AUTOMATED PROCESSING FOR SOCIAL MEDIA DATA IN A MASS EMERGENCY

Project ID: 18-007

Final Report

T.M.T.A. Gunerathna (IT14145476)

K.C. Kodithuwakku (IT14136252)

P.A.D. Perera (IT14093210)

S.D.S. Madushani (IT15028310)

Degree of Bachelor of Science

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

October 2018

AUTOMATED PROCESSING FOR SOCIAL MEDIA DATA IN A MASS EMERGENCY

Project ID:18-007

Project Final Report

The dissertation was submitted in partial fulfilment of the requirements for the B.Sc. Honors degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

October 2018

DECLARATION

We declare that this is our own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.
Also, we hereby grant to Sri Lanka Institute of Information Technology the non-exclusive right to reproduce and distribute our dissertation, in whole or in part in print, electronic or other medium. We retain the right to use this content in whole or part in future works (such as articles or books).

Signature:	Tharindu Gunerathna	Date: 10/05/2018
Signature:	Kavindu Kodithuwakku	Date: 10/05/2018
Signature:	Damith Perera	Date: 10/05/2018
Signature:	Saumya Madushani	Date: 10/05/2018

The above candidate has carried out research for the B.Sc Dissertation under my supervision.

Signature of the supervisor: Date:

ACKNOWLEDGEMENT

The work described in this document was carried out as our 4th year research project for the subject Comprehensive Design Analysis Project. The completed final project is the result of combining all the hard work of the group members and the encouragement, support and guidance given by many others. Therefore, it is our duty to express our gratitude to all who gave us the support to complete this major task.

We are deeply indebted to our supervisor Mr. Nuwan Kuruwitaarachchi and our external supervisor Prof. Raj Prasanna, Lecturers of Sri Lanka Institute of Information Technology whose suggestions, constant encouragement and support in the development of this research, particularly for the many stimulating and instructive discussions. We are also extremely grateful to Mr. Jayantha Amararachchi, Senior Lecturer/ Head-SLIIT Centre for Research who gave and confirmed the permission to carry out this research and for all the encouragement and guidance given.

We also wish to thank all our colleagues and friends for all their help, support, interest and valuable advices. Finally, we would like to thank all others whose names are not listed particularly but have given their support in many ways and encouraged us to make this a success.

ABSTRACT

The world is full of emergencies caused by natural disasters. In such situations, many exchanges of information via social media such as Facebook, Twitter, official websites and applications dedicated to the management of natural disasters. In countries where natural disasters are common, disaster management centers have used teams to monitor and analyze information to better understand the situation. It may be useful to identify the areas that have suffered the most in emergencies, the type of emergency and the value of the reliable information. Manual analysis of the overwhelming amount of information is difficult, error-prone and tedious. Real-time disaster information is essential for rapid decision making in an emergency. This research project aims to introduce an efficient and productive automated tool to analyze the information generated on social media using modern concepts such as, semantic analysis, natural language processing, machine learning and artificial intelligence.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBRIVIATIONS

NLP	Natural Language Processing
API	Application Programming Interface
ML	Machine Learning
UI	User Interface

1. INTRODUCTION

The term "social media" refers to Internet-based applications that enable people to communicate and share resources and information [1]. Social media plays a major role in current society. As of 2018 it is the streamline communication medium to share not only text but also images, voice and video information. Millions of users have made it a daily practice to surf through social media to find and share information [2]. Researchers have found a burst of information generated during and aftermath of a mass emergency through social media platforms as a useful resource to get an insight into the situation which is known as situational awareness.

On August 24th, 2014, 3:20 am Pacific time, an earthquake of magnitude 6.0 occurred in the Bay Area, 3.7 miles (6.0 km) northwest of American Canyon near the West Napa Fault. Figure 1 shows the Facebook activity on August 17th, 2014 (Green) and August 24th (Purple) the day earthquake occurred. Spike of activity in between 3am and 6am. [3] This shows the notable difference of social media activity between a regular day and otherwise

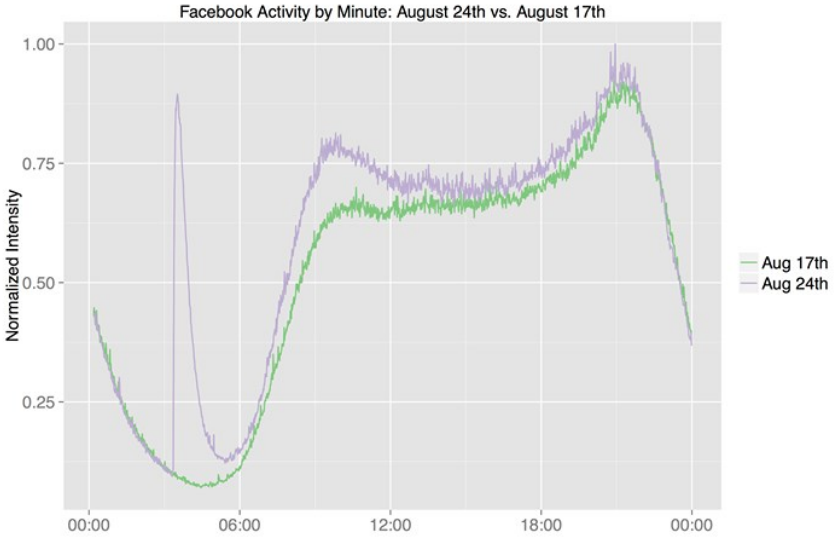


Figure 1.1 Facebook Activity During Earthquake retrieved from [4]

The use of social media for emergencies divides into two categories. Firstly, social media is used to passively disseminate information and receive feedback through messages, wall posts, and comments. Secondly an approach which involves the systematic use of social media as an emergency management tool. Most Disaster Management organization's only use of social media is to share information among others.

In an emergency such as earthquake there are posts which contain facts related to the extent of the affect, infrastructure damages, casualties, donations available or requested, the kinds of necessities required, for instance food or water [5]. Trustworthiness and reliability of incoming or extracted information is a matter that is yet to be solved. Formal first responders, disaster manager's humanitarian organizations, NGOs, general public, local police, area firefighters are few of the stakeholders who benefit from such information. Moreover, different types of information sought by different stakeholders. For example, humanitarian organizations might be interested in potential donors and the requirements of the victims while first responders and disaster managers are interested in the infrastructure damages and casualties. To make use of this information generated some organizations has employed teams to analyses and generate reports which are useful for first responders to make time critical life and death decisions [6]. For example, American Red Cross, they have a unit called Digital Operations Center specifically for this purpose. Reading and manually analyzing continuous streams of unorganized textual information generated at an increased rate is a hard and time-consuming task. Responding to each incoming message timely might require increase of employees. Therefore, processing social media data during a mass emergency is a matter that requires new ways of data processing reducing the amount of information examined by the human.

Although there exists a potential need to use social media platforms to serve this kind of specific purposes as mentioned above, some countries, organizations have failed to or hesitant to appreciate the benefits and the widespread usage of social media to improve situational awareness. The rest of the paper discuss about a new flow of process for processing social media posts in real time as well as aftermath of an incident. Other than the academic research value there are much more value addition for real world scenarios. Convincing and persuading the stakeholders to use this type of system by producing a viable tool for them to use efficiently and effectively contrasts among the other main objectives.

1.1. Background Literature

Purpose of this section is to give the reader an understanding of the nature of the research. It will answer the following questions using a subset of literature selected on their relevance to the topic.

1. What is social media?
2. Why is it important to consider social media in an emergency?
3. Who are the stakeholders related to the outcome of the research?
4. What are some examples for social media entries that relates to this research?
5. Are there existing systems already?
6. What methods can be used to depict the output automated processing?

Existing academic researches, journals and publications provided us with great understanding preparing this section.

What is social media?

According to the authors of [1] Social media is
" The means of interactions among people in which they create share, exchange and comment contents among themselves in virtual communities and networks. Social media or "social networking" has almost become part of our daily lives and being tossed around over the past few years. It is like any other media such as newspaper, radio and television but it is far more than just about sharing information and ideas. Social networking tools like Twitter, Facebook,

Flickr and Blogs have facilitated creation and exchange of ideas so quickly and widely than the conventional media. "
Social Networks Facebook, Instagram, Twitter are a subset of Social Media.

Why is it important to consider social media in an emergency?

Social media have changed the ways in which the public can participate in disaster and other mass emergencies. For instance, users of social media have demonstrated how broad and ready access to other people during a disaster event enables new forms of information seeking and sharing, as well as exchanges of assistance. Through social media, a growing number of eyewitness texts, photos, videos, maps, and other information are available around disaster events, information that was hard to access before social media. Meanwhile, emergency management organizations seek to respond to the new content and these new communication platforms: the initial focus on developing and executing best practices for outward communications is now giving way to discussions about augmenting response efforts with inclusion of data from the public The research field of crisis informatics has arisen in response. Researchers of crisis informatics investigate the nature of socio-behavioral phenomena in mass emergency mediated by social media environments and devise new methods for its investigation

Who are the stakeholders related to the outcome of the research?
First responders, Formal Response agencies, Humanitarian organizations, general public, local police or area firefighters.

What are some examples for social media entries that relates to this research?

Following shows some sample entries quoted from research papers.

- "OMG! The fire seems out of control: It's running down the hills!" (bush fire near Marseilles, France, in 2009, quoted from Twitter in De Longueville et al. [2009])
- "Red River at East Grand Forks is 48.70 feet, +20.7 feet of flood stage, −5.65 feet of 1997 crest. #flood09" (automatically-generated tweet during Red River Valley floods in 2009, quoted from Twitter in Starbird et al. [2010])

- "Anyone know of volunteer opportunities for hurricane Sandy? Would like to try and help in any way possible" (Hurricane Sandy 2013, quoted from Twitter in Purohit et al. [2013])

- "My mom's backyard in Hatteras. That dock is usually about 3 feet above water [photo]" (Hurricane Sandy 2013, quoted from Reddit in Leavitt and Clark [2014])

- "Sirens going off now!! Take cover ... be safe!" (Moore Tornado 2013, quoted from Twitter in Blanford et al. [2014])

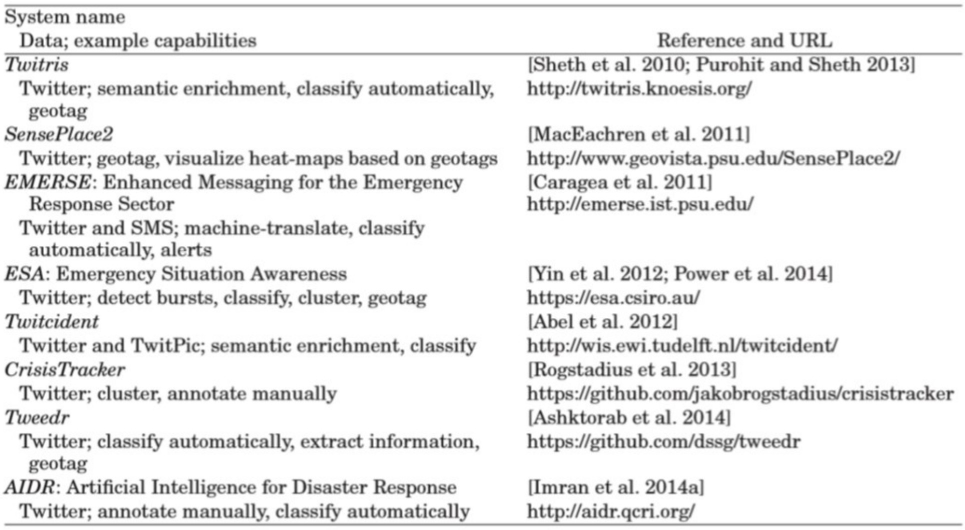
As we can see those quoted entries provides valuable information regarding the disasters from eyewitnesses and volunteers.

Are there existing systems already?

The answer is yes. Here is a comparison of existing systems extracted from [2].

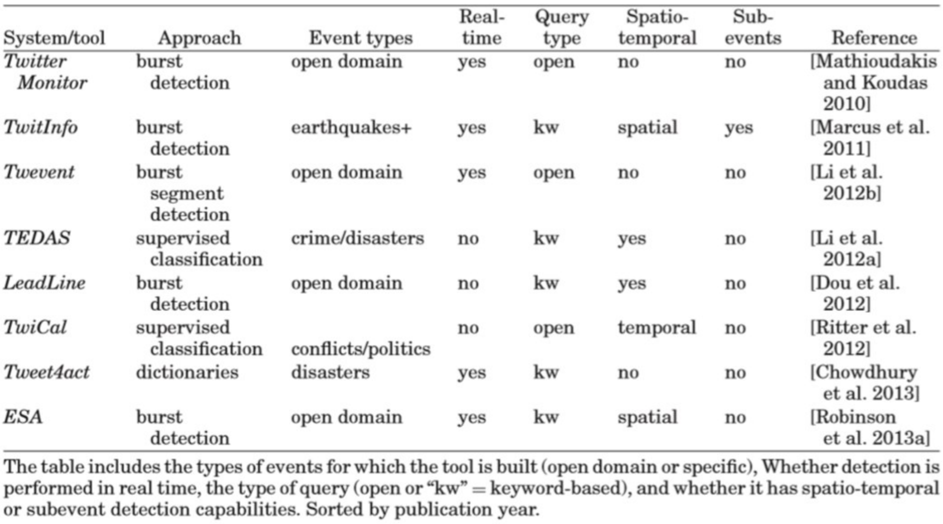
It is important to notice almost all of the existing systems are based on popular microblogging social networking platform called Twitter which allows users to publish entries with a maximum length of 140 characters.

Table 1 Existing Systems and where to find them



System name Data; example capabilities	Reference and URL
Twitris	[Sheth et al. 2010; Purohit and Sheth 2013]
Twitter; semantic enrichment, classify automatically, geotag	http://twitris.knoesis.org/
SensePlace2	[MacEachren et al. 2011]
Twitter; geotag, visualize heat-maps based on geotags	http://www.geovista.psu.edu/SensePlace2/
EMERSE: Enhanced Messaging for the Emergency Response Sector	[Caragea et al. 2011] http://emerse.ist.psu.edu/
Twitter and SMS; machine-translate, classify automatically, alerts	
ESA: Emergency Situation Awareness	[Yin et al. 2012; Power et al. 2014]
Twitter; detect bursts, classify, cluster, geotag	https://esa.csiro.au/
Twitcident	[Abel et al. 2012]
Twitter and TwitPic; semantic enrichment, classify	http://wis.ewi.tudelft.nl/twitcident/
CrisisTracker	[Rogstadius et al. 2013]
Twitter; cluster, annotate manually	https://github.com/jakobrogstadius/crisistracker
Tweedr	[Ashktorab et al. 2014]
Twitter; classify automatically, extract information, geotag	https://github.com/dssg/tweedr
AIDR: Artificial Intelligence for Disaster Response	[Imran et al. 2014a]
Twitter; annotate manually, classify automatically	http://aidr.qcri.org/

Table 2 Methods Used in Existing Systems



System/tool	Approach	Event types	Real-time	Query type	Spatio-temporal	Sub-events	Reference
Twitter Monitor	burst detection	open domain	yes	open	no	no	[Mathioudakis and Koudas 2010]
TwitInfo	burst detection	earthquakes+	yes	kw	spatial	yes	[Marcus et al. 2011]
Twevent	burst segment detection	open domain	yes	open	no	no	[Li et al. 2012b]
TEDAS	supervised classification	crime/disasters	no	kw	yes	no	[Li et al. 2012a]
LeadLine	burst detection	open domain	no	kw	yes	no	[Dou et al. 2012]
TwiCal	supervised classification	conflicts/politics	no	open	temporal	no	[Ritter et al. 2012]
Tweet4act	dictionaries	disasters	yes	kw	no	no	[Chowdhury et al. 2013]
ESA	burst detection	open domain	yes	kw	spatial	no	[Robinson et al. 2013a]

The table includes the types of events for which the tool is built (open domain or specific), Whether detection is performed in real time, the type of query (open or "kw" = keyword-based), and whether it has spatio-temporal or subevent detection capabilities. Sorted by publication year.

What methods can be used to depict the output automated processing?

Common elements in displaying output according to the authors of [2].
Lists/timelines showing recent or important messages, sometimes grouping the messages into clusters or categories.
Time series graphs representing the volume of a hashtag, word, phrase, or concept over time, and sometimes marking peaks of activity.
Maps including geotagged messages or interpolated regions, possibly layered according to different topics.
Pie charts or other visual summaries of the proportions of different messages

1.2. Research Gap

Recently it has given a lot of attention for the usage of social media in various aspects of industries. Since the beginning of the "Information Era" social media are a proven method in digital marketing and advertising it has helped the small businesses to grow. Social media monitoring and regulation is another topic that come to life time which is taking limelight at a slow pace.

Usage of social media for different reasons other than financial benefits is somewhat ignored comparingly. One might think that there are no other goals that can be achieved but simply that is not true. Information is powerful. Proper use of information will result in valuable outcomes. In 2018 March it has revealed that millions of Facebook user information has been used illegally to generate manipulative messages to influence voters in America during elections (Cambridge Analytica incident). User behavior analysis which is used in e-commerce sites to suggest products is another example which shows the power of information and big data.

Most researches conducted in social media usage related to disasters or in other words emergencies have used the popular microblogging platform Twitter which provides a streaming API to collect publicly available entries (Posts) each maximum length of 140 characters in real time. There is so much information generated elsewhere other than Twitter. For instance, forums blogs dedicated channels for disaster response.

Among the techniques used to filter the entries that are related to some specific event provided hashtags (for example #earthquakes) keyword filtering are more common. When identifying a trend (Trend analysis through social media) some systems use word count mechanisms and give the most repeated words as an output. Limitations in streaming API (Maximum number of requests per minute) slows down the process increasing the Latency.

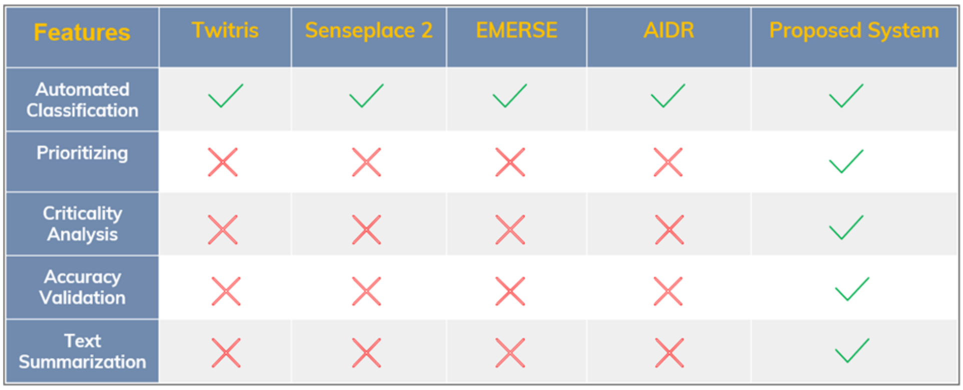
Although mechanisms have provided to filter out entries for a given event, the ability for the existing systems to evaluate the accuracy or the dependability of an entry is limited. Systems that use mix of human interaction and computation power are called "Hybrid systems". They use crowdsourcing to create a model to filter the future entries.

1.3. Research Problem

Tough social media is practically and widely used in financial business-oriented scenarios applications for other purposes are scarce Increasing widespread use, popularity and large user base of social media had lead the way for researchers to identify various other uses of social media platforms. In fact, there is a lot of work to be done for the context of social media usage in an emergency.

Some organizations and government agencies have identified the use of social media as an important role in emergency response. For example, American Red Cross has deployed so called Digital Response Center in order to provide situational awareness information and help who are in need. Due to the lack of manpower, lack of funds to conduct proper research and criticality of a situation stakeholders believe that it is resource wasting unachievable task

The task of processing social media entries requires new means of information filtering, classifying and summarization. The lacking feature of most current systems available is the accuracy and the dependability of a given entry. Hybrid systems highly depend on crowdsourcing which requires volunteers so called digital volunteers. This affects the latency of the process. Existing systems are highly dependent on the Twitter. Extracting data from numerous sources other than Twitter streaming API is a challenging task to be completed. The unstructured data needs to be cleaned in order to be used in other stages. Finding appropriate optimal number of categories to match the requirements of different parties (Organization, Government agencies etc.), identifying ways of calculating accuracy levels for entries, defining thresholds and finding the criticality of situation are major research areas which would be covered throughout the research project.



Features	Twitris	Senseplace 2	EMERSE	AIDR	Proposed System
Automated Classification	✓	✓	✓	✓	✓
Prioritizing	✗	✗	✗	✗	✓
Criticality Analysis	✗	✗	✗	✗	✓
Accuracy Validation	✗	✗	✗	✗	✓
Text Summarization	✗	✗	✗	✗	✓

Figure 1.2 Difference Between Existing Systems and Proposed System

1.4. Research Objectives

1.4.1. Main Objective

The main intention of this project is to deliver an open source web application which can generate social media posts in to meaningful information during emergency situation and first responders can give effective and efficient respond to emergency and saving life.

1.4.2. Specific Objective

- To identify major events happening around the world.
- To filter relevant post for one disaster

- To give validity level for the entry post which will make information more accurate
- To give criticality level for the entry post which will allow first responders to give more efficient respond to where, their help most wanted.
- To summarize the content so that only valuable information will view

Apart from above mention objectives following intentions could be identified

- To minimize the human loss during disaster

With this application first responders and rescue team can get information about critical site faster and attend to the most wanted sites first.

- To find help for the people who need help
- To give information for volunteers who are willing to help
- To find loved once are safe

2. METHODOLOGY

This chapter illustrates the methodology for handling the proposed tool. It’s a well-ordered approach to the research, gathering requirements, designing and implementation to create effective solution to an existing problem an area where improvement is required.

2.1. General Approach

Defining the problem.

Discussions were subsequently held with members and supervisors to properly define and assess problem to identify the scope. Initial problem was reduced to match with the time given and added extra functionality to match the required complexity. Next step was to identify the potential stakeholders.

Background Research / Literature Review

Once the problem was defined and stakeholders were identified, background research was required to distinguish what exists already, what are the ongoing projects and what they are lacking. This step was important as it gave us an overview of how other researchers has conducted their and what their outcome was.

Identifying the outcomes, defining objectives and goals.

After analyzing what needs to be implemented outcomes objectives and goals were defined. Next schedule was developed which contains the major milestones.

Time to Time Verification

Regular discussions with members and supervisors were conducted to verify the followed path and the methodology was valid as decided in prior stages. When unexpected problems occurred during the process they were discussed and resolved as they came.

Documentation

More high-level journal was maintained throughout the research including daily work done. SRS document containing all the requirements and a proposal with the proposed solution was produced as main design documents.

2.2. Machine Learning Approach

Machine learning is the process of finding patterns in data to predict potential outcome. Following process was followed when applying machine learning algorithms to the given problem. For machine learning it requires a lot of data, computing power and effective machine learning algorithms. All of these are now available more than ever thanks to the information era. Applying machine learning techniques to Business purpose is a common real-world use case. Some problems require domain knowledge. But in this case the required domain knowledge was less.

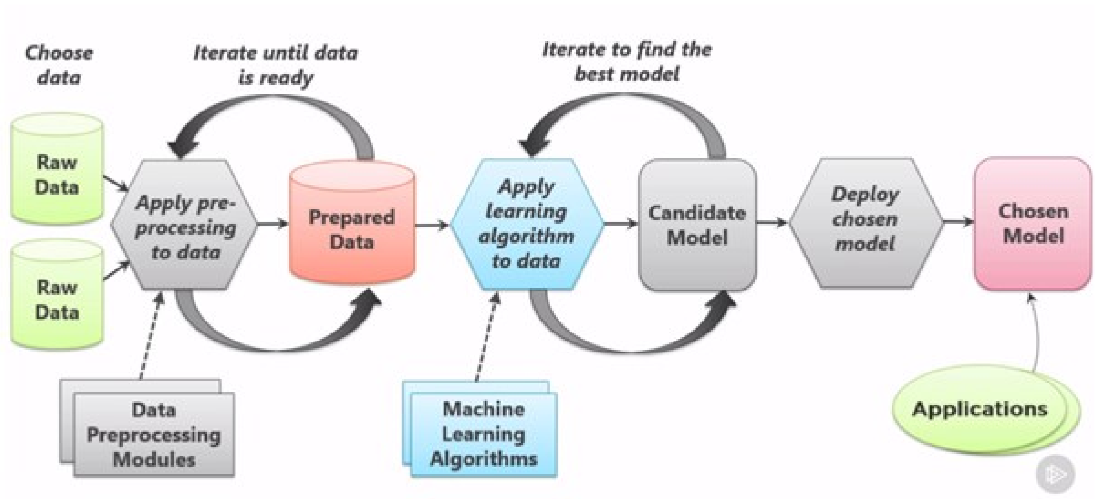

Choose data
Iterate until data is ready
Iterate to find the best model
Raw Data
Raw Data
Apply pre-processing to data
Prepared Data
Apply learning algorithm to data
Candidate Model
Deploy chosen model
Chosen Model
Data Preprocessing Modules
Machine Learning Algorithms
Applications

Figure 2.1 Process of Developing a Model Using Machine Learning

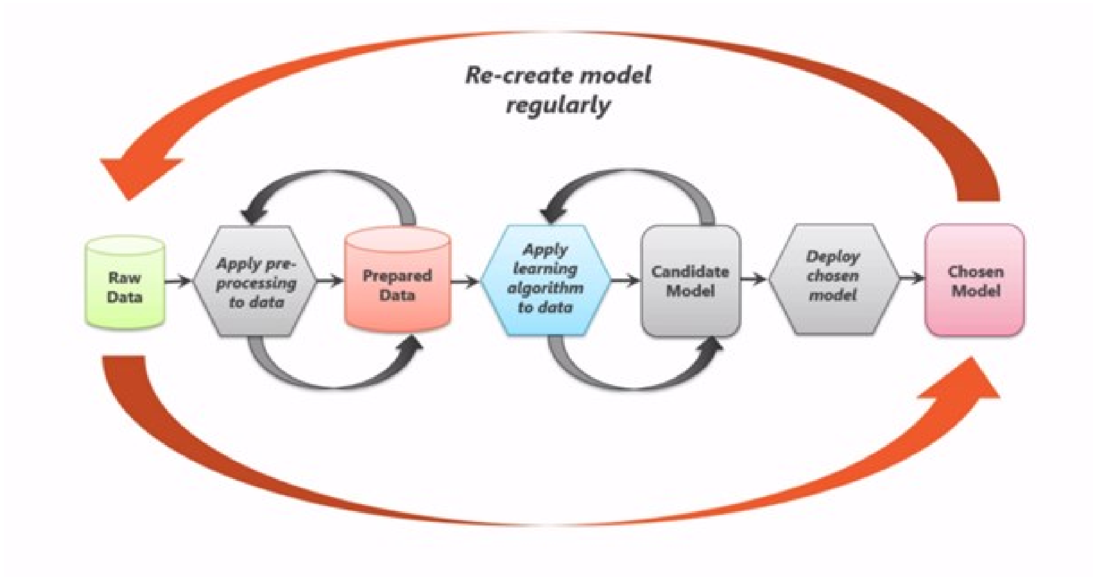

Re-create model regularly
Raw Data
Apply pre-processing to data
Prepared Data
Apply learning algorithm to data
Candidate Model
Deploy chosen model
Chosen Model

Figure 2.2 Iterative Process of Machine Learning

2.2.1. Choosing the right data/ collecting data

Since machine learning is so much dependent on data. Usually the data collected in the csv format. A good clean dataset usually results in a better trained model with higher accuracy. There are several methods which can be used to preprocess data.

Real world data are generally

- **Incomplete:** lacking attribute values, lacking certain attributes of interest, or containing only aggregate data
- **Noisy:** containing errors or outliers
- **Inconsistent:** containing discrepancies in codes or names

Tasks in data preprocessing

- **Data cleaning:** fill in missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies.
- **Data integration:** using multiple databases, data cubes, or files**.**
- **Data transformation:** normalization and aggregation**.**
- **Data reduction:** reducing the volume but producing the same or similar analytical results.

- **Data discretization:** part of data reduction, replacing numerical attributes with nominal ones.

Data transformation

1. Normalization:

   - Scaling attribute values to fall within a specified range. Example: to transform V in [min, max] to V' in [0,1], apply V'=(V-Min)/(Max-Min)
   - Scaling by using mean and standard deviation (useful when min and max are unknown or when there are outliers): V'=(V-Mean)/Standard Deviation

2. Aggregation: moving up in the concept hierarchy on numeric attributes.
3. Generalization: moving up in the concept hierarchy on nominal attributes.
4. Attribute construction: replacing or adding new attributes inferred by existing attributes

Data reduction

1. Reducing the number of attributes

- Data cube aggregation: applying roll-up, slice or dice operations.
- Removing irrelevant attributes: attribute selection (filtering and wrapper methods), searching the attribute space.
- Principal component analysis (numeric attributes only): searching for a lower dimensional space that can best represent the data

2. Reducing the number of attribute values

2.2.2. Training a Model

Learning requires

Identifying patterns

Recognizing those patterns when you see them again

Same applies for a machine to learn it requires past data. Depending on the historical data algorithm builds a model which will be used to label (in supervised learning) unseen new data.

2.2.3. Evaluating the model

Cross Validation

Cross validation attempts to avoid overfitting (training on and predicting the same datapoint) while still producing a prediction for each observation dataset. This is accomplished by systematically hiding different subsets of the data while training a set of models. After training, each model predicts on the subset that had been hidden to it, emulating multiple train-test splits. When done correctly, every observation will have a 'fair' corresponding prediction.

3. IMPLEMENTATION

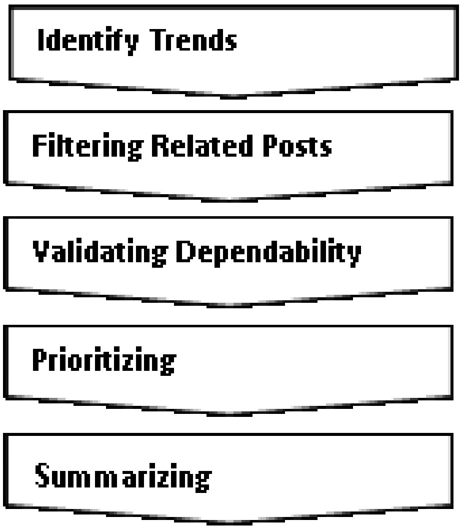


Figure 3.1 Flow of the Process

A. Identify Trends

Trend identification helps to detect major events happening around the world. To achieve this, we propose a threshold-based algorithm which will give its output as word clouds. The generic way is to count the most frequent words from text. When applied Bag of words algorithm for text we get the words and their frequencies. A word cloud (example Fig 3) can be generated hourly or minute by minute for better clarity. But to do this it costs more computational power.

B. Filtering Related Posts (Social Media Entries)

Entry Filter is responsible for identifying (Filtering) relevant posts. During a natural disaster relevant information are generated by eyewitnesses, people who are affected (victims), government and humanitarian groups/ agencies, people who wants to donate and people who are willing to take part in voluntary services. Processing these entries in a timely manner is a reoccurring matter in the disaster management aspect as this information get obsolete very quickly. To make use of the valuable information included in the entries (Posts) and to realize or visualize information, this component is introduced to the system. Filtering is a critical sub component since it is the starting point of the flow of automated processing for social media entries.

C. Validation

Social media is one of the major ways of dispersing information during a disaster situation. Sharing false information would lead to bad consequences, so that the validation of dependability of social media posts is crucial in a natural disaster. Studies have found that outdated, inaccurate, or false information has been disseminated via social media forums during disasters [8]. In some cases, the location of the hazard or threat was inaccurately reported. In the case of the March 2011 Japanese earthquake and tsunami, tweets for assistance were retweeted after the victims had been rescued [9]. The level of accuracy determines the level of dependability and trustworthiness of a piece of information extracted using the data generated during an emergency. This also involves analyzing metadata of entries such as number of comments shares and likes. If follow up comments are supporting the main entry them, it can be considered as an entry with higher dependability. Validation is a huge concern due to the misuse of social media to disperse of false information Posts that fails from this stage will not move to the next stage.

D. Prioritizing/ Finding Critical Level

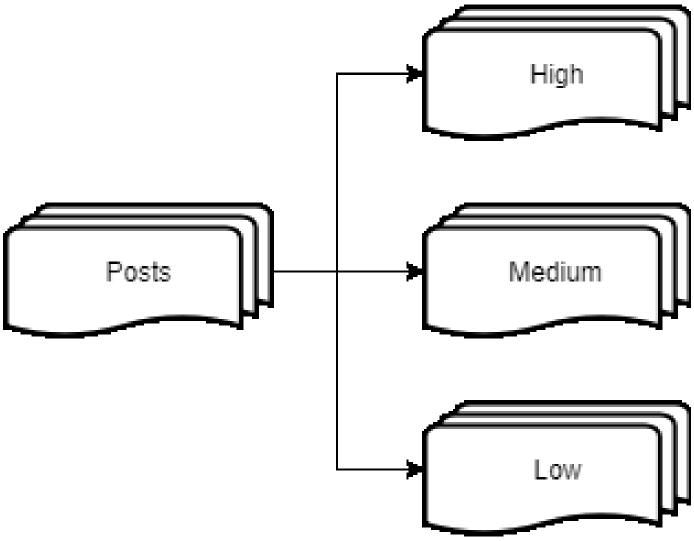


Figure 3.2 Posts Labeled Depending on the Critical Level

The evaluation of sentiment and the extent of effect of a post or the level of criticality will be important when referring to this kind of scenario. For example, if a post contains a request for immediate help that can be considered as a high priority. Intention of this component is to help the first responders with their critical decision making. Most of the time they must make important decisions rapidly during an emergency. They must decide what needs to be done, which issue needs to be attended first. So, by giving a level of importance high, medium or low, they will be able to get a chance to attend high priorities by spending less time. This is done using three separate datasets with labels “Low”, “Medium” and “High” and training a model using supervised machine learning.

E. Summarization

Summarization is focused on identifying the core meaning of a post and extracting the summarized content over bulk of social media posts. It is important to maintaining the core meaning of a post without damaging the actual meaning of it. This volume of text is an invaluable source of information and knowledge which needs to be effectively summarized to be useful [7] for the natural disaster supporting teams to take actions timely, effectively and efficiently.

Automatic text summarization is challenging, since as humans we summarize a piece of text by reading it entirely to develop our understanding, and then write a summary highlighting its main points. Since computers lack human knowledge and language capability, it makes automatic text summarization a very difficult and non-trivial task [7]. According to Radef et al. [6] a summary is defined as "a text that is produced from one or more texts, that conveys important information in the original text(s), and that is no longer than half of the original text(s) and usually, significantly less than that". Automatic text summarization is the task of producing a concise and fluent summary while preserving key information content and overall meaning. [7]

Text summarization approaches is divided into two groups: extractive and abstractive summarization. Extractive summarizations extract important sentences or phrases from the original documents and group them to produce a summary without changing the original text. Abstractive summarization consists of understanding the source text by using linguistic methods to interpret and examine the text. Abstractive methods need a deeper analysis of the text. These methods can generate new sentences, which improves the focus of a summary, reduce its redundancy and keeps a good compression rate.

Summaries produced by extractive summarization techniques are constructed by choosing a subset of sentences in the original text which is being the input for the text summarizer. The chosen sentences are supposed to be most important sentences of the input text corpus. According to the context of the research, the input text could be a social medial post with follow up comments. Extractive methods tend to be verbose and this is especially problematic as produced summaries should not be lengthy and be readable for natural disaster supporting teams. Thus, an informative and concise combination of extractive and abstractive summary is a better solution

System Architecture

EC2 (Elastic Cloud Compute) are Ubuntu Servers running on AWS cloud. The algorithm for filtering/ identifying, validation, finding critical level are written as a Python scripts runs on these servers as microservices.

All the tweets identified as related from the algorithm are saved in the DynamoDB. To access DynamoDB a RESTful service is developed using serverless. Serverless applications are helpful since they are managed by AWS (we don't have to worry about maintaining servers AWS does it for us).

Process starts when there are tweets available from Twitter Streaming API (is a freely available but limited service which provides tweets near real time). For more information about DynamoDB refer to APPENDIX A.

Global Secondary Index (GSI) has been created to reduce latency and to query using different partition key and sort key. DynamoDB Streams allow us to listen to DB operations such as INSERT, DELETE and UPDATE as it happens. Connected to stream is a lambda function which triggers for prior mentioned events. This lambda function calls other service components depending on the event type. Summarization happens on demand of the user through the web application.

Cognito user pool provides authentication services for the API. Web application is hosted on S3 (Simple Storage Service) and Served through CloudFront Distribution. In DynamoDB attribute for Time to Live is created in order to remove data automatically after expiration.

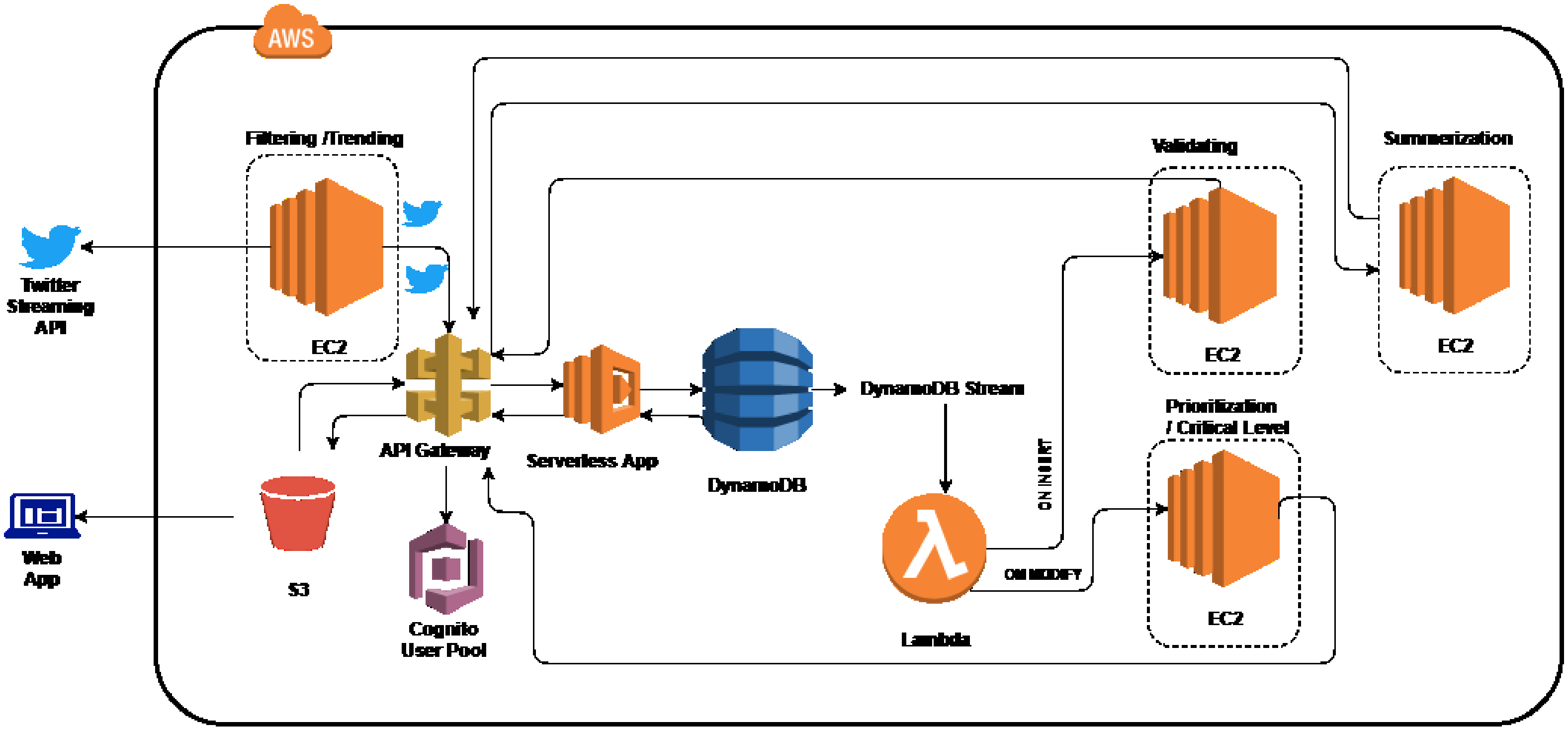


Figure 3.3 System Architecture

Tools and Technologies

Neo4J graph platforms, DynamoDB NoSQL databases with their own characteristics. It is known for its structure free Visual studio code is a general purpose integrated development environment which is freely available. PyCharm, Jupyter notebook are Integrated Development Environments for Python programming language.

NLTK (Natural language toolkit) is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum.

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms and can be freely distributed.

H**TML5, CSS3, Bootstrap 4, Angular, Cyper** are web technologies (Frameworks and markup languages) used for developing front end user interfaces.

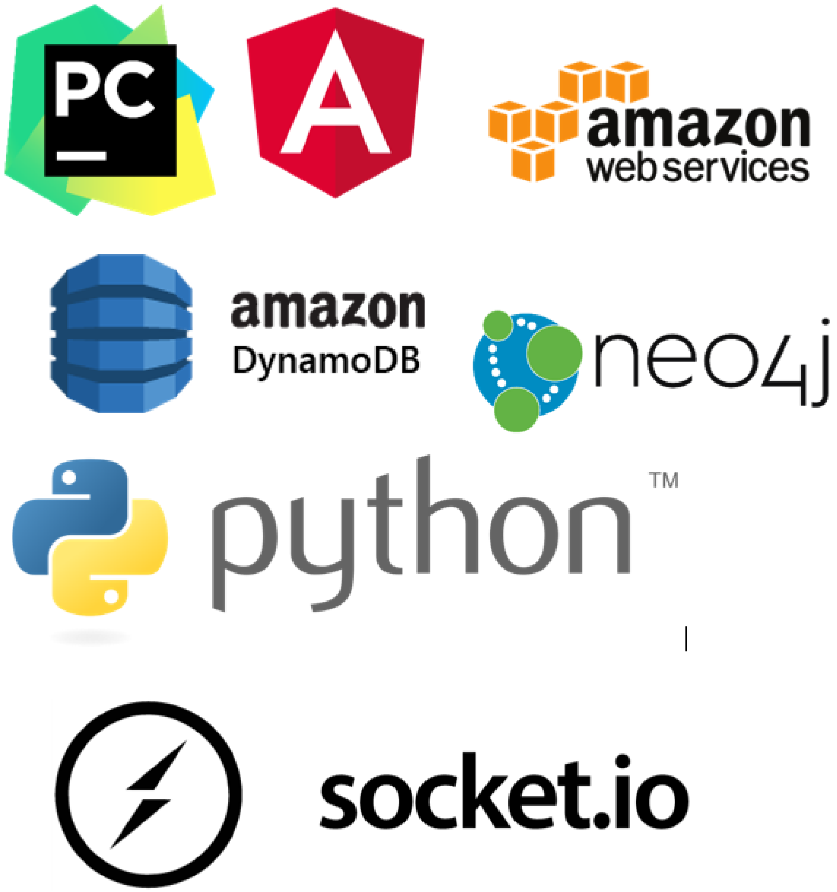


Figure 3.4 Tools and Technologies Used

System Constraints

Hardware Constraints

- Compatible with tablet pc and personal computer / laptop with required specifications function as means of displaying frontend which is more specific to the service providers who are intended to consume the use of proposed application programming interface (API). The latter mentioned hardware interfaces should have the capability of connecting to the internet which is supposed to be high speed and bandwidth while required web browsers are installed.

Communication Boundaries

- Modem (Built in GPRS/EDGE/3G/4G) and Wi-Fi Router - Used to connect to the internet to access web services.

Memory Constraints

The system will be developed as microservices for each module or component, deployed on Amazon Web Services (AWS). Memory and storage usages will be allocated according to the consumption of resources of proposed system. Virtual server instances for microservices will be initiated with 1GB of RAM and 1GB of storage.

2.1 Commercialization aspects of the product

The proposed application will be highly useful for disaster victims, all levels of government, first responders. Apart from that we can include general public, business community, media, elected officials, community officials, NGOs, law enforcement, medical communities, scientific communities, volunteer groups as our audience. This is a more of service than a business.

Accordingly, many significant benefits could be found in our product

- Give accurate up to date information

Testing

The main components of a web application testing strategy include usability, performance, security, functional and nonfunctional and database testing across multiple platforms, devices and browsers. A complete web application strategy must also consider testing across differing network connection speeds and geographical locations, as well as address the use of Wi-Fi, 3G or 4G connections. Testing must confront such issues as screen resolution and brightness,

CPU, memory and OS optimization.

A typical end-to-end web testing process, should start from creating test cases of the application, performing user acceptance

Table 3 Different Ways of Testing

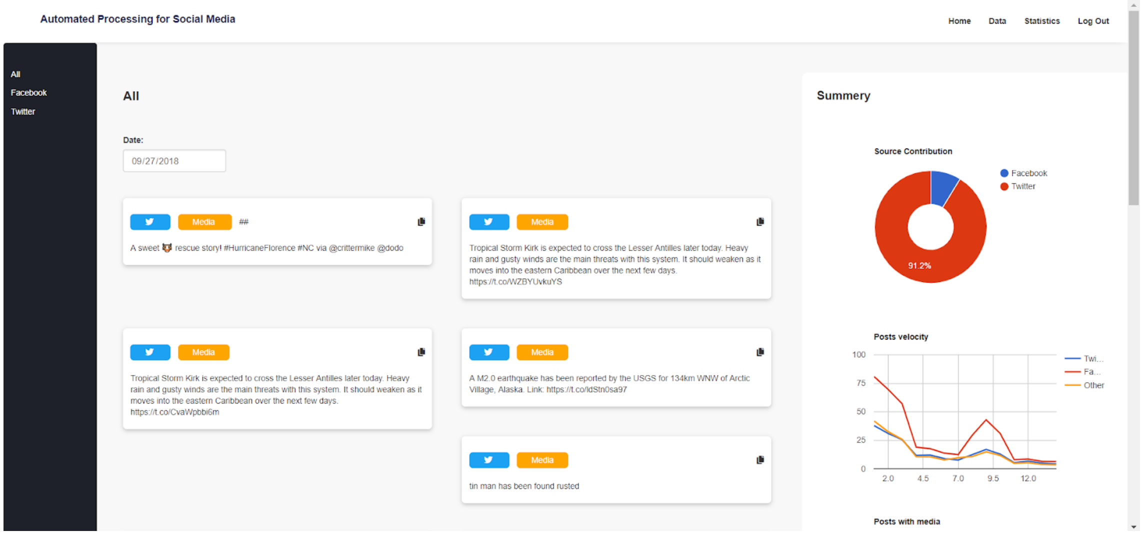
Prepare Test Case Specification	**Start the testing with writing test case specification for the system.**
Manual and Automate Testing	Execute test cases manually and automation.
Functional Testing	Test system against the functional requirements, feeding inputs and examine output.

Usability Testing	Test system from external users to get feedback for overall user experience.
Interface Testing	Interface testing ensures that all interactions between the web server and application server interfaces are running smoothly.
Compatibility Testing	Test application to find it’s compatible with all browsers and devices.
Performance Testing	Test the performance of the web application for its responsiveness, scalability, resource usage and stability based on standards.
Security Testing	Test that product is secure from harmful actions.
Database Testing	Test for data integrity while inserting updating database.

4. DISCUSSION

4.1 Results and research findings

The research Automated processing for social media data in a mass emergency is composed with four main components which are processed sequentially to generate satisfactory results for the individuals who are expecting. The process will be starting from the component which supposed to retrieve social media data from the cloud but according to the current status of the development of the research, social media posts will be only taken from Twitter using Twitter Stream API and processed to filter out relevant Twitter posts for the context of mass emergencies as the first phase of it. The social media data will be fed to the system in real time basis and each filtered social media post could be a tweet or a retweet according to the Twitter context. Later mentioned results will be published on Live Data Stream section of the frontend as a collection of cards and each card will be represented as a tweet card which refers to an individual tweet post. Each tweet card will only represent an individual tweet post or a root post according the research context and retweets which are supporting for a particular tweet will be taken as comments according to the research context while retweets for a particular tweet will not be displayed on a tweet card on Live Data Stream section.



Date: 09/27/2018

A sweet 🦊 rescue story! #HurricaneFlorence #NC via @crittermike @dodo

Tropical Storm Kirk is expected to cross the Lesser Antilles later today. Heavy rain and gusty winds are the main threats with this system. It should weaken as it moves into the eastern Caribbean over the next few days. https://t.co/WZBYUvkuYS

Tropical Storm Kirk is expected to cross the Lesser Antilles later today. Heavy rain and gusty winds are the main threats with this system. It should weaken as it moves into the eastern Caribbean over the next few days. https://t.co/CvaWpbbi6m

A M2.0 earthquake has been reported by the USGS for 134km WNW of Arctic Village, Alaska. Link: https://t.co/ldStn0sa97

tin man has been found rusted

Figure 4.1 Filtered Tweets

As the second phase of the first component of the process chain, twitter feeds will be gone through a process of identifying trends of current filtered social media posts and generates word clouds accordingly. Word clouds will be generated for a taken set of social media posts and identifies posts which contain words that are related to the context of natural disasters. The words clouds generated by the system will be very useful for end users as it supports for taking direct insights of the discussion going by received social media posts. In the word cloud, highly repeating words will take higher count and the count which refers to a particular word will be varied according to the repeating manner of that word. When referring to the word cloud, the words which takes higher count will be displayed with varied font sizes accordingly and words with increased font sizes will point out the important words of the discussion. In the system, word clouds will be generated for every period of hour using the collected filter social media posts which are fetched with in that period of time. So that the end user will be able to gain the sights by hourly basis.

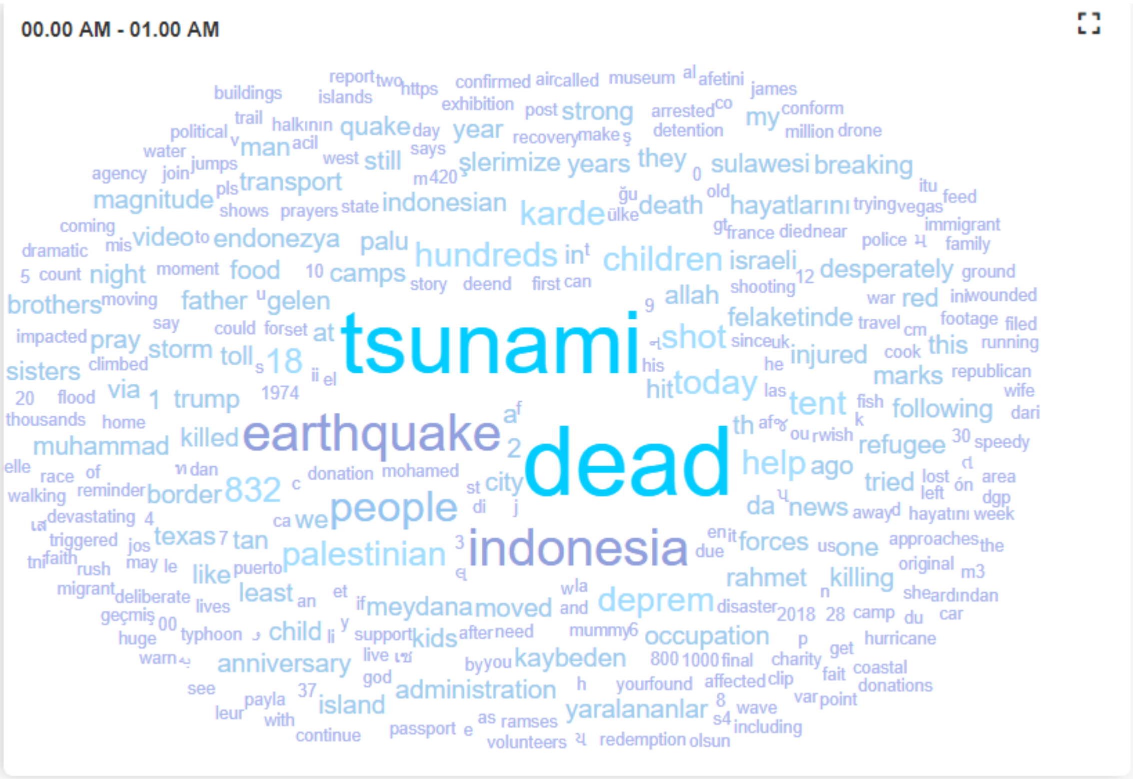
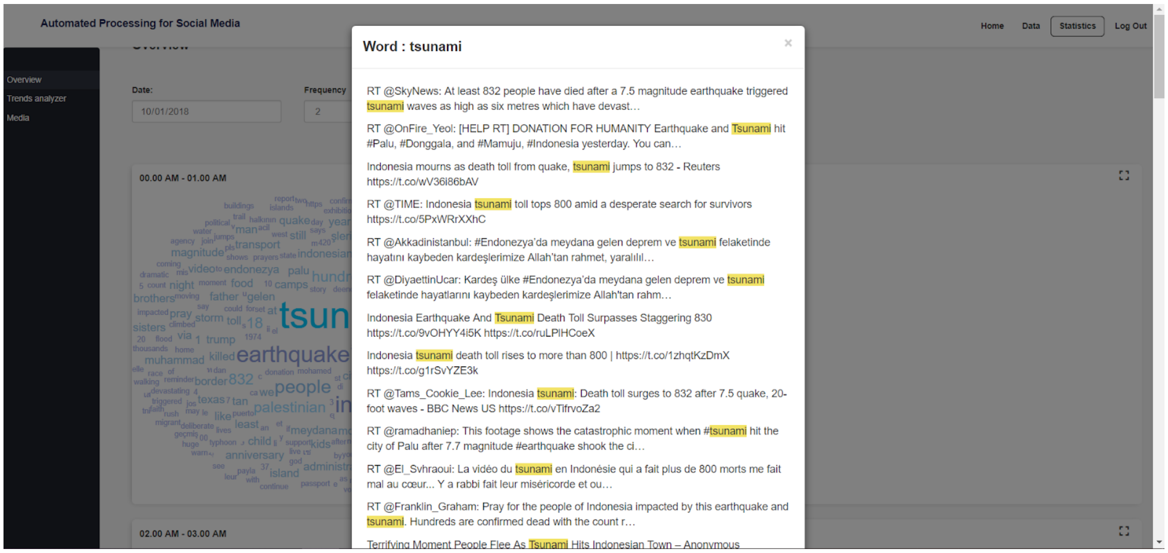

00.00 AM - 01.00 AM
report two https confirmed aircalled museum al afetini
buildings islands exhibition james
trail post strong arrested co my conform
political halkının quake day year recovery make ş detention million drone
water v man acil says
agency join jumps west still şlerimize years they 0 sulawesi breaking
pls transport m420 s itu
magnitude shows prayers state indonesian karde ğu ülke death old hayatlarını trying vegas feed
coming gt france diednear immigrant
dramatic mis video to endonezya palu hundreds in t children israeli police family
5 count night moment food 10 camps story deend first can 12 desperately ground
brothers moving father u gelen 9 allah shooting war red ini wounded
say could forset at felaketinde travel footage filed
impacted pray storm toll s 18 tsunami shot since uk. cm this running
sisters climbed ii el his he injured cook marks republican
20 flood via 1 trump 1974 hit today las tent fish wife
thousands home af th af k following dari
muhammad killed earthquake 2 dead ou rwish refugee 30 speedy
elle race of dan donation mohamed help ago
walking reminder border 832 c st city tried lost left ón area
devastating 4 ca we people di j da news away d hayatını dgp week
triggered jos texas 7 tan 3 indonesia en it forces us one approaches the
tni faith rush may le like puerto palestinian due original
migrant deliberate least an et rahmet killing m3
geçmiş lives if meydana moved w la and deprem disaster 2018 28 camp n she ardından
00 huge typhoon child li y support kids after need mummy6 occupation p du car
warn anniversary live byyou kaybeden 800 1000 final charity get hurricane
see god administration h yourfound affected clip fait coastal donations
leur payla 37 island yaralananlar 8 var point
with passport e as ramses s4 wave including
continue volunteers redemption olsun

Figure 4.2 Word Cloud Generated After Indonesian Tsunami (2018 Sept)



Date: 10/01/2018

Frequency 2

Word : tsunami

RT @SkyNews: At least 832 people have died after a 7.5 magnitude earthquake triggered tsunami waves as high as six metres which have devast…

RT @OnFire_Yeol: [HELP RT] DONATION FOR HUMANITY Earthquake and Tsunami hit #Palu, #Donggala, and #Mamuju, #Indonesia yesterday. You can…

Indonesia mourns as death toll from quake, tsunami jumps to 832 - Reuters https://t.co/wV36i86bAV

RT @TIME: Indonesia tsunami toll tops 800 amid a desperate search for survivors https://t.co/5PxWRrXXhC

RT @Akkadinistanbul: #Endonezya'da meydana gelen deprem ve tsunami felaketinde hayatını kaybeden kardeşlerimize Allah'tan rahmet, yaralılıl…

RT @DiyaettinUcar: Kardeş ülke #Endonezya'da meydana gelen deprem ve tsunami felaketinde hayatlarını kaybeden kardeşlerimize Allah'tan rahm…

Indonesia Earthquake And Tsunami Death Toll Surpasses Staggering 830 https://t.co/9vOHYY4i5K https://t.co/ruLPlHCoeX

Indonesia tsunami death toll rises to more than 800 | https://t.co/1zhqtKzDmX https://t.co/g1rSvYZE3k

RT @Tams_Cookie_Lee: Indonesia tsunami: Death toll surges to 832 after 7.5 quake, 20-foot waves - BBC News US https://t.co/vTifrvoZa2

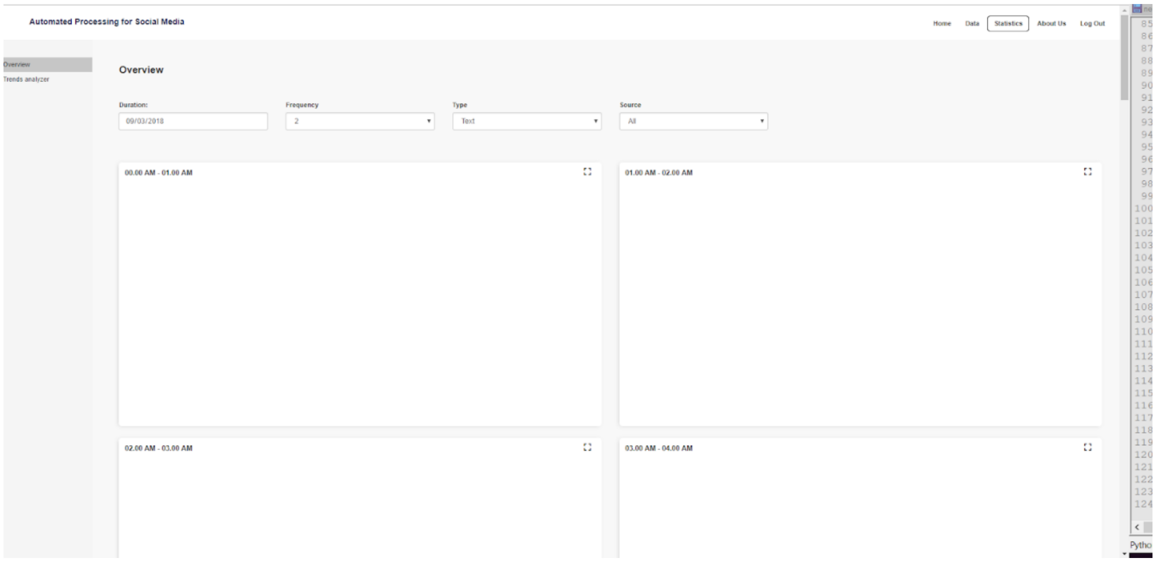
RT @ramadhaniep: This footage shows the catastrophic moment when #tsunami hit the city of Palu after 7.7 magnitude #earthquake shook the ci…

RT @El_Svhraoui: La vidéo du tsunami en Indonésie qui a fait plus de 800 morts me fait mal au cœur... Y a rabbi fait leur miséricorde et ou…

RT @Franklin_Graham: Pray for the people of Indonesia impacted by this earthquake and tsunami. Hundreds are confirmed dead with the count r…

Terrifying Moment People Flee As Tsunami Hits Indonesian Town – Anonymous

Figure 4.3 Detailed View of a Word Selected From a Word Cloud



Duration: 09/03/2018

Frequency 2

Type Text

Source All

Figure 4.4 Trending User Interface Where Word Clouds are Generated

In the second component of the process chain, each social media post will be validated by using the comments which have been received for that post. According to the implementation of this research, each twitter feed will be validated by the retweets received to that twitter feed. As a twitter feed to be validated by retweets received, a tweet should have one or more than retweets bind with it. If we assume, that there is a twitter feed which do not have retweets yet, the component will not be able to give an feedback on that twitter feed, but once the tweet receives retweets by the time, validating process will be performs for each time when the tweet gets retweets. As this real time behavior of the system, a tweet which has been validated in a way could be changed by the retweets received in the future as if recently received retweets are suggesting that the tweet which has been validated as true will be false. When the system is unable to give a specific feedback on a tweet in the absence of retweets for that tweet, a pending flag will be added on the tweet card in the Live Data Stream panel. So that validating process of social medial post will add three types of flags on a tweet card as a pending flag for showing that the system is unable to give an valid feedback on that tweet, a true flag for showing that a particular tweet is currently correct and a false flag for showing that a particular tweet is giving false information for the users.

In the third component of the process chain, each social media post will be priorities according to three levels to obtain the idea of taking immediate actions or addressing a particular event immediately by the end users of this system. This process will be joining hand with the process of second component of validating social media posts as this component will only do the process on social media posts which are validated as true by the second component or social media posts which are labeled as pending.

The fourth and final component of the process chain will be function in a isolate manner. This component will do the automatic text summarization on a given set of social media posts by the end user. On the process of automatic text summarization, this component will follow extractive summarization for identifying subset of social media posts which are more relevant and important over given set of social media posts. For generating a summary for a given set of social media posts, the user should be on the Live Data Stream panel and on that page, there is a button called summary and by clicking on that, automatic text summarization will be performing on currently available social media posts and selects a subset of social media posts as follows.

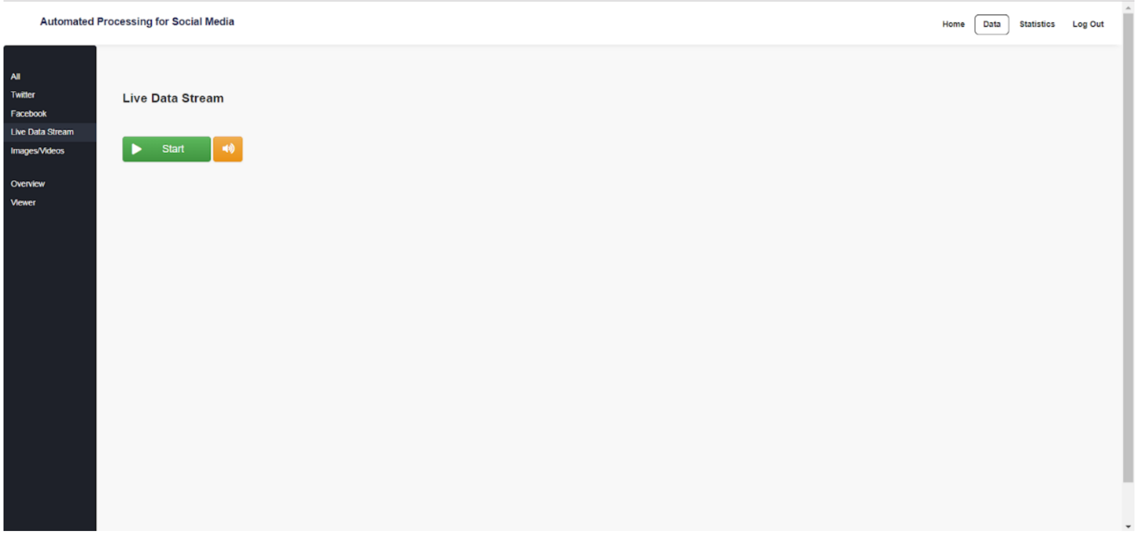
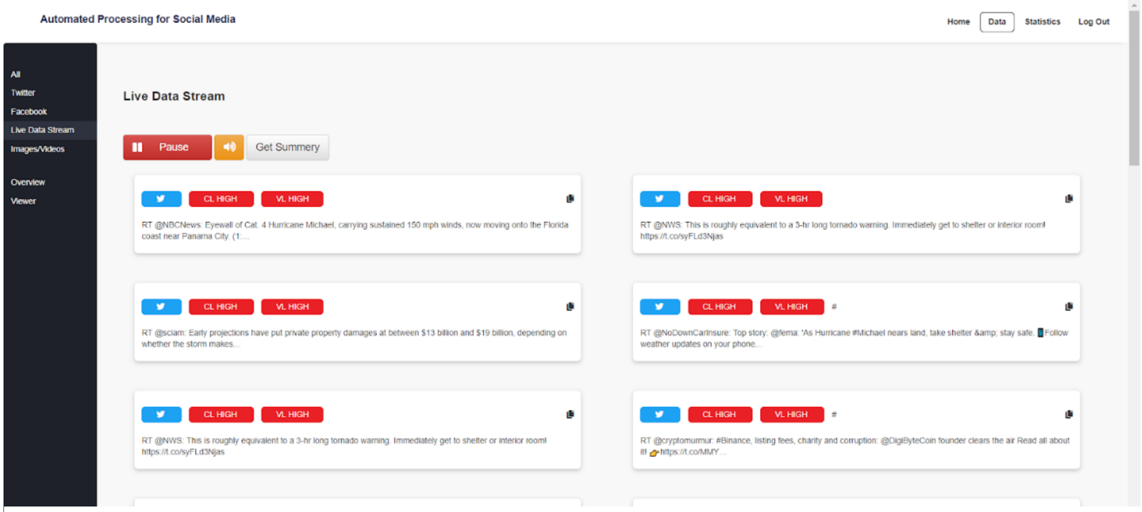


Figure 4.5 User Interface Where Live Data Will Be Populated



RT @NBCNews: Eyewall of Cat. 4 Hurricane Michael, carrying sustained 150 mph winds, now moving onto the Florida coast near Panama City. (1:…

RT @NWS: This is roughly equivalent to a 3-hr long tornado warning. Immediately get to shelter or interior room! https://t.co/syFLd3Njas

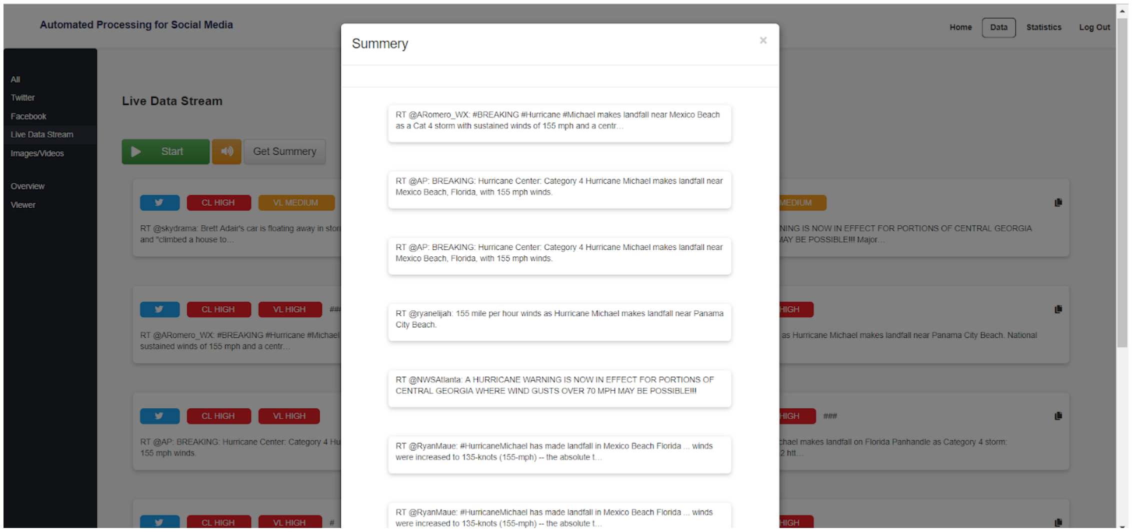
RT @sciam: Early projections have put private property damages at between $13 billion and $19 billion, depending on whether the storm makes…

RT @NoDownCarInsure: Top story: @fema: 'As Hurricane #Michael nears land, take shelter &amp; stay safe. 📱Follow weather updates on your phone…

RT @NWS: This is roughly equivalent to a 3-hr long tornado warning. Immediately get to shelter or interior room! https://t.co/syFLd3Njas

RT @cryptomurmur: #Binance, listing fees, charity and corruption: @DigiByteCoin founder clears the air Read all about it! 👉https://t.co/MMY…

Figure 4.6 Live Populated Data



Summery

RT @skydrama: Brett Adair's car is floating away in stor
and "climbed a house to...

RT @ARomero_WX: #BREAKING #Hurricane #Michael
sustained winds of 155 mph and a centr...

RT @AP: BREAKING: Hurricane Center: Category 4 Hu
155 mph winds.

RT @ARomero_WX: #BREAKING #Hurricane #Michael makes landfall near Mexico Beach as a Cat 4 storm with sustained winds of 155 mph and a centr...

RT @AP: BREAKING: Hurricane Center: Category 4 Hurricane Michael makes landfall near Mexico Beach, Florida, with 155 mph winds.

RT @AP: BREAKING: Hurricane Center: Category 4 Hurricane Michael makes landfall near Mexico Beach, Florida, with 155 mph winds.

RT @ryanelijah: 155 mile per hour winds as Hurricane Michael makes landfall near Panama City Beach.

RT @NWSAtlanta: A HURRICANE WARNING IS NOW IN EFFECT FOR PORTIONS OF CENTRAL GEORGIA WHERE WIND GUSTS OVER 70 MPH MAY BE POSSIBLE!!!

RT @RyanMaue: #HurricaneMichael has made landfall in Mexico Beach Florida ... winds were increased to 135-knots (155-mph) -- the absolute t...

RT @RyanMaue: #HurricaneMichael has made landfall in Mexico Beach Florida ... winds were increased to 135-knots (155-mph) -- the absolute t...

NING IS NOW IN EFFECT FOR PORTIONS OF CENTRAL GEORGIA
MAY BE POSSIBLE!!! Major...

as Hurricane Michael makes landfall near Panama City Beach. National

chael makes landfall on Florida Panhandle as Category 4 storm:
2 htt...

Figure 4.7 Generated Summery

Retweets which are referring to a tweet which could not be fetched at the start of data fetching process will be ignored as a valid reference for that tweet referred by those retweets is not available on our system side. Because of that, the retweets which have been accepted at the data filtering phase will be abandoned. Apart from that, system will filter out social media posts in a satisfactory manner to fetch relevant social media posts on the context of natural disasters and generate word clouds by hourly with the collected social media data within a period time (hour). The generated word clouds are very efficient as they help to gain insight of a situation instantly.

In the validation and prioritizing phases, each social media post will be validated and prioritized sequentially, and this will be mainly useful as the system is using the social media as the sources of data retrieval and any individual can post on them. Because of that the data or the social media post cans not be trusted as they are coming from untrusted source of information while very informative and important information can be derived. As well as in an event of natural disaster, time critical decisions should be made and should address highly prioritized events so that by categorizing of social media posts that could be achieve easily.

At an event of natural disaster, social media are flooded with social media posts coming for first responders of that situation so that going by each and every social media post will not be practical enough in a time critical situation so that the system will process the social media data to retrieve subset of input social media posts which are majorly discussed by the input text corpus.

5. CONCLUSION

Social media receives overwhelming number of posts during an emergency. Finding meaningful insightful information from this massive about of information with noise is hard. A lot of posts regarding emergency for example earthquake are about prayers and news which are not useful. Posts from victims and eyewitnesses are rare but very useful. However, for this information to be useful they must be process quickly since they get obsolete quickly. Existing solutions only provides the basic operation to process that information. Such as filtering related information. Organizations who responds to mass emergencies such and Disaster Management Centers (DMC) have employed teams called Intelligent Teams who go through these data as they come and make reports on the incident to improve the Situational Awareness. Doing this manually is inefficient and time-consuming. To make it more efficient in the using the same traditional way may require increase of the number of people in a team.

Although there exists a potential need for such a system which automates the process most of the organizations are hesitant to appreciate the value. This research project proposes a novel process capable of near real time analysis of social media data during mass emergency and generate useful meaning. The system would allow the decision makers, first responders with actionable information with higher dependability. Semantic analysis would give overall perspective for the status of the affected society.

REFERENCES

[1] Muhammad Imran, Carlos Castillo, Ji Lucas, Patrick Meier, and Jakob Rogstadius: Coordinating Human and Machine Intelligence to Classify Microblog Communications in Crises. *In proceedings of the ISCRAM'14. Pennsylvania, USA*.

[2] A. Bruns, J. E. Burgess, K. Crawford, and F. Shaw. # qldfloods and@ qpsmedia: Crisis communication on twitter in the 2011 south east queensland floods, 2012

[3] Sweetser, K. D., & Metzgar, E. (2007). Communicating during crisis: Use of blogs as a relationship management tool. *Public Relations Review, 33, 340–342*.

[4] Aude Hofleitner On Facebook when the earth shakes Retrieved from https://www.facebook.com/notes/facebook-data-science/on-facebook-when-the-earth-shakes/10152488877538859

[5] Fraustino, Julia Daisy, Brooke Liu and Yan Jin. "Social Media Use during Disasters: A Review of the Knowledge Base and Gaps," Final Report to Human Factors/Behavioral Sciences Division, Science and Technology Directorate, U.S. Department of Homeland Security.College Park, MD: START, 2012.

[6] Dragomir R Radev, Eduard Hovy, and Kathleen McKeown. 2002. Introduction to the special issue on summarization. Computational linguistics 28, 4 (2002), 399–408

[7] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. D., J. B., and K. Kochut, "Text Summarization Techniques: A Brief Survey," International Journal of Advanced Computer Science and Applications, vol. 8, no. 10, 2017.

[8] Tim Tinker and Elaine Vaughan, Risk and Crisis Communications: Best Practices for Government Agencies and Non-Profit Organizations, Booz Allen Hamilton, 2010, p.30,

[9] Adam Acar and Yuya Muraki, Twitter for Crisis Communication: Lessons Learned from Japan's Tsunami Disaster, International Journal of Web Based Communities, 2011 (forthcoming), p. 5.

[10] Bruce R. Lindsay, Social Media and Disasters: Current Uses, Future Options, and Policy Considerations Sep 6, 2011

GLOSSARY

Term	Definition
Natural Language Processing	Natural language processing (NLP) is a field of computer science, artificial intelligence and computational linguistics concerned with the interactions between computers and human (natural) languages.
Machine Learning	Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can change when exposed to new data.
API – Application Programming Interface	Application program interface (API) is a set of routines, protocols, and tools for building software applications. An API specifies how software components should interact.

APPENDICES

Appendix A – DynamoDB

DynamoDB is a fast and flexible NOSQL database which provides reliable performance. It is fully managed by Amazon Web Services (AWS) which means we don't have to manage servers, scaling, backup, security. AWS assures to provide consistent single-digit millisecond latency at any scale.

There are three core components to DynamoDB

- **Tables**

Like other DBs DynamoDB stores data in tables. There is a limitation to the number of tables that can be created region wise. (as of 2018 the initial limit is 256 table although you can request AWS for more). There is no limitation to the size of a table in DynamoDB.

- **Items**

An item is a group attributes that is uniquely identifiable among all other items. Each DynamoDB table consists of zero or more items. The maximum item size in DynamoDB is 400 KB, which includes both attribute name binary length (UTF-8 length) and attribute value lengths (again binary length). The attribute name counts towards the size limit.

- **Attributes**

Each item is composed of one or more attributes. An attribute is a fundamental data element, something that does not need to be broken down any further. Attributes in DynamoDB are similar in many ways to fields or columns in other database systems.

Partition Key / Sort Key

Partition key act as the primary key for a table. Sort Key is an optional field which is used with the partition key. When both Partition key and the sort key is used it becomes a Composite key for the table. If we are using Partition key alone then it must be unique. When we are using Partition key along with a Sort key the partition key can duplicate. But two Items with the same Partition key value cannot have the same Sort key value.

When decide to use DynamoDB we must first think about how we are going to retrieve data, and how we handle the provisioned capacity units.

- **Partition Key Only** (Primary key, must be unique)
- **Partition Key + Sort Key** (composite primary key)

DynamoDB Data Retrieval

There are two ways

- Scan

When used it goes through all the items in a table. Scan operation is less efficient, and it consumes more RCUs.

- Query

The query operation finds items based on primary key values. We can query a table of a secondary **index** that has a composite primary key (a partition key and a sort key)

We can optionally provide a “**FilterExpression**”. A “FilterExpression” determines which items within the results should be returned. All other results are discarded. Each of these operations returns only 1mb of data in a single request. A single Query operation will read up to the maximum number of items set (if using the Limit parameter) or a maximum of 1 MB of data and then apply any filtering to the results using FilterExpression. If “**LastEvaluatedKey**” is present in the response, we will need to paginate the result set.

Capacity Units

Capacity Units are the measurement of how much resources are used when using DynamoDB services. (Prices may vary depending on the region that is used to create the table)

- **Read Capacity Units (RCU)**

One RCU = one strongly consistent read per second, or two eventually consistent reads per second, for items up to 4 KB in size.

- **Write Capacity Units (WCU)**

One WCU = one write per second, for items up to 1 KB in size.

DynamoDB Eventual Consistency.

The eventual consistency option maximizes your read throughput. However, an eventually consistent read might not reflect the results of a recently completed write. All copies of data usually reach consistency within a second. Repeating a read after a short time should return the updated data.

DynamoDB Strong Consistency.

In addition to eventual consistency, DynamoDB also gives us the flexibility and control to request a strongly consistent read if your application, or an element of your application, requires it. A strongly consistent read returns a result that reflects all writes that received a successful response before the read.

DynamoDB Indexes

There are two types of indexes we can create in DynamoDB.

- Global Secondary Index (GSI)

The primary key of the index can be any two attributes from its table.

- Local Secondary Index (LSI)

The partition key of the index must be the same as the partition key of its table. However, the sort key can be any other attribute.

Time to Live

Time to Live (TTL) for DynamoDB allows you to define when items in a table expire so that they can be automatically deleted from the database. TTL is provided at no extra cost to reduce storage usage and reduce the cost of storing irrelevant data without using provisioned throughput. With TTL enabled on a table, you can set a timestamp for deletion on a per-item basis, allowing you to limit storage usage to only those records that are relevant.

TTL is useful if you have continuously accumulating data that loses relevance after a specific time period. For example: session data, event logs, usage patterns, and other temporary data. If you have sensitive data that must be retained only for a certain amount of time according to contractual or regulatory obligations, TTL helps us ensure that it is removed promptly and as scheduled.

DynamoDB Streams

DynamoDB Streams is an optional feature that captures data modification events in DynamoDB tables. The data about these events appear in the stream in near real time, and in the order that the events occurred. Each event is represented by a stream record. If you enable a stream on a table, DynamoDB Streams writes a stream record whenever one of the following events occurs:

- A new item is added to the table: The stream captures an image of the entire item, including all its attributes.

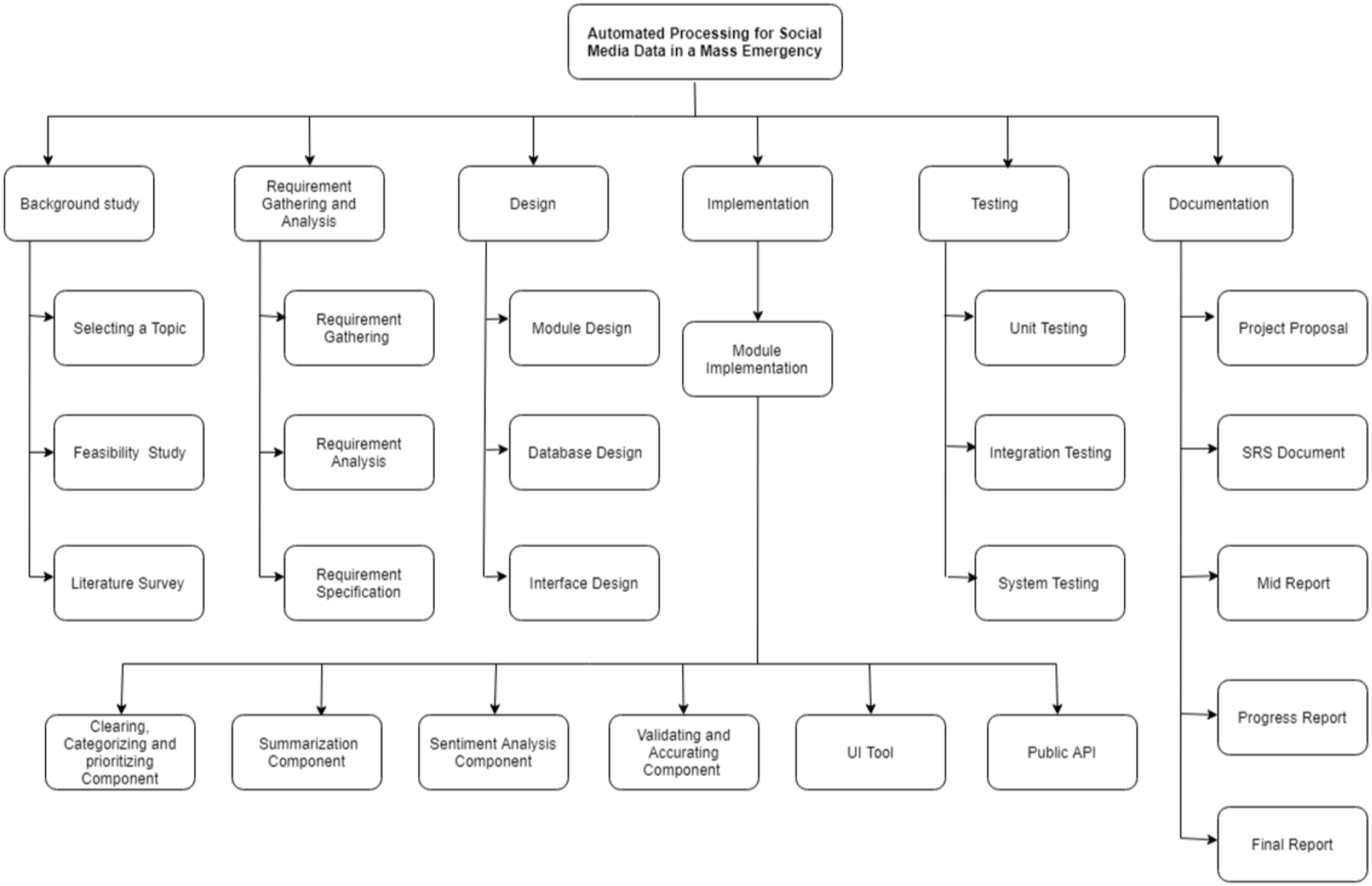
- An item is updated: The stream captures the "before" and "after" image of any attributes that were modified in the item.

- An item is deleted from the table: The stream captures an image of the entire item before it was deleted.

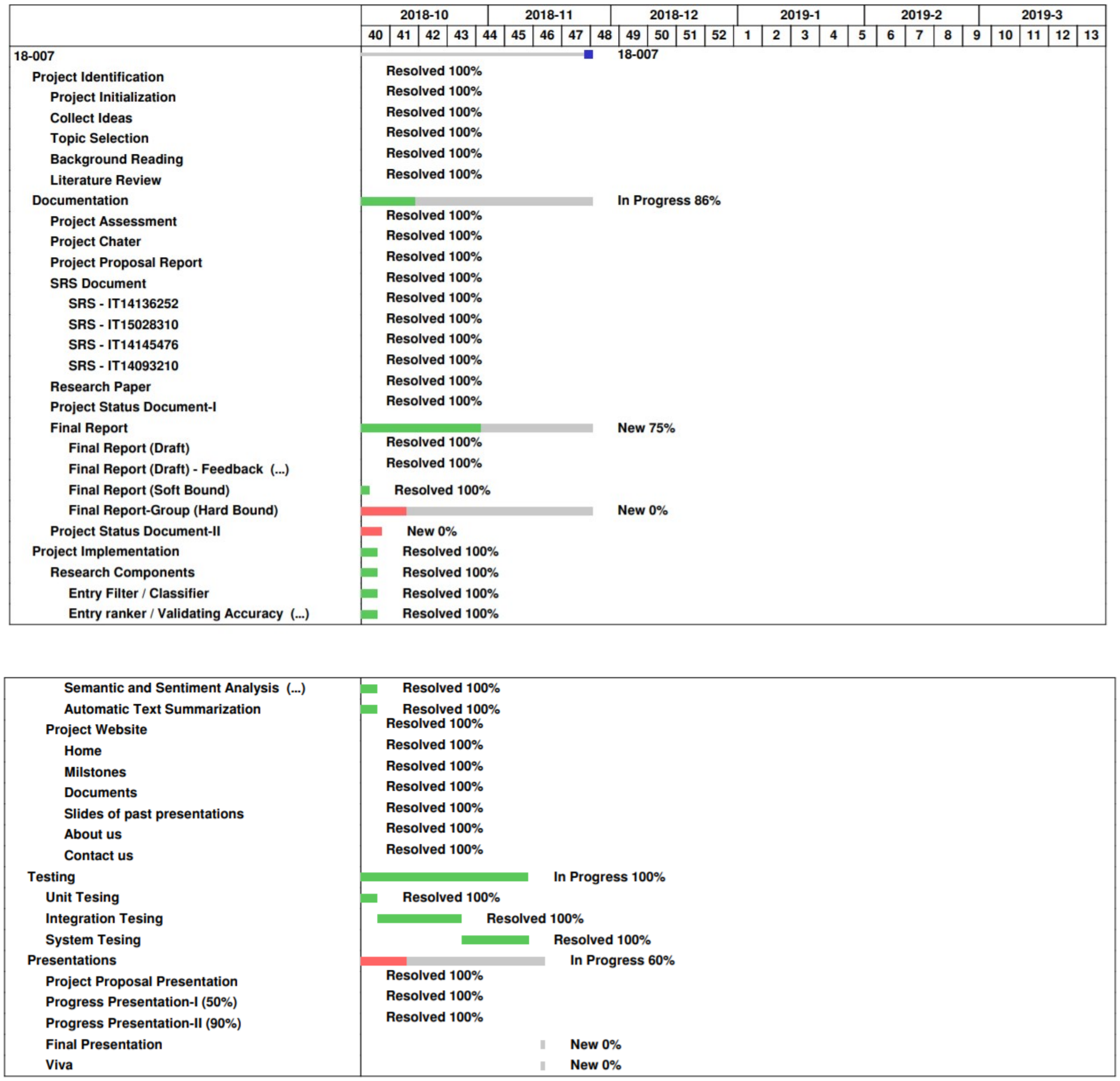
Each stream record also contains the name of the table, the event timestamp, and other metadata. Stream records have a lifetime of 24 hours; after that, they are automatically removed from the stream.

We can use Streams together with Lambda to create a trigger—code that executes automatically whenever an event of interest appears in a stream.

Appendix B – Work Breakdown Structure



Appendix C – Gantt Chart


Project Identification
Resolved 100%
Project Initialization
Resolved 100%
Collect Ideas
Resolved 100%
Topic Selection
Resolved 100%
Background Reading
Resolved 100%
Literature Review
Resolved 100%
Documentation
In Progress 86%
Project Assessment
Resolved 100%
Project Chater
Resolved 100%
Project Proposal Report
Resolved 100%
SRS Document
Resolved 100%
SRS - IT14136252
Resolved 100%
SRS - IT15028310
Resolved 100%
SRS - IT14145476
Resolved 100%
SRS - IT14093210
Resolved 100%
Research Paper
Resolved 100%
Project Status Document-I
Resolved 100%
Final Report
New 75%
Final Report (Draft)
Resolved 100%
Final Report (Draft) - Feedback (...)
Resolved 100%
Final Report (Soft Bound)
Resolved 100%
Final Report-Group (Hard Bound)
New 0%
Project Status Document-II
New 0%
Project Implementation
Resolved 100%
Research Components
Resolved 100%
Entry Filter / Classifier
Resolved 100%
Entry ranker / Validating Accuracy (...)
Resolved 100%
Semantic and Sentiment Analysis (...)
Resolved 100%
Automatic Text Summarization
Resolved 100%
Project Website
Resolved 100%
Home
Resolved 100%
Milstones
Resolved 100%
Documents
Resolved 100%
Slides of past presentations
Resolved 100%
About us
Resolved 100%
Contact us
Resolved 100%
Testing
In Progress 100%
Unit Tesing
Resolved 100%
Integration Tesing
Resolved 100%
System Tesing
Resolved 100%
Presentations
In Progress 60%
Project Proposal Presentation
Resolved 100%
Progress Presentation-I (50%)
Resolved 100%
Progress Presentation-II (90%)
Resolved 100%
Final Presentation
New 0%
Viva
New 0%