

Abstract

Digital Population Identity is a government application created to make it easier for people to manage citizenship files online. Many user reviews reflect various sentiments regarding people's satisfaction and disappointment with the app. However, Google Play Store only displays application ratings and raw reviews, making it difficult for the government to understand the level of satisfaction and problems experienced by users. This research aims to implement the Bidirectional Encoder Representation from Transformers (BERT) model with fine-tuning and analyze the effect of input data quality on the performance of the model. It is found that the fine-tuned IndoBERT model without stopword removal and stemming preprocessing performed best, with accuracy, precision, recall, and F1-score values of 98%. These findings indicate that the preprocessing stage affects the final performance of the model. This study is expected to help the government gain a more comprehensive understanding of public opinion in order to develop better applications.

Keywords: *IndoBERT, Sentiment Analysis, User Reviews, Text Preprocessing, Digital Population Identity.*