

ABSTRAK

Dalam penelitian ini akan dilakukan analisis sentimen publik terhadap *cryptocurrency* yang disampaikan melalui media sosial Twitter. Ada beberapa tahapan yang dilakukan dalam analisis sentimen diantaranya adalah tahap pengumpulan data, *preprocessing*, *labeling*, pembobotan kata, klasifikasi dan evaluasi. Proses pengumpulan data dengan teknik *scraping* menghasilkan data sebanyak 10000 data berupa *tweet* berbahasa Indonesia yang diambil dari bulan juni tahun 2020 hingga bulan juni tahun 2022 menggunakan *library* python bernama *snsrape*. Pada proses *preprocessing* yang bertujuan untuk membersihkan data terdiri dari 6 tahapan yaitu *remove duplicate*, *case folding*, *cleansing*, normalisasi kata, *tokenizing*, *stopword removal* dan *lemmatization*. Pada proses *labeling* data dilakukan menggunakan pemrograman python dengan bantuan *library* *textblob*, hasil *labeling* data terdiri dari 3 kelas yaitu positif, negatif dan netral. Metode klasifikasi yang digunakan dalam penelitian ini adalah Naive Bayes. Algoritma Naive Bayes dikombinasikan dengan fitur pembobotan kata menggunakan metode TF-IDF dalam satu kelas *pipeline* python. Data yang digunakan pada tahap klasifikasi sebanyak 5000 data *tweet* yang terdiri dari 2500 label positif dan 2500 label negatif yang kemudian dibagi menjadi 2 yaitu untuk data latih sebanyak 4000 data dan data uji sebanyak 1000 data. Pengujian dan evaluasi dilakukan menggunakan metode *confusion matrix* dengan algoritma naive bayes yang menghasilkan nilai akurasi mencapai 89%.

Kata Kunci : *Cryptocurrency*, Klasifikasi, Analisis sentimen, Naive bayes



KARAWANG

ABSTRACT

This study conducted an analysis of public sentiment on Twitter towards cryptocurrencies. There were several stages carried out in sentiment analysis starting from data collection, preprocessing, labelling, word weighting, classification and evaluation. The process of collecting data using the scraping technique produced 10000 data in the form of Indonesian-language tweets taken from June 2020 to June 2022 using a python library called snsrape. The preprocessing process which aims to clean the data consists of 6 stages, namely removing duplicates, case folding, cleansing, word normalization, tokenizing, stop word removal and lemmatization. The data labelling process is carried out using python programming with the help of the textblob library, the data labelling results consist of 3 classes, namely positive, negative and neutral. The classification method used in this study is Naive Bayes. The Naive Bayes algorithm is combined with the word weighting feature using the TF-IDF method in one python pipeline class. The data used in the classification stage is 5000 tweet data consisting of 2500 positive labels and 2500 negative labels which are then divided into 2, namely 4000 data for training and 1000 test data. Tests and evaluations are carried out using the confusion matrix method with the naive Bayes algorithm which produces an accuracy value of 89%

Keyword : Cryptocurrencies, Classification, Sentiment analysis, Naive bayes

