**Methodology**

**Data Collection**

Social networking sites such as Youtube, Facebook and Twitter were used as sources of data for the corpus. The dataset from Youtube contains comments from videos focusing on controversial events in the Philippines such as issues about Manny Pacquiao in LGBT, the pabebe girls, joga videos, aldub bashers on their videos, the issue about Denice Cornejo and Bong Navarro, Mitch Liggayu controversial videos after Jam’s death, the issue about Donald Trump and Rodrigo Duterte after winning the position for the new becoming President of America, the issue about Laila de Lima facing after knowing that she’s the master mind on the drug smuggling happening on the Philippines and the issue about Marcos defining as not a hero, these topics are often a rich source for objectionable and rude comments. Most comments on YouTube are described as stand-alone, with users expressing opinions about the subject and content of the video. There were no clear patterns of dialogue in the corpus and it has no conversational features because some of the comments were constructed as responses to previously posted ones. . In Facebook, the team collected posts from the different universities' confession pages because these pages allow anyone to share personal secrets, rumors, gossips, and anything else they might want others to know about but are hesitant to post publicly or in a way that is tied to their identity. Thus, the anonymity of the person posting a confession makes these pages vulnerable to cyberbullying activities. In Twitter, posts from random Filipino people were collected. Twitter is also prone to cyberbullying occurrences since users can easily create fake accounts to launch their bullying cyber-attacks against people they don’t like or disagree with. In 2011, a study conducted by the University of Wisconsin-Madison found that 15,000 abusive tweets per hour, which equals 100,000 abusive tweets a week.

To obtain this data, the team crawled a subset of Youtube, Facebook, and Twitter using Import.io and extracted statements from these sites. The only field that was used in collecting their data was the textual content of the post while disregarding the other features such as the user information, links, and others. A total number of 2000 statements from these social networking sites were collected.

**Data Annotation**

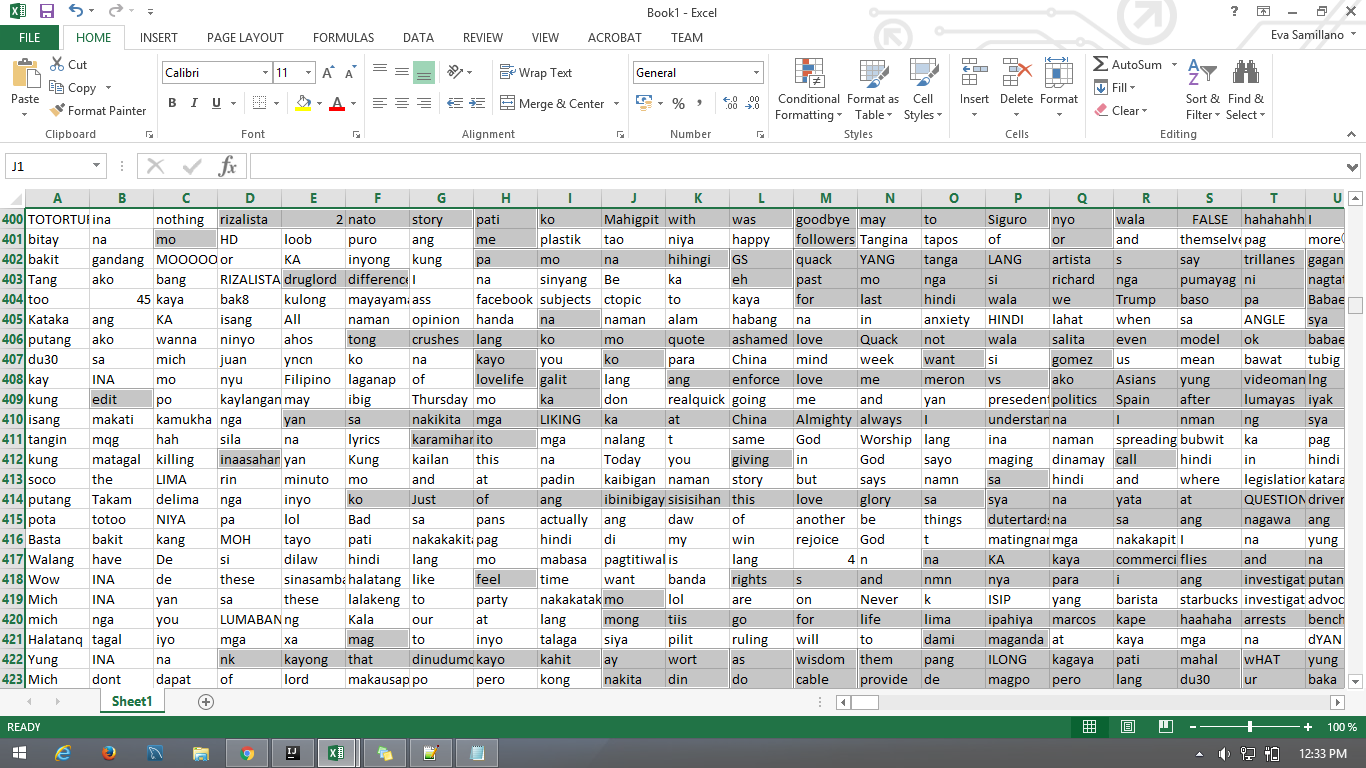
The initial part of the experimentation phase begins with the establishment of classes. After the team has successfully extracted information for their corpus, they formed a dataset of three classes: cyberbullying, non-cyberbullying and ambiguous cyberbullying. The previous data annotation process was accomplished by a single person therefore the researchers had decided to redo the whole process once again by conducting a questionnaire for 100 respondents. Participants will be given a sheet of paper containing 10 sentences taken from the corpus. They will be given the opportunity to annotate these sentences accordingly (whether they think the statement implies cyberbullying, does not imply cyberbullying, or they cannot identify at all). The project team members were also required to annotate the remaining 1000 dataset as well. This is to verify the accuracy of the model in terms of functioning side by side with features that are closely related to each other. This process will also determine which words can fall in more than one class so they can be given appropriate weights (feature weighting) – e.g. a particular word is more appropriate to be classified in this class compared to the other despite having the characteristics if being included in the latter class as well.

**Pre-Processing**

After obtaining data for the corpus, the dataset must undergo pre-processing primarily because it contains unstructured text. The purpose of this step is to transform messages into a uniform format that can be understood by the learning algorithm. In pre-processing, stripping of whitespaces and removal of unnecessary characters such as emoticons are performed through the use of Notepad++ and Excel.

**Feature Extraction**

One of the crucial steps in creating a cyberbullying detection model is feature extraction. By using machine learning algorithms for training the classifier, representation of text as a feature vector is required. For this process, the team used the Bag of Words(BoW) in a unigram technique model, which is one of the most commonly used representation in Natural Language Processing. The primary stage of this model is the creation of vocabulary of words which is in this approach indicates the collection of both abusive and non-abusive words. In BoW model, each word is associated with a count of occurrences. This vocabulary is defined as a set of non-redundant words wherein the order doesn’t matter. Each statement is represented as a feature vector composed of binary attributes for each word that occurs in that message Let {w1,…,wm} be a predefined set of m features (vocabulary of words) that can appear in a message. Let ni(d) be the number of times wi occurs in a message d. Then each messaged is represented by the message vector d:=(n1(d), n2(d),…,nm(d)). If a word present in the vocabulary appears in a given text message, its corresponding attribute is set i.e. 1, else it is set to 0.



*Bag of Words: Unigrams’*

**Folding**

In the dataset, the team used 75% of the dataset for training and 25%% for testing purpose. From the 2000 statements that were extracted from social networking sites, 1500 statements were used for training the classifier and 500 statements were used for testing the statements against the classifier.

**Analysis and Classification**

After the creation of Bag-of-Words (BoW) model, the statements were retrieved from Excel. The csv file was first converted into .arff (Attribute-Relation File Format) file format since it is the only format that can be used in WEKA. The team used a supervised learning approach to train the classifier in creating a cyberbullying detection model. They employed a Support Vector Machine (SVM) model in WEKA as a classification tool.