional. Customers will feel proud and confident that others will admire them if they use certain products that tend to have a higher level of satisfaction.

- Price. Products of the same quality but with a relatively lower price offer greater value to customers. If customers feel that the comparison between price and the quality received is not proportionate, they will feel dissatisfied.

• Cost. Customers who do not incur additional costs or do not need to waste time to obtain a product tend to be satisfied with that product.



Source : Tinjauan Pustaka 2025

3. Method

This research focuses on guest reviews for Hotel X on the TripAdvisor platform to understand customer sentiment towards hotel services and its impact on guest satisfaction levels. The study was conducted at a hotel located in Nusa Dua, Bali, with the research period spanning from November 2024 to January 2025. This timeframe is designed to ensure a more in-depth data collection and analysis of the phenomenon being studied.

The research began with data collection, where data was obtained through online reviews on TripAdvisor using Python. Python is an interpretive programming language that is easy to learn and can run on various platforms, with a focus on code readability. Duplicate data was then removed. This data formed the dataset used for the research. The test data will undergo data processing to remove unnecessary information.

The data used in this study consists of secondary sources in the form of guest reviews from the TripAdvisor platform, which were cleaned and processed to be valid and relevant for primary analysis. This "absolute population" includes all review data relevant to the scope of the research. Operational variables in this study include TripAdvisor reviews, satisfaction levels, time periods, sentiment classification, and observed trends over time. These variables are measured through indicators such as the number of reviews per period, review ratings, sentiment polarity (positive, negative, neutral), occupancy percentage, and fluctuations.

To collect data, this research employed an API-based scraping method, primarily utilizing Python programming for automation and data extraction. TripAdvisor review data was retrieved and organized for analysis, with the assistance of Google Colab to run scripts and manage data over a five-year span, which was then exported to Excel format.

Data cleaning is a crucial process to ensure the accuracy and consistency of the dataset. This process includes identifying and removing errors and irrelevant information. Cleaning was performed using Python, specializing in Natural Language Processing (NLP), which supports sentiment classification, topic identification, and deeper insights into customer feedback.

The data needs to be processed to be usable. After preprocessing, the dataset is ready to be implemented with several models for sentiment analysis. Then, Naive Bayes Classification (NBC) was used for classification, including Gaussian, Multinomial, and Bernoulli Naive Bayes.

Bayesian classification statistics refer to methods that can estimate the probability of data belonging to a particular class. This type of classification relies on Bayes' theorem, which operates similarly to trees and neural networks in its classification process. Bayesian classification has been shown to achieve significant precision and efficiency when used with large datasets. NBC is one of the simplest classification algorithms, characterized by its simplicity and strong data accuracy.

The Naive Bayes formula is used for the posterior probability of a class based on observed data. The Naive Bayes formula can be expressed as follows:

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

Explanation:

- P(Y|X) (Posterior Probability) = The probability of the hypothesis, representing the outcome we want to achieve.

- P(X|Y) (Likelihood) = The probability that data X is observed, given that hypothesis Y is true.

- P(Y) (Prior Probability) = The initial assessment of how likely hypothesis Y is to be true before considering any data.

- P(X) (Normalizing Constant / Evidence) = The probability that data X is completely independent of its category.

4. Result and Discussion

Text preparation procedures or TF-IDF (Term Frequency - Inverse Document Frequency), which consist of text cleaning, standardizing case format, grouping text into individual tokens, and removing stop words, play a crucial role in preparing text data for thorough analysis. Implementing these steps in sequence leads to finer, more uniform, and user-friendly text data for machine learning systems, ultimately enhancing the accuracy and efficiency of text data processing.

Tabel 2. Result Sentimen Analysis

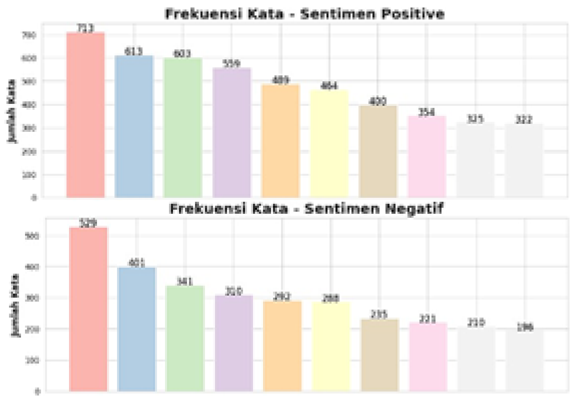
No	Sentimen	Hasil
1	**Positif**	1.001
2	**Negatif**	261
3	**Neutral**	108

Source: Data processed, 2025

Of all the reviews examined, 1,667 (72.99%) were considered Positive, indicating that a large number of guests were pleased with their experience and that their expectations were met. Conversely, 178 (19.08%) reviews were marked as Negative, showing dissatisfaction related to hotel services or facilities. Furthermore, 79 (7.93%) reviews were identified as Neutral, indicating a lack of clear positive or negative feelings. The prevalence of favorable sentiment analysis suggests that guest opinions about Hotel

X are quite optimistic; however, with the presence of negative sentiment, management must still take such feedback seriously.Gambar 1. Perbandingan Frekuensi kata sentiment positif dan negatif

Figure 1. Comparison of Positive and Negative Sentiment Word Frequencies



Source: Data Processing, 2025

The chart depicting frequently appearing terms in reviews with a positive tone shows words like "swimming pool", "villa", "staff", and "hotel", indicating that accommodation and services are usually cited favorably by those providing feedback.

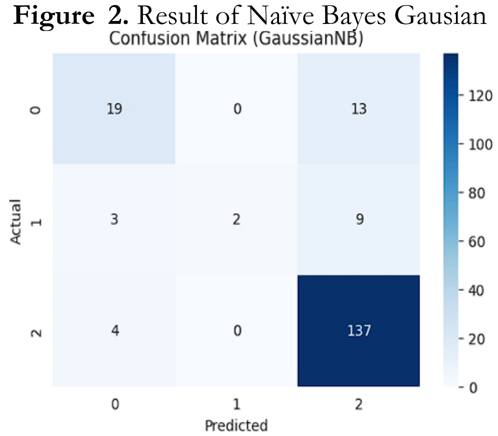
The chart depicting frequently appearing terms in reviews reflecting negative feelings shows words like "restaurant", "view", "bad", and "staff", indicating dissatisfaction with the accommodation. This suggests that most negative feedback comes from individuals using formal language patterns.

Naïve Baiyes Clasification (NBC)
Probabilitas Sebelumnya:

- **Positif : 0.729904**
- **Negatif : 0.190782**
- **Neutral : 0.079314**

The prior probabilities indicate the initial likelihood of sentiment before performing Naive Bayes classification. The probability of positive sentiment is the highest at 72.99%, indicating that most of the initial data consists of positive sentiment. Negative and neutral sentiments have probabilities of 19.08% and 7.93%, respectively. In Naive Bayes, these numbers serve as the basis for determining the probability of data affiliating with a particular sentiment class after evaluating its characteristics.

Naïve Baiyes Gausian
Figure 2. Result of Naïve Bayes Gausian

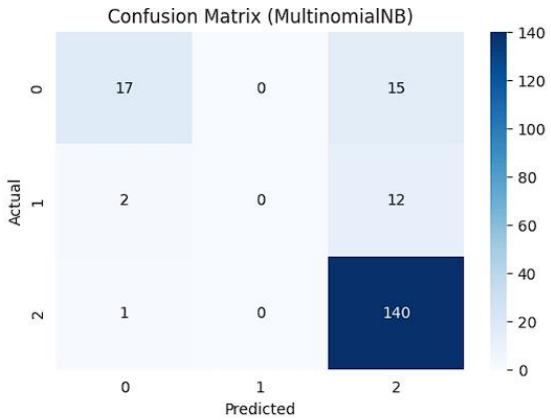


Source: Data Processing, 2025

Based on the results obtained from data processing with Gaussian Naive Bayes, this model achieved an accuracy rate of 84.49%. The model excels at identifying the "Positive" class (precision 0.86, recall 0.97, f1-score 0.91), but its effectiveness for the "Negative" class (f1-score 0.66) and "Neutral" class (f1-score 0.25) is still inadequate. This is illustrated in the matrix, which shows that many "Neutral" and "Negative" instances were inaccurately labeled as "Positive". Although the model demonstrates strong performance in recognizing positive sentiment, it has shortcomings in distinguishing between neutral and negative sentiments.

Naïve Baiyes Multinominal

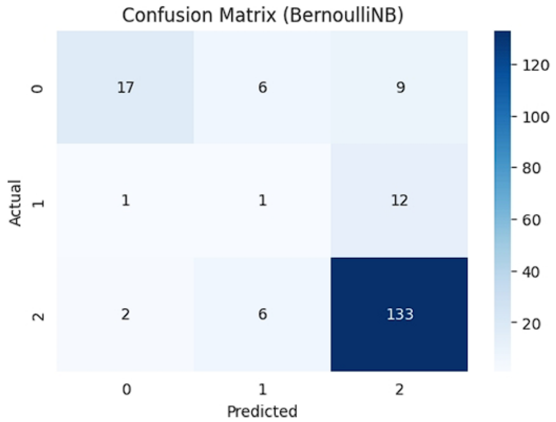
Figure 3. Result of Naïve Bayes Multinominal



Source: Data Processing, 2025

Based on the matrix depicted in the MultinomialNB graph, this model demonstrates strong capability in identifying class 2, successfully predicting 140 out of 141 instances. Specifically, 15 entries belonging to class 0 and 12 from class 1 were mistakenly identified as class 2, and there was no correct identification for class 1. Consequently, this model shows a preference for class 2, resulting in a reduced ability to accurately recognize classes 0 and 1.

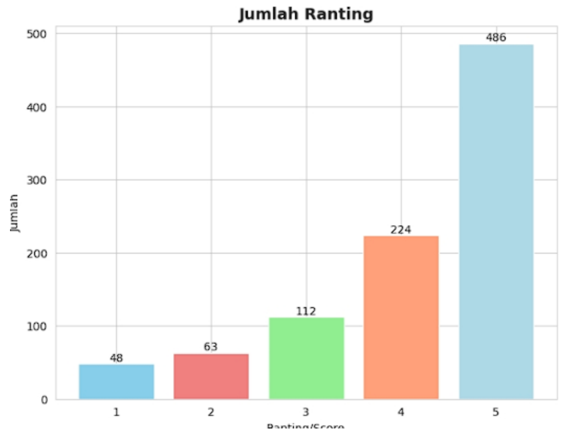
Naïve Baiyes Bernauli
Figure 4. Result of Naïve Bayes Bernauli



Source: Data Processing, 2025

From the analysis of the BernoulliNB matrix, it is evident that this matrix works quite effectively in identifying class 2, achieving 133 correct predictions out of a total of 141 actual instances. However, there are still errors in the classification of class 0 and class 1. The model accurately identified 17 instances of class 0, but it incorrectly categorized 6 instances of class 0 as class 1 and 9. In short, although the BernoulliNB model demonstrates strong performance in classifying class 2, it struggles more in differentiating between class 0 and class 1, especially in class 1, which has low prediction accuracy.

Overall Rating
Figure 5. Overall Rating



Source: Data Processing, 2025

According to the graph presented above, the distribution of ratings shows that the majority of data reflects high scores, with the peak being a score of 5 (486), followed by a score of 4 (224), while lower scores like 1 (48) and 2 (63) are much less frequent. When examining the results from using Naive Bayes Classification, the model shows increased accuracy in estimating classes with larger data volumes, particularly scores of 5 and 4, as the algorithm is significantly shaped by the class proportions in the training dataset. This can lead to scores with less representation, such as 1 and 2, being more susceptible to misclassification or less accurate identification by the model. Therefore, this uneven rating distribution impacts the effectiveness of the Naive Bayes model, leading to more reliable predictions for higher scores compared to lower ones..

5. Conclusion

Based on the results and discussion of this research, it's concluded that the sentiment analysis of guest reviews on TripAdvisor for Hotel X reveals a largely positive outlook, with 72.99% of guests expressing satisfaction with their accommodation experience. This positive feedback reflects appreciation for the hotel's facilities, swimming pool, and staff hospitality. However, the negative (19.08%) and neutral (7.93%) sentiments primarily indicate issues such as staff service quality, views, and service quality in the restaurant. This suggests that while overall guest sentiment is largely positive, management must continue to monitor, address, and improve areas that frequently receive complaints to enhance future guest satisfaction.Regarding model effectiveness, the Naive Bayes algorithm applied in the sentiment analysis achieved strong accuracy in classifying positive sentiments. However, it encountered difficulty in distinguishing between negative and neutral sentiments due to an imbalanced data distribution. Therefore, the continuous application of sentiment analysis remains crucial, serving as a basis for strategic decision-making related to service improvements in an effort to boost guest satisfaction amidst increasing competition in the hospitality sector.

Reference

1. Alnsour, Y. (2018). The effect of social media marketing on consumers' satisfaction. *International Journal of Business and Management Invention*, 7(3), 44–51.
2. Antara, M., & Sumarniasih, M. S. (2017). Role of tourism in economy of Bali and Indonesia. *Journal of Tourism and Hospitality Management*, 5(2). https://doi.org/10.15640/jthm.v5n2a4
3. Ariansyah, R. (2025). Pengaruh ulasan online terhadap keputusan pembelian wisatawan. *Jurnal Ekonomi Digital*, 4(1), 15–22.
4. Azzahra, S. A., & Wibowo, A. (2020). Analisis sentimen multi-aspek berbasis konversi ikon emosi dengan algoritme Naïve Bayes untuk ulasan wisata kuliner pada web TripAdvisor. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(4), 737. https://doi.org/10.25126/jtiik.2020731907
5. Chandradev, V., Suarjaya, I. M. A. D., & Bayupati, I. P. A. (2023). Analisis sentimen review hotel menggunakan metode deep learning BERT. *Jurnal Buana Informatika*, 14(2), 107–116.
6. Fahira, F., & Prianto, C. (2023). Prediksi pola kedatangan turis mancanegara dan menganalisis ulasan TripAdvisor dengan LSTM dan LDA. *Jurnal Tekno Insentif*, 17(2), 69–83. https://doi.org/10.36787/jti.v17i2.1096
7. Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 168–177). https://doi.org/10.1145/1014052.1014073
8. Kelly, R. (2020). Measuring customer satisfaction: Concepts and methods. *International Journal of Marketing Science*, 12(1), 11–21.
9. Khotimah, A. K. (2024). Analisis sentimen terhadap kualitas pelayanan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3044–3048. https://doi.org/10.36040/jati.v8i3.9520
10. LEVYTSKA, I. (2019). Revenue management in hotel business. *Economy. Finances. Management: Topical Issues of Science and Practical Activity*, 5(45), 108–117. https://doi.org/10.37128/2411-4413-2019-5-12
11. Mabrur, R., Indrasari, D., & Nuryanti, M. (2022). Determinants of customer satisfaction in the hospitality industry. *Journal of Business and Tourism Research*, 8(1), 34–46.
12. Maulidiah, L., Hasanah, U., & Fathurrochman, I. (2023). Customer satisfaction in digital hotel booking: A literature review. *Jurnal Manajemen dan Bisnis Digital*, 3(2), 25–33.

13. Sarudin, R., Ismail, A., & Dewi, D. (2021). Analisis online review TripAdvisor.com terhadap minat pembelian produk jasa akomodasi di Hotel Manhattan. *Jurnal Ilmu Sosial dan Ekonomi*, 7(4), 33–43. https://doi.org/10.30813/.v7i1.2634

14. Supriyanto, A., Wijayanti, R., & Nursidik, R. (2023). Penerapan analisis sentimen dalam pengelolaan ulasan pelanggan hotel berbasis teks. *Jurnal Teknologi dan Informasi*, 9(2), 54–61.

15. Wiastuti, R. D., & Kurnia, J. R. (2021). Atribut luxury hotel di Jakarta berdasarkan ulasan digital pada TripAdvisor. *Pringgitan*, 2(1). https://doi.org/10.47256/pringgitan.v2i01.161

16. Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51–65. https://doi.org/10.1016/j.tourman.2016.10.001