

ABSTRACT

The increase in fake reviews on e-commerce platforms is a serious problem that can affect user purchasing decisions and harm both sellers and buyers. Therefore, effective methods are needed to identify fake reviews and ensure the authenticity of the reviews provided. Product reviews significantly influence the reputation of sellers and the trust of buyers in the products offered. Studies show that most buyers read reviews before making a purchase, and trust in these reviews is almost equivalent to personal recommendations. However, the crucial role of reviews also triggers harmful practices, such as fake reviews intended to mislead buyers or unfairly boost product reputations. This research aims to develop a method to classify reviews into genuine or fake on the e-commerce platforms Shopee and Blibli using a machine learning approach, specifically Multilayer Perceptron (MLP). The proposed method involves several key steps to ensure accuracy and effectiveness in identifying fake reviews. These steps include preprocessing, feature extraction from review texts using Natural Language Processing (NLP) techniques, and manual labeling of reviews by three annotators based on predetermined criteria to ensure data validity. After preprocessing, features such as word frequency, sentiment, and review length are extracted from the review texts by calculating their weights using Term Frequency-Inverse Document Frequency (TF-IDF). These TF-IDF weights indicate the importance of a word in the analyzed document, and these features are used as input for the machine learning model. An MLP model with various parameter configurations such as hidden layers, L2 regularization, batch size, and learning rate is trained using the preprocessed and feature-extracted training data. The research results show that the model with a three-layer hidden configuration (25-25-100), L2 regularization of 0.1, batch size of 16, and learning rate of 0.001 provides the best performance with a validation accuracy of 0.847 and an f1-score of 0.90. These results indicate that the model can detect fake reviews with high accuracy and maintain consistent performance when applied to new, unseen data during the training process. The model was then evaluated using metrics such as accuracy, precision, recall, and f1-score. Additionally, the results of genuine and fake review predictions are illustrated in a confusion matrix showing the number of correct and incorrect

predictions for each class. The findings from this research are visualized through a website that provides information on the classification of reviews and displays the optimal model performance evaluation metrics. This data visualization is divided into three pages: the first page tests the model with user input reviews, the second page compares manual labels with model labels, and the third page compares models based on different data proportions. Training this MLP model is expected to be integrated into e-commerce platforms, providing benefits for all parties involved in the e-commerce ecosystem, including buyers, sellers, and marketplace platforms. This integration will ensure more accurate and trustworthy reviews, aiding better purchasing decisions.

Keywords—fake reviews, e-commerce, Machine Learning, MLP, NLP