

Big Data Project Report

SDG 5: Gender Equality

Topic	Page
Introduction	1
SDG and Background	1
Data	1
Methods	1
Results	2
Discussion	3
Conclusion	3
Team Member Contributions	4

Team

Jatin Garg (111462753)
Rahul Bhansali (111401451)
Aditya Tomar (111491409)
Shweta Bharti (111464627)

Analysis of Cyber Bullying Against Women on Twitter

1. Introduction

Violence and abuse in social media is a globally pervasive phenomenon affecting both developing and developed countries. In particular, violence against women in social media is a topic for which there has not been enough case studies performed. The property that this form of violence is “virtual” does not make it any less real. Online abuse and violence can lead to various physical, mental and emotional problems for women. Some of the impacts include, depression, anxiety, loneliness, low self-esteem as well as poor health and diseases. Thus, we as a society need to tackle this form of violence faced by women and try to come up with ways to reduce or put an end to this violence. So, in this project, we examine and explore violence and abuse faced by women in social media, specifically, on Twitter. Using Big Data technologies is a key aspect, if we want to understand Twitter violence against women as the volume of Twitter data can be huge. The motivation for the project is primarily driven by United Nations’ *Sustainable Development Goals(SDGs)* #5 which is Achieving Gender Equality and Empowering All Women and Girls.

2. Sustainable Development Goal and Background

The Sustainable Development Goals (SDGs) are a collection of 17 goals set by the United Nations. The SDGs cover a variety of social and economic issues such as poverty, hunger, climate action and health. Our project is based on SDG #5 which is Achieving Gender Equality and Empowering all Women and Girls. It consists of 9 broad targets as well as various indicators for each target to determine the progress made for each target. Some of the targets for SDG #5 include: (Target 5.1) ending all forms of discrimination against women, (Target 5.2) all forms of violence against women and girls everywhere and (Target 5.3) eliminate all forms of harmful practices, such as child marriage, early and forced marriage. Our project directly aligns itself to Target 5.2 as it analyzes violence faced by women on Twitter.

3. Data

We primarily used Twitter Data for analyzing online violence against women as raw data. We

collected around 8GB Twitter Data across a span of a week using the *Tweepy* Python library. Twitter Data consists of a *Json Object* corresponding to a tweet. The *Json Object* has sections corresponding to a tweet in a *Tweet Object (TO)*. *User Object* containing details related to user who produced the tweet and other fields. We filtered out the unnecessary attributes from the *Json Object* and converted it into a csv file using various Spark transformations and actions.

The number of tweets analyzed was approximately one million. For creation of a sentiment analyzer, another 600,000 tweets of a training data set were used.

4. Methods

One of the critical things that we need to figure out to analyze abuse and violence against women on Twitter is to classify a tweet as being violent against women or not. For this, we created a sentiment analyzer model that categorizes whether a tweet is violent against women or not.

The model is based on a neural network and we used the Keras library, TensorFlow and Word2Vec technologies to build it. The sentiment analyzer was trained on a training data set that contained approximately 600,000 tweets and for each tweet had a label of 1 if the tweet was deemed “violent against women” or 0 if the tweet was deemed “non-violent against women”. For creating the training dataset, we filtered out tweets from Twitter that contained expletives or other forms of abusive or insulting words towards women. We categorized such filtered tweets as violent and assigned them a value of 1. To get tweets that are “neutral” or “non-violent” we used the *Sentiment140* dataset. Overall, we had 600,000 tweets and we trained the model on this training data. After training, we tested the model on a completely unseen dataset (taken from the *Sentiment140 dataset*). The accuracy of the sentiment analyzer model was 61%.

Now, we used the sentiment analyzer to filter out the violent tweets against women from the raw data set.

We now wanted to classify the violent tweets into “categories. We came up with three

Analysis of Cyber Bullying Against Women on Twitter

categories of violence: *Mildly Severe*, *Moderately Severe* and *Extremely Severe*. Mildly Severe violent tweets involve name calling or insults. Moderately Severe violent tweets involve sentiments of slut-shaming. Extremely Severe violent tweets involve sentiments of molestation threats, death and rape threats, sexual harassments.

Given a tweet to the sentiment analyzer, it returns the probability [0,1] with which it thinks that a given tweet should be labelled “violent” or “non-violent”. So, we chose a threshold value of 0.7 for a tweet being labelled violent and classified the violent tweets into three distinct categories:

- Mildly Severe Violent Tweets: Probability of being violent is in the range of [0.7,0.8)
- Moderately Severe Violent Tweets: Probability of being violent is in the range of [0.8,0.9)
- Extremely Severe Violent Tweets: Probability of being violent is in the range of [0.9,1]

Using the above classification, we were able to categorize the violent tweets into separate categories.

We also wanted to find out the region-wise distribution of the violent tweets. For this we, used python matplotlib library with basemap library of world shapefile. We grouped the counts of tweets by cities and plotted a world heat map of violent tweets by region and its count. In order to find the location of cities on the world map, Google's geolocator api is used, which gives the longitude and latitude of the city on the map.

Using this, we were able to plot world maps, to view the distribution of violent tweets across the globe.

5. Results

We found interesting results during our analysis of violent tweets against women:

Total Number of Violent Tweets against women: 123,112

Total Number of Tweets Analyzed: 439,686

Percentage of Violent Tweets against women: 28%

Percentage of Violent Tweets against women of category:

- Mildly Severe: 4.91%
- Moderately Severe: 7.49%
- Extremely Severe: 15.9%

We plotted the following graphs:

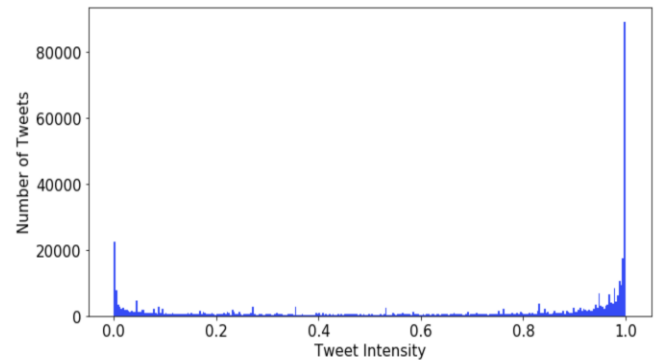


Fig -1: Probability Distribution Graph of Number of Violence Intensity of a Tweet (x-axis) v/s Number of Tweets (y-axis)

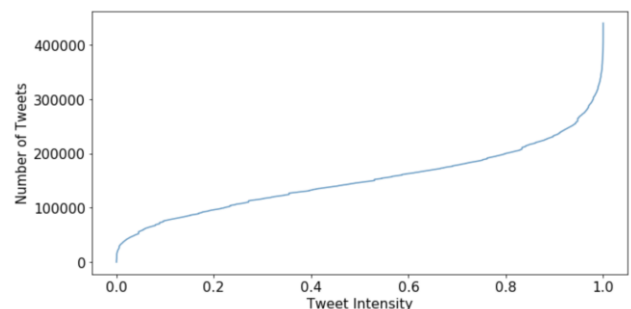


Fig-2: Cumulative Probability Distribution Graph of Violence Intensity of a Tweet (x-axis) v/s Number of Tweets (y-axis)

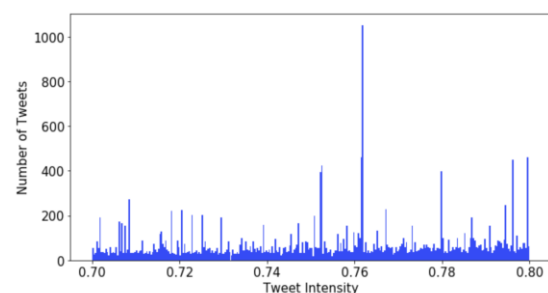


Fig-3: Distribution of Mildly Severe Violent Tweets.

Analysis of Cyber Bullying Against Women on Twitter

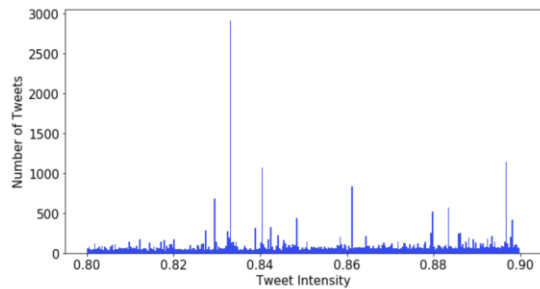


Fig-4: Distribution of Moderately Severe Violent Tweets.

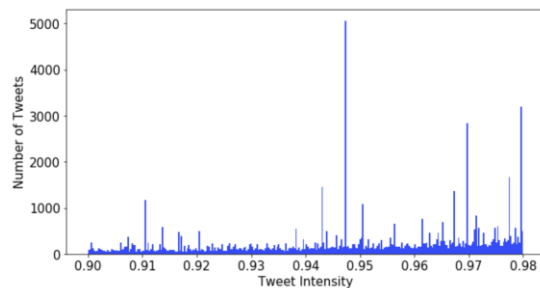


Fig-5: Distribution of Extremely Severe Violent Tweets.

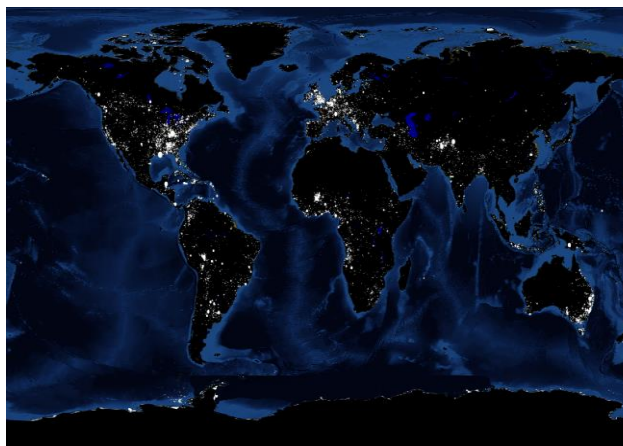


Fig-6: World map showing the distribution of violent tweets around the globe.

Here, the intensity of white colored portions signifies higher intensity of violent tweet against women.

Next, we tried to analyze the severity level of a person by analyzing his past tweets. So, this we gathered tweets according to the username using python script. We used our sentiment analyzer to

assign an intensity score to each of the tweet of the user and after that we averaged it out to get a final score. We used the same criterion that we used to categorize the tweets. A score of less than 0.7 means the person is neutral, otherwise the person is not neutral with increasing level of intensities (Level 1,2 and 3).

5.1 Interesting Results

We evaluated the score for certain users like Barack Obama and Donald Trump. We found that Barack Obama was neutral while Donald Trump was very mildly Level-1 intense tweeter. The score achieved by them are as follows:

Name	User Name	Intensity Score	Level
Barrack Obama	BarackObama	0.649	Neutral
Donald Trump	realDonaldTrump	0.716	Level 1
Pope Francis	pontifex	0.66	Neutral
Piers Morgan	piersmorgan	0.73	Level-1

6. Discussion

The world map of distribution of violent tweets against women and violent tweet data according to regions can be used by various countries or organizations to figure out the places or regions that need specific focus to curb online violence against women in those regions and possibly even form stricter laws and regulations to keep a check on online violence.

7. Conclusion

- Curbing online violence against women is necessary as online violence can have short and long term adverse effects on women.
- Sentiment Analyzer to classify a tweet as non-violent v/s violent against women was created.
- In total, 1 million tweets were used to train and test the sentiment analyzer, as well as to analyze online violence against women.
- We categorized violent tweets into three categories: mildly severe, moderately severe, extremely severe.

Analysis of Cyber Bullying Against Women on Twitter

- e) We plotted world maps to show distribution of violent tweets across the globe.
- f) Organizations and countries can make use of the world maps and region-wise distribution of violent tweets to concentrate on specific regions and maybe implement stricter laws or regulations in those areas.

Reference Links:

1. <https://ahmedbesbes.com/sentiment-analysis-on-twitter-using-word2vec-and-keras.html>
2. <http://help.sentiment140.com/for-students/>