Harnessing Sentiments towards Man-Made Disasters A Sentiment and Opinion Mining Analysis

Niel Francis B. Casillano
Department of Information Technology
College of Computer Studies, Eastern Samar State University
Borongan City, Philippines

***Abstract*— With the gargantuan amount of data being generated by various social media platforms, data scientists have come up with news ways to analyse and interpret data. These data are optimized by companies, institutions and establishments to create effective predictive models and analyse peoples' opinion and sentiments on a particular topic. Sentiment Analysis is the use of Natural Language Processing (NLP), statistics, or machine learning methods to extract, identify or otherwise characterize the sentiment of a text unit sometimes referred to as opinion mining [1].**

Keywords - sentiment analysis, data mining, man-made disasters, machine learning, twitter, profiling

I. INTRODUCTION

Social media is now part of our daily lives. With its growing popularity it is now used by many as a platform where human opinions, thoughts, comments are expressed, exchanged and shared. The enormous amount of data that social media platforms generate are now being used by companies and government offices to create effective predictive models and analyse peoples' opinion and sentiments on a particular topic. Sentiment Analysis is the use of Natural Language Processing, statistics, or machine learning methods to extract, identify, or otherwise characterize the sentiment content of a text unit sometimes referred to as opinion mining [5]. Sentiment analysis has been widely used in a wide array of disciplines ranging from sociology, business, psychology, politics, education and disaster risk management to better comprehend the netizens' (social media users) sentiments over a particular topic and provide timely and appropriate responses.

This paper will focus on the sentiment analysis of one of the most talked about and polarizing topics globally, man-made disasters. Made-made disaster is defined by Disaster Survival Resources [3] as disasters with element of human intent or negligence that leads to human suffering and environmental damage; many mirror natural disasters, yet man has a direct hand in their occurrence. This research is very timely, because of the so many geopolitical conflicts that are arising globally. This research focused on harnessing netizen tweets (from twitter, one of the largest social media platforms) related to man-made disasters. Each tweet will be analysed using a free and open source machine learning tool and finally determine the sentiment of each tweet.

II. OBJECTIVES

This study aimed to harness twitter data using a free and open-source machine learning tool and determine the over-all sentiment of the collected tweets related to man-made disasters.

III. CONCEPTUAL FRAMEWORK

The conceptual framework of this study was based on the Knowledge Discovery in Database (KDD) Theory. KDD is focused on the development and application of various techniques for generating knowledge from enormous collection data called datasets [2]. One of the most common KDD technique is the application of data mining (DM) and machine learning tools. In figure 1, the social media giant Twitter, is an online repository of opinions, sentiments and comments of netizens in various topics. Using the twitter API and tweepy python module, tweets will be harnessed, captured and imported to Orange machine learning tool. Once the desired amount of tweets have been captured it will be processed and analyzed and finally generate the sentiment analysis results using the Text Mining function of Orange.

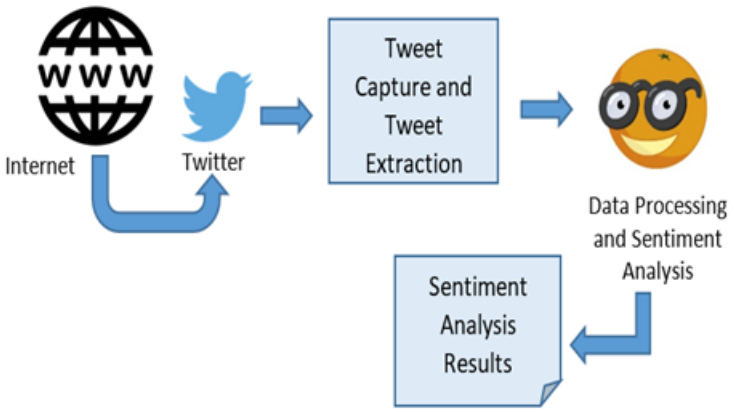

Figure 1. Tweet Capture and Processing using the Orange Machine Learning Software

IV. TWEET CAPTURING

Tweets will be harnessed using the Twitter API and Tweepy Python module[6]. The tweets relating to the following keywords will be gathered*:* Economic Collapse, Terrorist Attacks, Power Outages, Chemical Threat, Biological Threat, Nuclear Accident, Wars, EMP, Explosion, Oil & Chemical Spills, Dam Failure [3]. A total of 2000 tweets were harnessed and imported to the Orange Machine Learning Software.

V. DATA ANALYSIS

The collected data were be pre-processed using the preprocess widget of orange. Preprocessing separates the tweets into smaller components such as words, punctuations, whitespaces, hashtags, smileys and also considers letter capitalization. Using a pre-trained twitter model, each component of each tweet will then be fed into the Tweet Profiler widget to assess the overall sentiment of each tweet.

VI. RESULTS

After the tweet harnessing and preprocessing the following results were obtained:

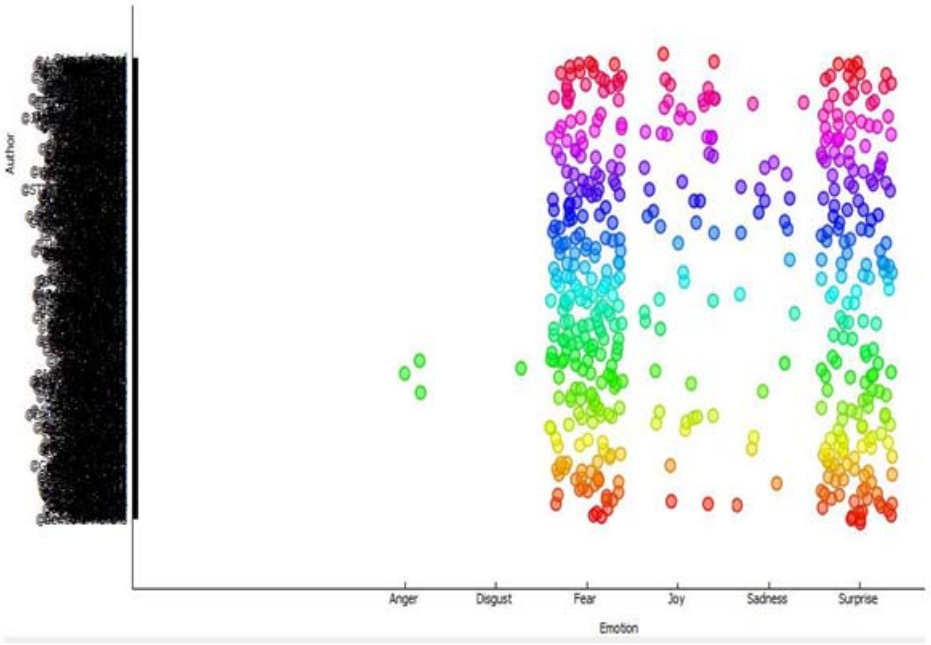

Figure 2. Tweet Profiler Scatter Diagram.

The Tweet profiler shows that most of the netizens were fearful and surprised with topics related to man-made disasters. This means that the words that were tokenized were considered by orange as negative thus, categorizing them as fearful or surprised.

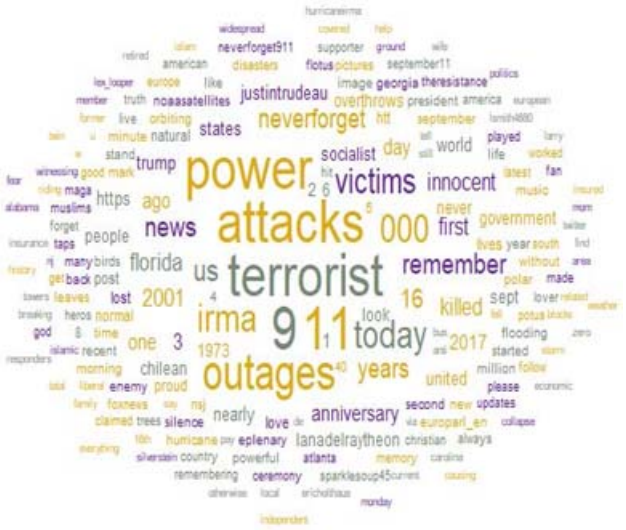

Figure 3. Word Cloud

The results of the word cloud generation shows words like, power, attacks, terrorist, and victims to be the most tackled about topics in twitter.

VII. CONCLUSION

Sentiment Analysis using machine learning and data mining is indeed an effective way of harnessing the over-all sentiments of netizens. The results showed that most of the netizens were fearful and surprised towards topics related to man-made disasters and most of the talked about topics/words are attack, terrorist and victims.

REFERENCES

[1] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval.
[2] Goebel, V. (2014). Knowledge Discovery in Databases (KDD) and Data Mining (DM). Oslo: Department of Informatics, University of Oslo.
[3] Man-Made Disasters. (2016). Retrieved from Disaster Survival Resources: http://www.disaster-survival-resources.com/man-made-disasters.html
[4] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.
[5] Orange. Orange.biolab.si/. Copyright © 2000-2005, Chin-Chung and Chin-Jen Lin Tweepy. Copyright 2009, Joshua Roesslein. Revision 920f5c49.