

Sentiment Analysis of Joglo Wifi UMKM Service with Naive Bayes Method

Jofan Fathurahman ^{a,1}, Dwi Hartanti ^{a,2}, Sopingi ^{b,3}

^a Informatics Engineering, Faculty of Computer Science, Universitas Duta Bangsa Surakarta

^b Information Systems, Faculty of Computer Science, Universitas Duta Bangsa Surakarta

¹ 202030247@mhs.udb.ac.id *, ² dwihartanti@udb.ac.id, ³ sopingi@udb.ac.id

*corresponding author

ARTICLE INFO

Article history:
Published

Keywords:

Customer Sentiment
Naive Bayes,
Sentiment Analysis,
Service Quality,
Customer Feedback

ABSTRACT

In today's digital era, many businesses face challenges in managing and responding to customer feedback and suggestions contained in comments. Joglo Wifi Corner UMKM Business Center & Coffee experiences similar difficulties, where the lack of customer feedback management negatively impacts service quality and customer satisfaction. In this study, we aim to develop and implement a sentiment analysis system that is able to classify customer sentiment into positive, negative, or neutral. The process begins with data collection. customer comments from the official Joglo Wifi Corner website, which are then processed through a preprocessing stage including tokenization, stop word removal, and lemmatization. Then, a Logistic Regression Naive Bayes model is drilled using the processed data to classify sentiment. The evaluation results show that this system achieves an accuracy of 91.67%, with an average precision of 0.646, an average recall of 0.805, and an average F1-Score of 0.701. The implementation of this system provides valuable insights for Joglo Wifi Corner managers in making strategic decisions to improve service quality. By responding more effectively to customer feedback, Joglo Wifi Corner promises that this research can improve customer satisfaction and expand the company's market coverage and identify areas for improvement based on customer feedback, thereby strengthening customer relationships. This approach also uses Agile Development system development methodology to ensure iterative and responsive development to changing user needs.

Copyright © 2024 by the Authors.

I. Introduction

Advances in information technology continue to increase every year, resulting in new innovations in the form of the internet. The internet system is a computer network that can be freely accessed by anyone and is widely connected using the TCP/IP packet-switching communication protocol. To connect to each other, computers require connecting media such as cables or optical fibres, satellites, or telephone lines. The impact of the internet in this digital era is very significant, especially in terms of high-speed internet access [1].

Micro, Small, and Medium Enterprises (MSMEs) are productive entities owned by individuals or legal entities, and fulfil the criteria to be classified as micro, small, and medium enterprises. MSMEs play a vital role in the Indonesian economy, accounting for almost 99.99 per cent of all businesses in Indonesia (BPS, 2016). MSMEs contribute 56.92 per cent to Indonesia's total Gross Domestic Product (GDP), equivalent to IDR 1,213.25 trillion. In addition, MSMEs also play an important role in absorbing labour (97.3 per cent of the total employed labour force) and have a significant share in the total number of business units in Indonesia. MSMEs also contribute significantly to investment in Indonesia, amounting to IDR 222.74 trillion or 51.80% of total investment in 2008 [2].



Joglo Wifi ID Corner aims to provide fast and quality Wi-Fi services to support work, study, and play activities in the Joglo area. However, despite receiving a lot of feedback from users, managing and analyzing this feedback is an obstacle that hinders the ability of Joglo Wifi ID Corner. This obstacle is caused by the volume and diversity of feedback, as well as the complexity and nuances of interactions that affect the way in which customers evaluate and respond to the services they receive [3].

In this research, the authors aim to develop and implement a sentiment analysis system that is able to classify customer sentiment into positive, negative, or neutral. on customer feedback obtained from the Joglo Wifi ID Corner web page. This system is designed to assist the owner in better understanding the comments and opinions given by service users. With this sentiment analysis system application, it is hoped that the owner can make more informed and data-based decisions regarding service improvements or adjustments based on sentiment analysis, so as to increase overall customer satisfaction in the research flow described.

In the next stage, the author uses the Python programming language to analyze and manage customer comment data. Python was chosen for its ability to process large data and its flexibility in developing sentiment analysis models. Using various Python libraries such as NLTK and scikit-learn, the authors were able to preprocess data, train classification models, and evaluate results effectively. The use of Python in this study enabled the implementation of sophisticated analysis techniques and provided an efficient solution to the customer feedback management challenges faced by Joglo Wifi Corner.

II. Method

This research applies the Agile Development methodology to facilitate iterative and collaborative development, which allows the team to adapt to changes and feedback during the development process as illustrated in Fig 1.

The purpose of this research is to develop a sentiment analysis system using the Naive Bayes algorithm. At this time, the Naïve Bayes classifier has been applied as one of the classification methods in sentiment analysis [4]. for customer feedback on the Joglo MSME site. The process starts with uploading two CSV files: one containing the word and sentiment dataset, and the other containing customer comments. These two files are processed by reading their contents into a Data Frame and creating a dictionary to store the words and sentiments. Next, a manual sentiment calculation is performed for each comment based on the dictionary that has been created. The comments are then processed with text preprocessing methods which include tokenization, stopwords removal. This process involves punctuation removal, tokenization, and stemming with the aim of improving the accuracy and consistency of the analysis results [5]and lemmatization.



Fig 1. Agile Development

The development stages used by the author for sentiment analysis with Agile Development include several steps.

a) Project Planning (Planning)

This stage aims to improve service quality and customer satisfaction through understanding feedback. customers through sentiment analysis. The method used is Naive Bayes for sentiment analysis, with data in the form of customer comments from the Joglo Wifi Corner website.

b) Data Collection

Data is collected from the official Joglo Wifi Corner website, with the type of data in the form of customer comments.

c) Data Preparation

At this stage, the data is cleaned by removing irrelevant or duplicate entries and handling missing values. The preprocessing process includes tokenization, stop words removal, and lemmatization. The whole text is converted to lowercase (case folding) It is said that common words that do not provide important information should be removed (stop words removal).

d) Data Visualisation

The data is visualized using word clouds to display a visual representation of the dominant words in the text.

e) Model Training

The algorithm used is Naive Bayes. The steps in model training include creating a word and sentiment dictionary, calculating sentiment for each comment, using TF-IDF Vectorizer to convert text into numerical features, and training the model using the processed data.

f) Model Evaluation

Models are evaluated using the metrics of accuracy, recall, precision, and F1-score and accuracy are both metrics used to measure model performance. Accuracy measures the extent to which the model's predictions match the actual data. Recall measures the ability of the model to find all relevant examples in the dataset. Precision measures how accurate the model is in finding relevant examples compared to the total positive predictions. F1-score is a combination of precision and recall giving an overall picture of the model's performance.

g) Implementation and Monitoring

The results of the analyses are used for strategic decision-making to improve service quality. The system is monitored continuously and adjustments are made where necessary.

III. Results and Discussion

At this stage, we will discuss the results of each stage of development and how they are interpreted.

A. Project Planning

At the planning stage of the project, the main objective set was to improve service quality and customer satisfaction through understanding their feedback through sentiment analysis. The Naive Bayes method was chosen as the sentiment analysis technique used, and the data to be analyzed were customer comments collected from the Joglo Wifi Corner website.

B. Data Collection

The data is collected from customer comments available on the official website of Joglo Wifi Corner in the form of CSV files. These comments include various feedback from customers regarding the services provided. The data obtained is text data which requires further preparation process before it can be analyzed.

Table 1. Data Set

No	Positive	Negative
1	Professional	blank
2	Friendly	mouse
3	Good	Dirty

Table 2. Comment

No	Comments
1.	experienced professional employees.
2.	The service staff is very helpful and friendly.
3.	B is not special

C. Data Preparation

In the data preparation stage, data cleaning is performed to eliminate invalid or irrelevant data by removing irrelevant or duplicate entries and handling missing values. Next, preprocessing is done which includes tokenization, stop words removal, and lemmatization to tidy up the text to make it more structured and processable. In addition, case folding is used to convert the entire text into lowercase letters and stop words removal is done to remove common words that do not provide important information. In text mining research, case folding preprocessing is necessary because the process is case sensitive and not all text uses capital letters consistently. Case folding is the process of transforming the entire document content into lowercase letters [6].

D. Data Visualisation

Data visualization is done using Word Cloud to display a visual representation of the dominant words in the text. Word cloud helps in identifying the most frequently occurring words in customer comments, providing an initial idea of the sentiment contained. The Word Cloud output is a visual representation of the dominant words in the text. The larger the size of the word that appears in the word cloud, the more frequently the word is used in the analysis [7].

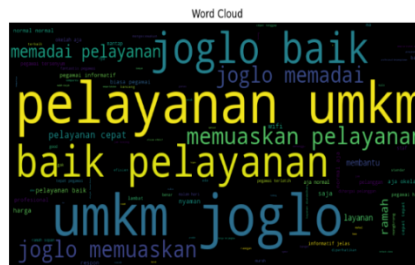


Fig 2. Word Cloud

E. Model Training

In the model training stage, the Naive Bayes algorithm is used to analyze sentiment. The steps include the creation of word and sentiment dictionaries, the calculation of sentiment for each comment, and the use of a TF-IDF Vectorizer to convert text into numerical features. The model is then trained using the processed data.

Table 3. DF Values of Training Data

Term	TF		DF
	D1	D2	
experienced	0	0	0
professional	1	0	1
employees	0	0	0
service	0	0	0
staff	0	0	0
very	0	0	0
helpful	0	0	0
friendly	1	0	1
not	0	0	0
Special	0	0	0

$$IDF = \log \frac{d+1}{df+1} + 1 \quad (1)$$

Table 4. IDF Values from Training Data

Term	DF	IDF
professional	1	2.098
friendly	1	2.098

$$W = TF \times IDF \quad (2)$$

Table 5. TF-IDF value of training data

Term	TF	
	D1	D2
experienced	0	0
professional	2.098	0
employees	0	0
service	0	0
staff	0	0
very	0	0
helpful	0	0
friendly	2.098	0
not	0	0
special	0	0

$$TFnorm(t, d) = \frac{TF(t,D)}{\sqrt{\sum_i (TF(t,d))^2}} \quad (3)$$

Table 6. Data Normalization

Term	DF	IDF
Professional	1	2.098
friendly	1	2.098

F. Model Evaluation

The trained model is evaluated using accuracy, recall, precision, and F1-score metrics. Accuracy measures the extent to which the model's predictions match the actual data, recall measures the model's ability to find all relevant examples in the dataset, precision measures how accurate the model is in finding relevant examples compared to the total positive predictions, and F1-Score provides an overall picture of the model's performance. Stage of data analysis, the activity is to analyze student ability test data using t-test, while the interview results are analyzed using descriptive analysis [8]. The evaluation results show that the model performs well in classifying sentiment from customer comments. What is illustrated in Fig 3 is the result of Validation Data and Test Data. and at this stage the author uses a systematic 70% Test Data, 15% Validation Data, 15% Test Data.

```

=== Classification Report (Validation) ===
      precision    recall  f1-score   support

negatif      0.25      1.00      0.40         3
netral       1.00      0.89      0.94       125
positif      0.75      1.00      0.86        15

 accuracy      0.90       143
macro avg     0.67      0.96      0.73       143
weighted avg  0.96      0.90      0.92       143

=====

=== Classification Report (Test) ===
      precision    recall  f1-score   support

negatif      0.38      0.75      0.50         4
netral       0.99      0.94      0.96       129
positif      0.57      0.73      0.64        11

 accuracy      0.92       144
macro avg     0.65      0.81      0.70       144
weighted avg  0.94      0.92      0.93       144

=====
    
```

Fig 3. Model Evaluation Results

The Naïve Bayes classification algorithm has an equation that can be used as a guideline for calculating values [9] This equation is used to calculate the probability in the decision-making process, using the following formula.

G. Implementation and Monitoring

The results of sentiment analysis are used for strategic decision-making to improve service quality. The implemented system monitors performance continuously and makes adjustments if necessary to ensure the service remains optimized. Continuous monitoring is also conducted to evaluate the effectiveness of the system and make improvements if necessary. The results of sentiment evaluation using Gaussian Naive Bayes show that the model has different performance for each sentiment label. For positive labels, precision is 0.57, recall is 0.73, and F1-score is 0.64. For negative labels, precision is 0.38, recall is 0.75, and F1-score is 0.5. For neutral labels, precision was 0.99, recall was 0.94, and F1-score was 0.96. The average precision was 0.65, the average recall was 0.81, and the average F1-score was 0.70. The overall accuracy of the model is 0.92, which shows that the model is able to classify customer comments quite well.

Label	Precision	Recall	F1-Score
Positif	0.5714285714285714	0.7272727272727273	0.64
Negatif	0.375	0.75	0.5
Netral	0.9918032786885246	0.937984496124031	0.9641434262948207
Average	0.6460772833723653	0.8050857411322528	0.7013811420982736
Accuracy	0.9166666666666666	0.9166666666666666	0.9166666666666666

Fig 4. Result Logistic Regression

At this stage, the sentiment analysis results produce the comment value listed in Fig 5. This value is the conclusion of all data obtained from a total of 955 comment data. From the analysis results, it was found that there were 95 comments with positive sentiments, which accounted for 9.95% of the

total data. Comments with negative sentiments totaled 22, which accounted for 2.30% of the total data. Most of the comments, 838 comments or 87.75% of the total, had neutral sentiments. These results provide an overview of customer perceptions of the services provided by Joglo Wifi Corner.

```

=== Pembagian Data ===
Total Data: 955
Data Latih (Train): 668 (69.95%)
Data Validasi: 143 (14.97%)
Data Uji (Test): 144 (15.08%)
=====

```

Fig 5. Value Result

At this stage, an evaluation is conducted to compare the algorithm's prediction results with the actual data classification. The confusion matrix method is used as the main evaluation tool in this case [10]. The result of the metric is explained in Figure 6

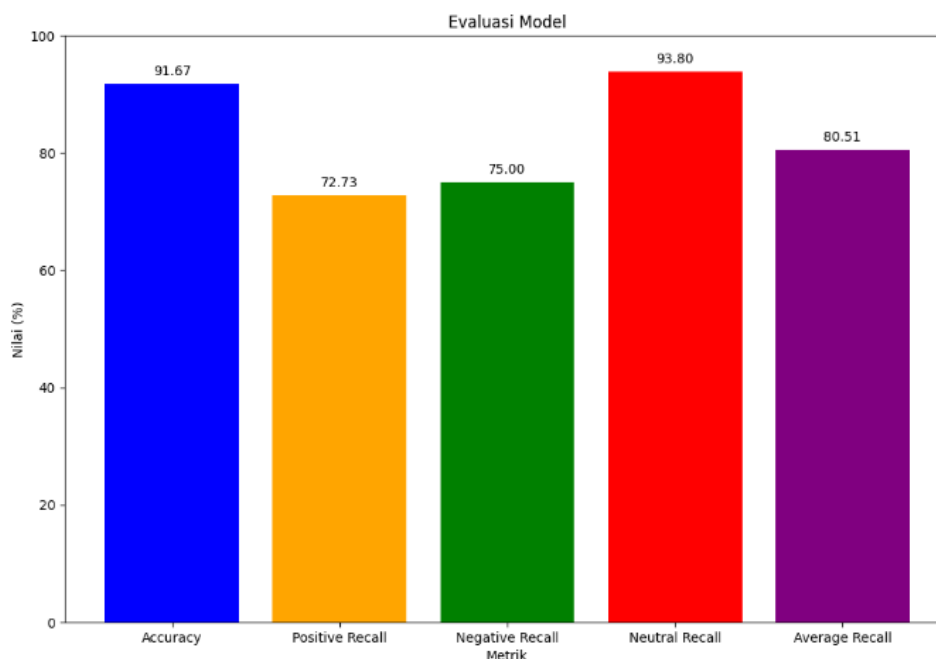


Fig 6. Metrick evaluation results

which displays the output results, including the model evaluation metrics: an accuracy of 91.67%, a positive recall of 72.73%, a neutral recall of 93.80%, a negative recall of 75%, and an average recall of 80.51%. A discussion on the 'Sentiment Analysis of Joglo Wifi MSMEs Service with the Logistic Regression Method' using the obtained data is necessary to provide further insights.

IV. Conclusion

The Logistic Regression algorithm has demonstrated excellent performance in classifying the sentiment of Joglo Wifi Corner customer comments, achieving 91.67% accuracy. It effectively identifies neutral comments with a precision and recall of 0.99 and 0.94, respectively. Positive sentiment detection shows a precision of 0.57 and recall of 0.73, while negative sentiment detection has a precision of 0.38 and recall of 0.75. The model's average precision, recall, and F1-score are 0.65, 0.81, and 0.70, respectively, indicating overall good performance. This sentiment analysis system was developed using the Agile Development methodology, which facilitated flexible, collaborative development and iterative improvements through sprint planning. The insights gained from this system enable Joglo Wifi Corner managers to better understand and respond to customer feedback, ultimately enhancing service quality and customer satisfaction.

References

- [1] D. Aditya Rachman, 'Analisis Kualitas Layanan Jaringan Internet Fiber To The Home Pt. Xyz Menggunakan Wireshark', *Jurnal Ilmiah Teknik dan Ilmu Komputer*, pp. 1–9, 2023.
- [2] M. Rapii, A. Riswanto, and L. Febriani, 'Analisis Faktor-Faktor Yang Mempengaruhi Pendapatan Umkm Pada Masa Pandemi Covid-19', *Jambura Economic Education Journal*, vol. 5, no. 1, 2023.
- [3] D. Merdika Mansur and A. Halim Perdana Kusuma Putra, 'YUME: Journal of Management Pengaruh Kualitas Pelayanan Terhadap Kepuasan Pelanggan Pada Bengkel Mobil Honda Dealer-Pluit', 2023.
- [4] F. A. Wenando, 'Analisis Opini Publik Terhadap Undang-Undang KUHP Tahun 2022 Menggunakan Algoritma Naïve Bayes Classifier', *JURNAL FASIKOM*, vol. Volume 13 No. 2, no. Agustus 2023: 334-339, pp. 1–6, 2023.
- [5] R. Aziz et al., 'Rizky Aziz 1) , Tresna Maulana Fahrudin 2)', *JURNAL FASILKOM*, vol. 14 No. 1 |, no. April 2024:, pp. 1–10, 2024.
- [6] S. Supriyatna and E. Fahrudin, 'Pemanfaatan Algoritma Text Mining Dalam Menemukan Pola Risiko Bencana Sebagai Pengetahuan Kebencanaan Dari Dokumen Kajian Risiko Bencana (Krb) 1*', *Jurnal Informatika Utama*, vol. 2, no. 1, 2024, doi: 10.55903/jitu.v2i1.xx.
- [7] A. Sri Widagdo et al., 'Analisis sentimen terhadap pelayanan Kesehatan berdasarkan ulasan Google Maps menggunakan BERT', *FASILKOM*, vol. Volume 13 No. 2, no. Agustus 2023: 326-333, pp. 1–7, 2023.
- [8] A. Ihsan and S. Achriadi, 'Analysis of Students Competency in the Industrial Engineering Study Program South Aceh Polytechnic in Mastering the Concept Survey and Mapping Courses', *Jurnal Inotera*, vol. 6, no. 2, pp. 89–98, Oct. 2021, doi: 10.31572/inotera.vol6.iss2.2021.id148.
- [9] K. Anwar, 'KLIK: Kajian Ilmiah Informatika dan Komputer Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes', vol. 2, no. 4, pp. 148–155, 2022, [Online]. Available: <https://djournal.com/klik>
- [10] Wijiyanto and Sopingi, 'Kontribusi Keluarga Dalam Prediksi Mahasiswa Lulus Tepat Waktu Menggunakan Model Support Vector Machine', 2024.
- [11] Salya Rater, Balkhaya, "The Influence of Social Media Digital Marketing Analysis on New Student Acceptance in a University" *J. Inotera*, Vol 8, No. 1, PP. 150-154, 2023