

ABSTRAK

Cryptocurrency merupakan mata uang digital yang saat ini terkenal dan banyak digunakan dalam berbagai aktivitas masyarakat. Bitcoin sebagai salah satu *cryptocurrency* terus meningkat. Fluktuasi harga *cryptocurrency*, termasuk Bitcoin, dipengaruhi oleh pasar dan opini publik (Zhang et al., 2018). Twitter menyampaikan bahwa lebih dari 1 miliar *tweets* tentang *cryptocurrency* sejak 2020 dan menjadi platform utama untuk diskusi *cryptocurrency*. Terdapat korelasi antara sentimen di media sosial dengan fluktuasi harga *cryptocurrency*, seperti yang ditunjukkan oleh penelitian (ElBahrawy et al., 2019) yang menyimpulkan bahwa aktivitas di Twitter berpengaruh pada fluktuasi harga *cryptocurrency*, serta penelitian (Kristoufek, 2015) yang menemukan korelasi antara sentimen di Twitter dengan fluktuasi harga Bitcoin.

Metode *machine learning* yang digunakan dalam usaha untuk melakukan analisa fenomena yang dipaparkan sebelumnya salah satunya pendekatan *hybrid* dan analisis regresi linier sederhana. Pendekatan *hybrid* ini menggabungkan *Text blob* dan *Naïve Bayes* serta analisis regresi linier sederhana untuk mencari korelasi pergerakan harga *cryptocurrency* dengan sentimen di sosial media Twitter. Menggunakan analisis regresi linier sederhana, hubungan antara sentimen dan fluktuasi harga *cryptocurrency* diukur. Pengumpulan data akan dapat dilakukan dari platform twitter untuk data sentimen dan CoinGecko untuk data *marketcap crypto* dengan melakukan scraping data melalui twitter *developer* dan API CoinGecko. Data yang dikumpulkan akan melalui tahap preprocessing lalu masuk ke tahap pembuatan model dengan *Textblob* lalu dilanjutkan dengan pelatihan menggunakan *Multinomial Naïve Bayes*. Tahapa selanjutnya normalisasi dengan menggunakan *min max normalization* dengan tujuan agar meminimalisir kesalahan komputasi. Tahapan dilanjutkan dengan analisis data menggunakan *Analisis Regresi Linear*.

Pembuatan model klasifikasi dilakukan menggunakan *Textblob* lalu akan di latih menggunakan *Multinomial Naïve Bayes*. Dari tahapan tersebut akan mendapatkan hasil keefektifan dari penggunaan metode tersebut dengan menggunakan evaluasi klasifikasi *confussion matrix*. Setelah itu akan di implementasikan metode Analisis Regresi Linear untuk mendapatkan hasil korelasi atau hubungan antara data sentiment dan data *crypto*. Berdasarkan pengujian dengan 50 data aktual *Textblob* dapat memberikan data yang lebih valid dan memberikan nilai akurasi yang tinggi yaitu 80% sedangkan *Multinomial Naïve bayes* dengan perbandingan 60/40 hanya memberikan akurasi sebesar 46%.

Kata Kunci: *Cryptocurrency*, Bitcoin, Klasifikasi, *Machine Learning*, *Textblob*, *Naïve Bayes*, *Analisis Regresi Linear*

ABSTRACT

Cryptocurrency is a digital currency that is currently well-known and widely used in various community activities. Bitcoin as one of the cryptocurrencies continues to increase. *Cryptocurrency* price fluctuations, including Bitcoin, are influenced by the market and public opinion, (Zhang et al., 2018). Twitter says that more than 1 billion tweets about cryptocurrencies since 2020 and become the main platform for discussion of cryptocurrencies. There is a correlation between sentiment on social media and cryptocurrency price fluctuations, as shown by research by (ElBahrawy et al., 2019) who concluded that activity on Twitter has an effect on cryptocurrency price fluctuations, as well as Kristoufek's research (2015) which found a correlation between sentiment on Twitter and Bitcoin price fluctuations.

The machine learning method used in an attempt to analyze the phenomena described earlier includes a hybrid approach and simple linear regression analysis. This hybrid approach combines Text blob and Naïve Bayes as well as simple linear regression analysis to find correlations between cryptocurrency price movements and sentiment on social media Twitter. Using a simple linear regression analysis, the relationship between sentiment and cryptocurrency price fluctuations is measured. Data collection will be carried out from the Twitter platform for sentiment data and CoinGecko for crypto marketcap data by scraping data through the Twitter developer and the CoinGecko API. The data collected will go through the preprocessing stage then enter the modeling stage with Textblob and then proceed with training using Multinomial Naïve Bayes. The next stage is normalization using min max normalization with the aim of minimizing computational errors. The stage is followed by data analysis using Linear Regression Analysis.

The classification model is made using Textblob and then it will be trained using Multinomial Naïve Bayes. From these stages, the results of the effectiveness of using this method will be obtained by using the confusion matrix classification evaluation. After that, the Linear Regression Analysis method will be implemented to get the results of the correlation or relationship between sentiment data and crypto data. Based on testing with 50 actual data Textblob can provide more valid data and provide a high accuracy value of 80% while Multinomial Naïve bayes with a ratio of 60/40 only gives an accuracy of 46%.

Keyword: *Cryptocurrency, Bitcoin, Classification, Machine Learning, Textblob, Naïve Bayes, linear regression analysis.*