

**TOXIC TWEET CLASSIFICATION WITH NATURAL LANGUAGE PROCESSING  
AND MACHINE LEARNING TECHNIQUES**

**MEDIATORS FOR GAME STREAMING**

An Undergraduate Thesis Portfolio  
Presented to the Faculty of the  
School of Engineering and Applied Science  
In Partial Fulfillment of the Requirements for the Degree  
Bachelor of Science in Computer Science

By

Tanapol Kosolwattana

April 28, 2020

## SOCIOTECHNICAL SYNTHESIS

A toxicity problem is a common issue in the online world especially in the gaming communities where users abuse their anonymity to attack each other through negative words. This problem prompts the technical project and STS project to mitigate the toxicity with different approaches. In the STS project, the problem and solution are explained through the sociotechnical framework. In the technical project, the alternative approach is used in the feature extraction process to optimize the accuracy of the classifying machine.

In the technical project, the main objective is to allow the classification machine to classify sentences into correct categories with high accuracy. There are four steps in this project including, data gathering, text cleaning and word preprocessing, feature extraction, and classifier training. The data acquired from the first step are from the sample tweets that are labeled by three categories (hate-speech, offensive, and non-toxic). Then, the data are cleaned by removing stop words, symbols, and numbers. In the feature extracting step, the technique that is implemented is called Global Vector for Word Representation or GloVe. It is an unsupervised learning technique that is used to extract the features of words by considering the local context information of words. Then, different classifiers are applied to train the dataset and provide results in scores including, accuracy score, precision and recall score, and f-score.

Based on the baseline result, after applying three classifiers; Linear SVM, Logistic Regression, and Random Forest. Linear SVM provides the best results in terms of accuracy and precision score. For the optimization process, there are some parameter tuning processes and more word preprocessing steps that are applied to get better results. However, Linear SVM still provides the best results among the three classifiers. In a further development, the classifiers can

give more accurate results if they know the insight meaning the words because currently, they just know the meaning based on the contexts which do not indicate the tone of the sentence. Also, the dataset should be larger so that the machine can generalize the process of doing the sentiment analysis for larger word groups.

The STS project focuses on how streamers get appropriate support from platform providers when they have specific issues which is hate speech in this project. The proposed solution is to have the mediators who provide support such as giving the information to streamers to access help in proper channels that platform providers offer, suggesting the technical tools to filter out toxic and hate speech comments during streaming periods, etc. Therefore, the research question would be how mediators can play an important role in game streaming. The problem is discussed and analyzed through the Technology and Social Relationships framework. The sources that are employed to answer the question are from the events that streamers face with hate speech issues, the streamers' opinions on the technical tools from platform providers, and the example model that employs a mediator to support the organization.

The example case of effective mediation is from the Swedish Urban Network Association or SUNA, an organization that provides information about network infrastructure. The mediator of SUNA acts as a supporter for members who need to access help, news, and report from the organization and provide an advisory session for trading knowledge within group members. This example relates to a mediator for game streaming communities in the way that a mediator encourages support from the streaming platforms and provides a guideline for streamers who want to have a reference for getting help. Therefore, the mediator is illustrated in the center of

the Social Construction of Technology framework where the mediator bridges all communication and support from a streaming platform to streamers.

In conclusion, the technical project and STS project are coupled in the way that the STS perspective focuses on the broad picture solution to mitigate toxicity within the socio-technical system while the technical research focuses on optimizing the tool that is used to classified toxic words or phrases. Even though there is no official mediator in the communities, if the community members stop typing toxic words and phrases to attack each other, the community will be healthier and safer.

## **TABLE OF CONTENTS**

### **SOCIOTECHNICAL SYNTHESIS**

#### **TOXIC TWEET CLASSIFICATION WITH NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING TECHNIQUES**

Technical advisor: N. Rich Nguyen, Department of Computer Science

#### **MEDIATORS FOR GAME STREAMING**

STS advisor: Catherine D. Baritaud, Department of Engineering, and Society

#### **PROSPECTUS**

Technical advisor: N. Rich Nguyen, Department of Computer Science;

STS advisor: Kent A. Wayland, Department of Engineering and Society