**Jurnal Teknik Informatika (JUTIF)**
P-ISSN: 2723-3863| E-ISSN: 2723-3871 Vol. 6, No. 6, December 2025, Page. 1530-1537
https://jutif.if.unsoed.ac.id DOI: https://doi.org/10.52436/1.jutif.6.6.3540

# Geo-Sentiment Analysis of Public Opinion of X Users towards the Documentary Film Dirty Vote using the Bidirectional Long Short-Term Memory Method

**Syifa Salsabila*[1], Yuliant Sibaroni[2], Sri Suryani Prasetiyowati[3]**

[1,2,3] School of Computing, Telkom University, Indonesia
Email: [1]SyifaSalsabilaa@student.telkomuniversity.ac.id, [2]yuliant@telkomuniversity.ac.id, [3]srisuryani@telkomuniversity.ac.id

Received : Jun 9, 2025; Revised : Nov 20, 2025; Accepted : Nov 22, 2025; Published : Dec 11, 2025
Phone Number : 081381682478 (**Whatsapp/Cellphone Number, For Corresponding**)
The cellphone number is only for ease of communication and will NOT be displayed in the article

**Abstract**

Presidential elections held every five years, often generates significant public discourse. The 2024 presidential election saw the release of the documentary Dirty Vote, which raised allegations of electoral fraud and sparked polarized opinions on social media, especially on X. This study aims to analyze public sentiment toward Dirty Vote using geo-sentiment analysis and the Bidirectional Long Short-Term Memory (Bi-LSTM) model. Data were collected from geotagged tweets, with sentiment classified as positive, negative, or neutral. The research explored various data processing techniques, including TF-IDF for feature extraction, FastText for feature expansion, and balancing methods like SMOTE and class weighting to address data imbalance. Results showed that the baseline Bi-LSTM model achieved an accuracy of 71.57% and an F1-Score of 74.05%. When enhanced with TF-IDF and FastText, accuracy increased to 77.07%, though the F1-Score dropped slightly to 72.95%. Applying SMOTE resulted in a decrease in accuracy to 76.45%, but significantly improved the F1-Score to 74.93%. Exploratory data analysis revealed that negative sentiment was most concentrated in Java Island, particularly Jakarta, and peaked during February 2024, coinciding with the documentary's release and the election period. This study significantly contributes to understanding how geographic locations influence public opinion on sensitive political issues. A lack of understanding of geographically-based sentiment patterns can hinder identifying regional needs, leading to poorly targeted policies. By integrating data analysis methods with geographical approaches, this research provides deep insights for designing more effective, data-driven public intervention strategies and supports policymaking that is more responsive to the dynamics of public opinion.

***Keywords :*** *Bi-LSTM, Class Weight, Dirty Vote, FastText, Geo-sentiment analysis, SMOTE, TF-IDF*

**This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License**

## 1. INTRODUCTION

Indonesia has adopted a democratic system since its independence in 1945. One of the key procedures in this democracy is the General Election (Pemilu), which holds significant meaning. The purpose of holding elections is to ensure that the transfer of power occurs safely and orderly, to uphold the sovereignty of the people, and to protect the human rights of every citizen [1], [2]. The presidential and vice-presidential elections are held every five years, making them a popular public event that is always widely discussed. Various perspectives emerge from different segments of society regarding the presidential and vice-presidential candidates. Amidst the festive atmosphere of this democratic celebration, the release of the documentary Dirty Vote on February 11, 2024, during the election's quiet period, was like a bomb exploding in the middle of the celebration. This film raises issues regarding alleged fraud in the 2024 Indonesian Presidential Election media Indonesia. The documentary's release sparked both support and opposition, generating diverse public opinions on social media, especially on

P-ISSN: 2723-3863| E-ISSN: 2723-3871 Vol. 6, No. 6, December 2025, Page. 1530-1537
https://jutif.if.unsoed.ac.id DOI: https://doi.org/10.52436/1.jutif.6.6.3540

X, which plays a significant role in shaping political views and broadening public discussions on political issues [3].

To understand the reviews of the documentary Dirty Vote on social media X, a sentiment analysis approach can be used to analyze the opinions and emotions contained in the sentences [4] and to uncover the sentiment polarity in the text, whether it is positive, neutral, or negative [5].

There has been extensive research on sentiment analysis [6], [7], [8], which has become a rapidly developing research topic, with many studies exploring its various aspects [9], [10]. One such study [11] by Sinanto et al. compared feature selection algorithms in sentiment analysis of movie reviews. Another study [12] analyzed sentiment regarding the 2024 election using Twitter data, employing the Naïve Bayes algorithm. Additionally, several previous studies have applied sentiment analysis using deep learning methods, which utilize layered architectures to handle complex and multidimensional data [13], [14]. For example, research by [15], [16] and [17] successfully demonstrated high accuracy using deep learning models such as Bi-LSTM. Overall, Bi-LSTM has the advantage of combining information from both the past and the future to understand context more comprehensively, thereby improving accuracy in natural language processing tasks [18], [19].

Another study [20] also conducted sentiment analysis on the performance of the General Election Commission (KPU) ahead of the 2024 election. Most previous studies tend to focus on sentiment analysis without incorporating the geographical dimension. To gain a more comprehensive understanding of public sentiment, there is a need to expand sentiment analysis by considering the geographic aspect, known as geo-sentiment analysis. This analysis can provide deep insights into how geographic location influences public sentiment [21].

Several researchers have demonstrated the potential of geo-sentiment analysis in providing deeper insights into public opinion [22], [23]. For example, in a study by Mushofy et al. [24], sentiment analysis was focused on COVID-19 vaccination using geo-tagged Twitter data, and the results showed that the dominant sentiment was positive, concentrated in the Karawang region. Tau Hu et al. [25] conducted another study that used data to analyze sentiment using a variety of methods, such as temporal and geographic analysis and the Local Indicators of geographic Association (LISA) method to find spatial clusters. Additionally, they classified geo-tweets using the Latent Dirichlet Allocation (LDA) model. Through geo-sentiment analysis, insights can be gained regarding the differences in sentiment distribution patterns based on users' geographic locations [25]. This also helps in understanding the geographical impact on users' emotions in a specific context [26].

This study addresses a gap in sentiment analysis research, which has predominantly focused on non-geographical aspects. While the majority of prior research has employed machine learning and deep learning techniques to examine public sentiment on social media, the geographical dimension, which significantly influences public perception, remains underexplored. By implementing geo-sentiment analysis, this research provides insights into how geographic location shapes sentiment toward the documentary "Dirty Vote" and identifies specific sentiment patterns across various regions. Additionally, this study enhances sentiment analysis methodologies by integrating the Bi-LSTM model with data imbalance handling techniques, such as class weight and SMOTE, to improve sentiment analysis performance on complex and imbalanced datasets.

The problem formulation of this study is to evaluate the performance of the Bi-LSTM model in analyzing geo-sentiment from Twitter data related to the Dirty Vote documentary, and to understand the geographical distribution of public sentiment toward the film across various regions in Indonesia. This study is limited to sentiment analysis of public opinion on Twitter regarding Dirty Vote using the Bi-LSTM method. The analysis focuses on the collection of geolocation data from Twitter users, extraction of relevant tweets about Dirty Vote, and sentiment classification (positive, negative, or neutral) using the Bi-LSTM model.

P-ISSN: 2723-3863| E-ISSN: 2723-3871 Vol. 6, No. 6, December 2025, Page. 1530-1537
https://jutif.if.unsoed.ac.id DOI: https://doi.org/10.52436/1.jutif.6.6.3540

The objective of this study is to analyze the geo-sentiment of X users toward the Dirty Vote documentary using the Bi-LSTM model, and to assess the effectiveness of the model in sentiment analysis classification. This study also aims to compare the effectiveness of various data imbalance handling techniques, such as the use of class weight and SMOTE, in improving the performance of the Bi-LSTM model on geo-sentiment data. Furthermore, this research seeks to deepen the understanding of how geographic location influences public sentiment, contributing to the development of geo-sentiment analysis methodologies. The findings of this study are expected to provide practical insights for political communication strategies and the analysis of public opinion on socially sensitive issues.

## 2. METHOD

Figure 1 shows the system design in this study, which aims to provide a clear depiction of the system developed.

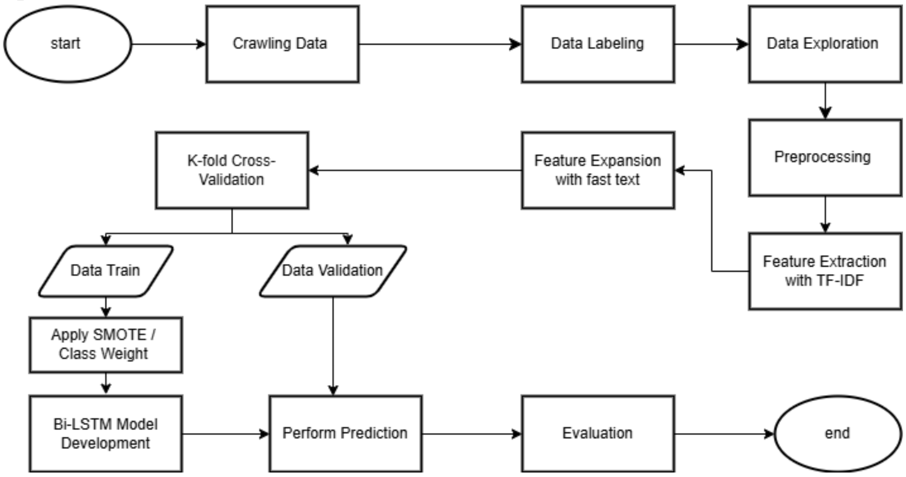


Figure 1. System Design Flowchart

### 2.1. Crawling Data

The data for this study were collected from tweets by users of the X application. The crawling process was carried out using the open-source Application Programming Interface (API) provided by the X application, with Python programming language. The collected tweets were from X users containing geolocation information from various regions, along with keywords or topics related to Dirty Vote.

### 2.2. Data Labeling

The collected data will then undergo labeling. Data labeling will be performed on each tweet and classified into several sentiment label categories: "Positive", "Negative", and "Neutral", with the aim of identifying the sentiment of each text in the dataset [24]. Data Exploration

The data collected in this study can be further analyzed to gain a better understanding of the sentiment and the geographic distribution of public opinion regarding the Dirty Vote documentary. The data exploration involves calculating the percentage of tweets with positive, negative, and neutral sentiments, as well as comparing sentiment across geographic regions. A histogram will be used to visualize the number of tweets with positive, negative, and neutral sentiments. As in the study by [27], a distribution map helps to understand the spread of sentiment across different regions. This visualization identifies sentiment distribution patterns and allows for comparison of sentiment across geographic areas.

### 2.3. Preprocessing