

ABSTRAK

ANALISIS SENTIMEN TERHADAP BAKAL CALON PRESIDEN 2024 DI INDONESIA MENGGUNAKAN DATA TWITTER DENGAN ALGORITMA NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE

Oleh

ATIQA HANIFAH SHALIAH

Indonesia merupakan negara demokrasi, hal ini ditunjukkan dengan adanya pemilihan presiden yang dilakukan setiap lima tahun sekali. Pada tahun 2024 Indonesia kembali akan melaksanakan pilpres, tentunya ini menjadi topik yang hangat diperbincangkan oleh masyarakat, khususnya mengenai bakal calon presiden yang akan mencalonkan diri. Penelitian ini bertujuan untuk melakukan klasifikasi *tweet* dan melakukan analisis sentimen terhadap opini yang disampaikan oleh masyarakat melalui twitter menggunakan *metode Cross-Industry Standard Process for Data Mining (CRISP-DM)*. Model klasifikasi ini dibangun dengan menggunakan algoritma *machine learning* yaitu *Naïve Bayes Classifier* dan *Support Vector Machine*. Penelitian ini berfokus pada dua kandidat calon presiden yaitu Anies Baswedan dan Ganjar Pranowo. Dataset pada penelitian ini terdiri atas 3000 data *tweet* pada masing-masing dataset dengan pembagian kelas pada data yang tidak seimbang, dimana persentase distribusi untuk label positif sebesar 46.60%, netral sebesar 16.10 %, dan negatif sebesar 37.30% untuk dataset Anies Baswedan. Sementara pada dataset tokoh Ganjar Pranowo memiliki persentase distribusi untuk label positif sebesar 58.83%, netral sebesar 16.77%, dan negatif sebesar 24.40%. Hasil penelitian menunjukkan bahwa dengan menggunakan metode evaluasi *classification report* algoritma *naïve bayes classifier* memiliki performa yang lebih tinggi pada dataset Ganjar Pranowo dengan nilai akurasi sebesar 92% dan nilai weighted avg *f1-score* sebesar 92% , lalu pada dataset Anies Baswedan diperoleh nilai akurasi sebesar 87% dengan nilai weighted avg *f1-score* sebesar 87%. Sementara pada algoritma SVM diperoleh nilai akurasi sebesar 88% untuk dataset Anies Baswedan dengan nilai weighted avg *f1-score* sebesar 88% dan diperoleh nilai akurasi sebesar 82% dengan nilai weighted avg *f1-score* sebesar 80% untuk dataset Ganjar Pranowo. Hasil tersebut menjadikan bahwa algoritma *naïve bayes classifier* cenderung memiliki performa yang lebih baik daripada algoritma SVM dalam melakukan pengklasifikasian sentimen. Selain itu hasil analisis data menunjukkan bahwa opini yang diberikan oleh masyarakat pada masing-masing dataset memiliki presentase sentimen positif yang cukup besar hal ini menunjukkan bahwa tokoh Anies Baswedan dan Ganjar Pranowo memperoleh respon yang baik dari masyarakat.

Kata Kunci: Analisis Sementimen, Calon Presiden, *Naïve Bayes Classifier*, *Support Vector Machine*, CRISP-DM, Twitter

ABSTRACT

SENTIMENT ANALYSIS ON 2024 PRESIDENTIAL CANDIDATES IN INDONESIA USING TWITTER DATA WITH NAÏVE BAYES CLASSIFIER AND SUPPORT VECTOR MACHINE ALGORITHMS.

By

ATIQA HANIFAH SHALIAH

Indonesia is a democratic country, this is shown by the presidential election that is held every five years. In 2024, Indonesia is set to hold another presidential election, which has become a hot discussed topic among the public, particularly regarding the potential candidates who will be running for office. Many people voice their thoughts and opinion regarding the presidential candidates vying for office by means of tweets on the popular social media platform, Twitter. This research aims to classify tweets and conduct sentiment analysis on public opinion using the Cross-Industry Standard Process for Data Mining (CRISP-DM) method. This classification model is built using machine learning algorithms, namely Naïve Bayes Classifier and Support Vector Machine. This study focuses on two presidential candidates, Anies Baswedan and Ganjar Pranowo. The dataset contains 3000 tweets in each dataset with an imbalanced class distribution. The percentage distribution for the labels in the Anies Baswedan dataset is 46.60% positive, 16.10% neutral, and 37.30% negative. Meanwhile, the percentage distribution for the labels in the Ganjar Pranowo dataset is 58.83% positive, 16.77% neutral, and 24.40% negative. The results show that using classification report evaluation method, the naïve bayes classifier algorithm has a higher performance on the Ganjar Pranowo, achieving an accuracy of 92%. and a weighted average f1-score of 92%. Meanwhile, for the Anies Baswedan dataset, an accuracy of 87% with a weighted average f1-score of 87% was obtained. For the SVM algorithm, an accuracy of 88% was obtained for the Anies Baswedan dataset with a weighted average f1-score of 88%, and an accuracy of 82% with a weighted average f1-score of 80% was obtained for the Ganjar Pranowo dataset. Based on the results, it can be observed that the naïve bayes classifier algorithm outperforms the SVM algorithm in sentiment classification. Furthermore, the data analysis reveals that the public's opinions on each dataset have a fairly large percentage of positive sentiments. This indicates that Anies Baswedan and Ganjar Pranowo figures have garnered a favorable response from the public.

Keywords: Sentiment Analysis, presidential candidates, Naïve Bayes Classifier, Support Vector Machine, CRISP-DM, Twitter