Analysis Of Public Sentiment Towards Crypto Exchange Using The Naïve Bayes Method

Qori' Hidayatulloh [1], Addin Aditya[2]*, Dedy Ari Purnomo[3]

[1,2,3] *Faculty of Science and Technology, Jl. Raya Tidar No.100, Karangbesuki, Kec. Sukun, Kota Malang, Jawa Timur 65146, Indonesia*

Keywords

Cryptocurrency; Crypto Exchange; Sentiment Analysis; Twitter; Investment

Correspondence Email:
addin@ubhinus.ac.id

Abstract

Cryptocurrency is a rapidly growing form of digital currency that serves as both an alternative payment system and an investment instrument. To conduct transactions, users require crypto exchanges as intermediaries. However, the large number of available exchanges, both local and international, often makes it difficult for users to select platforms that are safe, legal, and reliable. This problem is further complicated by technical issues, such as unstable systems during market fluctuations, which may cause financial losses for users. Therefore, this study aims to conduct sentiment analysis on various crypto exchanges by utilizing public opinion on social media, particularly Twitter. The results of this analysis are expected to provide more objective insights for users in choosing the right exchange, thereby minimizing risks and supporting a better investment experience.

1. Introduction

Cryptocurrency, or virtual currency, is a type of digital money that is not regulated by a central authority but rather managed by its owners. Its emergence offers solutions to the limitations of conventional payment systems while also presenting an attractive investment opportunity. To acquire and trade digital assets, users rely on crypto exchanges, which serve as platforms for exchanging conventional currency with cryptocurrency. As of 2021, there were more than 300 crypto exchanges worldwide (Coinmarketcap, 2021), highlighting the rapid growth of this industry.

Nevertheless, the abundance of exchange options creates challenges for users, who must carefully choose a platform that is safe, legal, and reliable in order to avoid losses and data breaches. The case of Indodax, whose system frequently went down during price declines, causing significant losses to its users (Media Konsumen, 2021), exemplifies the potential risks. This indicates the need for an objective evaluation of crypto exchanges to help users make wiser investment decisions.

At the same time, the rise of social media in Indonesia provides a valuable source of public opinion regarding crypto exchanges. According to a report by We Are Social, Indonesia ranks ninth globally in terms of social media usage, with Twitter being one of the most popular platforms (Kompas Tekno, 2021). Public opinions expressed on Twitter can be analyzed through sentiment analysis to identify positive, negative, and neutral perceptions of various exchanges. Therefore, this study focuses on leveraging Twitter data to provide meaningful insights that can support better decision-making for cryptocurrency investors.

1.1 Literature Review

Research on sentiment analysis in the context of cryptocurrency has been widely conducted, providing insights into public opinion and market behavior. Prasetiya et al. (2021) examined sentiment analysis based on comments and replies on Twitter for three major cryptocurrencies: Bitcoin, Ethereum, and Ripple. Their study compared the performance of Naive Bayes and Support Vector Machine (SVM) methods, showing that SVM achieved higher accuracy than Naive Bayes. This highlights the effectiveness of machine learning techniques in classifying sentiments into positive, negative, or neutral categories.

Another study [4] focused on Bitcoin price prediction by combining sentiment analysis with Long Short-Term Memory (LSTM) models. The research demonstrated that sentiment analysis could effectively separate positive and negative tweets, achieving an accuracy of 80%. By integrating historical price data with sentiment scores, the LSTM model was able to forecast future Bitcoin prices. Evaluation using Root Mean Square Error (RMSE) resulted in a value of 335.20 with 10 epochs, suggesting that incorporating sentiment features can enhance predictive performance.

Parlika et al. (2020) conducted a sentiment analysis of Bitcoin and cryptocurrency using Python's TextBlob library. By collecting 3,433 tweets via Twitter API, they classified the data into positive (41.3%), neutral (44.9%), and negative (13.7%) sentiments. Their study also employed word cloud visualization to highlight key terms associated with Bitcoin, thereby providing a clearer understanding of public perception.

From these previous studies, it can be concluded that sentiment analysis plays a crucial role in identifying public perceptions of cryptocurrency through social media platforms such as Twitter. The current research shares similarities in terms of using Twitter sentiment analysis but differs in scope by focusing on crypto exchanges rather than cryptocurrencies in general. Additionally, this study introduces enhanced data preprocessing stages and integrates the results into a web-based system to improve classification accuracy and practical usability.

2. Research Methods

This study applied a sentiment analysis approach to evaluate public perceptions of crypto exchanges in Indonesia. The research method consists of four stages: problem identification, system design, prototype design, and system testing.

Problem Identification

The increasing adoption of cryptocurrency in Indonesia has created challenges for new investors in selecting the right crypto exchange. With many platforms available, it is important to ensure the chosen exchange is legal, secure, and reliable. Sentiment analysis of Twitter data was used to capture public opinion, helping users choose exchanges with generally positive perceptions.

System Design

The workflow of the proposed system is illustrated through several flowcharts.

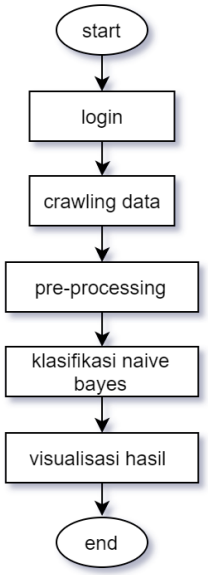

Fig. 1. System Flowchart

The Data Crawling Flowchart demonstrates the collection of up to 3,000 tweets (1,000 per exchange) using the Tweepy library.

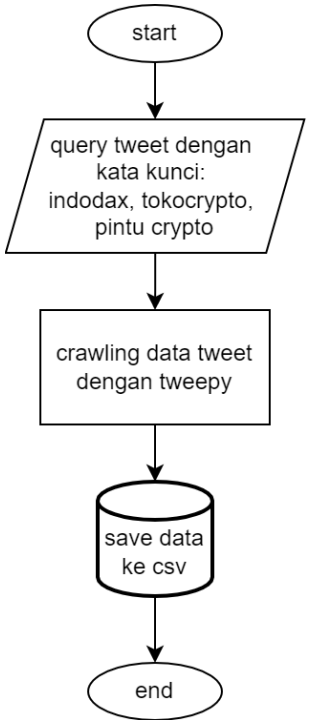

Fig. 2. Data Crawling Flowchart

The Pre-processing Flowchart outlines several steps: removing duplicates, case transformation, tokenization, token filtering by length, stopword removal using Sastrawi, and stemming using Sastrawi.

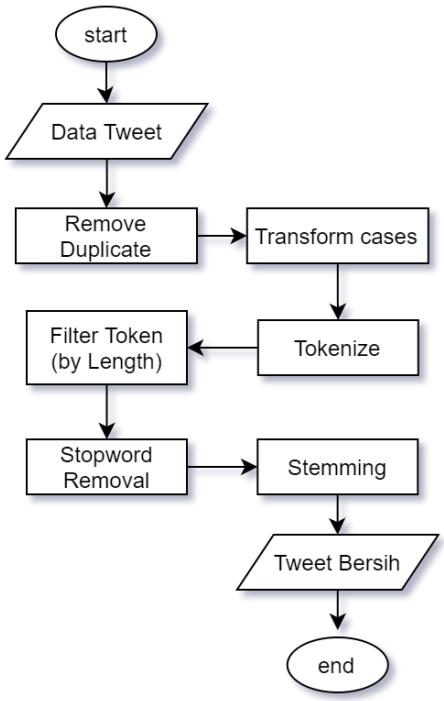

Fig. 3. Pre-processing Flowchart

The Naïve Bayes Method Flowchart shows the classification process with 700 tweets for training and 300 tweets for testing for each exchange.

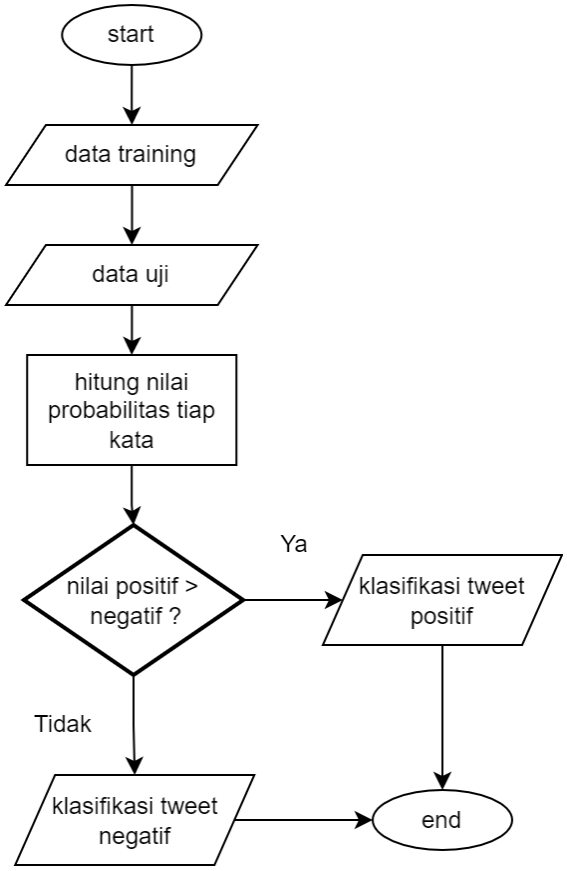

Fig. 4. Naïve Bayes Flowchart

Prototype Design

A prototype application was developed to display the sentiment results. The Homepage Interface presents the main web page where users can view sentiment classifications of each crypto exchange.

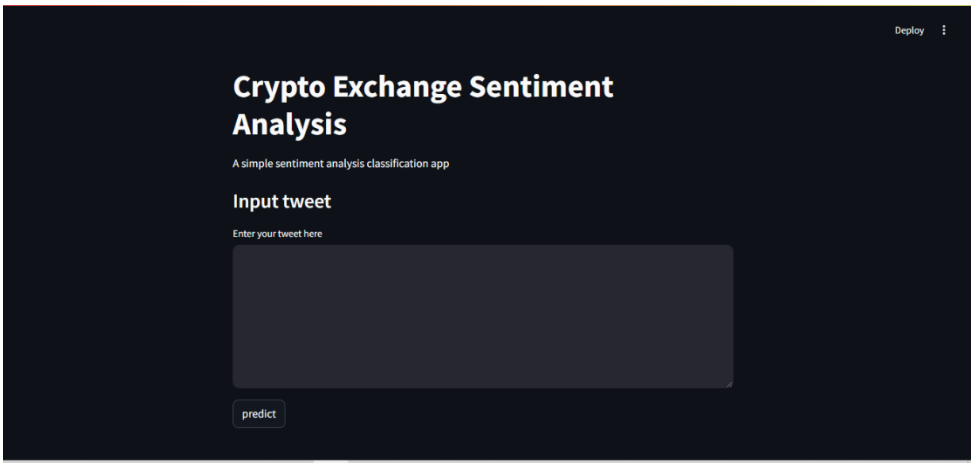

Fig. 5. Homepage Interface

System Testing

The system was evaluated using a confusion matrix, which measures classification performance by identifying true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These metrics are used to calculate accuracy, precision, recall, and F1-score, providing insight into the effectiveness of the Naïve Bayes classifier in classifying sentiments.

3. Result and Discussion

The evaluation measure was carried out using a 3x3 confusion matrix to assess the classification performance of the Naïve Bayes model across three sentiment classes: positive, neutral, and negative. The confusion matrix provided insights into true and false classifications for each category. True Positive (TP) indicates that the system correctly classified a positive sentiment, True Neutral (TNt) indicates correct classification of neutral sentiment, and True Negative (TN) represents correct classification of negative sentiment. Conversely, False Positive (FP), False Neutral (FNt), and False Negative (FN) indicate misclassifications among the three classes.

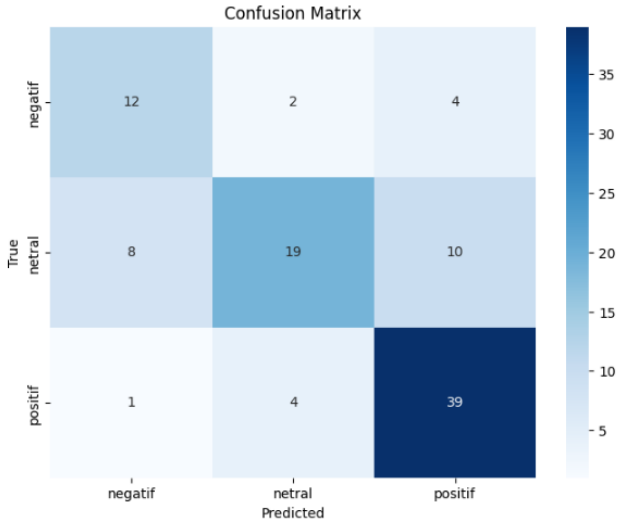

Fig. 6. Confusion Matrix for Tokocrypto

Based on the Tokocrypto confusion matrix, the model performed very well in predicting positive sentiment. However, it struggled with negative and neutral classifications, showing a significant number of errors.

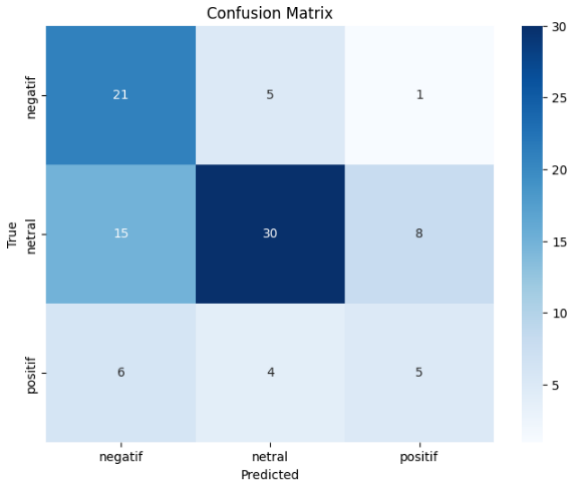

Fig. 7. Confusion Matrix for Indodax

For the Indodax exchange, the model achieved better performance in predicting negative sentiment compared to positive and neutral. Both positive and neutral classes still experienced high misclassification rates.

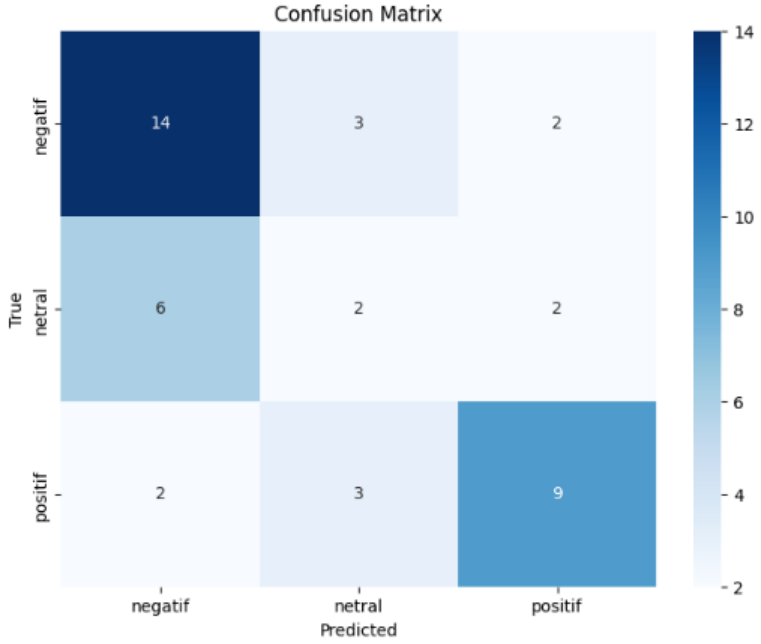

Fig. 8. Confusion Matrix for Pintu

The Pintu exchange results indicated that the model classified negative and positive sentiments fairly well, while neutral sentiment remained problematic.

Overall, using 1,500 tweets across three exchanges with the Naïve Bayes method, the sentiment analysis successfully captured public opinion. However, the average accuracy across the three exchanges was only 0.58, which is below the commonly acceptable threshold of 0.70–0.90. Among the three, Tokocrypto received the highest proportion of positive sentiment (53.3%), followed by Pintu (30.2%), and Indodax (14.7%).

4. Conclusions

Based on the conducted analysis, design, system implementation, and evaluation, it can be concluded that the sentiment analysis of Twitter data for Tokocrypto, Indodax, and Pintu has been successfully carried out. Although the system was able to classify sentiments, the resulting accuracy level remains relatively low, averaging 0.58. The findings highlight that Tokocrypto is perceived more positively by users compared to Indodax and Pintu.

For future work, several improvements are recommended. First, during data collection and labeling, many tweets contained spam and typographical errors, reducing the quality of training data. Future research should consider additional data sources beyond Twitter and implement more advanced pre-processing techniques to handle slang, typos, or abbreviations. Second, alternative weighting techniques should be explored to improve model accuracy and robustness.

5. References

Arisandi, A. D., & Atika, L. (2020). Prediksi mata uang Bitcoin menggunakan LSTM dan sentiment analisis pada sosial media. Jurnal Ilmiah Komputasi, 19(4). https://doi.org/10.32409/jikstik.19.4.370

Damsar, & Indrayani. (2018). Pengantar sosiologi pasar. Prenadamedia Group.

Devita, R. N., Herwanto, H. W., & Wibawa, A. P. (2018). Perbandingan kinerja metode Naive Bayes dan K-Nearest Neighbor untuk klasifikasi artikel berbahasa Indonesia. Jurnal Teknologi Informasi dan Ilmu Komputer, 5(4), 427. https://doi.org/10.25126/jtiik.201854773

Eka Sembodo, J., Setiawan, E. B., & Abdurahman Baizal, Z. (2016). Data crawling otomatis pada Twitter. INDOSC, 11–16. https://doi.org/10.21108/INDOSC.2016.111

Laurensz, B., & Sediyono, E. (2021). Analisis sentimen masyarakat terhadap tindakan vaksinasi dalam upaya mengatasi pandemi Covid-19. Jurnal Nasional Teknik Elektro dan Teknologi Informasi, 10(2), 118–123. https://doi.org/10.22146/jnteti.v10i2.1421

Nasukawa, T., & Yi, J. (2003, October). Sentiment analysis: Capturing favorability using natural language processing. Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP 2003) (pp. 70–77). https://doi.org/10.1145/945645.945658

Parlika, R., Pradika, S. I., Hakim, A. M., & K. R. N. M. (2020). Analisis sentimen Twitter terhadap Bitcoin dan cryptocurrency berbasis Python TextBlob. Jurnal Ilmiah Teknologi Informasi dan Robotika, 2, 33–37.

Prasetiyo, E. (2012). Data mining: Konsep dan aplikasi menggunakan Matlab. CV Andi Offset.

Prasetiya, A., Kunang, Y. N., Negara, E. S., & Chandra, W. (2021). Sentiment analisis terhadap cryptocurrency berdasarkan comment dan reply pada platform Twitter. Journal of Information Systems and Informatics, 3(2). http://journal-isi.org/index.php/isi

Yunita, D. (2017). Perbandingan algoritma K-Nearest Neighbor dan Decision Tree untuk penentuan risiko kredit kepemilikan mobil. Jurnal Informatika Universitas Pamulang, 2(2), 103. https://doi.org/10.32493/informatika.v2i2.1512