INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN PROGRESSIVE EDUCATION & DEVELOPMENT

OPEN ACCESS JOURNAL

Vol 14, Issue 4, (2025) E-ISSN: 2226-6348

Sentiment Analysis of Social Media Reactions to Indonesia's Free School Lunch Program Using Machine Learning Techniques

Rizki Syaputra
Master in Data Science, Faculty of Computing, Universiti Teknologi Malaysia, Malaysia
Email: syaputra@graduate.utm.my

Abdul Azzam Ajhari
Department of Informatics, University Cyber Asia,Jakarta, Indonesia
Email: abdulazzam@lecturer.unsia.ac.id

Liao Xiuya
Master in Data Science, Faculty of Computing, Universiti Teknologi Malaysia, Malaysia
Email: liaoxiuya@graduate.utm.my

Li Jing
Phd, Faculty of Computing, Universiti Teknologi Malaysia, Malaysia
Email: jingli@zuaa.zju.edu.cn

Prof. Madya. Ts. Mohd Shahizan Bin Othman
Lecturer, Faculty of Computing, Universiti Teknologi Malaysia, Malaysia
Email: shahizan@utm.my

DOI Link: http://dx.doi.org/10.6007/IJARPED/v14-i4/26681

Published Online: 06 October 2025

Abstract

The purpose of this study is to analysis public sentiment towards the free school meal program organized by the Indonesian Government using data from social media. This program aims to improve students' nutritional intake and help ease the burden on low-income families. However, this policy has drawn mixed reactions on social media, both in the form of support and criticism. Data was collected from various social media platforms and processed through data cleaning and modelling stages using machine learning algorithms, namely Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naive Bayes. The results of the analysis show that the SVM and Naive Bayes methods have a higher level of accuracy in classifying positive and negative sentiments than KNN. Overall, the majority of public sentiment towards this program is positive, reflecting support for the policy that is considered beneficial for students in need. However, there are also a number of negative sentiments that highlight issues of distribution and quality of service. These findings can

provide insights for the government to evaluate and improve the effectiveness of the program, as well as identify areas that need improvement.
Keywords: Sentiment Analysis, Social Media, Free Meal Program, Indonesian Government, Public Reaction

Introduction

Social media has grown in popularity and effectiveness in recent years as a means of rapidly disseminating information. A large number of individuals utilize it effectively, particularly in Indonesia. A lot of individuals oppose government initiatives on social media. The Free Meal Program has been the most talked about program and a trending issue on different social media platforms since the presidential election debate, the presidential election process, and the inauguration of Gibran Rakabumi Raka and Mr. Prabowo Subianto as Indonesia's eighth president in 2024. There have been mixed reactions to this program on social media; some have been complimentary, while others have been hostile or cynical. The Python programming language, Google Collab, and machine learning are used in this work to apply sentiment analysis approaches. Finding trends in the public's perception of this program is the aim of this study. The goal of this study is to shed light on the efficacy of the free meal program so that it may serve as a foundation for future program evaluations and improved laws.

A variety of sentiments, both positive and negative, appear on social media, reflecting support, criticism, or dissatisfaction. Negative sentiments often highlight issues such as uneven distribution of aid, questionable food quality, or lack of transparency in program implementation. On the other hand, positive sentiments indicate that many people appreciate this government initiative as a strategic step to reduce the economic burden on the underprivileged.

Unfortunately, this broad public response has not been systematically analyzed to provide reliable insights to policymakers. Without a deep understanding of public perception, the government risks facing challenges in increasing public acceptance and program effectiveness. Therefore, machine learning-based sentiment analysis is needed to identify sentiment patterns on social media accurately and efficiently. Thus, the results of the analysis can be the basis for continuous evaluation and improvement of policies. This issue is relevant to ensure that public policies such as the Free Meal Program are not only implemented well, but also receive positive acceptance from the wider community.

Literature Review

This chapter reviews existing literature and explores academic research issues, highlighting research issues within the broad scope of global scientific understanding. The chapter begins with a brief introduction to free meal school program, sentiment analysis data, and classification technique, as well as a brief overview of machine learning for classification, which provides a foundation for understanding the research effort.

Free Meal Program

The Free Meal Program or “Free Nutritious Meals” (MBG) initiated by the Prabowo-Gibran pair has officially been discussed in the 2025 budget planning. The Ministry of Finance (Kemenkeu), the Coordinating Ministry for Economic Affairs (Kemenko) and the Prabowo-

Gibran transition team have set the MBG budget at IDR 71 trillion in the first phase in 2025. This amount is considered to have taken into account the fiscal deficit target of 2.29% - 2.82% (Prabowo-Gibran, 2023; BBC Indonesia, 2024).

The MBG program in the first phase will focus on targeting elementary, junior high and senior high school students in the quintile 1 and 2 categories in the disadvantaged, outermost and remote areas (3T) in Indonesia (BBC Indonesia, 2024). According to the Prabowo-Gibran transition team, the target targets, budget size and program governance will continue to be evaluated and expanded to eradicate stunting in Indonesia.

The amount of budget to be spent and the effectiveness of the impact that will be generated from this program become the pros and cons of public discourse. Moreover, public budgets and public policies should be accounted for by policy makers. In addition, changes related to program names, targets, budgets and so on are known to civil society only through media coverage. There are no permanent, transparent and sustainable channels and mechanisms for public participation to ensure civil society participation in monitoring program developments. Various public concerns have emerged; from the quality of planning, limited fiscal space, to unclear governance (Suwastoyo, 2024). Given the urgency, the Center for Indonesia's Strategic Development Initiatives (CISDI) took the initiative to conduct a study of the MBG program which is divided into several series. The state budget is expected to be used transparently, measurably and have a positive impact on public health development.

From a social perspective, this program has a positive impact in reducing hunger levels, increasing access to education, and empowering local communities through the participation of BUMDes, UMKM, and cooperatives in the food supply chain. In addition, this program also provides real assistance to families in difficult economic conditions by reducing their economic burden and increasing access to adequate food. From an economic perspective, this program provides a significant stimulus to the food sector, food processing industry, and distribution sector, which results in local and national economic growth.

The provision of free meals in schools also helps increase consumption of goods and services, creates new business opportunities, and supports the growth of small and medium enterprises. However, to ensure the sustainability and effectiveness of this program, careful evaluation is needed. This evaluation must take into account various aspects, from the quality of the food provided to the efficiency of resource use. The implications of this evaluation will influence the direction and strategy of further program development, as well as ensuring that the program continues to provide maximum benefits to the Indonesian people as a whole. Thus, the free lunch program is not only a solution to the problem of hunger, but also a progressive step in building inclusive and sustainable social and economic welfare in Indonesia.

Twitter or currently known as the X application is a social media platform that allows its users to interact through what is called a tweet. Users can share information and views on various topics being discussed. X has features that allow unlimited submission of opinions, search for the latest news, share other people's tweets, and provide comments. In the X application, information can spread quickly and easily, making it a means of finding out someone's opinion sentiment, both positive and negative.

Sentiment analysis is a method for assessing the emotional tone in digital text to determine whether the tone is positive, negative, or neutral. The use of sentiment analysis involves analyzing opinions, feelings, evaluations, emotions, assessments, or attitudes towards a product, figure, organization, issue, service, and event. In addition, sentiment analysis is always related to society because the information obtained comes from social media where society acts as its users.

Social Media

Social media is a digital platform that is very influential in today's human life. Its main purpose is to facilitate communication between people, share information, and create social interactions without any limitations of space and time. In this digital era, public opinion is often expressed through social media, one of which is Platform X. This platform is the main place for the public to convey views, comments, and criticisms of current issues. The Free Lunch Program policy, for example, has received great attention on Platform X. The many opinions circulating on social media regarding this program show the importance of analyzing public responses. This method allows researchers to automatically identify and classify public opinion based on text generated by social media users. (Program et al., 2024)

Sentiment Analysis

Definition and Importance of Sentiment Analysis

Sentiment analysis is the process of analyzing digital text to determine whether the emotional tone of the message is positive, negative, or neutral.(Awazon, 2022)

The goal of sentiment analysis is to evaluate the emotions, attitudes, and opinions expressed by individuals across platforms regarding goods, brands, services, politics, or organizations. This approach includes machine learning and is lexicon-based, which allows grouping data into categories such as very positive, positive, neutral, negative, and very negative.

Machine Learning Technique For Sentiment Analysis

Various machine learning algorithms have been used for sentiment analysis, each with its strengths and limitations:

a. Naive Bayes: A probabilistic classifier based on Bayes' theorem, which performs well in text classification due to its simplicity and efficiency.
b. Support Vector Machines: A supervised learning model that finds the optimal hyperplane to separate classes in a feature space, robust in high-dimensional spaces.
c. Random Forest: This is an ensemble learning methodology where multiple decision trees are created and then their results are combined to perform better for classification tasks without overfitting.
d. Neural Networks: These models take inspiration from the way the human brain works; thus, neural networks are capable of learning very complex patterns in data. Variants of deep learning such as Convolutional Neural Networks and Long Short-Term Memory networks have performed impressively on sentiment analysis tasks.

A comparative study by (Kolchyna et al., 2015) demonstrated the superiority of machine learning methods, especially SVM and Naive Bayes classifiers over lexicon-based approaches when performing sentiment classification tasks.

Related Work
Sentiment analysis is increasingly used in public policy assessments to gauge citizen reactions. For example, studies on health policies during COVID-19 have highlighted important insights into public concerns and levels of support (Wang et al., 2021). In the education space, sentiment analysis has been used to assess programs such as school meal initiatives, offering data-driven insights for program optimization. A study by Kumar et al. in 2023 found evidence of the effectiveness of sentiment analysis in evaluating policies, integrating social media data to understand public satisfaction and areas for improvement.

Analyzing sentiments on social media, however, comes with its specific challenges. There is noisy data filled with slang, abbreviations, and grammatically incorrect sentences, complicating text processing. The informal and code-mixed varieties of Indonesian languages also add to the intricacies of sentiment analysis. These challenges would need specific preprocessing techniques and a strong linguistic model. Works like Rahimi et al. in 2020 discuss the challenges of analyzing informal language texts in low-resource settings. Table 1 summarizes previous work on sentiment analysis leveraging social media and public policy with several machine learning techniques.

Table 1
Previous work on Sentiment Analysis

Author /Year	**Title**	**Research Focus**	**Machine Learning Methods**
(Monselise et al., 2021)	Topics and sentiments of public concerns regarding COVID-19 vaccines: Social media trend analysis	Analysis of social media trends related to public sentiment towards vaccination	NLP, Sentiment Classification
(Adak et al., 2022)	Sentiment analysis of customer reviews of food delivery services using deep learning and explainable AI	Food delivery service customer review analysis using explainable AI	Deep Learning
(Ainin et al., 2020)	Sentiment analysis of multilingual tweets on halal tourism	Sentiment on halal tourism through multilingual tweet analysis	Random Forest, Naive Bayes
(Nguyen et al., 2019)	Pride, love, and twitter rants: Combining machine learning and qualitative techniques	Combination of machine learning and qualitative techniques for social media sentiment analysis	Sentiment Analysis with qualitative features
(Hudaefi et al., 2022)	Zakat administration in times of COVID-19 pandemic in Indonesia: A knowledge discovery via text mining	Text mining on zakat administration during the pandemic	Text Mining
(Tao et al., 2019)	Social media data-based sentiment analysis of tourists' air quality perceptions	Sentiment analysis of air quality perception by tourists	Neural Networks

(Landwehr et al., 2016)	Using tweets to support disaster planning, warning and response	Tweet analysis to support disaster planning	Logistic Regression, Support Vector Machine
(Sudo et al., 2020)	Robots, AI, and service automation (RAISA) in hospitality: Sentiment analysis of YouTube streaming data	Sentiment analysis of YouTube streaming data about RAISA	BERT, Contextual Analysis

Research Objectives

The research objectives of this study are follows:

1) To Gather Twitter information on the free meal program using crawling methods.
2) To Implement sentiment analysis to the gathered data in order to determine the public's perspective (positive, negative, or neutral).
3) To Categorize the sentiment-related primary subjects of public discourse.
4) To Present research to help policymakers and the government improve the calibre of social welfare initiatives.

Research Methodology

This chapter explains the research methodology used to analyze sentiment related to the "free meal" program promoted by Prabowo and Gibran. Also public reactions to the program through social media, especially through the X application or Twitter. This methodology includes the process of data collection, data pre-processing, data modelling, to classification using machine learning techniques to identify sentiment patterns (positive, negative, or neutral). This study aims to generate meaningful insights from social media data related to public sentiment towards the program.

Data Collection

Data was collected from the Twitter platform using Crawling Data Technique.
Table 2 describes the data explored using tweet with the keyword "makan siang gratis". The language filter used is "id" or Indonesian on Twitter since the launch of this free lunch program was announced during the Prabowo and Gibran presidential campaign in 2023.

Table 2
Dataset Crawling Parameters

No	Twit Param	Data Type	Description
1	Search	String	Search Keyword
2	Lang	String	Language used on Twit
3	Since	String	Filtering Tweets Posted data base on since date

Table 3
Dataset Explanation

Distributions This section shows the distribution of values for each column of the dataset.	
conversation_id_str	This distribution shows unique conversation IDs that are mostly distributed in a certain range. Large ID values indicate that this is data taken from Twitter, as IDs are usually long numbers.
favorite_count	Distribution of the number of "likes" or "favorites" on tweets. Most tweets have a low "like" value (close to zero), indicating that many tweets receive little attention or interaction.
id_str	Like conversation_id_str, this is a unique ID for a tweet. Its distribution follows a similar long ID pattern.
quote_count	Distribution of the number of "quote retweets". Most of the data has a value of zero, indicating that most tweets are not quoted by other users. However, there are some extreme values with higher "quote" numbers.
2-d Distributions This section shows the relationship between variables with a 2-dimensional distribution.	
favorite_count vs conversation_id_str	This graph shows that the number of “likes” is sporadically distributed across the conversation IDs. Most of the “like” values are low, with a few outliers having high “like” counts.
favorite_count vs id_str	Similar to the previous relationship, but focused on the unique ID of each tweet. The pattern is similar, with a few dots indicating popular tweets.
quote_count vs id_str	Most tweets have a low quote value, but there are a few outliers where tweets have a significant number of quotes. This suggests that only a small number of tweets attract the attention of other users to re-comment.
Values This section visualizes the distribution of values in the form of a line:	
conversation_id_str	The lines indicate sequential IDs. This confirms that the data may have been collected chronologically.
favorite_count	The distribution pattern shows that most values are close to zero with a few peaks (outliers).
quote_count	Most of the values are close to zero, indicating tweets that are rarely requoted, but there are a few peaks with higher values.

As presented in Figure 1, the total data obtained from Twitter from September 1, 2023 to March 26, 2025 is 10.531 data.The data contains fifteen attributes or features, namely conversation_id_str, created_at, favorite_count, full_text, id_str, image_url, in_reply_to_screen_name, lang, location, quote_count, reply_count, retweet_count, tweet_url, user_id_str, and username, as explained in Table 3.

Unfortunately, even though the language has been filtered to Indonesian, the dataset still has data in other languages (not Indonesian). Then the data is saved into a comma-separated values (csv) format document.

	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name	lang	location	quote_count	rep
0	1866872716541972550	Wed Dec 11 23:25:45 +0000 2024	4	@pramudyawdynto Target makan gratis e prabowo	1866987767919677789	NaN	pramudyawdynto	in	Vladivostok, Russia	0	
1	1866464297657975123	Wed Dec 11 21:52:21 +0000 2024	1	@Yudhi2024 @prabowo masih enak dan bergizi mak...	1866964263254032437	NaN	Yudhi2024	in	NaN	0	
2	1866464297657975123	Wed Dec 11 21:21:34 +0000 2024	0	@StevanFirman15 @prabowo Yakin pak makan grati...	1866956515414094156	NaN	StevanFirman15	in	Bekasi Barat, Indonesia	0	
3	1866754957552259364	Wed Dec 11 18:42:05 +0000 2024	0	@03__nakula Makanya program Prabowo makan sian...	1866916381872230896	NaN	03__nakula	in	Bandung, Jawa Barat	0	
...	...	...	...	...	...	...					
10527	1853711186329563542	Tue Nov 05 08:09:21 +0000 2024	0	https://t.co/oNxnzYO7qs Wapres tinjau uji coba...	1853711186329563542	NaN					N
10528	1853708711438799265	Tue Nov 05 07:59:31 +0000 2024	0	Polres Mahakam Ulu Laksanakan Program Pembagia...	1853708711438799265	https://pbs.twimg.com/ext_tw_video_thumb/18537...					N
10529	1853708646867505556	Tue Nov 05 07:59:16 +0000 2024	0	Polres Mahakam Ulu Laksanakan Program Pembagia...	1853708646867505556	https://pbs.twimg.com/media/GbmyKyxaMAA9uOC.jpg					N
10530	1853702824854741218	Tue Nov 05 07:36:08 +0000 2024	2	Roji menyebut dampak positif dari pemberian ma...	1853702824854741218	NaN					N
10531	1853568660868391232	Tue Nov 05 07:34:18 +0000 2024	0	Akun Tiktok @hariankompas dalam #AADK edisi Se...	1853702363539980766	https://pbs.twimg.com/ext_tw_video_thumb/18537...					hariankom

10532 rows × 15 columns

Figure 1. Total Data Tweets Free Meal on Twitter

Research Framework

This research framework includes the following steps: Problem Definition and Literature Review, Data Collection: Retrieve data from Twitter using specific keywords, Data Pre-processing: Cleaning and preparing data for further analysis., Feature Extraction: Applying stemming and vectorization techniques., Sentiment Classification: Using machine learning models (KNN, Naive Bayes, and SVM), and Model Evaluation: Compares model performance using evaluation matrices. The details of the research framework for this study are shown in Figure 2.

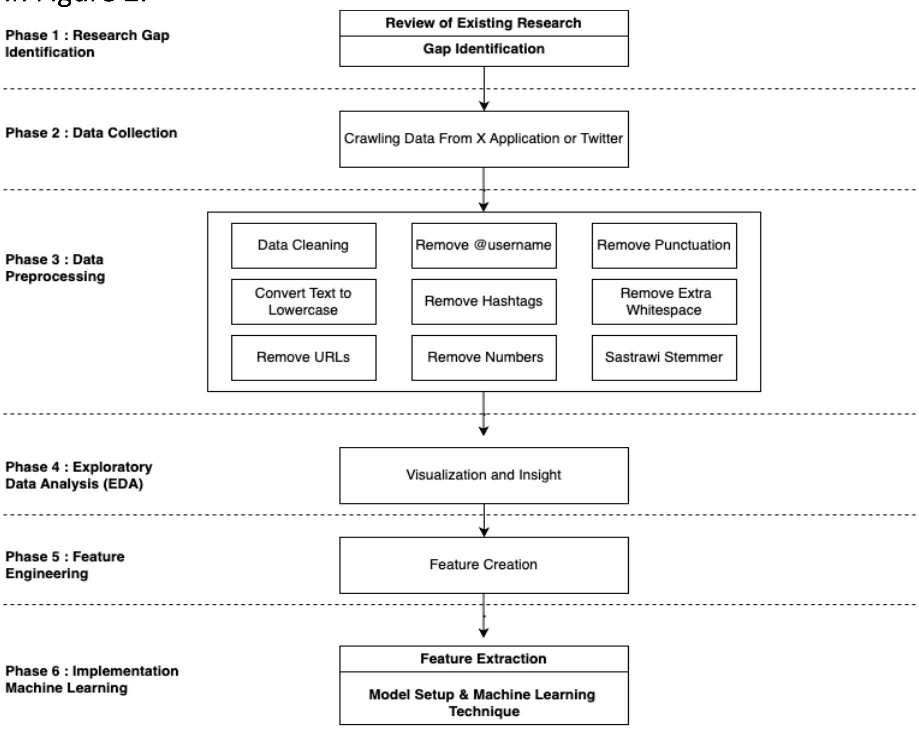

Figure 2. Research Framework For Sentiment Analysis

Problem Formulation

The main objective of this study is to use a sentiment analysis approach to public reactions on social media with machine learning technique classification, thus providing valuable data for further government policies. However, to ensure accurate and reliable analysis, several problems need to be solved. Main Problems are identifying public sentiment regarding the "Free Meal" program and Comparing the performance of KKN, Naïve Bayes, and SVM Algorithms in sentiment classification based on Twitter data

Data Collection

The following dataset was obtained by crawling data process on application X or twitter. The data obtained is related to tweets about Prabowo and Gibran's free meal program. This data collected from 2023 until January 2025.

Data Pre-processing and Data Cleaning

Initial analysis needs to be completed before moving on to further pre-processing. Data merging procedures are required to unify all the raw data into a single data frame once we have a good understanding of the features available in the data set. Several data processing and data transformation procedures will be used on the data set in an attempt to further unify the disorganized raw data.

Data cleaning is an important process in sentiment analysis, especially to ensure that the data used is clean, relevant, and can be processed well by the model. Here are the data cleaning steps carried out on the Twitter tweet dataset about the Prabowo-Gibran free meal program:

Initialize Sastrawi Stemmer

Sastrawi is used to perform stemming, which is changing affixed words into basic forms (root words). For example, "makannya" becomes "makan". Using Stemming helps simplify word variations so that the model can more easily recognize patterns in the data.

- Convert Text to Lowercase
  All letters in the text are converted to lowercase. Makes the analysis more consistent because uppercase and lowercase are treated the same. For example, "PRABOWO" and "prabowo" are considered the same.
- Remove URLs
  Removes links (URLs) from text such as "https://...". Links do not provide relevant sentiment information and can interfere with analysis.
- Remove @username
  Removes mentions or tags such as "@user". Mentions are usually not relevant for sentiment analysis because they only point to a specific account.
- Remove Hashtags
  Removing hashtags such as "#prabowo" or "#makangratis". Hashtags can be removed because they often do not contain the context needed in sentiment analysis, although there are certain cases where hashtags are analyzed separately.
- Remove Numbers
  Removes numbers from text. Numbers usually have no meaning in the context of sentiment, unless specifically relevant (can be processed separately if important).
- Remove Punctuation

Vol. 14, No. 4, 2025, E-ISSN: 2226-6348 © 2025

Removes punctuation such as ".", ",", "?", etc. Punctuation does not contribute directly to sentiment analysis.

- Remove Extra Whitespace
  Removes excess whitespace in text. Makes text neater and easier to read.
- Apply Stemming
  Uses the Sastrawi stemmer to convert words to their basic form. Reduces variations in words that have the same meaning.
- Apply Preprocessing
  Combines all the above steps into one preprocessing pipeline that is applied to the entire dataset. Ensures all data is processed in a uniform manner.
- Translate Data to Minimize English Words
  Translates English words to Indonesian using a library such as the Google Translate API. Standardizes the language so that all text is in one language (Bahasa Indonesia) to facilitate sentiment analysis.

In the Figure 3, it explains the flow of the data cleaning process with a literary stemmer to the process of minimizing words in English.

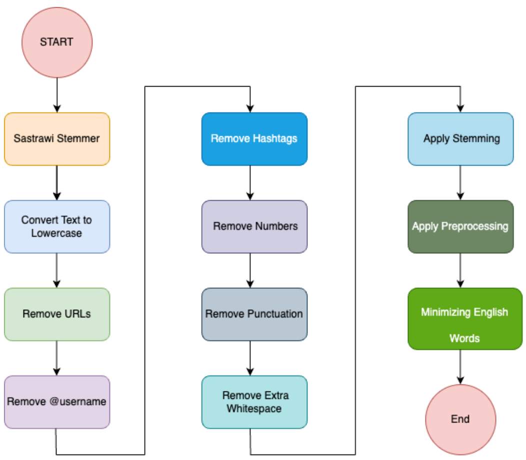

Figure 3. Flow Data Cleaning and Preparation

To identify missing values and remove rows and columns without values, data cleaning is done in this section. Figure 4 shows that in data pre-processing, several things are done such as converting text to lower case, removing URLs, removing @username, removing hashtags, removing numbers, removing punctuation and removing extra whitespace. Then apply all the pre-processing processes with the syntax data['full_text'] = data['full_text'].apply(preprocess_text)

```python
# Initialize Sastrawi stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()

# Preprocessing function for tweets
def preprocess_text(text):
    text = text.lower()  # Convert text to lowercase
    text = re.sub(r"http\S+|www\S+|https\S+", '', text, flags=re.MULTILINE)  # Remove URLs
    text = re.sub(r'@\w+', '', text)  # Remove @username
    text = re.sub(r'#\w+', '', text)  # Remove hashtags
    text = re.sub(r'\d+', '', text)  # Remove numbers
    text = re.sub(r'[^\w\s]', '', text)  # Remove punctuation
    text = re.sub(r'\s+', ' ', text).strip()  # Remove extra whitespace
    text = stemmer.stem(text)  # Apply stemming
    return text

# Apply preprocessing
data['full_text'] = data['full_text'].apply(preprocess_text)
```

Figure 4. Data Cleaning Process

```python
import matplotlib.pyplot as plt

# Count the number of duplicates
tweet_bot = len(data.loc[data['full_text'].duplicated() == True])
# Count the number of non-duplicates
tweet_normal = len(data.loc[~data['full_text'].duplicated()])
labels = 'Bot', 'Normal'
sizes = np.array([tweet_bot, tweet_normal])
colors = ['lightskyblue', 'pink']
explode= (0, 0.5)
def absolute_value(val):
    a  = np.round(val/100.*sizes.sum(), 0)

    a=  str(round(val,2))+"%"+'\n'+str(a) +" data"
    return a

plt.pie(sizes, labels=labels, colors=colors,
        autopct=absolute_value, explode=explode,shadow=True)

plt.axis('equal')
plt.title("Data Proportion")
plt.legend()
plt.show()
```

Figure 5. Process Cleaning Data and Create Graphs based on Data

Figure 6 shows that from the data that was previously collected from the 2023 – March 2025 datasets, a data cleaning process was carried out to obtain a data proportion of which around 80.73% or 7465.0 data were normal data while 19.27% or 1782 data were BOT data.

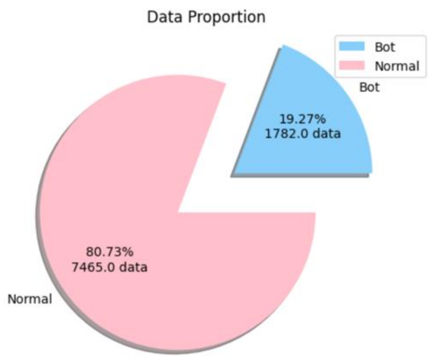

Figure 6. Data Proportion

Data Modelling

The cleaned data is converted into numerical format using vectorization techniques such as Term Frequency-Inverse Document Frequency (TF-IDF). This representation is used as input for the machine learning model. In Figure 7, this is the process of creating a data model. The resulting model will be entered into the machine learning technique to get the results. The syntax used for the data model creation process is : vectorizer = TfidfVectorizer(max_features=5000) X_vectorized = vectorizer.fit_transform(X)

```python
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report

# Divide the data into features (X) and targets (y)
X = data['full_text']
y = data['sentiment']
```

```python
# TF-IDF Vectorizer to convert text to vectors
vectorizer = TfidfVectorizer(max_features=5000)  # Adjust max_features as needed
X_vectorized = vectorizer.fit_transform(X)
```

```python
# Splitting data into training and test data
X_train, X_test, y_train, y_test = train_test_split(X_vectorized, y, test_size=0.2, random_state=42)
```

Figure 7. Process Data Modelling

Stemming Data

Stemming is done to reduce words to their basic form. For example, "eat," "the food," and "ate" all return to "eat." This process helps unite different forms of words that have similar meanings. And in this project we will use the Sastrawi library for the data stemming process.

```python
# Initialize Sastrawi stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()
```

Figure 8. Initialize Sastrawi Stemmer

Classification Models and Technique

The final stage to obtain sentiment analysis results is to apply and classify the data model into machine learning techniques. The machine learning techniques that will be used are KNN, SVM and Naive Bayes

Three machine learning algorithms are used for sentiment classification:

1. K-Nearest Neighbors (KNN):Classifies tweets based on the majority sentiment of their nearest neighbors in feature space.
2. Naive Bayes: Bayes' theorem based probabilistic model suitable for text classification.
3. Support Vector Machine (SVM): A supervised learning model that separates sentiment classes using hyperplanes in high-dimensional space.

Each model will be evaluated using metrics such as accuracy, precision, recall, and F1-score to determine the best performance. Model results are evaluated using the following metrics:

a. Accuracy: Percentage of correct predictions.
b. Precision: The accuracy of positive predictions.
c. Recall: The model's ability to detect all positive data.
d. F1-Score: Harmonic mean of precision and recall.

In Figure 9 is the model implementation process in each machine learning technique.

```
# Hyperparameter tuning untuk KNN
knn_params = {'n_neighbors': [3, 5, 7, 9], 'weights': ['uniform', 'distance']}
knn_grid = GridSearchCV(KNeighborsClassifier(), knn_params, cv=5, scoring='accuracy')
knn_grid.fit(X_train, y_train)
knn_best_model = knn_grid.best_estimator_
```

```
# Hyperparameter tuning untuk Naive Bayes
nb_params = {'alpha': [0.1, 0.5, 1.0, 1.5, 2.0]}
nb_grid = GridSearchCV(MultinomialNB(), nb_params, cv=5, scoring='accuracy')
nb_grid.fit(X_train, y_train)
nb_best_model = nb_grid.best_estimator_
```

```
# Hyperparameter tuning untuk SVM
svm_params = {'C': [0.1, 1, 10, 100], 'kernel': ['linear', 'rbf']}
svm_grid = GridSearchCV(SVC(), svm_params, cv=5, scoring='accuracy')
svm_grid.fit(X_train, y_train)
svm_best_model = svm_grid.best_estimator_
```

```
# Predicting test data and calculating accuracy
knn_pred = knn_best_model.predict(X_test)
nb_pred = nb_best_model.predict(X_test)
svm_pred = svm_best_model.predict(X_test)

# Displays accuracy results and classification reports
print("KNN Accuracy:", accuracy_score(y_test, knn_pred))

print("\nNaive Bayes Accuracy:", accuracy_score(y_test, nb_pred))

print("\nSVM Accuracy:", accuracy_score(y_test, svm_pred))
```

Figure 9. Implementation Model to Machine Learning Technique

Result Sentiment Analysis

In this sentiment analysis section, we will identify each word from social media tweets and categorize whether the word is positive, negative or neutral. Here are some examples of positive, negative and neutral sentences.

	full_text	sentiment
0	target makan gratis e prabowo	positive
1	masih enak dan gizi makan gratis jumat berkah ...	positive
2	yakin pak makan gratis rp sesuai gizi yang ada...	positive
3	makanya program prabowo makan siang gratisbiar...	positive
4	prancis dukung program makan gizi gratis prabowo	positive
...	...	...
7454	hai masih dalam suasana ingat hut ke menteri l...	positive
7455	wapres tinjau uji coba makan gizi gratis di pa...	positive
7456	polres mahakam ulu laksana program bagi makan ...	positive
7457	roji sebut dampak positif dari beri makan gizi...	positive
7458	akun tiktok dalam edisi selasa november ulas t...	positive

7459 rows × 2 columns

Figure 10. Top Positive Sentiment Prediction

```
negative_tweets.head(15)
```

	full_text
5	iya benar juga kata lu ya rakyat miskin mana y...
8	dasar dunia internasional ini radikal intolera...
18	ini yang saya khawatir sejak prabowo jadi menh...
31	presiden prabowo subianto klaim program makan ...
32	presiden prabowo subianto ungkap ada rp miliar...
85	brantas judi online saja tidak becus mengomong...
86	buat program makan siang gratis saja harus nge...
90	rugi negara ratus trilyun denda cuma satu mily...
94	bukan ancam cara langsung ya ini ancam nya sep...
105	jelas sekali sih mul ini takut takut penjara d...
107	mereka juga meminta sesuatu bu dok tak ada mak...
180	apa enggak malu itu pmbesar negeri ini kasih m...
186	hah cius tapi jalan mbak baru januari gue miki...
191	makan gizi gratis mbak jadi salah satu titik f...
194	jokowi meninggalkan hutang banyak yang tanggun...

dtype: object

Figure 11. Top Negative Sentiment Prediction

	full_text
11	mangkas anggar makan gizi gratis jadi rp ribu ...
35	dulu bapak jadi sales panci yang di jalan esdm...
45	cina saja langsung takut dan segera inves cair...
74	menteri laut dan ikan kkp sakti wahyu trenggon...
76	dari wawancara itu ungkap operasi oligarki jag...
...	...
7296	jaman susu impor di program makan gizi gratis ...
7312	program makan gizi gratis di bagai negara tunj...
7397	cah ada info recruitment awas program makan gi...
7404	ikut program makan gizi gratis dulu bro biar k...
7445	indonesia bakal impor juta sapi perah buat mak...

371 rows × 1 columns

dtype: object

Figure 12. Top Neutral Sentiment Prediction

The Figure 13 shows a word cloud for positive sentiment reviews. The word cloud analysis illustrates that “healthy”, “economic movement”, “prosperous”, “thank you”, “future”, “lunch”, “free lunch”, “industrial sector” and “great potential” are the most frequently used words in the reviews. The word “Lunch” is the most frequently used word, indicating that this program often receives positive reviews from public reactions on social media. While the words “healthy” and “prosperous” are the words that are the hopes of this free meal program when it is run by the government under the leadership of President and Vice President Prabowo Subianto and Gibran.

Figure 13 World Cloud of Positive Sentiment

The Figure 14 shows a word cloud for negative sentiment reviews. The word cloud analysis illustrates that "no", "prabowo", "problem", "fail", "stupid", "sarcasm", "poor", "corruptor" and "politics" are the most frequently used words in the reviews. The words "fail", "problem" and "poor" illustrate that many people will be pessimistic about this free meal program. They assume that this program will fail and not continue and will not be on target. So it will add new problems for the country, while currently those who need to be helped are people with poor economies to be more prosperous.

Figure 15 World Cloud of Negative Sentiment

The Figure 16 shows a word cloud for neutral sentiment reviews. The word cloud analysis illustrates that “nutrition”, “confused”, “program”, “campaign promise”, “help”, “realization”, “children”, “economy” and “need” are the most frequently used words in the reviews. These words may indicate that the meal program provides a little hope for children to get better nutrition. In addition, the community also hopes that the government can realize the program, not just make promises during the campaign.

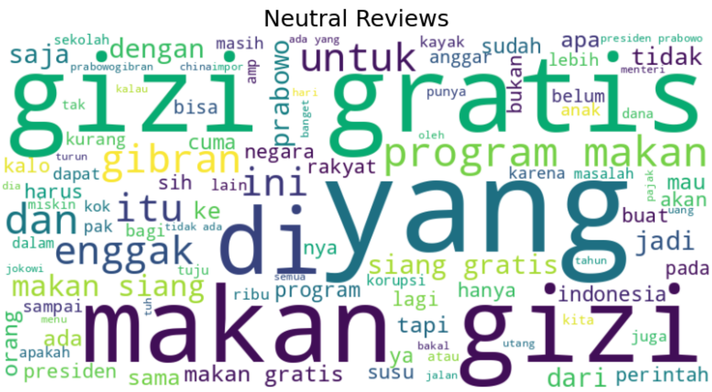

Figure 16 World Cloud of Neutral Sentiment

From the Table 3 and Figure 17 the balanced data size of the sentiment analysis, it can be seen that data sets are balanced. With a sample data 33.33% or 906 row. This will make it easier for us to get balanced results.

Table 3
Balanced Data Size Sentiment

No	Type	Data Size
1	Neutral	302
2	Positive	302
3	Negative	302

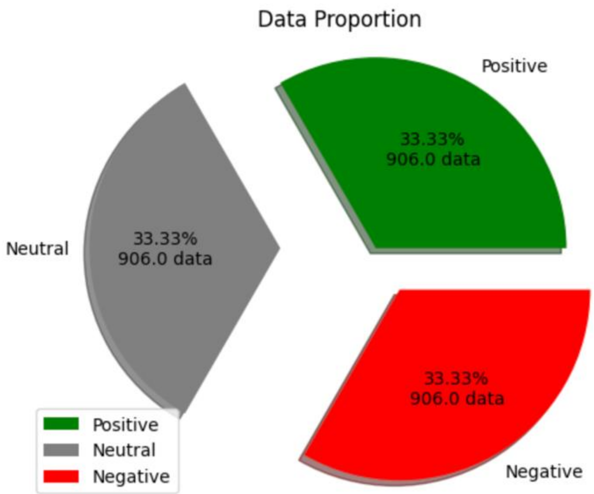

Figure 17 Graph of Proportion of data from sentiment analysis results

Feature Extraction

The feature extraction process is carried out to convert raw text data into numeric representations that can be processed by the machine learning model. The first step is to perform class balancing by ensuring that the amount of data for each sentiment class (positive, neutral, and negative) has the same proportion. This aims to avoid model bias towards classes with dominant data amounts. In this implementation, the under sampling method is applied, namely selecting the same number of samples based on the smallest number of classes, resulting in a balanced distribution with each class having 33.33% or 906 rows data samples.

The next step, the text data is converted into a vector using the TF-IDF (Term Frequency-Inverse Document Frequency) Vectorizer, with a maximum feature limit used of 5,000 features. TF-IDF calculates the weight value of each word based on its frequency of occurrence in a particular document (term frequency) and how unique the word is compared to other documents (inverse document frequency). This process produces a matrix with a size of (912, 3462), where 912 is the number of documents (data) after balancing, and 3462 is the number of unique features generated.

The last step is to encode the sentiment label using LabelEncoder. The sentiment labels “positive,” “neutral,” and “negative” are transformed into their respective numeric values (0 for negative, 1 for neutral, and 2 for positive). This process ensures that the target data conforms to a format acceptable to the machine learning algorithm. This transformation produces data ready for the model training process, with numeric representations of the text and encoded sentiment labels in Figure 18.

```
Balanced Data Size: sentiment
negative    304
neutral     304
positive    304
Name: count, dtype: int64
TF-IDF Shape (Features): (912, 3462)
Encoded Sentiments: ['negative' 'neutral' 'positive']
```

Figure 18. Result of Feature Extraction

Model Development

At this stage, we will help the model become Model X and Y. And then use the TF-IDF Vectorizer (Term Frequency-Inverse Document Frequency Vectorizer) technique to convert text data into numeric representations (vectors) before the data is processed using machine learning algorithms. Its main function is to give weight to each word in the document, so that relevant words are more prominent and common words are less influential. TF-IDF Vectorizer is an important step in the text-based machine learning pipeline. It helps capture important information from text data and ignores irrelevant information, thereby improving the performance and accuracy of the prediction model.

The Importance of TF-IDF in Machine Learning :

a) Reducing Data Dimensionality: TF-IDF allows the use of only relevant words (for example, with max_features=5000) without having to process all the words in the dataset, making the model more efficient.
b) Reducing Overfitting: Very common or irrelevant words (stopwords) are given low weight or ignored, which helps the model not to be affected by noise.
c) Highlight Relevant Words: Words that are specific and relevant to a particular class will get higher weights, increasing the model’s accuracy in understanding the relationship between words and target labels.
d) Compatible with Machine Learning: Machine learning models like SVM, Naive Bayes, or KNN can only process numeric data. TF-IDF converts text data into a numeric format that is acceptable to the model.

In Figure 19 is the code syntax for the model creation process and also the implementation of the TF-IDF Vectorizer technique for each model.

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report

# Divide the data into features (X) and targets (y)
X = data['full_text']
y = data['sentiment']
```

```
# TF-IDF Vectorizer to convert text to vectors
vectorizer = TfidfVectorizer(max_features=5000)  # Adjust max_features as needed
X_vectorized = vectorizer.fit_transform(X)
```

```
# Splitting data into training and test data
X_train, X_test, y_train, y_test = train_test_split(X_vectorized, y, test_size=0.2, random_state=42)
```

Figure 19. Creating Data Model and Implement TF-IDF

Model Evaluation and Implement Machine Learning Technique

The model that has been successfully created previously will then be processed for implementation in machine learning techniques. To get better accuracy, in this project researchers will use several machine learning techniques such as KNN (K-Nearest Neighbors), Naive Bayes, and SVM (Support Vector Machine). In this process, Hyperparameter Tuning will be used for each machine learning technique. Hyperparameter tuning is done to find the best parameters of the three machine learning models. Here are the details:

1) KNN (K-Nearest Neighbors)
Tested parameters:
a) n_neighbors: Number of neighbors considered (3, 5, 7, 9).
b) weights: How to give weight to neighbors (uniform for all neighbors have the same weight, and distance for weights based on distance).

The best model is stored in the variable knn_best_model.

2) Naive Bayes
Tested parameters:
alpha: Smoothing parameter (tested values: 0.1, 0.5, 1.0, 1.5, 2.0).
The best model is stored in the variable nb_best_model.

3) SVM (Support Vector Machine)

Tested parameters:
C: Regularization parameter (tested values: 0.1, 1, 10, 100).
kernel: Kernel function (linear for linear kernel, and rbf for radial basis function kernel).
The best model is stored in the variable svm_best_model.

After finding the best model for each algorithm, an evaluation is carried out using test data (X_test and y_test) with the following steps:

a) Test Data Prediction
KNN: Using knn_best_model to predict.
Naive Bayes: Using nb_best_model to predict.
SVM: Using svm_best_model to predict.

b) Calculating Accuracy
Accuracy is calculated with the accuracy score function, which compares the predictions to the original labels in the test data.

After conducting model evaluation and data test, the results were obtained as Table 4 for Cross Validation Results Accuracy and Table 5 For Confusion Matrix Results Accuracy:

Table 4
Cross Validation Results Accuracy

No	Model	Accuracy
1	KNN	91.02%
2	Naïve Bayes	91.02%
3	SVM	92.70%

Table 5
Confusion Matrix Results Accuracy

No	Model	Overall Accuracy	Precision	Recall	F1-Score
1	Naïve Bayes	50.27%	0.53	0.50	0.44
2	Support Vector Machine (SVM)	56.28	0.56	0.58	0.57

After conducting sentiment analysis using three Machine Learning techniques—Naïve Bayes, Support Vector Machine (SVM), and KNN—it can be concluded that Naïve Bayes has the lowest accuracy on the test data, which is 50.27%, with a low f1-score, indicating that this model is less effective in classifying sentiment related to the free meal program, while Support Vector Machine (SVM) shows better accuracy than Naïve Bayes on the test data, which is 56.28%, and has the highest f1-score in the positive sentiment class. This shows that SVM is better at capturing positive sentiment patterns than Naïve Bayes.

Based on the results of this evaluation, Support Vector Machine is recommended as the most optimal model for sentiment analysis of the government's free meal program.

Discussion

After conducting several processes for sentiment analysis, model evaluation, and data testing with machine learning data techniques, it was found that the accuracy of the machine learning technique, Support Vector Machine (92.70%) is better than KNN and Naïve Bayes with the same high accuracy (91.02%) in predicting sentiment. Hyperparameter tuning successfully improves model performance by selecting the best combination of parameters for each algorithm. All models show very good performance with accuracy above 90%, which means that the data has been processed well (for example: through TF-IDF Vectorizer and balanced training-test data division). Although Naïve bayes is slightly lower in accuracy, this algorithm is usually more stable for data with high dimensions and complex distributions, so it can be considered for larger or more varied datasets. With these results, SVM can be selected as the best model for sentiment analysis cases based on performance on test data.

Conclusion

Sentiment Analysis of the Free Meal Program voiced by Prabowo-Gibran tries to get public responses to the program. This analysis clarifies public perception of the policy through data taken from tweets on Twitter through the web scraping process. This project involves several phases, from data collection to final analysis. The data after being scraped will go through a cleaning stage which means that various preprocessing techniques can be carried out, including converting text to lowercase, removing URLs, usernames, punctuation, and irrelevant words. Then, stemming and foreign word translation are carried out to create uniform data.

The cleaned data is then converted into numeric form using the TF-IDF Vectorizer. This technique allows us to represent text in vector form so that it can be used by machine learning algorithms. The three algorithms used in this project are K-Nearest Neighbors (KNN), Naive Bayes, and Support Vector Machine. The results of the analysis show that SVM has the highest level of accuracy, which is 92.70%, while KNN and Naive Bayes are slightly lower at 91.02%. This means that the results show that the machine learning model is reliable in analyzing public sentiment based on social media text. With a deeper analysis, it can be seen that the trend of public opinion tends to be positive, which means that the public is still accepted in the 'Free Meals' initiative. It cannot be ignored that there are a small number of people who express negative and neutral sentiments, reflecting concerns or misunderstandings about the program.

From the success of the project, we can draw the following conclusions:

a) Data quality is crucial: Good data cleaning will play a big role in achieving better results after analysis.
b) Appropriate model selection: SVM performed very well, so it is the primary choices for this case.
c) Dominantly positive sentiment: The Free Meal Program has received strong positive support from the public.

Overall, the project successfully achieved its goal of measuring public response to a policy program in a structured and data-driven manner. The project also demonstrated that social media sentiment analysis can act as a powerful tool in evaluating public policies in real time.

References

Adak, A., Pradhan, B., & Shukla, N. (2022). Sentiment Analysis of Customer Reviews of Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence: Systematic Review. *Foods*, *11*(10). https://doi.org/10.3390/foods11101500

Ainin, S., Feizollah, A., Anuar, N. B., & Abdullah, N. A. (2020). Sentiment analyses of multilingual tweets on halal tourism. *Tourism Management Perspectives*, *34*. https://doi.org/10.1016/j.tmp.2020.100658

Ajhari, A. A. (2023). The Comparison of Sentiment Analysis of Moon Knight Movie Reviews between Multinomial Naive Bayes and Support Vector Machine. *Applied Information System and Management (AISM)*, *6*(1), 13–20. https://doi.org/10.15408/aism.v6i1.26045

Awazon. (2022). *Apa yang dimaksud dengan Analisis Sentimen?* Https://Aws.Amazon.Com/Id/What-Is/Sentiment-Analysis.

Hudaefi, F. A., Caraka, R. E., & Wahid, H. (2022). Zakat administration in times of COVID-19 pandemic in Indonesia: a knowledge discovery via text mining. *International Journal of Islamic and Middle Eastern Finance and Management*, *15*(2), 271–286. https://doi.org/10.1108/IMEFM-05-2020-0250

Klimczuk, A. (2021). Introductory Chapter: Demographic Analysis. In *Demographic Analysis - Selected Concepts, Tools, and Applications*. IntechOpen. https://doi.org/10.5772/intechopen.100503

Kolchyna, O., Souza, T. T. P., Treleaven, P., & Aste, T. (2015). *Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination*. http://arxiv.org/abs/1507.00955

Landwehr, P. M., Wei, W., Kowalchuck, M., & Carley, K. M. (2016). Using tweets to support disaster planning, warning and response. *Safety Science*, *90*, 33–47. https://doi.org/10.1016/j.ssci.2016.04.012

Monselise, M., Chang, C. H., Ferreira, G., Yang, R., & Yang, C. C. (2021). Topics and sentiments of public concerns regarding COVID-19 vaccines: Social media trend analysis. *Journal of Medical Internet Research*, *23*(10). https://doi.org/10.2196/30765

Nguyen, T. T., Criss, S., Allen, A. M., Glymour, M. M., Phan, L., Trevino, R., Dasari, S., & Nguyen, Q. C. (2019). Pride, love, and twitter rants: Combining machine learning and qualitative techniques to understand what our tweets reveal about race in the us. *International Journal of Environmental Research and Public Health*, *16*(10). https://doi.org/10.3390/ijerph16101766

Program, T., Siang, M., Tundo, G., & Rachmawati, D. N. (2024). Implementasi Algoritma Naive Bayes untuk Analisis Sentimen. In *Jurnal Indonesia : Manajemen Informatika dan Komunikasi (JIMIK)* (Vol. 5, Issue 3). https://journal.stmiki.ac.id

Ratna Patria. (2022). *Analisis Sentimen Media Sosial: Pengertian, Manfaat, Cara*. Https://Www.Domainesia.Com/Berita/Analisis-Sentimen-Media-Sosial/.

Sudo, K., Murasaki, K., Kinebuchi, T., Kimura, S., & Waki, K. (2020). Machine Learning–Based Screening of Healthy Meals From Image Analysis: System Development and Pilot Study. *JMIR Formative Research*, *4*(10), e18507. https://doi.org/10.2196/18507

Tao, Y., Zhang, F., Shi, C., & Chen, Y. (2019). Social media data-based sentiment analysis of tourists' air quality perceptions. *Sustainability (Switzerland)*, *11*(18). https://doi.org/10.3390/su11185070