World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
Vol:18, No:11, 2024

Hate Speech Detection in Tunisian Dialect

Authors : Helmi Baazaoui, Mounir Zrigui

Abstract : This study addresses the challenge of hate speech detection in Tunisian Arabic text, a critical issue for online safety and moderation. Leveraging the strengths of the AraBERT model, we fine-tuned and evaluated its performance against the Bi-LSTM model across four distinct datasets: T-HSAB, TNHS, TUNIZI-Dataset, and a newly compiled dataset with diverse labels such as Offensive Language, Racism, and Religious Intolerance. Our experimental results demonstrate that AraBERT significantly outperforms Bi-LSTM in terms of Recall, Precision, F1-Score, and Accuracy across all datasets. The findings underline the robustness of AraBERT in capturing the nuanced features of Tunisian Arabic and its superior capability in classification tasks. This research not only advances the technology for hate speech detection but also provides practical implications for social media moderation and policy-making in Tunisia. Future work will focus on expanding the datasets and exploring more sophisticated architectures to further enhance detection accuracy, thus promoting safer online interactions.

Keywords : hate speech detection, Tunisian Arabic, AraBERT, Bi-LSTM, Gemini annotation tool, social media moderation
Conference Title : ICLEMC 2024 : International Conference on Language Endangerment: Methodologies and Challenges
Conference Location : Singapore, Singapore
Conference Dates : November 18-19, 2024