

Abstract-In In the ever-evolving digital era, social media, especially platforms like TikTok, have become primary channels for users to share opinions, experiences, and expressions. However, the increasing prevalence of hate speech in reviews on the Google Play Store for the TikTok app indicates the need for a sophisticated approach to identify and classify harmful content. This research aims to optimize the classification of hate speech in Google Play reviews of the TikTok app by integrating Term Frequency-Inverse Document Frequency (TF-IDF), Differential Evolution, and Word2Vec within a Recurrent Neural Network (RNN) model. The TF-IDF technique is used to extract relevant features from reviews, while Differential Evolution efficiently optimizes the model parameters. Word2Vec enhances the representation of words in the context of app reviews, and the RNN model enables the recognition of temporal patterns in hate speech. The results of this research, achieving the highest accuracy of 88.63% and an F1 score of 88.62%, are expected to contribute significantly to improving hate speech classification on digital platforms focused on app reviews. The study demonstrates the effectiveness of combining advanced feature extraction and optimization techniques to develop a robust classification system for identifying and mitigating hate speech.

Keywords: Classification; Term Frequency-Inverse Document Frequency; Word2Vec; Differential Evolution; Recurrent Neural Network