HATE SPEECH DETECTION IN SOCIAL MEDIA USING LSTM ALGORITHM

GOBINATH S (AP)
Department of Information Technology
V.S.B. Engineering College, Karur

SUJITHKUMAR R, SURYA J, SUGUMAR K, SRIRAM S
Department of Information Technology
V.S.B. Engineering College, Karur

In this project, we propose a text-based hate speech detection framework utilizing Twitter datasets and Long Short-Term Memory (LSTM) algorithms. LSTM networks, a type of recurrent neural network (RNN), are well-suited for processing sequential data such as text due to their ability to capture long-range dependencies and contextual information. By training an LSTM model on annotated Twitter data containing examples of hate speech and non-hateful language, we aim to develop a robust classifier capable of automatically identifying hate speech in real-time. The proposed framework will involve several key steps, including data preprocessing, feature extraction, model training, and evaluation. In this study, we offer an LSTM (long short-term memory) algorithm-based text-based hate speech identification system using Twitter datasets. Because LSTM networks, a subtype of recurrent neural networks (RNNs), can capture contextual information and long-range relationships, they are well-suited for processing sequential input, like text. Our goal is to create a strong classifier that can automatically identify hate speech in real-time by training an LSTM model using annotated Twitter data that contains instances of both hateful and non-hateful language. A number of crucial processes, including feature extraction, model training, evaluation, and data preprocessing, will be involved in the suggested framework. To prepare the text data for analysis, preprocessing techniques like tokenization, stemming, and stop word removal may be used. To represent the textual information in a numerical manner that can be input into the LSTM model, feature extraction approaches like word embeddings or TF-IDF may be used. Using the labeled data as training data, the LSTM model will discover word connections and patterns suggestive of hate speech.

 ISBN: 978-81-971598-1-7