Contribution ID: **136** Contribution code: **145** Type: **Poster**

Multilingual Racial Hate Speech Detection Using Transfer Learning

The rise of social media eases the spread of hateful content, especially racist content with severe consequences. In this paper, we analyze the tweets written in French targeting the death of George Floyd in May 2020 as the event accelerated debates on racism globally. Using the Yandex Toloka platform, we annotate the tweets into categories as hate, offensive, or normal. Tweets that are offensive or hateful are further annotated as racial or non-racial. We build French hate speech detection models based on the multilingual BERT and CamemBERT and apply transfer learning by fine-tuning the HateXplain model. We compare different approaches to resolve annotation ties and find that the detection model based on CamemBERT yields the best results in our experiments.

Find me @ my poster

2: Monday afternoon

Keywords

Racial hate speech, offensive speech, transfer learning, Toloka

Authors: AYELE, Abinew Ali (Language Technilogy Group, University of Hamburg, Hamburg, Germany); Prof. BIEMANN, Chris (Language Technilogy Group, University of Hamburg, Hamburg, Germany); Dr YIMAM, Seid Muhie (House of Computing and Data Sceince); Ms DINTER, Skadi (Language Technilogy Group, University of Hamburg, Hamburg, Germany)