

ABSTRAK

Dalam penelitian ini akan dilakukan analisis sentimen publik terhadap cryptocurrency (Bitcoin) yang disampaikan melalui media sosial Twitter. Ada beberapa tahapan yang dilakukan dalam analisis sentimen diantaranya adalah tahap pengumpulan data, labeling, sampling, preprocessing, pembobotan kata, klasifikasi dan evaluasi. Proses pengumpulan data menghasilkan dataset awal sebanyak 71.430 data berupa tweet berbahasa Inggris. Pada proses labeling data dilakukan menggunakan pemrograman python dengan bantuan library textblob, hasil labeling data terdiri dari 3 kelas yaitu positif, negatif dan netral. Pada proses preprocessing yang bertujuan untuk membersihkan data terdiri dari 6 tahapan yaitu cleansing, case folding, normalisasi kata, tokenizing, stop word removal dan lemmatization. Metode klasifikasi yang digunakan dalam penelitian ini adalah komparasi antara algoritma Naïve Bayes, Logistic Regression, dan Support Vector Machine (SVM). Ketiga algoritma tersebut dikombinasikan dengan fitur pembobotan kata menggunakan metode TF-IDF. Data yang digunakan pada tahap klasifikasi sebanyak 14.000 data tweet (setelah melalui proses sampling dan penghapusan kelas netral) yang terdiri dari label positif dan label negatif yang kemudian dibagi menjadi 2 yaitu untuk data latih sebanyak 11.200 data dan data uji sebanyak 2.800 data. Pengujian dan evaluasi dilakukan menggunakan metode confusion matrix di mana algoritma Support Vector Machine (SVM) keluar sebagai model terbaik yang menghasilkan nilai akurasi mencapai 85,82% dan F1-Score sebesar 85,34%.

Kata Kunci : *Support Vector Machine, Logistic Regression, Naïve Bayes, Analisis Sentimen, Bitcoin*

ABSTRACT

This study conducts a public sentiment analysis on cryptocurrency (Bitcoin) expressed through the social media platform Twitter. Several stages are carried out in the sentiment analysis process, including data collection, labeling, sampling, preprocessing, word weighting, classification, and evaluation. The data collection process resulted in an initial dataset of 71,430 English-language tweets. The data labeling process was performed using Python programming with the assistance of the TextBlob library, producing three sentiment classes: positive, negative, and neutral. The preprocessing stage, which aims to clean the data, consists of six steps: cleansing, case folding, word normalization, tokenizing, stop-word removal, and lemmatization. The classification methods used in this research involve a comparison of three algorithms: Naïve Bayes, Logistic Regression, and Support Vector Machine (SVM). These algorithms are combined with a word weighting feature using the TF-IDF method. The dataset used in the classification stage consists of 14,000 tweet data (after sampling and removal of the neutral class), comprising positive and negative labels. The dataset is then divided into two parts: 11,200 data for training and 2,800 data for testing. Testing and evaluation are conducted using the confusion matrix method, where the Support Vector Machine (SVM) algorithm emerges as the best-performing model, achieving an accuracy of 85.82% and an F1-Score of 85.34%.

Keyword: *Support Vector Machine, Logistic Regression, Naïve Bayes, Sentimen Analysis, Bitcoin*

KARAWANG