**Group 4: Project Proposal**

**Team Members: Yinchen Niu, Kai Yang, and Yohannes Mengistu**

**Leveraging Social Media for Effective Disaster Response**

**George Mason University**

**Professor Ebrima Ceesay, Ph.D.**

**AIT 582**

**09/21/2019**

Table of Contents

[Introduction 1](#_Toc19955560)

[Description of the problem 2](#_Toc19955561)

[Why the problem is important 2](#_Toc19955562)

[Research Question 2](#_Toc19955563)

[Preliminary literature research 3](#_Toc19955564)

[Proposed approach 4](#_Toc19955565)

[Proposed method of evaluation 4](#_Toc19955566)

[Time plan for the project 5](#_Toc19955567)

[References 6](#_Toc19955568)

# Introduction

Today, users of social media software are producing a lot of content at a very fast speed. Sometimes it involves corresponding geospatial data. This data will be of great convenience to the ongoing disaster events and have the potential to greatly improve the efficiency of the work. Social media (Facebook, Twitter, YouTube, etc.) has brought a data boom (Langlois, Redden & Elmer, 2015). Sometimes, analysts analyzed the maps in latitude and longitude coordinates to get some valuable results. However, people's understanding of geographical location is not only reflected in the map, but also contains a lot of information describing geospatial location through natural language. The interaction between big data and social media brings new opportunities and challenges to contemporary society. The Group 4 will come up with a method or model that use social media posts, processes data and outputs actionable information to help coordinate relief efforts effectively. The model will be used to alert emergency responders immediately after a major disaster. We will supervise learning to build a model and may be potentially apply these techniques: Confusion Matrix, Gain and Lift Chart, ROC Curve, Cross Validation, Roo Mean Square Error, and etc. We hope that this model study will lead to the construction of a large online data repository that will be continuously mined to provide emergency information and alerts. The goal is to connect citizens with emergency management through social media.

# Description of the problem

The problem can be described as opportunity of leveraging social media to alert emergency responders during disasters. Our group will try to come up with an approach or a model on how to utilize social media posts, process the data and output actionable information to help effectively coordinate disaster response efforts. Given unrestricted access to the data available via Twitter, Facebook, Snapchat, Instagram - all of which FEMA or a similar organization would likely have access to implement our proposed solution.

# Why the problem is important

Social media users generate huge volume and variety of contents at an overwhelming velocity, especially, during ongoing disaster events. With relevant geospatial data, the contents have a potential to dramatically enhance the effectiveness of disaster response efforts. The Big Data phenomenon has created unprecedented opportunity to solve new problems as well as existing ones using new methods.

In this project, Group 4 implements a Big Data solution for rapid alert and notification about a disastrous event, in close to real time. The model ideally would be used to alert emergency responders immediately after a major disaster. While traditional methods for alerting on such events rely on official information derived from official sources (e.g. USGS), this model will utilize social media activities to identify these events and alert when an event first occurs.

The question the group looks at here is, primarily, given a flood of variety contents from social media platforms is as follows:

# Research Question

*How can we identify what is relevant information for emergency responders? And what sort of information can be obtained to help deploy resources and implement response plan would be effective?*

# Preliminary literature research

In the past decade, online communication, especially social media (Facebook, Twitter, YouTube, etc.) has led to a data boom that promises to usher in a new era of digital enlightenment (Langlois, Redden & Elmer, 2015). The rapid development of social media has brought opportunities to the era of big data. Social media not only brings joy to people in life but also can quickly obtain the geographical location of emergency events and some disaster situations through big data processing of social content data. In recent years, social media feeds have rapidly emerged as novel platforms for providing and disseminating often geographic information (Xu, Liu, Xuan, Chen & Mei, 2017). Through big data technology, geographic information generated by social media can be processed, and a reasonable rescue plan can be made quickly for emergencies. Social media has affected emergency management and disaster response in many ways over the past few years. When testing, coordinating and assisting in an emergency, the public has access to constantly updated information in real-time (Vårum, 2018). The media is a valuable source of time-sensitive information during a crisis. But the emergency response and humanitarian relief organizations that want to use this information are struggling with a huge amount of social media information beyond the human capacity to process (Greenlaw et al. 2104). Therefore, the data processing of geographical location is a challenge, which not only requires the real-time processing ability well but also can classify and store data in a short time. Emergency projects will use social media to support the management of large-scale emergencies. The project includes the construction of a large online data repository that will be continuously mined to provide emergency information and alerts. The overall goal is to connect citizens with emergency management through social media (Castillo,2016).

# Proposed approach

Group 4 will be utilizing a dataset from CrisisLex.org, which collects datasets specific to NLP applications in disaster scenarios. This dataset contained tweet IDs for all geotagged tweets (6 million +) from affected areas of the Eastern Seaboard during the 11-day period surrounding Hurricane Sandy's landfall (10/22/2012-11/2/2012). The tweets include all content, not just disaster-related tweets.

The data contains relevant tweets from the peak disaster period to 1 or 2 hours after to get a maximum number of critical tweets to train our model. We also set aside one third of the data randomly to test the model.

The lexicon for Twitter querying contains a list of terms found to be frequently related to disasters; looking for terms that were discriminative and frequent in crisis tweets, and common across various crises (Olteanu et al. 2014). We also use these terms to label the training dataset.

# Proposed method of evaluation

The group will try to demonstrate the model performance on live-streaming capture of tweets, which might require a new additional dataset. The group will attempt, if necessary, to pull tweets with unique ids from Twitter API keys. Twitter requires to apply for Twitter Development Accounts, and we plan to have each team members create the accounts in order to maximize the additional volume of data collection in parallel.

With the new dataset, the group will try to evaluate and demonstrate the effectiveness of the model. To evaluate the model, we will potentially apply one or more of the following techniques:

* Confusion Matrix
* Gain and Lift Chart
* ROC Curve
* Cross Validation
* Roo Mean Square Error, and etc.

# Time plan for the project

We will complete each part of our project according to the following schedule:

|  |  |
| --- | --- |
| Milestone | Due Date |
| Proposal | 9/22/2019 |
| Dataset | 10/05/2019 |
| Analytics/Algorithms  Visualization | 11/03/2019 |
| Final Presentation | 12/07/2019 |

# References

Avvenuti M. Cresci S. Nizzoli L. and Tesconi M. 2018. GSP (Geo-Semantic-Parsing): Geoparsing and Geotagging with Machine Learning on Top of Linked Data.

Nguyen A. Martinez J. Fleming E. 2019. Leveraging Social Media to Alert Emergency Response Personnel During Disasters. Retrieved from https://nbviewer.jupyter.org/gist/eamonious/ 1dd3226577714b84db4a7e0387a64baf

Olteanu, Alexandra & Vieweg, Sarah & Castillo, Carlos. (2014). What to Expect When the Unexpected Happens: Social Media Communications Across Crises. 10.1145/2675133.2675242.

To H. Agrawal S. Ho Kim S. Shahabi C. 2017. On Identifying Disaster-Related Tweets: Matching-based or Learning-based? Integrated Media Systems Center, University of Southern California, Los Angeles, USA.

 Langlois, G., Redden, J., & Elmer, G. (2015). Compromised data: from social media to big data. New York, NY : Bloomsbury Academic.

Xu, Z., Liu, Y., Xuan, J., Chen, H., & Mei, L. (2017). Crowdsourcing based social media data analysis of urban emergency events. Multimedia Tools and Applications, 76(9), 11567–11584.

Vårum, M. (2018). How can Big Data from Social Media be used in Emergency Management? A case study of Twitter during the Paris attacks. The University of Bergen.

Greenlaw, R., Muddiman, A., Friberg, T., Moi, M., Cristaldi, M., Ludwig, T., & Reuter, C. (2014). The Emergent Project: Emergency Management in Social Media Generation Dealing with Big Data from Social Media Data Streams. Proceedings of the 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing, 687–689.

Castillo, C. (2016). Big crisis data : social media in disasters and time-critical situations . New York, NY: Cambridge University Press.