1. INTRODUCTION

On the era of online interactions, social media platforms like TikTok have become central entertainment hubs where users can share their opinions and experiences. These platforms have revolutionized the way people connect, interact, and entertain themselves. With the convenience and widespread reach of these platforms, millions of users, including teenagers, are engaged daily in creating and consuming content. This massive user engagement has brought about a significant shift in digital communication and content-sharing practices. However, along with this rapid growth and popularity, a negative phenomenon related to hate speech has emerged, particularly in reviews discussing TikTok. The increase in user activity, especially among teenagers, has unfortunately led to a rise in negative interactions. Hate speech, characterized by derogatory, offensive, and inflammatory language, not only disrupts user experiences but also creates an unsafe and negative online environment [1]. This problem is not just a minor inconvenience but a serious issue that affects the mental well-being of users and the overall health of online communities.In light of research on hate speech in TikTok app reviews, the development of advanced classification methods becomes increasingly important [1].

The importance of addressing hate speech cannot be overstated. Hate speech has the potential to escalate into real-world consequences, including bullying, harassment, and even physical violence. Therefore, in light of research on hate speech in TikTok app reviews, the development of advanced classification methods becomes increasingly important [1]. Accurate and efficient identification of hate speech is crucial for maintaining a positive user experience and ensuring the safety of users. This challenge has spurred researchers and technologists to explore innovative approaches to detect and mitigate hate speech effectively [2]. One of the primary tools in this endeavor is the development of sophisticated machine learning algorithms capable of analyzing and classifying vast amounts of text data with high accuracy.

In research [3] conducted by Ravinder Ahuja et al., sentiment analysis was performed on the SS-Tweet dataset using two key features, namely TF-IDF and N-Gram. Sentiment analysis involves determining the sentiment expressed in a piece of text, which can range from positive and neutral to negative. This type of analysis is particularly useful in understanding public opinion and identifying problematic content. In this study, six classification algorithms were employed: Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), Logistic Regression (LR), and Naive Bayes (NB). The results indicated that using TF-IDF features at the word level performed 3-4% better than N-Gram features. This finding underscores the effectiveness of TF-IDF in capturing the relevance and importance of words within a document. Based on these results, the study recommends using TF-IDF feature extraction at the word level as the best choice for sentiment analysis using machine learning algorithms [3]. TF-IDF, which stands for Term Frequency-Inverse Document Frequency, is a statistical measure used to evaluate the importance of a word in a document relative to a collection of documents. By weighing the frequency of words inversely with their occurrence in the overall corpus, TF-IDF helps in emphasizing words that are significant in a specific context.

Another noteworthy approach to enhancing classification performance is highlighted in research conducted by K. Vijayaprabakaran et al. This study explored the use of the Differential Evolution (DE) approach to find the optimal activation function in the Long Short-Term Memory Network (LSTM). LSTM is a type of Recurrent Neural Network (RNN) that is particularly effective in handling sequential data and capturing long-term dependencies. The researchers conducted experiments on the IMDB dataset for sentiment classification and the UCI HAR dataset for human activity recognition. The study revealed that the combination of comb-H-sine activation functions yielded better results than other activation functions in terms of accuracy. The statistical test results indicated that comb-H-sine is significantly different from several other activation functions. The conclusion drawn from this study is that searching for the optimal activation function can be effectively achieved using the Differential Evolution approach. This optimization method, which mimics the process of natural selection, is considered to enhance the classification system's performance significantly [4].

By an efforts to expand classification capabilities, the implementation of Word2Vec for feature expansion in RNN (Recurrent Neural Network) represents a significant step [5]. This method utilizes the occurrence of paired words in occurrence matrices to train text, making it efficient in utilizing statistics while also maintaining the linear structure of the Word2vec method [6]. Word2Vec has proven to perform well in tasks like word analogy [6]. It provides the ability to embed the semantic context of words.

Based on research [7] conducted by Yequan Wang et al., this research proposes a capsule model based on Recurrent Neural Networks (RNN-Capsule) for sentiment classification. The model was evaluated using two benchmark datasets, namely Movie Review (MR) and Stanford Sentiment Treebank (SST), as well as one proprietary dataset. Experimental results showed that the RNN-Capsule model achieved the highest accuracy with an accuracy value of 91.6% compared to several other comparative methods on both datasets. This research successfully demonstrated that RNN can provide high accuracy results for predictive modeling. While the RNN approach allows the model to understand the temporal context in text [5].

Combining TF-IDF feature extraction technology, Differential Evolution optimization, and the application of Word2Vec for RNN feature expansion in the classification system this research is aiming to create an effective

approach in identifying and classifying hate speech in TikTok app reviews. A deep understanding of this context is expected to contribute significantly toward efforts to create a more positive and safe online environment for the TikTok community.