HASIL CEK_Sentiment_Analysis_of_the_Waste_Problem

by UNIVERSITAS AHMAD DAHLAN 1

Submission date: 06-Feb-2024 07:58AM (UTC+0700)
Submission ID: 2186303391
File name: Sentiment_Analysis_of_the_Waste_Problem.pdf (1.4M)
Word count: 7128
Character count: 39817

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 663-673
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6918

Sentiment Analysis of the Waste Problem based on YouTube comments using VADER and Deep Translator

Herman Yuliansyah[1,*], Surahma Asti Mulasari[2], Sulistyawati[2], Fanani Arief Ghozali[3], Bambang Sudarsono[4]

[1]Faculty of Industrial Technology, Department of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2]Faculty of Public Health, Department of Public Health, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[3]Faculty of Teacher Training and Education, Department of Electronics Engineering Vocational Education, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[4]Faculty of Teacher Training and Education, Department of Automotive Technology Vocational Education, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
Email: [1,*]herman.yuliansyah@tif.uad.ac.id, [2]surahma.mulasari@ikm.uad.ac.id, [3]sulistyawati.suyanto@ikm.uad.ac.id, [4]fanani.ghozali@pvte.uad.ac.id, [5]bambang.sudarsono@pvto.uad.ac.id
Correspondence Author Email: herman.yuliansyah@tif.uad.ac.id

Abstract—The waste problem is a severe problem that significantly affects the environment and public health. To effectively determine the public's perception of the waste problem, it is necessary to examine public sentiment toward waste management. This research aims to develop a sentiment analysis model using VADER and deep-translator and analyze the Yogyakarta waste emergency problem. This research was conducted in two phases, namely, the first phase was developing a sentiment analysis model by evaluating its performance based on public data. Then, the second phase classifies public comments from YouTube regarding the waste problem to understand public perceptions and evaluations by identifying positive, negative, and neutral sentiments. The model evaluation results show that sentiment analysis using VADER and deep translator can achieve Accuracy, Precision, Recall, and F1-score values of 0.716, 0.837, 0.853, and 0.738, respectively. The sentiment results from YouTube comments obtained positive, neutral, and negative sentiments of 30.0%, 31.7%, and 37.3%, respectively. The results of the sentiment analysis are neutral sentiment discussing waste management, disappointment in negative sentiment, and hope for waste management in positive sentiment.

Keywords: Waste Problem, Sentiment Analysis; Lexicon-Based; VADER; Deep-Translator

1. INTRODUCTION

The waste problem is a severe challenge faced by many countries worldwide, including Indonesia. Rapid population growth, urbanization, and lifestyle changes have increased the volume of waste produced [1]. To overcome this problem, a deep understanding of public sentiment and perceptions regarding waste is the key to formulating effective strategies and policies. Social media data provides information that has great potential for gaining insight into people's views regarding waste problems [2]. Sentiment analysis collects, processes, and assesses opinions or emotional evaluations of individuals or groups about a particular topic or entity [3]. Sentiment analysis aims to understand and measure general attitudes, sentiments, or feelings related to the analyzed subject.

The waste problem in Yogyakarta is a critical issue that requires immediate action to take quick steps to handle waste [4]. The high urban waste production in Yogyakarta is due to suboptimal waste management, especially in dealing with the enormous problem of waste production from various community activities [5]. Rapid population growth, tourists coming every year, and a lack of environmental awareness have also resulted in a drastic increase in waste produced daily. Several innovations that have been conducted to turn problems into opportunities include the waste bank, advanced recycling, innovative education, great collaboration, and the movement to process waste and rubbish with biopores [6]. Thus, serious efforts are needed from the government, society, and industry players to jointly solve this emergency waste problem, focusing on environmental education, improving waste management infrastructure, and promoting sustainable lifestyles.

Sentiment analysis has become a handy tool in various research conducted today. The sentiment analysis of the waste problem in Yogyakarta is hoped to show that many people feel concerned and worried about this situation. Many expressed concerns about the negative impacts that arise, such as environmental pollution and disruption to the tourism sector, which is the livelihood of many people in this city. However, some also expressed hope that concrete steps and solutions would soon be taken by the government and local communities.

Many studies have adopted sentiment analysis techniques to understand people's sentiments in various domains, such as social media [7], mental health [8], social crisis detection [9], climate change [10], restaurant reviews [11], response to the pandemic COVID-19 [12]–[14], product reviews [15], [16], Amazon reviews [17], movie reviews [17], airline reviews [18], student responses to teaching [19], [20], electric vehicle reviews [21] and more. In the social media domain, sentiment analysis helps researchers understand public opinion on contemporary issues and monitor changes in sentiment in real time. In environments such as climate change, this research helps understand how people perceive environmental change and its impacts [10]. Thus, sentiment analysis has become essential in understanding and responding to currently developing social issues.

Sentiment analysis has also played an essential role in industry and business, especially in understanding product and service reviews [15], [16]. Customer reviews on platforms like Amazon provide valuable insights into customer satisfaction and product quality [17]. Sentiment analysis in restaurant and airline service reviews helps

Submitted: 19/10/2023; Accepted: 31/01/2024; Published: 31/01/2024

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 663-673
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6918

companies improve the quality of their services based on customer feedback [11], [18]. In addition, sentiment analysis is also used in education to understand students' responses to teaching and curriculum [19], [20]. All this shows that sentiment analysis is a versatile and valuable tool used in various research and industrial contexts to understand people's feelings, opinions, and responses.

Two main approaches can conduct sentiment analysis: lexicon-based and machine learning/deep learning-based [10], [22]. The lexicon-based approach relies on a predefined dictionary of words and sentiment weights. This approach is advantageous when data resources are limited or when wanting to understand the specific sentiment toward certain words or phrases. The main advantage of the lexicon-based system is its ability to provide easy-to-understand interpretations and high readability for humans. However, this approach tends to be less adaptive in dealing with language variations and complex contexts.

On the other hand, machine learning/deep learning-based approaches use algorithms that can learn independently from data to identify sentiment patterns. This approach has the advantage of handling diverse languages and contexts. However, the disadvantage is that it requires a large training dataset, and the training process is time-consuming.

While machine learning/deep learning-based approaches have advantages in dealing with language variations and complex contexts, lexicon-based has a clarity and speed of interpretation that makes it especially useful in specific sentiment analysis or when time is a crucial factor. Meanwhile, machine learning/deep learning-based approaches require more significant resources and time for training, which may only sometimes be available in certain situations. With a lexicon-based system, we can quickly understand the sentiment in the text without having to involve a complicated training process.

There are several lexicon-based sentiment analysis methods, including SentiWordNet [23], Affective Norms for English Words (AFINN) [24]–[26], Valence Aware Dictionary and sEntiment Reasoner (VADER) [27], [28], TextBlob[29], Pattern [30], dan NRC Emotion Lexicon [31]. SentiWordNet is a sentiment-based dictionary that assigns positive, negative, and neutral sentiment scores to each word in its dictionary. SentiWordNet calculates the overall sentiment of the text by balancing the scores of the terms that appear in the text. AFINN is an English sentiment dictionary that assigns a sentiment score on a scale from -5 (negative) to +5 (positive) for each word [25]. AFINN calculates the overall sentiment of the text by adding up the scores of the terms found. Later, VADER is a very sophisticated and context-based sentiment dictionary. This algorithm considers punctuation, case, and word context to calculate the overall sentiment of the text. TextBlob is a Python library that uses lexicon-based and machine-learning methods. It utilizes a dictionary of words with emotion and can perform grammatical analysis, such as multiple phrase identification. Pattern is another Python library that supports lexicon-based sentiment analysis. It can recognize sentiment in text and provides various other text-processing features. Lastly, the NRC Emotion Lexicon focuses on sentiment-related emotions like anger, happiness, sadness, and others. Identify emotional aspects in the text that can use the NRC Emotion Lexicon.

This research uses VADER because of VADER's ability to overcome complex variations in context and language and recognize subtle sentiments in text. The VADER is proposed by C. Hutto and E. Gilbert [27] in 2014. VADER utilizes a lexicon containing preassigned sentiment values for words. Each word is assigned a sentiment score, indicating its level of positivity or negativity. The creation of this lexicon involved a manual process of gathering and annotating words by human evaluators. Each word in the VADER lexicon is assigned a sentiment polarization score, ranging from -1 (indicating negativity) to +1 (indicating positivity), with a neutral score hovering around 0. To address issues of intensity and exception, VADER considers the weight of words accompanying sentiment words. For instance, "excellent" receives a higher positive score than "good." The overall sentiment score for a text is computed by amalgamating the scores of individual words, factoring in intensity and exception. VADER also takes into consideration the interrelation between words in a sentence. The algorithm employs heuristic rules to manage instances like negation (words conveying opposing sentiment) and contrast (shifts in sentiment within the same context). VADER assigns importance to punctuation and capitalization, where punctuations like exclamation marks (!) can elevate positive scores, and uppercase letters may signify emotional intensity. Furthermore, VADER acknowledges the contextual impact in sentiment assessment; for instance, using positive words following negative ones can alter the interpretation of the overall sentiment. VADER's high accuracy can be an excellent choice for sentiment analysis research. This advantage makes VADER a popular and effective lexicon-based algorithm for various sentiment analysis research purposes. In this research, the model also added a language translator with a deep translator because the dictionary used by VADER is English.

The main objective of this research is to develop a sentiment analysis model using VADER and deep translator by testing it based on public datasets. Furthermore, the proposed sentiment analysis mode is used to conduct sentiment analysis based on social media data in the form of YouTube comments from videos discussing the Yogyakarta waste emergency problem to understand the public's perception and evaluation of the waste problem in the form of identifying positive, negative and neutral sentiments. The two contributions made in this research are the sentiment analysis model using VADER with a deep translator. The analysis results in positive, negative, and neutral sentiments and visualization of comments for each emotion.

The remainder of this research is structured by defining the research methods carried out in Section 2, presenting the results of the evaluation of sentiment analysis models from public data, the results of sentiment

Submitted: 19/10/2023; Accepted: 31/01/2024; Published: 31/01/2024

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 663-673
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6918

analysis, and word cloud visualization based on the data examined in Section 3. This research is concluded with the results of sentiment analysis in Section 4.

2. RESEARCH METHODOLOGY

2.1 Research Stages

This research was designed with two stages: phase 1 and phase 2. The first phase was developing a sentiment analysis model using VADER and Deep-Translator based on a public dataset, and the second phase was conducting sentiment analysis based on a dataset of YouTube comments. The stages in phase 1 aim to find out the accuracy, precision, recall, and F1-Score values of the proposed sentiment analysis model and apply this model to phase 2. In more detail, the stages of this research are shown in Figure 1.

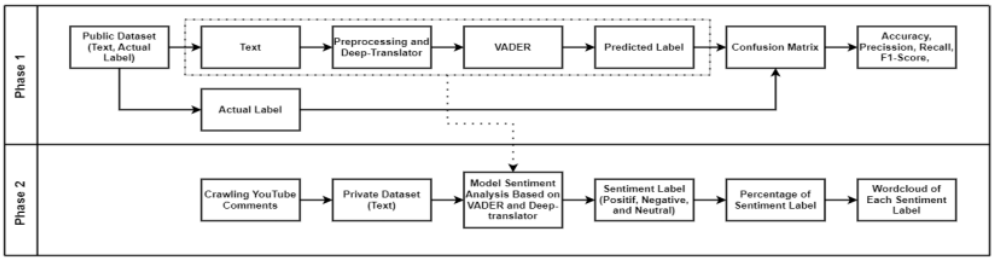

Figure 1. Research Design

In phase 1, the dataset consists of several data attributes. However, only two data attributes are used to evaluate the sentiment analysis model: the text attribute and the label attribute. The text attribute is preprocessed by changing the text to lower and removing mentions, URLs, and hashtags. Furthermore, the text that has been cleaned is translated into English for texts that contain language other than English. VADER is used to classify text into sentiment based on compound values according to variations in sentiment from the actual label. C. Hutto and E. Gilbert [27] determined compound values for three types of sentiment, namely positive, negative, and neutral, based on equation (1).

$$\text{Compound Value (CV)} = \begin{cases} \text{positive}, & \text{CV} \geq 0.05 \\ \text{neutral}, & -0.05 < \text{CV} < 0.05 \\ \text{negative}, & \text{CV} \leq -0.05 \end{cases} \quad (1)$$

Meanwhile, the compound value is adjusted to equation (2) for variations in two sentiments, namely positive and negative.

$$\text{Compound Value (CV)} = \begin{cases} \text{positive}, & \text{CV} \geq 0 \\ \text{neutral}, & \text{CV} < 0 \end{cases} \quad (2)$$

This label attribute is referred to as the actual label, which is used for evaluation with the confusion matrix by comparing the results to the predicted label. Next, the sentiment analysis model examined in Phase 1 was used to analyze the sentiment dataset related to the waste problem in Yogyakarta in Phase 2. The dataset examined in phase 2 was crawled based on YouTube comments. The result of phase 2 is a classification of comment text divided into three sentiments: positive, negative, and neutral. Then, the classification results are visualized in WordCloud to see the words that frequently appear in each classification result.

2.2 Dataset

Phase 1 of this research uses a public dataset obtained from Kaggle, shown in Table 1. The Sentiment140 dataset is small and imbalanced, while the IMDB dataset is large but balanced. These two datasets were chosen because they represent different data characters even though the data context is still similar, namely tweets and reviews. In phase 2, a dataset based on YouTube comments was used crawled based on the URL shown in Table 2. This YouTube URL was obtained based on search results related to "Jogja Waste Emergency (Jogja Darurat Sampah)," and 464 comments were obtained. The dataset obtained based on Table 2 is in the form of comment text and is then analyzed based on the sentiment analysis model proposed in Phase 1.

Table 1. Public Dataset Dataset

Dataset	Number of Text	Attribut	Label	Percentage Label
Sentiment140	497	{Polarity, Time, Type, User, Tweet}	{positive, neutral, negative}	{positive: 36.4%, neutral: 35.6%, negative: 28.0%}

Submitted: 19/10/2023; Accepted: 31/01/2024; Published: 31/01/2024

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 663-673
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6918

Dataset	Number of Text	Attribut	Label	Percentage Label
IMDB	40000	{Text, Label}	{positive, negative}	{positive:50%, negative:50%}

Table 2. Private Dataset Dataset

#	YouTube URL
1.	https://www.youtube.com/watch?v=JAmjAsiZ2Z8
2.	https://www.youtube.com/watch?v=iamXWPUyENs
3.	https://www.youtube.com/watch?v=tiA1kx_nh9M
4.	https://www.youtube.com/watch?v=yTaBfKiMdp8
5.	https://www.youtube.com/watch?v=J67bhP4uLHY
6.	https://www.youtube.com/watch?v=g2o1AIjYl3o
7.	https://www.youtube.com/watch?v=PSnkQuVw608
8.	https://www.youtube.com/watch?v=cKQjzjmqgr8
9.	https://www.youtube.com/watch?v=qccg19p723E
10.	https://www.youtube.com/watch?v=vFj4eu2Sq6g
11.	https://www.youtube.com/watch?v=kANTxJtgCSw
12.	https://www.youtube.com/watch?v=FfMN7Vj2UOM
13.	https://www.youtube.com/watch?v=-rbfGM-_3JA
14.	https://www.youtube.com/watch?v=qVbDHm-ogxw
15.	https://www.youtube.com/watch?v=0UAWWqx2AF8
16.	https://www.youtube.com/watch?v=-6eA-ePgsiU
17.	https://www.youtube.com/watch?v=u01bnoN0el0
18.	https://www.youtube.com/watch?v=PAuEWVlCyA4

3. RESULT AND DISCUSSION

This research has high urgency due to urgent environmental problems, the potential of social media data, and better decision-making. Waste is a pressing environmental problem that negatively impacts ecosystems and human health. Understanding public sentiment towards this problem will help formulate appropriate solutions for sustainable waste management. Social media has become the leading platform for people to share their opinions and experiences about waste. Sentiment analysis using social media data can provide valuable insights into understanding people's mindsets and preferences regarding waste problems. By understanding community sentiment, stakeholders can make better decisions in policy formulation and implementation of waste management programs. This information can help increase the effectiveness of addressing waste problems. The results and discussion section is divided into two: the evaluation results of the proposed sentiment analysis model and the implementation of the proposed sentiment analysis model to obtain sentiment analysis results based on a dataset obtained from YouTube comments.

3.1 Model Sentiment Analysis using VADER and Deep-Translator

The proposed sentiment analysis model is shown in Figure 1 regarding phase 1, then the program code implementation in Python is obtained as shown in Figure 2-7. The sentiment140 dataset was obtained from the keitozoumana GitHub URL (https://github.com/keitazoumana/VADER_sentiment-Analysis). Keitozoumana also published the VADER sentiment analysis program code at URL [32]. Keitozoumana's sentiment analysis model was used as the basis for this research to be further developed by improving it on a deep translator. The IMDB dataset was obtained from Kaggle with the URL https://www.kaggle.com/datasets/yasserh/imdb-movie-ratings-sentiment-analysis as a comparison dataset. Figure 2 and Figure 3 is the program code to read these two datasets, stored in files with the CSV extension.

```
In [ ]: import pandas as pd

        # Read the data set
        data_url = "https://raw.githubusercontent.com/keitazoumana/VADER_sentiment-Analysis/main/data/testdata.manual.2009.06.14.csv"
        data = pd.read_csv(data_url)

        data.head(3)
```

Figure 2. Read dataset Sentiment140

```
In [ ]: import pandas as pd

        # Read the data set
        data_url = 'dataset/data imdb.csv'
        data = pd.read_csv(data_url)
        data.sample(5)
```

Figure 3. Read dataset IMDB

Submitted: 19/10/2023; Accepted: 31/01/2024; Published: 31/01/2024

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 663-673
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6918

The first preprocessing stage after reading the dataset is label transformation. In the Sentiment140 dataset, labels are symbolized by three integers {0,2,4} and transformed into positive, neutral, and negative strings. Furthermore, for the IMDB dataset, there are two integers {0, 1} where these labels are transformed into positive and negative label strings. This label is needed in the model evaluation process with the confusion matrix.

```
In [ ]: def format_data(data):

            last_col = str(data.columns[-1])
            first_col = str(data.columns[0])

            data.rename(columns = {last_col: 'tweet_text', first_col: 'polarity'}, inplace=True)

            # Change 0, 2, 4 to negative, neutral and positive
            labels = {0: 'negative', 2: 'neutral', 4: 'positive'}
            data['polarity'] = data['polarity'].map(labels)

            # Get only the two columns
            return data[['tweet_text', 'polarity']]

        # Apply the transformation
        data = format_data(sentiment_data)
        data.head(3)
```

Figure 4. Change Label Polarity

After the label transformation process, it continues with the cleaning process. Because VADER considers punctuation and word context to calculate the overall sentiment of the text, the cleaning process is conducted with just four things: changing the text to lower, removing mentions, removing URLs, and removing hashtags. The text cleaning process is performed as shown in Figure 5.

```
In [ ]: import re
        import pandas as pd

        def clean(text):
            text = str(text)
            text = text.lower() #Lower
            text = re.sub(r'@\w+', '', text) # Replace mentions with an empty string
            text = re.sub(r'https?://\S+|www\.\S+|t\.co/\S+|#\S+', '', text) # Remove URLs
            text = re.sub(r'#\w+', '', text) #remove hastag
            return text

        data['tweet_text'] = data['tweet_text'].apply(clean)
        data
```

Figure 5. Text Cleaning

VADER is a text classification for sentiment analysis that uses English, so a translation process is needed for the text, which needs to calculate the compound value and then determine the polarity of the text into positive, negative, or neutral sentiment. This translation process is conducted by installing Deep Translator and then importing Google Translator. Each checked text can be translated into English based on the language source with the auto parameter. This auto parameter is intended to adjust the text's source language, as shown in Figure 6. This translation is conducted by calling the translation method based on GoogleTranslator.

```
In [ ]: !pip install deep-translator
```

```
In [ ]: from deep_translator import GoogleTranslator

        df['tweet_text'] = df['tweet_text'].apply(lambda x: GoogleTranslator(source='auto', target='english').translate(x))
        df
```

Figure 6. Translation Process

The main stage in the sentiment analysis process is text classification. In this text classification, the aim is to classify text by calculating compound values. The positive, neutral, and negative sentiment results are calculated based on equations (1) and (2). This text classification process is conducted by installing and importing NLTK and downloading the VADER lexicon. An instance of SentimentIntensityAnalyzer is created to calculate the polarity score. The polarity score result is a value consisting of four categories: positive value, negative value, neutral value, and compound value. This compound value becomes a reference value to be transformed into positive, negative, and neutral sentiment. The text classification process for sentiment analysis is conducted in Figure 7.

```
In [ ]: # Install and import nltk
        !pip install nltk
        import nltk

        # Download the lexicon
        nltk.download("vader_lexicon")
```

```
In [ ]: # Import the lexicon
        from nltk.sentiment.vader import SentimentIntensityAnalyzer

        # Create an instance of SentimentIntensityAnalyzer
        sent_analyzer = SentimentIntensityAnalyzer()
```


Submitted: 19/10/2023; Accepted: 31/01/2024; Published: 31/01/2024

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 663-673
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6918

```
In [ ]: def format_output(output_dict):

            polarity = "neutral"

            if(output_dict['compound']>= 0.05):
                polarity = "positive"

            elif(output_dict['compound']<= -0.05):
                polarity = "negative"

            return polarity

        def predict_sentiment(text):

            output_dict =  sent_analyzer.polarity_scores(text)
            return format_output(output_dict)

        # Run the predictions
        data["vader_prediction"] = data["tweet_text"].apply(predict_sentiment)

        # Show 5 random rows of the data
        data.sample(5)
```

Figure 7. Text Classification Process

Evaluation results are conducted with the confusion matrix on two datasets, namely Sentiment140 and IMDB. Four evaluation measures are Accuracy, Precision, Recall, and F1-score. This value is considered suitable for making a sentiment analysis model because it obtains an accuracy and f1-score value of more than 70%. Furthermore, the precision and recall values reached more than 80%. This sentiment analysis model performed best on the Sentiment140 dataset, as shown in Table 3.

Table 3. Confusion Matrix Results

Dataset	Accuracy	Precision	Recall	F1-score
Sentiment140	0.716	0.837	0.807	0.734
IMDB	0.698	0.788	0.853	0.738

3.2 Sentiment Analysis and Visualization

The sentiment analysis model proposed in Phase 1 examined the dataset from YouTube comments. Based on this model's results, an even sentiment distribution is obtained, as shown in Figure 8. However, negative sentiment is the most common, followed by neutral and positive sentiment.

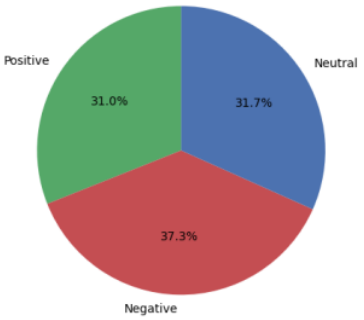

Figure 8. Percentage of Predicted Sentiment Analysis

This polarization of sentiment is evenly spread, making it attractive to discuss. Based on the compound value, a candlestick with a comprehensive standard deviation gap is obtained, as shown in Figure 9. The most considerable compound value of the positive comments examined is 0.944, which states that

"In fact, there is a solution for everything, there is just the will and commitment to hold the budget, and the money also comes from the community. I hope that the councils who are in control of the budget will think about buying a waste incinerator device, which is only zero point zero a certain percentage of the price (as long as there is no mark up on the budget), thankfully each RW will be given a grant of this equipment, so that it can be more useful. better yet, add waste processing equipment too, along with land belonging to the district/provincial government in each district/city. because if a citizen wants to buy such equipment, they also require various permits which of course in this country require an 'envelope' and are complicated. People, yes, there are many campuses with a lot of research, provide space and funding for research results to be produced, so that the waste in this country since ancient times has a solution, and isn't always a mess."

Meanwhile, the smallest compound value of the negative comments examined was -0.987, which states that

Submitted: 19/10/2023; Accepted: 31/01/2024; Published: 31/01/2024

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 663-673
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6918

"If you don't know what the problem is, at least research it first, it has been explained there that people throw rubbish on the street because the rubbish dump is closed. Don't pay the dues and throw it on the bridge? Wow, this is ridiculous again, where I live there is a garbage charge every month, but I dispose of the trash myself, there's no one from the government to collect it. The point is that the Jogja city government is incompetent in dealing with this waste crisis, Jogja is a densely built city, if it is told to independently process what is there, it will make noise among the community because burning smoke is everywhere (this has happened before). So where do people throw their rubbish if everything is closed? cook waste stored at home? The existing ones will set and smell if stored for a long time. The point is that the people of Jogja are actually good (it can be seen that before the ban on throwing rubbish there was no rubbish on the side of the road). Now they are confused about where to throw it away, it's stored smelly. OK, just throw it away in an empty place. This is purely the government's fault."

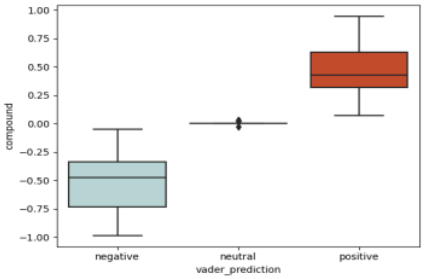

Figure 9. Text Classification Process

A visualization of the terms that frequently appear from each sentiment is shown in Figure 10-12. In neutral sentiment, the words that appear more dominant than other words are "governor," "rubbish," "Jogja," and "people." These words show the impression of the people of Jogja towards the governor regarding rubbish management. Furthermore, for negative sentiment, the words that appeared more frequently were the words "rubbish," "waste," "problem," and "people." These words show that "rubbish" and "waste" are a "problem" for "people." Later, for this positive sentiment, there are the words "solution," "waste" and "rubbish." These words show a discussion regarding solutions to overcome "waste" and "rubbish." For more details, these comments are presented in Table 6-8.

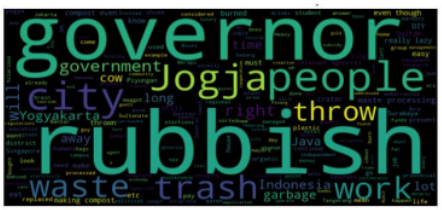

Figure 10. WordCloud Neutral Sentiment

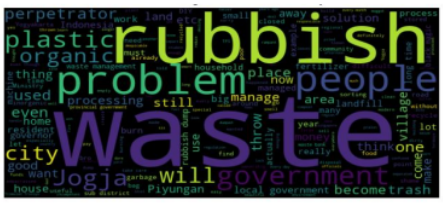

Figure 11. WordCloud Negative Sentiment

Figure 12. WordCloud Positive Sentiment

Submitted: 19/10/2023; Accepted: 31/01/2024; Published: 31/01/2024

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 663-673
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6918

Several comments that can be highlighted as neutral sentiment are more inclined toward waste management, as shown in Table 6. Several words hope that there will be community awareness regarding waste processing using incinerators, softeners, or crushing machines to be used as waste generators and plant fertilizer. This hope also raises the potential to bring in investors to this business.

Table 6. Neutral Sentiment

English Text	Compound
"creative economy" should be able to turn waste into gold...	0.0258
Ideal waste management means that every district/city has at least 1 integrated processing facility, more than 1 waste processing system is better, as well as community cooperation for awareness about independent waste processing.	0
managed with an incinerator. the term burnt can still be managed	0
can be burned to generate electricity	0
Actually, can it be burned or not? Have you ever heard that abroad there are softening or crushing machines like that?	0
why not make a pltsa (garbage power plant)	0
I haven't burned the ashes to make fertilizer for joss plants... instead of piling up for days it will cause disease	0
build a waste processing factory that satisfies the boss, look for investors.	0

Waste management is an important issue, and the creative economy has great potential to turn waste into valuable resources. Ideally, every district or city should have at least one integrated waste processing facility, even better if there is more than one. Collaboration between the government community and awareness of independent waste processing is critical. One approach is to use an incinerator, where the waste is burned, and the resulting energy can be used to generate electricity. However, the question is whether there are other methods besides burning. Several countries have developed softening or crushing machines to manage waste. A more sustainable solution is to consider a waste power plant (PLTSA), which can convert waste into energy. And, of course, think about innovative measures such as turning the remaining ash from burning into valuable plant fertilizer. With adequate investment and waste processing plants, we can solve the waste problem while creating sustainable economic opportunities.

Table 7. Sentiment Negative

Later, in Table 8, there are expectations from people regarding waste management. This hope is realized in the form of waste management and recycling regulations. Hope for accelerating the resolution of the waste problem and improving it to keep the environment clean. Garbage and waste are an integral part of the life cycle. However, in dealing with this problem, Indonesia's waste management often relies on traditional methods, which are no longer effective. With population growth increasing, waste management must keep up with more advanced technological developments. Many developed countries have successfully overcome the waste problem with excellent and efficient management. The government should show serious intentions in dealing with the waste problem by setting effective regulations, providing the necessary facilities and infrastructure, and using advanced technology, such as smelting machines, to produce valuable products for everyday life. Preparing a budget that suits your needs is also required, and experts and educational institutions design the right solution. With these steps, the waste problem in this city can be resolved quickly, maintain the city's beauty, and provide other benefits such as the production of compost fertilizer and environmentally friendly industrial processing of non-organic materials.

Table 8. Sentiment Positive

English Text	Compound
It has been explained there that people throw rubbish on the street because the rubbish dump is closed. don't pay the dues and throw it on the bridge? Wow, this is ridiculous again, where I live there is a garbage charge every month, but I dispose of the trash myself, there's no one from the government to collect it. The point is that the Jogja city government is incompetent in dealing with this waste crisis, Jogja is a densely built city, if it is told to independently process what is there, it will make noise among the community because burning smoke is everywhere (this has happened before). So where do people throw their rubbish if everything is closed? cook waste stored at home? The existing ones will set and smell if stored for a long time. The point is that the people of Jogja are actually good (it can be seen that before the ban on throwing rubbish there was no rubbish on the side of the road). Now they are confused about where to throw it away, it's stored smelly. OK, just throw it away in an empty place. This is purely the government's fault. critical.....	-0.986

Submitted: 19/10/2023; Accepted: 31/01/2024; Published: 31/01/2024

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 663-673
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6918

English Text	Compound
There are many problems in this country, one of which is waste but it is not managed efficiently by the regional and central governments. will become a time bomb throughout Indonesia in the future. Let's see how serious management is from the small ones or the big ones. a definite impact on society twice more, especially small and poor communities. The number 1 leader of the village school must think from small things to big things because the government is a servant, not to be served.	-0.9158
Every household should manage the waste they produce themselves, not just throw it away in a landfill, I feel sorry for the waste collectors and dump sites, whose waste is it anyway????	-0.9054
Waste is an industrial opportunity...where is this government...economic potential is not always the beautiful things that have been managed only for the beautiful and productive ones...but forget to process the impact of beautiful development and growth.	-0.8895
The government should have a policy that waste that is difficult to decompose should be reduced or even eliminated if it is piled up in piles, think long-term, the longer it goes, the more you will feel sorry for our children and grandchildren... we will inherit the waste management... don't act if there are victims, don't act If it's not viral yet... think ahead	-0.8658
The impact is that the burden on the environment increases, as well as health threats. The worry is that if a landslide occurs it will cause casualties, possibly loss of life, or property.	-0.8316
Why aren't Indonesia trying to build a waste melting plant? So build a closed warehouse to reduce unpleasant odor pollution, and a smelter that can turn waste into dust.	-0.8271
closing rubbish dumps is ridiculous. There is rubbish every day, if the government had thought about it this would not have happened. The governor should issue a gubernatorial regulation that each region must have its own TPS because the biggest contributor of waste is Sleman, even though Sleman has the largest area but doesn't have TPS and still dumps it in Piyungan. Why is the DIY government slow???	-0.822
It's time for the DIY provincial government to have integrated waste technology. The population is increasing and land for final waste disposal is limited	-0.7579
The real problem is that this waste is mixed (organic/non-organic). If it were separated, it might be easier to turn it into recycled material or even fertilizer, automatically the amount of waste wouldn't be as much.	-0.7506
trash... waste... that's the chain of life. If the solution continues to be done using traditional methods,... while life is becoming more and more advanced, the population is increasing... how come waste management seems to be stagnant/running in place. even though the era is high technology... many city-states are developing rapidly, but waste has been regulated, and the end result is good, clean, no problem. The government seems to see that this waste is not/is not yet optimal.... shows the intention to deal with waste with regulations, but the facilities must be made easier to procure, and without using machines, equipment, which is effective and efficient by using melting machines, which produce more useful downstream products , for life. create a budget that is appropriate/measurable to your needs.. involve experts and educational institutions to design it.	0.8741
Hopefully it will be resolved quickly, so that my hometown remains beautiful	0.8074
Piyungan tpst can be a mine for worm protein, maggots & organic compost, and industrial scale non-organic enumeration. I hope you are in good health.	0.7003
If the Piyungan waste allocation is looking for another allocation in the vast Kulonprogo area, what do you do, Mr. Municipal Government? But try to break it down neatly and not pollute the surrounding environment.	0.6959
Come on, residents of Jogyakarta must be able to clean up so that Jogyakarta remains clean and its citizens healthy..	0.6597
just make it into compost or something else, friend	0.4939
If the waste is processed and recycled, not collected or thrown away, Indonesian people should have self-awareness and not just blame each other on related parties.	0.2349

4. CONCLUSION

This research proposed a sentiment analysis model using VADER and deep-translator. The evaluation results shown the model has good performance and can be used to analyze sentiment from YouTube comments regarding the waste problem in Yogyakarta. The positive, negative, and neutral sentiment analysis results shown that people's opinions about waste management vary. Several processing technologies, such as incinerators and waste softening machines, are expected to be supported by each district/city and must also have integrated waste management. Negative sentiment is related to disappointment with waste management and separation. The waste

Submitted: 19/10/2023; Accepted: 31/01/2024; Published: 31/01/2024

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 663-673
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6918

problem is worrying and is compounded by many residents throwing rubbish carelessly. The government must enforce effective regulations, use advanced technology, and appropriate budgets to solve the waste problem.

ACKNOWLEDGMENT

The authors thanks to Universitas Ahmad Dahlan, Indonesia for support the research.

REFERENCES

[1] L. M. Heidbreder, I. Bablok, S. Drews, and C. Menzel, "Tackling the plastic problem: A review on perceptions, behaviors, and interventions," Sci. Total Environ., vol. 668, pp. 1077–1093, Jun. 2019, doi: 10.1016/j.scitotenv.2019.02.437.
[2] M. Liu, X. Luo, and W.-Z. Lu, "Public perceptions of environmental, social, and governance (ESG) based on social media data: Evidence from China," J. Clean. Prod., vol. 387, p. 135840, Feb. 2023, doi: 10.1016/j.jclepro.2022.135840.
[3] P. Jiang, J. Zhou, Y. Van Fan, J. J. Klemeš, M. Zheng, and P. S. Varbanov, "Data analysis of resident engagement and sentiments in social media enables better household waste segregation and recycling," J. Clean. Prod., vol. 319, p. 128809, Oct. 2021, doi: 10.1016/j.jclepro.2021.128809.
[4] P. D. I. Yogyakarta, "DIY Terus Upayakan Perbaikan Pengelolaan Sampah." https://jogjaprov.go.id/berita/diy-terus-upayakan-perbaikan-pengelolaan-sampah (accessed Oct. 17, 2023).
[5] D. D. I. Yogyakarta, "Ketua Komisi A Ingatkan Pentingnya Selesaikan Masalah Sampah Perkotaan dari Hulu." https://www.dprd-diy.go.id/ketua-komisi-a-ingatkan-pentingnya-selesaikan-masalah-sampah-perkotaan-dari-hulu/ (accessed Oct. 17, 2023).
[6] B. K. dan P. S. D. Manusia, "Mengatasi Darurat Sampah Di Kota Yogyakarta Kreativitas Menuju Kebersihan Yang Memikat Hati." https://bkpsdm.jogjakota.go.id/detail/index/28545 (accessed Oct. 17, 2023).
[7] S. Rahman, N. Jahan, F. Sadia, and I. Mahmud, "Social crisis detection using Twitter based text mining-a machine learning approach," Bull. Electr. Eng. Informatics, vol. 12, no. 2, pp. 1069–1077, Apr. 2023, doi: 10.11591/eei.v12i2.3957.
[8] N. Khalid, S. Abdul-Rahman, W. Wibowo, N. A. S. Abdullah, and S. Mutalib, "Leveraging social media data using latent dirichlet allocation and naïve bayes for mental health sentiment analytics on Covid-19 pandemic," Int. J. Adv. Intell. Informatics, vol. 9, no. 3, 2023, doi: https://doi.org/10.26555/ijain.v9i3.1367.
[9] A. Samih, A. Ghadi, and A. Fennan, "Enhanced sentiment analysis based on improved word embeddings and XGboost," Int. J. Electr. Comput. Eng., vol. 13, no. 2, p. 1827, Apr. 2023, doi: 10.11591/ijece.v13i2.pp1827-1836.
[10] N. Mohamad Sham and A. Mohamed, "Climate Change Sentiment Analysis Using Lexicon, Machine Learning and Hybrid Approaches," Sustainability, vol. 14, no. 8, p. 4723, Apr. 2022, doi: 10.3390/su14084723.
[11] N. R. Bhowmik, M. Arifuzzaman, M. R. H. Mondal, and M. S. Islam, "Bangla Text Sentiment Analysis Using Supervised Machine Learning with Extended Lexicon Dictionary," Nat. Lang. Process. Res., vol. 1, no. 3–4, p. 34, 2021, doi: 10.2991/nlpr.d.210316.001.
[12] M. Mujahid et al., "Sentiment Analysis and Topic Modeling on Tweets about Online Education during COVID-19," Appl. Sci., vol. 11, no. 18, p. 8438, Sep. 2021, doi: 10.3390/app11188438.
[13] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, and G. S. Choi, "A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis," PLoS One, vol. 16, no. 2, p. e0245909, Feb. 2021, doi: 10.1371/journal.pone.0245909.
[14] A. A. Reshi et al., "COVID-19 Vaccination-Related Sentiments Analysis: A Case Study Using Worldwide Twitter Dataset," Healthcare, vol. 10, no. 3, p. 411, Feb. 2022, doi: 10.3390/healthcare10030411.
[15] H. Zhao, Z. Liu, X. Yao, and Q. Yang, "A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach," Inf. Process. Manag., vol. 58, no. 5, p. 102656, Sep. 2021, doi: 10.1016/j.ipm.2021.102656.
[16] R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, "Sentiment Analysis on Product Reviews Using Machine Learning Techniques," 2019, pp. 639–647. doi: 10.1007/978-981-13-0617-4_61.
[17] S. A. S. Neshan and R. Akbari, "A Combination of Machine Learning and Lexicon Based Techniques for Sentiment Analysis," in 2020 6th International Conference on Web Research (ICWR), Apr. 2020, pp. 8–14. doi: 10.1109/ICWR49608.2020.9122298.
[18] A. M. Rahat, A. Kahir, and A. K. M. Masum, "Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset," in 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), Nov. 2019, pp. 266–270. doi: 10.1109/SMART46866.2019.9117512.
[19] A. T. Mahmood, S. S. Kamaruddin, R. K. Naser, and M. M. Nadzir, "A Combination of Lexicon and Machine Learning Approaches for Sentiment Analysis on Facebook," J. Syst. Manag. Sci., vol. 10, no. 3, pp. 140–150, Sep. 2020, doi: 10.33168/JSMS.2020.0310.
[20] Z. Nasim, Q. Rajput, and S. Haider, "Sentiment analysis of student feedback using machine learning and lexicon based approaches," in 2017 International Conference on Research and Innovation in Information Systems (ICRIIS), Jul. 2017, pp. 1–6. doi: 10.1109/ICRIIS.2017.8002475.
[21] Y. Pratama, D. T. Murdiansyah, and K. M. Lhaksmana, "Analisis Sentimen Kendaraan Listrik Pada Media Sosial Twitter Menggunakan Algoritma Logistic Regression dan Principal Component Analysis," J. Media Inform. Budidarma, vol. 7, no. 1, pp. 529–535, 2023, doi: http://dx.doi.org/10.30865/mib.v7i1.5575.
[22] Q. He, "Recent Works for Sentiment Analysis using Machine Learning and Lexicon Based Approaches," in 2022 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), Apr. 2022, pp. 422–426. doi: 10.1109/AEMCSE55572.2022.00090.
[23] V. A. Rao, K. Anuranjana, and R. Mamidi, "A Sentiwordnet Strategy for Curriculum Learning in Sentiment Analysis," 2020, pp. 170–178. doi: 10.1007/978-3-030-51310-8_16.

Submitted: **19/10/2023**; Accepted: **31/01/2024**; Published: **31/01/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 663-673
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6918

[24] F. ˚Arup Nielsen, "Afinn Project." https://www2.imm.dtu.dk/pubdb/edoc/imm6975.pdf (accessed Oct. 17, 2023).
[25] F. Å. Nielsen, "AFINN Sentiment Lexicon." AFINN Sentiment Lexicon (accessed Oct. 17, 2023).
[26] F. Å. Nielsen, "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs," in Proceedings of the ESWC2011 Workshop on "Making Sense of Microposts": Big things come in small packages 718 in CEUR Workshop Proceedings, 2011, pp. 93–98. doi: https://doi.org/10.48550/arXiv.1103.2903.
[27] C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," in Proceedings of the International AAAI Conference on Web and Social Media, May 2014, vol. 8, no. 1, pp. 216–225. doi: 10.1609/icwsm.v8i1.14550.
[28] A. Amin, I. Hossain, A. Akther, and K. M. Alam, "Bengali vader: A sentiment analysis approach using modified vader," in 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 2019, pp. 1–6.
[29] S. Loria and others, "textblob Documentation," Release 0.15, vol. 2, no. 8, p. 269, 2018.
[30] T. De Smedt and W. Daelemans, "Pattern for python," J. Mach. Learn. Res., vol. 13, no. 1, pp. 2063–2067, 2012.
[31] S. M. Mohammad and P. D. Turney, "Nrc emotion lexicon," Natl. Res. Counc. Canada, vol. 2, p. 234, 2013.
[32] Z. Keita, "Social Media Sentiment Analysis In Python With VADER — No Training Required!," 2022. https://towardsdatascience.com/social-media-sentiment-analysis-in-python-with-vader-no-training-required-4bc6a21e87b8 (accessed Aug. 01, 2023).

Submitted: **19/10/2023**; Accepted: **31/01/2024**; Published: **31/01/2024**

HASIL CEK_Sentiment_Analysis_of_the_Waste_Problem

ORIGINALITY REPORT

13%	13%	6%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Kristen Satya Wacana Student Paper	6%
2	www.stmik-budidarma.ac.id Internet Source	5%
3	stmik-budidarma.ac.id Internet Source	2%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		