

## INTISARI

### **Analisis Sentimen pada Data Tidak Seimbang Menggunakan Model BERT: Studi Kasus pada Opini Publik Mengenai Pasangan Calon Presiden dan Wakil Presiden Indonesia 2024 di Media Sosial X**

Oleh

Sania Rizka Ramadhani

20/459373/PA/20034

Pemilihan umum merupakan momen penting bagi negara demokratis seperti Indonesia, dengan setiap masyarakat memiliki peran penting dalam membentuk arah perubahan masa depan. Penelitian ini bertujuan untuk menganalisa pendapat publik terkait dengan pasangan calon presiden dan wakil presiden yang maju dalam pemilihan umum 2024 dengan menggunakan model BERT. Data yang digunakan diambil dari media sosial X yang dikumpulkan menjelang pemilihan umum. Selanjutnya, dilakukan evaluasi terhadap performa hasil klasifikasi model dan dibandingkan dengan model Bi-LSTM sebagai tolok ukur.

Berdasarkan hasil pembentukan model BERT dengan kombinasi antara ratio pembagian data, ukuran *batch size*, *epoch*, dan nilai *learning rate*, diperoleh bahwa model BERT memiliki performa paling baik dengan nilai akurasi 95,69%, presisi 95,68%, *recall (weighted)* 95,69%, dan *F1 score (weighted)* sebesar 95,68%. Sementara itu, hasil yang diperoleh dari model Bi-LSTM adalah akurasi 81,30%, presisi 81,09%, *recall (weighted)* 81,30%, dan *F1 score (weighted)* sebesar 81,13%.

## **ABSTRACT**

### **Sentiment Analysis on Imbalanced Data Using BERT Model: Case Study on Public Opinions Regarding the Indonesian 2024 Presidential and Vice Presidential Candidates on Social Media X**

By

Sania Rizka Ramadhani

20/459373/PA/20034

General elections are crucial moments for democratic countries like Indonesia, with every community playing a significant role in shaping the direction of future changes. This research aims to analyze the opinions of public regarding the presidential and vice-presidential candidates participating in the 2024 general elections using the BERT model. Data utilized are sourced from social media platform X, collected prior to the general elections. Subsequently, the performance of the classification model is evaluated and compared with the Bi-LSTM model as a benchmark.

Based on the results of constructing the BERT model with a combination of data split ratio, batch size, epoch, and learning rate values, it was found that the BERT model performed best with an accuracy of 95.69%, precision of 95.68%, weighted recall of 95.69%, and weighted F1 score of 95.68%. Meanwhile, the results obtained from the Bi-LSTM model were an accuracy of 81.30%, precision of 81.09%, weighted recall of 81.30%, and a weighted F1 score of 81.13%.