IV. Design and Methodology

In this section, the researchers illustrate their understanding on the process of creating a cyberbullying detection mechanism.

**The Creation of the Corpus**

**Data Collection and Annotation**

The team acquired data from the most used social media networking site which is Youtube, Facebook and twitter. A total of 2000 comments (both cyberbullying and non cyberbullying) were obtained and evaluated according to their corresponding level of how they define cyberbullying (0 - for comments containing no cyberbullying occurrences, 1 - for comments expressing cyberbullying content, 2 - for comments ambiguous regarding cyberbullying events).

Moreover, the team decided to use Import.io as their primary tool for the annotation process. It is a powerful and easy-to-use tool for data extraction that has the aim of getting data from any website in a structured way. The YouTube dataset contains comments from videos on controversial events in the Philippines that were considered controversial as these topics would most likely encourage viewer to post their opinions about it. Most of the cyberbullying statements that the researchers obtained came from YouTube. In Facebook, they collected posts from different universities secret pages, wherein a user can post his/her story about controversial school experiences. In Twitter, posts from random Filipino people were collected.

The dataset was cleaned in order to remove unnecessary symbols present in text using notepad ++. These symbols were removed as so they will not be extracted as features later on.

Annotating data before was only done by a one person before but then the researchers plan to make a manual annotation, voluntary participants will be given a piece of paper containing 10 sentence from the corpus that the researchers gathered then participants will identify if the sentence existing some marks/hints of cyberbullying.

**Pre-Processing**

.Bag-of-words and Ngram which is the most popular and simple representation in Natural Language Processing and Information Retrieval is the feature extraction that the researcher used for thier text document.

After cleaning the data, using excel and notepad ++ the researchers start to splice the 50 random sentences into unigram that they get in the corpus for the first experiment,then from the splice word they get the unique words to become their bag-of-words or bag-of-unigram. On of the feature of Bag-of-words is it can calculate various measures to characterize text, term frequency is one of its feature that calculate the number of times that the term appears in a text.

For the next process, Bag-of-words that we collect will become attribute, the number of times that the term appeared in the text will be the data later on, and lastly the annotated part before will become the classes. This input will be in a format of .arff (Attribute-Relation File Format)so that Weka could read the file.

Weka is an open source that define as the collection of machine learning algorithms for data mining tasks, using this researchers could visualize the document text that researchers process before also using Weka researchers could view the currently performance of their system and the occurring error.

In the succeeding phases of the project, the researchers are planning to continue with the text pre-processing tasks that are needed in order to lessen the difficulty of extracting linguistic features from the text (which is the phase of the project that will come after the pre-processing phase). Feature selection will be implemented using the following methods: Bag of Words – which will yield bigrams and trigrams, and application of TF-IDF scores – which will aid in the process of determining the most relevant words in each cyberbullying category, as these scores will be used to evaluate the bigrams and trigrams of the instances resulting from the Bag of Words approach (the higher the score, the more relevant the term is). Profane words and words written in uppercase letters are going to be considered offensive by default.

The initial part of the experimentation phase begins with the establishment of classes involving the predefined classifications of cyberbullying expressions based on sensitive issues. 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. The second part of the experimentation phase will take into consideration the final features that will be selected (or the most determining words for each cyberbullying classifier based on offensive issues). The group may either remove or add more features depending on the results the experiments will yield regarding the cyberbullying detection model’s accuracy

*Table III. Proposed plan of activities for the accomplishment of the project*

V. Results and Discussions

This section presents the outcome of the proponent's preliminary experiments.

The first result show that the currently performance of the system is 20%, but for this experiment researchers not yet complete the number of data that should be in the file(10/50)data.

The next result shows that having more data in the system could increase the capability to meet the accuracy that the researchers planning to had.The data was completed (50/50).