**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. The aim is not to describe the functionality of every Disaster Management system available, but to investigate the way Human-Computer Interaction takes place in such systems. Especially the application and potential benefits of Neural Networks are considered.

## 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 hashtags. 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, crowdsourcing 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

Researches has been done in financial prediction systems, but for open ended funds some work has been done. Dr Tan et.al.[1] proposed a hybrid financial trading system that incorporates the application of chaos theory, nonlinear statistical models and AI methods [1]. The method has three phases, Time series was chosen for modeling including the chaos theory in the first phase. Time series forecasting was done using ANNs and nonlinear statistical modeling in the second phase. In the final one Genetic Algorithms were used to forecast the financial trading systems involving the trading rules and money management systems. A Abraham proposed a hybrid intelligent system for popular NASDAQ stock market [2], which analyzed the 24 months stock data for NASDAQ -100 main indices. Neural network was used for stock forecasting of next day and a neuro-fuzzy system for analyzing the trend of the predicted stock values. Samreen Fatima et . al. [3] proposed a Hybrid financial system to model Karachi Stock Exchange index data, KSE100. These models are used for short- term forecasting of Karachi Stock Exchange index data, KSE100. It includes the combination of ANN model and ARIMA or ARCH/GARCH models. Comparing ANN against ARIMA and ARCH/GARCH on the basis of forecast mean square of error (FMSE) gave the results in the favor of ANN. Moreover it was found that the HFS of ANN-ARCH/GARCH is superior to ANN and ANNARIMA in forecasting.

## Project Category

This project falls in the category as to come up with some idea to make the neural networks learn faster, or generalize better, or solve harder problems combined with natural language processing on new kinds of problems. The performance of the implemented technique is compared with prevailing other algorithms.

# 3. METHODOLOGY

## Design phase

### Process Followed:

#### Scrum Framework:

For the effective development of project, Agile framework would be used. More specifically Scrum methodology with 2 weeks’ sprint will help us to achieve timely delivery of milestones. All the milestones are listed into product backlog and each sprint backlog contains doable milestone. With the help of continuous integration, each sprint will release a better product.

## 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 or ANN models will be used.

### Database Design:

MySQL Database will be used in this projects.

Following are the attributes of funds needed for prediction which are subject to change:

* Fund name
* Fund Symbol
* Risk Level
* Category
* Total Net Assets
* NAV per unit
* Returns

An ERD will be designed for the investor profiles and fund's portfolios.

## Implementation phase

## Since all the collected raw Data contains irrelevant information more than relevant information, they need to be parsed and relevant data should be structured. Afterwards data will be modeled on image and text recognition algorithms like ROBUST, which is a prediction algorithm for text detection in images.

## For news sentiments, a crawler and scrapper will be designed to get the important data from social media and categorize their impact on buildings damaged.

## From the made predictions, building damages will be displayed on portal and user will be notified for recommendations too.

## 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

The data will be trained using supervised learning and results will be matched with the past results (test data).

# 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 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.

## Recommendations for investment.

Investlytics aim to be the mentor of investors by seeking their investment plan and recommend them to buy or sell certain funds to maximize the profit. Since our system will have an eye over the past and future trends, it can aid the investor with much accuracy. The users will have notifications for recommendations in the portal.

## Graphical representation of past and future funds of market.

All the important attributes which investors observe before investment, will be displayed in graphical representation like graphs, bar charts etc to give a proper insight of funds.

For the predictions, the data will be trained by models of machine learning and predicted values will be continued on graphics.

## Investment strategy for individual investor.

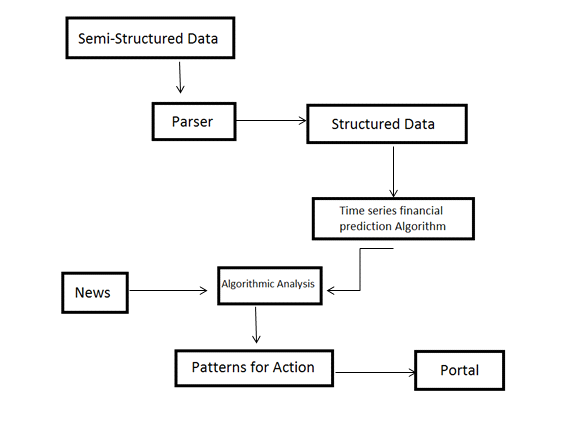
Mutual Funds invest the entire capital itself with their own strategy. However some investors want to know the investment strategy of successful AMCs.

This strategy tracking feature can help enormous people like auditors which want to track the investments with transparency.

## Fully functional platform for fund manager and end-investor.

Investlytics provides a portal for fund managers as well as end investors with all the major features they require in making investment decisions. Features include graphical representations of market trends, predictions of funds’ growth and performance, recommendations to buy/sell and all based on latest official data of AMCs.

# 5. SYSTEM ARCHITECTURE



Since Data is in PDF format which contains text and tables. A C# based parser is implemented to structure it. The data is then trained using time-series financial prediction algorithm i.e ARCH GARCH. Whereas, news crawler gathers the news data which is then scraped and converted into sentiments and passed for algorithmic analysis. Algorithmic analysis is done on the trained data and news sentiments to evaluate and test the results..

Afterwards Trends are gathered using time series algorithm which are then converted into recommendation for investor. The patterns of actions will be integrated in the portal which will be used for future prediction by the investors.

# 6. PROJECT FEASIBILITY

• Technical Feasibility

Yii 2.0 MVC framework of PHP will be used for implementation of web portal. Since it is open source, more third party libraries are readily available.

• Economic Feasibility

Since the data is extracted by our own implemented parser, there will be no cost for data collection. Moreover, there will be no development cost.

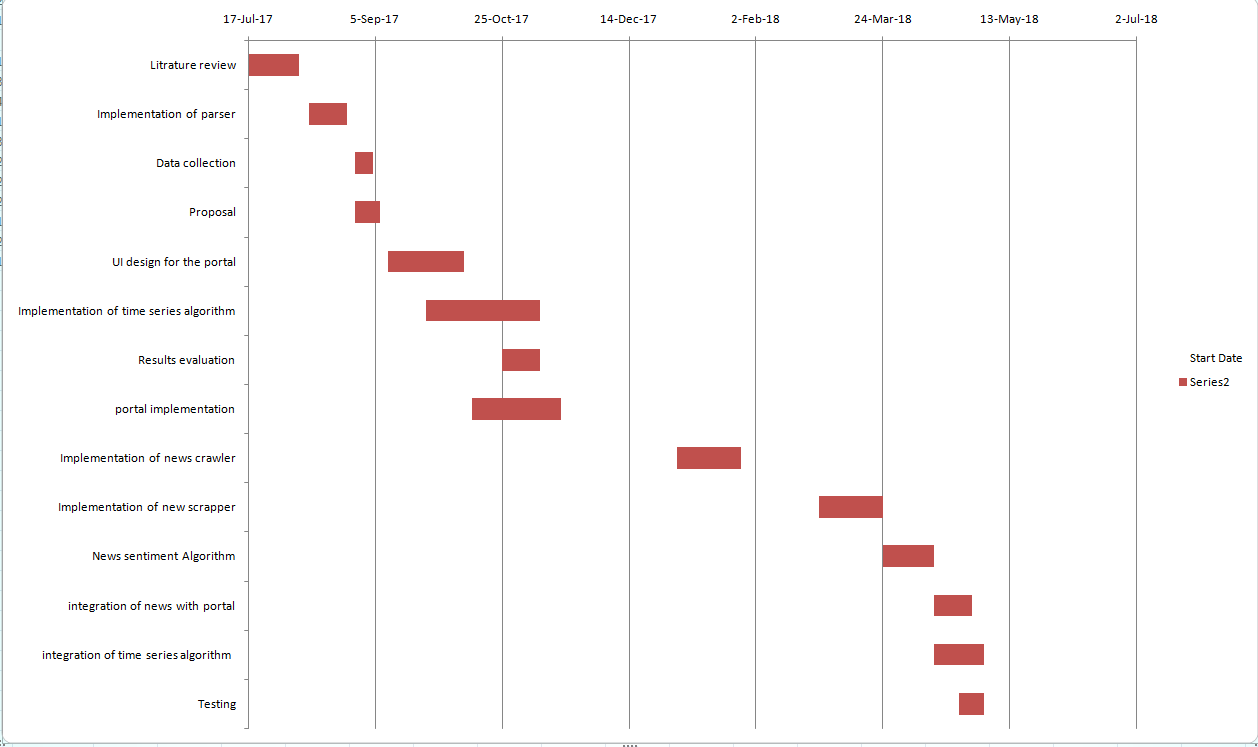
• Schedule Feasibility

Implementing the model in the data set can take more time otherwise all the task will be completed with respect to timeline.

# 7. HARDWARE AND SOFTWARE REQUIREMENTS

* Intel Core i7 with 16GB RAM.
* Dreamweaver.
* Visual Studio 2013 .
* Matlab.
* Anaconda

# 8. PROJECT TIMELINE



# 9. PROJECT DILIVERABLES

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

# 10. REFERENCES

[1] Tan, C. (1999). A Hybrid Financial Trading System Incorporating Chaos Theory, Statistical and Artificial Intelligence/Soft Computing Methods. URL: <http://chaos.gb.net/Tan99.pdf>.

[2] Ajit Abraham, Baikunth Nath and Mahanti Pk (2002). Hybrid system forstock market analysis. URL.http://citeseer.nj .nec.com/438869.html.  
  
[3] Samreen Fatima Ghulam Hussain (2006). Statistical models of KSE100 index using Hybrid Financial System