World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:19, No:09, 2025

A Comparative Study of Natural Language Processing Models for Detecting Obfuscated Text

Authors : Rubén Valcarce-Álvarez, Francisco Jáñez-Martino, Rocío Alaiz-Rodríguez

Abstract : Cybersecurity challenges, including scams, drug sales, the distribution of child sexual abuse material, fake news, and hate speech on both the surface and deep web, have significantly increased over the past decade. Users who post such content often employ strategies to evade detection by automated filters. Among these tactics, text obfuscation plays an essential role in deceiving detection systems. This approach involves modifying words to make them more difficult for automated systems to interpret while remaining sufficiently readable for human users. In this work, we aim at spotting obfuscated words and the employed techniques, such as leetspeak, word inversion, punctuation changes, and mixed techniques. We benchmark Named Entity Recognition (NER) using models from the BERT family as well as two large language models (LLMs), Llama and Mistral, on XX_NER_WordCamouflage dataset. Our experiments evaluate these models by comparing their precision, recall, F1 scores, and accuracy, both overall and for each individual class.

Keywords : natural language processing (NLP), text obfuscation, named entity recognition (NER), deep learning
Conference Title : ICMLC 2025 : International Conference on Machine Learning and Cybernetics
Conference Location : Prague, Czechia
Conference Dates : September 06-07, 2025