Buzzer Account Detection in Political Hate Tweets Using IndoBERT and Ensemble Learning: Case Study of the Indonesian Presidential Election 2024

Fizio Ramadhan Herman[1*], Ade Romadhony[2]

[1,2]*School of Computing, Telkom University,
Bandung, Indonesia*
*fizioramadhan@gmail.com

Abstract

The Indonesian Presidential Election of 2024 has seen a widespread use of social media such as Twitter for political campaigning and discussion. However, this has also enabled the spread of hate speech from buzzer accounts that are created to influence public opinions. This study implements a machine learning approach to classify buzzer accounts that are spreading hate speeches during the presidential election period. By utilizing IndoBERT for hate speech classification and a traditional machine learning model to classify buzzer accounts. This study analyzes 62,341 tweets for hate speech classification and 961 accounts for buzzer account classification. Our implementation of IndoBERT achieved a strong performance with 91.12% of precision and recall, and 91.19% accuracy and F1-score in hate speech classification. While for buzzer account classification, we compared Decision Tree, Random Forest, and XGBoost, with Decision Tree achieving the highest performance of 64% precision, recall, accuracy, and F1-Score. Our results demonstrate the effectiveness of combining deep learning for hate speech classification with traditional machine learning for buzzer account classification, contributing to the development of more effective content filtering for election discourse on social media.

Keywords: buzzer detection, ensemble learning, IndoBERT, presidential election, sentiment analysis, social media

I. INTRODUCTION

The 2024 Indonesian Presidential Election has captured the attention of Indonesians for these past few months. Since late November 2023, there has been a growing discussion regarding this topic, especially on the internet, due to the political party campaigns have started. Social media such as Twitter and Facebook, play a significant role in politics [1] by providing spaces for individuals to share views and engage in political discussions [2]. As the campaign season intensifies, the use of social media by Indonesian political parties to promote an advertise candidates is one example that demonstrates the role that social media occupies in politics. However, the open nature of Twitter can both encourage informed political discourse and create opportunities for the dissemination of hate speech, which could potentially shape public opinions about presidential candidates.

Increasing amount of political hate speech on Twitter has emerged as a notable concern. Based on a study conducted by Mozafari et. al [3], hate speech is commonly defined as any form of communication that involves the criticism of an individual or a group based on some characteristics such as gender, nationality, religion, race, etc. It is important to consider the potential consequences of this issue, as it could lead to division within communities, incite violence, and undermine the integrity of fair elections. In order to tackle this problem, the proposed research seeks to develop a robust deep learning approach to detect and identify buzzer accounts that are specifically created or used to spread political hate speech during the 2024 Indonesian Presidential Election.

While previous studies have explored the role of Twitter in political polarization [2] the influence of e-campaigns on voters [4] [5] and the detection of malicious social bots [6], there is a need for a comprehensive study focused on the detection of buzzer accounts actively engaged in spreading political hate speech during the Indonesian Presidential Election. Previous research has also demonstrated significant progress in buzzer detection. Suciati et al. [1] evaluated multiple ensemble learning algorithms for buzzer detection, achieving 62.3% accuracy and 61.3% precision using AdaBoost, while Ibrahim et al. [8] developed an automatic buzzer detection system combined with sentiment analysis to predict election results with a mean absolute error of 0.61%. However, while these studies have broadened our understanding of buzzer detection and political sentiment analysis separately, these leaves a gap in the integrated study of hate speech-spreading buzzer accounts. A detailed review of related studies is provided in Section 2.

This research aims to bridge this gap by leveraging ensemble learning for buzzer detection and deep learning approaches, specifically IndoBERT, for hate speech classification. Through collaboration with *NoLimit Indonesia*, we analyze a dataset of 62,341 tweets and 961 accounts to evaluate our proposed approach in identifying buzzer accounts that spread hate speeches during election periods. The study focuses on developing a framework that integrates ensemble learning and IndoBERT, assessing its effectiveness on a real Twitter dataset collected during the election period, and establishing quantitative benchmarks for identifying hate speech-spreading buzzer accounts. Additionally, the dataset used in this research is expected to serve as a resource for future studies on hate speech classification and buzzer account detection.

II. LITERATURE REVIEW

This section will analyze the recent research carried out in the field of buzzer account detection and hate speech classification. We will study this body of work by splitting it into several components to conduct a comprehensive review.

A. Buzzer Account

1) Definition of Buzzer Account: According to a study conducted by Rismi Juliadi in 2017 [9], a "buzzer" on Twitter is anticipated to spread information that will have an impact on their followers. The widespread adoption of buzzers can be understood through the lens of the Diffusion of Innovation Theory by Everett Rogers [10]. This theory explains that communicators or messengers who get information from the mass media have the
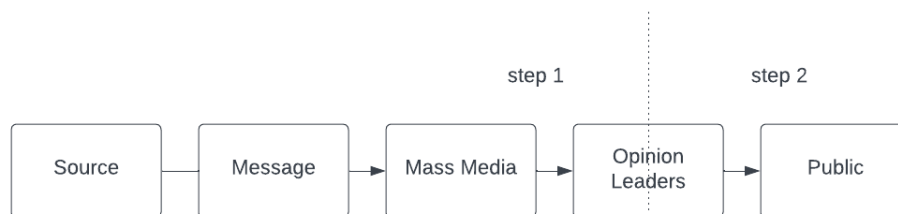

Fig. 1. Two Step Flow Model (Katz & Lazarsfeld, 1955)