

Modeling Social Media Narratives about Caste-related News

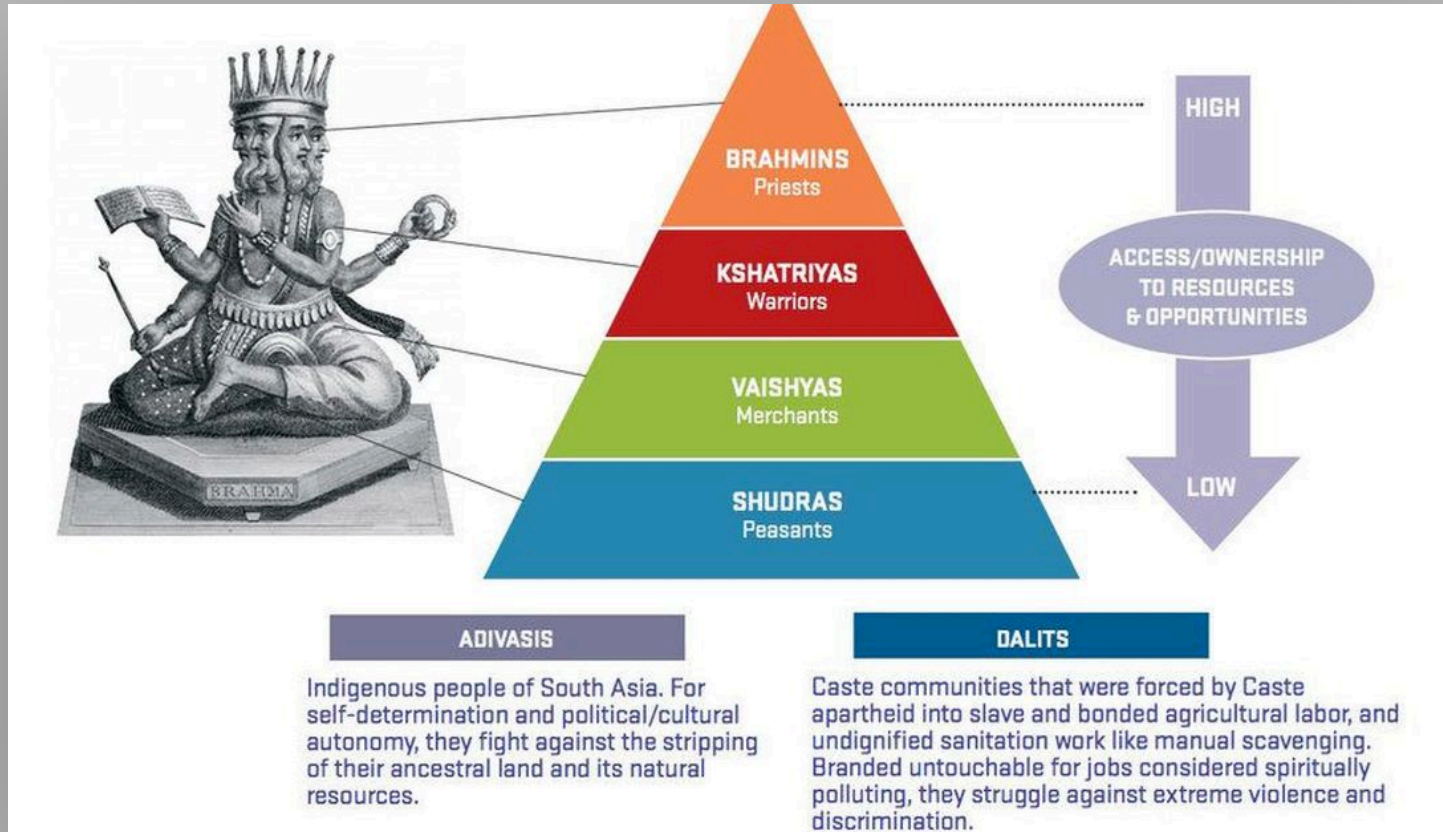
Prashanth Vijayaraghavan and Lavanya Vijayaraghavan

==



About Caste System

”



*The outcaste is a
bye-product
of the caste
system.
Nothing can
emancipate
the outcaste
except the
destruction of
the caste
system.*

— Dr. B.R. Ambedkar

Key Contributions



Cast

An automated data collection pipeline for aggregating caste-related stories.

A cross-platform corpus, Censor, that contains characteristic narratives aggregated using extraction of value judgments from Reddit content related to caste-related stories.

Ongoing modeling approach to infer value judgments and generate counter-narratives for user-generated comments.

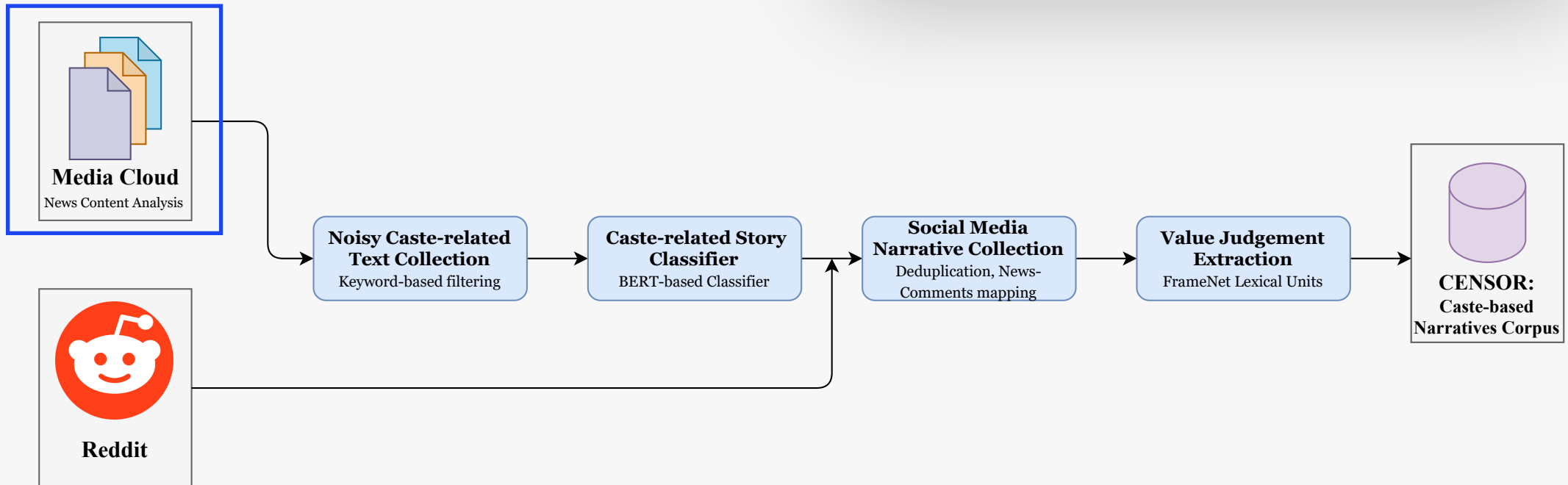
Data Sources

Explorer



Get a quick overview of how your topic of interest is covered by digital news media by exploring attention, language, and entities.

- **Media Cloud** tracks millions of stories and provides instant **analysis of digital news** over various topics.
- Media Cloud allows users to **choose media collections** and **submit boolean queries** that match these sources' stories.

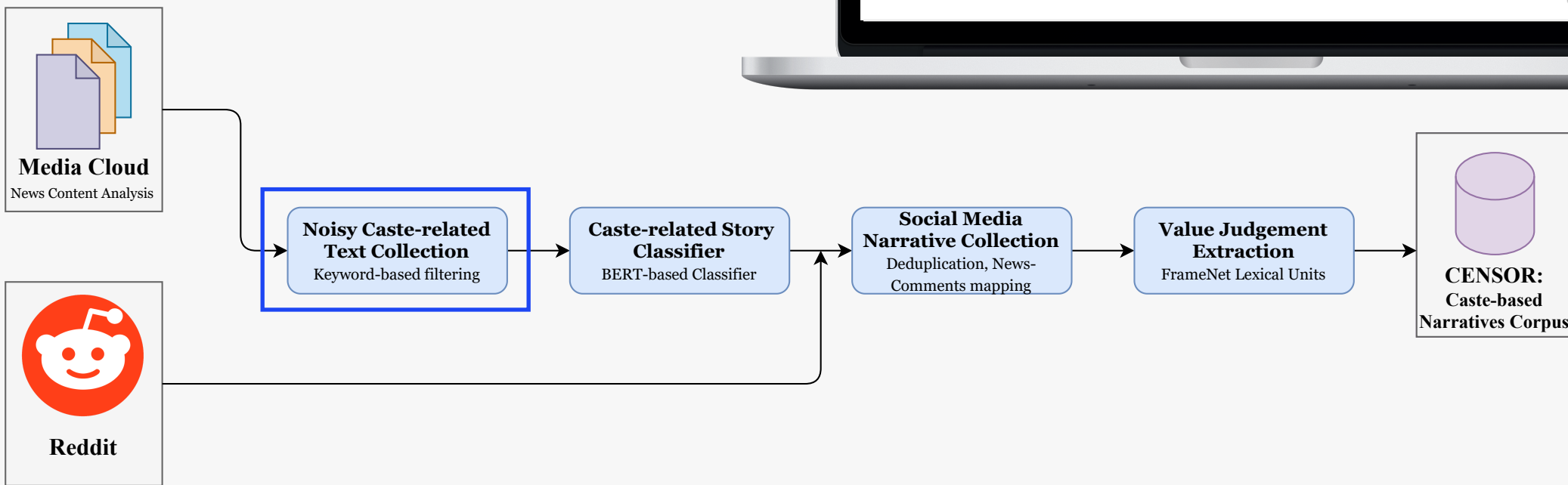
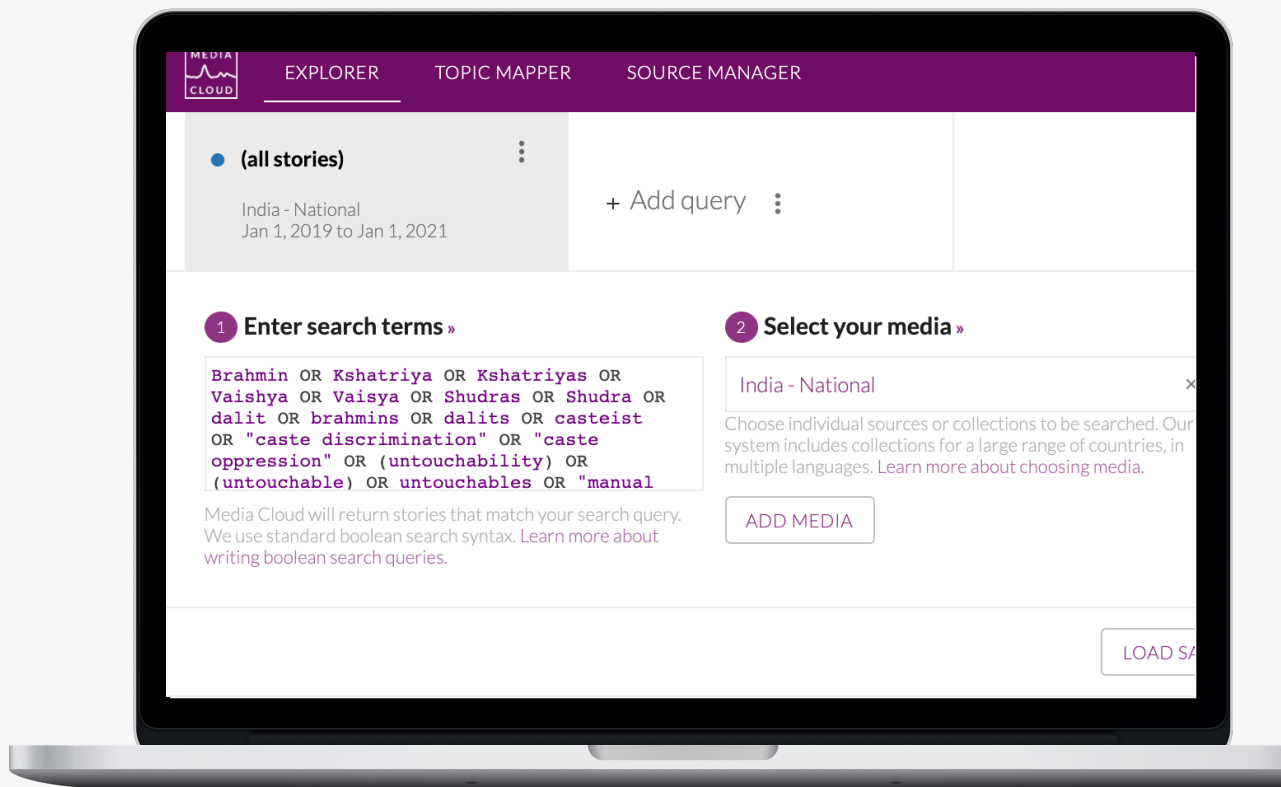


Noisy Caste-related Text Collection

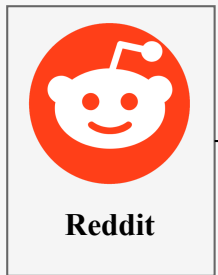
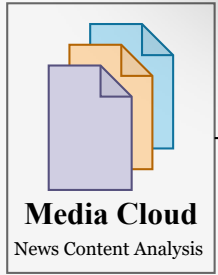
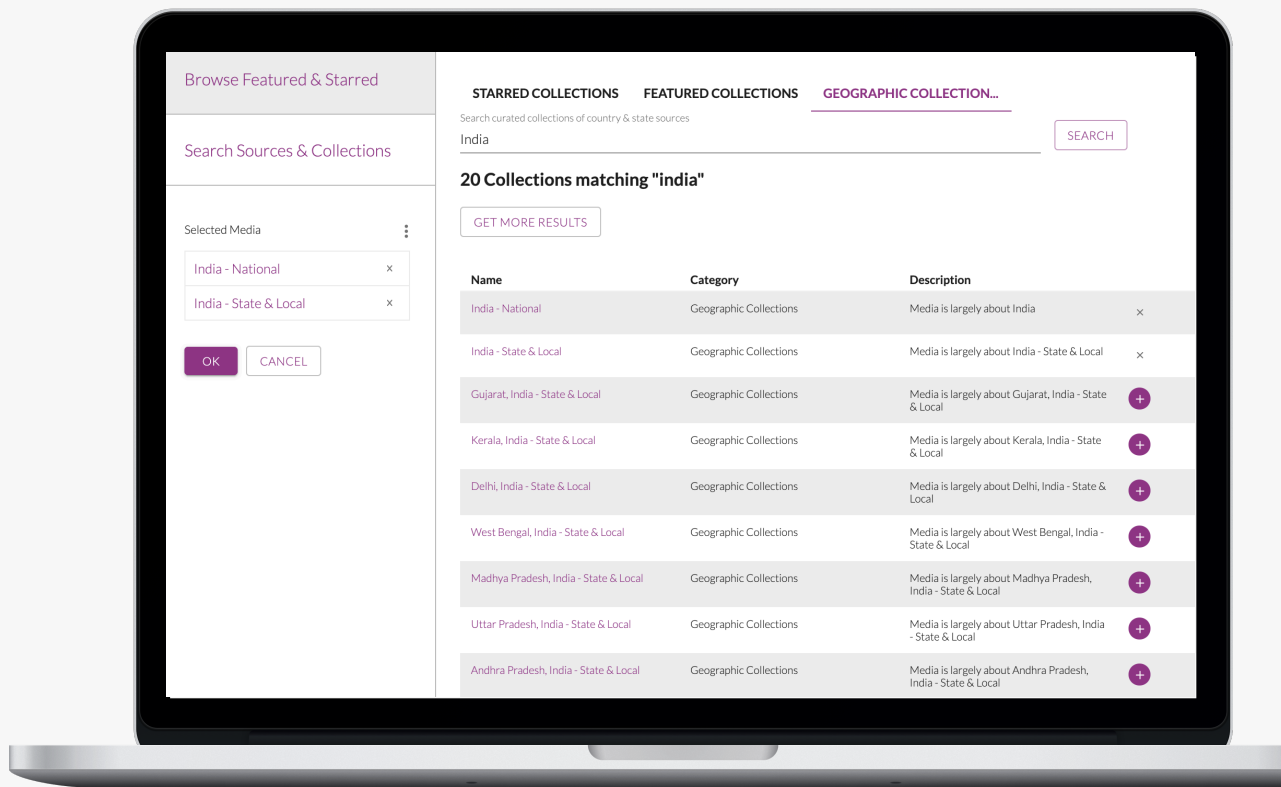
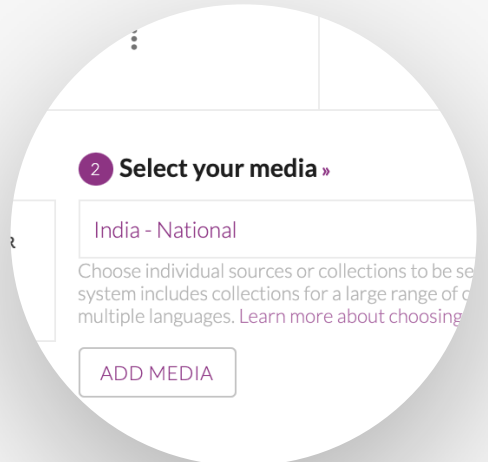
Writing Media Cloud Queries

You can query Media Cloud with boolean searches that match stories, a lot like you Google. Here are some examples:

- Boolean Connectors
 - OR
 - cheese OR cheesy - using "OR" searches for stories that use "cheese" or the word "cheesy," or stories that contain both words
 - Note: OR is the default connector. This means that unless you specifically put quotations around words or add an "AND" or between them, the system will treat your search as though the OR



Noisy Caste-related Text Collection



Noisy Caste-related Text Collection
Keyword-based filtering

Caste-related Story Classifier
BERT-based Classifier

Social Media Narrative Collection
Deduplication, News-Comments mapping

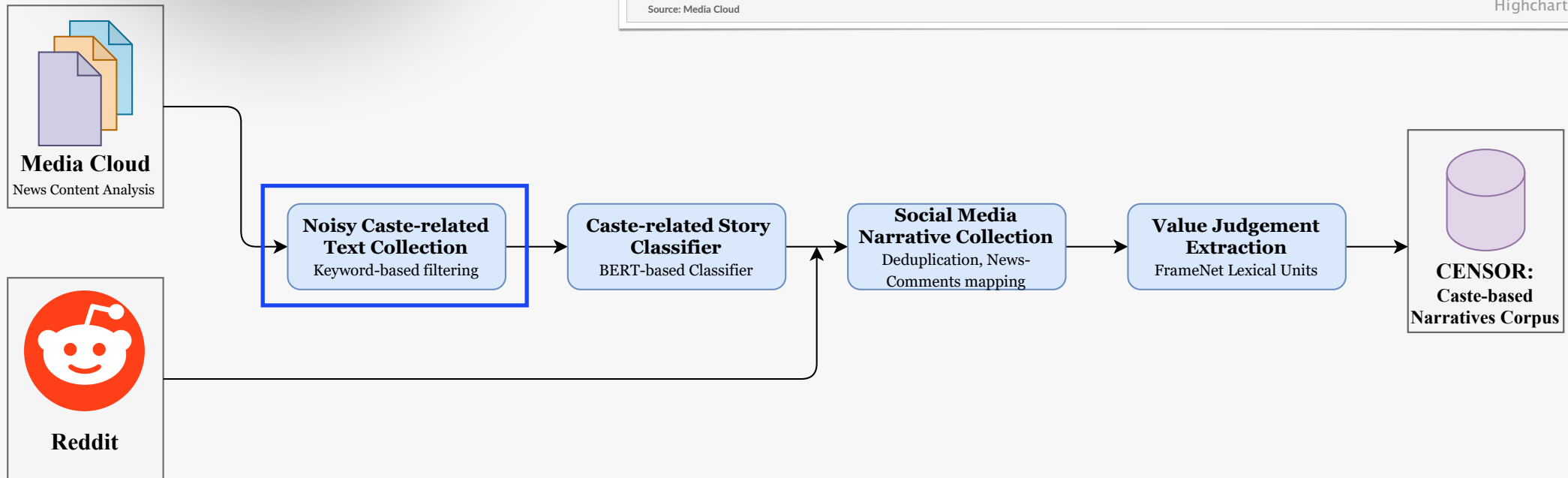
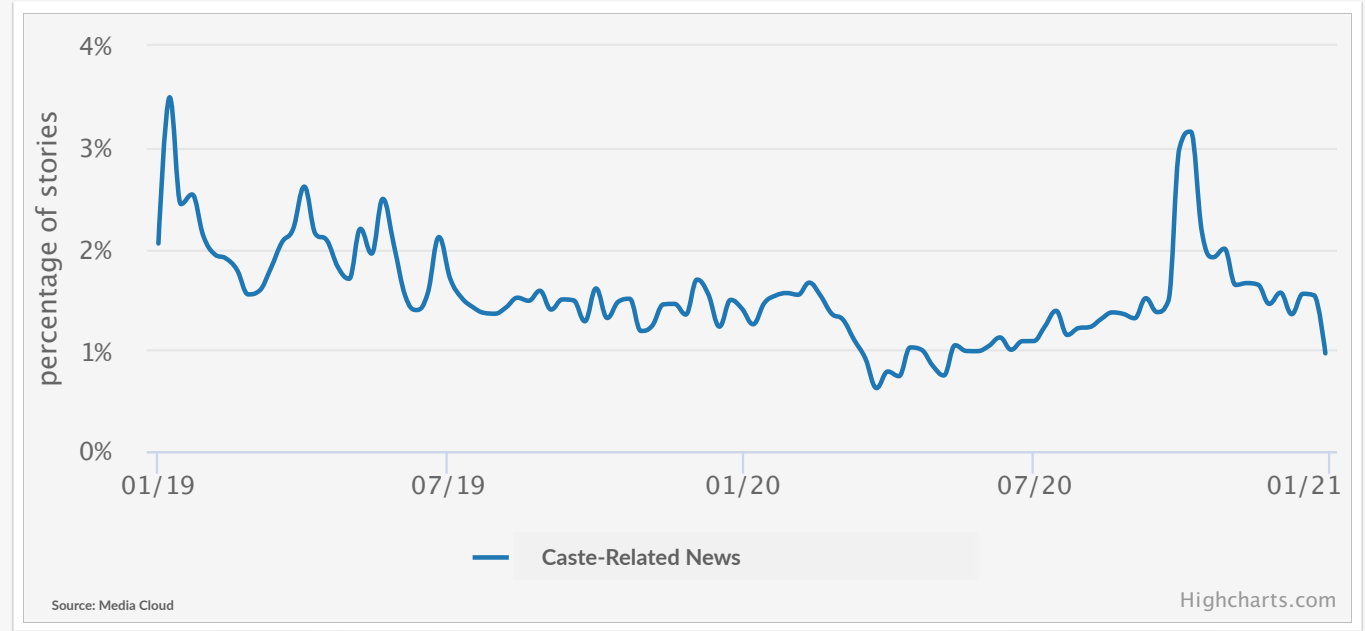
Value Judgement Extraction
FrameNet Lexical Units

CENSOR: Caste-based Narratives Corpus

Noisy Caste-related Text Collection

Attention Over Time

Compare the attention paid to your queries over time to understand how they are covered. This chart shows the number of stories that match each of your queries. Spikes in attention can reveal key events. Plateaus can reveal stable, "normal", attention levels. Click a point to see words and headlines for those dates. Use the "view options" menu to switch between story counts and a percentage.

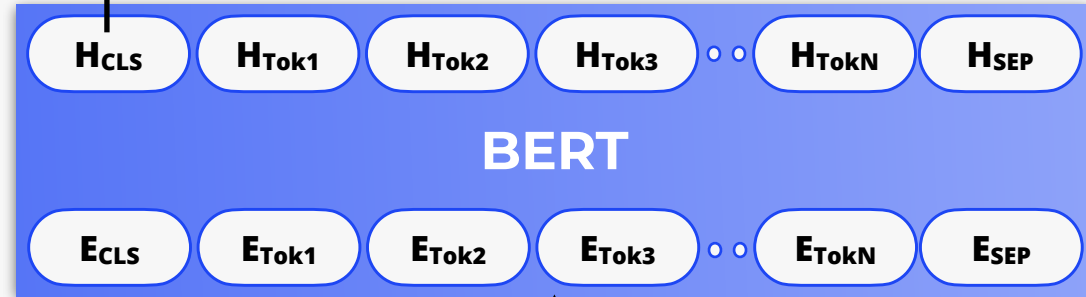


Caste-related Story Classifier

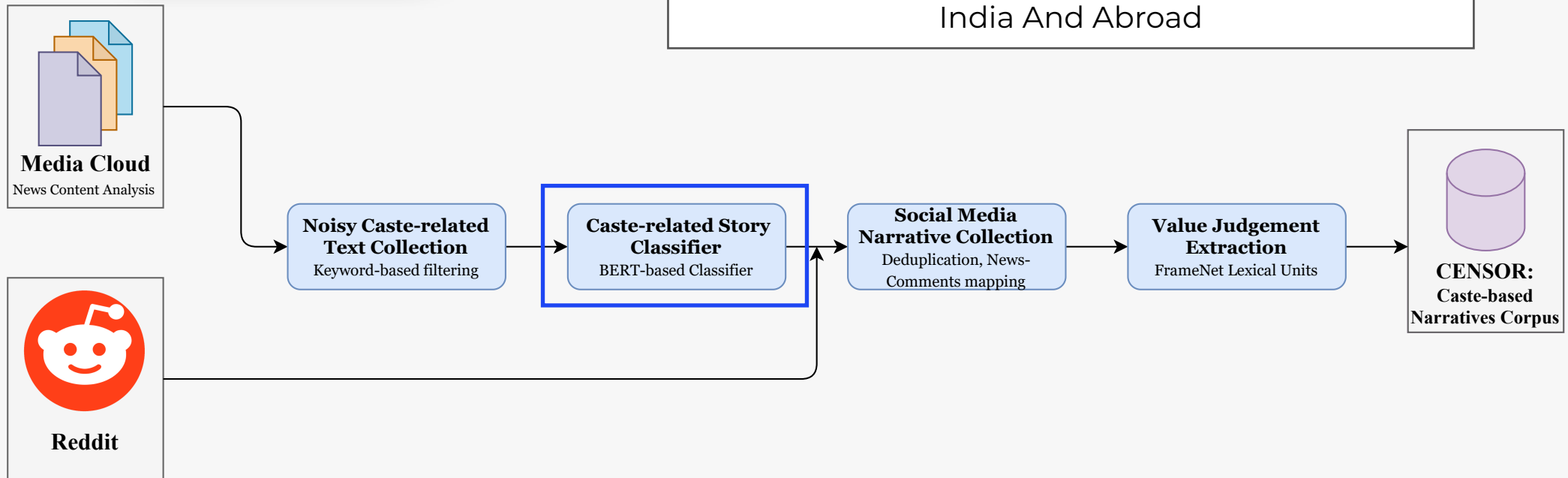
89.6%

F1-Score

Caste-related?

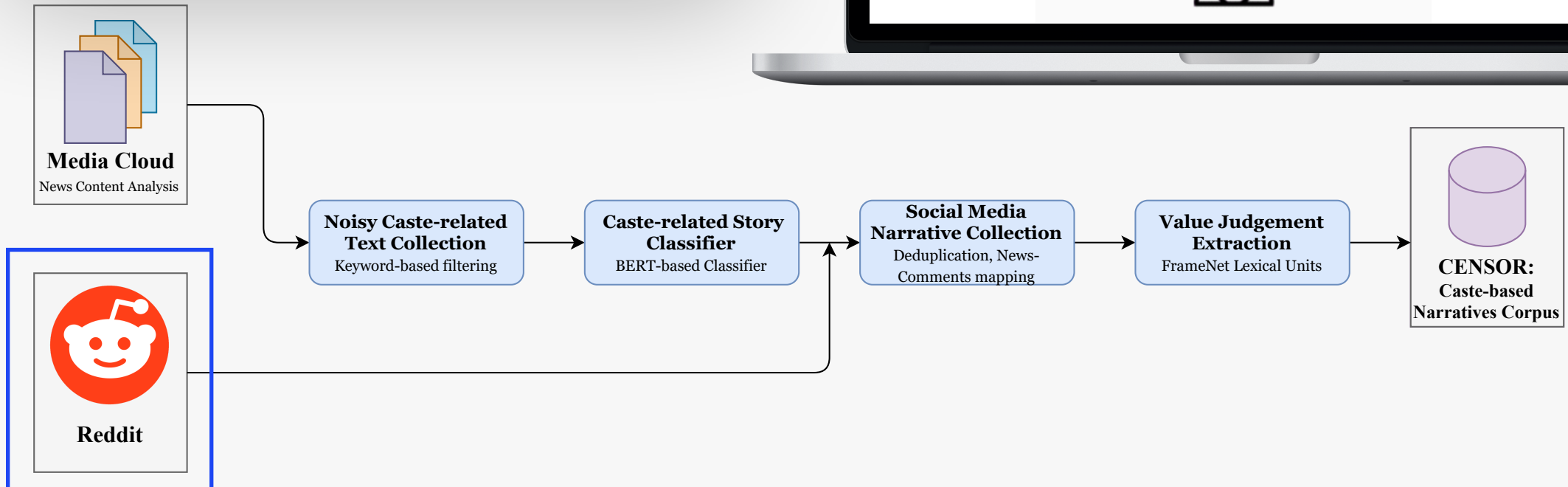


The Evidence Is Clear, Caste Hurts Corporations In India And Abroad

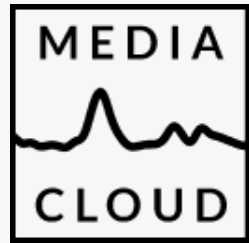


Data Sources

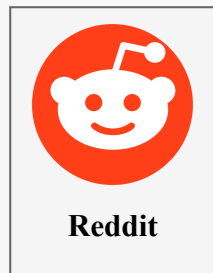
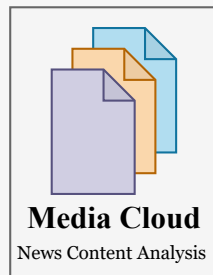
- **PushShift API** is used to find public subreddits based on keywords to view the top subreddits and use other parameters (like NSFW) to discover the top subreddits in various categories.
- **API Endpoints:**
 - /reddit/comment/search
 - /reddit/submission/search
 - /reddit/subreddit/search



Social Media Narrative Collection



The Evidence Is Clear, Caste Hurts Corporations In India And Abroad

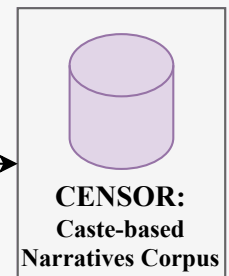


Noisy Caste-related Text Collection
Keyword-based filtering

Caste-related Story Classifier
BERT-based Classifier

Social Media Narrative Collection
Deduplication, News-Comments mapping

Value Judgement Extraction
FrameNet Lexical Units



↑ 22 ↓

r/india · Posted by u/rokosbasslisk 8 months ago

The Evidence Is Clear, Caste Hurts Corporations In India And Abroad

huffingtonpost.in/entry/...

Politics

rokosbasslisk · 8 months ago

>Despite upper castes constituting only 14% in India's overall population share, they occupy over 94% of corporate positions in India

↑ 10 ↓

Share Report Save

justicekatjukatli 8 months ago

No shit.

↑ 3 ↓

Share Report Save

Severe-Boss 8 months ago

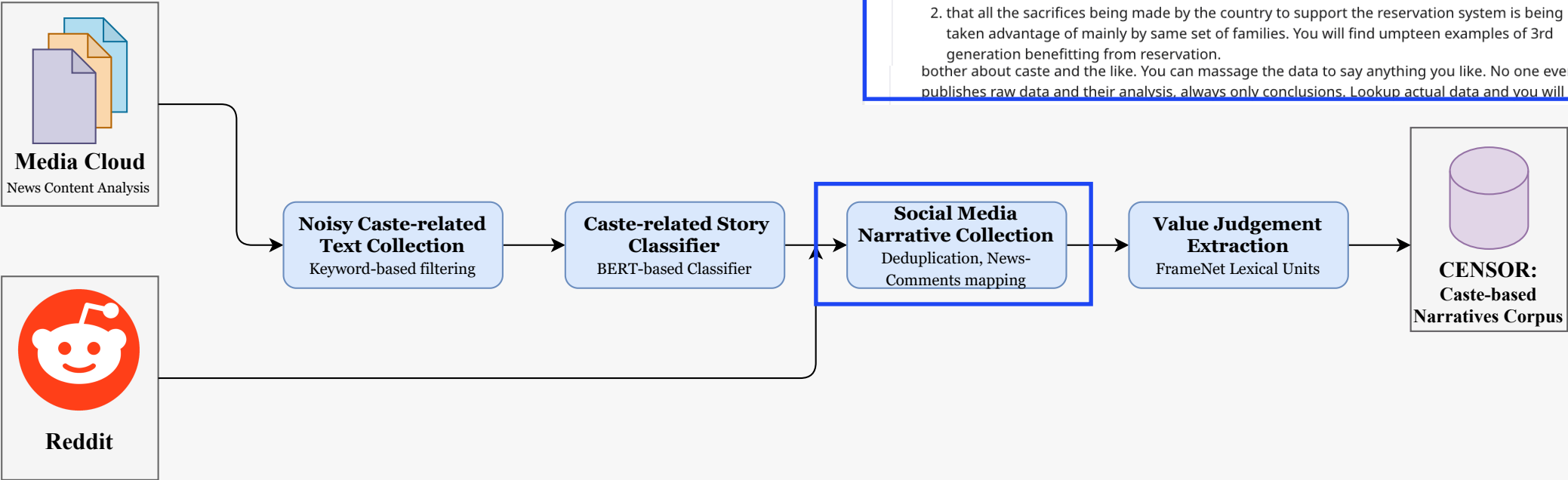
This is such nonsense bunkum. In a globalised competitive world, companies are not going to bother about caste and the like. You can massage the data to say anything you like. No one ever publishes raw data and their analysis, always only conclusions. Lookup actual data and you will know:

1. that Upper castes have higher representation in pvt sectors but lower in public sector because the educated among upper castes didn't have the opportunities in public services
2. that all the sacrifices being made by the country to support the reservation system is being taken advantage of mainly by same set of families. You will find umpteen examples of 3rd generation benefitting from reservation.

bother about caste and the like. You can massage the data to say anything you like. No one ever publishes raw data and their analysis, always only conclusions. Lookup actual data and you will


Social Media Narrative Collection

Dataset Statistics	
#Noisy Caste-related Collection	180,423
#Caste-related Stories	138,848
#Matched Reddit Posts	21,589
#Total Comments	863,560



↑ 22 ↓


r/india · Posted by u/rokosbasslisk 8 months ago



The Evidence Is Clear, Caste Hurts Corporations In India And Abroad


huffingtonpost.in/entry/...

Politics

rokosbasslisk · 8 months ago


>Despite upper castes constituting only 14% in India's overall population share, they occupy over 94% of corporate positions in India

↑ 10 ↓ Share Report Save

justicekatjukatli 8 months ago

No shit.

↑ 3 ↓ Share Report Save

Severe-Boss 8 months ago

This is such nonsense bunkum. In a globalised competitive world, companies are not going to bother about caste and the like. You can massage the data to say anything you like. No one ever publishes raw data and their analysis, always only conclusions. Lookup actual data and you will know:

1. that Upper castes have higher representation in pvt sectors but lower in public sector because the educated among upper castes didn't have the opportunities in public services
2. that all the sacrifices being made by the country to support the reservation system is being taken advantage of mainly by same set of families. You will find umpteen examples of 3rd generation benefitting from reservation.

bother about caste and the like. You can massage the data to say anything you like. No one ever publishes raw data and their analysis, always only conclusions. Lookup actual data and you will

Value Judgements Extraction

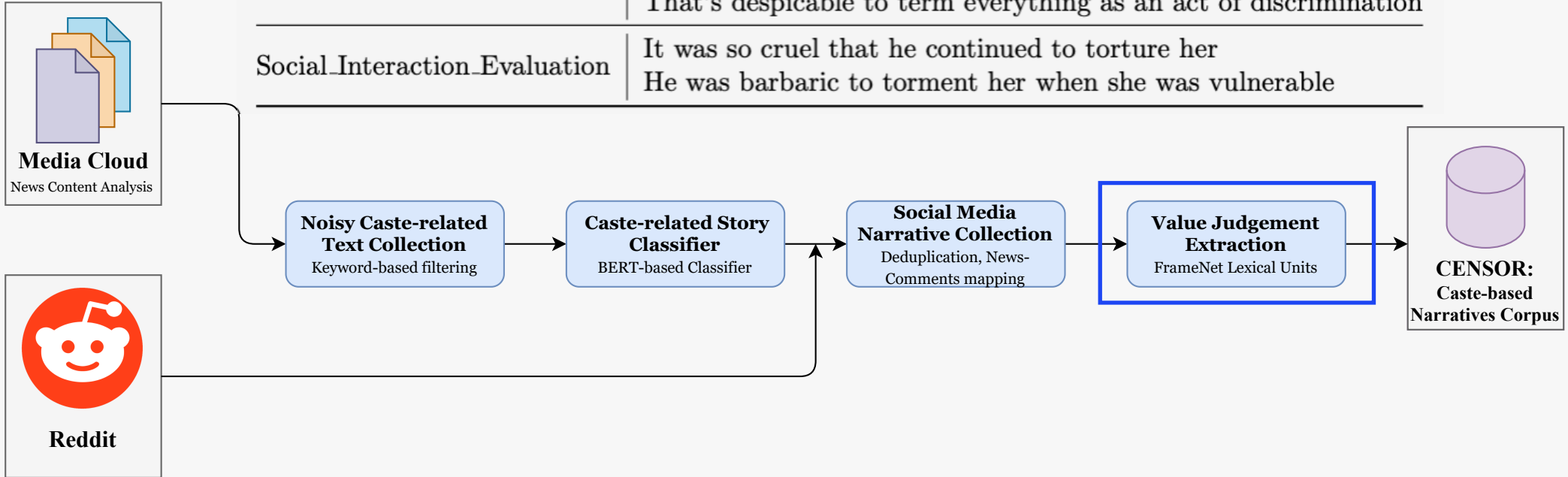
About FrameNet ▾

reddit

Search

Frame-frame Relations:
Inherits from: [Gradable attributes](#)
Is Inherited by: [Attitude description](#), [Being questionable](#), [Compliance](#), [Disgraceful situation](#), [Expertise](#), [Frugality](#), [Mental property](#), [Morality evaluation](#), [Praiseworthiness](#), [Rashness](#), [Social interaction evaluation](#), [Trendiness](#)

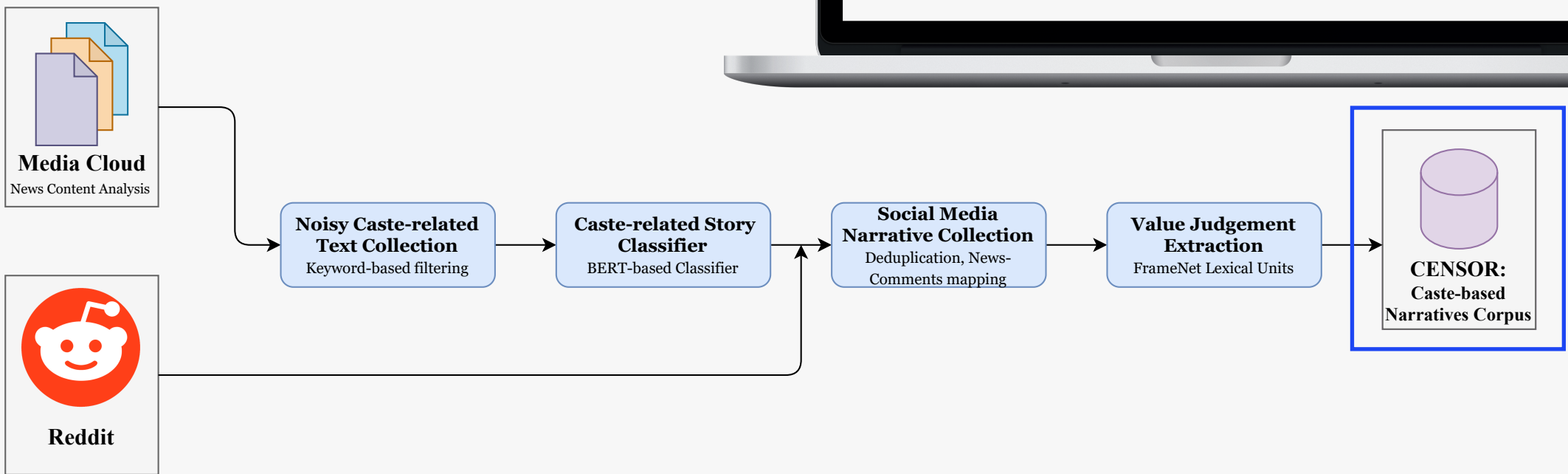
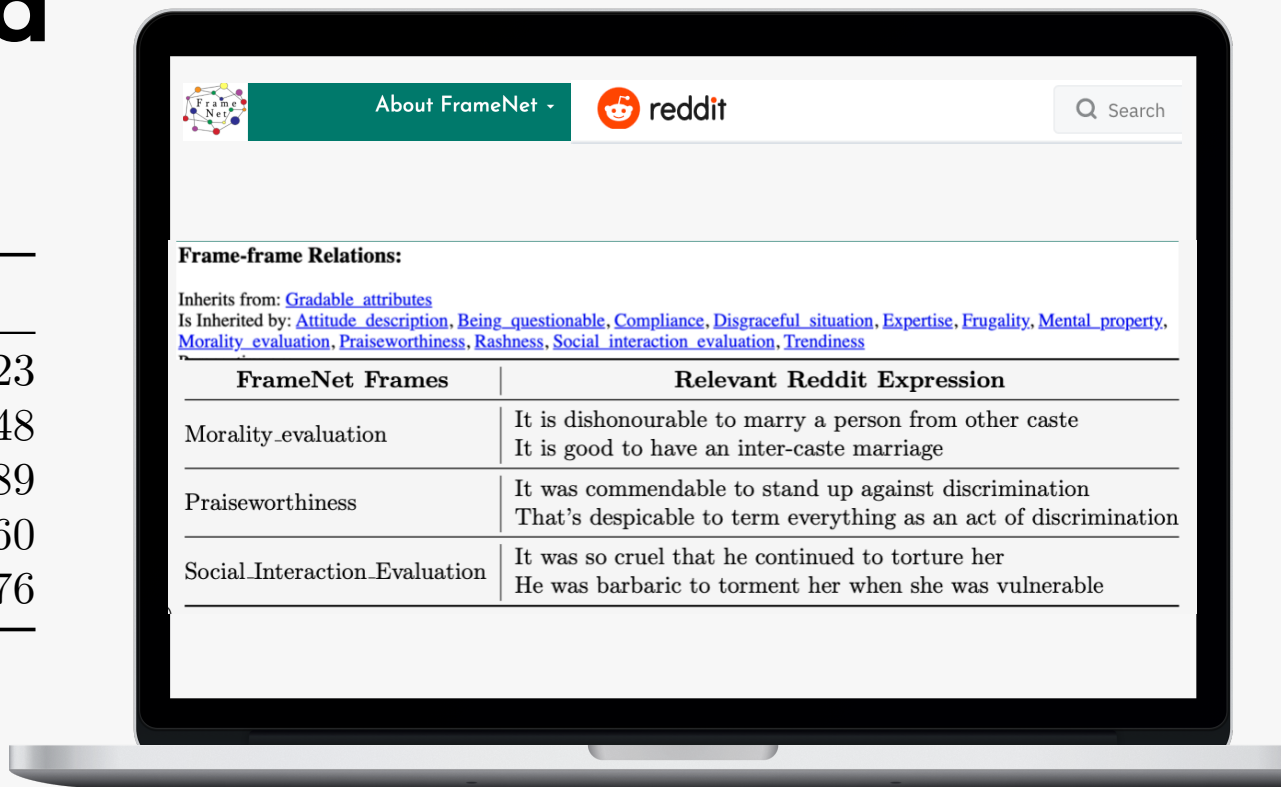
FrameNet Frames	Relevant Reddit Expression
Morality_evaluation	It is dishonourable to marry a person from other caste It is good to have an inter-caste marriage
Praiseworthiness	It was commendable to stand up against discrimination That's despicable to term everything as an act of discrimination
Social_Interaction_Evaluation	It was so cruel that he continued to torture her He was barbaric to torment her when she was vulnerable



CENSOR: Caste-based Narratives Corpus

Dataset Statistics


#Noisy Caste-related Collection	180,423
#Caste-related Stories	138,848
#Matched Reddit Posts	21,589
#Total Comments	863,560
#Total Comments w/ Value Judgements	118,776



Ongoing Modeling Work:

Embedding Comments

=



rokosbasslisk

8 months ago


>Despite upper castes constituting only 14% in India's overall population share, they occupy over 94% of corporate positions in India

10

Share

Report

Save



justicekatjukatli

8 months ago


No shit.

3

Share

Report

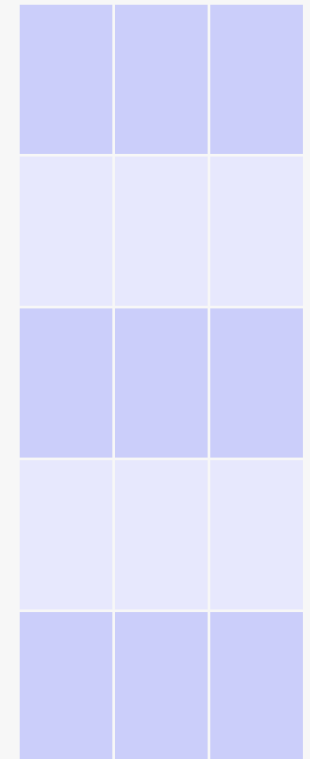
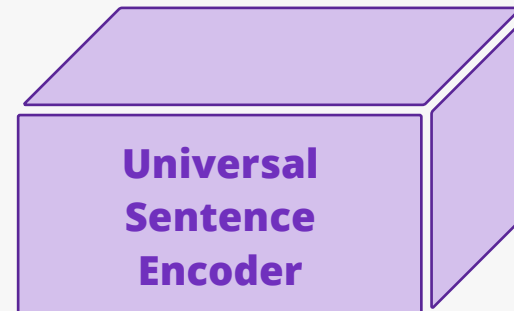
Save



Severe-Boss

8 months ago

This is such nonsense bunkum. In a globalised competitive world, companies are not going to bother about caste and the like. You can massage the data to say anything you like. No one ever publishes raw data and their analysis, always only conclusions. Lookup actual data and you will know:



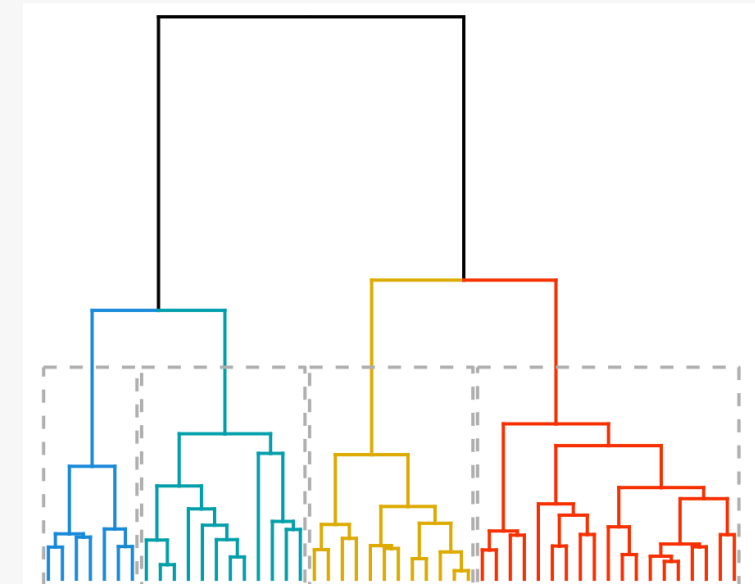
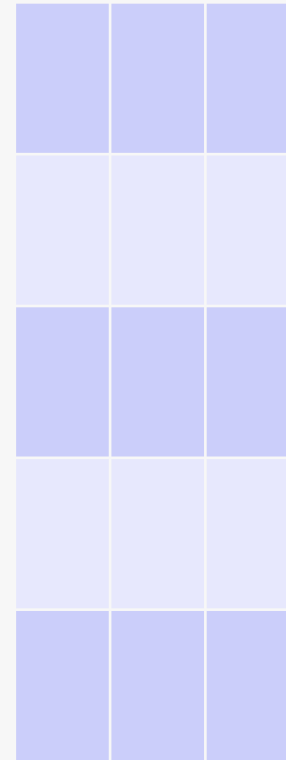
Ongoing Modeling Work: Hierarchical Clustering

=

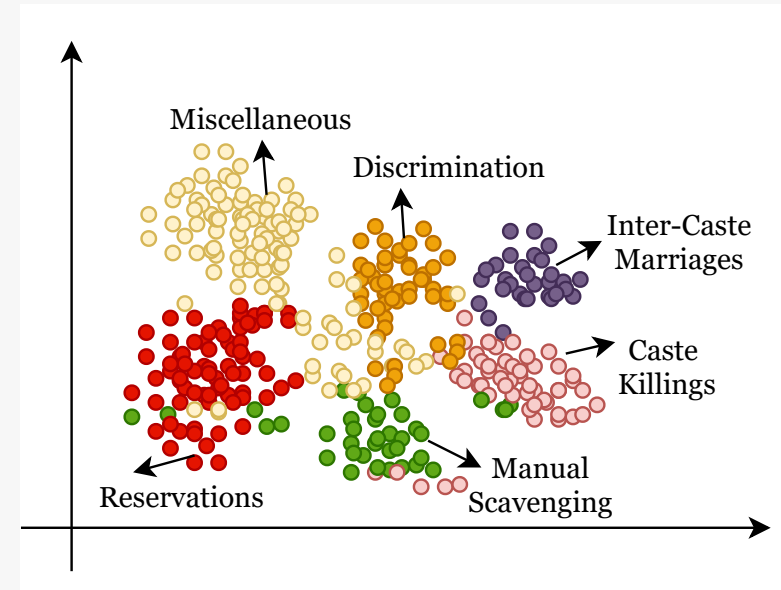
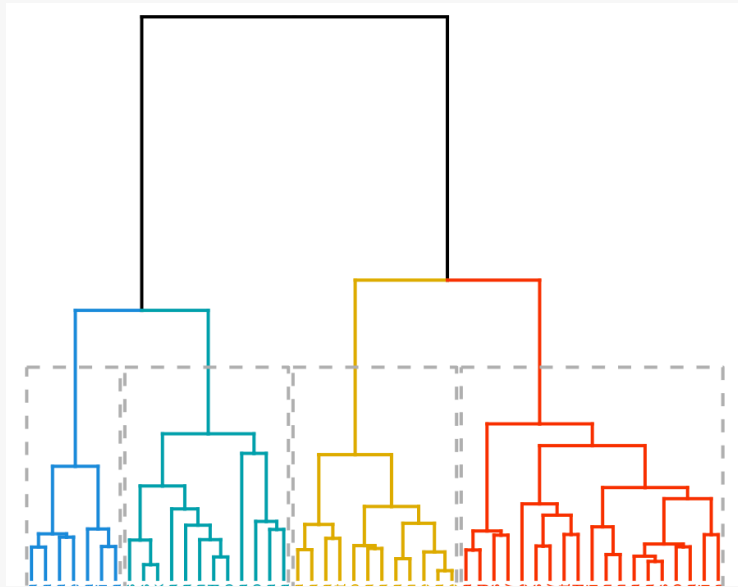
rokosbasslink 8 months ago
>Despite upper castes constituting only 14% in India's overall population share, they occupy over 94% of corporate positions in India
↑ 10 ↓ Share Report Save

justicekatjukatti 8 months ago
No shit.
↑ 3 ↓ Share Report Save

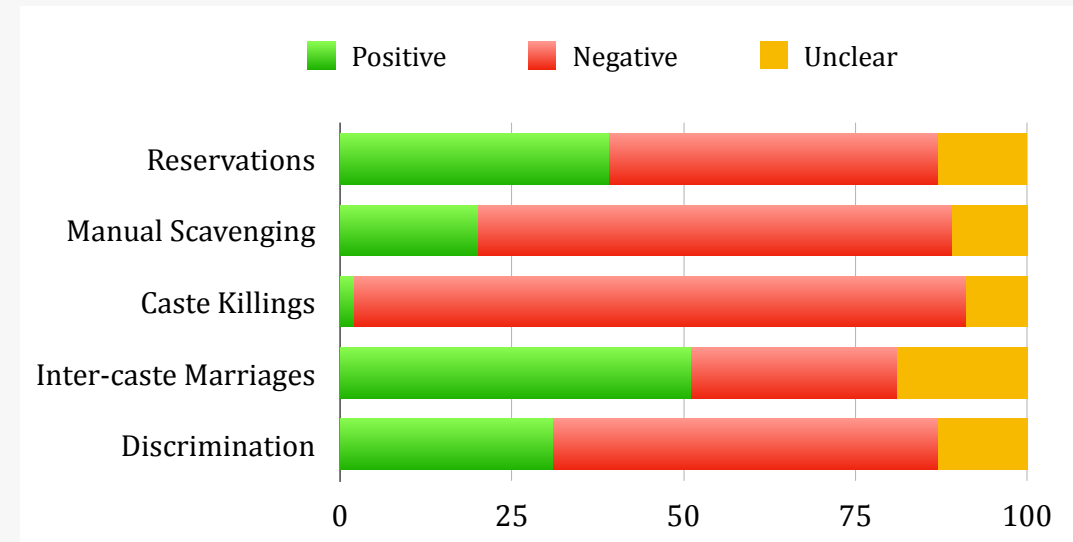
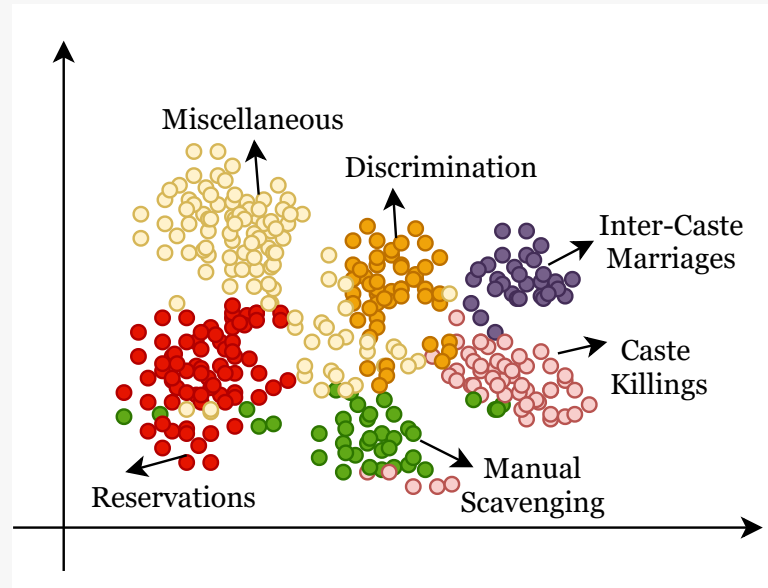
Severe-Boss 8 months ago
This is such nonsense bunkum. In a globalised competitive world, companies are not going to bother about caste and the like. You can massage the data to say anything you like. No one ever publishes raw data and their analysis, always only conclusions. Lookup actual data and you will know:



Ongoing Modeling Work: Hierarchical Clustering

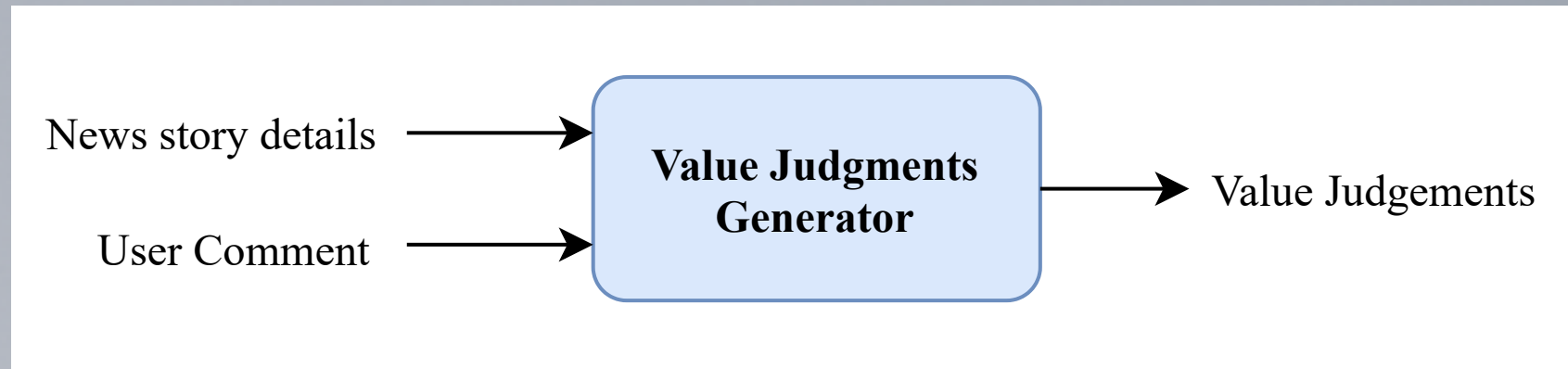


Ongoing Modeling Work: Emotional Valence



Ongoing Modeling Work:

Automatic Value Judgement Extractor/Generator



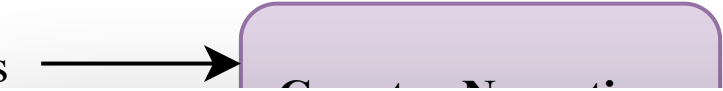
Ongoing Modeling Work:

Generate Counter Narratives

News story details

User Comments

Value Judgements



Counter-narratives

- Retrieval and Generation methods
- Generation methods — data and decoding strategies.
- Pretrained LMs using DAPT/PPLM
- Automatic and Manual Evaluation Metrics

1. Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. Plug and play language models: A simple approach to controlled text generation. arXiv preprint arXiv:1912.02164, 2019.
2. Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A Smith. Don't stop pretraining: Adapt language models to domains and tasks. arXiv preprint arXiv:2004.10964, 2020.



Thank You!

Contact: pralav@mit.edu