Name(s)	Project Number
Sneha Revanur	**J0814**

Project Title

Analyzing Gender-Based Violence and Aggressive Behavior through Social Media Data

Abstract

Objectives/Goals

My project goal was to develop a computational model to classify Tweets for relevance to gender-based violence (GBV), a prevalent human rights issue that transcends a variety of demographics.

Methods/Materials

I defined four classes that pertain to GBV: Physical Violence, Sexual Violence, Harmful Practices, and Other. I used the Python programming language. I extracted data from the Twitter API based on class-specific search criteria and employed the Natural Language Toolkit (NLTK) for natural language processing on Tweet text. Of the 4,000 filtered Tweets, 80% were used as training data and the rest were used as testing data. I used the Naive Bayes classification algorithm to train the machine learning model. I went on to conduct a comparative analysis of two feature sets consisting of unigrams and bigrams. I also constructed a confusion matrix to better analyze the model's performance.

Results

The feature set based on my search criteria had the highest accuracy, with over 85%. For the NLP-based features, Harmful Practices had the highest precision and Other had the lowest. For the search criteria-based features, Harmful Practices had the highest precision and Physical Violence had the lowest. The countries that most frequently discussed GBV included the US, UK, Canada, India, and Australia.

Conclusions/Discussion

I was able to meet my project goal, and successfully leveraged computational linguistics, machine learning, and computational social science to develop a highly accurate Tweet classification model for GBV.

Summary Statement

I developed a computational model that can independently classify Tweets into one of four classes germane to gender-based violence (GBV) while harnessing natural language processing and machine learning.

Help Received

I developed this model myself with the support of my parents and teacher sponsor.