

Sentiment Analysis of 2024 Presidential Candidates in Indonesia: Statistical Descriptive and Logistic Regression Approach

1st Fadlan Akmal Ramadhan
School of Computing
Telkom University
Bandung, Indonesia
fadlanar@student.telkomuniversity.ac.id

2nd Putu Harry Gunawan
School of Computing
Telkom University
Bandung, Indonesia
phgunawan@telkomuniversity.ac.id

Abstract— *Sentiment analysis is a technique for analyzing public sentiment or a particular segment of sentiment toward an object. One of the implementations of sentiment analysis is toward potential presidential and vice-presidential candidates in the upcoming 2024 Indonesian Presidential Election (Pilpres). This upcoming Presidential Election is a topic that is frequently discussed in Indonesian politics, by voters, various political parties, and people's community daily conversation anywhere. Many prospective candidates are rumored to be supported by various political parties to become their presidential and vice-presidential candidates, without any concern about their reputation and sentiment among targeted voters. However, the sentiment towards these candidates was not always clear, whether it leans towards positive or negative. To address this issue, this sentiment analysis uses Logistic Regression and Descriptive Statistics to do classification and analysis of the sentiment of tweets regarding each of the prospected candidates. The result of the sentiment analysis was a polarity score of tweets for these prospective candidates. The Logistic Regression method has an overall average accuracy score of 79% on all five prospected 2024 presidential candidate datasets. The highest accuracy is obtained at 93% on the Prabowo Subianto dataset and the lowest at 70% on the Anies Baswedan and Muhaimin Iskandar datasets. This analysis also provides valuable insights into the public sentiment toward potential candidates for the 2024 Presidential Election in Indonesia.*

Keywords—*sentiment analysis, 2024 presidential election, logistic regression, statistical descriptive, classification*

I. INTRODUCTION

In this modern era, the influence of advanced technologies on various fields is inevitable and cannot be denied, especially in the communication field. communication field was one of the fields that has the fastest technological developments, ranging from service providers, communication devices, to the media that serves as a communication and socialization platform. Social Media has become a place for people to socialize instantly, practically, and at a low cost. One of the social media platforms that has many users is Twitter, a text-based social media platform that allows users to post and interact with messages known as “tweets”. The presence of Twitter as a social media platform for posting tweets in the form of opinions, debates, and points of view can be advantageous for parties that have some interests. In political terms, politicians tend to be more honest on social media rather than in talk shows or news interviews [1] and it has become popular to provide a stable political movement [2].

Sentiment analysis is a popular method that is used to gather information on social media platforms such as judgments, preferences, or people's feelings and opinions for an entity [3], in the form of positive or negative sentiment.

Sentiment analysis can be implemented in various fields and social media platforms, ranging from cryptocurrency price [4], COVID-19 vaccination responses [5], COVID-19 conspiracy theories [6], government policies [7], movie reviews [8], stock price predictions [9], to political issues discussed on Twitter [10]. Twitter has emerged as an influential platform, especially for politicians [11] as a way to campaign through digital communication technologies [12]. The political issues are closely related to the opinions, reputations, and sentiment of those with particular interest, such as political parties, legislative candidates, and presidential candidates. Through sentiment analysis, these parties can easily obtain information on the reputation and sentiment of their targeted supporters or segments. The issue of the 2024 presidential election is currently being widely discussed in Indonesia. According to the reputable social media analysis website, Drone Emprit (<https://pers.droneemprit.id>), there are several strong potentials and prospected presidential candidates based on their popularity on Twitter, including Anies Baswedan (105.110 tweets), Puan Maharani (30.475 tweets), Prabowo Subianto (9,500 tweets), Ridwan Kamil (21.631 tweets), and Muhaimin Iskandar (6.480 tweets).

There are multiple binary classification methods in sentiment analysis. In study [13] Naïve-Bayes, Support Vector Machine, Decision Trees, and Stochastic Gradient Descent are used as the method for classification in movie reviews, while the study [8] only use Support Vector Machine for the same purpose. Study [14] uses Logistic Regression algorithm for performing sentiment analysis, specifically for sentiment detection on Twitter. Meanwhile, study [15] uses Logistic Regression to predict cardiovascular disease and analyze the factors of cardiovascular disease.

In a related political election topic research, study [10] focused on analyzing the data without performing any classifications. Study [16] only conducted a comparative approach between the use of Twitter and Instagram during the 2015 Norwegian Elections. Meanwhile, study [17] focused solely on sentiment analysis using Naïve Bayes without performing any descriptive statistical analysis. Study [18] specifically focuses on analyzing the sentiment towards 2020 election during the COVID-19 pandemic using Naïve Bayes classifier and Support Vector Machine without explicitly mentioning the candidates. On the other hand, study [19] focused solely on analyzing the tweet data based on tweet polarity without discussing or providing the prediction result.

The aim of this study is to conduct sentiment analysis on potential and prospective candidates for the 2024 Presidential Election using the Statistical Descriptive and Logistic