ABSTRACT

E-commerce is a transaction process or business activity involving goods and services using information technology. In Indonesia, many e-commerce companies have developed, one of which is Bukalapak. Bukalapak ranks third in monthly website visits with a total of 29,460,000 as of Quarter 2 (April, May, June) 2021. Seeing the increasingly fierce e-commerce competition today, Bukalapak is competing in planning and making various efforts in order to survive and improve the quality of service to consumers. One way that can be done is to evaluate the results of the review. To be able to make a decision from the results of the review, one of the steps that need to be taken is by classifying the review with the aim of categorizing the data against comments or reviews so that it can assist business actors in drawing conclusions regarding the trend of comments. The dataset used in this study amounted to 96,329 data. Before conducting the analysis, the data that has been collected is pre-processed so that it has a total of 87,241 data. Because the data has a missing value, it is overcome by the imputation method using mode () or mode. After overcoming the missing value, the dataset is weighted using the tfidfVectorizer, then resampling is carried out with SMOTE so that the data is balanced. The review was analyzed using the K-Nearest Neighbors algorithm with three scenarios, namely 60:40, 70:30, and 80:20, and had three types of k_neighbors, namely k=3, k=5 and k=7. The distance used in this study is Euclidean. The results of the analysis show that the best KNN is at the training and testing ratio of 80:20 with k=3. The results of the analysis show that the training accuracy is 82 % and the evaluation results are 76%. The results of the classification with KNN show more negative comments than positive ones. The results of this study are expected to be used as evaluation material for Bukalapak to improve service quality.

Keyword: E-commerce, Classification, K-Nearest Neighbors, Euclidean