Submission Summary

Conference Name
European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases 2025 - Journal Track

Track Name
ECMLPKDDJT2025-Second Track

Paper ID
3

Paper Title
BERT-Based Toxic Comment Detection for Online Platforms

Abstract
In the digital age, the proliferation of user
generated content on online platforms has necessitated effective mechanisms to detect and mitigate toxic comments. This research focuses on leveraging the Bidirectional Encoder Representations from Transformers (BERT) model for toxic comment detection.
The proposed system employs BERT's advanced natural language processing capabilities to classify comments as toxic or non-toxic, aiming to improve the moderation process on social media and other user-centric platforms. Our approach involves fine-tuning the pre-trained BERT model on a labeled dataset of comments, optimizing it for high accuracy in identifying harmful language patterns.
The implementation demonstrates significant improvements in precision and recall compared to traditional machine learning methods. By using BERT, we can capture nuanced linguistic features and context, enabling more accurate detection of subtle forms of toxicity. This research not only contributes to the field of natural language processing but also offers practical applications for enhancing online community management and fostering safer digital interactions. The results indicate that BERT-based models are a promising solution for real-time toxic comment detection, ensuring a healthier and more respectful online environment.

Created
6/19/2024, 8:38:38 PM

Last Modified
6/19/2024, 8:38:38 PM

Authors

Bhavesh S (BMS college of engineering) <bhavesh.ai21@bmsce.ac.in>

Submission Files

NLP_PAPER_PDF.pdf (430.1 Kb, 6/19/2024, 8:38:19 PM)