JoCoSiR
Journal of Computer Science Research
ISSN: 2986-2337(Online)

SENTIMENT ANALYSIS OF E-COMMERCE REVIEWS WITH NATURAL LANGUAGE PROCESSING (NLP)

Rahmat Idhami [a], Andri Saputra [b], Taufa Fadly [c], Robet Silaban [d].

[a] *Magister Teknologi Informasi, Universitas Pembangunan Panca Budi.*

email: [a] idhami531@gmail.com, [b] asaputra14183@gmail.com , [c] taufan.fadly@gmail.com, [d] robert.silaban96@uhn.ac.id.

ARTICLEINFO

Keywords:
sentiment analysis
Natural Language Processing
Indonesian e-commerce

IEEE style in citing this article:
Rahmat Idhami, Andri Saputra, Taufa Fadly, Robert Silaban "E-Commerce Review Sentiment Analysis Using Natural Language Processing (NLP)" JoCoSiR: Scientific Journal of Information Systems Technology, vol. 7, no. 1, pp. 1-10, 2021. [Fill citation heading]

ABSTRACT

E-commerce in Indonesia is growing rapidly, with Shopee as a leading platform. This study uses Natural Language Processing algorithms to analyze customer satisfaction sentiment from reviews on the Google Play Store. The results identify issues related to courier services and provide recommendations for improving service quality, delivery tracking systems, and overall customer satisfaction and loyalty towards Shopee. This chapter describes the research methodology for sentiment analysis of Shopee reviews using Natural Language Processing methods. These stages include data collection, cleaning, pre-processing, labeling, data separation, classification, and negative word analysis. This study aims to identify the dominant negative sentiment in Google Play Store reviews. This study outlines data scraping, cleaning, pre-processing, labeling, and Natural Language Processing classification to identify negative words in Shopee user reviews. This method provides insights into courier service issues and recommendations for couriers frequently highlighted in reviews, with a focus on future service improvements. Based on the study, Natural Language Processing is effective in identifying positive and negative sentiment in Shopee with an accuracy of 86-87%. Negative sentiment was dominant (62.5%), particularly regarding "recommended couriers," with complaints about delays and unprofessionalism. Recommendations included improving courier service quality, delivery tracking systems, customer communication, and courier training and supervision to improve customer satisfaction.

1. Introduction

In today's digital age, the internet has made it much easier for consumers to search for information before deciding to purchase a product. One factor consumers consider when purchasing a product is reviews from other customers. These reviews provide insight into the product's user experience and provide information about its quality.[1]

However, the number of reviews available for a product can be quite large, especially for well-known products. In this case, manually analyzing each product review individually would be time-consuming and resource-intensive. Therefore, a fast and accurate sentiment analysis method is needed to help obtain information about consumer views on a product.[2]

Natural Language Processing (NLP) is a technology that enables computers to understand and analyze human language. One application of NLP is sentiment analysis, the process of identifying whether a text contains positive, negative, or neutral sentiment. In the context of product reviews, sentiment analysis can be used to extract customer views on purchased products[3].

Shopee is one of the largest and most popular online stores in Indonesia. This platform provides a variety of products at affordable prices and offers an easy and secure online shopping experience. Over time, Shopee has become the platform of choice for sellers to market their products online.[4] However, with the large number of product reviews provided by consumers, it is difficult for sellers to understand and manage these reviews effectively. In this case, the application of sentiment analysis on product reviews using the NLP method can be a solution for sellers on Shopee to understand consumer opinions on the products they sell and make necessary improvements or changes to enhance the quality of their products and services.[5]

The author used Shopee, a store that sells electronics and household goods, as the research object. The author chose this store because he saw potential for further development, given that several buyers had left negative reviews. Analyzing each customer's review can help the store owner take further steps to address customer issues.

In this study, the author will focus on four products sold by Shopee. These four products are electronic equipment and household appliances that have varied reviews, both positive and negative. Each product review will be processed by a model to classify it as positive or negative. The author will count the number of positive and negative reviews for each product and generate a report or visualization explaining consumer views on each product. The final result of this study is a sentiment analysis model for product reviews in the Shopee online store, which can be used as a basis for developing applications or systems related to sentiment analysis on e-commerce platforms.

ISSN 2502-3357 (Online)

2. State of the Art

2.1. Sentiment Analysis

"Sentiment analysis is a data cleaning technique of irrelevant words and symbols and changing qualitative data into quantitative data, then user review data will be classified to get positive and negative reviews" [1]. "Sentiment analysis is classifying each pole of text on various internet sources and social media in the form of documents or sentences then determining whether the word is included in the positive or negative category" [2]. Sentiment analysis is comprehensive enough for this research. What is described by the two researchers above also illustrates how sentiment analysis is used to group text, identify positive and negative.

2.2. Web Scraping

"Web scrapping is a technique for collecting data from the internet, applications, or social media, for example, the Google Play Store. Web scraping takes semi-structured document data, web scraping has the goal of taking information that will be used for needs, either in whole or in part" [3]. This dataset will also be converted into a CSV file after being taken to make it easier for researchers to apply it in Google Colab.

2.3. Google Colab

Google Colab is a cloud-based service provided by Google to run the Python language through a web browser. Google Colab is used for machine learning, data analysis, and can also be used for deep learning development. Google Colab can also run Python code without having to install additional software. As explained above, this study uses Google Colab for the sentiment analysis being carried out.

2.4. Google Playstore

Google Play Store is an official app store developed by Google that allows users to search and download apps for the Android system. Tokopedia app review data will be obtained using Web Scraping techniques, which is the process of extracting data from website pages. Tokopedia review data is taken using the Google-playscraper API, Google-play-scraper is an API for extracting app information data and app reviews from the Google Play Store more easily without external dependencies [4].

2.5. Customer satisfaction

"Customer satisfaction is related to positive or negative results of a service and product consumed according to customer expectations. If the performance of the product or service produced by the company is the same as what the consumer imagined, the consumer will be very satisfied" [5]. Customer satisfaction is very important for e-commerce such as Tokopedia because it can influence the quality of service available at Tokopedia. Customer satisfaction is generally recognized as the main factor influencing the formation of loyalty in e-commerce.

3. Method

In this study, sentiment analysis was conducted using natural language processing on product reviews at Zalika Store 88, Shopee. The Sentiment Analysis method is an approach and technique used to analyze opinions, sentiments, and emotions contained in text. This method can be used to understand users' views on a topic, product, service, or brand. Meanwhile, Natural Language Processing (NLP) is a branch of computer science and artificial intelligence that focuses on the understanding, processing, and generation of human language by computers [6]. NLP allows computers to interact with humans through natural language, which can be in the form of speech or text. The main goal of NLP is to understand, analyze, and generate text in human language, as well as to develop systems that can process and communicate with humans in a way similar to human communication [7].

NLP involves solving several challenges in human language processing. Human language has a complex structure, including grammar, ambiguity, the use of different words in different contexts, figurative expressions, and more. Therefore, NLP uses a computational approach to understand and process human language[8]. Stages of NLP The stages in natural language processing through NLP include several main steps, including[9]:

1. Tokenization: The process of separating text into smaller units such as words, phrases, or sentences. Tokenization is important because it helps computers understand the structure and meaning of a given text.
2. Preprocessing: This stage involves steps to clean and prepare the text for further analysis. Preprocessing includes text normalization (changing capital letters to lowercase, removing punctuation, and so on), filtering (removing common words that don't add significant meaning), and stemming (removing affixes to get the root word).
3. Parsing: The process of grammatical analysis that aims to break down text into structured units such as phrases or sentences. Parsing helps in understanding the syntactic structure of the text being analyzed.

 http://doi.org/10.65126/jocosir.v2i2.61

4. Information Extraction: This stage involves extracting relevant information from the given text. This information can be entities (such as names of people, places, or dates), relationships between entities, or other attributes that are important in the text.
5. Sentiment Classification: One important aspect of NLP is sentiment analysis, which is the recognition and classification of sentiment in text, such as positive, negative, or neutral. Sentiment classification allows us to understand the feelings or opinions contained in the text.

4. Results and Discussion

The results obtained were obtained by extracting review data from four products using web scraping techniques. The following script implements web scraping to collect product reviews from four different Shopee product links. Web scraping is the process of automatically extracting data from web pages using a computer program.

```
import re
import json
import requests
urls = [
   'https://shopee.co.id/-MULUS-KIPAS-ANGIN-MODEL-AC-KIPAS-ANGIN-AC-2PK-3PK-(REMOTE)(LED)-i.605092469.15108321157',
   'https://shopee.co.id/-BAGUS-Aquarium-Mini-Tabung-Mika-Cocok-Cupang-2IN1-LAMPU-TIDUR-D-10cm-T-i.605092469.14623940621',
   'https://shopee.co.id/-BARU-RAK-DINDING-MINIMALIS-MODEL-ANGKLUNG-SINGLE-RAK-KAYU-i.605092469.11174914826',
   'https://shopee.co.id/-PREMIUM-Kurma-SUKARI-EMBER-Al-Qossim-850-GR-sukari-ember-Premium-Saudi-Ruthob-i.605092469.21465230217'
]
for url in urls:
   r = re.search(r'i\.(\d+)\.(\d+)', url)
   shop_id, item_id = r[1], r[2]
   ratings_url = 'https://shopee.co.id/api/v2/item/get_ratings?filter=0&flag=1&itemid={item_id}&limit=20&offset={offset}&shopid={shop_id}&type=0'
   offset = 0
   while True:
      data = requests.get(ratings_url.format(shop_id=shop_id, item_id=item_id, offset=offset)).json()
      for i, rating in enumerate(data['data']['ratings'], 1):
         print(rating['author_username'])
         print(rating['comment'])
         print('-' * 80)
      if i % 20:
         break
      offset += 20
```

Figure 1 Shopee Data Scrapping

The script generated a list of various Shopee product reviews. The next step was to analyze the sentiment of each review using natural language processing methods.

A. Data Reading

The first step in the analysis is to read the dataset into Python using the pandas library, which allows data management in tabular form (dataframe).

```
import pandas as pd
# Reading CSV dataset
data = pd.read_csv("shopee_reviews.csv")
# Displays the first five lines
print(data.head())
```

Main columns of the dataset:

Column	Information
user	Shopee username
review	Original review text from customers
label	Sentiment categories: 0 = Negative, 1 = Positive
remove emoji	Review text after emojis are removed

The number of reviews used was 2937 rows, which included both positive and negative reviews. Initial inspection (`data.head()`) ensure the dataset structure is read correctly.

Vital Records: Checking the initial data is important to ensure there are no missing values or incorrect character encoding, so that further analysis can run smoothly.

B. Data Preprocessing

Review texts are usually unstructured and contain variations such as:

- Capital letters and lowercase letters

ISSN 2502-3357 (Online)

- Punctuation (!, ?, ,)
- Symbols and emojis

If raw data is fed directly into an ML model, the results will be suboptimal because the model cannot understand the meaning of the same word if it is spelled differently. Therefore, preprocessing is performed, which includes:

2.1 Lowercasing
All letters are changed to lowercase. Example:

- "Good" → "good"
- "GOOD" → "good"

This ensures that the same words are considered identical by the model.

2.2 Removing non-alphabetic characters
Characters such as numbers, punctuation, symbols, or emojis are removed to leave only the relevant words. Example:

- "Good!!! 😊 " → "good"

2.3 Removing excess spaces
Some words can be separated by more than one space, so they need to be cleaned up to a single space.

2.4 Indonesian Stemming
Derived words are converted to their base form using the Sastrawi library. Example:

- "good" → "good"
- "buy" → "buy"
- "late" → "slow"

Python Implementation:
```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import re
stemmer = StemmerFactory().create_stemmer()
def clean_text(text):
    if pd.isna(text):
        return ""
    text = str(text).lower() # lowercase
    text = re.sub(r'[^az\s]', ' ', text) # remove non-alphabetic characters
    text = re.sub(r'\s+', ' ', text).strip() # remove excess spaces
    return stemmer.stem(text)

data['clean'] = data['review'].apply(clean_text)
```

Example of clean & stem results:

Original Text	Clean & Stem
"This product is great"	"this product is good"
"Buying it is easy"	"easy buy"
"Fast & friendly shopping app"	"fast friendly shopping app"

Benefits of preprocessing:

- Reducing word redundancy
- Improve text consistency
- Makes it easier for the model to recognize important word patterns

C. TF-IDF Feature Extraction
Once the text is cleaned, the next step is to convert the text into numbers so that it can be processed by the Machine Learning model.
The method used is TF-IDF (Term Frequency-Inverse Document Frequency):

- **TF (Term Frequency):**Counts how often a word appears in a review.
- **IDF (Inverse Document Frequency):**Gives higher weight to words that rarely appear across reviews, and lowers the weight of common words (e.g. "product" or "good").

Python Implementation:
```
from sklearn.feature_extraction.text import TfidfVectorizer
```

ISSN 2502-3357 (Online)

```
vectorizer = TfidfVectorizer(max_features=3000)
X = vectorizer.fit_transform(data['clean'])
y = data['label']
```

The result is a numerical matrix with the top 3,000 words as features. Each row of the matrix represents a review, and each column represents the word's weight.

Important reasons: Naïve Bayes models cannot work directly with text, so this numerical representation is very important.

D. Model Training

The dataset is divided into:
- **Training data (80%)**: for the model to learn word patterns related to sentiment
- **Test data (20%)**: to test the generalization ability of the model

The model used is Multinomial Naïve Bayes, which is effective for text because:
- Calculating word probabilities per class
- Fast and stable for medium sized datasets

Python Implementation:
```
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
model = MultinomialNB()
model.fit(X_train, y_train)
```
This model produces a predicted probability for each review, whether it is positive (1) or negative (0).

E. Model Evaluation

After the model is trained, it is evaluated using previously unseen test data. The evaluation metrics used are:
1. **Accuracy**– percentage of correct predictions from all test data
2. **Precision**– model accuracy when predicting a particular class
3. **Recall (Sensitivity)**– the model's ability to detect all data from the correct class
4. **F1-score**– combination of precision and recall to assess model balance

Python Implementation:
```
from sklearn.metrics import accuracy_score, precision_score, recall_score,
f1_score, classification_report, confusion_matrix

y_pred = model.predict(X_test)

acc = accuracy_score(y_test, y_pred)
prec = precision_score(y_test, y_pred, average='weighted')
rec = recall_score(y_test, y_pred, average='weighted')
f1 = f1_score(y_test, y_pred, average='weighted')

print("Accuracy:", acc)
print("Precision:", prec)
print("Recall:", rec)
print("F1-Score:", f1)
print(classification_report(y_test, y_pred))
```

Model evaluation results:

Evaluation Metrics	Mark (%)
Accuracy	87.59
Precision	87.88
Recall	87.59
F1-Score	87.64

Classification report per class:

Label	Precision	Recall	F1-Score	Support
0	0.82	0.89	0.86	243
1	0.92	0.87	0.89	345

Confusion Matrix:

Actual\Predictions	0	1
0	216	27
1	45	300

Interpretation of results:

- Of the 243 negative reviews, 216 were classified as true, 27 as false.
- Of the 345 positive reviews, 300 were classified as true, 45 as false.
- Prediction error is relatively even → the model is not biased towards one class.
- High precision and recall indicate a stable and balanced model.

5. Conclusions

This study successfully applied the Natural Language Processing (NLP) method to analyze the sentiment of Shopee customer reviews taken from the Google Play Store. Based on this case study, it can be concluded:

- **Model Effectiveness:**The NLP method using Multinomial Naïve Bayes classification showed high accuracy in identifying positive and negative sentiments, with an accuracy value reaching 87.59%.The model was also assessed as stable and balanced, as indicated by high and even Precision, Recall, and F1-Score values.
- **Dominance of Negative Sentiment:**Negative sentiment is the dominant sentiment (62.5%) in customer reviews.
- **Courier Service Issues:**Negative complaints mainly relate to "recommended couriers", with the main issues highlighted being late deliveries and unprofessionalism of the couriers..
- **Methodology:**The research stages include data collection through web scraping, data pre-processing (such as lowercasing and Sastrawi stemming), feature extraction using TF-IDF, and sentiment classification.. The total number of reviews analyzed was 2937 rows

6. Acknowledgment

These suggestions aim to address the dominant negative sentiments identified through NLP sentiment analysis and improve Shopee's overall service quality.

1. Focus on Improving Courier Services
   Improvements need to be made to the quality of courier service which is often highlighted in reviews..
   Special attention should be paid to complaints regarding late deliveries and unprofessionalism of couriers.
2. Improvement of Logistics and Communication Systems
   - **Tracking System:**Improve the shipment tracking system so that customers get more accurate information..
   - **Communication:**Improve customer communication regarding shipping status and any issues that may arise..
3. Training and Supervision
   Conduct better training and supervision of couriers to ensure professional service, which will ultimately increase customer satisfaction and loyalty towards Shopee.

Discussion

This study successfully implemented the **Natural Language Processing (NLP)** method to analyze the sentiment of Shopee customer reviews, with data collected from the Google Play Store. The research focused on identifying the dominant sentiment (positive or negative) and pinpointing the specific issues most frequently highlighted by customers.

1. Dominance of Negative Sentiment

The main finding reveals that **negative sentiment was dominant** among customer reviews, accounting for **62.5%** of the total reviews analyzed. The total number of reviews used in this analysis was **2937** rows. The dominance of negative sentiment indicates that, despite Shopee being one of the largest and most popular platforms in Indonesia , there is an urgent need for service improvements to enhance customer satisfaction.

2. Primary Courier Service Issues

In-depth analysis of the negative sentiment showed that complaints were overwhelmingly dominated by issues related to **"recommended couriers"**. The main problems highlighted were:

- **Late deliveries**.
- **Unprofessionalism of the couriers**.

This finding provides very specific and actionable insights, suggesting that logistics and delivery services are the primary hurdles to Shopee's customer satisfaction.

3. Effectiveness of the NLP Method

The NLP method used, which employed **Multinomial Naïve Bayes** classification and **TF-IDF (Term Frequency-Inverse Document Frequency)** feature extraction, proved to be highly effective.

- **Model Accuracy:** The model achieved a high accuracy value of **87.59%** in identifying positive and negative sentiments.
- **Stability and Balance:** The evaluation showed that the model was **stable and balanced**, as indicated by the high and consistent values for Precision, Recall, and F1-Score (e.g., Precision 87.88%, Recall 87.59%, F1-Score 87.64%). The Confusion Matrix analysis further confirmed a relatively even prediction error, suggesting the model was not biased toward one class.
- **Methodology Stages:** The research process included key NLP steps such as data collection through **web scraping** , **data pre-processing** (including lowercasing, removal of non-alphabetic characters, and Indonesian *stemming* using the Sastrawi library) , feature extraction , data splitting (80% training, 20% testing) , and model training.

4. Recommendations for Follow-up

Based on the sentiment analysis, the study presents several strategic recommendations for Shopee to address the dominant negative sentiment and enhance customer loyalty:

- **Focus on Improving Courier Services:** Improvements must be made to the quality of courier service, with special attention paid to complaints regarding late deliveries and unprofessionalism.
- **Improvement of Logistics and Communication Systems:** Enhance the shipment tracking system so that customers receive more accurate information, and improve customer communication regarding shipping status and any issues that may arise.
- **Training and Supervision:** Conduct better training and supervision for couriers to ensure professional service, which will ultimately increase customer satisfaction and loyalty towards Shopee.

In conclusion, this research demonstrates that NLP is an accurate and effective tool for transforming large volumes of textual review data into actionable business insights.

7. References

[1] VR Prasetyo, N. Benarkah, and VJ Chrisintha, “Implementation of natural language processing in creating chatbots in the information technology program at the University of Surabaya,” J. Tek., vol. 10, no. 2, pp. 114–121, 2021.
[2] HE Rosyadi and others, “Design and construction of a WhatsApp-based job vacancy information chatbot using the NLP (Natural Language Processing) method,” BRILIANT J. Research and Conceptual, vol. 5, no. 1, pp. 619–626, 2020.
[3] PR Togatorop and others, “Generating Entity Relationship Diagrams from Requirements Specifications Using Natural Language Processing for Indonesian,” J-Icon J. Comput. and Inform., vol. 9, no. 2, pp. 196–206, 2021.

[4] DO Sihombing, “Implementation of Natural Language Processing (NLP) and Cosine Similarity Algorithm in Automatic Essay Exam Scoring,” J. Comput. and Inform. System, vol. 4, no. 2, pp. 396–406, 2022.
[5] N. Munasatya and S. Novianto, “Natural Language Processing for Sentiment Analysis of President Jokowi Using Multi Layer Perceptron,” Techno. Com, vol. 19, no. 3, pp. 237–244, 2020.
[6] AY Chandra, D. Kurniawan, and R. Musa, “Chatbot Design Using Dialogflow Natural Language Processing (Case Study: Ordering System in a Coffee Shop),” J. Media Inform. Budidarma, vol. 4, no. 1, pp. 208–215, 2020.
[7] I. Puspasari and PH Rusmin, “Classification of Wazan in Al-Qur'an Words Using Natural Language Processing,” J. Technol. Informatics, vol. 3, no. 2, pp. 41–48, 2022.
[8] E. Wita, “Application of Natural Language Processing to Identify Ambiguous Sentences in Newspapers Using Shift Reduce Parsing Method,” BEES Bull. Electr. Electron. Eng., vol. 2, no. 2, pp. 63–66, 2021.
[9] IGR Kusnanda, IM Sukarsa, and AANH Susila, “Hotel Chatbot Design with Natural Language Processing Chatbot Model and Button-Based Chatbot,” JITTER-Jurnal Ilm. Teknol. dan Komput., vol. 3, pp. 711–720, 2022.
[10] AA Chandra and others, “Development of Telegram-Based Student Information Chatbot Using Natural Language Processing Method,” J. ICTEE, vol. 3, no. 1, pp. 20–27, 2022.
[11] YA Alfaridzi, “IoT Device Control System Using Voice Command Based on Natural Language Processing.” 2021.