Shakti : Portal for Detection of Sexual Harassment and Chat Predators

Dreamy Pujara, 202211005, Vipasha Vaghela, 202211002,

Abstract—**The vast increase in using social media sites like Twitter and Facebook led to frequent sexual harassment on the Internet, which is considered a major societal problem. The project aims to predict the types of sexual harassment given a description of the incident that may belong to one, more or none of the three categories of sexual harassment namely commenting, groping/touching, and ogling/staring. We used Random Forest with binary relevance. Word representations are carefully reviewed in text specific to mapping to real number vectors. The chat sexual predators Detection approach with the proposed model.**

Index Terms—**Sexual Harassment, NLP, Random Forest etc**

I. Introduction

O UR aim in this project is to use Machine Learning Algorithms to classify the text description that a victim has experienced into category of sexual harassment (Commenting/Ogling/Groping). This is a Multi Label Problem which means the description may belong to one class , more than one class or no class at all.

The data was collected from safecity.in . Safecity is a platform as a service product that powers communities, police and city government to prevent violence in public and private spaces.

II. Matrices Used

For judging the performance of the model we will be using 2 metrics :

F1 Score , Precision and Recall Matrix. We need to know the fraction of data points predicted to be in a certain class (precision) and also of all the data points predicted what fraction of them were correctly classified(recall) .

As a constraint for us False positives and False Negatives can be very costly , hence we are using F1-Score as a performance metric. Hamming Loss. Since we are having a multi label classification problem , hamming loss will help was predict the average fraction of labels that is incorrectly predicted.

Description	Commenting	Ogling/Facial Expressions/Staring	Touching /Groping
Was walking along crowded street, holding mums...	0	0	1
This incident took place in the evening.I was ...	0	1	0
I WAS WAITING FOR THE BUS. A MAN CAME ON A BIK...	1	0	0
Incident happened inside the train	0	0	0
I witnessed an incident when a chain was bruta...	0	0	0

Fig. 1. Displays the classification of various sentences into one or multiple classes as it is a multi-class problem.

III. Data Analysis

Since this is a multi-label classification, there is obviously data overlap and the data points that do not belong to any class are missed out from here. So this can be converted to a multi-class problem (for the sake of analysis). The data distribution is observed to understand the depth of the imbalance situation.
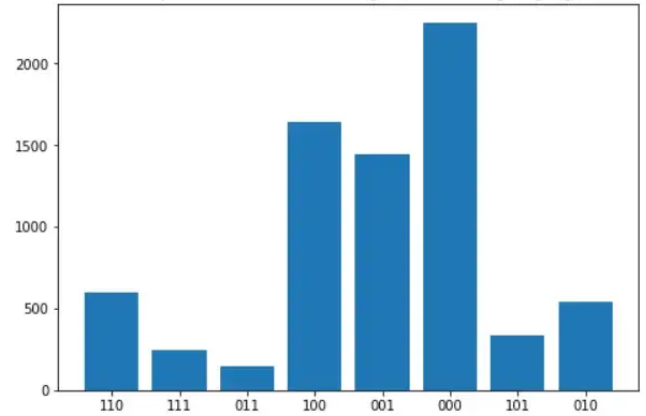

Figure 2

Now this is alarming , out of 7200 points 2000+ points are not classified to any class [0,0,0] . Also there are only very few data points where 2 or more classes coexist.

A boxplot was plotted to see the distribution of lengths of the sentences , and it was observed that 95% of the descriptions lie in and under 25 words but there are descriptions that have more than 400 words also , so there is a very skewed distribution of description lengths.
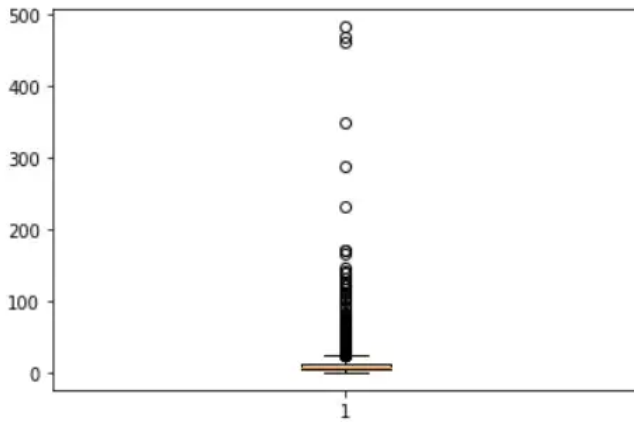
Figure 3

IV. Data Cleaning and Preprosessing

The most basic data clean up where the stop words were removed (a , the , am , are ..), any special symbols, and contractions (don't -¿ do not) .

It was observed that there are words like look , looking , looked which basically means the same thing but are in just different forms. So it could be extremely beneficial if I performed stemming on the data where words like [Looking , looked , looks , look] will be converted to a basic form [look].

when it was checked to see if it has any impact , It was observed that the same final model gave significantly lower performance without stemming.

Figure 4

V. Exploratory Data Analysis

first the balance of the target classes are observed. As is evident from the below graph, the data is highly imbalanced in all three datasets(train, test, and validation) we have descriptions that are not actual incidents way more than the ones that are harassment incident descriptions.
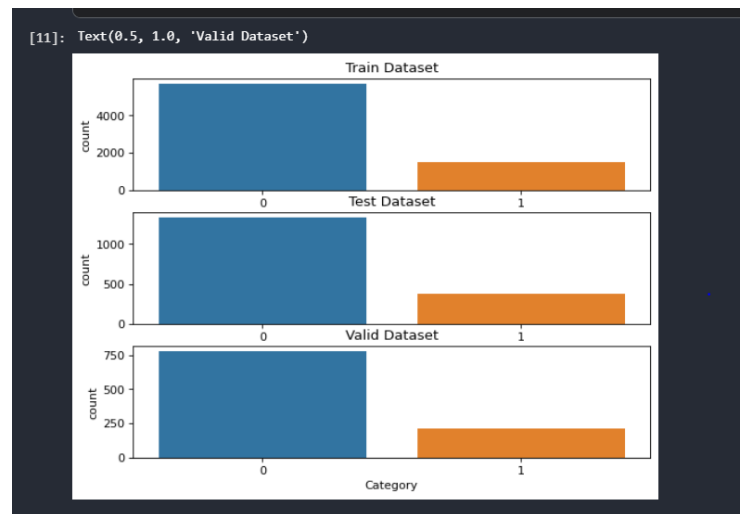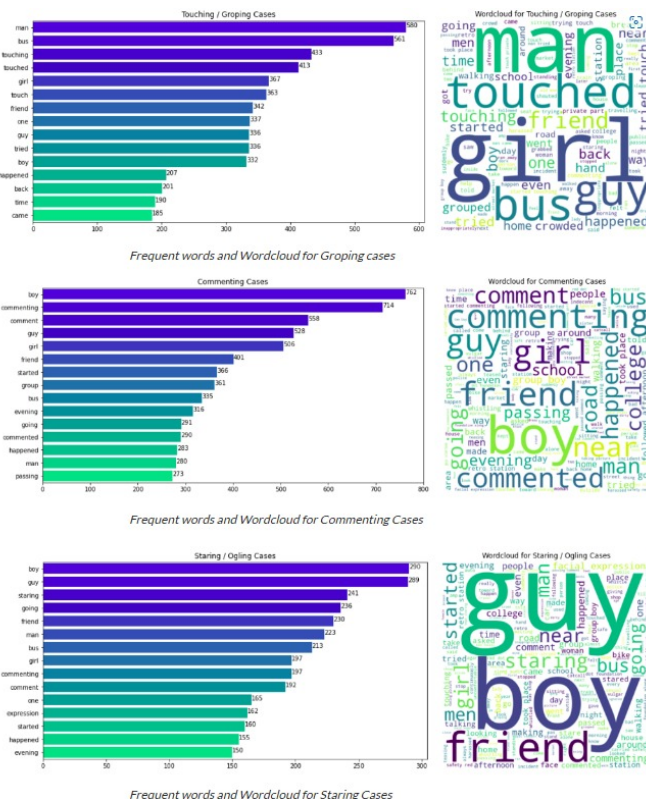

Figure 5

On further analyzing the text data, it was observed that it contains duplicate data which is erroneous, the same description of the text is duplicated and assigned to a different category. As you would be able to see below the description of misbehaved has a category of 0 in row 199 and a category of 1 in row 7175.

VI. Tools and Technologies

There are various technologies that we used in this project. for the frontend part we used HTML,CSS,REACT JS, BOOTSTRAP. for the backend part we used python flask.and we used NLP as our machine learning domain.
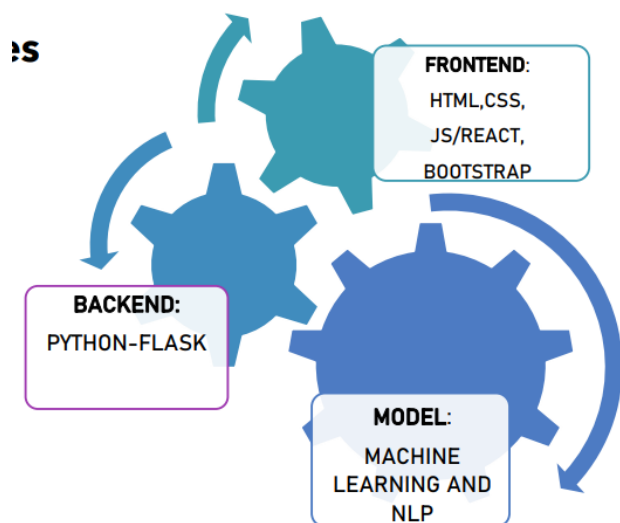
Figure 6

Figure 8

VII. WORKFLOW

In the first phase of our project, we preprocessed textual data using NLP and going to find some patterns in the data. Exploratory data analysis of data and visualization using different graphs and charts.

In the second phase we developed a machine -learning model for classification, we implemented some core models with the estimation of accuracy. For the evaluation of our models, we developed some performance matrices.

In the third phase, we deployed our machine learning model in binary pickle format. We integrated this deployed model into real-time applications using the flask framework.

The stories have to be placed in the space provided and it will classify it as sexual harassment or not and will label it into 3 different categories that are Commenting , Ogling/Facial Expression , Staring and touching/groping.
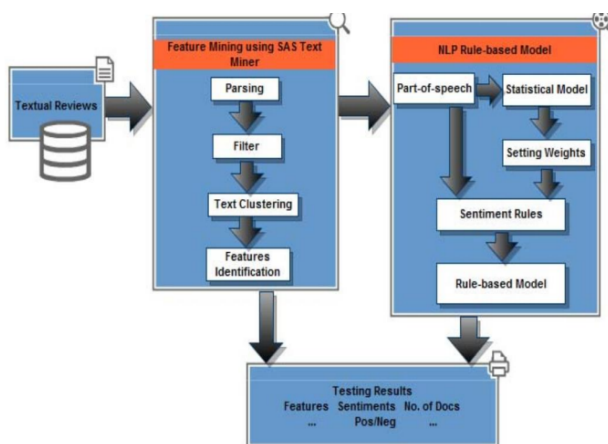

Figure 9

Figure 7

VIII. WEBSITE

The Model is embedded on the website that is made for it. React is used for the frontend part and Flask is used for the backend part. The website is named Shakti.
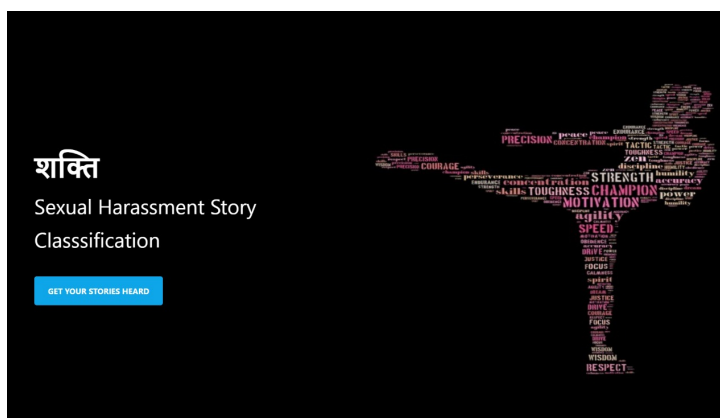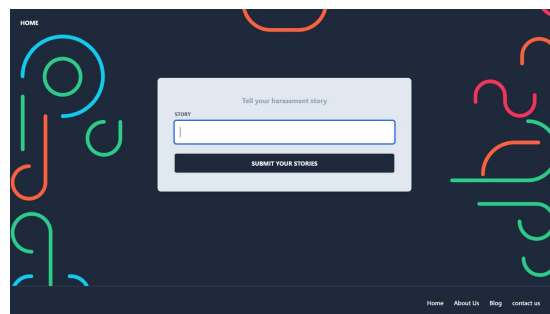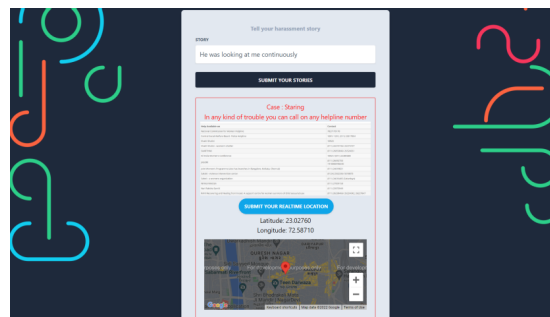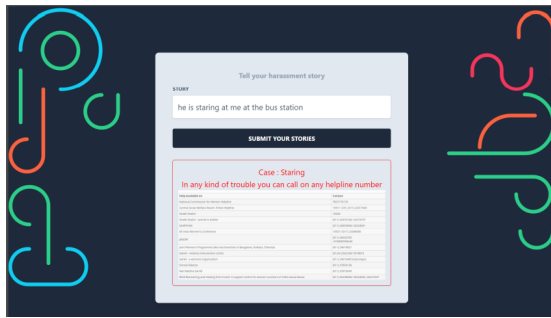

Figure 10

Figure 11

IX. DISCUSSION AND SUMMARY

Shedding light on the perspective of "women development" is as equally important as "women's equality". Hence, this portal aims to empower women by opening up the doors to various opportunities and assistance in varied areas by providing a web portal equipped with different features. Our motto is to develop a woman-driven ecosystem that encourages women to equally contribute to society.

APPENDIX A
OTHER FEATURES WHICH ARE PROVIDED ALONGSIDE THE DETECTION PORTAL

There are various features which can be embedded to the application alongside the Sexual detection portal which are as follows:

- Financial Literacy: providing earning opportunities

- Legal Assistance: Guiding them to take further action in case they are victimized.

- Self Defence : Self Defence guidance

- Network Building: To connect women to join the cause.

- Encourage users using reward system.

- Make Women Independent

APPENDIX B
REPOSITORY LINK

https://github.com/vipashaaV321/Sexual_Harassment_Story_Classification

REFERENCES

[1] Karlekar, Sweta, and Mohit Bansal. "Safecity: Understanding diverse forms of sexual harassment personal stories." arXiv preprint arXiv:1809.04739 (2018).

[2] https://tanmoyblogs.medium.com/more-complicated-model-means-better-accuracy-not-really-7038d4653c1f

[3]Alawneh, Emad, et al. "Sentiment analysis-based sexual harassment detection using machine learning techniques." 2021 international symposium on electronics and smart devices (ISESD). IEEE, 2021.

[4] https://www.analyticsvidhya.com/blog/2022/02/classifying-sexual-harassment-using-machine-learning

[5]https://kijoms.uokerbala.edu.iq/cgi/viewcontent.cgi?article=3157context

[6]Borj, Parisa Rezaee, and Patrick Bours. "Predatory conversation detection." 2019 International Conference on Cyber Security for Emerging Technologies (CSET). IEEE, 2019.