Check-Worthiness Estimation

To determine whether a claim in a tweet or transcriptions is worth fact-checking.

Typical approaches to make that decision require to either resort to the judgments of professional fact-checkers or to human annotators to answer several auxiliary questions such as "does it contain a verifiable factual claim?", and "is it harmful?", before deciding on the final check-worthiness label

I have decided to use the machine learning approach by training the model on the train data sets and finding the best F1 score over the positive class of the classification using the dev set.

Three Languages have been used (English, Dutch and Arabic) and various pre-trained models along with some traditional methods were used depending on the language.

The following were the models that were used for:

1)BERT

2)RoBERTa

3)xlm-RoBERTa

4)DistilBERT

5)AlBERT

6)CamemBERT

7)Electra

5)BigBird

6)AraBERT(Arabic)

8)MultinomialNB

9)SVC-linear

10)Logistic Regression

The following were the best models according to the f1 score that was obtained when finding the accuracy with the dev set.

English-DistilBERT

Arabic-AraBERT

Dutch-MultinomialNB

The predictions for the test data are predicted using these models.

The following is the GitHub link for the repository

https://github.com/TheHolyOrange/Check-Worthiness-Estimation.git