

1. INTRODUCTION

TikTok is a social media application that allows users to create and innovate creative content and share short videos. Content on TikTok is numerous and diverse, from creative and educational to inappropriate behavior such as cyberbullying that can affect users' mental health [1]. Content on the TikTok application often reaps criticism (sentiment) because of the negative effects it causes, so it is important to carry out sentiment analysis of the TikTok application in understanding the emotional response of public perception towards the TikTok application [2], so that it can help policy makers and application developers in dealing with negative issues, to improve features to create a more positive and comfortable user experience.

Research on sentiment analysis has been conducted by [3], who reported that deep learning classifiers like Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) consistently outperform traditional machine learning algorithms such as SVM in terms of precision, recall, F-measure, and accuracy. Effective sentiment analysis involves feature extraction and expansion such as word embedding, because word embeddings retain the order and semantic nature of sentences in aspect-based sentiment analysis [4]. Feature extraction plays a crucial role in transforming initially unstructured text data into a structured format, facilitating sentiment analysis. By detailing the key characteristics of the text, feature extraction not only helps reduce information loss but also enhances the discriminative ability in sentiment classification [5]. In this way, feature extraction not only affects the accuracy of sentiment analysis but also increases its effectiveness by providing a better representation of the nuances and complexities within the text [6]. Text feature expansion is a process that involves adding dimensions or additional features to the text representation for machine learning models, aiming to improve representation quality and the model's ability to understand complex text. This process contributes to an enriched vocabulary that the model can recognize, including the recognition of ironic or sarcastic statements that often appear [4][7][8].

Some researchers studying sentiment analysis frequently use optimization methods due to their numerous advantages. According to [9], without the application of optimization methods, lexicon labeling would be manually performed by humans, requiring significant time and cost. If optimization techniques such as the Barebones Particle Swarm Optimization (BBPSO) are implemented, lexicon labeling is automated by the system, making it more effective and efficient, as well as reducing time and cost. In the research conducted by [10] and [11], SGD proved to be effective in overcoming complex problems in sentiment analysis on certain topics. The use of SGD in training models with extracted features facilitates the grouping of tweets into positive and negative sentiment categories with accurate results. However, it should be noted that SGD has some drawbacks, such as high dependency on proper parameter selection and accurate features, as well as the potential to generate noise in parameter updates due to the use of a single example at each iteration.

Based on the background that has been described, the research was conducted with the aim of analyzing the sentiment of the TikTok app by exploring the benefits of integrating feature extraction techniques with Word2Vec, followed by feature expansion using FastText, and the optimization algorithm Stochastic Gradient Descent (SGD) to enhance the performance of the LSTM model through four scenarios.