Live Monitoring Sexual Harassment Reports In Social Networks

Dr. Naeemul Hassan, Amrit Poudel

Department of Computer and Information Science, The University of Mississippi

Email: nhassan@olemiss.edu, apoudel1@go.olemiss.edu

dear.lab

INTRODUCTION

Motivation

Examples:

"I was 16 when I was groped from behind while waiting in the queue"

- This tweet reports the sexual harassment.

"Don't ever tell a rape survivor that rape culture doesn't exist"

- This tweet does not report the sexual harassment.

Rise in Me Too Movement:

- People are more open than ever in sharing their sexual harassment experience.

Lack of Research on Tracking Sexual Harassment Experiences:

- No comprehensive database containing sexual harassment experiences
- No automated classification of such experiences

OBJECTIVES

- Which classification model performs better under varying max_features:2000, 5000, 10,000 and 20,000?
- What are the most common words used while expressing the harassment experience?
- Which classification model performs better on 10-class classifier model under varying value of max_features?

OVERVIEW

DATA COLLECTION

Twitter[3]: 520,761 tweets, tweeted by 363,090 unique users

Harassmap.org[1] and Safecity.in[2]: 12,000 sexual reports.

→ 24,000 samples taken for further processing through some heuristics approach

Data Processing and Labeling

Initial Phase:

Manual Labelling of data as 0 and 1 by different individuals

- 0 → Not related to sexual harassment
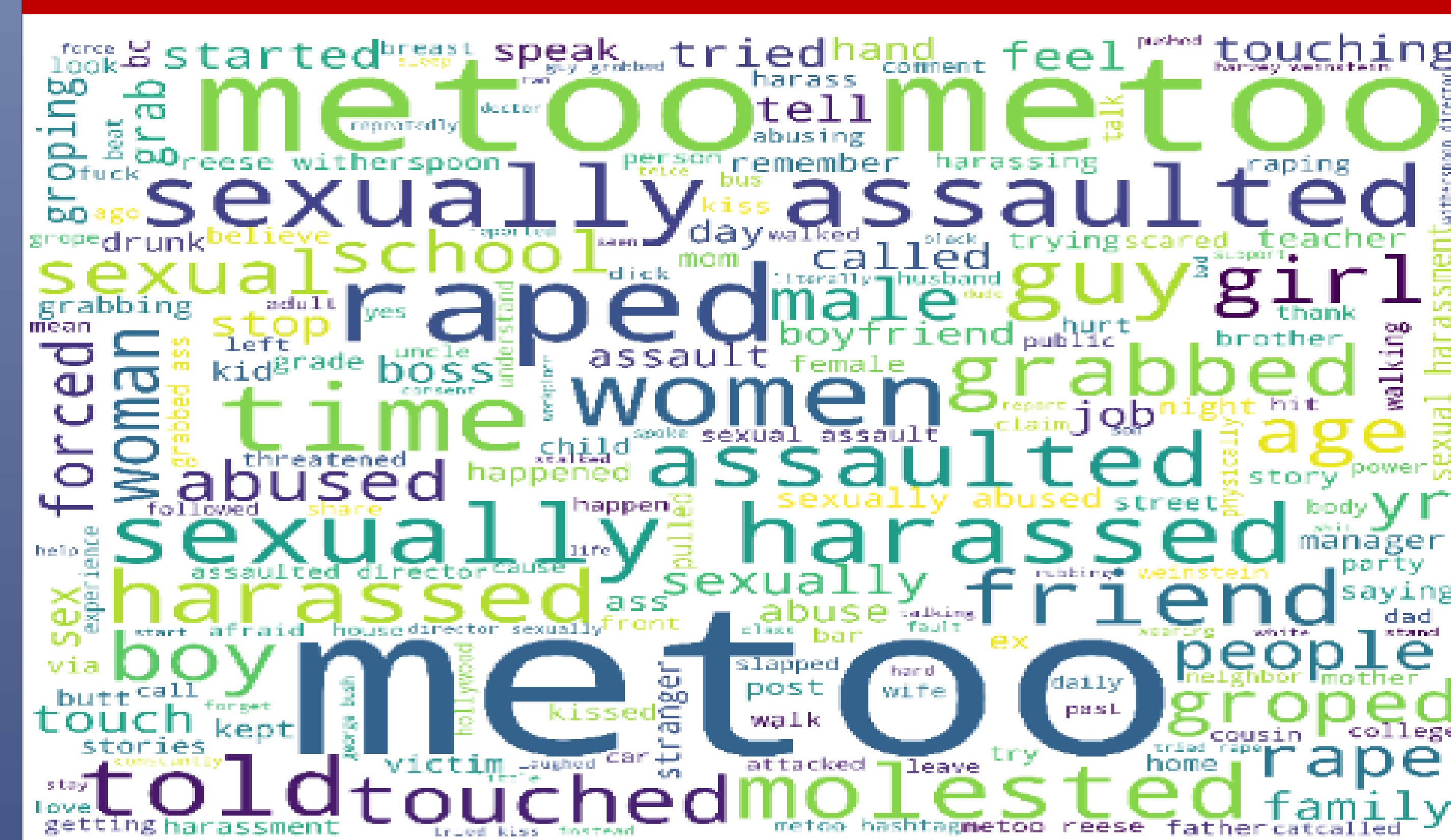- 1 → Related to Sexual Harassment

Amazon Mechanical Turk

Second Phase: Processing of data

- Lemmatization
- Stemming
- Removal of StopWords
- Language Detector (langdetect)
- Count Vectorizer

Final Phase:

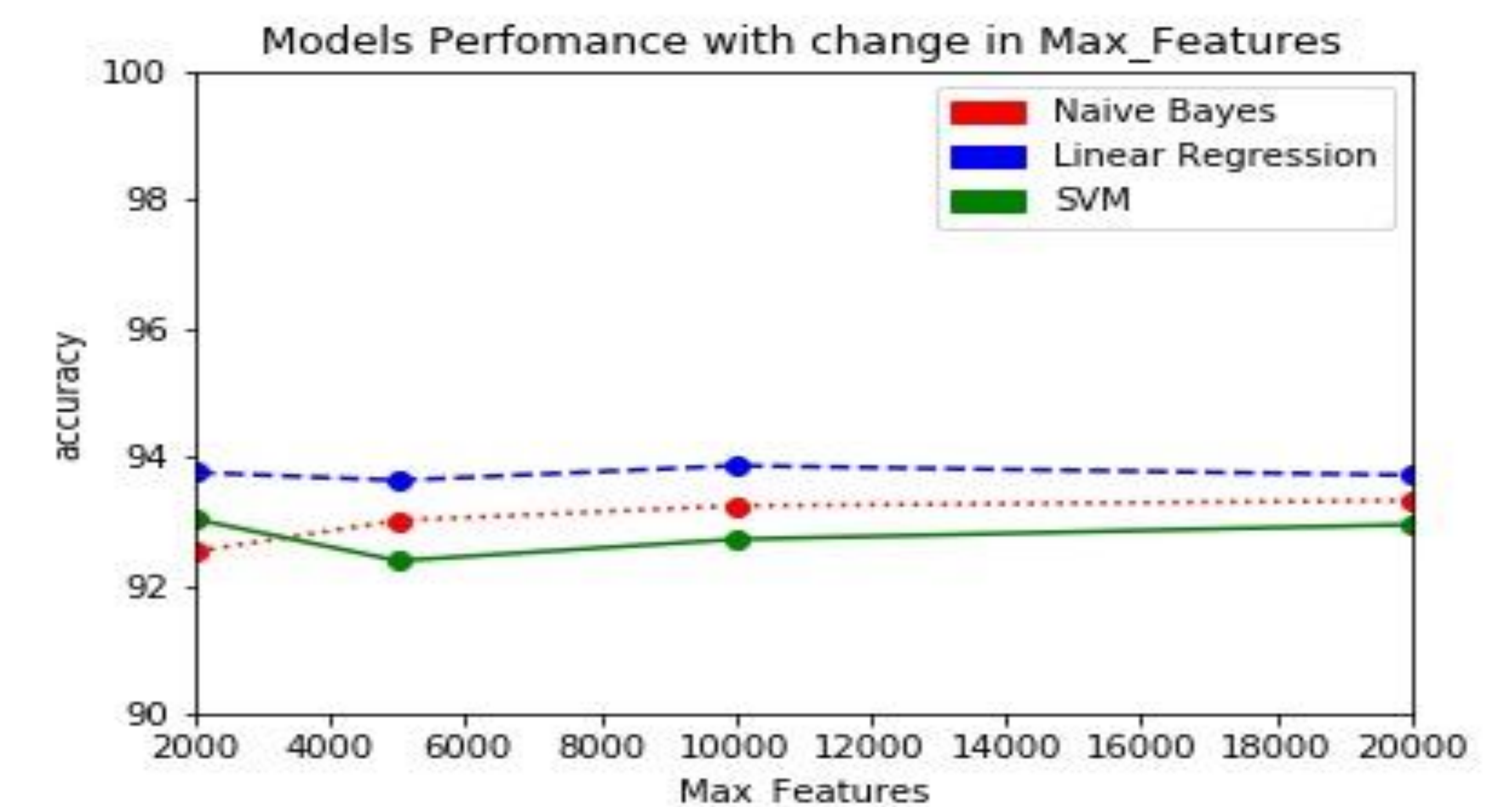Training and Testing of dataset in the classification models.

RESULT AND DISCUSSION

Word Cloud

REFERENCES

1. https://harassmap.org/en/
2. http://www.maps.safecity.in/
3. https://twitter.com/

RESULT AND DISCUSSION

Binary Classifier

10-class Classifier

Classification Report

Categories	Precision	Recall	F1-Score	Support
Catcalls	0.55	0.50	0.52	419
Comments	0.51	0.56	0.54	453
Indecent Exposures	0.12	0.06	0.08	17
Ogling/ Facial Expressions	0.37	0.33	0.35	206
Others	0.52	0.53	0.52	265
Rape/ Sexual Assault	0.56	0.44	0.49	82
Sexual Invites	0.40	0.20	0.27	83
Stalking	0.36	0.27	0.30	120
Taking Pictures	0.76	0.67	0.71	57
Touching	0.58	0.77	0.66	396
Average/ Total	0.51	0.52	0.51	2098

Fig: Classification Report for 10-class classifier using Logistic Regression under the parameter max_features = 20,000. Accuracy of 52.287% was obtained.

FUTURE WORK

- Visualization of Results in Web-based platform in real time on US Map is in progress.
- The research can be further expanded after collection of more relevant data, and modification of the classification models.