An Intelligent Detection Approach for Drug Abuse in Arabic Social Media Posts.

By

Amani AbedAlghafer

Supervisor**:**

Dr. Bilal Hawashin

Co-Supervisor**:**

Dr. Tarek Kanan

Al - Zaytoonah University of Jordan, 2022

Abstract

lately, the use of the Internet has mainly led to an increase in social networking sites. The world has become an open environment, which has led to the adoption of social media users on these sites to exchange medical experiences and their adoption in many cases as basic references in obtaining medical advice, which leads to the use of various medicines in an incorrect way and thus the abuse of medicines such as pain relievers, antibiotics, and dermatology medicine.

Social media users have become dependent on the experiences of others, which has led to the abuse of some medications prescribed by doctors.

A growing problem, abuse of prescription medications can have a negative impact on all age groups and come with adverse health consequences, as individuals in societies become susceptible to many drug interactions and serious side effects and reduced drug efficacy, which makes a simple health problem turn into a complex health problem; our study aims to classify drug use in Arabic content in social media (use, abuse) by using both Machine Learning (ML) algorithms and AraBERT model with Deep Learning Many studies detect the drug abuse in the English language. There is lack studies on Arabic language. We used preprocessing tools to handle

the Arabic language; NLP techniques (Stop Word Removal, Normalization, and Stemming) were applied. Arabic social media dataset were created from Facebook with nearly 7,000 posts.

We used different ML classifiers; Support Vector Machine (SVM), Decision Tree (J48), Naive Bayes (NB), K-Nearest Neighbor (KNN), and Random Forest (RF). We also applied AraBERT based model with DL; CNN-AraBERT, RNN-AraBERT, and LSTM-AraBERT. The classifier's accuracy was evaluated by calculating the F1-Measure, Recall, and Precision measurements**.**

The results indicated that CNN-AraBERT classifier is given the highest value of F1-Measure for Facebook dataset for both classification tasks with (98.3%) for binary classification and (90.99%) for multi-classification.

Keywords: Machine Learning, NLP, Arabic Language, AraBERT, Social Media.