1. **Introduction**

**1.1 Background of the Problem**

0

Recent technological advancements have extended the way people communicate. With the rise of Web 2.0, people can easily connect with one another simultaneously through chat rooms, email, instant messaging, forums, and social networking sites.[2] However, alongside with the modern advancements in communication, an old pervasive issue arises with a new form in a new environment known as cyberbullying. [3] Cyberbullying is defined as an aggressive, intentional act carried out by an individual or a group through electronic means of communication, repeatedly and overtime against someone who cannot easily defend himself and can take multiple forms (e.g., threats, discrimination, and insult) in different contexts (Van Royen, Poels, Daelemans, & Vandebosch, 2015). Furthermore, it is a form of harassment that occurs via the Internet which includes vicious forum posts, name calling in chat rooms, creating fake profiles on social networking sites, and sending cruel messages. [4] Unlike the traditional form of bullying, the perpetrators of cyberbullying may use different types of communication technologies such as social networking sites to inflict harm on someone repetitively and deliberately. [1] Since information spread fast in the cyberspace and the number of audience is limitless, it can have deeper and longer-lasting effects than physical bullying (Campbell, 2005). According to Smith et al (2008), victims of cyberbullying may experience severe depression, low self-esteem, or even suicidal attempts.

Philippines was recognized as the social media capital of the world, with more and more Filipinos getting inclined to different social networking sites. [5] A study conducted by We Are Social in 2017 found that Filipinos spent an average of 4 hours and 17 minutes per day on social media sites. [6] However, as the number of Filipino social media users continuously increases, it consequently intensifies the problem of cyberbullying in the Philippines. [7] A survey administered by Stairway Foundation Inc. revealed that 80% of Filipinos have been cyberbullied through social media. [8] Popular cyberbullying incidents in the Philippines are Paula Jaime Salvosa’s “Amalayer” incident [9], Raymond Malinay’s prank involvement [10] and DJ Karen Bordador’s cyberbullying experience, following her arrest with her boyfriend in a drug-related buy bust operation. [11]. However, these are only few of the cyberbullying instances that has been formally reported.

The growing cases of cyberbullying led to the introduction of Anti Bullying Act of 2013, which requires all elementary and secondary school to adopt policies that will prevent and address cyberbullying in educational institutions. [13] In 2015, House Bill 5718 was proposed to provide consequences for cyberbullying act wherein perpetrators shall face a penalty of six months to six years of imprisonment. [14] Social media administrators also play a crucial role in the process of combating cyberbullying by ensuring a safe environment, deleting harmful contents, and identifying perpetrators of online bullying. Furthermore, they have adopted various strategies to protect their users by preventing and intervening in cyberbullying situations. Their current practice involves having a moderator that will monitor inappropriate content which will allow them to detect cyberbullying in an early stage and to take actions thereafter. One of the most common methods used by these sites is privacy settings which allows users to limit the amount of information that can be viewed publicly. A reporting tool page was also used wherein users can report instances of online bullying directly to the administrators. Safety Mode, an opt-in setting, was introduced by YouTube to filter search results. Facebook has moderation and profanity blocklist that can be used to filter a set of harmful words on a page. Twitter offers Mute Feature that allows a user to remove a person’s tweets from his timeline without them knowing. Despite the efforts made by the authority and administrators of social networking sites, these methods were deemed to be inefficient because it is impossible to monitor all activities in the cyberspace given the vast amount of information available online. In addition to this, their methods rely heavily on the users to submit a report before taking an action. Since Philippines remains to be on a conservative level, Filipinos are often reluctant to admit that they have been cyberbullied and report a cyberbullying instance. [8] Thus, there is a need for technology to intervene in the process of mitigating online bullying.

To facilitate the process of monitoring online information and to track cyberbullying instances automatically and accurately, several studies were conducted towards the development of an automatic cyberbullying detection model (Dadvar et al., 2012; Dinakar et.al, 2011). Moreover, several machine learning approach to text categorization were applied to automate this process. Two of the most popular methods were Naive Bayes (Sintaha, M. Satter, S. Zawad, N. Swamaker, C. & Hassan, A, 2016; Marathe, S. & Shirsat, K, 2015) and Support Vector Machines (Van Hee et al, 2015). Their methods significantly reduced the task of the moderator in monitoring the activities in social media.

The current research aims to improve the previous cyberbullying detection models that were proposed by the researchers in the past in order to generate a flexible and a more accurate model that is primarily suited for Filipinos and to aid in addressing the problem of cyberbullying in the Philippines.

**Statement of the Problem**

How we can easily mitigate cyberbullying occurrences among public Filipino social media posts

**Objectives**

**Main Objective**

To create a cyberbullying detection model

**Specific Objectives**

* To gather textual data for the corpus
* To perform text preprocessing
* To perform text annotation on the dataset
* To implement machine learning algorithm using weka
* To generate a cyberbullying detection model
* To develop a cyberbullying detection system
* To test the system’s accuracy