


**SPEAK OUT AND BE HEARD**

# **UNCOVERING THE SHADOWS OF GENDER-BASED VIOLENCE AMONG FILIPINO REDDITORS**

This initiative seeks to explore and expose the experiences of gender-based violence as shared by individuals in Filipino Reddit communities. Our project analyzes the nature of these experiences, investigates patterns based on demographics, geography, and special



dates, and examines the impact of community engagement on these critical issues.

 Explore Our Study

# OUR RESEARCH FOCUS

Overview   Problem   Hypothesis   Solution

Gender-based violence is a critical issue that affects communities worldwide, and social media platforms often become spaces where these experiences are shared. Our project focuses on analyzing narratives around gender-based violence within Filipino Reddit communities. By examining these discussions, we aim to identify patterns and insights that could inform better responses and preventative measures.

title	body	link	author	created	upvotes	subreddit	search_input
['I, "nt", 'care', 're', 'gift', 'unsolicited', 'pics', 'unsolicited', 'pics', '']	['So', 'like', 'many', 'people', 'subreddit', 'T, 'take', 'nudes', 'T, 'posted...']	<a href="https://www.reddit.com/r/alwaysjicy/comments/1086dx/l_dont_care_if_youre_a_girl_unsolicited_pics_are/">https://www.reddit.com/r/alwaysjicy/comments/1086dx/l_dont_care_if_youre_a_girl_unsolicited_pics_are/</a>	t2_uup1eor9	2023-01-10 18:35:34	127	alwaysjicy	unsolicited pics
['Among', 'meron', 'sa', 'unsolicited', 'pics', '']	['Hello', 'T, 'So', 'curious', 'lang', 'ako', 'sa', 'sagot', ...]	<a href="https://www.reddit.com/r/alwaysjicy/comments/124xpta/among_meron_sa_unsolicited_pics/">https://www.reddit.com/r/alwaysjicy/comments/124xpta/among_meron_sa_unsolicited_pics/</a>	t2_t6imt3mv	2023-03-28 23:40:43	10	alwaysjicy	unsolicited pics
['I, 'felt', 'thrown', 'bus', 'abused', 'exhausted']	['Nagkwento', 'ako', 'sa', 'ate', 'ko', 'pero', 'iba', 'lang', ...]	<a href="https://www.reddit.com/r/OffMyChestPH/comments/1ajcqqj/i_fell_thrown_under_the_bus_abused_and_ano/">https://www.reddit.com/r/OffMyChestPH/comments/1ajcqqj/i_fell_thrown_under_the_bus_abused_and_ano/</a>	t2_96uz8figs0	2024-02-05 17:39:56	3	OffmychestPH	abused
['Cyberbullying']	['Parang', 'na', 'normalize', 'sa', 'sa', 'twitter', ...]	<a href="https://www.reddit.com/r/OffMyChestPH/comments/183h0q/cyberbullying/">https://www.reddit.com/r/OffMyChestPH/comments/183h0q/cyberbullying/</a>	t2_j4epqse2o	t2_j4epqse2o	2	OffmychestPH	Cyberbullying
...	...	...	...	...	...	...	...

## ABOUT OUR DATASET

## DATA COLLECTION PROCESS

The posts were filtered based on specific search inputs related to different forms of gender-based violence. The search inputs used are as follows:

abused

sexual harassment

Cyberbullying

domestic violence

emotional abuse

physical abuse

unsolicited pics

stalking

The originally scraped dataset contains:

- **title:** The title of the post.
- **body:** The main content of the post.
- **link:** The link URL that directs to the post.
- **author:** Author of the post
- **created:** Date when the post was created.
- **subreddit:** The subreddit from which the post originated.
- **searchinput:** The keyword used for the search input.

## PREPROCESSING

Our preprocessing efforts aim to refine the raw dataset into a clean, analyzable format that supports reliable data

The dataset was collected by scraping posts from Philippine subreddits and using keywords related to various forms of gender-based violence for the search inputs. The scraped data had 1,900+ rows in total.

preprocessed\_dataset.csv

ph\_holidays.csv

analysis and insight generation. This process ensures data quality and facilitates effective data exploration.

## STEPS IN PREPROCESSING

### 1. Cleaning Text Data

We strip out irrelevant characters, such as URLs, special characters, and formatting elements from the posts' titles and bodies to focus purely on textual content. Text data is converted to a uniform case (all lower case) to ensure consistency across all text entries.

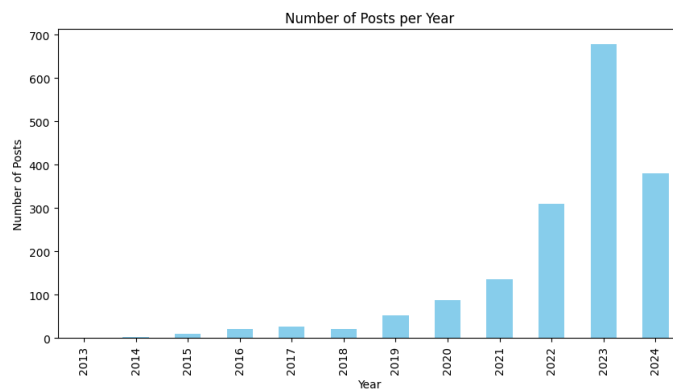
▼ Read More

## EXPLORATORY DATA ANALYSIS

POST PER YEAR   POST PER DAY   MONTHLY ANALYSIS  
WEEKLY ANALYSIS

# POSTS PER YEAR

Our team has first explored the number of reddit posts per year that mentions gender-based violence keywords.



The chart shows a general increase in the number of posts from 2013 to 2024. There is a noticeable spike in posts starting from 2020, with the highest number of posts recorded in 2023.

Based on the chart, we can infer the following detailed answer to the key question regarding temporal patterns:

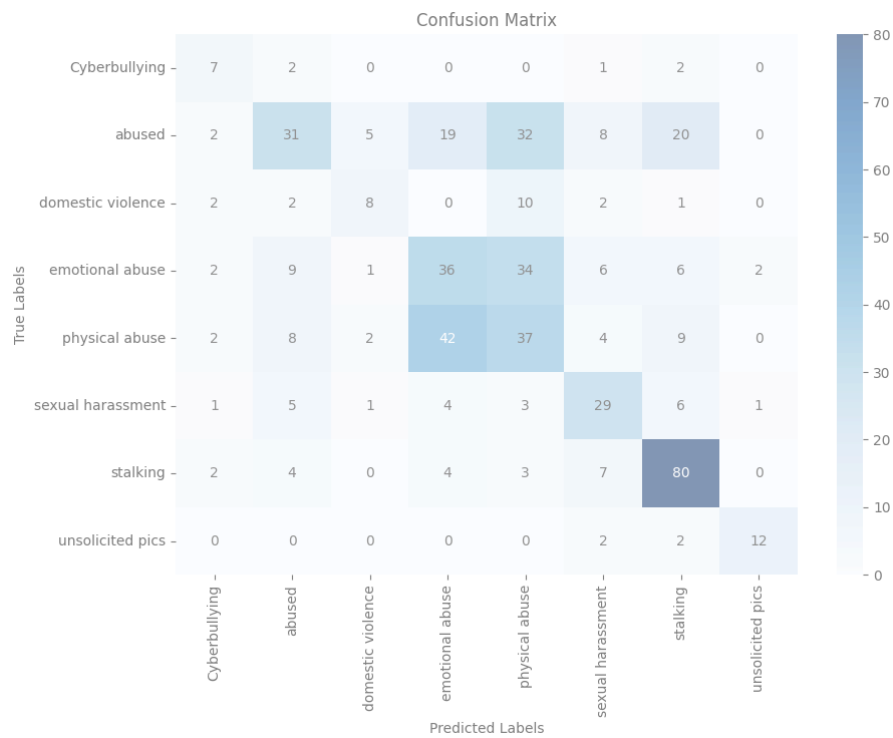
1. **Gradual Increase Over Time** — From 2013 to 2019, the number of posts regarding gender-based violence remains relatively low and stable, indicating that these discussions were not as prominent or not as frequently shared on Filipino Reddit communities during these years.
2. **Significant Spike Starting 2020** — Beginning in 2020, there is a marked increase in the number of posts. This could be attributed to various factors such as increased awareness, more active community engagement, or external societal events prompting more discussions on this topic. The year 2020 itself might have seen more discussions due to the global COVID-19 pandemic, which had significant social impacts, including increased domestic stress and potential increases in gender-based violence incidents.
3. **Peak in 2023** — The year 2023 records the highest number of posts. This peak suggests a continued and growing awareness and willingness to discuss gender-based violence on these platforms. It could also

indicate that the community engagement has become stronger, or there might have been significant events or movements related to gender-based violence that prompted more discussions.

The increasing trend of posts over the years, especially the spike starting in 2020, indicates that:

1. **Growing Awareness** — There is increasing awareness and discussion about gender-based violence in Filipino Reddit communities.
2. **Community Engagement** — Community engagement on this issue has likely improved, as more people are willing to share their experiences and discuss this topic.
3. **External Factors** — External factors, such as the COVID-19 pandemic or significant social movements, could have influenced the increase in posts.

## IN A NUTSHELL



### HIGH ACCURACY

“Stalking” and “Unsolicited Pics” categories have the highest number of correct classifications (80 and 12, respectively), indicating the model is highly accurate in predicting these categories.

### MODERATE ACCURACY

“Cyberbullying,” “Sexual Harassment,” and “Physical Abuse” categories have moderate accuracy, with significant correct classifications but also notable misclassifications.

### LOW ACCURACY

“Abused,” “Domestic Violence,” and “Emotional Abuse” categories have lower accuracy, with many instances misclassified as other types of violence. For example, “Abused” is frequently misclassified as “Physical Abuse,” and “Emotional Abuse” is often misclassified as “Physical Abuse.”

## MEET THE TEAM

### ARIAN REY

Hi there! I'm Arian Rey. 🙋 I am a 4th-year Computer Science student at the University of the Philippines - Diliman, specializing in UI/UX

### CESS VENTURES

Just a simple girl and a CS132 Student

### CHRISTIAN QUINZON

Just a simple girl? and a CS132 Student

design and software development. Currently, I am passionate about developing Keller, a browser extension that helps low-vision and blind people generate alternative text for images.

With this project, I aim to raise awareness and increase efforts to make technology more accessible in the Philippines. I've also worked on other cool projects. If you're interested, please visit my website to check them out:

<https://www.arianrey.com/>