



الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la
Recherche Scientifique
المدرسة العليا للإعلام الآلي. 08 ماي 1945. بسيدي بعباس
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès



MEMOIRE

To obtain the diploma of **State Engineer**
Field: **Computer Science**
Specialty: **Ingénierie des Systèmes Informatiques (ISI)**

Theme

**Sarcasm Detection In Arabic Tweets Using Deep Learning
and BERT-based Models with Data Augmentation**

Presented by :
Touhami Wided Ahlem
Taleb Douaa

Submission Date : **4th July , 2022** In front of the jury composed of :

Mr. KESKES Nabil
Mme. ALIANE Hassina (CERIST)
Mr. KHALDI Belkacem
Mr. ALIANE Ahmed Amine (CERIST)
Mme. Klouche Badia

President
Supervisor
Supervisor
Co-Supervisor
Examiner

Academic Year : 2021/2022

Abstract

Automatic sarcasm detection from text is an important classification task that can help identify the actual sentiment in user-generated data, such as reviews or tweets. Despite its utility, detecting sarcasm remains a difficult problem due to the lack of vocal intonation or facial expressions in textual data. To date, the majority of solutions have relied on hand-crafted affect characteristics such as emojis and hashtags, or pre-trained models of non-contextual word embeddings, such as Word2vec. However, the inherent limitations of these models make them unsuitable for detecting sarcasm. We present in this study a deep neural network sarcasm detection application. As a starting point, we used a set of deep learning models. Moreover, we have used three main Transformer-based Models for Arabic Language Understanding. The proposed model has been evaluated using the ArSarcasm-v2 dataset, in addition to a manually collected dataset containing sarcastic AraCOVID19 tweets. Extensive experiments on different datasets demonstrate that the proposed models give a good outcome. As a result, the proposed system is capable of achieving an accuracy of 98% for the AraCOVID19 dataset and accuracy of 79% for the ArSarcasm-v2 dataset.

Keywords: Sarcasm Detection, Sentiment Analysis, Natural Language Processing, Data Preprocessing, Deep learning ,Transformers, AraBERT, MARBERT, CAMELBERT, LSTM, BiLSTM , GRU , CNN ,BiLSTM-CNN, Data Augmentation.