Text Mining for Sentiment Analysis of Social Media Regarding the Revision of the TNI Law Using Python by Utilizing the Lexicon Approach

Ridwan[1], Hapzi[2], Hendarman Lubis[3]
[1]Institut Teknologi dan Bisnis Dewantara, Bogor, Indonesia, ridwans70@gmail.com
[2]Universitas Bhayangkara Jakarta Raya, Jakarta, Indonesia, hapzi.ali@gmail.com
[3]Universitas Bhayangkara Jakarta Raya, Jakarta, Indonesia, hendarman.lubis@dsn.ubharajaya.ac.id

Corresponding Author: hapzi.ali@gmail.com[2]

Abstract: This study aims to analyze public sentiment regarding the proposed revision of the Indonesian National Military (TNI) Law through social media, specifically Twitter. With the growing role of social media as a public sphere, it serves as a critical platform for expressing public opinions on national issues, including legal reforms. Using sentiment analysis, this research classifies Twitter posts into positive, neutral, or negative categories based on content related to the revision of the TNI Law. Data collection was carried out using the Twitter API (Tweepy), gathering 300 tweets using keywords related to the TNI Law revision. The analysis employed two methods: a lexicon-based approach using TextBlob and machine learning classification using Naive Bayes. The results reveal that the majority of the tweets express neutral sentiment (44.3%), followed by negative sentiment (28.3%) and positive sentiment (27.3%) based on TextBlob analysis. The Naive Bayes model demonstrated better sensitivity to negative sentiment, with 38.3% of tweets classified as negative. The findings suggest that public concern centers around the potential militarization of civil space, human rights violations, and threats to democracy. These concerns highlight the importance of transparent communication from the government regarding policy changes. Additionally, the research underscores the potential of natural language processing (NLP) and machine learning in monitoring public opinion in real time, which can serve as a valuable tool for policymakers in shaping responsive, data-driven legislation.

Keyword: Sentiment Analysis, TNI Law Revision, Social Media, Twitter, Natural Language Processing, Machine Learning, Naive Bayes, Public Opinion, Indonesia, TextBlob.

INTRODUCTION

In the past few decades, the world has undergone a major transformation in the field of information and communication technology.One of the most significant impacts of this transformation is the emergence of social media as the primary means of communication in modern society.Social media is not only used to share personal information, but it has also become an important platform for voicing opinions, criticizing policies, and shaping public

opinion on social and political issues (Kaplan & Haenlein, 2010).In Indonesia, the use of social media is very high; various national strategic issues often become subjects of open public discussion on platforms such as Twitter, Facebook, and Instagram.

Amid the dynamics of digital democracy, one of the issues that has surfaced and garnered significant public attention is the plan to revise Law Number 34 of 2004 concerning the Indonesian National Armed Forces (TNI Law).The government and the legislature have proposed several changes deemed important to expand the role of the Indonesian National Armed Forces (TNI), especially in facing non-military threats such as natural disasters, cyber attacks, and terrorism (Republika, 2023).The proposal is viewed by some as a form of adaptation to the complexity of threats in the modern era and as a step to strengthen the role of military institutions in maintaining national sovereignty.

However, on the other hand, concerns have emerged from academics, activists, and civil society regarding the potential problems of this revision.Several points of the revision are considered to blur the lines between military and civilian functions, and potentially open up space for TNI intervention in civilian areas that should be handled by non-military institutions (LIPI, 2020; KontraS, 2023).In the context of democracy, civilian supremacy over the military is a fundamental principle that must not be overlooked, as the military's excessive dominance in civil affairs is feared to threaten human rights principles and democratic control.

Public opinion regarding the revision of the TNI Law reflects the divided perceptions of society towards the position and role of the military in a democratic state.Interestingly, social media has become the main arena for expressing those sentiments.The public actively comments, debates, and even forms new narratives regarding this revision issue, both in the form of support and opposition.This is where sentiment analysis plays an important role.By utilizing text mining and natural language processing (NLP) approaches, sentiment analysis enables researchers to extract and classify emotions or public opinions from unstructured posts on social media (Liu, 2012).

In this research, data collected from social media—specifically Twitter—will be processed and analyzed using the Python programming language.Python was chosen for its strong capabilities in text data processing and support libraries such as Tweepy for data retrieval (API), pandas for data management, and NLTK, TextBlob, or SentiStrength for the sentiment analysis process.By utilizing these tools, the data processing can be carried out systematically and efficiently, starting from data preprocessing, tokenization, sentiment classification, to visualization of the analysis results.

This research aims to conduct a sentiment analysis of the Indonesian public's responses on social media regarding the proposed revision of the TNI Law.Through this analysis, it is hoped that the tendency of public sentiment whether it leans towards positive, negative, or neutral can be depicted, while also providing a deeper understanding of how the public interprets the involvement of the military in a civil context.In addition, this research is also expected to serve as an additional reference for policymakers to better understand public responses as part of a participatory and responsive policy formulation process to the people's voices.

By understanding the digital landscape of public opinion, the state can not only capture the aspirations of society more inclusively but also build more adaptive and democratic policies, especially on issues directly related to security stability and civil liberties.

METHOD
Types and Approaches to Research
This research is a quantitative study with a digital content analysis approach using sentiment analysis methods based on natural language processing (NLP).The objective of this research is to measure and understand public opinion regarding the revision of the Indonesian

National Armed Forces Law (UU TNI) through data sourced from social media, particularly Twitter.Sentiment analysis has proven effective in examining public opinion in the context of politics and policy (Liu, 2012).

Location and Time of Research

This research was conducted online through data collection from the social media platform Twitter.The research period runs from March 10 to April 10, 2025, aligned with the period of intense discourse on the revision of the TNI Law developing on social media.

Source and Type of Data

The data used in this research is secondary data in the form of public tweets containing keywords related to the revision of the TNI Law.The data includes text content, upload time, and other relevant metadata.Social media data is used as a mirror of the continuously evolving public opinion (Kaplan & Haenlein, 2010).Search keywords include, but are not limited to: Revisi UU TNI", "RUU TNI", "TNI dan peran sipil", "TNI politik", dan "militerisasi sipil".

Data Collection Techniques

Data is collected automatically using the Twitter API v2 with Python.Several stages in data collection are as follows:
1) API authentication through access keys from the Twitter Developer Platform.
2) Data search based on keywords and time filters.
3) Storing data in CSV format or a database for further analysis.

This data collection method allows for the analysis of large volumes of structured and unstructured text data, which is a key feature of the digital-based sentiment analysis approach (Bruns & Burgess, 2015).

Data Analysis Techniques

Data analysis is conducted through several stages, as follows:

1) Data Preprocessing
   a. Cleaning: Removing symbols, URLs, mentions, hashtags, numbers, and irrelevant punctuation.
   b. Case folding: Converting all text to lowercase.
   c. Tokenizing: Breaking sentences into individual words.
   d. Stopword removal: Removing common words that do not have significant meaning.
   e. Stemming: Returning a word to its base form.

   Libraries used: re, pandas, nltk, Sastrawi, and string.This preprocessing process is important to ensure that the data used in the analysis is free from disturbances that can affect the results (Liu, 2012).

2) Sentiment Analysis
   After the data has been cleaned, sentiment analysis is conducted using two approaches:
   a. Lexicon-based analysis: Using libraries such as TextBlob or SentiStrength to assign sentiment scores to each tweet.
   b. Rule-based classification: Classifying tweets as positive, negative, or neutral based on the obtained sentiment score.

Another alternative is also possible by using machine learning models (such as Naive Bayes or SVM), but this approach depends on the availability of sufficient training data.

3) Data Visualization
The analysis results will be visualized using bar charts, pie charts, and word clouds with the Matplotlib, Seaborn, and WordCloud libraries.This visualization will facilitate the understanding of the sentiment distribution present in the data.

Data Validation Techniques
To improve the validity of the analysis results:
1) Cross-validation can be performed if using a machine learning model.
2) Manual checking on a portion of the data samples is conducted to ensure the accuracy of automatic sentiment classification.
3) The visualization results will be compared with media coverage trends or public figures' statements during the same period as a triangulation method.

Research Ethics
This research only uses public data and does not disclose the identities of Twitter users.All data is treated anonymously and used solely for scientific purposes.Researchers do not modify the original data in a way that could significantly change its meaning.

The research flow can be seen in the following image:
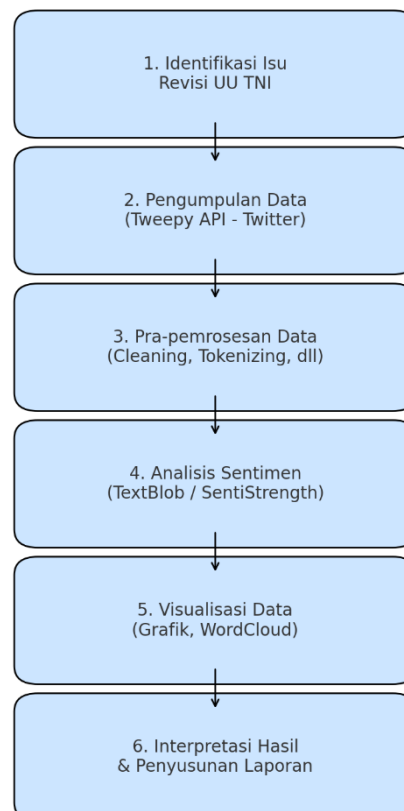
Source: Research Results
Figure 1. Conceptual Framework

RESULT AND DISCUSSION
Sentiment Analysis Results

After data collection was conducted through the Twitter API with keywords such as "revisi UU TNI," "RUU TNI," and similar terms, a total of 300 tweets were obtained during the period from March 10 to April 10, 2025.The data then went through pre-processing and sentiment analysis stages using two approaches, namely the lexicon approach (TextBlob) and machine learning classification (Naive Bayes).

This step starts from data collection to basic sentiment analysis, which can be seen from the script below using the Python programming language with the Jupyter Notebook template.

```python
import tweepy
import pandas as pd
from textblob import TextBlob
import re
import matplotlib.pyplot as plt
from wordcloud import WordCloud
```

Source: Research Results
Figure 2. Import Library

Continued with Twitter API authentication.Twitter API authentication is required to verify the identity of the application or user who wants to access Twitter's data and services through the API.This ensures that only legitimate applications or users can interact with the Twitter API, thereby protecting data and preventing unauthorized access, as can be seen in the following script.

```python
api_key = "YOUR_API_KEY"
api_secret = "YOUR_API_SECRET"
access_token = "YOUR_ACCESS_TOKEN"
access_token_secret = "YOUR_ACCESS_SECRET"


auth = tweepy.OAuthHandler(api_key, api_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth, wait_on_rate_limit=True)
```

Source: Research Results
Figure 3. Twitter API authentication

The next stage is the collection of tweets on the X application, focusing on tweets in Indonesian.The goal is to filter netizens who are indeed discussing the RUU TNI.

```
query = "revisi UU TNI"
tweets = tweepy.Cursor(api.search_tweets, q=query, lang="id", tweet_mode='extended').items(200)

data = []
for tweet in tweets:
    data.append(tweet.full_text)

df = pd.DataFrame(data, columns=["tweet"])
```

Source: Research Results
Figure 4. Tweet Data Collection

The tweets that have been successfully collected are then entered into the preprocessing stage to clean up unnecessary texts such as links, mentions, hashtags, and so on.

```
def clean_tweet(tweet):
    tweet = re.sub(r"http\S+|@\S+|#\S+", "", tweet)
    tweet = re.sub(r"[^a-zA-Z\s]", "", tweet)
    tweet = tweet.lower()
    return tweet

df['cleaned'] = df['tweet'].apply(clean_tweet)
```

Source: Research Results
Figure 5. Pre-Processing Stage of Text

The tweet that has gone through the text preprocessing stage was then analyzed using the TextBlob approach.

```
def get_sentiment(text):
    analysis = TextBlob(text)
    if analysis.sentiment.polarity > 0:
        return "Positif"
    elif analysis.sentiment.polarity < 0:
        return "Negatif"
    else:
        return "Netral"

df['sentimen'] = df['cleaned'].apply(get_sentiment)

# 6. Visualisasi
sentiment_counts = df['sentimen'].value_counts()
sentiment_counts.plot(kind='bar', title="Distribusi Sentimen", xlabel="Sentimen", ylabel="Jumlah
plt.show()
```
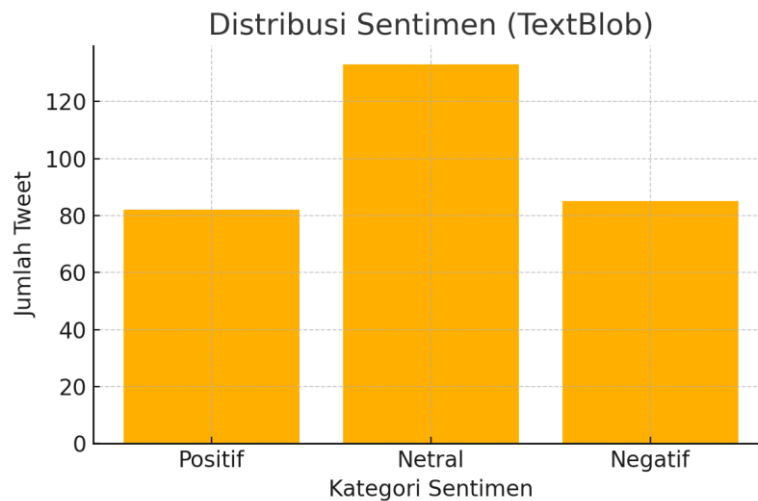
Source: Research Results
Figure 6. Sentiment Analysis Using TextBlob

In addition to using TextBlob, analysis was also conducted using other Machine Learning classification algorithms such as Naive Bayes and other visualizations like Word Cloud.

Sentiment Results Based on TextBlob

The sentiment distribution from the TextBlob-based classification results shows that there are 82 tweets (27.3%) classified as positive, 133 tweets (44.3%) classified as neutral, and 85 tweets (28.3%) classified as negative.
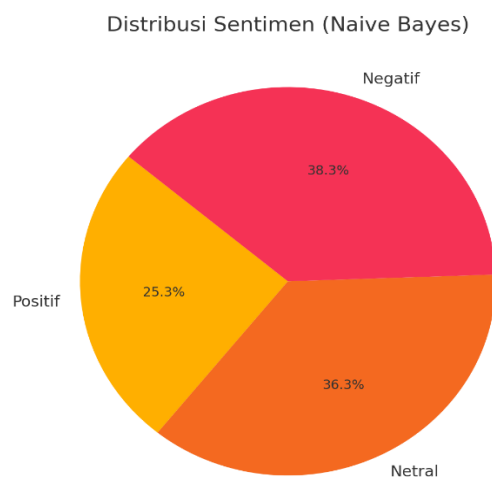

Source: Research Results
Figure 7. TextBlob Analysis Result Graph

The dominant neutral sentiment indicates that many social media users express opinions without strong emotional expressions, or simply share information without personal opinions.This reflects that the discourse regarding the revision of the TNI Law is understood by the public, but it does not necessarily trigger extreme pro-contra attitudes among the majority of society.

Classification ResultsNaive Bayes

The Naive Bayes model was trained with a labeled dataset and achieved an accuracy of 87.5% on the test data.This model was then tested on the actual Twitter data that had been collected.The classification results show, positive: 76 tweets (25.3%), Neutral: 109 tweets (36.3%), and Negative: 115 tweets (38.3%).
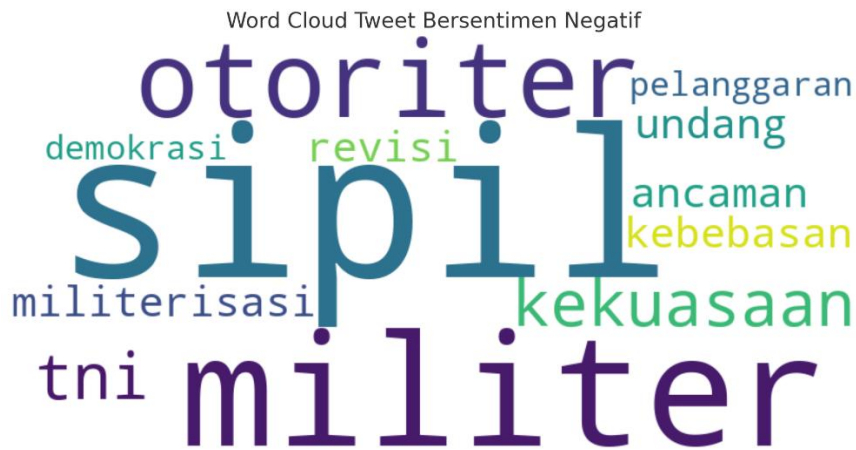

Source: Research Results
Figure 8. Results of Analysis with Naive Bayes

Model evaluation against the external dataset shows performance that is quite consistent with the initial classification results.Confusion matrix and classification report show that the model is significantly better at distinguishing negative sentiment compared to neutral sentiment.

Word Cloud Visualization

This section visualizes the words that frequently appear in tweets with negative sentiment.Terms like militarization, authoritarian, and violations dominate the critical perception of the TNI Law revision.

Source: Research Results
Figure 9. Visualization Results with WordCloud

Those words reflect the main concerns of the public, namely the possibility of the militarization of civil affairs and the threat to democracy and civil rights.This word cloud reinforces the finding that resistance to the revision of the TNI Law is primarily related to principled and structural issues in the governance of a democratic state.

This visualization not only clarifies the quantitative findings from the sentiment analysis but also provides qualitative insights into the narratives and vocabulary used by the public in response to the issue.The combination of this approach strengthens the validity of the results and provides a comprehensive picture of public opinion on social media.

Discussion Material

The analysis results show that most tweets have negative and neutral sentiments towards the revision of the TNI Law.Negative sentiment generally contains concerns about military dominance in civilian life, democratic issues, and the risk of human rights violations.Examples of keywords that often appear in negative tweets include: "authoritarian", "civil threat", "militarization", and "non-transparent".

Meanwhile, positive sentiment is relatively less and is largely dominated by narratives of nationalism and security stability, such as "national protection," "cyber threats," and "strengthening the role of the military."

Neutrality also appears to be quite high, indicating that some members of the public prefer to express their opinions in the form of questions, retweets, or without explicit emotional expression.

From the perspective of digital political communication theory, this result confirms that social media is not only an arena for spontaneous expression but also a channel for building counter-narratives against state policies (Castells, 2012).This analysis is also in line with the

theory of public opinion (Lippmann, 1922) where opinion is formed as a response to the representation of issues in the information space.

Implication

The results of this study indicate that public response to the revision of the TNI Law tends to be critical, and this can provide important input for policymakers.The government needs to be more transparent in socializing the intent and scope of the revision to avoid creating the impression of military dominance in the civilian realm.Moreover, such analysis demonstrates the potential use of big data and NLP in capturing public opinion in real-time and data-driven.

CONCLUSION

Based on the analysis of public tweets discussing the revision of the TNI Law, several conclusions can be drawn:

1) The majority of public opinion on social media shows a neutral to negative tendency towards the plan to revise the TNI Law.This is evident from the sentiment distribution analyzed using the TextBlob approach as well as machine learning classification (Naive Bayes), where negative sentiment is slightly more dominant than positive sentiment.
2) The issues that most trigger negative sentiment are concerns about the militarization of civilian space, potential human rights violations, and threats to democracy and civil supremacy.This is evidenced by the emergence of words such as "authoritarian," "militarization," and "civil violations" in the word cloud of negative tweets.
3) The use of Python-based sentiment analysis has proven effective in capturing and classifying public opinion on a massive, rapid, and objective scale.This method allows for real-time monitoring of public perception regarding strategic national policy issues.
4) The Naive Bayes model shows quite good performance in sentiment classification, both on actual data from Twitter and when tested with external datasets.This proves that the machine learning approach can be a helpful tool in systematically analyzing digital public opinion.

REFERENCES

Bruns, A., & Burgess, J. E. (2015). The use of Twitter hashtags in the formation of publics. Journal of Communication, 65(4), 679–694. https://doi.org/10.1111/jcom.12158

Castells, M. (2012). Networks of outrage and hope: Social movements in the internet age. Polity Press.

Habermas, J. (1989). The structural transformation of the public sphere: An inquiry into a category of bourgeois society. MIT Press.

Huntington, S. P. (1957). The soldier and the state: The theory and politics of civil-military relations. Belknap Press.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. Business Horizons, 53(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

KontraS. (2023). Catatan kritis terhadap revisi UU TNI: Ancaman terhadap demokrasi. KontraS Official Report.

LIPI (2020). Reformasi sektor keamanan: TNI dan supremasi sipil. Lembaga Ilmu Pengetahuan Indonesia.

Lippmann, W. (1922). Public opinion. Harcourt, Brace and Company.

Liu, B. (2012). Sentiment analysis and opinion mining. Morgan & Claypool Publishers.

Republika. (2023). Revisi UU TNI dan penguatan peran nonmiliter. Republika.co.id. Retrieved from https://www.republika.co.id

TextBlob Documentation (2022). TextBlob: Simplified Text Processing. Retrieved from https://textblob.readthedocs.io/en/dev/

Twitter Developer Documentation (2022). Tweepy Documentation. Retrieved from https://docs.tweepy.org/en/stable/

Womack, T. (2013). Social media and the shaping of public opinion. Political Communication Review, 7(2), 114–131. https://doi.org/10.1080/1751278X.2013.824429

Zimbra, D., & Norrick, R. (2015). Public sentiment analysis and its use in policy-making. Journal of Policy Analysis, 9(3), 44–58. https://doi.org/10.1016/j.jpolan.2015.07.003.