# Emotionally Informed Hate Speech Detection: A Multi-target Perspective (Paper Summary)

Authors -
Patricia Chiril · Endang Wahyu Pamungkas · Farah Benamara · Véronique Moriceau · Viviana Patti

Summary by
Mohammad Aflah Khan
IIIT Delhi

# The problem

## Hate Speech

Huge amounts of Hate speech is being dumped on Social Media platforms

## Lacking Approach

Currently it’s treated as a binary classification problem and it does not utilize the topical focus or the target-oriented nature of hate speech

# Contributions of the Paper

## Contribution 1

**Use topic specific Hate Speech data to find patterns and be able to recognize specific forms of Hate Speech**

## Contribution 2

**Instead of Binary Classification the paper tries to identify the target group and the topic of the hate speech**

## Contribution 3

**By using Sentic Computing and Semantically structured hate lexicons the paper tries to determine the specific manifestation of hate speech**

# Some Major Results of the Paper

| Result 1 | Result 2 | Result 3 |
| --- | --- | --- |
| **Topic specific dataset trained models perform better than general trained models** | **Multi task approach outperforms single task model in detecting both the class and target of hate speech** | **Using EmoSenticNet emotions improves results** |

# Characteristics of Existing Models

First, they are trained to predict the presence of general, target-independent HS, without addressing the problem of the variety of aspects related to both the topical focus and target-oriented nature of HS.

Second, systems are built, optimized, and evaluated based on a single dataset, one that is either topic-generic or topic-specific

# Conclusions of the Paper

## Conclusion 1

**Training on topic-generic datasets generally fails to account for the linguistic properties specific to a given topic**

## Conclusion 2

**Combining topically focused datasets enabled the detection of multi-target HS even if the topic and/or target are unseen**

## Conclusion 3

**Affective knowledge encoded in sentic computing resources and semantically structured hate lexicons improve finer-grained HS detection**

# Some Improvement Windows

Deal with inherent bias in Dataset

Accommodate prediction of related classes like Sexism and Misogyny

Handle Sarcasm and Irony

Complex inference of views to not make naive classifications

Model misclassified by picking up words while the tweet might be combating it

Testing new HourGlass Model