

Nabil Beshay

STUDENT, MASTER OF SCIENCE IN INFORMATION
TECHNOLOGY

A Deep Learning Based Multilingual Hate Speech Detection for Resource Scarce Languages

Over the last decade, the increased use of social media has led to an increase in hateful activities in social networks. Hate speech over online social networks is a worldwide problem that leads to diminishing the cohesion of civil societies. Due to the lack of restrictions set by these sites for its users to express their views as they like, anybody can make adverse and unrealistic comments in abusive language against anybody with an ulterior motive to tarnish one's image and status in society. Hate speech is one of the most dangerous of these activities, so users have to protect themselves from these activities from social media sites such as YouTube, Facebook, Twitter, etc. Large-scale social platforms are currently investing important resources into automatically detecting and classifying hateful content, without much success. The rapid spread of social media websites is accompanied by an increasing number of social media users which showed a higher rate of hate speech, as well.

This research introduces a method for using deep learning algorithms to

predict hate speech from social media websites. Especially for a country like Canada with a multilingual and bilingual population, this hate content would be in a code-mixed form, which makes the task difficult. We implement proposed algorithms to detect hate speech in two social media text datasets: Arabic and English. This study employs a variety of feature engineering techniques and a comparative study among different machine and deep learning algorithms to automatically detect hate speech messages on Arabic and English datasets. After hate speech data is collected, as a part of the preprocessing steaming, token splitting, character removal, and inflection elimination are carried out before performing the hate speech recognition process through deep learning algorithms. In future, we would like to deploy proposed algorithms to other low resource languages and explore more language specific features in deep learning framework.

Research Advisor: Dr. Baidya Saha,
Assistant Professor