

4-21-2017

# Measuring Gender-Based Violence Attitude on Twitter

Goonmeet Bajaj

Wright State University - Main Campus, bajaj.3@wright.edu

Tanvi Banerjee

Wright State University - Main Campus, tanvi.banerjee@wright.edu

Amir Hossein Yazdavar

Wright State University - Main Campus

Valerie L. Shalin

Wright State University - Main Campus, valerie.shalin@wright.edu

Amit Sheth

Wright State University - Main Campus, amit.sheth@wright.edu

Follow this and additional works at: [https://corescholar.libraries.wright.edu/urop\\_celebration](https://corescholar.libraries.wright.edu/urop_celebration)



Part of the [Arts and Humanities Commons](#), [Engineering Commons](#), [Life Sciences Commons](#), [Medicine and Health Sciences Commons](#), [Physical Sciences and Mathematics Commons](#), and the [Social and Behavioral Sciences Commons](#)

## Repository Citation

Bajaj, G., Banerjee, T., Yazdavar, A. H., Shalin, V. L., & Sheth, A. (2017). *Measuring Gender-Based Violence Attitude on Twitter*. .

This Article is brought to you for free and open access by the Office of the Vice President for Research at CORE Scholar. It has been accepted for inclusion in Browse All Celebration of Research, Scholarship, and Creative Activities Materials by an authorized administrator of CORE Scholar. For more information, please contact [corescholar@www.libraries.wright.edu](mailto:corescholar@www.libraries.wright.edu), [library-corescholar@wright.edu](mailto:library-corescholar@wright.edu).



# Measuring Gender-Based Violence Attitude on Twitter

Goonmeet Bajaj, Tanvi Banerjee, Amir Hossein Yazdavar, Valerie Shalin, Amit Sheth

Kno.e.sis Center, Wright State University, Dayton, OH

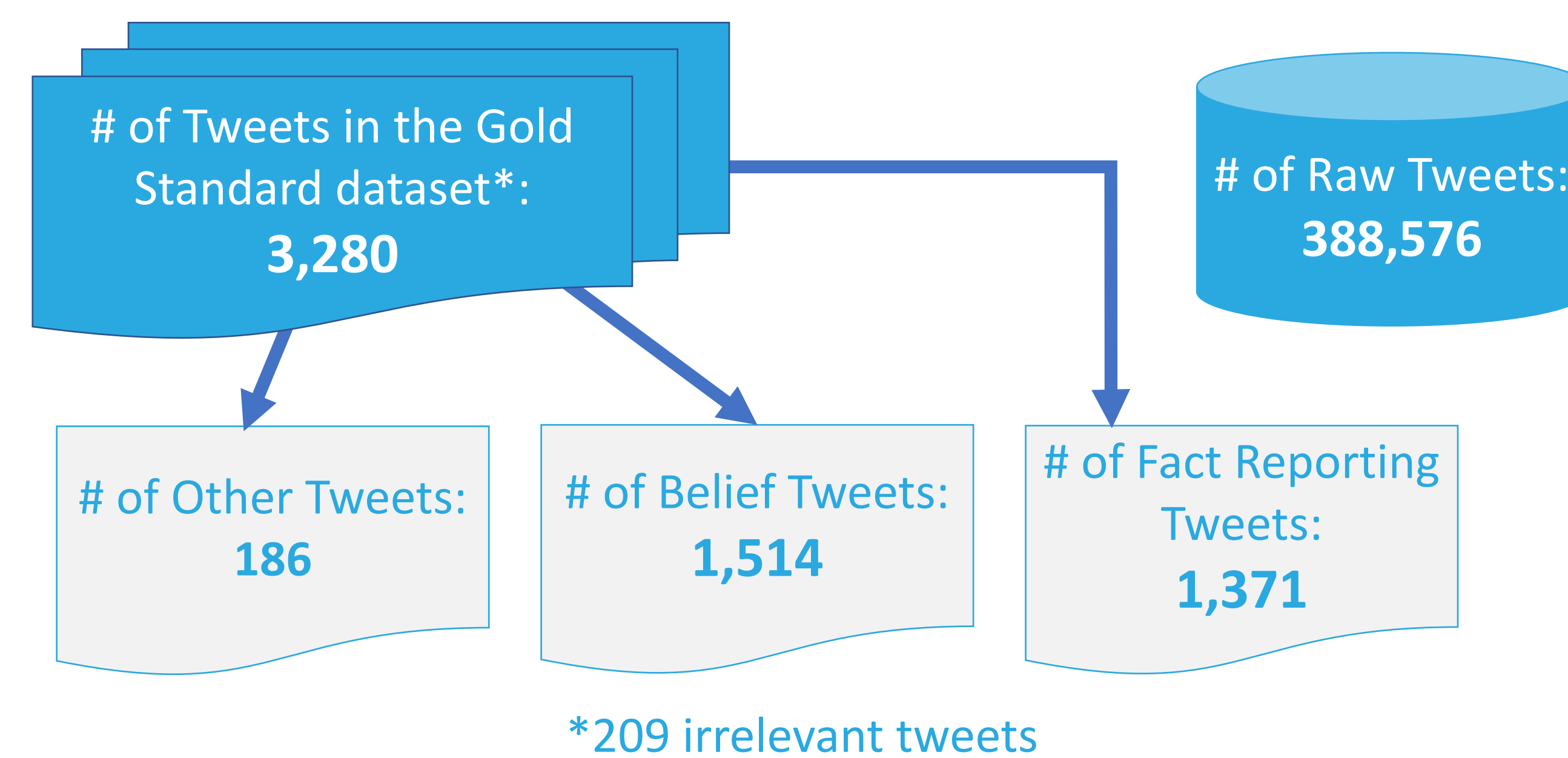


## Overview

This study examines social media to analyze communication patterns, specifically in the Gender-Based Violence domain. It explores the pragmatic content of tweets to categorize them into one of the following groups: **Belief**, **Fact Reporting**, or **Other** to determine the tweeting practices prevalent in this domain.

- Key note:  $Fact\ Reporting + Belief = Belief$

## Dataset



## Gold Standard Guidelines

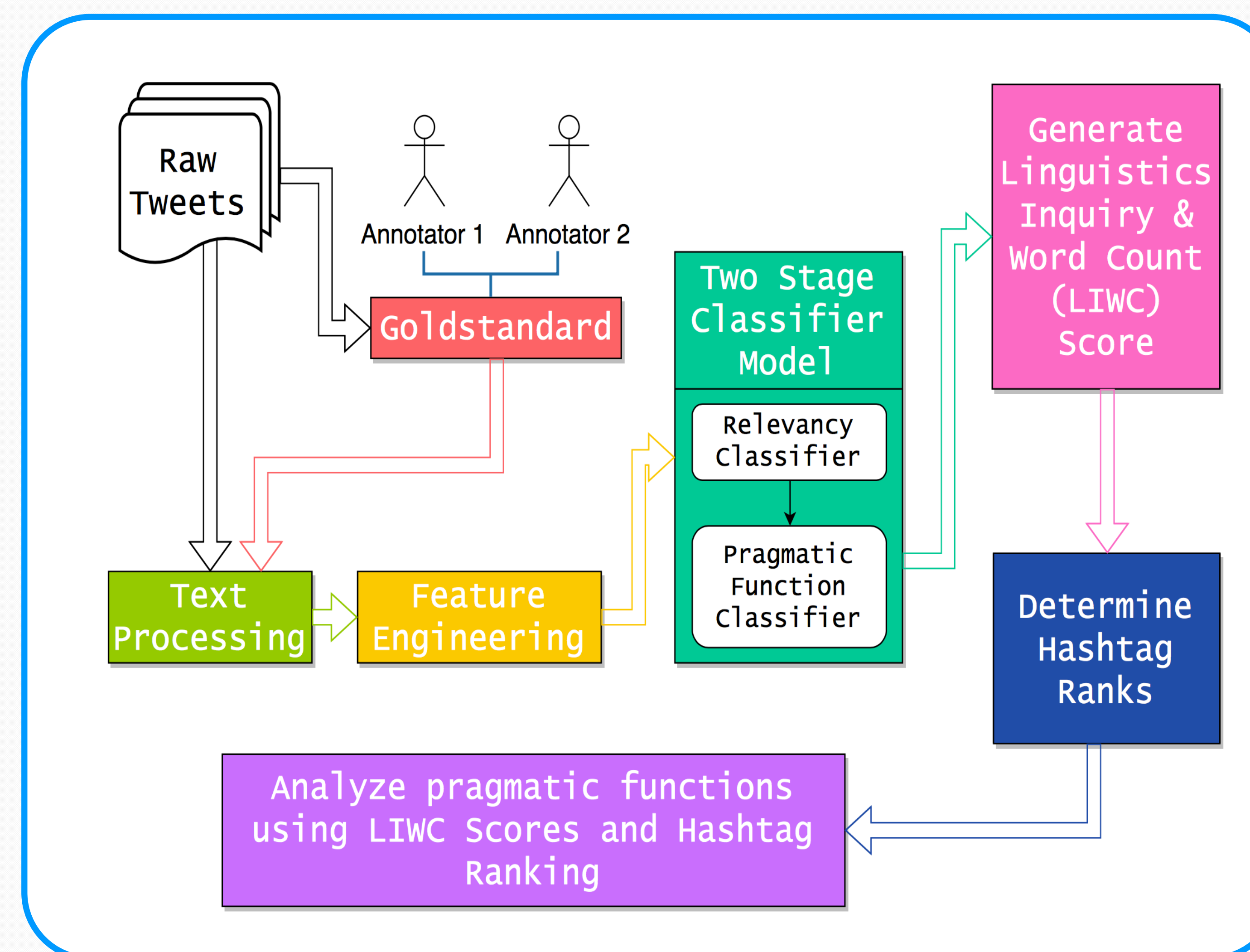
**Belief**: The major factor behind #Rape in #India is the #Bollywood which incites feeling 2 cross the moral limits #BanBollywood

**Other**: Question of the day, "Is there any law protection of rape victim?"

**Fact Reporting**: Ohio high school rape convict released -USA Today http://...

## Text Processing

- Spell checking
- Word-breaking
- Part-of-speech tags
- Removal of non-ascii characters
- Screen name removal
- \*Stopword removal
- Hyperlink removal
- \*Punctuation removal
- \*Stemming
- \*These techniques were proposed but not applied to the dataset



## Classification Performance

### Two-stage classifying model:

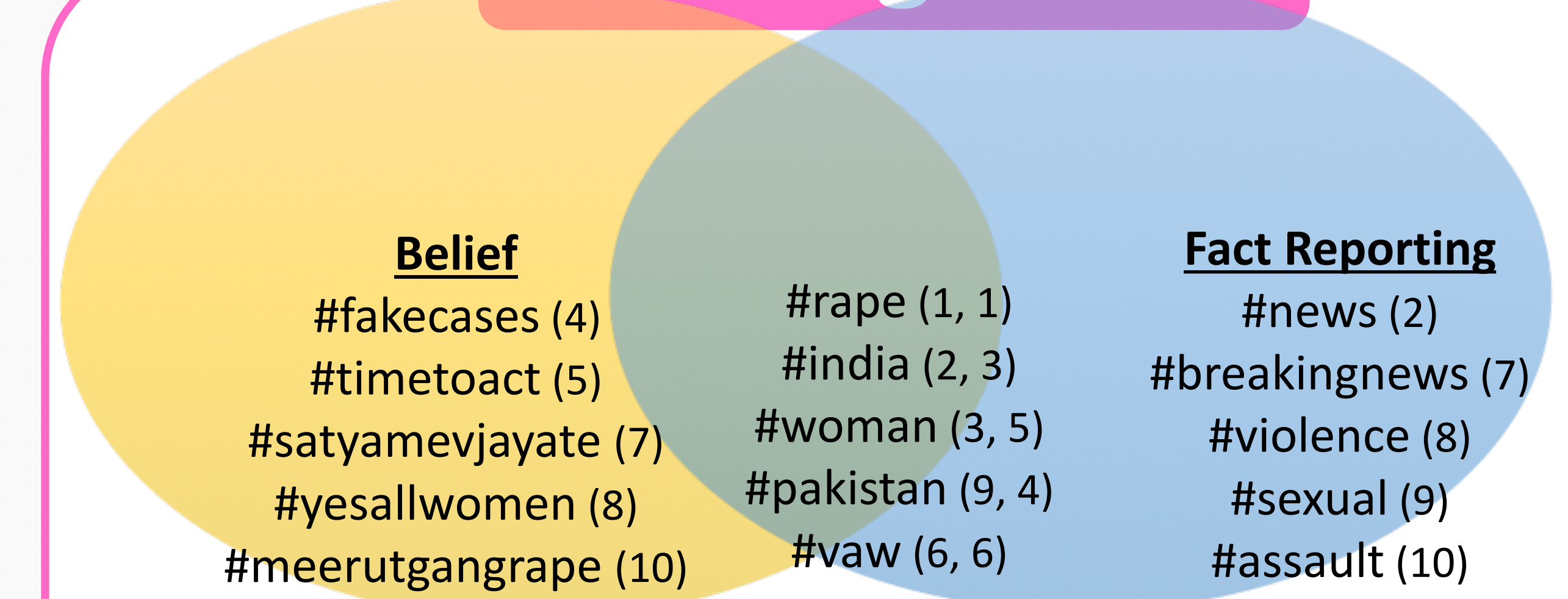
- one classifier is trained to detect tweets related to the GBV domain
- second classifier for the pragmatic function detection of each GBV related tweet

Feature Set	Precision		Recall		F-Measure	
	MNB	SVM	MNB	SVM	MNB	SVM
1 N-gram (N=1)	0.652	0.679	0.634	0.693	0.638	0.676
2 N-gram (N=2)	0.619	0.623	0.552	0.636	0.572	0.621
3 N-gram (N=3)	0.597	0.613	0.398	0.599	0.446	0.574
4 <b>N-gram (N=1,2,3)</b>	0.629	<b>0.700</b>	0.587	<b>0.695</b>	0.599	<b>0.679</b>
5 N-gram (N=1,2,3) + POS	0.638	0.672	0.592	0.693	0.607	0.672

	Belief	Fact Reporting	Other	Total Tweets
Relevant	264,390	111,753	4783	<b>380,926</b>
Non-relevant	-	-	-	7,650

Annotator \ Model	Fact Reporting	Belief
Fact Reporting	69	31
Belief	12	88

## Insights



### Hashtag Ranking:

- Belief hashtags are more specific to the GBV domain and indicate a desire to resist GBV
- Fact reporting hashtags are more generic, such that they are often used to broadcast facts to a mass audience

	Belief (B)	Fact Reporting (FR)
Analytic	72.075	93.953
Pronouns	7.983	1.889
Cognitive processes	7.003	3.436
Swear words	0.495	0.098
Netspeak	3.752	2.433
Assent	2.220	1.838
All Punctuation	52.561	67.083

### LIWC scores takeaways:

- The **analytic** scores for the two pragmatic function confirms that fact reporting tweets reflect formal communication, whereas, belief tweets are more informal and personal.
- The higher **swear words** score for belief tweets indicates that belief tweets more frequently contain swear words when compared to fact reporting tweets.
- The lower **netspeak** score for fact reporting tweets in demonstrates that these tweets are less likely (than belief tweets) to contain words like "lol" and "4ever".

## Brainstorm

Please tell me your interest and your thoughts about this study?

- How can we better study Gender-Based Violence?
- How can the detection of pragmatic function improve?

Email me @ goonmeet.bajaj@gmail.com

Project website:

[http://wiki.knoesis.org/index.php/Gender-%C2%ADbased\\_violence\\_\(GBV\)](http://wiki.knoesis.org/index.php/Gender-%C2%ADbased_violence_(GBV))

