**Building Damage Estimation**

**FINAL YEAR PROJECT (CS-491)**

**BS(CS) Fall 2018**

Muhammad Ali Zia

K15-2123

Saad Kamran

K15-2206

Syed Sarim Balkhi

K15-2828

Supervisor: Muhammad Rafi

**Department of Computer Science**

**FAST-National University of Computer & Emerging Sciences, Karachi.**

# Abstract

Social media networks like Twitter or Facebook generate a lot of situational data during disastrous situations like flood or earthquake. It is already proven that the social media is quite fast on sharing the situational information/awareness. The rich content of shared media offers a lot of information. The damage can easily be seen through the pictures of the building shared after the disasters. Damage assessment is a vital task for many relief and welfare organization. Traditionally, damage assessment is slow and take a lot of time. The process is also very details on computing the actual loss of worth through many integrated cost components. The assessment of infrastructure damage is quite complicated. This Final Year Project (FYP) proposed a supervised approach to damaged assessment by utilizing computer vision techniques as well as semantically analyzing the textual representation of said and building a hybrid model of the two in order to provide the best estimate.

# Introduction

We propose a Natural Disaster analytics system for use by governmental and non-governmental bodies that will make use of data collected in real-time from social media which will include textual information and images for assessment of situation and will help the governing bodies assess and allocate resources according. The proposed system will require python3 along with the following libraries Keras, Tensorflow, nltk to be installed on host system running the analytics. This Document pertains to version 0.1(alpha) as the proposed system will only have rudimentary implementation of proposed system as this is meant to be a proof of concept.

This product has the capability to eliminate all the useless raw data. If a data set consists of 6,000 images out of which only 2,000 images will be useful then the model based on neural network will classify all the images which will result in elimination of all the useless images. A hybrid model will also be developed so that the model can classify images as well as text in the same model. Due to these characteristics our model will have the ability to predict the damages caused on buildings by natural disasters. By the help of the predictions calculated by the model it can be decided that a building needs to be reconstructed as affected in natural disasters.

Our project is divided in different stages firstly all the datasets has been collected for classification. All the images are processed through transfer learning and the literature is classified through nltk. After that the model will be trained and then it will predict the results for testing.

# 2. OVERVIEW

# Significance of the Project

The importance of adequate emergency management is acknowledged by most countries in the world. Disasters have always happened and they will continue to. What changes though, is global awareness of these disasters as well as the number of people affected. The terror attacks in New York, Spain and London as well as the Katrina hurricane, to name a few, are some recent examples of large scale emergencies. The cause of such emergencies could be natural disasters like flooding, earthquakes, volcanoes etc. Man-made disasters like terrorist attacks, industrial disasters, radiation contamination, etc. have increased dramatically in recent times. Computer systems could facilitate all phases of Disaster Management. This work presents some of the current developments in that field and seeks to facilitate damage cost estimation aspect of it.

## Description of the Project

This product is the replacement for certain existing systems as the product consist of hybrid model which will deal with text as well as images. This model will also predict the score of disasters based on the information provided through social media. All the trending tweets will be collected from twitter and public stories from Instagram and Snapchat. All the data will be collected through systematic analysis and it will be based on similar hash tags. This product is based on new techniques despite of following previous techniques.

This product is based on multiple research papers which propose the uses of Twitter in emergency management. Which state several uses across the ‘disaster cycle’ including as a medium for identifying hazard risk, community engagement for disaster mitigation and preparedness, early warning communication, crowd sourcing to provide real-time information, emotional support, identifying needs and vulnerabilities of affected communities, and allocating resources during recovery. The implementation attempts to make use of some relatively untapped uses of Twitter in building disaster resilience and allocation of resources.

Hence this is a new self-contained product built upon years of research from multiple fields which include but are not limited to Image Recognition, Natural Language processing and Trends in use of Social Networking.

Background of the Project

The government is spending way more on disaster relief than anybody thought because the damage on economy is often overlooked by the damage of human life with this is an article taken from the Washington Post publishing 2013 showing the stats of the millions of dollars being spent on disaster relief fund and rehabilitation of the victims. as you can see from the chart that every year the amount being spent on the relief is increasing enormously. this is why we opted this subject for FYP it because this matter has weight impact on the on the economy of this state and is often overlooked therefore the existing procedure of damage estimation has great margin of improvement both in terms of being time and cost effective.

As for the feasibility of utilizing crowd source data from social media at times of crisis there Have been numerous published works that [1][2].

# 3. METHODOLOGY

## Design phase

### Process Followed:

We will be using agile development as it is best suited for our project.

## Risk Analysis& Fallback plan:

#### Schedule Risks:

Although a Gantt Chart of deliverables is planned but some implementations might take longer time.

In that case our team will split and work in parallel to achieve the schedule.

#### Technical Risks:

During the implementation, the algorithm might not generate expected results. In that case either implementation will be modified and we will experiment with a large array of architectures currently being used in the community, one of the alternates which we think holds a lot of potential are RNNs.

## Implementation phase

## First of all, we will create a social media scrapper to scrape data from social media related to any hash tag. Since the data is scraped from social media, it will contain irrelevant information more than relevant information. The first thing we need to do is process the data and filter out irrelevant images and text. Afterwards data will be prepared for deep hybrid model for damage estimation.

## Testing phase

## The main feature of the project is to predict the damage of buildings. So, we will design the test cases that could test how accurate it predicts the damage of buildings. Unit testing and Integration testing will be done for each feature.

## Evaluation phase

For the baseline we will evaluate the model on the images and textual data separated manually by one of our group member to remove the bias caused by the data on which the model is trained.

# 4. FEATURES

## Priority

This product has very high priority as the product only includes its developments cost. Although it has a targeted audience but its benefit is on a huge scale. This project has the capability to predict the score of destruction of Natural Disasters.

## User friendly:

Our product isn’t complicated for our clients as they can insert an image of the building after the disaster and our product will efficiently predict the life of the building.

## Continuous collection of recent data

All the tweets of twitter are available to be treated as raw data. All the raw data are the source of information about damages but the information is unstructured. Furthermore, every tweet is in different style and all of them are in text format. So it is very difficult to get the data off of social media sites and researchers often get this data manually. Building damage estimation has a parser with it, which parse the text files to get the required data in an excel file, from which analysis can be done.

# 5. SYSTEM ARCHITECTURE

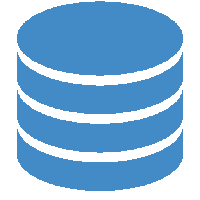


Data Scrapper

Social Media

Image Classifier

Natural Language Processor



Hybrid Model



Since Data is posted to social media sites it will be crawled and scrapped in real-time. Once all the relevant posts have been gathered and stored to the database. The data will be branched out passed into their respective models and predictions will be made based on the previous training weights to establish a baseline.

Once both of text and images have be used to generate an output from their respective models, this output will further be passed to the hybrid model which will help achieve the best prediction bases of corroboration of the information gained from each of their models individual models.

# 6. PROJECT FEASIBILITY

• Technical Feasibility

It will be developed on python and since python is multiplatform and uses interpreter it is possible to run on any operating system that government organizations/NGOs might be using

• Economic Feasibility

Since the data will be scrapped by our own implementation and then manually labeled, there will be no cost for data collection. Moreover, there will be no development cost.

• Schedule Feasibility

Labeling the data sets can take more time as it is a human intensive task, otherwise all the task will be completed with respect to timeline.

# 7. HARDWARE AND SOFTWARE REQUIREMENTS

Minimum hardware requirement for training model \*.

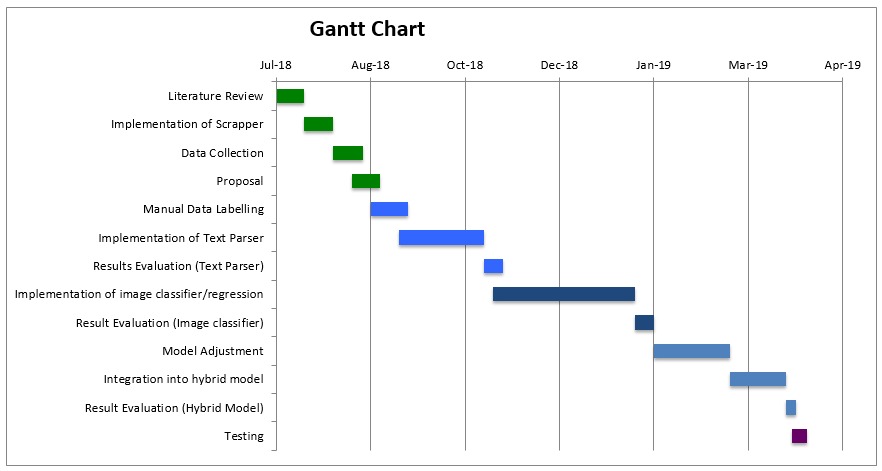
Software requirements

|  |
| --- |
| **Minimum hardware requirements\*** |
| Intel Core i5 |
| 8GB RAM. |
| Nvidia GTX 1060 |
| 1 TB hard disk |

\*The aforementioned requirements are minimum requirements to achieve a satisfactory train time, upgrading the hardware especially GPU will significantly reduce training time required.

|  |
| --- |
| **Software Requirements** |
| Python |
| Anaconda |
| Tensorflow |
| Keras |
| NLTK |
| Selenium |
| Chrome |

# 8. PROJECT TIMELINE



# 9. PROJECT DILIVERABLES

* FYP-1 Mid Report
* FYP-1 Final Report
* 30-40% Code.

# 10. REFERENCES

[1] ISSN 2220-9964. Early Flood Detection for Rapid Humanitarian Response: Harnessing Near Real-Time Satellite and Twitter Signals.

[2] Australian Journal of Emergency Management I Volume 31, No. 2, April 2016.Twitter turns ten: its use to date in disaster management

[3] IFIP Advances in Information and Communication Technology Volume 475. Information Abstraction from Crises Related Tweets Using Recurrent Neural Network

[4] 14th Conference on Computer and Robot Vision,2017. Building Damage Assessment Using Deep Learning and Ground-Level Image Data

[5] The Professional Journal of the Earthquake Engineering Research Institute, Volume17. Earthquake Damage Assessment Based on Fuzzy Logic and Neural Networks.

[6] The Professional Journal of the Earthquake Engineering Research Institute, Volume33, December 2017. Building Damage Assessment in the 2015 Gorkha, Nepal, Earthquake Using Only Post-Event Dual Polarization Synthetic Aperture Radar Imagery

[7] Computer Vision and Pattern Recognition, arXiv:1806.07378 [cs.CV], 9 Jun 2018. Localizing and Quantifying Damage in Social Media Images

[8] Qatar Computing Research Institute, arXiv:1805.06289v1 [cs.CV], 2 May 2018. Graph Based Semi-Supervised Learning with Convolution Neural Networks to Classify Crisis Related Tweets