

# **Detecting Depression Tone that Leads to Suicide in Arabic content on Social Media using Machine Learning and Deep Learning Techniques**

**By**

**Odai Mamoun Sadaqa**

Supervision

**Dr. Tarek G Kanan**

Co-supervisor

**Dr. Mohammed ElBes**

**Al-Zaytoonah University of Jordan, 2020**

## **Abstract**

This study aims to suggest a system to detect the depressing tone that leads to suicide from Arabic content on social media, in particular, Facebook and Twitter, by using two famous methods, Machine learning and Deep learning techniques.

Our research focuses on the Arabic language, despite the difficulties we face in the Arabic language, we have simplified the concepts of the Arabic language using three of the most famous Natural Language Processing (NLP), which include stemming, normalization, and stop word removal. We collect the dataset from Arabic social media, Facebook, and Twitter, the total number was collected approximately 10,000 publications. Then, we applied five classifiers in machine learning techniques, which include Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (J48), and Random Forest (RF). In addition, to evaluate the performance of the classifiers, we calculated F1-measure, Recall, and Precision measurement accuracy.

The result showed superiority for the RF classifier the highest value of F1-measure with (99.1%). On the other hand, the best result achieved with all datasets without applying pre-processing techniques, through SVM with F1-measure (93.4%).

In addition, when separating the dataset into the twitter dataset and Facebook dataset, the result showed superiority for RF classifier the highest value of F1-measure with (99.3%) and (99.1%) respectively with all datasets and by applying preprocessing techniques.

On the other hand, we apply Deep Learning techniques, which include Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Deep Neural Networks (DNN). We use the Python programming language, in particular, the Keras library to apply Deep Learning. In addition, we apply the feature extraction method, which includes Word Embedding and Bag Of Words (BOW) on the dataset.

The result showed superiority for DNN Algorithm the highest value of F1-measure with (96%), followed by CNN with (94%), and RNN with (62%) respectively with all datasets.