



Sentiment Analysis of Hotel Guests Using the Naive Bayes Classifier

Ahmad Zamsuri*,¹ Tengku Ardiansyah Saputra²

^{1,2}Informatics Engineering, Faculty of Computer Science, Universitas Lancang Kuning, Pekanbaru, Indonesia.

*Ahmad Zamsuri

Email: ahmadzamsuri@unilak.ac.id

Received 13/06/2024, Revised 20/06/2024, Accepted 25/06/2024, Published 30/06/2024



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Abstract

The reviews given by visitors are critical and should be meticulously analyzed to enhance the quality of hotel services. Traditionally, reviewing each guest's feedback manually is time-consuming and inefficient. Therefore, there is a need for an effective technique to aggregate and analyze large volumes of reviews. This research aims to provide a sentiment analysis of hotel visitor reviews using the Naïve Bayes algorithm. This study's findings are based on its comprehensive approach to automating sentiment analysis, which significantly reduces processing time and increases accuracy. For this purpose, we used Google Colaboratory to implement and evaluate the Naïve Bayes algorithm. The results reveal that the sentiment analysis model achieves varying levels of accuracy across different hotels: Hotel Arena with 0.74, K Hotel George with 0.94, Hotel Claridge Paris with 0.92, Suite Hotel 900 m zur Oper with 0.84, and Atlantis Hotel Vienna with 0.86. These findings underscore the potential of the Naïve Bayes algorithm for effectively capturing and analyzing customer sentiment, offering a valuable tool for hotel management to understand and improve guest satisfaction. This study is pioneering in its application of the Naïve Bayes algorithm in the context of sentiment analysis for multiple hotel brands, providing a scalable solution for the hospitality industry.

Keywords: Sentiment Analysis, Naïve Bayes Algorithm, Hotel Reviews

Introduction

Information technology, encompassing both software and hardware, serves as a crucial tool for supporting and enhancing the rapid and effective dissemination of information for individuals and society at large. The advent of advanced information technology has necessitated a closer examination of whether the management of information technology and business functions achieves the anticipated outcomes.

With the exponential growth of the Internet and social media, a vast array of information and public opinion have become readily available. A survey conducted by the Internet Service Entrepreneurs Association (APJII) in 2014 indicated a significant increase in Internet users in Indonesia over recent years (Darwis and Siskawati, 2021).

The hospitality industry, which is a cornerstone of the tourism sector, relies heavily on information technology to meet the demands of its clientele. Prospective visitors often seek to ascertain the quality of hotels without physically visiting them, making online reviews an invaluable resource. Websites that aggregate reviews about



global accommodations and hotel issues offer a wealth of information, but the sheer volume of reviews can pose challenges for users attempting to gauge hotel quality accurately. This research aims to provide actionable recommendations to enhance the facilities and services offered by hotels, thereby ensuring that they meet and exceed guest expectations.

Materials and Methods

The research method used in this study is one of the data mining tasks for data classification, which is mapping (classifying) data into one or several predefined classes, thus enabling the determination of hotel reviews.

Data collection was carried out by identifying the research problems faced. Next, data collection was conducted through observation and literature review.

Results and Discussion

Data Analysis

The data analysis process in this study used the Naïve Bayes method to find sentiment values and accuracy from classifying hotel reviews using Google Colaboratory. Naïve Bayes Algorithm Calculation on formula 1.

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

$P(H|X)$ = Posterior probability

$P(H)$ = Prior probability

$P(X|H)$ = Likelihood

$P(X)$ = Marginal probability

Using the formula above, review sentence values can be calculated. The manual calculation of the Naïve Bayes method is as follows:

Table 1. Data Review

Doc	Comments	Label
D1	staf yang luar biasa kamar yang bersih nyaman akan inap lagi	Positive
D2	staf senang tempat tidur besar sangat akomodatif tidak jauh dari pusat kota	Positive
D3	staf tipu kami dengan harga	Negative
D4	kamar dingin tidak privasi kamar mandi	Negative
D5	kamar pertama milik tangga yang curam tempat	?

tidur loteng tidak praktis
bahaya

Number of Negative Terms (D1, D2) = 18

Number of Positive Terms (D3, D4) = 11

Here is the manual calculation method of Naïve Bayes:

a. Probability of Occurrence of Positive and Negative categories (Class Probability):

$$P(V_j) = \frac{|docs\ j|}{|D|}$$

$$P(Positive) = \frac{\text{Number of positive training data}}{\text{Total training data}} = \frac{2}{4} = 0,5$$

$$P(Negative) = \frac{\text{Number of negative training data}}{\text{Total training data}} = \frac{2}{4} = 0,5$$

b. Calculation of Word Probability for each class:

$$P(X_i | Y_j) = \frac{nk+1}{n+|kata|}$$

$$P(\text{"kamar"}|\text{positif}) = \frac{1+1}{18} = 0,11$$

$$P(\text{"kamar"}|\text{negatif}) = \frac{2+1}{11} = 0,27$$

$$P(\text{"tempat"}|\text{positif}) = \frac{1+1}{18} = 0,11$$

$$P(\text{"tempat"}|\text{negatif}) = \frac{0+1}{11} = 0,09$$

$$P(\text{"tidur"}|\text{positif}) = \frac{1+1}{18} = 0,11$$

$$P(\text{"tidur"}|\text{negatif}) = \frac{0+1}{11} = 0,09$$

$$P(\text{"tidak"}|\text{positif}) = \frac{1+1}{18} = 0,11$$

$$P(\text{"tidak"}|\text{negatif}) = \frac{1+1}{11} = 0,18$$

c. Category Classification Calculation:

$$= P(\text{Positif}) + P(\text{"kamar"}|\text{positif}) + P(\text{"tempat"}|\text{positif}) + P(\text{"tidur"}|\text{positif}) + P(\text{"tidak"}|\text{positif}) = 0,94$$

$$= P(\text{Negatif}) + P(\text{"kamar"}|\text{negatif}) + P(\text{"tempat"}|\text{negatif}) + P(\text{"tidur"}|\text{negatif}) + P(\text{"tidak"}|\text{negatif}) = 1,13$$

Based on calculations, Hotel Arena received a positive classification accuracy of 0.94, while the negative classification accuracy was 1.13. Each word has a positive and negative probability value. It can be seen that the product of the positive classification category is smaller than that of the negative classification category, so the comment is categorized as negative. The classification probability value for words in the positive and negative labels is $0.94 < 1.13$, thus it can be concluded that the comment on D5 is classified as negative.

Testing Naïve Bayes Accuracy Using Google Colaboratory

Preprocessing Testing Results

	nama hotel	alamat	sentimen	review
0	arena hotel	irlandia	positif	tidak keluh nyata hotel bagus lokasi yang bagu...
1	arena hotel	belgium	negatif	meski gambar tunjuk kamar yang sangat bersih k...
2	arena hotel	italia	positif	hotel sedang renovasi dengan sangat hati denga...
3	arena hotel	italia	positif	hotel luar biasa saya ambil dengan tulus karen...
4	arena hotel	belanda	positif	area umum indah kamar bagus tapi jendela rusak...
...	...	...	...	...
109	arena hotel	israel	negatif	jalan jauh dari pusat kota
110	arena hotel	britannia raya	positif	ruang indah staf sangat bantu
111	arena hotel	britannia raya	positif	ruang cerah lapang staf sangat ramah
112	arena hotel	britannia raya	positif	staf sangat ramah sangat bantu
113	arena hotel	britannia raya	positif	hotel takjub bangun yang indah kamar sangat mo...

114 rows × 4 columns

Fig 1. Proprocessing Result

Performance Settings

The results of the data calculation are divided into 80% training data and 20% testing data. The calculation results can be seen in Figures 2 and 3 below.

	nama hotel	alamat	review	sentimen
0	arena hotel	australia	lokasi bagus belah trem tetapi sedang dalam ba...	positif
1	arena hotel	britannia raya	sarap	positif
2	arena hotel	britannia raya	ruang cerah lapang staf sangat ramah	positif
3	arena hotel	britannia raya	staf sangat sopan sedia bantu tiap saat hotel ...	positif
4	arena hotel	irlandia	staf tahan adalah terima yang bagus agak loyo ...	positif
...	...	...	...	...
86	arena hotel	austria	transparan buruk	negatif
87	arena hotel	britannia raya	kena biaya untuk paket ulang tahun ketika ikla...	negatif
88	arena hotel	italia	hotel bintang biasa tidak yang istimewa tidak ...	positif
89	arena hotel	britannia raya	kamar yang luar biasa hotel yang tarik lokasi ...	positif
90	arena hotel	turki	bangun dengan suasana luar biasa dengan desain...	positif

91 rows × 4 columns

Fig 2. Training Data Result

	nama hotel	alamat	review	sentimen
0	arena hotel	britannia raya	staf yang luar biasa kamar yang bersih nyaman ...	positif
1	arena hotel	belanda	area umum indah kamar bagus tapi jendela rusak...	positif
2	arena hotel	argentina	kami suka kami inap arena hotel hotel sangat b...	positif
3	arena hotel	britannia raya	staf senang tempat tidur besar sangat akomodat...	positif
4	arena hotel	britannia raya	seluruh hotel sangat bersih staf sangat ramah ...	positif
5	arena hotel	britannia raya	hotel super gaya belah taman yang indah cukup ...	positif
6	arena hotel	britannia raya	staf tipu kami dengan harga	negatif
7	arena hotel	britannia raya	kami tinggal kamar studio kamar luas bersih te...	positif
8	arena hotel	rusia	kamar standar adalah lokasi yang luar biasa ge...	positif
9	arena hotel	britannia raya	hotel sangat bagus staf sangat ramah bantu say...	positif
10	arena hotel	britannia raya	akan suka tempat tidur ganda	positif
11	arena hotel	spanyol	saya suka desain dekorasi hotel staf sangat ba...	positif
12	arena hotel	britannia raya	staf fantastis yang tidak bisa lebih bantu ram...	positif
13	arena hotel	britannia raya	staf sangat ramah sangat bantu	positif
14	arena hotel	amerika serikat	kamar pertama milik tangga yang curam tempat t...	negatif
15	arena hotel	irlandia	tidak keluh nyata hotel bagus lokasi yang bagu...	positif
16	arena hotel	britannia raya	lokasi nuansa hotel	positif
17	arena hotel	britannia raya	hotel luar biasa interior yang indah kami tida...	positif
18	arena hotel	irlandia	kamar dingin tidak privasi kamar mandi	negatif
19	arena hotel	britannia raya	staf sangat bantu hotel lokasi yang bagus	positif
20	arena hotel	taiwan	sangat dekat dengan area museum gedung konser	positif
21	arena hotel	britannia raya	kerja bangun kamar	negatif
22	arena hotel	britannia raya	kami telah tinggal sini beberapa kali selalu n...	positif

Fig 3. Testing Data Results

Naïve Bayes Accuracy Results

Arena hotel accuracy results

The arena hotel test results with a split percentage of 80% training data and 20% testing data show an accuracy level of 74% or 0.74 as shown in Figure 4 below.

	precision	recall	f1-score	support
negatif	0.50	0.33	0.40	3
positif	0.76	1.00	0.86	16
negatif	0.00	0.00	0.00	1
positif	0.00	0.00	0.00	3
accuracy			0.74	23
macro avg	0.32	0.33	0.32	23
weighted avg	0.60	0.74	0.65	23

Fig 4. Arena hotel accuracy results

KK Hotel George Accuracy Results

Testing the George hotel KK hotel with a split percentage of 80% training data and 20% testing data results in an accuracy level of 94% or 0.94 as shown in Figure 5 below.

	precision	recall	f1-score	support
negatif	1.00	0.75	0.86	8
positif	0.93	1.00	0.97	28
accuracy			0.94	36
macro avg	0.97	0.88	0.91	36
weighted avg	0.95	0.94	0.94	36

Fig 5. KK Hotel George Accuracy Results

Claridge Paris Hotel Accuracy Results

The results of testing the Claridge Paris hotel below in Figure 6 with a split percentage of 80% training data and 20% testing data results in an accuracy rate of 92% or 0.92.

	precision	recall	f1-score	support
negatif	0.00	0.00	0.00	0
positif	1.00	0.92	0.96	12
accuracy			0.92	12
macro avg	0.50	0.46	0.48	12
weighted avg	1.00	0.92	0.96	12

Fig 6. Claridge Paris Hotel Accuracy Results

Accuracy Results Suite Hotel 900 M Zur Oper

Hotel Suite Hotel 900 M Zur Oper testing with a split percentage of 80% training data and 20% testing data results in an accuracy level of 84% or 0.84 as shown in Figure 7 below.

	precision	recall	f1-score	support
negatif	0.67	0.33	0.44	6
positif	0.86	0.96	0.91	25
accuracy			0.84	31
macro avg	0.76	0.65	0.68	31
weighted avg	0.82	0.84	0.82	31

Fig 7. Accuracy Results Suite Hotel 900 M Zur Oper

Atlantis Hotel Vienna Accuracy Results

Testing the Atlantis Hotel Vienna hotel below in Figure 8 with a split percentage of 80% training data and 20% testing data results in an accuracy rate of 86% or 0.86.

	precision	recall	f1-score	support
negatif	1.00	0.73	0.84	11
positif	0.77	1.00	0.87	10
accuracy			0.86	21
macro avg	0.88	0.86	0.86	21
weighted avg	0.89	0.86	0.86	21

Fig 8. Atlantis Hotel Vienna Accuracy Results

Diagram Evaluation Results

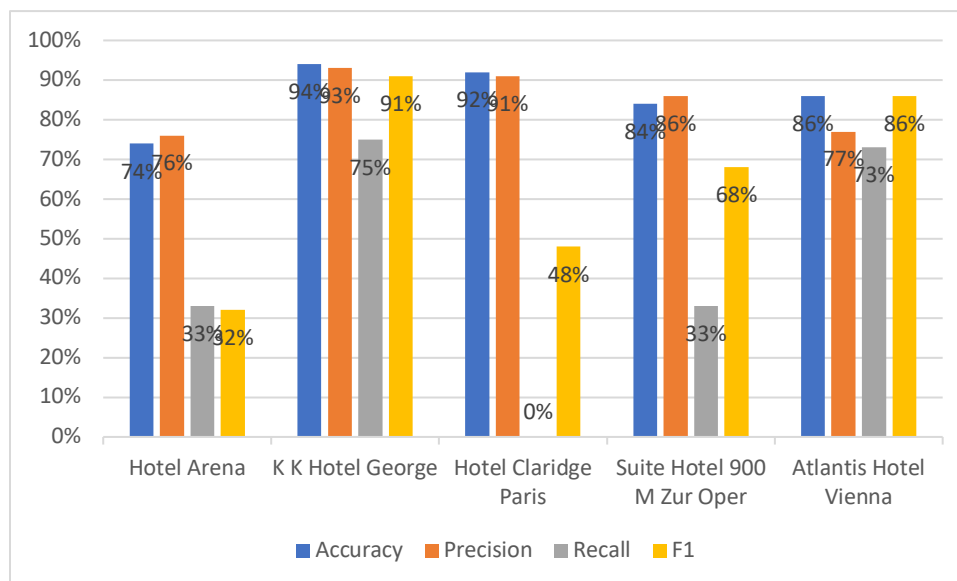


Fig. Diagram Evaluation Results

Based on Figure 9, the diagram of the Accuracy of the naïve bayes classification can be seen in the diagram above with the results of the percent form with the Accuracy value in blue, Preciosion in orange, Recall in gray, and F1 in yellow.

Conclusion

Hotel review analysis resulted in the classification values of the naïve bayes algorithm with accuracy in hotels as follows: Hotel Arena has an accuracy of 0.74, K K Hotel George 0.94, Hotel Claridge Paris 0.92, Suite Hotel 900 m zur Oper 0.84, Atlantis Hotel Vienna 0.86. For future research, the author suggests conducting experiments using other algorithms to achieve better or lesser results and also applying them to other tools. For future research, the author suggests testing at tourist attractions, restaurants, and other places with sentiment classification not limited to positive or negative sentiments.



References

- Amrullah, A. Z., Sofyan Anas, A., Adrian, M., & Hidayat, J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal*, 2(1). <https://doi.org/10.30812/bite.v2i1.804>
- Azza Amira, S., Utama, S., & Muhammad Hanif Fahmi, dan. (2020). Edu Komputika Journal Penerapan Metode Support Vector Machine untuk Analisis Sentimen pada Review Pelanggan Hotel. *Edu Komputika*, 7(2). <http://journal.unnes.ac.id/sju/index.php/edukom>
- Bijaksana, A., Negara, P., Muhandi, H., & Putri, I. M. (2020). Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes Dan Seleksi Fitur Information Gain Sentiment Analysis On Airlines Using Naive Bayes Method And Feature Selection Information Gain. 7(3), 599–606. <https://doi.org/10.25126/Itiik.202071947>
- Cahyani, R., & Pandu Adikara, P. (2019). Analisis Sentimen terhadap Ulasan Hotel menggunakan Boosting Weighted Extreme Learning Machine (Vol. 3, Issue 8). <http://j-ptiik.ub.ac.id>
- Darwis, D., Siskawati, N., & Abidin, Z. (n.d.). Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional. 15(1).
- Dini Utami, L., Rachmi, H., Nurlaela, D., & Informasi Kampus Kota Bogor, S. (2018). Komparasi Algoritma Klasifikasi Pada Analisis Review Hotel. 14(2), 261. www.bsi.ac.id
- Khofifah, W., Rahayu, D. N., & Yusuf, A. M. (2022). Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps. *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, 16(4). <https://doi.org/10.35969/interkom.v16i4.192>
- Kusuma, A., & Nugroho, A. (2021). Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naive Bayes. *Jurnal Ilmiah Teknologi Informasi Asia*, 15(2). <https://doi.org/10.32815/jitika.v15i2.557>
- Nasution, N., Djahara, K., & Zamsuri, A. (2015). Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes (Studi Kasus: Fasilkom Unilak). *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 6(2), 1-11.
- Program, A. T., Sistem, S., Stmik, I., Mandiri, N., Jl, J., Damai, N., Barat, W. J., & Selatan, J. (2017). Optimasi Particle Swarm Optimization Sebagai Seleksi Fitur Pada Analisis Sentimen Review Hotel Berbahasa Indonesia Menggunakan Algoritma Naive Bayes. *Jurnal Teknik Komputer*, 11(2). <http://www.tripadvisor.com>
- Taufik STMIK Nusa Mandiri Jakarta Jl Damai No, A., Jati Barat Jakarta Selatan, W., & Sitasi, C. (2018). Komparasi Algoritma Text Mining Untuk Klasifikasi Review Hotel. *Jurnal Teknik Komputer*, 2, 112–118. <https://doi.org/10.31294/jtk.v4i2.3461>
- Thomas, V. W. D., & Rumaisa, F. (2022). Analisis Sentimen Ulasan Hotel Bahasa Indonesia Menggunakan Support Vector Machine dan Tf-Idf. *Jurnal Media Informatika Budidarma*, 6(3), 1767. <https://doi.org/10.30865/Mib.V6i3.4218>
- Utami, D. S., Erfina, A., & Id, M. A. (2022). Analisis Sentimen Ulasan Terkait UNESCO Global Geopark Di Google Maps dengan Algoritma Naive Bayes. In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 6, Issue 2).