Arabic Digital Harmful Text Detection

American University of Beirut

مؤذٍ-ضارّ
Harmful

Problem Definition
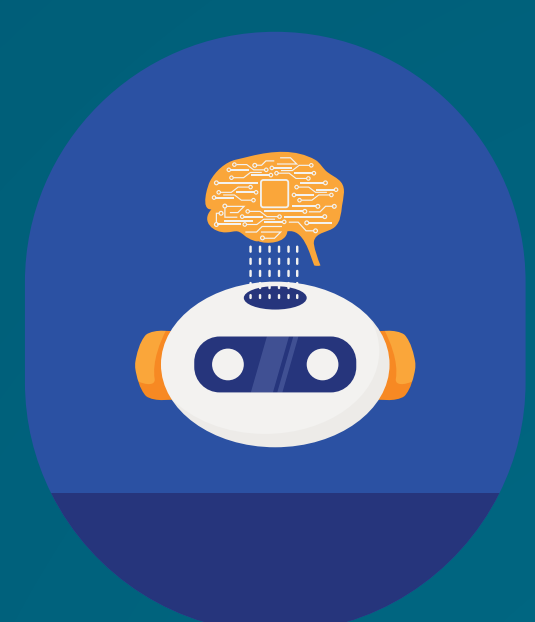
The severe effects of digital harmful text on the Arab society by increasing social conflict, crime and suicide rates. We have chosen to tackle this problem by classifying written arabic text into 5 categories, 4 of which describe different forms of harmful text including hateful text and bullying text.

Dataset Description

Our dataset consists of 12211 arabic tweets extracted from 4 sources:

-Arabic Hate Speech detection by Raghadash (GitHub)

-Multilingual and Multi-Aspect Hate Speech Analysis by HKUST (Github)

-Arabic Levantine Hate Speech Detection by Haithem Hermessi (Kaggle)

-Terrorism and Jehadism Speech Detection by Haithem Hermessi (Kaggle)

Investigated Models

We used 4 different models, using 60% of our dataset for training each one, 20% for validation and 20% for testing:

- Naive Bayes
- Logistic Regression (Multinomial)
- SVM
- Neural Network

Results

Best accuracy for our models were:

Naive Bayes: 55%

Logistic Regression: 62% for C = 1.0

SVM: 59% for a RBF Kernel

Neural Network: 52%

62%

Conclusion

We have obtained a maximum accuracy of 62% after experimenting on 4 different models with different parameters on validation. We believe a larger dataset for our models would have helped increase the accuracy. Lowering the number of classes we had would have also had the same effect but would drift us away from the project's main goal.

Refrences

1- Arabic Hate Speech detection by Raghadash on GitHub
https://github.com/raghadsh/Arabic-Hate-speech

2- Multilingual and Multi-Aspect Hate Speech Analysis by HKUST
https://github.com/HKUST-KnowComp/MLMA_hate_speech

3- Arabic Levantine Hate Speech Detection by Haithem Hermessi
https://www.kaggle.com/datasets/haithemhermessi/arabic-levantine-hate-speech-detection

4- Terrorism and Jehadism Speech Detection by Haithem Hermessi
https://www.kaggle.com/datasets/haithemhermessi/terrorism-and-jihadist-speech-detection?resource=download

5- Arabic Text Pre-Processing
https://github.com/saobou/arabic-text-preprocessing

Project members: Johnny Lamaa, Amjid Khurwat, Nacer Boudad