

HASOC

Hate Speech and Offensive Content Identification in Indo-European Languages

Introduction

This is the call to participate in the Shared Task on Hate Speech and Offensive Content Identification in Indo-European Languages. We invite everyone from academia and industry to participate in the Shared Task on the Identification of Offensive content for Indo-European languages.

HASOC is inspired from two evaluation forums, [OffensEval](#) and [GermanEval](#) 2018, and try to leverage synergies of both the forum.

There has been significant work in several languages in particular for English. However, there is a lack of research in this recent and relevant topic for most other languages. This track intends to develop data and evaluation resources for several languages. The objectives are to stimulate research for these languages and to find out the quality of hate speech detection technology in other languages. In the long run, the track aims at supporting researchers to develop robust technology which can cope with multilingual data and to develop transfer learning approaches which can exploit learning data across languages. For future editions, we envision the integration of further languages.

Data

Dataset will be created from the Twitter and Facebook and distributed in tab separated format. Participants are allowed to use external resources and other datasets for this task. Dataset will be prepared in 3 languages (German, English and code-mixed hindi).

The size of Training data corpus is approximately 8000 posts for each language.

Tasks

Our objective behind the HASOC shared task is to leverage the synergies of both forums. HASOC shared task is offered in 3 sub-tasks.

Participants in this year's shared task can choose to participate in one, two or all of the subtasks.

• Sub-task A :-

Sub-task A focus on Hate speech and Offensive language identification offered for English, German, Hindi. Sub-task A is coarse-grained binary classification in which participating system are required to classify tweets into two class, namely: Hate and Offensive (HOF) and Non- Hate and offensive (NOT).

- (NOT) Non Hate-Offensive - This post does not contain any Hate speech, offensive content.
- (HOF) Hate and Offensive - This post contains Hate, offensive, and profane content.

In our annotation, we label a post as HOF if it contains any form of non-acceptable language such as hate speech, aggression, profanity otherwise NOT.

• Sub-task B :-

Sub-task B is a fine-grained classification. Hate-speech and offensive posts from the sub-task A are further classified into three categories.

- (HATE) Hate speech :- Posts under this class contain Hate speech content.
- (OFFN) Offensive :- Posts under this class contain offensive content.
- (PRFN) Profane :- These posts contain profane words.

HATE SPEECH

Describing negative attributes or deficiencies to groups of individuals because they are members of a group (e.g. all poor people are stupid). Hateful comment toward groups because of race, political opinion, sexual orientation, gender, social status, health condition or similar.

OFFENSIVE

Posts which are degrading, dehumanizing,insulting an individual,threatening with violent acts are categorized into OFFENSIVE category.

PROFANITY

Unacceptable language in the absence of insults and abuse. This typically concerns the usage of swearwords (Scheiße, Fuck etc.) and cursing (Zur Hölle! Verdammt! etc.) are categorized into this category.

We expect most posts to be OTHER, some to be HATE and the other two categories to be less frequent. Dubious cases which are difficult to decide even for humans, should be left out.

• **Sub-task C :-**

Sub-task c will check the type of offense. Only posts labeled as HOF in sub-task A are included in sub-task C. The two categories in sub-task C are the following:

- Targeted Insult (TIN): Posts containing an insult/threat to an individual, group, or others.
- Untargeted (UNT): Posts containing nontargeted profanity and swearing. Posts with general profanity are not targeted, but they contain non-acceptable language.

The multilingual HASOC Corpus will be sampled from Facebook and Twitter and distributed in tab separated format. Participants are allowed to use external resources and other datasets for this task. Dataset will be prepared in 3 languages (German, English and code-mixed hindi).

The size of Training data corpus is approximately 8000 posts for each language and Test data is approximately 1000 posts for the each language. Classification systems in all tasks will be evaluated using either macro-averaged F1-score or weighted F1-score.

Organisers

1. Thomas Mandl :- University of Hildesheim, Germany
2. Sandip Modha :- DA-IICT, Gandhinagar, India
3. Chintak Mandlia :- infoAnalytica Consulting Pvt. Ltd.
4. Daksh Patel :- Dalhousie University, Halifax, Canada
5. Aditya Patel :- Dalhousie University, Halifax, Canada
6. Mohana Dave :- LDRP-ITR, Gandhinagar, India