ABSTRACT

In Indonesia, the number of cryptocurrency investors has experienced a significant increase, reaching a total of 18.83 million people by January 2024, up from 16.83 million people in January 2023. With the growing public interest in cryptocurrency, various cryptocurrency investment platforms have rapidly developed, including Indodax and Tokocrypto, which dominate the market in Indonesia. Despite the increasing popularity of cryptocurrency investment applications, several issues may arise, such as slow transaction speeds, hacking risks, and security uncertainties. These problems can lead to dissatisfaction and doubt among app users, ultimately harming the reputation of the cryptocurrency industry. Therefore, this study aims to analyze the sentiment of user reviews of cryptocurrency investment applications using the Multinomial Naive Bayes method. This method was chosen for its simplicity, speed, and effectiveness in text classification, especially with large datasets. The study involves collecting user review data from the Google Play Store for the Indodax and Tokocrypto applications through web scraping. The collected data is then processed through several stages: text preprocessing, sentiment labeling, word weighting, and finally, sentiment classification using the Multinomial Naive Bayes algorithm. In this study, various dataset splitting scenarios and random state parameter settings were tested to evaluate the developed model. Testing was conducted with three data splitting scenarios: 90:10, 80:20, and 70:30. The results show that model accuracy is greatly influenced by the test data size and random state value. Smaller test data sizes tend to provide higher and more consistent accuracy, while larger test data sizes show lower and more unstable performance. This study found label imbalance in the dataset, so the SMOTE technique was applied. The use of the SMOTE technique to balance labels significantly improved model performance in terms of accuracy, precision, recall, and F1-score for each dataset. For the Indodax dataset, the highest accuracy was achieved in scenario 2 with an 80:20 split ratio, where accuracy increased from 82.48% to 86.62%, precision from 81.41% to 87.68%, recall from 82.48% to 86.62%, and F1-score from 76.78% to 86.90%. For the Tokocrypto dataset, the highest accuracy was achieved in scenario 1 with a 90:10 split ratio, where accuracy increased from

82.53% to 85.72%, precision from 82.04% to 86.92%, recall from 82.53% to 85.72%, and F1-score from 77.74% to 86.12%. Combining the datasets from both applications resulted in an accuracy of 85.72%, precision of 86.92%, recall of 85.72%, and an F1-score of 86.12%. This model demonstrates good performance and can be used on website inputs to detect the sentiment of other cryptocurrency investment applications. The study shows that the Multinomial Naive Bayes algorithm was successfully implemented in sentiment analysis of user review data from cryptocurrency investment applications. The model is effective not only for Indodax and Tokocrypto but also applicable to other applications. This research provides significant insights into user perceptions of the Indodax and Tokocrypto cryptocurrency investment applications. The results of this study can be used by application developers to improve services, better respond to user issues, and enhance customer satisfaction and loyalty. Additionally, this research provides valuable information for potential investors in selecting cryptocurrency investment applications based on sentiment analysis. Thus, this study not only helps improve the quality of cryptocurrency investment application services but also contributes to increasing public trust and interest in cryptocurrency investment in general.