Sistemasi: Jurnal Sistem Informasi ISSN:2302-8149
Volume 13, Nomor 3, 2024: 1044-1053 e-ISSN:2540-9719

Sentiment Analysis of pegipegi.com Review on Google Play Store with Naïve Bayes

[1]Muhamad Naufal Burhanuddin Balit*, [2]Fandy Setyo Utomo
[1,2]Informatics Study Program, Faculty of Computer Science, Amikom University Purwokerto
Jl. Letjend Pol. Soemarto No. 126, Banyumas 53127, Central Java, Indonesia
*e-mail:20sa1166@mhs.amikompurwokerto.ac.id

(received: 24 Januari 2024, revised: 20 Maret 2024, accepted: 20 Maret 2024)

Abstract

In the current era, a shift in consumer behavior is evident in the use of online platforms for booking tickets, involving various services such as flights, hotels, trains, buses, and entertainment. PegiPegi.com, as a rapidly growing online travel agent in Indonesia, demonstrates success by understanding the value of technology and maintaining strong partnerships. This phenomenon also impacts sentiment analysis, where users of this platform often provide reviews. This research aims to apply the Naïve Bayes classification method in sentiment analysis of PegiPegi.com reviews, focusing on understanding customer satisfaction and service improvement. By combining these approaches, the study contributes to a deeper understanding of user responses to OTA services and presents the evaluation results of the Multinomial Naive Bayes classification model with an accuracy rate of 89.5%. The high precision in the Negative class indicates the model's ability to identify negative reviews. However, there are challenges in classifying the Neutral class, suggesting potential for further improvement. Nevertheless, the F1-score of 0.522 reflects a good balance between overall precision and recall.

***Keywords**: Sentiment Analysis, Naive Bayes Algorithm, PegiPegi.com*

1 Introduction

In this era, we are witnessing a major shift in people's behavior towards the convenience of ordering tickets online through various booking platforms. This phenomenon is not limited to flight tickets, but also involves reservations for other services such as hotels, trains, buses and entertainment.[1]. This shift in consumer behavior shows an increasing preference for convenience and efficiency in the ordering process, which is replacing more traditional conventional methods. The main attraction of using online platforms is better accessibility and ease of navigation, allowing users to plan travel and entertainment activities more practically. One clear example of an online platform that meets these needs is PegiPegi.com. The company known as PT. Go Online Destination, or more familiarly known as Pegipegi, is an online travel agent (OTA) that is experiencing rapid development in Indonesia[2]. PegiPegi.com is an application and website that provides booking services for various types of travel and entertainment. Through this platform, users can easily search, compare and book tickets for hotels, flights, trains, buses and various entertainment activities.

Pegipegi takes effective and innovative strategic steps to ensure customer comfort and satisfaction on their journey. They understand the value of the latest technology in the online travel industry and continually update their platform to be responsive and easy to access from a variety of devices, making them the top choice for customers. Strong partnerships with various parties such as hotels, airlines and railway companies provide customers with a variety of travel options. An attractive rewards program, 24/7 customer service, and quality content through inspirational articles and hotel reviews add value and strengthen Pegipegi's reputation in the industry. With a commitment to innovation, customer comfort and service quality, Pegipegi has succeeded in creating a strong position as one of the leaders in the online travel industry[3].

Apart from providing comfort and convenience in the ordering process, the phenomenon of increasing use of online ordering sites such as PegiPegi.com also has an impact on sentiment analysis. Sentiment analysis is an automatic process that involves extracting, processing and understanding data

 ISSN:2302-8149
e-ISSN:2540-9719

in the form of unstructured text. The aim of this process is to automatically identify and retrieve information related to the sentiment contained in an opinion or opinion sentence[4]. Users of this platform often provide reviews and testimonials regarding their experiences while using the service. Sentiment analysis of these reviews can provide valuable insight into customer satisfaction, deficiencies that need to be addressed, and service advantages that users appreciate. Through comments and reviews, PegiPegi.com can identify positive or negative sentiment trends which can be important feedback to improve the quality of its services. Sentiment analysis can also help in developing marketing strategies and making smarter business decisions based on customer perceptions and experiences. Thus, public involvement in ordering tickets online not only creates changes in consumer behavior, but also opens up opportunities to improve and optimize services based on real-time customer feedback.

In general, research can be explained as a process of collecting and analyzing data that is carried out systematically and logically in order to achieve specific goals[5]. In this context, research does not only focus on collecting data alone, but also aims to achieve specific goals such as developing knowledge, improving certain practices, or providing concrete solutions to identified problems. One approach that can be used to provide a solution is to apply classification methods, such as Naive Bayes, in analyzing sentiment from structured text data. This research seeks to overcome the problems faced by investigating and digging up specific information, understanding certain phenomena, and testing the proposed hypotheses. The role of Naive Bayes in sentiment analysis is expected to make a significant contribution in identifying and understanding the sentiment of the opinion or opinion sentences being analyzed.

The purpose of this research is to investigate or dig up specific information, understand certain phenomena, or test proposed hypotheses. Apart from that, research also aims to contribute to our understanding of a particular topic or problem. The significance or benefits of research can include several aspects. First, research can provide new insights and additional knowledge into a field of knowledge. Second, research results can be used as a basis for developing policies or best practices in various fields, such as health, education or technology. Apart from that, research can also contribute to scientific literature and become a reference for further research. The significance of research can also be seen in the possible practical application of the findings found, providing solutions to certain problems, or improving people's quality of life. Thus, research is not only beneficial for the development of science, but also has a positive impact on society and the surrounding environment. A brief conclusion from this research includes findings that provide new insights and additional knowledge to certain areas of knowledge. The Naive Bayes method in text sentiment analysis also contributes to policy development, scientific literature, and becomes a reference for further research. The significance of research is also seen in the possible practical application of the findings, which can provide concrete solutions to certain problems and improve people's quality of life.

2 Literature Review

Sentiment Analysis is a field in text mining research which aims to classify text documents. This process involves extracting a person's written opinions, emotions, and evaluations regarding a specific topic. In carrying out its duties, Sentiment Analysis uses natural language processing techniques to understand and interpret text in more depth. Essentially, this method allows the identification of positive, negative, or neutral attitudes contained in a text document, helping in understanding human views and responses to the topic. By leveraging advances in natural Language processing, Sentiment Analysis becomes a powerful tool in exploring opinions and feelings conveyed through text, providing valuable insights for decision making and deep understanding of user responses[6].Naive Bayes is a classification method based on simple probability and applied to classes that are independent of each other[7]. The literature can discuss the advantages and limitations of this method, as well as examples of its use in previous studies.

Research conducted by Fauziah Afshoh Journal entitled "Sentiment Analysis with Naïve Bayes on Tokopedia Application Comments" presents the results of research using the Naïve Bayes method to classify Tokopedia application user comments into positive and negative categories. Research steps include identifying problems through observing comments on Google Playstore, determining objectives to clarify the research framework, and evaluating the level of accuracy of data

 ISSN:2302-8149
e-ISSN:2540-9719

classification results using the Naïve Bayes method on test data. Sentiment analysis results show a certain percentage of comments are classified as positive and negative. Apart from that, this research provides benefits in providing information to Tokopedia management so that they can maintain consistent service to consumers, maintain the level of user trust, and ultimately increase profits in accordance with the targets set. Thus, this journal contributes to applying the sentiment analysis method using Naïve Bayes to comments from users of the Tokopedia application, which can provide valuable information for management in improving service quality and user trust.[8].

The second research was conducted by Achmad Ridok with the title "Sentiment Analysis of Tweets Data on Handling Covid-19 in Indonesia using the Naïve Bayes Method and Selection of Sentiment Words using Lexicon Based". The results of this journal include sentiment analysis of the policy for Implementing Restrictions on Community Activities (PPKM) in Indonesia using the Naïve Bayes method and selecting sentiment words using the Lexicon Based feature. From the sentiment analysis process carried out, it was found that sentiment analysis of PPKM policies in Indonesia can be carried out through various stages or processes, including preprocessing with the Lexicon Based feature, weighting using raw term frequency, and classification using the Naïve Bayes method. The essence of this journal is the application of sentiment analysis methods to understand the public's views on policies for handling Covid-19 in Indonesia based on Tweets data. By using the Naïve Bayes method and the Lexicon Based feature, this research provides insight into how sentiment analysis techniques can be used to understand public sentiment regarding policies for handling Covid-19. Apart from that, this journal also discusses the use of Twitter as a popular social media platform for expressing opinions and interacting regarding issues that are currently being discussed. Thus, this journal contributes to understanding the public's views on policies for handling Covid-19 in Indonesia through sentiment analysis. Tweets data-based using the Naïve Bayes method and Lexicon Based features[9].

The latest research was conducted by Oktavia Putri Zusrotun. The results of this research are that sentiment analysis using Twitter data with the keyword "online learning" in Indonesian shows that 50.75% of sentiment is positive, 83.3% of sentiment is negative, and 1% of sentiment is neutral. This research also develops sentiment analysis to classify tweets made by Twitter users as positive, negative or neutral using the Naïve Bayes method. The essence of this research is to find out the public's perspective on online learning based on public sentiment via Twitter, to be able to classify the polarization of positive and negative sentiments regarding online learning, to find out how much accuracy is obtained from the prediction results of the Naïve Bayes algorithm classification regarding online learning, and to create a model. classification using the Naïve Bayes method regarding online learning. This research also shows that social media, especially Twitter, can be used as a forum for voicing complaints about education[10].

By referring to the three studies above, the author conducted research with the aim of combining the concept of Sentiment Analysis using the Naïve Bayes method in the context of OTA services, especially PegiPegi.com. The steps of this research include identifying problems through observing user comments, setting goals to clarify the research framework, and evaluating the level of accuracy of data classification results using the Naïve Bayes algorithm on test data. This research was conducted to gain a deeper understanding of user responses to the PegiPegi.com OTA service, as well as to provide valuable information to management. By applying the Sentiment Analysis method using Naïve Bayes, it is hoped that this research can help maintain consistent service to consumers, increase the level of user trust, and ultimately achieve increased profits in accordance with the targets set. In line with the contributions of previous research, this research also aims to expand understanding of the application of Sentiment Analysis in a broader context, integrating concepts that have been successfully applied in previous research. Thus, it is hoped that this research can make a positive contribution in improving service quality and user trust in the OTA platform, PegiPegi.com, as well as enriching the literature regarding the application of Sentiment Analysis using the Naïve Bayes method. By utilizing advances in natural language processing, Sentiment Analysis has become a tool which is invaluable in exploring opinions and feelings expressed through text, providing valuable insights for decision making and deep understanding of user responses[5].

 ISSN:2302-8149
e-ISSN:2540-9719

3 Research methods

Figure 1 illustrates the research stages for conducting sentiment analysis of the OTA PegiPegi.com service using the Naive Bayes algorithm.

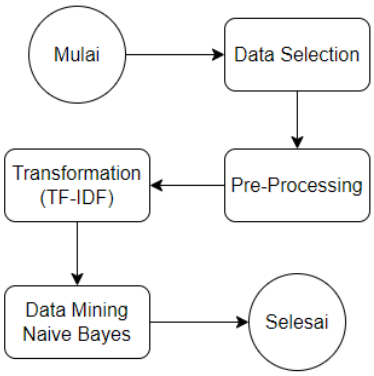

Figure 1. research steps

Figure 1 illustrates the process of analyzing text data using the Naive Bayes algorithm. The process begins with selecting relevant data, including data sources and required variables. After that, the data goes through a preprocessing stage to clean and prepare it so it is ready for analysis. The next step is the transformation of the data into a numerical representation using the TF-IDF method, which is the key to describing the weight of words in the context of the document. The choice of numerical representation allows the use of the Naive Bayes algorithm, which basically performs data classification based on the probability of key words in the text. This process helps in identifying patterns and sentiments contained in the dataset. After the classification results evaluation stage, this text data analysis provides an in-depth understanding of the dataset and can be the basis for further decisions or recommendations based on the analysis findings.

3.1. Data Selection

In this initial stage, the author carries out information collection and statistical labeling. Data was taken from the Shopee application on the Google Play Store using web scraping techniques using the Python programming language[11].

3.2. Data Pre-Processing

The pre-processing step in sentiment analysis is the initial stage of the sentiment analysis process. This involves processing unstructured input data into structured data before the main process, such as classification or sentiment analysis, is carried out. At this stage, several steps will be taken, including the following:

1. Cleaning: Cleaning is a method for cleaning emoticons and symbols carried out in this research. Emoticons and symbols in reviews were ignored because this research focused more on the text contained in the reviews. Ignored characters include "~", " `", " !", " $", " %", " ^", " &", " *", " (", ")", " _", "-", "+", "=", ":", "'", "comma", "period", "?". Reviews that initially contained characters like "please min, repair the delivery service" after being processed became "please min for the delivery service repair".
2. Case Folding: Case folding is a stage that involves converting the text in a file into a standard form, i.e. lower case. For example, the comment "Please improve the payment feature" would become "please improve the payment feature", the uppercase "T" is changed to lowercase "t".
3. Word Normalization (Word Normalizer): Word normalization is a technique used to correct words in a review in order to produce sentences that comply with correct and accurate Indonesian grammar. For example, if there is a comment "why did the update process take so long", after going through the Word Normalization process, it will become "why did the update process take so long". The change in the word "why" to "why" was done to make it easier to understand the message.

 ISSN:2302-8149
e-ISSN:2540-9719

4. Stopword Removal: Stopword deletion is the stage of eliminating words based on a list of conjunctions. Words like “in”, “and”, “the” will be removed.
5. Tokenization: Tokenization is a technique used to break text into words, taking into account punctuation and spacing boundaries. For example, the sentence "shopee is slow now" was then changed to "shopee", "now", "slow".
6. Stemming: Stemming is the process of reducing words to their base form or “root” word. For example, the word "cancel" would be changed to "cancel". Stemming is used in text processing and text analysis to help reduce word variation and increase consistency.

By applying these steps, sentiment analysis can be performed more effectively and more accurate results can be obtained[12].

3.3. Transformation

The transformation or attribute formation stage in text refers to the process of obtaining the required representation. At this stage, the author carried out feature extraction using the TF-IDF method. The following is the formula for TF – IDF.

$$y: W(d,t) = TF(d,t) \quad (1)$$

In this equation it can be explained that TF (d, t) is the frequency of term t in text d[11].

3.4. Data Mining

In the field of computer science, data mining refers to data mining techniques that aim to discover hidden patterns in a data set with the aim of generating new knowledge. Specifically, data mining involves various methods adapted to the intended use of the data, such as estimation, prediction, classification, clustering, and association[13].

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2)$$

Data mining is a data analysis process that has the aim of finding significant and previously unknown patterns or relationships. By applying the latest methods, data mining plays a role in uncovering valuable information from data, providing deeper understanding, and producing conclusions that can provide benefits to the data owner.[14].

4 Results and Discussion

The results of the study indicate a noteworthy performance of the Multinomial Naive Bayes classification model in analyzing sentiment from PegiPegi.com reviews. The study also addresses the broader implications for the online travel agency industry, providing a basis for future research and developments in sentiment analysis methodologies.

4.1. Data Selection

This step is part of the preparation in the data selection process. Information obtained from interviews will undergo an attribute selection stage before then undergoing a preprocessing process[15].

4.1.1. Data Scrapping

*Web scraping*is a technique used to retrieve large amounts of data from a website and store it in the form of a local file or database in table format. The web scraping process is included in the data mining stage category, which involves the integration of various scientific disciplines such as machine learning, pattern recognition, statistics, databases, and visualization. Data mining aims to uncover hidden information in data collections or datasets[16].

 ISSN:2302-8149
e-ISSN:2540-9719

```
from google_play_scraper import Sort, reviews

result, continuation_token = reviews (
    'com.pegipegi.dmd.pro',
    lang='id', # defaults to 'en'
    country='id', # defaults to 'us'
    sort=Sort.MOST_RELEVANT, # defaults to Sort.MOST_RELEVANT
    count=1000, # defaults to 100
    filter_score_with=None # defaults to None(means all score)
)
```

✓ 0.7s

Figure 2. Data scrapping process

Figure 2 utilizes the `google_play_scraper` library to access reviews of the Pegipegi application in the Google Play Store. Through the `reviews` function, applications are identified based on the package ID 'com.pegipegi.dmd.pro', with the review language set to Indonesian ('id') and the country set to Indonesia ('id'). Reviews are sorted by relevance using the `Sort.MOST_RELEVANT` parameter, and 1000 reviews are taken with the `count=1000` parameter. No review score filter is applied (`None`), so all review scores are taken. The result of this function includes two values, namely `result` which contains a list of reviews and related information, and `continuation_token` which can be used to continue fetching reviews if necessary. With this configuration, we can collect user reviews of the Pegipegi application from the Google Play Store for further analysis.

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion
0	16ce1d6e-c9b2-4f95-a980-1364960eda32	Aden	https://play-lh.googleusercontent.com/a/ACg8oc...	Aplikasi travel agen paling burukl.. Saya book...	1	25	12.00.0	2024-01-16 05:29:49	Kami memohon maaf atas ketidaknyamanannya. Kam...	2024-01-16 07:54:12	12.00.0
1	c5fa4148-11cc-43eb-a878-0984303a0ea9	Siska Dianisa	https://play-lh.googleusercontent.com/a-/ALV-U...	Sangat kecewa,wkt si buka aplikasi masih ters...	1	305	11.46.0	2023-11-23 21:39:30	Kami mohon maaf atas ketidaknyamanan yang Anda...	2023-11-24 07:06:33	11.46.0
2	a3b35fb8-7449-407d-a0d2-492649decf53	Fikri Lana	https://play-lh.googleusercontent.com/a/ACg8oc...	Saya booking villa dan sudah bayar via debit d...	1	290	11.47.0	2023-12-11 17:53:00	Yth. Fikri Lana,\nKami mohon maaf atas ketidak...	2023-12-11 18:16:51	11.47.0
3	3ef1c87a-14aa-4927-bfb9-c615db2bfb84	Insani Miftahul Janah	https://play-lh.googleusercontent.com/a-/ALV-U...	Sangat mengecewakan. Tampilan aplikasi membing...	2	81	11.45.0	2023-11-21 12:52:03	Hai Insani Miftahul Janah, Kami mohon maaf ata...	2023-11-21 12:49:54	11.45.0
4	ddaa5e18-f1c3-43b0-9a54-176166801155	Shakila Aqueena	https://play-lh.googleusercontent.com/a/ACg8oc...	Kecewa sekali sama aplikasi ini. Sudah pesan ...	1	97	11.49.0	2023-12-30 22:06:57	Kami mohon maaf atas ketidaknyamanan yang Anda...	2023-12-31 07:07:29	11.49.0

Figure 3. results from scrapping data

Data from Figure 3 can be used by developers or researchers to gain further insight from user reviews of the Pegipegi application on the Google Play Store. This information can serve as a basis for evaluating application performance, planning feature updates, or improving services based on feedback received from users.

4.1.2. Labeling data

In the context of machine learning, data labeling is the process of identifying and adding informative labels to raw data, such as images, text, or video. The goal of data labeling is to provide the necessary context so that machine learning models can learn and produce more accurate predictions or results. For example, labels can indicate whether an image contains objects such as birds or cars, translate spoken words in an audio recording, or identify the presence of a tumor in an X-ray. The data labeling process is critical in a variety of machine learning use cases, including computer vision, natural language processing, and speech recognition[17].

	content	score	Label
43	proses pembatalan (refund) disetujui sejak tgl...	1	Negative
500	mantabs banyak diskon dan easy	5	Positive
498	Aplikasi penipu sampai si hotel saya di mintai...	1	Negative
234	Saya booking hotel bulan nov 2023, saldo sudah...	1	Negative
238	Sudah boking dan bayar jauh jauh hari ternyata...	1	Negative
797	Ini aplikasi paling gak jelas, punya promo unt...	1	Negative
504	Agoda harga berubah ubah sebel	3	Neutral

Figure 4. Results from Data Labeling

 ISSN:2302-8149
e-ISSN:2540-9719

Figure 4 shows the results of the data labeling process. The results of data labeling are very instrumental in developing a sentiment analysis model that is accurate and appropriate to the specific context of Pegipegi.com. By using data that has been properly labeled, the model can achieve better learning and adapt to classify sentiment more accurately.

4.2. Data Pre-Processing

The pre-processing process in the Naive Bayes algorithm includes a series of steps involving data cleaning by handling missing values, tokenization to separate word units, converting letters to lowercase, eliminating common words, and applying a stemming or lemmatization process. Additionally, feature extraction such as TF-IDF is used to convert data into vector representation. It is also important to handle outliers and convert categorical variables with the encoding method to ensure that the data is ready to be used for training or prediction using the Naive Bayes algorithm.

	content	score	Label
0	proses pembatalan (refund) disetujui sejak tgl...	1	Negative
1	mantabs banyak diskon dan easy	5	Positive
2	Aplikasi penipu sampai si hotel saya di mintai...	1	Negative
3	Saya booking hotel bulan nov 2023, saldo sudah...	1	Negative
4	Sudah boking dan bayar jauh jauh hari ternyata...	1	Negative
5	Ini aplikasi paling gak jelas, punya promo unt...	1	Negative
6	Agoda harga berubah ubah sebel	3	Neutral

Figure 5. Results from data pre-processing

Figure 5 shows the output of the pre-processing process, where the text has been processed to become more structured, standardized, and ready to be used in the next analysis stage. Data that has undergone pre-processing has a higher potential to produce accurate and relevant analysis, especially when applied in machine learning models such as the Naive Bayes sentiment classification algorithm.

4.3. Transformation

Data transformation in the Naive Bayes algorithm refers to the steps to change or convert text data into a representational form that can be utilized by the algorithm in carrying out classification analysis. In this context, the data transformation process involves converting text into feature vectors that can be understood and used by Naive Bayes for classification purposes.

```
from sklearn.feature_extraction.text import TfidfVectorizer

tfidf_vectorizer = TfidfVectorizer()
tfidf_train = tfidf_vectorizer.fit_transform(X_train)
tfidf_test = tfidf_vectorizer.transform(X_test)
```

✓ 0.0s

```
print(X_train.shape)
print(y_train.shape)
print(X_test.shape)
print(y_test.shape)
```

✓ 0.0s

```
(800,)
(800,)
(200,)
(200,)
```

Figure 6. TF-IDF weighting and results

 ISSN:2302-8149
e-ISSN:2540-9719

Figure 6 shows the weighting process using the Term Frequency-Inverse Document Frequency (TF-IDF) method using the `scikit-learn` library. In the code, the `TfidfVectorizer` object from `scikit-learn` is used to convert text to a numerical representation based on the frequency of words in the document and the inverse document frequency of those words. The training data (`X_train`) and test data (`X_test`) are then converted into a TF-IDF matrix using the `fit_transform` and `transform` methods in the `TfidfVectorizer` object. The results of this transformation produce a numerical representation that can be used as input for training and testing machine learning models, such as the Naive Bayes classification algorithm. The final print statement displays the dimensions of the TF-IDF matrices for the training and test data, as well as the dimensions of the associated labels (`y_train` and `y_test`). This TF-IDF representation can be used in a variety of text processing tasks, including sentiment analysis.

4.4. Data Mining

The evaluation results of the Multinomial Naive Bayes classification model show generally good performance, with an accuracy rate of 89.5%. The high precision in the Negative class (164 out of 164) indicates that the model is very good at identifying truly negative reviews. However, it should be noted that the model tends to face difficulties in classifying the Neutral class, with the precision and recall of this class each reaching 0.00. The overall F1-score of 0.522 reflects a good balance between precision and recall.

```
MultinomialNB Accuracy: 0.895
MultinomialNB Precision: 0.6288288288288288
MultinomialNB Recall: 0.48484848484848486
MultinomialNB f1_score: 0.5216093600764088
confusion_matrix:
 [[164   0   0]
 [  3   0   0]
 [ 18   0  15]]
==================================================

              precision    recall  f1-score   support

    Negative       0.89      1.00      0.94       164
     Neutral       0.00      0.00      0.00         3
    Positive       1.00      0.45      0.62        33

    accuracy                           0.90       200
   macro avg       0.63      0.48      0.52       200
weighted avg       0.89      0.90      0.87       200
```

Figure 7. research results

Figure 7 is the result of research that has been carried out**.**The confusion matrix presents a comparison between the actual classification results and the predicted results. From the confusion matrix, it can be seen that the model tends to focus more on the Negative class, but has challenges in recognizing the Neutral class. This model evaluation provides valuable insight into model performance and can be used as a foundation for further improvements, particularly in addressing class imbalance and improving the ability to recognize reviews with neutral sentiment.

5 Conclusion

The conclusions of this research include an in-depth understanding of consumer trends turning to online ticket reservations through platforms such as PegiPegi.com. With the increasing preference for convenience and efficiency in the booking process, PegiPegi.com has succeeded in creating a strong position as a fast-growing online travel agent in Indonesia. Strategic steps, partnerships with service providers, and technological innovation are key factors for the company's success. The importance of sentiment analysis in the context of using platforms such as PegiPegi.com is also revealed. Sentiment analysis allows companies to understand user views and experiences through reviews and testimonials. By using the Naïve Bayes classification method, this research aims to contribute to improving service quality, maintaining consistency, and increasing user trust.

By referring to previous research related to sentiment analysis, especially those carried out on the Tokopedia application and public responses to PPKM policies in Indonesia, this research combines

 ISSN:2302-8149
e-ISSN:2540-9719

these concepts in the context of the PegiPegi.com OTA service. Through this research, it is hoped that new insights can be obtained regarding user responses to PegiPegi.com services, as well as making a positive contribution in improving service quality and user trust in the platform. In conclusion, this research focuses not only on the application of sentiment analysis methods, but also on a broader understanding of the use of these concepts in improving user experience on OTA platforms. Furthermore, the evaluation results of the Multinomial Naive Bayes classification model show generally good performance, with an accuracy level of 89.5%. The high precision in the Negative class (164 out of 164) indicates that the model is very good at identifying truly negative reviews. However, it should be noted that the model tends to face difficulties in classifying the Neutral class, with the precision and recall of this class each reaching 0.00. The overall F1-score of 0.522 reflects a good balance between precision and recall, indicating the model's ability to provide balanced performance in identifying positive and negative reviews.

Reference

[1] A. B. P. Negara, H. Muhardi, and I. M. Putri, "Analisis Sentimen Maskapai Penerbangan menggunakan Metode Naive Bayes dan Seleksi Fitur Information Gain," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, p. 599, 2020, doi: 10.25126/jtiik.2020711947.

[2] I. A. Wulandari, F. Umar, and E. Setyawati, "Strategi Peningkatan Kualitas Pelayan Online Travel Agent Pegipegi dengan Quality Function Deployment," *J. REKAYASA dan OPTIMASI Sist. Ind.*, vol. 2, no. October, pp. 32–37, 2020, [Online]. Available: https://journal.univpancasila.ac.id/index.php/jrosi/article/view/2474

[3] Fiaz, "Strategi Bisnis Pegipegi: Menghadirkan Kemudahan dalam Perjalanan Anda," https://perpusteknik.com/. [Online]. Available: https://perpusteknik.com/strategi-bisnis-pegipegi/

[4] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.

[5] M. Arsyam and M. Yusuf Tahir, "Ragam Jenis Penelitian dan Perspektif," *Al-Ubudiyah J. Pendidik. dan Stud. Islam*, vol. 2, no. 1, pp. 37–47, 2021, doi: 10.55623/au.v2i1.17.

[6] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen terhadap Penggunaan Aplikasi Shopee mengunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjeee.v5i1.16830.

[7] S. Informasi, U. Darwan Ali, J. Batu Berlian, and K. Tengah, "Penerapan Metode Naive Bayes dalam Mengklasifikasi Penerima BLT pada Desa Pelangsian," *J. JUPITER*, vol. 14, no. 2, pp. 64–70, 2022.

[8] R. Apriani *et al.*, "Analisis Sentimen dengan Naïve Bayes terhadap Komentar Aplikasi Tokopedia," *J. Rekayasa Teknol. Nusa Putra*, vol. 6, no. 1, pp. 54–62, 2019, [Online]. Available: https://rekayasa.nusaputra.ac.id/article/view/86

[9] A. A. A. Sumanjaya, Indriati, and A. Ridok, "Analisis Sentimen Data Tweets terhadap Penanganan Covid-19 di Indonesia menggunakan Metode Naïve Bayes dan Pemilihan Kata Bersentimen menggunakan Lexicon Based," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 4, pp. 1865–1872, 2022, [Online]. Available: http://j-ptiik.ub.ac.id

[10] O. P. Zusrotun, A. C. Murti, and R. Fiati, "Analisis Sentimen terhadap Belajar Online pada Media Sosial Twitter menggunakan Algoritma Naive Bayes," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 3, pp. 310–319, 2022, doi: 10.23887/janapati.v11i3.49160.

[11] K. M. Elistiana, B. A. Kusuma, P. Subarkah, and H. A. A. Rozaq, "Improvement of Naive Bayes Algorithm in Sentiment Analysis of Shopee Application Reviews on Google Play Store," vol. 4, no. 6, pp. 1431–1436, 2023, [Online]. Available: https://doi.org/10.52436/1.jutif.2023.4.6.1486

[12] 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.

[13] J. Dongga, A. ` Sarungallo, N. Koru, and G. Lante, "Implementasi Data Mining menggunakan

Algoritma Apriori dalam Menentukan Persediaan Barang (Studi Kasus: Toko Swapen Jaya Manokwari)," *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, pp. 119–126, 2023, doi: 10.33379/gtech.v7i1.1938.

[14] N. Noviyanto, "Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 berdasarkan Negara di Benua Asia," *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 183–188, 2020, doi: 10.31294/p.v22i2.8808.

[15] F. Alghifari and D. Juardi, "Penerapan Data Mining pada Penjualan Makanan dan Minuman menggunakan Metode Algoritma Naïve Bayes," *J. Ilm. Inform.*, vol. 9, no. 02, pp. 75–81, 2021, doi: 10.33884/jif.v9i02.3755.

[16] A. Z. Rizquina and C. I. Ratnasari, "Implementasi Web Scraping untuk Pengambilan Data pada Website E-Commerce," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 4, pp. 377–383, 2023, doi: 10.47233/jteksis.v5i4.913.

[17] Amazon, "Apa Itu Pelabelan Data?," https://aws.amazon.com/. Accessed: Jan. 30, 2024. [Online]. Available: https://aws.amazon.com/id/what-is/data-labeling/