

# A Predictive Model to Detect Online Cyberbullying

Abhijeet Kasture

A thesis submitted to Auckland University of  
Technology for the fulfilment of the requirements  
for the degree of  
Master of Computer and Information Science  
(MCIS)



2015

School of Computer and Mathematical Sciences

## Abstract

Cyberbullying is prevalent in most countries across the globe. The aim of this research was to develop a predictive model to identify the occurrence of cyberbullying tweets on Twitter. The paradigm shift in the Internet of Things was observed a decade ago, which resulted in enormous growth in the number of active Internet users. Today, this number has exceeded three billion. Social networking websites are classic examples of Internet applications that have large numbers of active users. Twitter, for instance, is one of the most famous social networking portals, with more than 300 million active users at any given time. However, unfortunately it is also a stage for users who are involved in unethical use of the Internet, such as cyberbullying. With such a staggering number of active users on the Internet, cyberbullying has become a widespread global phenomenon. It has extremely adverse effects on its victims. In some cases victims have committed suicide in response to the shame and hatred that is associated with cyberbullying<sup>1</sup>.

In this research, 1313 unique tweets were collected from Twitter. With the help of psychological studies referring to, the behavior of individuals and the use of dialects pertaining to verbal aggressiveness, 376 tweets were manually tagged as cyberbullying tweets in the first phase. In the next phase, every word in a tweet was individually categorised based on the pragmatics of language. In order to achieve this, tweets were categorised using Linguistic Inquiry and Word Count (LIWC), a psychometric evaluation tool that categorises text based on Linguistic Processes, Psychological Processes, Personal Concerns and Spoken Categories. Collectively, they add up to 67 sub-word-categories. In the next step of the psychometric evaluation, LIWC calculated the degree to which different word-categories were used by people in cyberbullying. Psychometric evaluation therefore aided in effective text categorisation and quantifying the degree of word usage, which was observed to be a gap in previous studies. As a result, tweets were converted to a multi-dimensional attribute relational numeric dataset. This dataset was very rich in terms of the information that it carried.

This dataset was then used to train machine learning classifiers in Weka to develop a predictive model to detect cyberbullying. The data was randomly segmented 66% for training the predictive model and 34% for testing it. It was seen that the Random Forest classifier built the predictive model with a precision value of 0.97, indicating that binary classifiers outperformed the multiclass classifiers in detecting cyberbullying tweets.

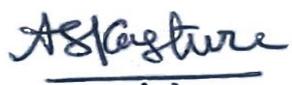
---

<sup>1</sup> <http://nobullying.com/amanda-todd-story/>

## Attestation of Authorship

I hereby declare that this thesis is my own work. To the best of my knowledge and belief, it contains no materials previously written or published by any other person, except for the authors defined in the acknowledgements. I also declare that this work has not been submitted to any other institution or university for the award of any other degree or diploma in any university or other institution of higher learning.

The thesis work was conducted from March 2014 to August 2015 under the supervision of Dr. Parma Nand at Auckland University of Technology.

A handwritten signature in blue ink that reads "Abhijeet Kasture". The signature is written in a cursive style and is underlined with a single horizontal line.

Abhijeet Kasture  
Auckland, New Zealand  
August 2015

## **Acknowledgements**

My acknowledgements extend to all the individuals involved directly and indirectly for their help in accomplishment of my thesis.

Most importantly, I thank Dr. Parma Nand. As my primary supervisor, his guidance and direction in various brain storming sessions helped me overcome complications throughout the entire course of my thesis. I thank him for his counsel and resolute help. I want to express gratitude toward our senior lecturer, Shoba Tegginmath, for being my secondary supervisor and controlling me with her remarks enhancing my written work.

I thank Ksenya Nefiodova, our Program Administrator for the School of Computing and Mathematical Sciences at AUT, for her kind assistance throughout the course of my thesis.

I would like to express deepest gratitude to my mother, Rekha Kasture, for unwavering inspiration and encouragement throughout the entire journey of my thesis.

Lastly, I would like to thank my dear friends Vishal and Preeti for their constant motivation and trust that uplifted me in difficult times.

# Table of Contents

<i>Abstract</i> .....	<i>ii</i>
<i>Attestation of Authorship</i> .....	<i>iii</i>
<i>Acknowledgements</i> .....	<i>iv</i>
<i>List of Figures</i> .....	<i>viii</i>
<i>List of Tables</i> .....	<i>ix</i>
<b>1. Introduction</b> .....	<b>1</b>
<b>1.1 Outline of thesis</b> .....	<b>5</b>
<b>2. Literature Review</b> .....	<b>6</b>
<b>2.1 Cyberbullying</b> .....	<b>6</b>
<b>2.1.1 Major components of ever-developing cyberbullying</b> .....	<b>7</b>
<b>2.1.2 Statistics of bullying on Twitter</b> .....	<b>9</b>
<b>2.1.3 Cyberbullying statistics</b> .....	<b>9</b>
<b>2.2 Twitter</b> .....	<b>10</b>
<b>2.2.1 Cyberbullying on Twitter</b> .....	<b>11</b>
<b>2.3 Related Works on Text Classification Techniques for Cyberbullying</b> .....	<b>11</b>
<b>2.4 Psychometric Analysis</b> .....	<b>13</b>
<b>2.4.1 Related work on text classification using psychometric analysis</b> .....	<b>15</b>
<b>2.5 Literature Review of the Tools Used</b> .....	<b>16</b>
<b>2.5.1 Psychometric evaluation using LIWC</b> .....	<b>16</b>
<b>2.5.2 TAGS archiving tool</b> .....	<b>19</b>

2.5.3 Weka toolkit .....	21
2.5.4 Overview of classification algorithms used in Weka .....	21
<b>3 Methodology .....</b>	<b>25</b>
3.1 System Architecture.....	25
3.1.1 System flow .....	27
3.2 Data Pre-processor.....	30
3.2.1 Purpose of data pre-processor .....	30
3.2.2 Data pre-processor functionality .....	31
3.2.3 Operational use of processed data .....	33
3.3 LIWC Psychometric Evaluation on Tweets using LIWC .....	36
3.3.1 Parameter settings for LIWC processing .....	36
3.4 Weka Classification .....	39
<b>4 Results and Discussion .....</b>	<b>42</b>
4.1 LIWC Results .....	42
4.1.1 The preliminary results of LIWC .....	44
4.2 Weka Results .....	46
4.2.1 Attribute selection .....	47
4.2.2 Weka classifier algorithms .....	48
4.3 Weka Experimenter .....	54
<b>5 Conclusion.....</b>	<b>55</b>
<b>6 Scope for Future Research .....</b>	<b>57</b>
<b>7 References.....</b>	<b>59</b>
<b>Appendix 1 .....</b>	<b>64</b>

*Appendix 2* ..... 66

*Appendix 3* ..... 69

# List of Figures

Figure 1 – Text Categorisation in LIWC .....	18
Figure 2 – TAGS Functionality Diagram .....	20
Figure 3 – Multilayer Perceptron, Neural Network Classifier .....	22
Figure 4 – System Architecture .....	25
Figure 5 – System Flow Diagram .....	27
Figure 6 - Schematic Representation of Data Pre-Processor .....	32
Figure 8 – Schematic Representation of LIWC Functionality .....	38
Figure 9 – Weka Classification Overview Diagram .....	39
Figure 10 - J48 Decision Tree.....	53
Figure 11 – Threshold Curve of Random Forest .....	77
Figure 12 – Attribute Selection using CFS Subset Eval Filter.....	78
Figure 13 – Attributes Ranked 1-22 using Info Gain Eval Filter.....	79
Figure 14 – Attributes Ranked 23-44 using Info Gain Eval Filter.....	80
Figure 15 – Attributes Ranked 45-66 using Info Gain Eval Filter.....	81
Figure 16 – Attributes Ranked 66-67 using Info Gain Eval Filter.....	82
Figure 17 – GUI for Multilayer Perceptron (Epochs Calculation) .....	83
Figure 18 – GUI for Multilayer Perceptron (Epochs Calculation) at Epoch 340 .....	84
Figure 19 – GUI for Multilayer Perceptron (Epochs Calculation) at Epoch 500 .....	85

## List of Tables

Table 1– Bullying ... then and now ("Traditional bullying vs. cyberbullying," 2011) .....	8
Table 2– Cyberbullying and Social Networks ("Cyberbullying Statistics," 2014).....	10
Table 3– LIWC Example Output .....	18
Table 4 – Snippet of LIWC Results .....	43
Table 5 – Weka Classifier Outputs .....	46
Table 6 – Cost-matrix Comparison.....	54

## List of Equations

Equation 1– LIWC Word Category Output .....	17
Equation 2 – Pattern Generated by LIWC for Instance $i$ .....	45

# *1. Introduction*

Predictive analytics is a field of study whereby valuable information can be extracted from existing or new datasets. This extracted information could be a pattern that defines a particular event, or it could be a prediction of an event based on similar events in the past or, it could be the identification of a particular trend that was previously unknown (Larose & Larose, 2015). It works by computing the likelihood of an event by linking the relation and performance of its certain attributes prevalent in different samples of data. Different datasets can be linked by developing data attributes in primary datasets specifically for what needs to be identified. For example, if one wanted to identify addresses of places, then the dataset is searched for attributes (or words) indicating different places. Then, addresses of places in another dataset can be found by linking the same data attributes that were identified in the primary dataset. In this way a relational link is developed. The performance of these links is based on the way in which these data attributes are connected depending on how strongly they are correlated. Predictive analytics also creates the provision for quantifying the relationships of these data attributes across different samples. As a result, predictive analytics describes the relationships between all the entities and attributes of the data, which can be useful in decision making (Siegel, 2013).

This research investigates the use of predictive analytics to detect cyberbullying on Twitter. Predictive analytics can be used to study and identify various attributes of conversational data pertaining to cyberbullying. As a matter of fact, the use of language is always personalised and shaped according to the personality and character of an individual. Therefore, different people use different styles of writing to bully someone online. However, the usage of word-categories in which these styles of language are tailored remain the same. In similar manner, different people use different words to indicate a specific thing. Although the usage of words can be different, the categories of words that they belong to remain the same. For example, consider the following two sentences: “*This cake is so yummy*”, and, “*This is a delicious cake*”. These two sentences indicate the same feeling of “liking” the cake, however with different word syntaxes. The words ‘tasty’ and ‘delicious’ fall in the same word-category that indicates perceptual process. Therefore, this study investigates the usage of word-categories to detect cyberbullying as opposed to the direct usage of words. By linking the relation of data

attributes obtained from word-categories indicating cyberbullying on the data used in this research, predictive analytics can aid in the detection of cyberbullying tweets on Twitter.

Harassing someone on the internet is called cyberbullying. It is a widespread social phenomenon, having a negative impact on the lives of people in countries across the globe. Online bullies generally target an individual or group of individuals as their victims. It is mainly found on social networking websites, where victims are most vulnerable to humiliation. Many government bodies and not-for profit-healthcare organisations have highlighted the harmful effects of cyberbullying on the victims based on psychological surveys across different countries. In some cases, the repercussions of cyberbullying on victims have been suicidal tendencies, and in quite a few extreme cases, cyberbullying has also led victims to commit suicide (Bauman, Toomey, & Walker, 2013; Litwiler & Brausch, 2013; Luxton, June, & Fairall, 2012; Schneider, O'Donnell, Stueve, & Coulter, 2012; Van Geel, Vedder, & Tanilon, 2014). In recent times, cyberbullying has taken its toll on one of the most popular social networking websites, Twitter.

Twitter is a micro-blogging, social networking website wherein the users can write short 140 character messages called Tweets. As of May 2015, there were over 302 million active users on Twitter. A user can address tweets directly to other users as well as broadcast messages. Any tweets posted on Twitter have a chain effect. A user posting a tweet can be viewed by his followers and the followers of the receiver of the tweet. In addition, if one of the followers re-tweets the original tweet, then it can be seen by his respective followers. The rising number of re-tweets spread the message on Twitter like wild-fire, which has been termed “going viral” in social media.

Cyberbullying on Twitter is a global phenomenon because of its huge volumes of active users. The trend shows that cyberbullying on Twitter is growing rapidly every day, immeasurably. This hike could be related to the major paradigm shift in the Internet that was observed a decade ago, termed Web 2.0 (Graham & Haarstad, 2014; Shuen, 2008). It unleashed substantial progress in the usability, availability, flexibility and portability of internet applications. As a result, general internet use has seen gigantic development. The number of internet users in 2014 surpassed three billion, which is close to half the population in the world (ITU, 2014). In 2000, this number was roughly 400 million (ITU, 2015). This gigantic rise in users has led to a confounding amount of information flow on the internet, generally referred to as “big data”. The flexibility of Web 2.0 applications has bridged the communication gap between users across continents. In addition, users

have access to mobile internet devices that are used to share information from anywhere due to state-of-the-art wireless technology. Internet today, as a service, is considered to be a boon. However, with the growing number of users it is seen that the internet use is often abused by many users who indulge in malpractices such as cyberbullying, cyberterrorism, E-Commerce fraud, misleading marketing and advertising, privacy breaches, unethical hacking and identity theft, just to name a few.

In the light of the above discussion, this research used predictive analytics to detect cyberbullying on Twitter. Predictive analysis predominantly requires effective text classification of the dataset, which can then be used to develop a model to detect cyberbullying tweets. It is first necessary to define the attributes of the data that could act as predictors to detect cyberbullying from tweets. This has been identified as a major gap in previous studies. In this research, the main focus was to define the attributes of the data that could be used to classify bullying tweets. Three decades of psychological studies have proved that the writing patterns of individuals can give insights into their social behaviour patterns (Beck, 2011; Izard, 2013; Miller & Rollnick, 2012). Therefore, the text classification of data in this research was conducted based on the pragmatics of language. Every individual has his or her own unique style of writing, which means the pragmatics of language being used differs from person to person. However, the words used in written text fall into four main categories: Linguistic Process, Psychological Processes, Personal Concerns and Spoken Categories. These are known as psychometric properties. These four psychometric properties collectively add up to 67 sub-categories of words.

In this research, 1313 unique tweets were collected out of which 376 tweets were manually separated as cyberbullying tweets, using cyberbullying-identified categories of verbal aggressiveness. The next goal was the individual psychometric evaluation of these unique tweets. Psychometric evaluation has two functional aspects. Initially, it aided in the classification of text based on the pragmatics of language. In this case every tweet was classified into 67 word-categories according to psychometric properties. For example, social process, cognitive process or positive and negative emotions are considered as psychometric properties. These psychometric properties are explained further in section 2.5. Individual classification of tweets can help identify the various psychometric behaviour patterns of different people associated with the intention of cyberbullying. Then, based on the set of words used in tweets, psychometric evaluation calculated the degree to which each word-category was used by people in cyberbullying.

As a result of psychometric evaluation on tweets, the tweets were converted to a numeric dataset that carried more information related to cyberbullying. In technical terms, this dataset contained 376 unique patterns of word-category usages that conveyed cyberbullying. These psychometric patterns acted as predictors for cyberbullying in tweets. Every cyberbullying tweet has a different syntactical and pragmatic structure and hence every tweet classified was based on psychometric properties that acted as a unique predictor for detecting cyberbullying. This type of text classification was applied on every conversational tweet using a tool called Linguistic Inquiry and Word Count, abbreviated as LIWC. Further, it was observed that an information rich numeric dataset made machine learning more efficient for detecting cyberbullying. Using Weka, a data and text mining tool, machine learning algorithms such as Random Forest, Sequential Minimal Optimization (SMO), J48 Decision Tree and Multilayer Perceptron were applied to the numeric dataset, obtained after psychometric evaluation of tweets, to build a predictive model to detect cyberbullying tweets.

The numeric dataset was randomly segmented, 66% for training the predictive model and 34% for testing it. The testing dataset comprised of 446 instances out of which 123 instances were cyberbullying tweets. It was seen that the Random Forest classifier built the predictive model to detect cyberbullying with 97% accuracy, with precision and recall values of 0.983 and 0.935 respectively. The overview of the results for the dataset used in this research indicates that binary classifiers such as the Random Forest classifier outperformed the multiclass classifiers in detecting cyberbullying tweets.

## **1.1 Outline of thesis**

Chapter 1 introduced the prevailing issue of cyberbullying and provides an overview of the research conducted to detect cyberbullying from a machine learning perspective. Chapter 2 gives a detailed explanation of the cyberbullying issue and its prevalence on Twitter, followed by reviews of previous works on cyberbullying using text classification techniques. The chapter continues by reviewing text classification using psychometric analysis and related works on Twitter, which is followed by a detailed review of the tools and algorithms used to train the machine learning classifiers. Chapter 3 starts by defining the system architecture followed by system flow design. The chapter then continues to experiment on the three stages developed in the system architecture. Chapter 4 gives a detailed analysis of the results from the LIWC and Weka classifier experiments, followed by a conclusion and discussion of future scope in Chapter 5 and Chapter 6 respectively.

## ***2. Literature Review***

The literature review covers comprehensive information on the prevalent issue of cyberbullying associated with the global influence of Twitter. It continues by reviewing related works on detecting cyberbullying. The following chapter then describes the psychometric evaluation in detail, followed by review of the tools used in this research.

### **2.1 Cyberbullying**

Bullying is an aggressive human-behavioural pattern of intentionally causing another person discomfort or harm (Kowalski, Limber, & Agatston, 2012).

The current form of development in communications technology has a flip side to it, whereby cyberbullying has taken place with increasing frequency every day around the globe. Cyberbullying can be defined as the use of technological advancement via cell phones, e-mails, chat rooms or social networking platforms like Twitter or Facebook to humiliate or threaten others (Kowalski et al., 2012).

According to the research at the University of British Columbia, cyberbullying is a bigger problem than traditional bullying. The statistics, generated with the help of surveys involving 733 youths, show that 25 to 30 percent of the young people participated in cyberbullying but only 12 percent mentioned traditional bullying. Out of those participants, 95 percent found online taunting language as a joke, and the rest meant to humiliate or harm someone (Shapka, 2012).

In most cases of traditional bullying, the bully plans the next attack on the victim whereas it may not be planned and could be impulsive in cyberbullying. Similar to traditional bullying, cyberbullying has the following characteristics:

- A need for control or power
- Aggression
- Proactively targeting the victim.

With the ease of creating or participating in groups online through a valid e-mail address,

it is very easy to generate fake accounts and bully anonymously. As the anonymous comments and messages are not enough to trace the individual who is bullying, cyberbullies are free to do so as they please without repercussions. Recently, it has been observed that cyberbullies target random people across the globe. This is an upward trend amongst the cyberbullies, as they know that they cannot be harmed physically or in any other way (Kowalski et al., 2012).

To increase awareness about cyberbullying, many non-profit organisations try to reach out to people over the internet. For example, the Nemours Foundation is dedicated to improving the health of children and teens. They have developed an informative website<sup>2</sup> concerning total health, behaviour and development from the pre-natal stage to teenage years. Similarly, the governments of most countries try to highlight such sensitive issues and reach out to their people by addressing the harmful facts about cyberbullying based on the latest data. It is observed that cyberbullying has increased over time and has an extremely negative psychological impact on victims.

### **2.1.1 Major components of ever-developing cyberbullying**

- **Novel social repercussions**

Bullies on the internet frequently hide behind false identities and the victims of cyberbullying may not even know the attacker. At the pinnacle of cyberbullying, a victim can have multiple attackers using the same technology resulting in the act of gang-bullying (Slonje, Smith, & Frisé, 2013).

Medical editor Larissa Hirsch states, “*Bullies and mean girls have been around forever, but technology has given them a whole new platform for their actions.*” (“Cyberbullying,” 2014). This highlights the fact that victims are now more vulnerable to being attacked by bullies.

---

<sup>2</sup> <http://kidshealth.org/>

- **Novel psychological repercussions**

Traditional bullying has always resulted in physically, mentally and emotionally affecting the victims, which is observed to be long-lasting (Xu, Zhu, & Bellmore, 2012). However, and contrary to traditional bullying, the results of cyberbullying can be exceedingly long-lasting. Cyberbullying has been shown to cause serious psychological damage including depression, anxiety and emotional disorders (Slonje et al., 2013). By comparing traditional face-to-face bullying to cyberbullying, it is found that they have a similar psychological effect on victims' health. However, the upward trend in cyberbullying shows increased adverse effects in most cases. Cyberbullying follows the victim everywhere, causing high distress and extremely negative outcomes (Schneider et al., 2012). In the most extreme cases, such as that of *Amanda Todd*, the bullying has ended after the victim committed suicide (Dufour, 2012).

- **Novel technological repercussions**

Cyberbullying can be done at any time, from anywhere and mostly it is totally anonymous. To address the gravity of the issue and its consequences, government agencies collate information and try to reach out to people through a website<sup>3</sup> managed by the US Department of Health & Human Services and warn them:

*“Bullying online is very different from face-to-face bullying because messages and images can be: Sent 24 hours a day, 7 days a week, 365 days a year, Shared to a very wide audience, Sent anonymously.”* ("Traditional bullying vs. cyberbullying," 2011).

<b>Then</b>	<b>Now</b>
Face to face	Anonymous
Schoolyard	At school and home
During the school day	All day, every day
Smaller audience	Larger, possibly global audience

Table 1– Bullying ... then and now ("Traditional bullying vs. cyberbullying," 2011)

It is easy to bully online as it does not involve face to face interaction. The tendency to become desensitised to a computer screen triggers bullying as there is no spectacle of a reader's reaction after reading the text or post and hence there is no awareness of whether

---

<sup>3</sup> <http://www.stopbullying.gov/cyberbullying/>

they are going too far in joking (Litwiller & Brausch, 2013).

Sadly, one of the most popular social networking platforms, Twitter, has become a stage for bullying, harassment and abuse as it is easy for people to bully online by launching their cyber-attacks against people they don't like or disagree with.

### **2.1.2 Statistics of bullying on Twitter**

According to the study conducted by the University of Wisconsin-Madison in 2011, there were 15,000 abusive tweets per hour. According to the Bureau of Justice statistics, US Department of Health and Human Services and the Cyberbullying Research Center, 52% of teens have reported being cyberbullied. The study shows that cyberbullies come from all the age groups (Fitzgerald, 2012).

### **2.1.3 Cyberbullying statistics**

An online forum<sup>4</sup> plays a vital role in gathering surveys and also collects statistics of bullying and cyberbullying worldwide. The bullying statistics gathered by this forum in 2014 covered major polls published worldwide and included many participants. The focus was on the major trends and shifts in cyberbullying and its effects worldwide.

#### **Highlights of the statistics from surveys collected from more than 10,000 youths**

- 70% of young people are victims of cyberbullying
- 20% of them are experiencing extreme cyberbullying on a daily basis
- 37% of them are experiencing cyberbullying on a very frequent basis
- New research suggests that young males and females are equally at risk of cyberbullying
- Facebook, Ask.FM and Twitter were found to be the most likely sources of cyberbullying, having the highest traffic of all social networks ("Cyberbullying Statistics," 2014).

---

<sup>4</sup> <http://nobullying.com/>

Online Social Networking Platform	Percentage of youths using this platform	Percentage of youths as victims
Facebook	75%	54%
YouTube	66%	21%
<b><u>Twitter</u></b>	<b><u>43%</u></b>	<b><u>28%</u></b>
Ask.Fm	36%	26%
Tumblr	24%	22%
MySpace	4%	89%

Table 2– Cyberbullying and Social Networks ("Cyberbullying Statistics," 2014)

## 2.2 Twitter

Twitter is an online social networking platform, also referred to as a microblogging site, where users can share information via messages up to 140 characters. These short messages are called '*tweets*'. Twitter allows users to tweet through its website or through its applications developed for various external compatible devices. In most countries, users can also use SMS services to tweet.

Tweets can be read by anyone unless the users restrict access strictly to their followers. When a user subscribes to another user account, the subscription is termed '*following*' and the subscriber is called the '*follower*'. Twitter, as a social networking platform, spins around the term 'followers'. For example, if user A is following user B, user A as a follower gets access to read and retweet user B's tweets.

Out of all the tweets generated on Twitter, roughly 40% of tweets are conversational tweets (Kelly, 2009). Users make use of hashtags to tweet about trending topical information. Similarly, users make use of '@' followed by a username, for example '@*username*' to post a tweet mention or reply to another user.

### **2.2.1 Cyberbullying on Twitter**

One thing that makes it easy for bullies to harass someone online is that they can retain their anonymity by creating fake accounts to bully someone. Due to the functionality of 'hashtags' and '@username', victims are more vulnerable to direct online attacks. In addition, the victims are totally exposed, as their followers can witness the entire cyberbullying episodes.

Twitter provides a system to reduce cyberbullying, but unfortunately it is not effective. Twitter has a 'report abuse' form that users affected by cyberbullying must fill out if they wish some action needs to be taken. Twitter needs a more intelligent system to detect cyberbullying, which is more efficient in detecting cyberbullying tweets on Twitter.

## **2.3 Related Works on Text Classification Techniques for Cyberbullying**

Researchers have been trying to address the issue of cyberbullying using text-mining techniques for over a decade. In their studies, they have implemented text mining to detect vandalism, spam, internet abuse and cyberterrorism (Kontostathis, 2009; Simanjuntak, Ipung, Lim, & Nugroho, 2010; Smets, Goethals, & Verdonk, 2008; Tan, Chen, & Jain, 2010).

Dinakar, Reichart and Lieberman conducted a supervised machine learning approach to develop a model to identify cyberbullying (2011). They started by collecting YouTube comments as corpus, further labelled it manually and implemented various binary and multiclass classifiers. Their study revealed that binary classifiers outdid the multiclass classifiers. In their approach, they applied practical knowledge to develop an application for identifying cyberbullying. The data was analysed in segments, where every segment was an individual comment. It can be observed that the pragmatics of conversational data were left out. However, they concluded by addressing the fact that identification of cyberbullying on social networking platforms could be addressed more accurately if those features were included (Dinakar et al., 2011).

In another study by Yin et al., a model to detect harassment with a supervised learning approach was developed (2009). They collected a corpus by extracting online feeds and trained this data on a support vector machine classifier using data dimensions classified

based on contents, context and sentimental features. Yin et al. (2009) trained the classifier to detect harassment solely based on the content of the feeds, but failed to analyse the characteristics of the user posting these feeds. The baselines of this study comprised the frequency of foul words used, implication of N-grams and TF-IDF weighting. In Natural Language Processing (NLP), N-grams are a contiguous sequence of 'n' number of syntactical characteristics found in a sequence of text and TF-IDF stands for Term Frequency- Inverse Document Frequency. The results established on these baselines showed improvements.

Dadvar and De Jong proposed an approach in 2012 to develop a model to detect cyberbullying that incorporated users' information relating to their characteristics and post harassment behaviour in parallel to the conversational data exchanged. They introduced a cross-system analysis study wherein the users' activity across different online social platforms could be monitored to identify cyberbullying behaviour. Furthermore, their study revolved around the application of vocabulary, gender involvement and second and third person pronouns. They collected a corpus of 2200 manually labelled dataset out of which 34% and 66% feeds were generated by females and males respectively. Dadvar and De Jong used a support vector machine classifier to detect cyberbullying (2012). The baseline comprised profane words, second person pronouns, all other person pronouns and TFIDF weighting. The classifier trained the male and female posts individually and resulted in a fair improvement in accuracy. However, this study fell short in contextual features and the complete pragmatics of conversational data.

Reynolds, Kontostathis and Edwards used supervised machine learning to develop a model for cyberbullying detection (2011). The data for this research was retrieved from a question-answering networking website named 'formspring.me'. In addition, the retrieved data contained information about user profiles. The data was divided into 10 files for the training set and testing set respectively. The data was labelled using Amazon Mechanical Turk for identification of true positives. After labelling the dataset, it was observed that this dataset had a class imbalance where 173 out of 1219 posts were identified as cyberbullying. Reynolds et al., used textual features to develop their model. They created a list of swear and insult words and categorised each word based on a scale of severity. The features for input to the classifiers were determined based on the number of bad words, the density of bad words and overall "badness" of the post. Furthermore, to avoid class imbalance, Reynolds et al. copied the positive training set of cyberbullying

several times (2011). The classifiers used for this study included J48, JRIP, IBK and SMO. However, the results obtained for testing set significantly deviated from that of the training set. By comparing the two features, namely, the number of bad words in a post and density of the bad words in a post, the research concluded that the density of bad words in a post greatly determined the accuracy in detecting cyberbullying.

The aforementioned studies highlighted a major gap in detecting cyberbullying. Although the researchers in their experiments exploited machine learning classifiers to their full potential, the studies indicate that there was a struggle to construct a comprehensive set of instruments to measure all the possible attributes of the data to predict cyberbullying. Therefore, this study introduces the worth of psychometric evaluation, which when implemented on the dataset results in

- Effective text classification of the dataset based on the psychometric properties of the pragmatics of language (67 word-categories) and,
- Measurement of the degree to which individuals use different word-categories in cyberbullying.

## **2.4 Psychometric Analysis**

Psychometrics, generally, is considered to be a field of study that quantifies characteristic differences of humans. It involves two major tasks: construction of instruments to measure psychological variables, and estimations derived after analysing the data obtained from these measurements. In brief, the construction of a behavioural or psychological scale, and analysing the resulting data from this scale, is considered to be the field of psychometrics. Hence a measurement study using statistical methodology, and deriving estimations by analysing this statistical data, is defined as the study of psychometrics (Browne, 2000).

For any scientific study to advance, that requires quantification process, its methodology must be based on a solid construct of instruments for measurement. These instruments are generally used to measure the variables relevant to the study and further help to make important estimations and compare the overall significance of the study being conducted. Usually, construction of these instruments lacks precise definition and hence results in inaccurate estimations. Additionally, the resulting estimations are open to significant

errors. Thus, the measurement mechanism used for the purpose of quantification involves repeated attempts in various different ways. When developing relations between the wide ranges of variables, each variable is measured repeatedly, amassing a huge number of calculations. Therefore, analysing psychological measurements based on a statistical approach is generally multivariate (Browne, 2000).

It is generally confusing for the naked eye to make estimations due to the ambiguity in the construction of these instruments of measurement. Researchers are aware of the existence of a hidden or latent variable that leads to inaccurate measurements. This is inferred, as researchers cannot make accurate estimations by creating relationships from the observed variables (Browne, 2000).

Psychometric evaluation has useful applications. For instance, it can be used to determine an individual's personality (Chamorro-Premuzic & Furnham, 2014). This can be used to evaluate one's strengths and weaknesses that generally result in exact bearing of their cognitive abilities and general social behavioural style (Kline, 2013). Due to this, many companies while hiring perform psychometric evaluation on candidates to identify potential match for specific job role. It also aids in identifying aptitudes of individuals wherein, they can specify certain career domains specific to individual test. Hence, psychometric evaluation can identify various psychological implications behind every individual's style of writing (Chamorro-Premuzic & Furnham, 2014).

Further, psychometric evaluation can be implied on texts, collectively generated by various people on a similar topic. Twitter provides its users the ability to use 'hashtags' that generally redirects to a specific trending topic. Different people may have different opinions pertaining to that specific trend. NLP is used to differentiate between those opinions of different individuals (Cambria, Schuller, Xia, & Havasi, 2013). Different NLP modules can differentiate and classify different opinions, but effective implementation of psychometric evaluation can help identify the degree with which these individual opinions differ (Zhang, 2014). This is based on differentiating usage of words by various people wherein, every word falls in a word-category pertaining to a specific psychometric property. The four primary psychometric properties comprise of Linguistic process, Psychological process, Personal concerns and Spoken categories (Tausczik & Pennebaker, 2010).

Hence, psychometric evaluation can distinguish between psychometric attributes of different people that triggers a specific opinion in them about a specific thing. In addition,

it can be used to measure the degree with which these psychometric attributes differ for different people. Going back to the same example from introduction; “*This cake is so yummy*”, and “*This is a delicious cake*”. These statements possess different syntactical structure but both indicate a certain ‘liking’ towards a cake. Psychometric evaluation tells us that these similar opinions trigger perceptual process which is also ‘positive’.

Similarly, in cyberbullying, different people use different styles of writing. In short, the word usage differs but they eventually indicate cyberbullying. In this research, the main focus is to identify specific psychometric properties that are used to convey cyberbullying. Understanding bullying behavior or mentality is a broad scope of study due to inclusion of multiple factors related to pragmatics of language. These psychometric properties that trigger cyberbullying differ between individuals in terms of degree of word usage. Therefore, psychometric evaluation aided in development of a scale that generated a pattern that can be used by machine learning to develop a model to detect cyberbullying. Due to this approach, the probability of occurrences of a cyberbullying tweets on Twitter was detected based on the psychometric properties that it carried. This is contrary to the traditional methods of detecting cyberbullying based on raw text attributes such as just foul or swear words.

#### **2.4.1 Related work on text classification using psychometric analysis**

Bollen, Mao and Zeng conducted research to investigate the degree to which the collective moods of users on Twitter can predict changes in the stock market (2011). The idea behind this research was to validate the notion that individual behavioural actions and decision-making abilities are emotionally driven. They collected around 10 million tweets from three million users over the time span of 10 months as their dataset. At first, they used Opinionfinder, a tool that classifies tweets on a daily time series as positive versus negative to determine the collective moods of users. Secondly, they used another tool, GPOMS, which classifies tweets based on six dimensions of moods, namely: happy, vital, alert, calm, sure and kind. Then Granger Causality Analysis was used to correlate the collective public mood to the Dow Jones Industrial Average (DJIA). Finally, a fuzzy neural network classifier was used to make improved prediction accuracy in DJIA prediction models using measurements obtained from the collective moods of users on Twitter. This research made extensive use of psychometric analysis to make predictions on such a confounding size of data, resulting in impressive accuracy of 87%.

## 2.5 Literature Review of the Tools Used

### 2.5.1 Psychometric evaluation using LIWC

Linguistic Inquiry and Word Count (LIWC) compute the extent to which individuals use diverse classes of words over a wide exhibit of writings, including messages, emails, speech, poems, and every-day discourse. It provides you with the ability to focus on the degree to which any text content uses positive or negative feelings, inter-personal mentions, causal words, and multiple other dialect measurements (Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007).

LIWC can break down many standard ASCII content records and Microsoft Word archives in terms of various linguistic and behavioural dimensions. For example, pronouns, prepositions, articles are linguistic dimensions of text. In addition, categories such as positive and negative emotions, anger, and sadness are psychological dimensions of text. The descriptive list of these 67 word categories can be found on LIWC website<sup>5</sup>. It likewise permits you to fabricate your own word category references that can be built upon these linguistic and behavioural dimensions to break down and analyse the text particularly significant to your study (Pennebaker et al., 2007).

LIWC is proficient and powerful for anticipating the different structural, emotional, perceptual and cognitive components existing in people's verbal and composed discourses (Pennebaker et al., 2007).

LIWC applications are intended to investigate and analyse written content on a word-to-word premise, and compute the rate at which the words appear in the content by matching them to 67 word categories provided in LIWC default dictionary, and create the results as an output file which is a tab-delimited document that can be specifically read into applications such as Microsoft Excel (Pennebaker et al., 2007).

The LIWC application has an inbuilt dictionary of words and word categories, previously mentioned as linguistic dimensions, which classifies which words ought to be tallied in the target text file. Words read and examined by LIWC are target words. Words in the LIWC inbuilt default dictionary are dictionary words. Gatherings of dictionary words

---

<sup>5</sup> <http://liwc.net/descriptiontable1.php>

relating to a specific dimension, for example positive feeling words, are characterised as word categories (Pennebaker et al., 2007).

For classification purposes, LIWC analyses the text document word by word from start to finish. It then searches its dictionary for a match with a dictionary word and then assigns it to the respective word category. If a target word matches a dictionary word, then a respective word category scale is incremented (Pennebaker et al., 2007).

To understand the LIWC text processing module, let us scale it down to the following example. Consider a sentence that has a word count of ten ( $w_1, w_2, \dots, w_{10}$ ) and it needs to be categorised in five word categories, namely  $wc_1, wc_2, \dots, wc_5$ . The text-processing module will match every target word to a dictionary word as shown in Figure 1. As mentioned earlier, for every match the appropriate word scale category is incremented. On the other hand, if a target word ( $w_4$ ) does not match any dictionary word entry, then the target word is skipped and therefore not categorised in any word category. As seen in Figure 1,  $wc_4$  is not incremented because none of the target words match that particular category.

Since,  $w_1, w_2$ , and  $w_3$  were categorised under  $wc_1$ , it simply indicates three out of ten words have incremented that particular word category. As seen in Table 3, LIWC generates output for every word category based on the following formula in Equation 1.

$$V(wc_n) = \frac{N}{TWC} \%$$

Equation 1– LIWC Word Category Output

Where,

- ‘ $V(wc_n)$ ’ is the output value generated by LIWC for a particular word category,
- ‘ $N$ ’ is the number of words categorised in that particular word category, and
- ‘ $TWC$ ’ stands for total word count.

By implementing this formula for every word category, the output result file of LIWC looks like that shown in Table 3.

wc1	wc2	wc3	wc4	wc5
30%	20%	20%	0%	20%

Table 3– LIWC Example Output

So far we have seen how LIWC aids in effective text categorization by using its internal default dictionary to analyse text and categorise every target word to its respective word-category. In addition LIWC computes total percentage of words in any given text that belong to a specific word-category. Therefore, it enables the user to understand the style of writing prevailing in textual records. It is these styles of writing as opposed to specific word usage that individuals use in cyberbullying, on which the predictive model is built upon in this research.

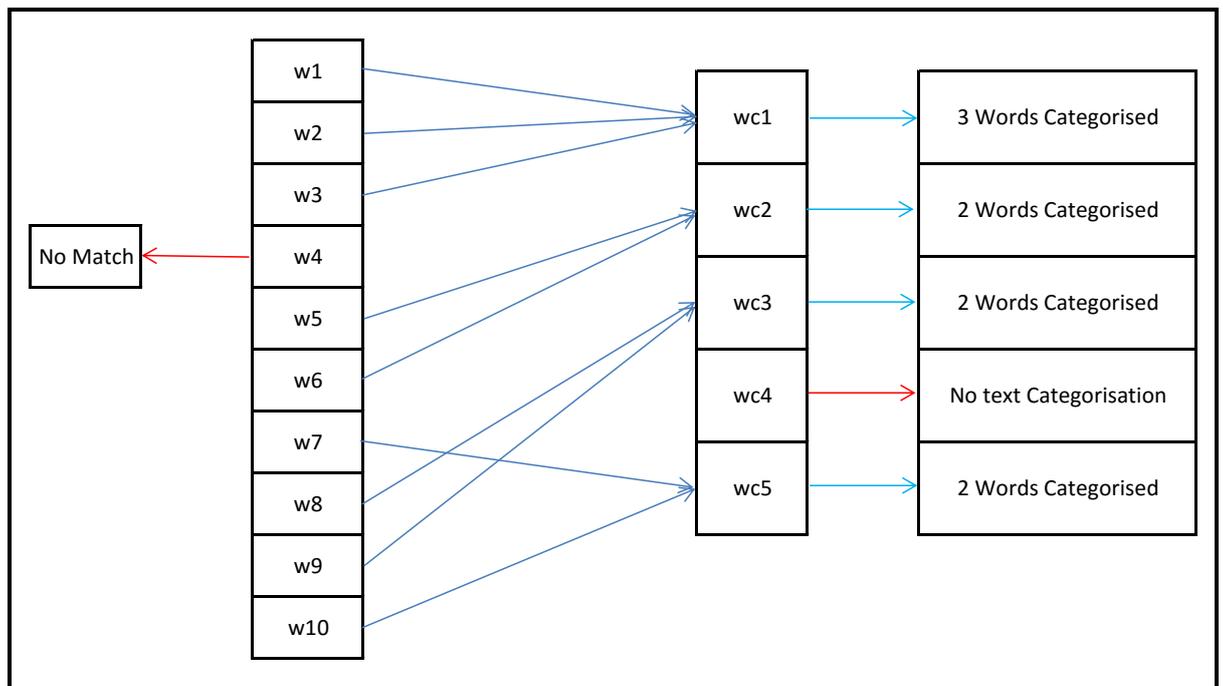


Figure 1 – Text Categorisation in LIWC

The aforementioned example is scaled down significantly to understand the operational use of LIWC text categorization. However, Tausczik and Pennebaker (2010) provide comprehensive large scale examples in which they have used LIWC for effective text categorization of text documents containing over tens of thousands of words. Further, Tausczik and Pennebaker (2010) explained the development of LIWC dictionary over the years since its conception. Multiple judges with relevant knowledge in pragmatics of language combined with numerous brain storming sessions, collectively decided which

words must be included in specific word-categories while developing the dictionary of LIWC. The decisions made were based on: the agreement by majority judges' votes and the inter-reliability of words falling in similar category using Cronbach's alpha. Finally, they used Pearson correlational analyses to validate LIWC text categorization consistency, or as they term it 'LIWC external validity'. By using LIWC categorization on several hundred million words, an extremely high correlation was revealed between LIWC scales and judges' ratings (Tausczik & Pennebaker, 2010). These comparisons can be viewed on LIWC website<sup>6</sup>.

### 2.5.2 TAGS archiving tool

TAGS is a Twitter archiving tool in the form of a Google Sheet template, that enables you to automatically collect search results from Twitter. To set up a personal Google Spreadsheet for TAGS, visit the URL<sup>7</sup> created by Martin Hawksey. Figure 2 provides a comprehensive overview of the TAGS archiving tool functionality.

After setting up TAGS application to retrieve tweets, keywords were entered in 'Enter term' to retrieve tweets. The default number of tweets retrieved will be 3000. However, the user can change the limit to 18000.

The keywords that were used to retrieve the tweets are, *nerd, gay, loser, freak, emo, whale, pig, fat, wannabe, poser, whore, should, die, slept, caught, suck, slut, live, afraid, fight, pussy, cunt, kill, dick, bitch*.

By clicking 'Run now' from 'TAGS' menu bar after entering the keywords, the data will be archived from Twitter on Google Spreadsheet extension provided on the TAGS GUI. Twitter provides the user with API keys for functionality of TAGS. These API keys allows the users to retrieve and archive Twitter data based on some authorization steps which are explained in detail in Appendix 1.

---

<sup>6</sup> <http://liwc.net/descriptiontable1.php>

<sup>7</sup> <https://tags.hawksey.info/>



### **2.5.3 Weka toolkit**

Weka is a classic repository of machine learning classifiers used for data mining purposes. It allows the user to perform data mining tasks on single or multiple datasets directly through its interactive graphical user interface, or the tool can be called from the user's own java code. Apart from Weka's classic repository of algorithms, it has inbuilt tools for users to perform tasks like data pre-processing, individual classification, regression, clustering and association rules parameter settings, and has an excellent interactive visualization function for the data mining tasks (Bouckaert et al., 2013).

In addition, Weka has an inbuilt experimenter that can be utilised for comparing the reliability of different classifiers used on the same datasets. This enables the user to validate the data mining experiments carried out in the Weka environment (Bouckaert et al., 2013).

### **2.5.4 Overview of classification algorithms used in Weka**

#### **2.5.4.1 Random Forest**

The Random forest classifier is a collection of multiple decision trees. Assuming there are  $T$  numbers of decision trees generated by the classifier, a random vector  $v^T$  is generated for every decision tree, which is also unique in comparison to all the past vectors  $v^1, \dots, v^{T-1}$ . All the generated vectors have the same distribution. Then using  $v^T$  and the training dataset, a decision tree is grown that generates a classifier  $c(i, v^T)$ , where  $i$  is the input class. When Random Forest completes growing  $T$  number of decision trees, these decision trees cast a vote to decide the most efficient input class  $i$ . The input class with the most votes becomes the final choice for classification by Random Forest (Breiman, 2001).

Random Forest outperforms other classifiers in accuracy and is highly efficient in eliminating the overfitting issue on the training dataset. It is also highly efficient in analysing highly dimensional datasets containing many instances and has methods to overcome the common problem of class imbalance. In addition, the classification process is unbiased as it works on generating random vectors that build the final classifier output (Breiman, 2001).

### 2.5.4.2 Multilayer Perceptron

Multilayer Perceptron (MLP) is a supervised neural network classifier. To train MLP it is necessary that it have a desired output for training the dataset. Like a human brain, MLP's knowledge acquisition primarily takes place through learning. Secondly, it uses synaptic weights to store knowledge between the strengths of inter-neuron connections (Du & Swamy, 2014).

The major task of MLP modelling involves accurate mapping of the input data to obtain the desired output based on past instances. The neural network model created by MLP must be able to generate accurate output even though the desired output is sometimes missing or unknown (Du & Swamy, 2014).

Backpropagation (BP) algorithm forms the backbone of MLP modelling. BP is responsible for continuously feeding the input data to the neural network. For every instance, the model is responsible for comparing the classifier output generated to the desired output. The error between these two is fed back or backpropogated to the neural network, which then recalibrates the weighting system with the intention of reducing the error rate at each iteration. This is how MLP is trained to reduce error and generate the desired output (Du & Swamy, 2014).

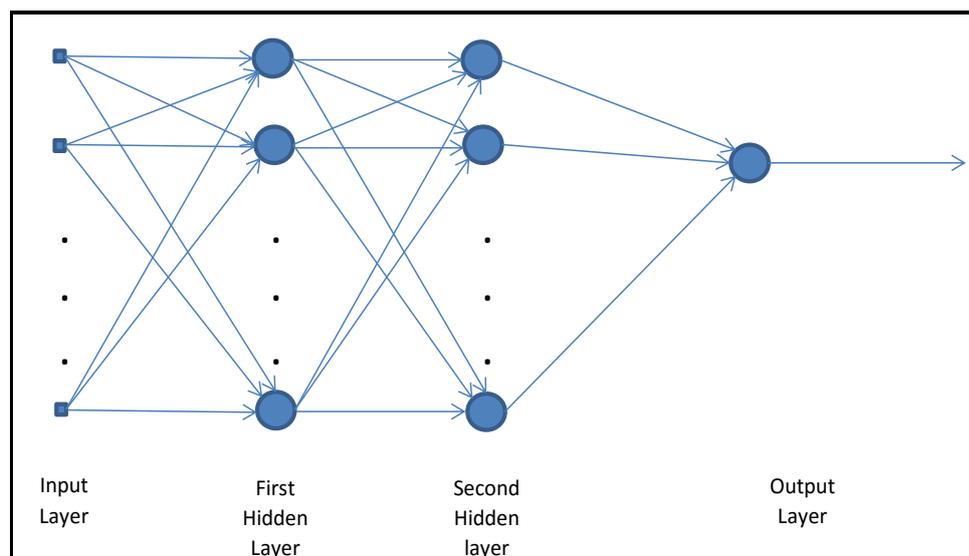


Figure 3 – Multilayer Perceptron, Neural Network Classifier

MLP's adaptive learning ability tends to get smarter with each iteration, hence discarding any occurrences of misclassification (Du & Swamy, 2014).

#### **2.5.4.3 Sequential Minimal Optimization**

Sequential Minimal Optimization (SMO) is a classifier used to train Support Vector Machine (SVM). SMO is widely used to solve the problem of quadratic programming (Meyer & Wien, 2014).

SVM is a differential classifier characterised by an isolating hyperplane. The classifier that is trained on a supervised labelled dataset yields an ideal hyperplane, which classifies new samples. The objective of SVM is to achieve an optimal hyperplane (Meyer & Wien, 2014).

Given a multi-dimensional labelled dataset being used to predict a nominal class category, where each instance belongs to either one of the categories, SVM modelling represents the instance of two classes as points in space. SVM further does mapping of these points until a clear gap, as broad as possible, is observed between them. This gap is the largest minimum distance between the two class instances and is hence called a margin. New instances are mapped into this largest minimum distance and classification occurs based on the nearest class instances. SVM provides the user with a regularisation that addresses the over-fitting problem of the dataset (Meyer & Wien, 2014).

#### **2.5.4.4 J48 Decision Tree**

A decision tree learning algorithm, as the name suggests, classifies the target variables based on the decisions made by the classifier on the input variables. Leaves and branches are common terminologies used to describe the decision tree classifier, where leaves are the target or class variable and branches are conjunctions of input variables that predict the class variable (Maimon & Rokach, 2008).

Decision trees predict the class variables depending on the attribute values of the input variables. Decision trees divide the input variables into smaller subsets based on the weight or value of the input variables at the attribute level. This means that the classified smaller subset has rich information that predicts the class variables. The decision tree continues to split subsets recursively until the value of the end node is the same as the

class value. The decision tree classifier uses only those input variables that contain the most information to predict the class output and discards the rest of the variables (Maimon & Rokach, 2008).

# 3 Methodology

## 3.1 System Architecture

This section describes the architecture that was used to classify cyberbullying and non-cyberbullying tweets. The approach proposed in this research aims to identify the psychometric properties associated with the words in tweet texts. In order to achieve this, the study was conducted in three stages. These three stages form the baseline for this research. This section in particular provides an overview of the entire methodology. The three stages are elaborated on in Figure 4.

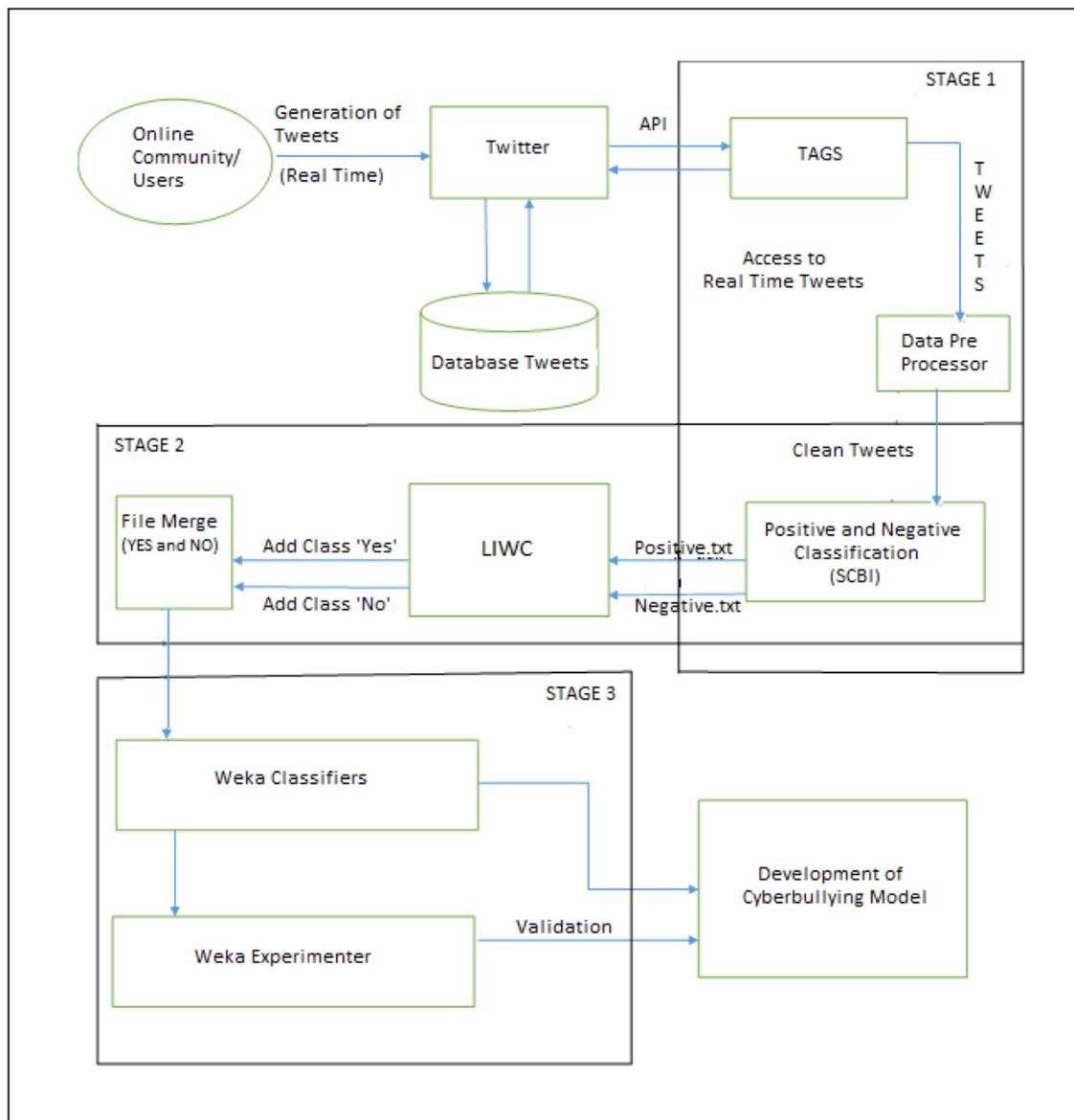


Figure 4 – System Architecture

The three stages that form the baseline for this research are as follows:

- Stage 1: Obtaining the dataset using TAGS archiving tool, Data pre-processing (D-PP) techniques and supervised cyberbullying identification of tweets (SCBI).
- Stage 2: Output of SCBI as input for LIWC, labelling LIWC output for every instance of tweet and file merger.
- Stage 3: Merged file data as input to Weka classifier and Weka experimenter.

The working of these three stages in parallel with each other can be explained by mapping the flow of this system architecture. Figure 5 illustrates the system flow, and the functionality of each stage.

### 3.1.1 System flow

This section along with Figure 5 provides a detailed explanation of the functionality of the three stages forming the baseline for this research. In Figure 5, P1, P2, P3, P4,..., PN are the patterns generated by different Weka classifiers.

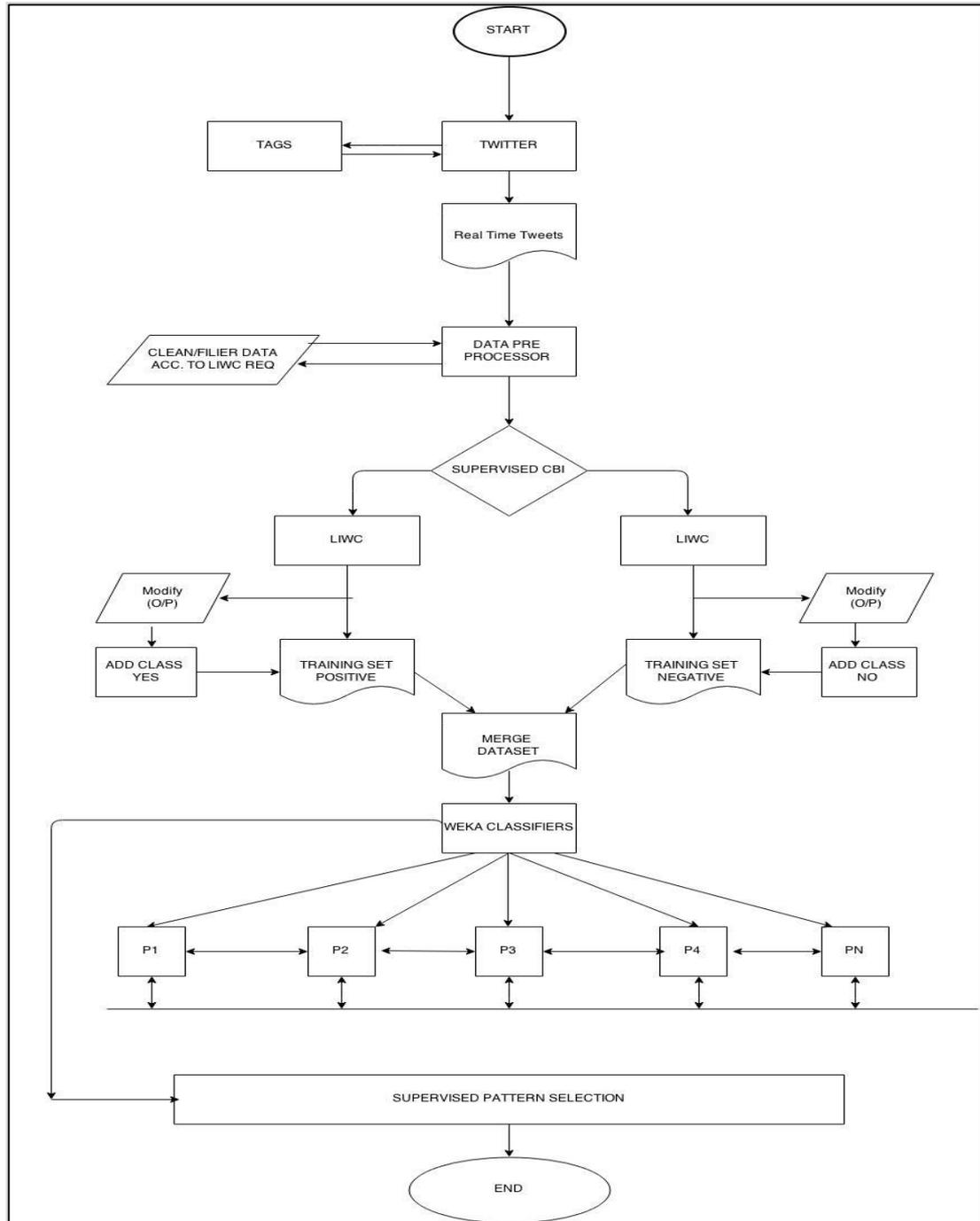


Figure 5 – System Flow Diagram

### 3.1.1.1 System Stage 1

Users from all over the globe collectively generate a massive number of tweets that are posted on Twitter in real time. These tweets are collected in real time and stored in a larger data repository. This data repository is a relational database where tweets are stored along with their meta data. This meta data could be the time stamp, geo stamp, profile information and so on.

The TAGS archiving tool, which is described in section 2.5.2, uses an API key provided by Twitter that enables it to retrieve tweets and archive them in a Google Spreadsheet. The data archived by TAGS is in tab-delimited format, in which one column has the tweets retrieved and the rest are the meta data (Appendix 1). TAGS can retrieve up to a maximum of 18000 tweets in a single run. For the purpose of this research, only the tweets for building the dataset were collected, hence all the meta data associated with it was discarded.

The tweets archived in the Google Spreadsheet were copied and saved in a new Microsoft Excel file. The data in the Excel file was stored in a single column where every single row was a different tweet. The archived Twitter conversational data is very noisy in nature and hence it required cleaning before proceeding to make it the training set. The character encoding of these retrieved tweets is not fully compatible with plain text. Therefore, in data preprocessing the tweets were converted to plain text with the help of Excel Macros.

In the next step of Supervised Cyberbullying Identification, the definitions of cyberbullying derived from psychological research on verbal aggressiveness were applied on the data to categorise them as cyberbullying or non-cyberbullying. Section 3.2.3.1 clarified all the definitive characteristics of cyberbullying. These tweets were stored separately in individual text files and were named 'Positive.txt' and 'Negative.txt'. Positive indicates that the tweet was indeed a cyberbullying tweet and negative indicates that it was not. Hence all the tweets identified as positive for cyberbullying were stored in 'positive.txt' and tweets that were not identified as cyberbullying were stored in 'negative.txt'.

### 3.1.1.2 System Stage 2

The application of LIWC is a potent way to categorise any written text in 67 language dimensions in total. These language dimensions are called the psychometric properties of text. (Section 2.4 covered the significance of psychometric properties and the text classified based on psychometric evaluation.) The output of LIWC was a tab-delimited file that categorised text in 67 psychometric categories of words for this research. Every row indicates an individual tweet. Every individual tweet is referred to as an instance. In the next step, LIWC was used to analyse the previously saved 'Positive.txt' and 'Negative.txt' individually and then two separate output files were generated. The outputs of these individual files were in '.xls' format that could be read using Microsoft Excel. These files are separately saved for further modification.

LIWC classified the 'Positive.txt' and 'Negative.txt'. These output files were modified by adding a variable called 'Class' for every instance. The class value in each instance for positive output file was 'yes' and similarly 'no' for negative ('yes' for cyberbullying tweets and 'no' for non-cyberbullying tweets). These files were then merged and saved as a combined training dataset for Weka classification. Table 4 in section 4.1 shows the class variable that was added to the dataset after LIWC categorisation.

### 3.1.1.3 System Stage 3

The training dataset developed in the previous stage was used in Weka for data mining tasks. Weka is a classic repository of classifier algorithms. The training dataset consisted of 1313 instances, out of which 376 instances were cyberbullying tweets. Similarly, the testing dataset comprised of 446 instances, out of which 123 instances were cyberbullying tweets.

Random Forest, Support Vector Machine, Multilayer Perceptron and J48 Decision Tree classifiers were used to train the classifiers to generate a predictive model to detect cyberbullying. In the final step of this research, the accuracy of each of the classifiers was validated using Weka Experimenter. In this environment, the accuracy of all the above classifiers was tested by estimating the standard deviations. This means calculating the degree to which the classifier output deviated in terms of its precision and recall testing. The significance of precision and recall is described later in section 4.2.

## 3.2 Data Pre-processor

This section is divided into three parts that aim to address:

- Purpose of data pre-processor,
- Functionality of data pre-processor, and
- Operational use of processed data.

### 3.2.1 Purpose of data pre-processor

To address this, it is necessary to understand the nature of data that is archived from TAGS.

Every character of text in most of the tweets is single byte. The character count in a tweet and the value of byte length is therefore always less than or equal to 148. If a user wants to tweet by using characters beyond the scope of the basic alphabet, numbers and standard punctuation, then the situation becomes tricky. In addition, it is observed that users tend to be more expressive and use emojis and accented characters in tweets (Schnoebelen, 2012). Kanji characters, for example, use multi-byte character encoding that allows users to represent text that is beyond single-byte text encoding. In short, Twitter allows only 148 bytes per tweet irrespective of single or multi-byte encoding. However, on the other hand, Twitter highlights the issue of character handling caused by the use of accented vowels and emojis generated using multi-byte encoding. This is because Twitter API accepts only UTF-8 encoding to ensure overall uniformity of data. In most cases, it is observed that tweets are encoded in an 8-bit multi-byte-character set (MBCS) rather than an 8-bit single-byte-character set (SBCS). A MBCS string may contain a combination of single-byte as well as double-byte characters. In addition, a two-byte MBCS consists of a lead and a trail byte and often they overlap with another MBCS or SBCS string in the text. In this case, they need to be monitored in order to understand their functionality in terms of which are the leading and trailing bytes so that data uniformity is maintained and subsequently corruption can be avoided.

Google Spreadsheet support UTF-8 encoding and often the above-mentioned overlap of leading and trailing bytes corrupts the data that is retrieved from Twitter API using TAGS. As a result, the data archived consists of @username, hashtags, hyperlinks, accented characters, excess carriage returns, unexpected leading and trailing spaces, hidden spaces,

line breaks, borders, over usage of emoticons and punctuation, spelling mistakes and redundancy. These unwanted dimensions in the data make it noisy in nature and difficult to analyse, hence it needs to be cleaned.

On the other hand, LIWC accepts only plain ASCII (Range 32-127) text files for the psychometric analysis of text. Moreover, to achieve improved accuracy using this tool it is necessary to rectify spelling mistakes in the document and get rid of all the grammatical errors. Hence, the data needs to be pre-processed to make it compatible for LIWC to further process it.

### **3.2.2 Data pre-processor functionality**

Data archived in Google Spreadsheet is exported in Microsoft Excel for pre-processing. Microsoft excel is a powerful tool in which to format data. It provides macro functions to perform a particular task that requires a single instruction, which further automatically expands to a set of instructions. Hence, extensive use of macros is the core of data pre-processor functionality. Figure 6 gives a comprehensive outlook of data pre-processor functionality.

#### **3.2.2.1 Microsoft Excel macros functionality**

The details of the macro code are attached in Appendix 2.

The features for text cleaning are summarised below.

1. Noise elimination:
  - @username - to preserve anonymity.
  - #hashtags - to reduce the jargon in tweets. Any type of jargon can hamper LIWC accuracy.
  - Hyperlinks - to avoid more jargon.
  
2. Cleaning and formatting:
  - Trim spaces – to remove excess space between words.
  - Remove borders – to keep the dataset uniform and maintain compatibility with plain text.

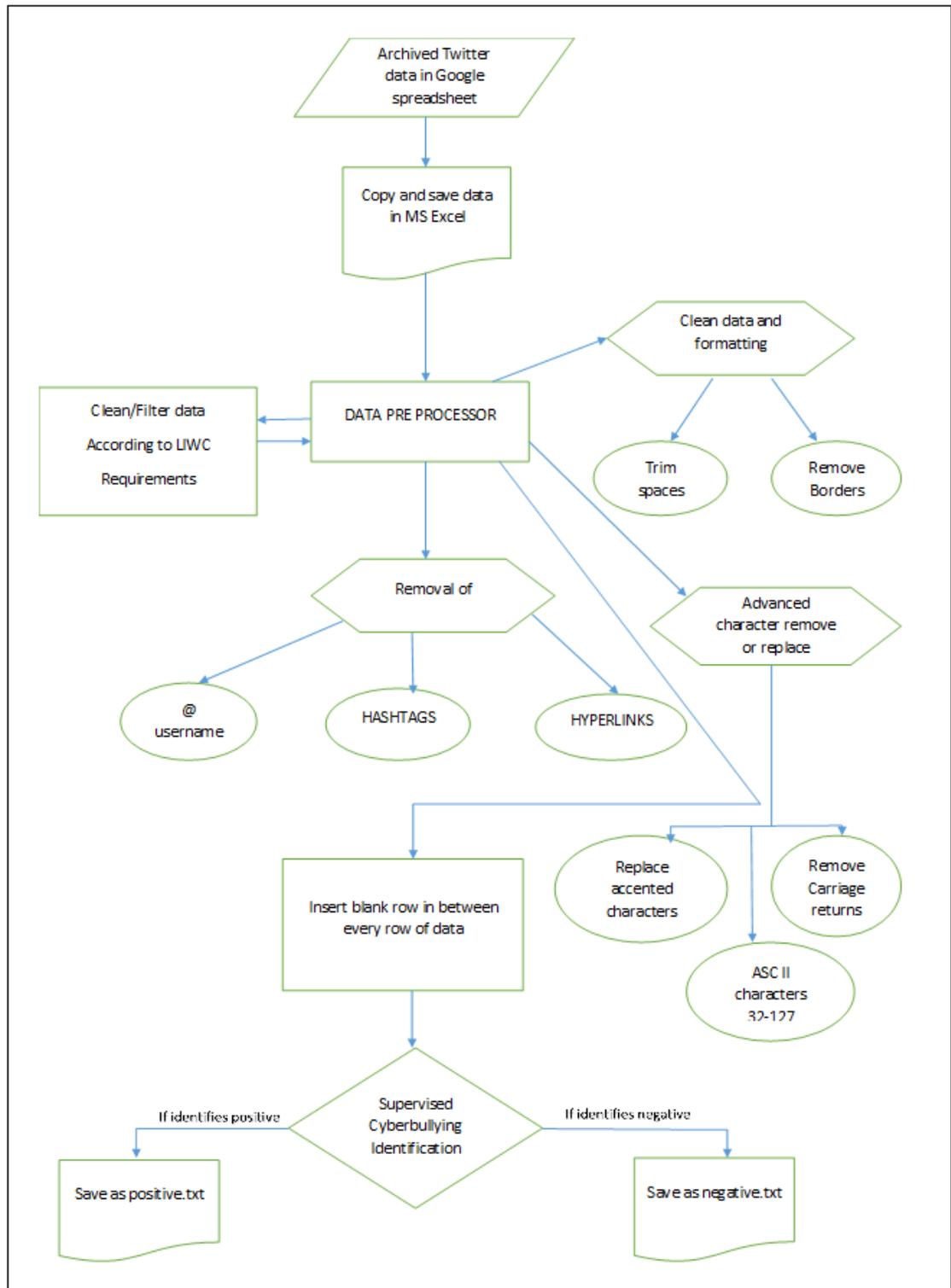


Figure 6 - Schematic Representation of Data Pre-Processor

3. Advanced character removal or replacement:
  - Replace accented characters – accented characters are multi-byte, which are not compatible with plain text format.
  - Remove carriage returns – to ensure that text in each cell remains intact.
  - ASCII (32-127) – convert the text into plain text that is readable by LIWC.
4. Insert a blank row after every row of data.

### **3.2.3 Operational use of processed data**

After the data was converted to plain text, the next task involved supervised cyberbullying identification (CBI). (Please refer to Appendix 3 to view this dataset.) In this step, the plain text data was manually categorised as cyberbullying or not-cyberbullying. As explained previously, these files were saved separately as ‘positive.txt’ for cyberbullying and ‘negative.txt’ for non-cyberbullying. This was done because each of these files were classified individually using LIWC for psychometric categorization of text in tweets. Then a ‘class’ variable indicating ‘yes’ for cyberbullying and ‘no’ for non-cyberbullying were assigned to individual instances categorized by LIWC. In this way, no text was misclassified and in addition it simplified machine learning. In order to classify each tweet, it was necessary to define rules for categorisation.

Psychology research by Salmivalli and Peets (2009) concerning the field of bullying confirms certain characteristic traits of bullies. Bullies tend to be highly aggressive, hostile and domineering towards others. In addition, bullies tend to have a positive attitude towards aggression and a negative attitude towards peers. Bullies also tend to have a low level of behavioural conduct and low cooperation with others (Salmivalli & Peets, 2009). Therefore, it can be inferred that aggression, hostility and negativity are key to understanding a bully’s mind. Furthermore, the verbal dialects of a bully are highly driven along the lines of bully mentality (Salmivalli & Peets, 2009).

According to various psychological studies conducted based on interviews of victims who were cyberbullied, it was identified that bullies post rumours, random threats and personal information of victims (Finley, 2014; Heirman & Walrave, 2008; Riebel, Jaeger, & Fischer, 2009; Sleglova & Cerna, 2011). In addition, the aforementioned bully mentality of anger, hostility, aggression and negativity collectively result in verbal aggressiveness in real life and on their internet activities (Salmivalli & Peets, 2009).

Psychological research reveals a strong link between verbal aggressiveness in written dialects and the language used by bullies on the internet (Dooley, Pyżalski, & Cross, 2009; Kowalski et al., 2012). In several events of cyberbullying, victims have mentioned that bullies use strong abusive verbal dialects. These dialects of speech are related with verbal aggressiveness (Vaillancourt et al., 2008).

Since the conception of psychological research pertaining human behavior, it was found that written and verbal dialects of humans are triggered by their current state of moods (Salmivalli & Peets, 2009). The mood of a bully, as mentioned previously, are dominated by the feelings of anger, hostility, aggression and negativity. Further, it was found that these moods trigger verbal aggressiveness in humans. Hence, based on these relational links, a tweet can be classified as a cyberbullying tweet if verbal aggressiveness is found to be pre-dominant in it.

In this research, to classify a tweet as a cyberbullying tweet, various underlying components of cyberbullying were studied. The main highlights of the components of verbal aggressiveness used in this research to classify tweets as cyberbullying are: -

- *Character attacks*, wherein, the reputation, integrity and morals of an individual are targeted with the purpose of defamation.
- *Competence attacks* are types of attacks wherein bullies denigrate individual's ability to do something.
- *Malediction* is used as an attack in which bullies curse and express a wish for some type of misfortune or pain to materialize in an individual's life.
- *Physical appearance attacks* are targeted on an individual's look and bodily structures. Typically, physical attributes of humans are found to shape and develop their personality and social behavioural relations. Due to the need of an individual to

be socially accepted, these types of attacks make victims feel socially neglected making a long lasting negative impact on their self-esteem.

- *Insults*, which are typically intentional, are attacks targeted to disrespect an individual in their social circles.
- *Profanity* is used as an attack wherein bullies use extremely offensive language that typically include foul, lewd, vulgar language in addition to swearing and cursing words.
- *Verbal abuse* is a type of attack that includes false accusations or blames, extreme criticisms and judgements about an individual and or statements that negatively define the victim.
- *Teasing*, if hurtful in nature and done as a spectacle for others to witness results in harassment and humiliation for the victim. It is perceived as a form of emotional abuse.
- *Threats*, are generally anonymous in cyberbullying. Due to this anonymity victims tend to live in constant 'fear' that leads to long-lasting depression, low self-esteem and delinquent behaviours.
- *Name-calling* is a type of attack wherein bullies use denigrating, abusive names and associate them to the victims leaving them extremely humiliated in front of others.
- *Mockery* is a type of attack wherein bullies pass comments on victims making them feel worthless, disrespecting them and make fun of them in front of everyone. Escalated form of mockery leads to low self-esteem of the victims.

The psychological literature provides a link between above mentioned types of attacks with verbal aggressiveness. As discussed previously, verbal aggressiveness is linked to bully mentality. Verbal aggressiveness is a very broad topic of study. However, based on various surveys and interviews of the victims of cyberbullying, it can be inferred that cyberbullying text indeed contains verbal aggressiveness. These types of attacks are very well defined in the literature and it would be difficult to get ordinary human annotators to classify it since they would require extensive study to be able to make the judgement.

All the tweets that were identified as cyberbullying were exported to a new text file and saved as 'Positive.txt', and the tweets that were not identified as cyberbullying were exported to a separate text file and saved as 'Negative.txt'.

### 3.3 LIWC Psychometric Evaluation on Tweets using LIWC

When you launch the LIWC application, a pop-up screen shows the currently loaded dictionary. We used the *Internal Pennebaker 2007 Dictionary*. Section 2.5.1 explained in detail the functionality of this dictionary with an example. The clean and noise-free data ('Positive.txt' and 'Negative.txt') obtained from the Data Pre-processor became the input text for LIWC.

#### 3.3.1 Parameter settings for LIWC processing

For the purpose of this study, it was necessary to process the text using all 67 word categories.<sup>8</sup> More word categories would imply that the text categorization would provide extensive information about word-categories prevalent in cyberbullying. The output of LIWC was the input for the data mining classifiers. Classifier output results in high efficiency for detecting cyberbullying if provided with large number of attributes in the dataset.

LIWC analyses text in segments. Therefore, for this study the text segment delimiter was set on two or more returns. This means that LIWC treated any text after two returns as a new instance and processed it independently. Based on these parameter settings, LIWC processed text and generated an output that was divided over 67 numeric word categories and every tweet processed was a unique instance. In the next step, as shown in Figure 7, 'positive.txt' and 'negative.txt' files were processed through LIWC and the results were saved as 'positive.xls' and 'negative.xls'. Then the text input data was converted to a numeric dataset as shown in Table 4.

The results of the LIWC process on 'positive.xls' and 'negative.xls' were manually labelled in order for the classifier to understand the two different classes of 'yes' and 'no'. For example, 'positive.xls' had 67 existing word categories, with 376 unique instances predicting cyberbullying tweets. Similarly, 'negative.xls' had 67 existing word categories, with 937 unique instances that did not signify cyberbullying. Every individual tweet is a single instance. Adding one more variable called 'Class' for every instance in 'positive.xls' and 'negative.xls' with values of 'yes' and 'no' respectively concluded the

---

<sup>8</sup> <http://www.liwc.net/descriptiontable1.php>

task of data labelling. The 'positive.xls' and 'negative.xls' were then merged together and saved as 'combined results.xls'. This file became the training dataset for the machine learning classifiers. This file now had 68 variables or word categories (67 numeric and 1 Class) and 1313 unique instances, out of which 376 instances acted as predictors for cyberbullying.

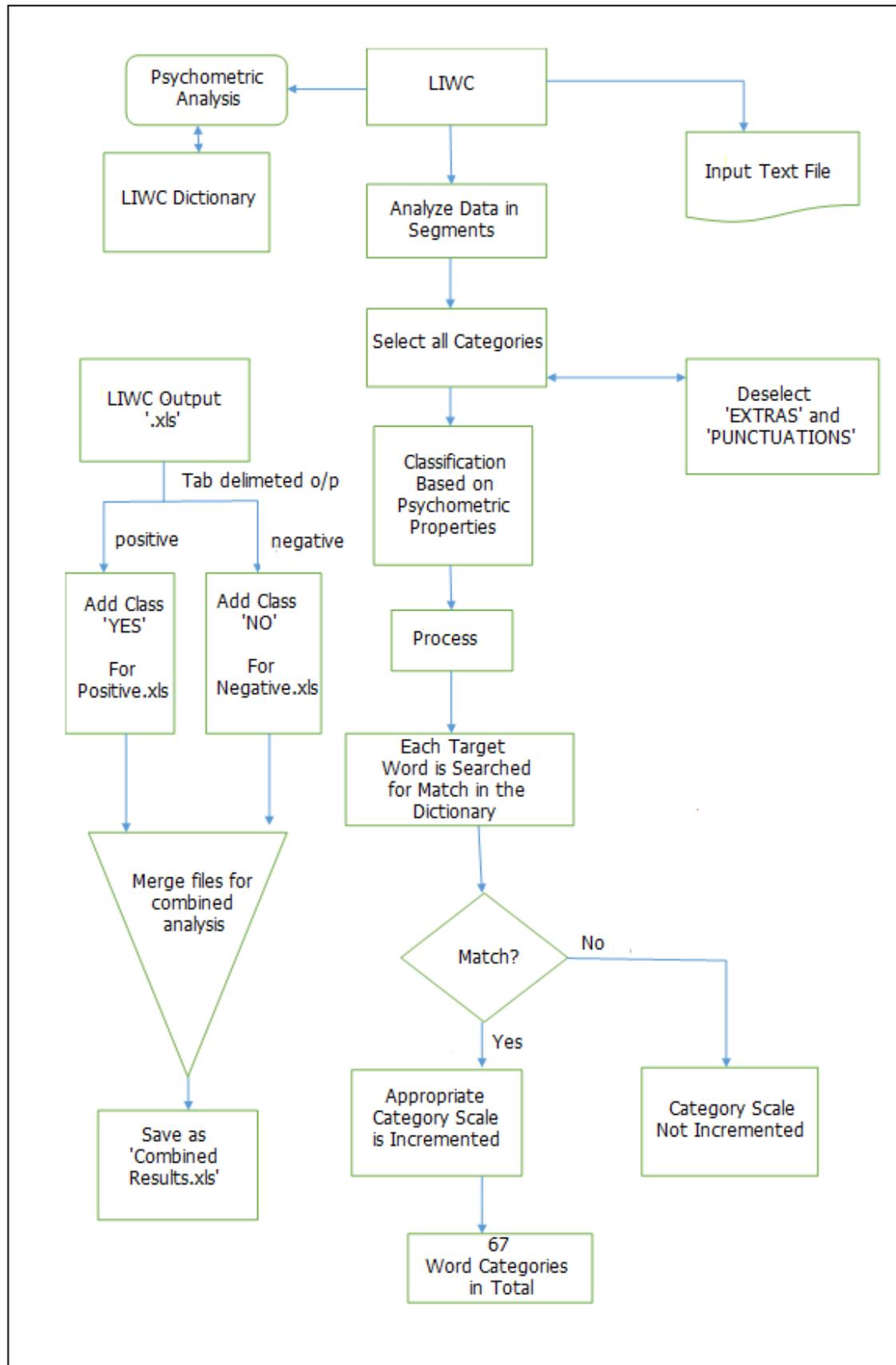


Figure 7 – Schematic Representation of LIWC Functionality

### 3.4 Weka Classification

Weka classification was used to develop the predictive model to detect cyberbullying. Figure 8 shows the experiments performed in Weka Explorer. The merged output files of LIWC that were labelled with a class attribute of value 'yes' and 'no' were used as the training datasets for the classifier algorithms. This training dataset had 1313 instances out of which 376 were labelled cyberbullying instances. It consisted of 67 psychometric variables that were numeric in nature and one class instance which was nominal. This dataset was extremely overlapping and had class imbalance.

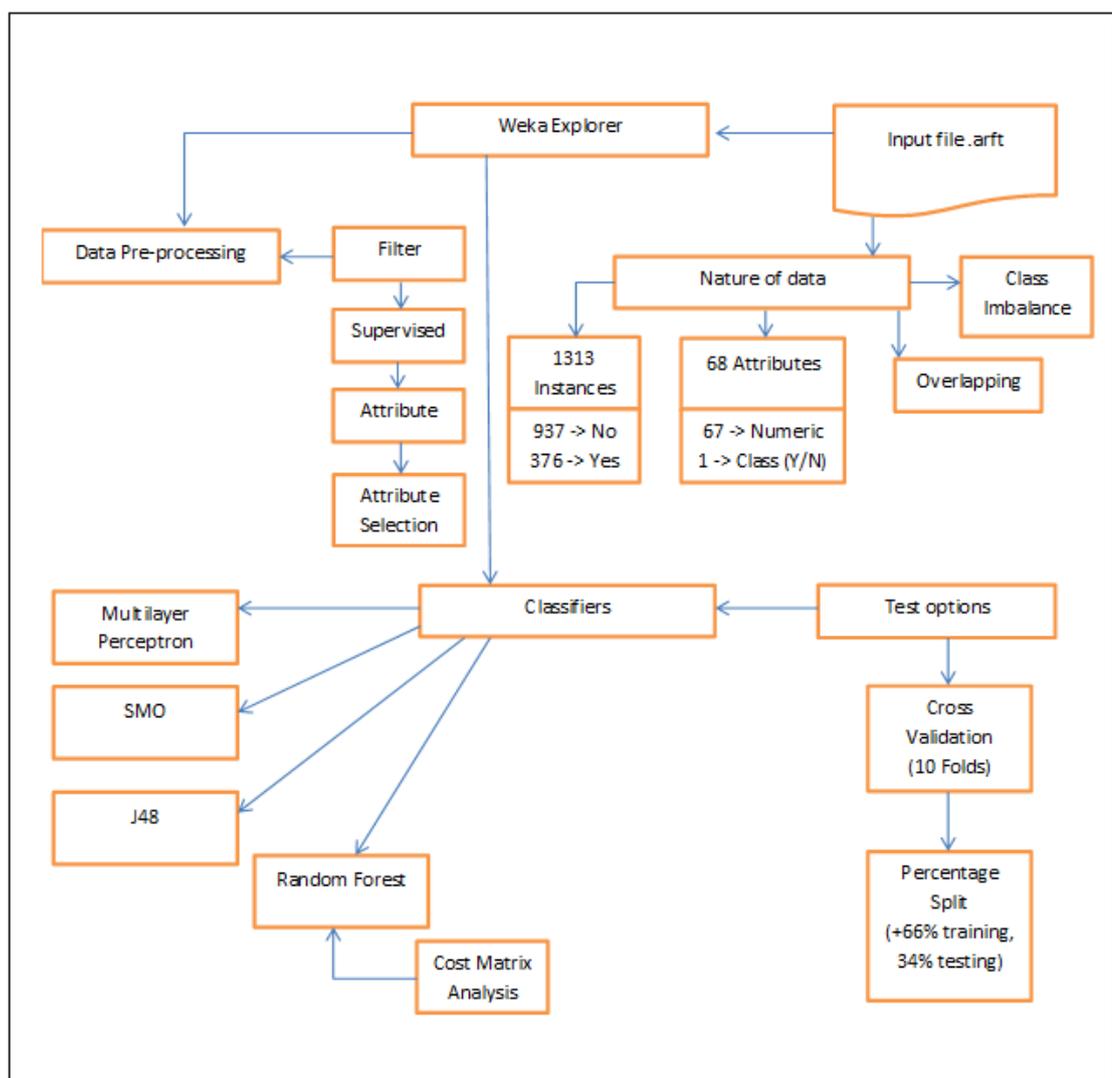


Figure 8 – Weka Classification Overview Diagram

This training dataset was loaded into Weka Explorer so that the classifiers could be trained to develop a model to detect cyberbullying in Twitter. The data pre-processing window in Weka Explorer provides filters for attribute selection. Using these attribute selection filters, the training dataset could be trimmed of unnecessary attributes that were insignificant in decision making as they exhibited less information for detecting the class attribute.

In this research, two attribute selection filters, Correlation feature selection (CFS) Subset Evaluator and Info Gain Attribute Evaluator, were used to determine the predictive ability of the classifiers for accurate classification. Firstly, CFS Subset Evaluator estimated the value of the selected features and considered the individual predictive ability of those features and redundancy between them. On the other hand, Info Gain Attribute Evaluator calculated the value of a feature depending on how rich its predictive information was in determining the class.

The classification window in Weka Explorer contains a repository of classification algorithms that can be used pertaining to the interest of study. For the purpose of this research, Random Forest, Support Vector machine (Sequential Minimal Optimization), Multilayer Perceptron and J48 Decision trees were trained on the dataset provided. These classifiers have exceptional computing abilities on a multivariate dataset that have numeric variables to determine the final class variable which is nominal. Section 2.5.4 explains the significance of the unique value propositions of each of these classifier algorithms.

For this study, there were 67 numeric variables in the dataset that lead to the classification of every instance to a nominal class variable indicating 'yes' and 'no'. The test options for each of these classifiers were, first, 10-fold cross validation and, secondly, percentage split in which the data was randomly divided as 66% for the training set and 34% for the test set. The test set comprises of 446 instances out of which 123 instances are cyberbullying tweets. Further, data mining using Weka aided in machine learning, where the classifiers extracted useful information by linking the significant data attributes of cyberbullying from the training set to generate a detection model.

After applying the classification techniques using data mining, the results of these classifier outputs were statistically compared to find out which classifier best suited the training set to develop the most accurate model. This was done by using a statistical approach to calculate the standard deviations observed across classifier outputs in Weka

Experimenter by adding all algorithms and the training dataset. The experiment was 10-fold cross-validated in conjunction with 10 iterations per experiment. This resulted in a total of 400 experiments. To analyse these results we performed a paired T-Test that showed us the comparison between different classifiers on the same dataset.

# *4 Results and Discussion*

## **4.1 LIWC Results**

The accuracy of LIWC is directly proportional to the quality of the text document it analyses. Here, quality implies to the compliance of certain text editing pre-requisites for analysing text documents, as explained previously in section 3.3. This quality, for the purpose of LIWC, can be termed clean and noise-free.

LIWC analysed the clean, noise-free text document and categorised words from individual tweets into a total of 67 subcategories of psychometric text properties, namely; Linguistic Process, Psychological Process, Personal Concerns and Spoken Categories.

Table 4 shows a snippet of LIWC output for the first 13 instances. In total 1313 such instances were analysed in this study. It is seen that LIWC converts every text instance into numerical format. Due to its ability to quantify any written text, it is simple to visualise the degree to which different word-categories are used. It provided a range of numerical values for different word-categories that assisted in the prediction of cyberbullying tweets. Such an information-rich document made machine learning easier and therefore resulted in a robust predictive model to detect cyberbullying.

The first column of Table 4 contains a full list of the 67 word-categories in LIWC. The following columns refer to individual tweet instances that were categorised respectively in each of those 67 word-categories. The categorisation of tweets was based on the method explained in section 2.5.1. The last category, called ‘class’, contains the nominal values ‘yes’ and ‘no’, which identifies the individual instance as a cyberbullying tweet or non-cyberbullying tweet respectively. (Appendix 3 contains the full LIWC output for all 1313 instances.)

Categories	Instance 1	Instance 2	Instance 3	Instance 4	Instance 5	Instance 6	Instance 7	Instance 8	Instance 9	Instance 10	Instance 11	Instance 12	Instance 13
WPS	19	27	9	8	12	26	21	15	13	6	5	25	12
Sixltr	0	3.7	0	0	8.33	3.85	9.52	13.33	7.69	0	20	4	16.67
Dic	84.21	92.59	77.78	87.5	75	92.31	85.71	93.33	84.62	100	60	100	75
funct	47.37	55.56	11.11	37.5	33.33	53.85	47.62	46.67	38.46	0	0	52	16.67
pronoun	10.53	37.04	11.11	12.5	0	23.08	23.81	26.67	15.38	0	0	20	0
ppron	5.26	37.04	11.11	0	0	19.23	19.05	20	15.38	0	0	20	0
i	0	11.11	0	0	0	11.54	4.76	0	7.69	0	0	12	0
we	0	0	0	0	0	0	0	0	0	0	0	0	0
you	0	25.93	0	0	0	7.69	14.29	20	7.69	0	0	4	0
shehe	5.26	0	11.11	0	0	0	0	0	0	0	0	0	0
they	0	0	0	0	0	0	0	0	0	0	0	4	0
ipron	5.26	0	0	12.5	0	3.85	4.76	6.67	0	0	0	0	0
article	10.53	0	0	0	0	3.85	4.76	6.67	7.69	0	0	4	8.33
verb	15.79	22.22	0	25	16.67	7.69	14.29	6.67	7.69	0	0	20	16.67
auxverb	10.53	11.11	0	12.5	8.33	7.69	4.76	0	7.69	0	0	16	0
past	15.79	11.11	0	0	0	0	4.76	0	0	0	0	0	0
present	0	3.7	0	12.5	8.33	7.69	4.76	6.67	0	0	0	8	16.67
future	0	3.7	0	12.5	8.33	0	4.76	0	0	0	0	4	0
adverb	10.53	7.41	0	0	8.33	11.54	0	0	0	0	0	4	0
preps	0	0	0	0	16.67	3.85	9.52	13.33	15.38	0	0	4	8.33
conj	5.26	7.41	0	0	0	11.54	4.76	0	0	0	0	8	0
negate	0	0	0	12.5	0	0	0	0	0	0	0	0	0
quant	0	0	0	0	0	0	0	0	0	0	0	0	0
number	5.26	0	0	0	0	0	0	0	0	0	0	0	0
swear	5.26	11.11	22.22	12.5	0	0	0	6.67	15.38	16.67	20	4	0
social	10.53	25.93	11.11	0	8.33	7.69	28.57	26.67	7.69	16.67	0	12	8.33
family	0	0	0	0	0	0	0	0	0	0	0	0	0
friend	0	0	0	0	0	0	4.76	0	0	0	0	0	0
humans	0	0	0	0	0	0	0	0	0	0	0	4	8.33
affect	10.53	18.52	0	12.5	0	34.62	9.52	13.33	15.38	16.67	60	12	16.67
posemo	0	3.7	0	0	0	7.69	4.76	0	0	16.67	20	4	0
negemo	10.53	14.81	0	12.5	0	26.92	4.76	13.33	15.38	0	40	8	16.67
anx	0	0	0	0	0	0	0	0	0	0	0	0	0
anger	5.26	11.11	0	12.5	0	7.69	4.76	6.67	7.69	0	20	8	16.67
sad	0	0	0	0	0	0	0	0	0	0	0	0	0
cogmech	0	11.11	22.22	12.5	25	7.69	19.05	13.33	7.69	0	0	36	0
insight	0	0	0	0	0	0	0	6.67	0	0	0	4	0
cause	0	3.7	0	0	0	3.85	4.76	0	0	0	0	0	0
discrep	0	7.41	0	0	16.67	0	4.76	0	0	0	0	8	0
tentat	0	0	0	12.5	0	0	0	0	7.69	0	0	0	0
certain	0	0	0	0	0	0	0	6.67	0	0	0	8	0
inhib	0	0	22.22	0	0	0	0	0	0	0	0	0	0
incl	0	0	0	0	8.33	0	9.52	0	0	0	0	8	0
excl	0	0	0	0	8.33	3.85	0	0	0	0	0	8	0
percept	15.79	0	44.44	0	8.33	0	4.76	6.67	0	0	0	4	0
see	15.79	0	22.22	0	0	0	0	0	0	0	0	0	0
hear	0	0	0	0	8.33	0	4.76	0	0	0	0	0	0
feel	0	0	22.22	0	0	0	0	0	0	0	0	4	0
bio	5.26	14.81	22.22	25	0	23.08	0	20	15.38	33.33	40	12	8.33
body	5.26	0	22.22	12.5	0	0	0	20	7.69	33.33	0	8	8.33
health	0	0	0	12.5	0	0	0	6.67	0	16.67	0	8	8.33
sexual	0	7.41	22.22	12.5	0	23.08	0	0	7.69	16.67	40	0	0
ingest	0	7.41	0	12.5	0	0	0	6.67	7.69	16.67	0	12	8.33
relativ	5.26	3.7	0	12.5	16.67	3.85	4.76	6.67	15.38	16.67	0	0	33.33
motion	0	0	0	0	8.33	0	0	0	0	0	0	0	16.67
space	0	0	0	0	8.33	3.85	4.76	6.67	7.69	0	0	0	16.67
time	10.53	3.7	0	12.5	0	0	0	0	7.69	16.67	0	8	0
work	0	0	0	0	0	0	0	0	0	16.67	0	0	0
achieve	0	0	0	0	0	0	4.76	0	0	16.67	0	0	0
leisure	5.26	3.7	0	0	0	0	4.76	0	0	16.67	0	0	0
home	0	0	0	0	8.33	0	0	0	0	0	0	0	0
money	0	0	0	0	0	0	0	0	7.69	0	0	0	0
relig	0	0	0	0	8.33	0	0	0	0	16.67	0	0	0
death	0	3.7	0	0	8.33	0	0	0	0	0	0	4	8.33
assent	0	0	0	0	0	7.69	4.76	0	0	0	0	0	0
nonfl	0	0	0	0	0	0	0	0	0	0	0	0	0
filler	5.26	0	0	0	0	0	0	0	0	0	0	8	0
class	yes	yes	yes	yes									

Table 4 – Snippet of LIWC Results

### 4.1.1 The preliminary results of LIWC

- The mean value of '*word count*' in all the 1313 tweets was 18 words per tweet.
- Every cyberbullying tweet comprised 13 words on an average.
- The mean value of '*dictionary words*' calculated for individual tweets shows that the dictionary could match 89% of '*target words*' with its '*dictionary words*'.
- This number dropped down to 87% while estimating the mean value of dictionary words used in cyberbullying tweets.
- As a result, for every cyberbullying tweet, 11 out of 13 words were categorised into its respective word categories.

The target words that did not match any dictionary words comprised proper nouns and internet slang words. The proper nouns mostly indicated a place, whereas most of the internet slang words could be categorised as euphemistic, vulgar or non-vulgar colloquialisms that are not used in daily verbal discourse.

The output of LIWC became the training dataset for classifier algorithms in Weka to predict the cyberbullying tweets. Complying with the pre-requisites of LIWC guaranteed that individual tweets were not misclassified. The dictionary in LIWC, which is its core for text categorisation, served as a perfect instrument to measure the psychometric properties behind every tweet accurately. This training set was more information rich in comparison to feeding the classifiers with raw data.

#### 4.1.1.1 Significance of psychometric evaluation using LIWC

It is extremely important to note that LIWC did not classify the tweets as cyberbullying or non-cyberbullying. It only categorised the words in the text into their respective word-categories and computed the degree of usage of these word-categories within the text. The class variable, as seen in Table 4, was manually added to every instance of LIWC output based upon the previous supervised cyberbullying identification (SCBI) of tweets described in the section 3.2.3.

As described in section 2.5.1, LIWC generated a unique pattern of 67 word-category values for every instance it categorised. This can be mathematically illustrated in Equation 2.

$$\mathbf{P}(\mathbf{I}_i) = \mathbf{V}(\mathbf{wc}_1), \mathbf{V}(\mathbf{wc}_2), \mathbf{V}(\mathbf{wc}_3) \dots \mathbf{V}(\mathbf{wc}_{67})$$

Equation 2 – Pattern Generated by LIWC for Instance  $i$

Where,  $\mathbf{P}(\mathbf{I}_i)$  is a pattern generated for an Instance ' $i$ ' and  $\mathbf{V}(\mathbf{wc}_1), \mathbf{V}(\mathbf{wc}_2), \mathbf{V}(\mathbf{wc}_3) \dots \mathbf{V}(\mathbf{wc}_{67})$  are the values ranging from zero to 100 derived from Equation 1 established in section 2.5.1.

In this way, LIWC generated patterns for 1313 instances out of which 376 indicated unique individual patterns of cyberbullying tweets. These individual patterns formed a concrete baseline for Weka classifiers to train a detection model by computing the fluctuation of each word-category value for every tweet. As a result, every classifier built an intelligent model by computing the significance of each word-category based on the value that they carried. In turn, the classifiers computed the inter-influence of all word-category values, mutually and reciprocally, in parallel to the prevailing significance of fluctuations in those values for every instance.

As stated in section 3.2.3, bullies tend to be aggressive, hostile and project negativity in written discourse. Psychometric evaluation of tweets aids in comprehending the relationship between word-categories and bully characteristics in a quantifiable way. The evaluation of four main psychometric properties in conversational dialects, namely Linguistic Process, Psychological Process, Personal Concerns and Spoken Categories, differ greatly according to an individual's style of writing. For example, it was observed that 40 word-categories indicating cyberbullying were incremented in instance number 294, compared to just six word-categories incremented in instance number 99. Therefore, it was not recommended to identify the exact word-categories that indicated cyberbullying in tweets. Instead, it was recommended to compute the collective significance of all word-categories and their influence on each other in detecting the likelihood of cyberbullying in tweet occurrences. The following section explains each classifier output model in detail.

## 4.2 Weka Results

The LIWC results output file ('combined results.txt') was loaded into Weka as input for training the classifiers to develop a predictive model to detect cyberbullying on Twitter.

Table 5 shows the classifier outputs based on individual classifier parameter settings.

Classifier	Pre-Processing Attribute Selection				Without Pre-Processing		Class (yes)
	Cross-Validation 10 Folds (Training)		% Split (66-34) (Testing)		Cross Validation 10 folds (Training)	% Split (66-34) (Testing)	
	CFS Subset Evaluator	Info Gain Evaluator	CFS Subset Evaluator	Info Gain Evaluator			
Random Forest	0.978	0.978	0.967	0.951	0.984	<b><u>0.983</u></b>	Precision
	0.96	0.952	0.943	0.943	0.963	<b><u>0.935</u></b>	Recall
SMO	0.982	0.986	0.965	0.974	0.986	0.965	Precision
	0.894	0.91	0.886	0.902	0.912	0.902	Recall
Multilayer Perceptron	0.951	0.964	0.94	0.866	0.963	0.873	Precision
	0.939	0.918	0.894	0.894	0.912	0.894	Recall
J 48 Decision Tree	0.954	0.944	0.932	0.935	0.947	0.935	Precision
	0.931	0.941	0.894	0.935	0.941	0.935	Recall

Table 5 – Weka Classifier Outputs

Table 5 contains the classifier outputs used in this research to develop a predictive model to detect cyberbullying on Twitter. The output for every classifier is stated in terms of precision and recall. Precision value for a classifier can be defined as the ratio of true positive elements with respect to the total selected elements. It calculates how many selected items are relevant. On the other hand, recall can be defined as the ratio of true positive elements with respect to the false negative elements. Therefore, it calculates how many relevant elements are selected by the classifier.

The next section explains the significance of the attribute selection process followed by a section that explains their influence on each classifier and output generated.

## **4.2.1 Attribute selection**

The pre-process tab in Weka allows the user to choose filters for supervised attribute selection. ‘Cfs subset eval’ and ‘Info gain eval’ were the two filters used on the training dataset individually.

### **4.2.1.1 Cfs Subset Eval**

‘Cfs Subset Eval’ calculates the worth of every attribute according to its individual predictive ability. The evaluator selects subsets of attributes with high correlation to the class and low inter-correlation between each other. ‘Cfs Subset Eval’ uses the ‘Best-first’ search method to assign weights to the attributes selected. Using this filter on the training dataset, it selected 20 attributes that had the highest correlation to the class attribute. A screenshot of the Weka attribute selection pane exemplifies this, as shown in Figure 9 in Appendix 3.

### **4.2.1.2 Info Gain Eval**

‘Info Gain Eval’ calculates the worth of every attribute by correlating the information gained with respect to class. This filter ranks the attributes hierarchically, starting from maximum information gain to minimum information gain. ‘Info Gain Eval’ uses the ‘ranker’ search method to evaluate the attributes individually and rank those according to information gained with respect to class attribute. It considers all data attributes, unlike ‘Cfs Subset Eval’. Figure 10 shows first the 22 ranked attributes, followed by Figure 11, Figure 12 and Figure 13, showing attributes ranked from 23–44, 45–66 and 66–67 respectively. These figures can be found in Appendix 3.

## 4.2.2 Weka classifier algorithms

### 4.2.2.1 Random Forest Classifier

The Random Forest classifier predicted the occurrences of cyberbullying tweets for this training dataset with the highest precision rate. According to the parameter settings, 100 trees are randomly generated with similar categorisation rules that consider seven random features for every tree that is generated. Every tree casts a vote for a tree other than itself that has more efficient classification rules. The tree that wins the maximum votes is the classifier output.

However, Random Forest has a down side in terms of comprehending the classifier output. Literature from previous studies confirms that Random Forest classification is treated as a 'Black Box' (Azoff, 1994). This indicates that the classifier output is difficult to interpret because it produces a large number of complex trees based on random feature selection techniques for every individual tree. As a result, it is not feasible to comprehend the exact features selected by the classifier output for the unanimously voted tree. This means that the classifier output is generally based on the feature selection techniques for one specific dataset. In turn, if the classifier is trained on a specific dataset, one cannot be sure if the same feature selection technique for the classifier can work on different datasets. This suggests that the model developed using Random Forest was not expected to generate the same results if deployed on a fresh dataset. The new dataset would have to first be trained using this classifier, followed by effective implementation on the test set.

In this research, the main focus was to detect the cyberbullying events that occur on Twitter. The Random Forest classifier features derived in this research were expected to work across any dataset as long as it had been extracted from Twitter. As Twitter supports only 140 characters per tweet, one can be sure that data retrieved from Twitter in the form of tweets is uniform throughout. Hence, the degree of word usage by users on Twitter is based on a limited word count, with a total character count of 140.

To validate whether the performance of Random Forest classifier was maintained across different datasets, the dataset for this research was randomly split into 66% for training the model and 34% to test the model. It was found that the classifier feature selection based on the training set performed exceptionally well on the test dataset also. The

classifier produced the precision value of 0.98 on the test set, which was the same as the value generated on the training set. Therefore, it can be said that the feature selection of the Random Forest classifier was robust and could be deployed on any other fresh dataset obtained from Twitter.

Attribute selection filters decrease the classifier precision, especially in the test set. It means that tree classifiers work more efficiently if trained on a dataset with more attributes and instances, such as in this dataset. However, it was expected that Random Forest could generate more efficient output given that it was trained on a larger number of instances.

The threshold curve set a benchmark of probability assigned to the true positives and true negatives. The threshold curve was plotted by sorting classifier predictions in descending order of probability values assigned to class category (in this case, 'Yes' and 'No'). The ROC curve values of class 'Yes' and class 'No' were 0.997 respectively. ROC curve generally aids in separating the two class values. Sometimes high ROC curve values lead to misclassification of instances. In spite of high ROC curve values, the classifier does not misclassify the occurrences of cyberbullying tweets. Please refer to Appendix 3 for detailed results for this classifier.

#### **4.2.2.2 Sequential Minimal Optimization (SMO)**

SMO trains the support vector machine (SVM). For the dataset supplied to the SVM classifier in this study, the SVM generated a hyperplane based on the 67 attributes by assigning individual attribute weights, which can be seen in the complete classifier output in Appendix 3. In order to create this hyperplane, SVM created a gap of largest minimum distance to separate multi-dimensional instances with 120,345 kernel evaluations. This suggests that SVM classification rules are extremely reliable, because it developed a precise gap that separated patterns of word-categories indicating cyberbullying versus non-cyberbullying patterns.

The results of SVM in Appendix 3 show the weights assigned to individual attributes. These attribute weights represent the vector coordinates generated by SVM in order to develop the hyperplane that separates two classes. These weights are orthogonal to the hyperplane generated. Hence, SVM generated vector coordinates for the 67 attributes relevant to the dataset supplied for training.

These individual vectors generated by SVM gave the direction to the predicted class attribute. In short, if we calculate the dot product of a vector with any point, it will state which side it belongs to with respect to class attribute. A positive dot product signifies that it is associated to non-cyberbullying class (no), whereas a negative dot product signifies affiliation to cyberbullying class (yes).

The SVM classifier generated the predictive model with excellent precision. The precision value on the test dataset was validated as being almost as good as that of the Random Forest classifier. However, unlike the Random Forest classifier feature selection output ambiguity, SVM's classification output is extremely comprehensive. The individual weights assigned for every attribute can be visualised and the nature of the hyperplane separating the two classes is easily graspable.

The attribute selection filters aided in improving the overall classification accuracy, especially for Info Gain Eval. As stated above, Info Gain Eval calculates the worth of every attribute and ranks them based on the amount of valuable information that they carry. With this attribute selection technique in place, it was easy for the SVM classifier to assign individual attribute weights. In turn, it resulted in the classifier generating extremely high precision and recall values, which can be seen in Table 5.

This proves that SVM is a very robust and reliable classifier for multi-dimensional datasets. Although it requires high processing power to generate the output model, it was seen that the model trained can be deployed on smaller test datasets and results in generating high precision and recall values, such as in this study. Please refer to Appendix 3 for detailed results.

#### **4.2.2.3 Multilayer Perceptron**

MLP works on neural networks that train the model on the statistical and mathematical baseline. To train the model on this dataset, MLP tried to learn every individual pattern that led to the final class attributes of 'yes' and 'no'. In order to learn every pattern associated with cyberbullying, the MLP initially assigned individual synaptic weights to every attribute with a unique threshold value. As stated previously, MLP uses the back propagation technique to learn and develop the model.

For every back propagation, MLP generated a hidden layer and followed the same

technique of assigning new individual attribute synaptic weights along with a new threshold value. In this case, MLP generated 36 hidden input layers based on the learning process that requires MLP to compute the probability of the values occurring in the next learning instance.

This shows the correlation between all hidden input layers and the link generated by MLP between all the attribute values for every input instance. These individual values can be seen in Appendix 3.

Due to a multivariate dataset, MLP was found to be a very time-consuming classifier for the training set supplied in this study. Table 5 shows that MLP trained the model with a high precision rate; however it failed in terms of the recall rates generated. This made the classification techniques of MLP a bad choice for this dataset. Moreover, if deployed on the test set, the recall value was seen to deteriorate even further.

However, the predictive precision ability elevated if the training set was processed using attribute selection filters. The results show that MLP can be made less expensive and less time-consuming if data is processed using attribute selection filters. However, in terms of relevance of the elements selected after classification, MLP's recall value did not elevate in any desirable way for this study, again making it a poor choice for developing a model to detect cyberbullying on Twitter.

Furthermore, classifier error was minimised by introducing 500 epochs. One epoch corresponds to 1313 training instances; therefore, the classifier had to go through 656,500 (1313 times 500) individual training trials. It was expected that at the end of every epoch, the error would be minimised to a certain extent. Figure 14 shows the Weka GUI pane for MLP epoch calculations. As discussed in section 2.5.4, the number of epochs was set to 500. Figure 15 (Appendix 3) shows the error value per epoch as 0.005967 at the end of 340 epochs. In comparison, the error value per epoch at the end of 500 epochs reduced to 0.0059432, as seen in Figure 16 (Appendix 3). This suggests that MLP modelling becomes more efficient if it is fed with a greater number of instances. Please refer to Appendix 3 for detailed results.

#### 4.2.2.4 J48 Decision Tree

A decision tree is a binary classifier that generates an output model based on certain categorisation rules. Decision trees build a model based on the attribute values of only those input variables that contain the most information. The J48 decision tree classifier trained on the dataset provided for this study generated a tree that started with a single node called '*shehe*' and continued branching through the input variables. A subset of 14 other input variables branched under this node was generated using this classifier. Therefore, the categorisation rules generated by the tree to predict the classification were based on a subset of 15 input variables as seen in the Figure 18. The decision tree is self-explanatory as it is built on binary classification rules. The size of the decision tree in this study was 31, which means the tree was divided into 31 nodes and it contained 16 leaves.

A pruned tree in comparison to an unpruned tree generates more efficient output with higher precision and smaller subsets of variables. Appendix 3 contains detailed results for this classifier.

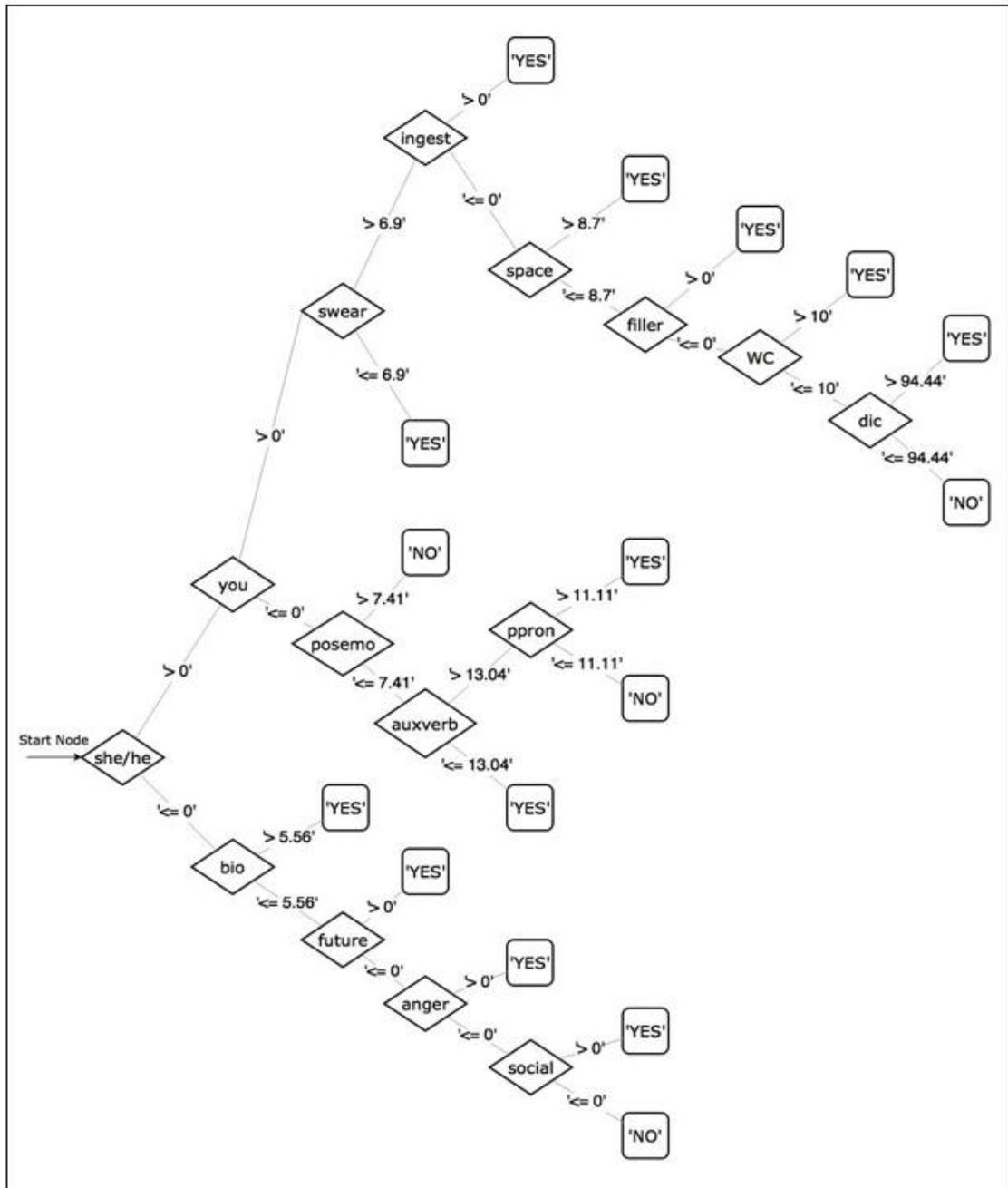


Figure 9 - J48 Decision Tree

#### 4.2.2.5 Cost-Sensitive Evaluation

A common limitation observed in this study was the small size of the training set for classifier learning methods. In spite of a small dataset, the predictive ability of the classifiers was extremely efficient. In this study, the focus lay on predicting the true positives as accurately as possible. To a certain extent, it would be fine for a non-cyberbullying tweet to be misclassified as a cyberbullying tweet but a cyberbullying tweet could not be misclassified as a non-cyberbullying tweet.

To eliminate this, the classifiers in Weka were trained using the cost-sensitive analysis technique. The cost matrix was set to make classification ten times more sensitive for predicting true positives. This reduced the overall classifier accuracy, but the Kappa Statistic (true positive rate) was significantly improved. Table 6 shows the classifier output comparison for cost-sensitive analysis performed on the Random Forest classifier. For detailed results, please refer to Appendix 3.

Cost-sensitive analysis			Without cost-sensitive analysis	
a	b	← classified as →	a	b
<b>368</b>	8	a = yes	<b>362</b>	14
37	900	b = no	6	931

Table 6 – Cost-matrix Comparison

### 4.3 Weka Experimenter

The Weka Experimenter calculates the degree to which each classifier output deviates in comparison to the desired output. The experimenter runs the classifier algorithms on the same dataset with a cross-validation of ten folds and ten iterations respectively. Hence, the experimenter conducts 400 experiments (4 Algorithms \* 10 Iterations \* 10-fold cross-validation:  $4*10*10 = 400$ ) and calculates the standard deviation of classifier output to the desired output. The Random Forest classifier outperformed the rest of the classifiers with the least standard deviation of 1.25. It was followed by SVM, Decision Tree and MLP classifiers, with standard deviations of 1.45, 1.47 and 1.64 respectively. This proves that binary classifiers such as Random Forest outperform multiclass classifiers like Multilayer Perceptron. For detailed results for the Weka Experimenter, please refer to Appendix 3.

## *5 Conclusion*

This research examined the effectiveness of using a multi-dimensional training dataset on machine learning classifiers to predict cyberbullying tweets on Twitter. The multi-dimensional dataset was created based on the pragmatics of language. LIWC analysed tweets individually in segments and categorised words from each tweet into 67 psychometric categories. However, LIWC required the text file to be cleaned according to its rules. This enabled the tool to generate an output with maximum accuracy. However, the data archived from TAGS was not compatible with LIWC. The data pre-processing core of this system was responsible for cleansing the dataset as per LIWC pre-requisites. The LIWC output became the training set for the Weka classifiers. The validity of the output models generated by the Weka classifiers depended on how correctly the dataset was labelled. Moreover, to label it correctly, the LIWC classification had to be precise and error free. In short, the efficiency of the data pre-processing core determined the validity and efficiency of the final predictive models generated by the Weka classifiers.

The predictive models generated using Weka show that binary classifiers outperformed multiclass classifiers. The predictive model trained using the Random Forest classifier yielded 98.5% accuracy with a precision rate of 0.983 and a recall rate of 0.935. It is seen that the predictive ability of all the classifiers deteriorated slightly when implemented on the testing dataset. However, the true positive rate could be improved by applying cost-sensitive analysis to the classifiers. In addition, the F-measures of all the classifiers except the MLP classifier on the test dataset were near 0.95. This indicates that the classifier algorithms had a high precision rate. MLP's precision rate fell to 0.88 on the test dataset, making it a poor choice for datasets with fewer instances.

The dictionary containing 67 psychometric categories and weighting based on Category Frequency Inverse Word Count (CF-IWC) formed the baseline of this study, where, CF-IWC is the same as  $V(wcn)$ , which is explained in the literature review section 2.5. This baseline provided a very information-rich text classification for tweets, which helped to analyse the behaviour patterns of cyberbullying.

This research identified the gap of ineffective text categorisation techniques in previous works related to the detection of cyberbullying. Psychometric evaluation assisted in understanding the degree of word-usage by different people in cyberbullying events. In this case, the tweets were categorised into 67 psychometric properties of written dialects,

which in turn aided in the effective classification of conversational data from Twitter based on the pragmatics of language. It was seen and highlighted in the study that there is no fixed pattern of word usage when people try to bully victims on the internet. However, it was possible to correlate the inter-influence of every word category of text that leads to cyberbullying.

As the text data was converted to a numeric relational dataset after psychometric evaluation, machine learning techniques using Random Forest, Support Vector Machine, Multilayer Perceptron and J48 decision tree classifiers were used to develop predictive models to detect cyberbullying on Twitter. In spite of being tagged as a '*black box*' when tested on the test dataset, Random Forest's feature selection techniques worked best in comparison to other classifiers.

The performance of SVM in training the predictive model was almost on par with Random Forest. But SVM's classification rules in comparison to Random Forest were extremely comprehensive, which made it a good alternative choice to develop the predictive model.

## *6 Scope for Future Research*

People from different communities have different writing styles. For example, individuals from two different communities separated geographically have a unique writing style, which is prevalent in their own community. On an abstract level people from different countries in the world use different writing styles to express the same idea. It also means that the writing styles of individuals is determined by their society. Hence, the pragmatics of cyberbullying language originating from different countries differs slightly from each other.

Additional information about the geographical origin of a tweet could be provided to the existing system architecture. By introducing instances with spatial information, identifying the unique cyberbullying pattern for every country is achievable. As explained above, the writing style of an individual is determined by their society and its collective behaviour. A support vector machine classifier could be used to separate instances from different countries and their uniqueness could be calculated.

In addition, by introducing a spatio-temporal variable along with the use of hashtags to analyse trending topics, a predictive model could be trained to analyse certain events occurring in that spatial boundary which could trigger a potential cyberbullying threat to an individual or a group of individuals. The hashtag variable is subject to change with respect to time and events.

Furthermore, this predictive model will have the ability to generate higher accuracy in varying conditions of the size of dataset.

### **Example by comparison**

Cyberbullying is prevalent in countries like New Zealand and India. New Zealand is not a densely populated country, whereas India is one of the most populous countries in the world. As a result, New Zealand has less internet users compared to India.

Therefore, the occurrences of cyberbullying tweets on Twitter in New Zealand are extremely low in comparison to India. In addition, based on the spatial classification, the cyberbullying patterns are unique in these two countries.

The predictive model to detect cyberbullying patterns in New Zealand would be based upon fewer instances in conjunction with the local writing styles of New Zealanders. In

contrast to New Zealand, the Indian model could be built upon a larger number of instances in conjunction with the writing styles of the Indian community.

The validity and accuracy of the predictive models to detect cyberbullying on Twitter in this case is primarily based on the correct psychometric categorisation of text.

It is observed that different countries follow different cyber laws. It is therefore possible to build a data repository containing the cyber laws of different nations. If the predictive model detected a cyberbullying threat, it could be matched to the degree to which the law was breached. The authority could then be notified about the breach of law so that further appropriate actions could be taken.

## *7 References*

- Azoff, E. M. (1994). *Neural network time series forecasting of financial markets*. New York, NY: John Wiley & Sons, Inc.
- Bauman, S., Toomey, R. B., & Walker, J. L. (2013). Associations among bullying, cyberbullying, and suicide in high school students. *Journal of adolescence*, 36(2), 341-350.
- Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond*. Texas, TX: Guilford Press.
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.
- Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., & Scuse, D. (2013). *WEKA Manual for Version 3-7-8*: January.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Browne, M. W. (2000). Psychometrics. *Journal of the American Statistical Association*, 95(450), 661-665.
- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*(2), 15-21.
- Chamorro-Premuzic, T., & Furnham, A. (2014). *Personality and intellectual competence*. Mahwah, NJ: Psychology Press.
- Cyberbullying. (2014). Retrieved from <http://kidshealth.org/parent/positive/talk/cyberbullying.html>
- Cyberbullying Statistics. (2014). Bullying facts, Bullying statistics. Retrieved from <http://nobullying.com/cyber-bullying-statistics-2014/>
- Dadvar, M., & De Jong, F. (2012). Cyberbullying detection: a step toward a safer Internet yard. Paper presented at the Proceedings of the 21st International Conference Companion on World Wide Web at New York, NY.

- Dinakar, K., Reichart, R., & Lieberman, H. (2011). Modeling the detection of textual cyberbullying. Paper presented at the The Social Mobile Web at Boston, MA.
- Dooley, J. J., Pyżalski, J., & Cross, D. (2009). Cyberbullying versus face-to-face bullying: A theoretical and conceptual review. *Zeitschrift für Psychologie/Journal of Psychology*, 217(4), 182-188.
- Du, K.-L., & Swamy, M. (2014). Multilayer perceptrons: Architecture and error backpropagation. In *Neural Networks and Statistical Learning* (pp. 83-126), New Delhi, India: Springer.
- Dufour, K. (2012, October 16). Amanda Todd case highlights issue of online bullying, *The Telegraph*. Retrieved from <http://www.telegraph.co.uk/news/worldnews/northamerica/usa/9612030/Amanda-Todd-case-highlights-issue-of-online-bullying.html>
- Finley, L. L. (2014). *School Violence: A Reference Handbook*. Santa Barbara, CA: ABC Clio, Inc.
- Fitzgerald, B. (2012). Bullying on Twitter: Researchers find 15,000 bully-related tweets sent daily (study). Retrieved from [http://www.huffingtonpost.com/2012/08/02/bullying-on-twitter\\_n\\_1732952.html](http://www.huffingtonpost.com/2012/08/02/bullying-on-twitter_n_1732952.html)
- Graham, M., & Haarstad, H. (2014). Transparency and development: Ethical consumption through Web 2.0 and the internet of things. *Open Development: Networked Innovations in International Development*, 79.
- Heirman, W., & Walrave, M. (2008). Assessing concerns and issues about the mediation of technology in cyberbullying. *Cyberpsychology: journal of psychosocial research on cyberspace*, 2(2), 1-12.
- ITU. (2014). *Measuring the Information Society Report*. Geneva, Switzerland: International Telecommunication Union.
- ITU. (2015). *Statistics*. Retrieved from <https://www.itu.int>
- Izard, C. E. (2013). *Human emotions*. New York, NY: Springer Science & Business Media.

Kelly, R. (2009). *Twitter Study-August 2009*. San Antonio, TX: Pear Analytics.

Kline, P. (2013). *Handbook of psychological testing*. Exeter, England: Routledge.

Kontostathis, A. (2009, May). *Chatcoder: Toward the tracking and categorization of internet predators*. Paper presented at the Proc. Text Mining Workshop 2009 held in conjunction with the Ninth International Conference on Data Mining (SDM 2009). Sparks, NV. May 2009.

Kowalski, R. M., Limber, S. P., & Agatston, P. W. (2012). *Cyberbullying: Bullying in the digital age*. West Sussex, UK: John Wiley & Sons.

Larose, D. T., & Larose, C. D. (2015). *Data mining and predictive analytics*. Retrieved from <http://AUT.ebib.com.au/patron/FullRecord.aspx?p=1895687>

Litwiller, B. J., & Brausch, A. M. (2013). *Cyber bullying and physical bullying in adolescent suicide: The role of violent behavior and substance use*. *Journal of youth and adolescence*, 42(5), 675-684.

Luxton, D. D., June, J. D., & Fairall, J. M. (2012). *Social media and suicide: A public health perspective*. *American Journal of Public Health*, 102(S2), S195-S200.

Maimon, O., & Rokach, L. (2008). *Data mining with decision trees: theory and applications: USA*: World Scientific Publishing.

Meyer, D., & Wien, F. T. (2014). *Support vector machines. The Interface to libsvm in package e1071*. Zhongli, Taiwan.

Miller, W. R., & Rollnick, S. (2012). *Motivational interviewing: Helping people change*. New York, NY: Guilford Press.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The development and psychometric properties of LIWC2007*.

Reynolds, K., Kontostathis, A., & Edwards, L. (2011). *Using machine learning to detect cyberbullying*. Paper presented at the Machine Learning and Applications and Workshops (ICMLA), 10th International Conference, Honolulu, HI.

Riebel, J., Jaeger, R. S., & Fischer, U. C. (2009). *Cyberbullying in Germany—an exploration of prevalence, overlapping with real life bullying and coping strategies*.

Psychology Science Quarterly, 51(3), 298-314, Germany.

Salmivalli, C., & Peets, K. (2009). Bullies, victims, and bully-victim relationships in middle childhood and early adolescence. *Handbook of peer interactions, relationships, and groups*, 322-340. New York, NY.

Schneider, S. K., O'Donnell, L., Stueve, A., & Coulter, R. W. (2012). Cyberbullying, school bullying, and psychological distress: A regional census of high school students. *American Journal of Public Health*, 102(1), 171-177.

Schnoebelen, T. (2012). Do you smile with your nose? Stylistic variation in Twitter emoticons. *University of Pennsylvania Working Papers in Linguistics*, 18(2), 14.

Shapka, J. (2012). Cyberbullying and bullying are not the same: UBC research. Retrieved from <http://news.ubc.ca/2012/04/13/cyberbullying-and-bullying-are-not-the-same-ubc-research/>

Shuen, A. (2008). *Web 2.0: A Strategy guide: Business thinking and strategies behind successful Web 2.0 implementations* (1 ed.). Sebastopol, Canada: O'Reilly Media.

Siegel, E. (2013). *Predictive Analytics : The power to predict who will click, buy, lie, or die*. Retrieved from <http://AUT.ebib.com.au/patron/FullRecord.aspx?p=1124085>

Simanjuntak, D. A., Ipung, H. P., Lim, C., & Nugroho, A. S. (2010). Text classification techniques used to facilitate cyber terrorism investigation. Paper presented at the *Advances in Computing, Control and Telecommunication Technologies (ACT), Second International Conference*, Jakarta, Indonesia.

Sleglova, V., & Cerna, A. (2011). Cyberbullying in adolescent victims: Perception and coping. *Cyberpsychology: journal of psychosocial research on cyberspace*, 5(2).

Slonje, R., Smith, P. K., & Frisén, A. (2013). The nature of cyberbullying, and strategies for prevention. *Computers in Human Behavior*, 29(1), 26-32.

Smets, K., Goethals, B., & Verdonk, B. (2008). Automatic vandalism detection in Wikipedia: Towards a machine learning approach. Paper presented at the *AAAI Workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy*, Palo Alto, CA.

- Tan, P.-N., Chen, F., & Jain, A. (2010). Information assurance: Detection of web spam attacks in social media. Paper presented at the Proceedings of Army Science Conference, Orlando, Florida.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, 29(1), 24-54.
- Traditional bullying vs. cyberbullying. (2011, August 8). Cyberbullying. Retrieved from <https://sites.google.com/site/cyberbullyingawareness/traditional-bullying-vs-cyberbullying>
- Vaillancourt, T., McDougall, P., Hymel, S., Krygsman, A., Miller, J., Stiver, K., & Davis, C. (2008). Bullying: Are researchers and children/youth talking about the same thing? *International Journal of Behavioral Development*, 32(6), 486-495.
- Van Geel, M., Vedder, P., & Tanilon, J. (2014). Relationship between peer victimization, cyberbullying, and suicide in children and adolescents: A meta-analysis. *JAMA pediatrics*, 168(5), 435-442.
- Xu, J.-M., Zhu, X., & Bellmore, A. (2012). Fast learning for sentiment analysis on bullying. Paper presented at the Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, Beijing, China.
- Yin, D., Xue, Z., Hong, L., Davison, B. D., Kontostathis, A., & Edwards, L. (2009). Detection of harassment on Web 2.0. *Proceedings of the Content Analysis in the WEB*, 2, 1-7.
- Zhang, L. (2014). Review of Handbook of Automated Essay Evaluation: Current Applications and New Directions. *Language, Learning & Technology*, 18(2), 65.

# Appendix 1

List of TAGS column names that can be included in Archive sheet

id	user_listed_count
id_str	user_created_at
text	user_favourites_count
source	user_utc_offset
truncated	user_time_zone
metadata	user_geo_enabled
created_at	user_verified
in_reply_to_status_id	user_statuses_count
in_reply_to_status_id_str	user_lang
in_reply_to_user_id	user_contributors_enabled
in_reply_to_user_id_str	user_is_translator
in_reply_to_screen_name	user_is_translation_enabled
geo	user_profile_background_color
coordinates	user_profile_background_image_url
place	user_profile_background_image_url_https
contributors	user_profile_background_tile
retweet_count	user_profile_image_url
favorite_count	user_profile_image_url_https
entities	user_profile_link_color
favorited	user_profile_sidebar_border_color
retweeted	user_profile_sidebar_fill_color
possibly_sensitive	user_profile_text_color
lang	user_profile_use_background_image
user_id	user_default_profile
user_id_str	user_default_profile_image
user_name	user_following
user_screen_name	user_follow_request_sent
user_location	user_notifications
user_profile_location	from_user
user_description	from_user_id_str
user_url	profile_image_url
user_protected	status_url
user_followers_count	time
user_friends_count	entities_str

## Set-up instructions for TAGS archiving tool

1. To start using TAGS click ‘Get TAGS’<sup>9</sup> and File > Make a copy (make sure you are logged in to your Google account to do this).
2. After your copy has been made, open TAGS > Setup Twitter Access.
3. A pop up window appears requesting your Twitter application credentials for authorization purposes.
4. Follow the URL<sup>10</sup> to create a custom Twitter application.
5. The mandatory fields for application details include the name of your application, the description of your application and your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens. In the final field of Call back URL copy and paste the URL<sup>11</sup> from the pop up window that appeared previously in the copy of TAGS. Click ‘Agree’ on the check box next to the ‘Developer rules of the road’ at the end to activate your Twitter application.
6. The Twitter application then provides the user with two keys; Consumer Key (API Key) and Consumer Secret (API Secret), for authorization purposes. Copy and paste the above keys in the respective fields on the pop up window to get successful authorization to TAGS archiving tool.
7. The application is now ready to use. Click ‘Enable custom menu’ to activate your application. A pop up screen requesting authentication appears; continue to complete the process.

---

<sup>9</sup> <https://tags.hawksey.info/get-tags/>

<sup>10</sup> <http://dev.twitter.com/apps/new>

<sup>11</sup> <https://spreadsheets.google.com/macros>





```

Application.ScreenUpdating = True
Application.DisplayAlerts = True
Exit Sub

```

ErrorHandler:

```

MsgBox "Sorry, something unexpected when wrong." & vbNewLine & _
    "Error " & Err.Number & ": " & Err.Description, vbCritical
End Sub

```

**c) Remove carriage returns:**

```

Sub RemoveCarriageReturns()
    Dim MyRange As Range
    Application.ScreenUpdating = False
    Application.Calculation = xlCalculationManual

    For Each MyRange In ActiveSheet.UsedRange
        If 0 < InStr(MyRange, Chr(10)) Then
            MyRange = Replace(MyRange, Chr(10), "")
        End If
    Next

    Application.ScreenUpdating = True
    Application.Calculation = xlCalculationAutomatic
End Sub

```

**4) Insert blank row in between every row of data:**

```

Sub Insert_Blank_Rows()

    'Select last row in worksheet.
    Selection.end(xldown).select

    Do Until ActiveCell.row = 1
        'Insert blank row.
        ActiveCell.EntireRow.Insert shift := xldown
        'Move up one row.
        ActiveCell.Offset(-1,0).Select
    Loop

End Sub

```

# Appendix 3

## Tweets used for developing the predictive model

There once was a cunt named Beau Who s eyes were so pink he looked like a freak show  
 How fuckin dare you you ungrateful bitch I should have let you die I fuckin fed you when you needed food  
 you tryna play me Got You  
 Drill Her Tight Black Pussy black drill pussy tight  
 No chance that fat fuck will live past 75  
 Dutch PM says jihadis should die in Syria rather than come home  
 aaron sorry for raping you me my locals OK BUT your A whore SO HOW IS IT EVEN RAPE HAHAAAAH  
 whore SLUT WHORE I am livid  
 I heard you beat your girlfriend keem oh and nice fight with pittsburgh you should make a drama alert on  
 these  
 Everybody knows from your head to your toes you re a big fat stinking cunt  
 I bet you'd eat keemstars ass for a spot in eRa nerd cunt  
 Fat pussy good night team karma  
 good Fucking pig slut Oink  
 I feel like fat girls should except the fact their fact and just be like I'm fat and i could easily kill you skinny  
 bitches  
 Fat girls go kill yourself Go cut yourself in the corner emo  
 Fat cunt  
 My friend in class saw a pic of you and asked me Who is this gay cunt lol  
 why are you suck a loser  
 You going to lick this pussy just like you should  
 did you get caught up watching gay porn  
 emos suck dicks  
 implying you don't want to suck that big fat bull dick  
 if these fucking pigs really believe in killing then they should kill themselves first hateislam islam sucks  
 fight me nerd  
 kill yourself nerd  
 If that CUNT is talking about me I kick the shit out the little fat wanker must be HE SHIT  
 Ate this fat bitch pussy once and it tasted like cinnamon buns  
 FUCKING GAY COCK SUCKER GO SUCK your OWN DICK  
 i should watch kill your darlings tonight  
 Girls that suck dick without being asked are goddesses and should not be fucked with  
 If you crash you die because it ain t enough seat belts for your fat as  
 Pig's blood for a pig Okay Assholes If you don't die by the end of this I'm gonna be really fucking  
 disappointed  
 If you are a man and you give me a weak ass handshake then ima assume that you are a gay pussy faggot  
 Die you are ugly okay so what should I do now  
 We'll kill you white pig I will fucking kill you you black bitch Wits student  
 I should see love some how karkoub you suck  
 We Should Kill His Whole Family  
 caught this nigger tryna touch my butt wtf some how Hatfield and gay people  
 Fat Brunette Bitch With Butterfly  
 These young whores kill me swearing up and down that they gay when in reality they just fuckin confused  
 no don't you have a fat pussy to attend to  
 If only Darren Wilson s pussy ass could be behind bars like this pig Scandal  
 I'm tired of you but fr pussy is pussy oh well what if she too fat to ride Da dick  
 I hate ugly and fat people They deserve to die  
 bitch ass nigger I hate you I really wanna fight You gay asf for sending me these emojis in my DMs And  
 die lowk  
 They should all die All of them Everything annihilated  
 You should die all together  
 pussy should taste like pussy no amount of pineapples is gonna change that drink water eat healthy and  
 your pussy should  
 I wonder if you can see this you fat thick acne ridden bitch hohoho slut  
 Kimi is a true whore

aye little ginger fat cunt i love your videos  
 or maybe I should kill them WITH FIRE  
 they should all die an horrible death  
 you are jealous from everything I do your condition is very serious you should go and see a doctor or you  
 will die because of me  
 I suck at drawing but that pig came out nice I'm happy as fuck  
 You should Suck My Dick How about that one  
 she should not be scared to kiss you after you eat her pussy  
 Fuckin Gay Cunt Swallow my Dick Faggot  
 As an indigenous Irish citizen I will not be told where to live by a wannabe Brit  
 Muslims Take Christian Children And Behead Them Rape And Kill Their Mothers And Hang Their Fat  
 via  
 you should die for that  
 Boy you suck You thanks Boy your ugly You yup Boy Justin Bieber is gay You roll up sleeves WANT TO  
 GO PUNK  
 I think you should get to say I will find you And I will kill you  
 baby you the only one on here being gay We talking about DICK And you putting pussy o  
 leave my mentions you fat cunt babe don't call her a fat cunt when you don't know her  
 I ll eat your pussy like a fat nigger eat the last piece of chicken  
 Oysa we should kill them all degil mi  
**JUST DO IT DO IT LICK IT GOOD SUCK THIS P Y JUST LIKE YOU SHOULD**  
 I'm not saying she s a slut but her vagina should be in the NFL Hall of Fame for greatest wide receiver  
 lmao idk you should know that feeling since it seems that you suck dick  
 Fat cunt  
 I wanna hurt my sister so bad she can be a total fuck face cunt sucking dick ass hypocritical shit hole freak  
 Craig Foster just talks to hear his own voice He can suck a fat one too SydneyVersusEverybody  
 jake ryan looks like he d be the kind of guy to let a gay kid suck his dick under the bleachers  
 aaron sorry for raping you me my locals OK BUT your A whore SO HOW IS IT EVEN RAPE HAHAHAAH  
 whore SLUT WHORE  
 If you litter suck a big fat one Seriously  
 My ceiling fan has 3 settings 1 Very slow 2 Slow 3 I'm about to detach from the ceiling and kill you in a  
 freak ceiling fan  
 Scholes should stfu the ginger cunt  
 Don t be afraid on my big black strapon dildo  
 I should thank you But it would be more fun to kill you  
 Suck it WHORE haha  
 stop being so fucking gay you pussy  
 Japanese DarkSoulsMindEye Hey jblackmel suck my dick loser  
 Fat Athabaskan pussy  
 wow what a n00b you should fight me in the wildy and then throw your computer out the window  
 what a gay cunt  
 that's a fat pussy  
 Fat girl pussy taste like a cheese puff  
 and that s when he caught gay  
 all because his wife is a fat cunt that is disabled when she isn't pt  
 You should be able to fight soon right  
 Ungrateful idiots should just die  
**HAPPY BIRTHDAY SLUT** Hope you have a gay day xoxoxo  
 sara threatens to fight me all the time but she never has is she a poser found out at 6 kick my ass sara  
 what a fat cunt  
 Awkward when ya only just realise that your ex is a fat cunt wannabe gangster Please die  
 want stick dick in a nigger ass amp then a bitch pussy go head with that gay shit my nigger  
**GAY PEOPLES are SUPER DOOPER ANNOYING TOO MANY SIDE COMMENTS GO to hell Well**  
 your going to hell by the time you DIE  
 yeah bitch i dont want your fat stomach on my mention i am not cameron i am Agung fight her  
 The whiter the whore the harder a black master will pound her cunt  
 Wink pussy  
 You should thank me for the honour of letting you watch me fill your hot wife tight cunt snow bunny bbc  
 owned  
 Why should I have to tolerate these bastards Why can I not kill them all with fire or blunt knives and spoons  
 this is my freak account I am somebody you know amp I have to tell you that I'd love to nut in your pussy  
 raw balls deep Repeatedly  
 hurry up and die terry you won t be missed you fat fuck

Fat girl pussy taste like a cheese puff  
 happy birthday gay boy you still suck  
 suck my fat cock  
 If your happy or think it's funny to be a home wrecking whore your the fuck weird and proly have a roast  
 beef pussy  
 happy birthday gay boy you still suck  
 damn pussy fat then a mother fucker  
 check out this x Ray type thing You and should do this to show how deep his cock is in your pussy  
 She a keeper i know her type she got that fat warm pussy guinnys  
 Broke People Should Just Kill Themselves  
 Fat girls will take one good pic and use it for 5 years word to Venny let us live  
 neck yourself you fat cunt I hope you test positive for aids you legit need to fucking hang yourself NO ONE  
 WOULD MISS YOU  
 aww thank you slut my big fat head hahahahha  
 because you are gay  
 No good identity stealing whore Pussy selling slut man stealing pussy eating whore ThadijahShaw  
 PETA can suck a fat one I believe I first saw activism with Defenders of Wildlife  
 Annoying ugly fat slut ahahahah  
 get a fucking grip you wog any proof I scam No so go suck on your dads chode you fat wanker  
 You should be in bed you slut  
 its not gay if you suck my dick  
 much pussy such gay  
 Calling a girl a slut whore whore bitch doesn t make you any more of a man  
 Fat pussy  
 ok can go suck a fat dick  
 Gay Activist to Suck My Dick  
 Fuck Mr Nimmo Fuck Mr Kim Both of you should suck a dick  
 dirty slut mmmmm turn around whore  
 Fat thin either way will always be a cunt via  
 You re a loser So go ahead and live your loser life make a bunch of loser friends Then together you can  
 lick each other s loser wounds  
 Suck it loser PrussiaBot  
 Dress like a Slut Get fucked liked a Whore I am a bareback cum whore BBBH  
 it damm go suck some bryce papinbrook cock or maybe demarco oh maybe paul you gay as fuck your  
 Bald fat pussy  
 BAEK IS GAY FIGHT ME  
 He mocked her talking by doing a simple hand movement Blah blah blah crust shut yer whore mouth and  
 let s fight already He  
 About what That your dumb dog should get AIDS and die  
 you should all die  
 oomf pussy smell like burnt rubber goodnight shut whore  
 I'LL FIGHT YOU WHORE  
 you should kill yourself dad  
 Only bitches with good pussy should be allowed to act crazy  
 That bitch should die in her sleep for lying  
 this was me I texted you this haha sucks to suck huh loser  
 are you a cunt too Or just a whore lol  
 As long as you have a pussy you should never be broke Thats what my pops always told me A women  
 should never be broke  
 Sjw coward Next thing you ll do is suck whale dick  
 You have literally ruined bashurs life You should seriously fucking kill yourself you re the teal child  
 molester  
 You all really hype up Bowser because she apparently got a fat ass meanwhile her pussy come with the  
 stomach flu  
 cunt hangs out with weird people everyone teased the fat cunt  
 kill the pig cut her throat bash her in  
 why you dont fuck with me noww fat pussy  
 Shut the fuck up stop pretending you re from the fucking hood because you live up pant and listen to fucking  
 wannabe ra  
 Happy birthday have a lovely day you slut love smella X thank you whore bag  
 Suck a dick suck a dick suck a dick suck a motherfucking dick suck a dick suck a dick suck a big fat dick  
 He can suck three fat dicks at once and choke I want my fucking money  
 you should die now Haru

Fucking pissed stupid ass dj can suck a fat one  
 Bill DeMott can suck my Jordanian zibby Get colon cancer you fat fuck  
 maybe you should fight a little harder  
 big fat pussy MUFASA  
 I can't decide whether you should live or die  
 I don't want to live in a world without elephants Why couldn t the fat chicks have gone extinct instead  
 we should fight him for being so gay  
 Everyone should take my warning very serious I will kill you behind what s mines  
 faggot gay ass cunt dick shitter if you ask me  
 You eat more pussy den me Girl you should've been uh nigger  
 and i said hey HEY what a wonderful kind of day where we can learn to all be gay and suck a dong with  
 each other  
**YOU SHOULD PROBABLY KILL YOURSELF**  
 If I should die I d tell Big they re still hearin his songs run into Pac ask him where we went wrong Jay Z  
 gay af suck his d or die  
 Well that s horrific this couple should kill themselves  
 Jackson That mutt should die  
 They should all die All of them Everything annihilated  
 Is this bitch asking to die **CALL ME A WHORE AND FAT ONE MORE FUCKING TIME I DARE YOU**  
 hey fat fuck did you seperate your shoulder when you tried to fire a gun is that why you hate guns pussy  
 Big fat pussy MUFASA  
 thanks you fat ginger cunt  
 hey fat fuck did you seperate your shoulder when you tried to fire a gun is that why you hate guns pussy  
 You shouldn t be pissed at me for my 3 inch dick you should be pissed at the big ass pussy you got  
**SWAY your VIDEOS SUCK SO MUCH SCARCE FANBOY your FUCKING UGLY AND your MOM SHOULD DIE**  
 hacked anyone s phone lately you fat cunt  
 you are not capable of any thing Hustle your way to your grave Cunt Pathetic loser  
 you can't laugh you fat cunt  
 I want handy to beat up mendecees assistant but she can't so kimbella should  
 you suck fat chicken cock  
 Gay af you werent saying that when you were eating my pussy last weekend  
 Still talking shit after I beat that ass sore loser  
 If jaden don't beat this alexis ting he s gay  
 Fuck off you fat tree swinging Banana eating cunt  
 whats wrong with slut shaming you should be ashamed to be a slut  
 your a gay cunt Mikee  
 get the hint he doesn t wanna call you fat cunt  
 Big fat cocky mek di pussy talk up  
 you don't look scared but you should be afraid  
**There s NO WAY In Tha World Any Of You Bitches Should Be Broke Cause Yall Sittin On AT LEAST**  
**60 Worth Of Pussy**  
 Oh for sure he is fat as fuck but I tend not to use it as an insult I prefer just cunt  
 This hentai whore has a cock and a pussy so she can fuck and get fucked at the same time  
 fat bitch tried fighting me in Walmart like ill gladly lay your ass out any day of the week cunt  
**Young Pretty Indian Whore Loves To Suck And Fuck**  
 you are one fat smelly ugly cunt  
 My sissy slut has gone on her 1st meet to earn me cash tonight Hope she enjoys cock deep in her throat for  
 the 1st time Cu  
**EACH AND EVERY SINGLE ONE OF YOU SHOULD DIE THE MOST HORRIBLE DEATHS EVER**  
**BECAUSE NONE OF YOU DESERVE TO LIVE**  
 And Pablo Sandoval is a fat cunt Deal with it  
 She a slut she a whore she a freak Ain t got a job but everything on fleek  
**YO PUSSY GAY**  
 suck it nerd  
 Her pussy should ve came wit a guess list  
 Im sorry that your girlfriend is an ocean pussy whore  
 Like Some are gay because they didn't get pussy early  
 I should just beat the shit out of them then lmao  
 Even a blind eye couldn t miss this fat cunt  
 how about you go suck a fat one  
 Isn t he just A right proper cunt Fat too  
 we should of kept score to see how bad i actually beat you

Good pussy Pretty face and your ass fat but your a whore Who let anybody hit it It It It  
 you re just the right amount of fat and i ll definitely treat you like the slut you are  
 DIE you FUCKING TRANNY TRICK TRAP WHORE DIE  
 I should just kill  
 Believe it or not this is a can of beer shoved up my cunt Classy Manchester whore  
 KILL URSELF FAT KKKUNT EVIL SHILL WHORE DIE you FUCKKKEN SLUT DIE  
 That s how a sluts pussy should be whipped That s how a sluts pussy should be whipped  
 suck my dickkk gay bitch  
 Big fat cocky mek di pussy talk up  
 She is a bitch a slut and a whore Her body was used but the users paid no attention to her and she was free  
 to find out why  
 She a slut she a whore she a freak aint got no job but everything on fleek  
 wants us to put a cock in each of my holes at 1 time  
 A fat ass doesn t always mean you got good pussy  
 Why are you so fat Hahaha  
 Ite pussy lol I should beat your ass  
 Yes she does You should see her pussy Immaculate  
 SUCK YOUR NANS DRY CUNT YOU FAT GINGER VIRGIN PEDOS  
 slut puppy cunt whore dog bitch  
 Should I knock his noodles out or should I let him live  
 suk your dead nanny and grandad you fat cunt you had 12 fuking men  
 your daughter should go choke on a dick and die  
 Should I beat up your boyfriend  
 pussy nigger wanna size me up i shall snap your fucking neck  
 suck your fucking mum you fat cunt  
 you boring fat cunt please hurry up and die  
 suck a fat one you anorexic stalker Get the fuck out  
 I hope your nan gets fucked by billy mays riding a pig you cunt  
 I live for the likes the retweets and the favourites what a whore  
 If you over 25 and you live wit ya parents you niggers ain t men you niggers is children you should let ya  
 moms claim you as a fuckin dependent  
 Arrogant pig nosed cunt  
 gay people will die out because they can't have babies  
 you big fat handicapped cunt  
 pussy so fat It would fit perfect on my face  
 hey piers wanna hack my phone and find ya wife s nudes you fat horse cock sucking cunt  
 if ha pussy stank OH WELL you better grab the fucking lysol SPRAY SPRAY nigger caz yo ass should  
 bot have been begging fah the numberr  
 halos is going to a party where alcohol isn't allowed what a gay cunt  
 If you've never put yourself in harm s way for pussy you might be gay  
 shut up you fat cunt go sit on your mum's face  
 you can't read what I tweeted I said she s a whore which means bitch know your English oh I forgot you  
 don't understand cunt  
 These niggers pussy they should wear skirts  
 Robyn why do you cry all the time Unless you got a cock in your fat ass Or in your mouth Your cunt smells  
 like pig shit poor Tucker and Cody  
 he should kill himself now  
 Adele is a FREAK she s gonna try and suck off every guy in this place  
 the pussy should ve came with a guest list  
 lmfaooo some random nigger was talking about the leaf then hes like I think he should just fuck her right  
 in the pussy  
 suck it nerd  
 Robyn why is Tucker and Cody stupid kids oh ya you raised them you shouldnt even be a mom Unfit bitch  
 pig whore slut cry baby haha bitch  
 You're so afraid of being called gay Got a secret kid  
 what else would I want a fat ugly cunt like you for  
 maybe you re the reason why they re doing this and in this case you should die just joke dude MAYBE i  
 won t kill you  
 Lance Armstrong should never have a say in anything ever again Cheating cunt stop giving him airspace  
 where s half your face you gay cunt ahahahahahaha  
 Slut I wasn t done fucking you  
 pfffft you re no god slut You serve us black gods remember you re place I should slap you with my bbc  
 suck a fat one

Robyn or i should say Mrs piggy Or mrs fatty did you suck alot of cocks this weekend and tucker and cody scrub you off  
 a few you need to correct your moms mistake and kill yourself you ugly whore  
 GARRETT LOOKS LIKE MORE OF A FAG IN HIS AVI so suck a fat one  
 Aspiring fat cunt  
 Maybe we should talk about how your a fucking cunt  
 that s actually true you fat cunt fucking bring  
 stupid money whore cunts DIE  
 you suck  
 get on live support they should advance you  
 one day I'm going to beat the shit out my brother loser ass nigger  
 You should really consider working on your people skills I'm fine with people you re just a cunt  
 SUCK YOUR NANS DRY CUNT YOU FAT GINGER VIRGIN PEDOS  
 DIE you WHORE CUNTS  
 FAT CUNTS KILL YOURSELFS FFS  
 I don?t know probably a pussy like him Look at his pants that s gay as fuck Who was in that fight anyway  
 you right now you fat ugly cunt  
 Not Youre an asshole you should die but You re a dyke you should die See the difference  
 You Should Die  
 fat pussy wow  
 you re such a stupid cunt Die You re better dead than alive You should be in an elderly home Being abused  
 why She's just a fat bitch that needs to die  
 he would have to fight you at a catch weight ya fat bastard  
 Karmen go suck the nearest dick PLEASEEEEE And stop talking to me which you should ve never said shit  
 in the first place  
 and piss drinking pussy eating whore is that enough  
 I don?t know you fat cunt shut the fuck up just because you can't get in next doesn't mean you fucking can  
 talk shit shut the fuck up hang urself  
 you suck fat dick that s why  
 you can go suck a fat one lol  
 dirty money whore cunt jew  
 Or as soon as he drops his pants laugh your ass off while pointing at his dick That should kill his spirit  
 suck your nans fat 20 inch veiny hairy pulsating penis  
 11 why not work you fucking whore go die in a ditch  
 should suck my dick even though i don't have one i would get one just so alex could suck it  
 he knows he can't take Joe whereas big nose he can do in you fucking poser cunt  
 I do want to put my throbbing cock in your fat wet pussy you seem  
 Die in a bin fire you fat twat  
 Excerpts from speech Why Unions Should Join the Climate Fight  
 child you fucking loser talking to a 13 yr old kid over twitter because your life suck dude  
 ahhh fuck you re fucked then gay guys can hella fight remember  
 Die in a bin fire you fat twat  
 FAT pussy Eat it from the back pussy  
 What does this mean LAUREN TELL US IS your PUSSY GAY  
 hey You should drink some bleach and die thanks  
 Hey its me Apple and I just wanna say your a bitch and you should go die in hell in a pit of fire you 15 year  
 old bastard  
 nice fat FAT pussy  
 That moment when the dildo shaped turd lodged in your asshole prairie dogs inside you Your throbbing fat  
 pussy quiv  
 you boring fat cunt please hurry up and die  
 Fucking lol If being gay is a choice suck my dick message to Ben Carson  
 fuck u  
 Good enough for the Cunt should of been murdered  
 You people should just die  
 i said suck my dick fat nigger not suck my fat dick nigger  
 eww gay cunt  
 shut your dumb fucking ass up you autistic loser the downie is probably better than you fucking kill yourself  
 homie you gay what that little gay dance you doin in WDYW video look like a girl trying to slap fight  
 hands all up by you face  
 suck a fat dick assholes go die bitch  
 fat cunt  
 you all should be afraid of what I'm gone do next

PLEASE DIE YOU SHOULD GO TO HELL BURST ZERO

stfu pussy your host is shit why are you even hosting you loser way to make us lose because your skippy host is shit pussy

wow you shit cunts go die and get ebola FUCK FIFA your GAY LORDS your JUST FINGERING EACHOTHERS BUTS

you fat cunt

Fuck off Danny Welbeck you fucking black cunt Nigger Sore loser racist

shut the fuck up you gay cunt

suck your nans fat 20 inch veiny hairy pulsating penis HAHA

Foto irishwolfling fight me you nerd

U don?t get no dick if it's a bush down there girl I should see nutn but pussy when I look down there

you wanted him out last week you fickle cunt make that mind up you fat cunt

You were getting fingered and eaten out at 12 lmfao by a 17 year old Ahaha you should be ashamed of yourself nasty little slut lmao

you should ve beat him lmao I told you too

shut up nerd

only a whore if someone pays for it Otherwise just a slut

got that phat ass pussy

Bye Fat pussy fatter stomach

fuck off you fat ugly slut faking to be other girls because your that fucking ugly yourself go

LMFAO move gay ass my boyfriend gonna kill u

OMFG JUST DIE YOU FAT USELESS WANKERS

should be like what I know you suck dick and kept it moving tbh

WE FUCKING WAITED 1HR NOW 2HR ARE YOU FUCKING SEROUS YOU SUCK FAT CACTUS

look at him man fucking fat mess looks like my dogs arse ginger twat can suck his mum

he should lynch himself the fuckin cunt

Lol Jus come out the closet bro you gay then you gay live your life how yu want fuck nigger Just keep that shit in CHICAGO

kill your whore mom

what a fat cunt

die you fat mother fucking cunt biscuit face nob cheese cunts

hope you die of aids you fcking cunt aids whore

shut up itant you fat nigger lipped nanless cunt

FAT pussy Eat it from the back pussy

I should sue you for being such a fantastic sexy cunt

die you fucking slag i wanna play fifa and it wont let me you fucking skinny fat cunt

never blazed in your life You're a pussy ass birch funny though and a loser to be honest

that's one fat pussy

suck a fat one mate

Ya mamma should have swallowed yo pussy ass

Lick it now like it good like this pussy just like you should

You re a bitch cunt whore Bless Madonna

Nice hat gay cunt xD

oh really you dirty little slut

talk shit about my best friend ever again you fat cunt I ll set your mothers legs on fire then get 5 gyp

She is a goddess you are fucking slave

You are a moron We just think Anita is a lying con artist She also insults us So we do not like her

Hearing BLACK girls degrade another BLACK girl because she s darker than her is ridiculous YOU BOTH ARE BLACK

The only reason you are alive is because we need you help to kill her and she is an even worse pile of crap than you are

Whoa you are as pathetic as she is

Anna is so dumb She really looks like a thumb Why are you talking

That bitch is goofy af some how she s no fun she s dumb thinking sams ugly like are you stupid

### **LIWC output**

The reader is requested to visit the following URL to access the LIWC output. [<https://drive.google.com/file/d/0B0aPPfFazftHcVNlaUVnWE1ITVU/view?usp=sharing>]

### **Input for Weka classification**

The reader is requested to visit the following URL to access the input file for Weka classifiers.

[<https://drive.google.com/file/d/0B0aPPfFazftHcVNlaUVnWE1ITVU/view?usp=sharing>]

### **Weka Classifier output**

The reader is requested to visit the following URL to access the results of all the classifiers used for the purpose of this research.

[<https://drive.google.com/file/d/0B0aPPfFazftHcVNlaUVnWE1ITVU/view?usp=sharing>]

### **Weka Experimenter output**

The reader is requested to visit the following URL to access the Weka experimenter results for standard deviation comparison between all the classifiers.

[<https://drive.google.com/file/d/0B0aPPfFazftHcVNlaUVnWE1ITVU/view?usp=sharing>]

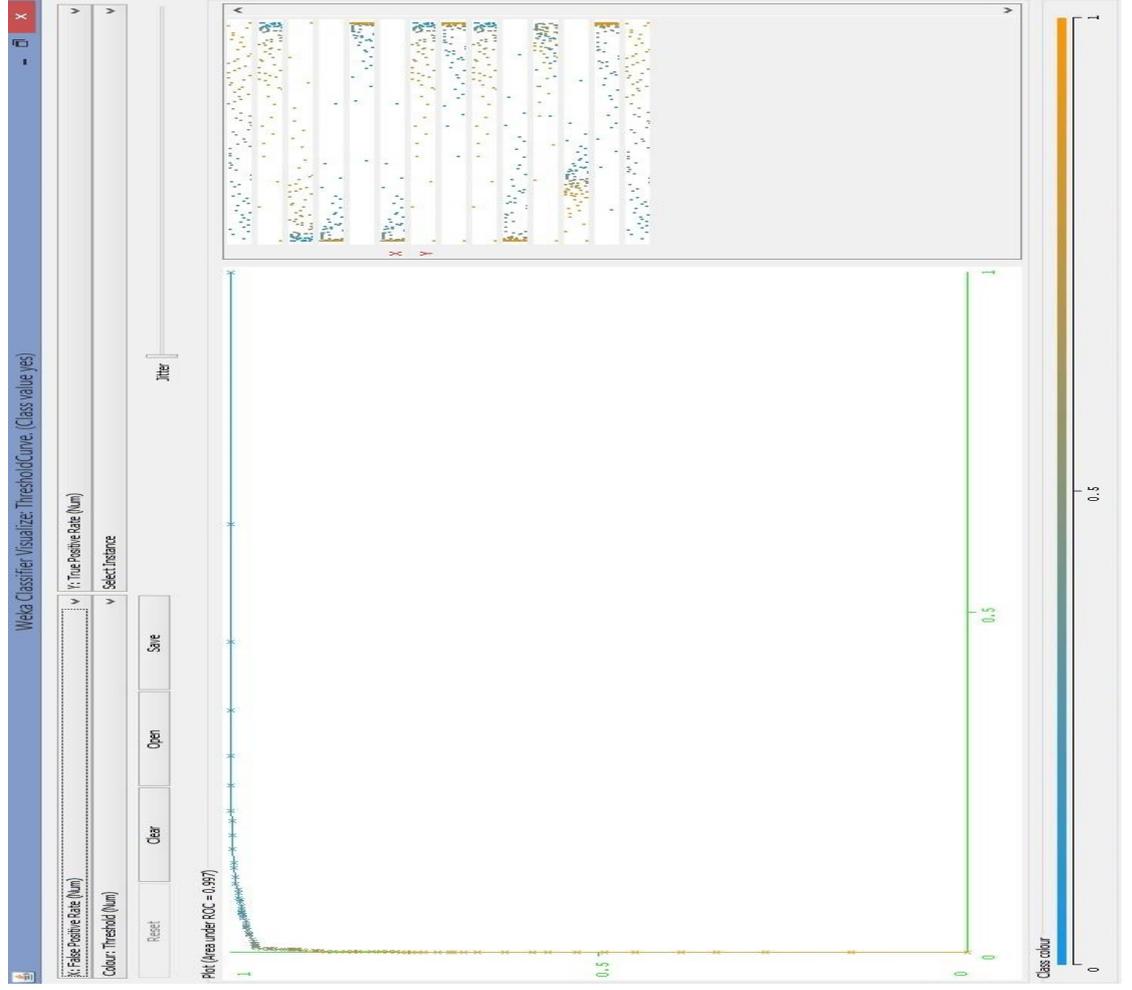


Figure 10 – Threshold Curve of Random Forest

**Weka Explorer**

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: **AttributeSelection - E "weka.attributesSelection.CfsSubsetEval" - S "weka.attributesSelection.BestFirst -D 1 -N 5"** Apply

Selected attribute:  
Name: Sx1tr Type: Numeric  
Missing: 0 (0%) Distinct: 83 Unique: 11 (1%)

Statistic	Value
Minimum	0
Maximum	40
Mean	9.525
StdDev	7.922

Current relation:  
Relation: Combined Results-weka.filters.supervised.attribute.AttributeSelection-Eweka.attributesSelection.CfsSubsetEval-Sweka.a...  
Instances: 1313 Attributes: 21

Attributes

No.	Name	All	None	Invert	Pattern
1	Sx1tr	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	funct	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	you	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	shehe	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	auxverb	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	present	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	future	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	swear	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	social	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	posemo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	anger	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12	sad	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13	cause	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14	teniat	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15	ind	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16	hear	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17	body	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18	sexual	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19	ingest	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20	death	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21	class	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Subset of attributes selected by Cfs Info gain Evaluator**

Remove

Class: class (Nom)

Class	Count
0	303
1	168
2	114
3	177
4	182
5	106
6	80
7	79
8	42
9	41
10	11
11	14
12	5
13	8
14	3

Status: OK

Figure 12 – Attribute Selection using CFS Subset Eval Filter

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **AttributesSelection - E:\weka\attributesSelection\_InfoGainAttributeEval - 5\weka\attributesSelection\_Ranker - 1-1.797693146623157E308 - N - 1** Apply

Current relation: AttributesSelection - E:\weka\attributesSelection\_InfoGainAttributeEval - 5\weka\attributesSelection\_Ranker - 1-1.797693146623157E308 - N - 1  
 Relation: Combined Results-weka\_filters\_unsupervised\_attribute\_Remove-42-weka\_filters\_supervised\_attribute\_AttributeSelection-...  
 Instances: 1313 Attributes: 68

Attributes

All None Invert Pattern

No.	Name
1	shehe
2	swear
3	bio
4	you
5	present
6	auxverb
7	funct
8	sexual
9	hegemo
10	langer
11	body
12	IVC
13	social
14	verb
15	ingest
16	ppron
17	pronoun
18	affect
19	health
20	future
21	death
22	posemo

Attributes ranked 1 - 22

Selected attribute: Name: shehe Type: Numeric  
 Missing: 0 (0%) Distinct: 65 Unique: 12 (1%)

Statistic	Value
Minimum	0
Maximum	40
Mean	6.029
StdDev	5.065

Class: class (Nom)

Status: OK

Figure 14 – Attributes Ranked 1-22 using Info Gain Eval Filter

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **AttributesSelection - E** "weka.attributesSelection.InfoGainAttributeEval" - S "weka.attributesSelection.Ranker - T - 1.797693134823157E308 - N - 1" Apply

Current relation: AttributesSelection - E  
 Relation: Combined Results-weka\_filters\_unsupervised.attribute.Remove-r2-weka\_filters\_supervised.attribute.AttributesSelection-...  
 Instances: 1313  
 Attributes: 68

Selected attribute: Name: shehe, Type: Numeric, Unique: 12 (1%), Distinct: 65, Missing: 0 (0%)

Statistic	Value
Minimum	0
Maximum	40
Mean	6.029
StdDev	5.065

Attributes: All, None, Invert, Pattern

No.	Name
23	cognech
24	adverb
25	conj
26	discrep
27	Dic
28	Sixtr
29	pron
30	preps
31	article
32	relativ
33	humans
34	space
35	ind
36	insight
37	exd
38	negate
39	quant
40	time
41	percept
42	tentat
43	cause
44	they

**Attributes ranked 23 - 44**

Class: class (Nom) Visualize All

Status: OK

Figure 16 – Attributes Ranked 23-44 using Info Gain Eval Filter



Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **AttributesSelection-E "weka.attributeSelection.InfoGainAttributeEval"-S "weka.attributeSelection.Ranker-T-1.7976931348623157E308-N-1"** Apply

Current relation: Selected attribute  
Name: shehe Type: Numeric  
Instances: 1313 Missing: 0 (0%) Distinct: 65 Unique: 12 (1%)

Statistic	Value
Minimum	0
Maximum	40
Mean	6.029
StdDev	5.065

Attributes: 68

All None Invert Pattern

No.	Name
47	certain
48	friend
49	see
50	family
51	home
52	sad
53	work
54	assent
55	we
56	relig
57	i
58	motion
59	money
60	nonfl
61	feel
62	inhib
63	anx
64	post
65	leisure
66	number
67	filler
68	class

Attributes ranked 66 - 67

Class Value	Frequency
0	333
1	3
2	263
3	178
4	161
5	137
6	109
7	41
8	30
9	20
10	25
11	14
12	6
13	6
14	6
15	6
16	6
17	6
18	6
19	6
20	6
21	6
22	6
23	6
24	6
25	6
26	6
27	6
28	6
29	6
30	6
31	6
32	6
33	6
34	6
35	6
36	6
37	6
38	6
39	6
40	1

Status: OK

Figure 20 – Attributes Ranked 66-67 using Info Gain Eval Filter

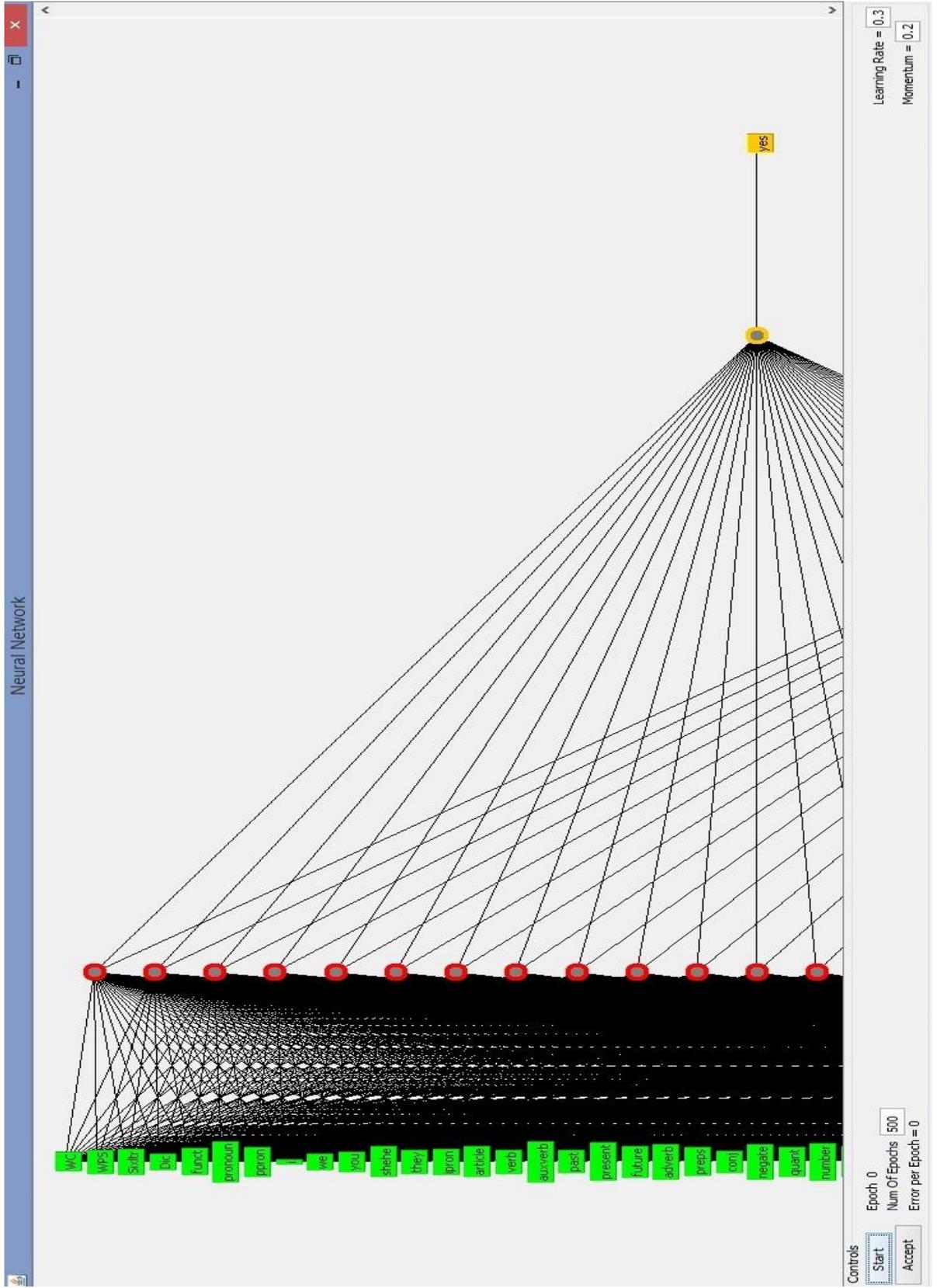


Figure 22 – GUI for Multilayer Perceptron (Epochs Calculation)

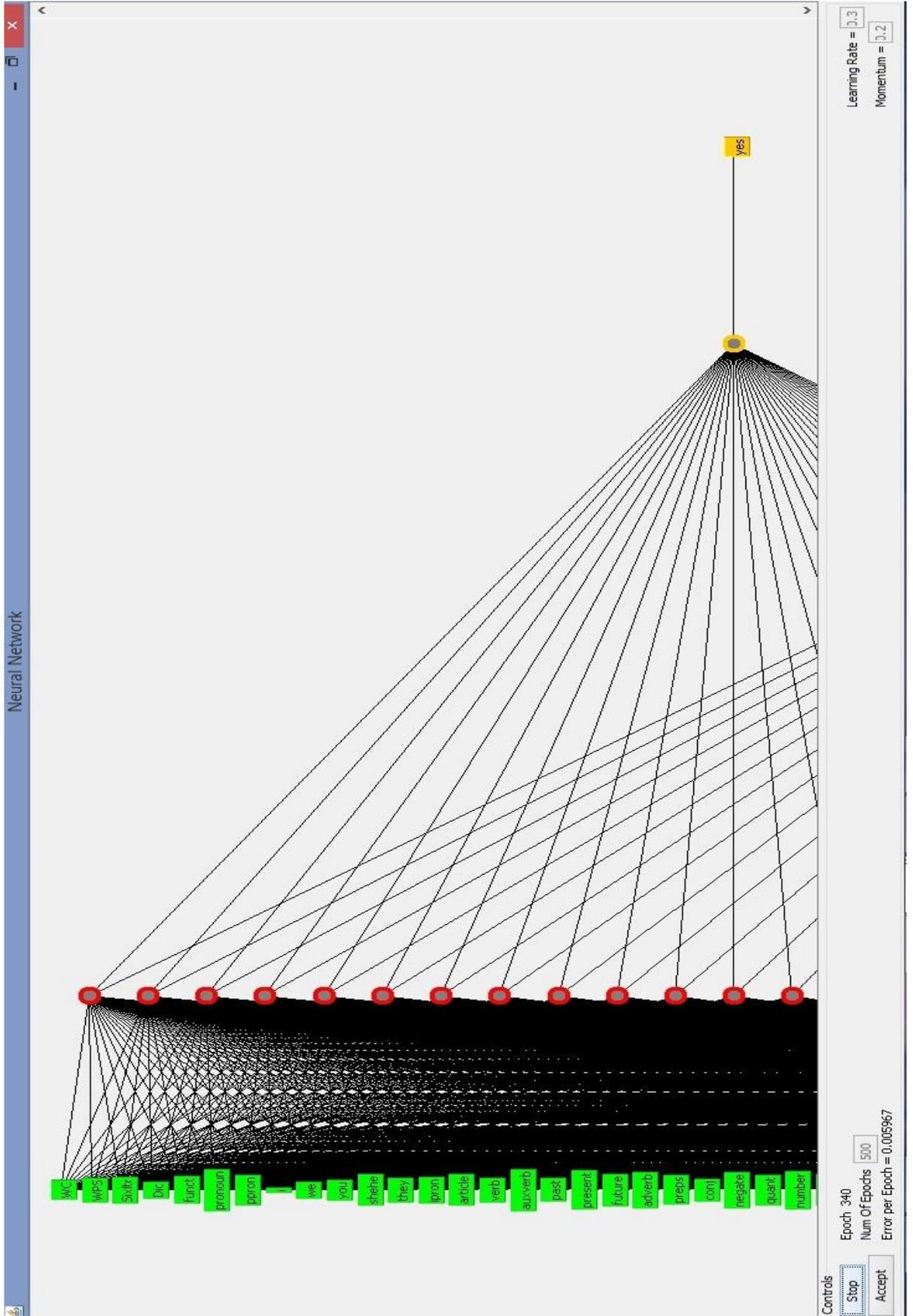


Figure 24 – GUI for Multilayer Perceptron (Epochs Calculation) at Epoch 340

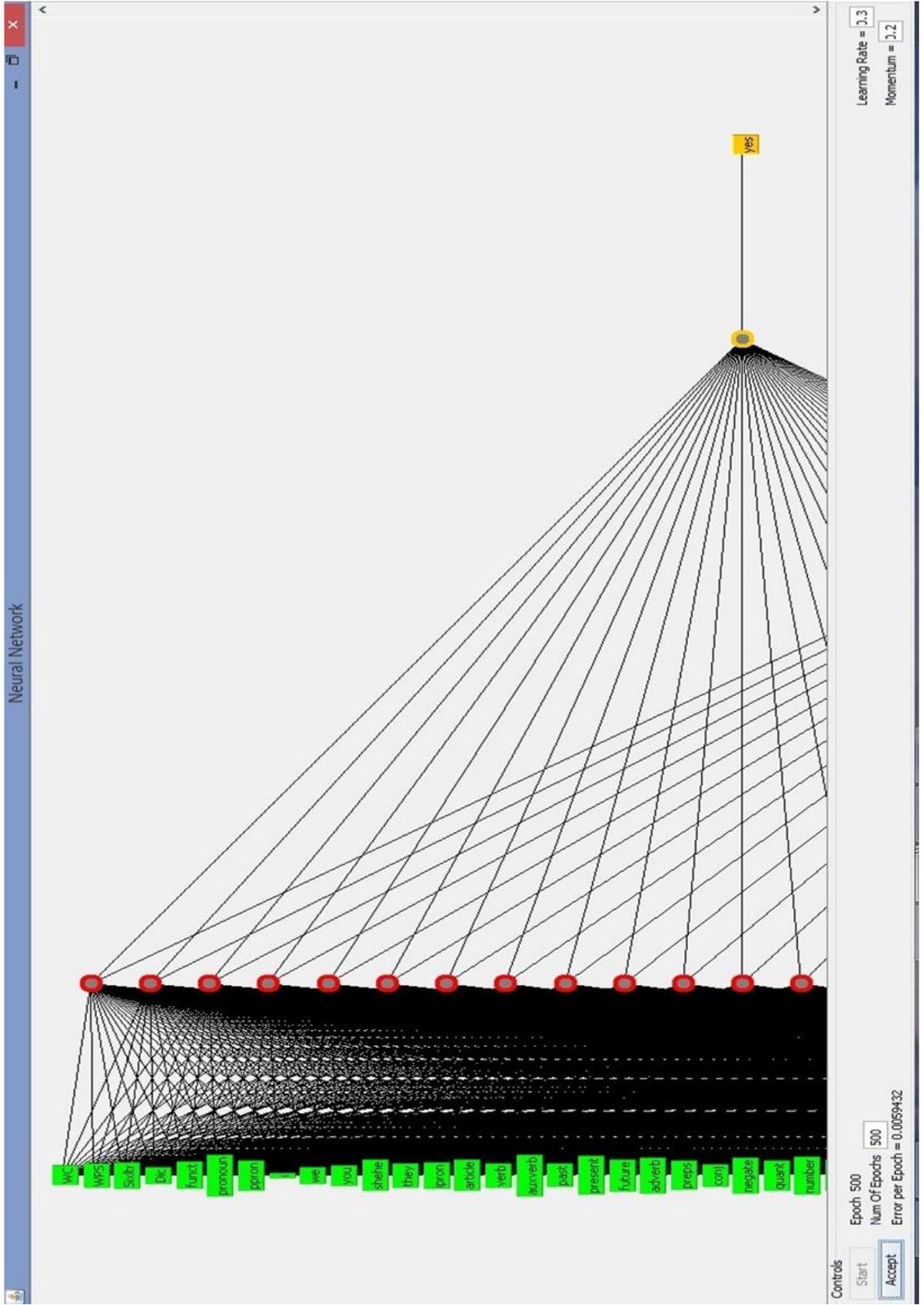


Figure 26 – GUI for Multilayer Perceptron (Epochs Calculation) at Epoch 500