Social Media Text Mining for Decision Support in Natural Disaster Management in Sri Lanka

Submitted by:

K.H.J.Imalka

158761R

Supervised by: Mr. Saminda Premaratne

Faculty of Information Technology

University of Moratuwa

May 2018

Social Media Text Mining for Decision Support in Natural Disaster Management in Sri Lanka

Submitted by:

K.H.J.Imalka

158761R

Supervised by: Mr. Saminda Premaratne

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Master Degree of Science in Information Technology.

May 2018

Declaration

I declare that this is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of reference is given.

Name of the student : K. H.J. Imalka

Student Number : 158761R

Signature of the student : ……. Date: ……

Supervised by:

Name of the supervisor : Mr. Saminda Premarathne

Signature of the supervisor : ……. Date: …….

Dedication

This Dissertation is dedicated to my loving parents and husband for being part of me and encouraging me always by being my side.

Acknowledgement

First I express my heartfelt gratitude to my supervisor Mr. Saminda Premarathna for his most valued guidance, commitment and kind support to make this research a success.

Also I sincerely thank Prof. Asoka S Karunananda who taught us all the research related document preparation which was a great support to manage all work with busy schedules.

It's my pleasure to thank Mr. Chaman Wijesiriwardena and all other Senior Lecturers, Lecturers, Instructors and staff members who helped us in many ways to successfully complete this research.

Then I would like to thank all the batch mates of MSc. In Information Technology batch 9 for their companionship and various kinds of support given throughout the program.

In addition, I would like to thank my work mates and friends for encouraging me with their support and best wishes.

Last but not least, my sincere thank goes to my beloved parents, husband and his parents for helping me to conduct this work without much stress and encouraging me to complete this research.

Abstract

With the popularity of internet and smart devices, social media is very popular today among individuals in almost all the ages which help them to create and share their personal feelings, experiences, ideas as well as information with others connected to them over a computer mediated technologies Individuals use these social media applications such as Facebook and twitter which are popular most to share their experiences, opinions, day today activities as well as achievements. Due to this nature when there are emergencies and natural disasters these social media applications tend to be flooded with content generated from public who affected, who are looking for their family members and friends, who are looking for information as well as with the people engage in humanitarian activities.

Therefore social media has become the first to generate related information when there is a catastrophic event before any of news sites or government bodies engage in disaster management. These social media content is quick accurate and subjective during disaster situations therefore we can use this information as an asset to reduce risk and build awareness among public about the disaster as well as to provide decision making support to relief efforts.

This research focuses on building decision making support using social media content generated during disaster situations in Sri Lankan context. Mainly the content will be tweets posted by public during a natural disaster and consisting with text written in English. Therefore situational awareness building will be done using text mining which natural language processing in this study since the content is unstructured.

Content will be analyzed using techniques to scrape, clean, classify and generate real information about the disaster and to visualize them to support decision making for authorities engage in disaster management as well as volunteers engage in relief efforts.

Contents

Table of Figures

Table of Tables

Chapter 1

Introduction to Social Media Text mining for Decision Support in Natural Disaster management

1.1 Introduction

Social media is very popular today among individuals in almost all the ages which help them to create and share their personal feelings, experiences, ideas as well as information with others connected to them over a computer mediated technologies. Social media services are (currently) Web 2.0 Internet-based applications and user-generated content is the lifeblood of social media. Individuals and groups create user-specific profiles for a site or app designed and maintained by a social media service and social media services facilitate the development of social networks online by connecting a profile with those of other individuals and/or groups.[1] As at August 2017 most popular social networking site are Facebook, YouTube, Instagram, Twitter, Reddit, Vine, Ask.fm, Pinterest, Tumblr, Flickr, Google+, Linkedln etc. [2]

According to the statistics of TRC[1], Sri Lanka's increased internet connectivity have given a boost to the Sri Lankan's presence in the social media, especially on the Facebook, the favorite local online hangout. Starting from mid-2016, the Sri Lankans number on Facebook has increased from a 4 million to a 5 million.

Due to this nature of social media services during times of disasters online users generate a significant amount of data, some of which are extremely valuable for relief efforts. [3] The increasing use of social media, such as Twitter and Facebook, by humanitarian organizations, public authorities and citizens preparing for and responding to disasters generates vast quantities of information.[4] Crowdsourcing of data during such a disaster can aid in the task of decision making.[5]. During disasters, the public is even more active online, increasingly turning to social media for the most up to date information. Social media, however, are used for more than information seeking or sharing during disasters; the public increasingly expects

[1] http://www.trc.gov.lk/2014-05-13-03-56-46/statistics.html

emergency managers to monitor and respond to their social media posts.[6] Social media improves situational awareness, facilitates dissemination of emergency information, enables early warning systems, and helps coordinate relief efforts[7]

According to Fraustino and all disaster communication deals with disaster information disseminated to the public by governments, emergency management organizations, and disaster responders as well as disaster information related and shared by journalists and the public. Disaster communication increasingly occurs via social media in addition to more conventional communication modes such as traditional media (e.g., newspaper, TV, radio) and word of mouth (e.g., phone call, face-to-face, group). Timely, interactive communication and user generated content are hallmarks of social media, which include a diverse array of web and mobile based tools.

Further they reveal multiple reasons the public uses social media during disasters:

- Because of convenience
- Based on social norms
- Based on personal recommendations
- For humor & levity
- For information seeking
- For timely information
- For unfiltered information
- To determine disaster magnitude
- To check in with family & friends
- To self-mobilize
- To maintain a sense of community
- To seek emotional support & healing

Social media such as Twitter have emerged as a new data source for disaster management and flood mapping by leveraging Twitter data in geospatial processes.[8]

Therefore we can identify that social media is emerging as an important information-based communication tool for disaster management and online information can enable effective disaster preparedness and reduce losses. This information can be used to discourse real picture of the disaster as well as an opportunity to understand needs of

victims and linking victims with helping hands monitoring risks and building community awareness.

1.2 Background of Study

1.2.1 Disasters and Disaster Communication

Disaster is a sudden, calamitous and unfortunate event that brings with it great damage, loss, destruction and devastation to human life as well as property and also hampers the ongoing developmental projects in a particular area being affected by the disaster. [9] Disaster has been defined in many ways; World Health Organization has defined disaster as any sudden occurrence of the events that causes damage, ecological disruption, loss of human life, deterioration of health and health services, on a scale sufficient to warrant an extraordinary response from outside the affected community or area.

There are different types of disasters out of them natural disasters are the most unpredictable type and the type affected to millions of citizens such as high intensity earthquake, floods, cyclone, flash floods, some major landslides and event of draught. These disasters generally cause a high loss of life and property and also lead to displacement of a lot of people from their shelters. Generally these disasters pose a major threat to the developmental projects as well as infrastructure of a particular area. Preparedness against these disasters should be on the top of the priority list.

Therefore disaster management is very important to survive in the case of a natural or a major man-made disaster and can be defined as the organization and management of resources and responsibilities for dealing with all humanitarian aspects of emergencies, in particular preparedness, response and recovery in order to lessen the impact of a sudden disaster.

1.2.2 Social Media

Social media is the collective of online communications channels dedicated to community-based input, interaction, content-sharing and collaboration. Websites and applications dedicated to forums, micro blogging, social networking, social

bookmarking, social curation, and wikis are among the different types of social media. [10]

Although Facebook or Twitter might come to mind, the social media realm includes a multitude of web- and mobile-based technologies ranging from photo and video sharing sites to rating and review forums. Table 1 (below) outlines a variety of social media types and some popular examples of each.

Social Media Type	**Examples**
Blogs	Blogger, WordPress
Discussion Forums	LiveJournal, ProBoards
Micro-blogs	Tumblr, Twitter
Photo/Video Sharing & Podcasting	Flickr, iTunes Podcasts, Youtube, Pinterest
Social Bookmarking	Del.icio.us, Diigo
Social Discovery Engines & News Sources	Reddit, StumbleUpon, Slashdot
Social/Professional Networking	Facebook, Google+, LinkedIn, MySpace
Social Rating/Reviews	AngiesList, Yelp
Video/Text Chatting	Skype, AIM, mobile texting
Wikis	Wikipedia, Wikispaces

Table 1.1: Social Media Types with Examples

In a world of increasing interconnectedness between individuals and companies across the globe, social media continues to evolve and play a larger role in day-to-day life. The increasing use of social media, such as Twitter and Facebook, by humanitarian organizations, public authorities and citizens preparing for and responding to disasters generates vast quantities of information.[4]

According to the statistics we can see that there is a significant usage in social media and following figure presents some valuable statistical indicators for the world's internet and social media users.

Figure 1.1: Global Digital Snapshot

Facebook is still the site with the most active users (1.860 billion) per month and 1.74 billion active users on mobile. That's roughly 22% of the world's population. (Statista)[2]

1.3 Motivation

In recent years, Twitter has been used to spread news about casualties and damages, donation efforts and alerts, including multimedia information such as videos and photos. [3] In responding to disasters, including the 2010 Haiti earthquake, the 2012 Sandy super storm and the 2013 Boston Marathon bombings, social media was used for relaying information, one and two-way communication, offering/requesting assistance and organizing disaster response. [4]

[2] https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

Figure 1.1 presents popularity of social media platforms in the world in millions

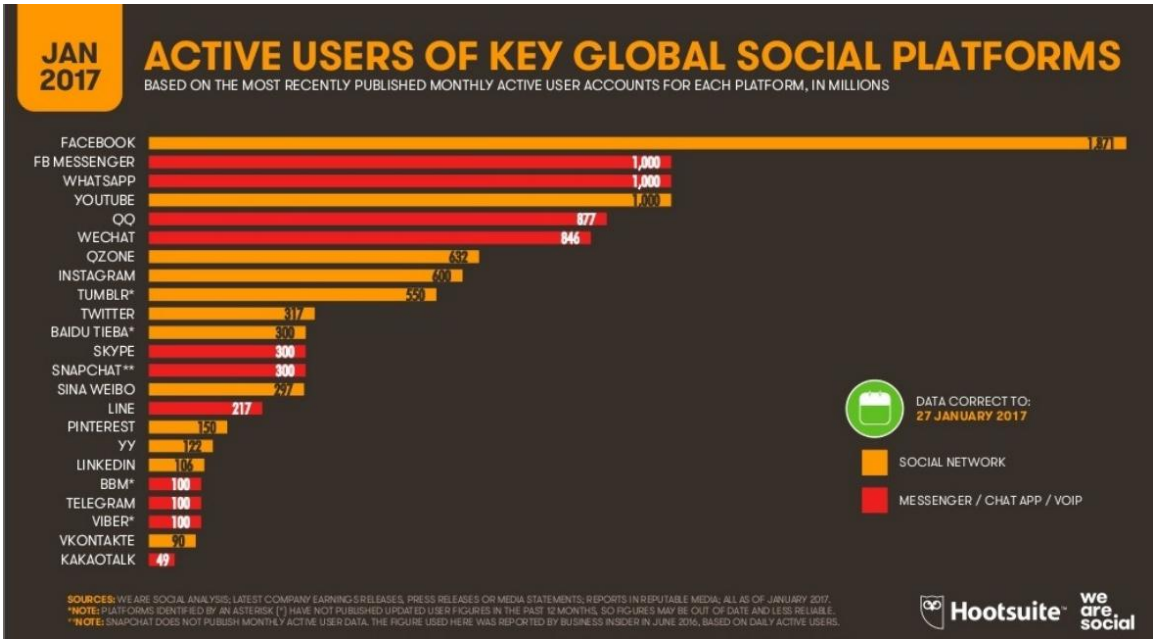


Figure 1.2: Active Users of Key Global Social Platforms

With the massive use of social media during disaster situations incoming information filtering is very much important for situational awareness. This is a real time analysis therefore this includes text mining, natural language processing with sentiment analysis, and social network analysis.

Text mining is used in every field be it for business intelligence, social media analysis, sentiment analysis, biomedical analysis, software process analysis and even for security analysis. [11] 55 popular text mining tools with their features are mentioned in the paper titled “Comparison of Text Mining Tools” by Arvinder Kaur and Deepti Chopra. Therefore we can use those tools to extract data to achieve the objective of this research work.

The potential application of social media to disaster and crisis management is attracting the research community’s attention either as a tool or as a source of data, e.g., for the creation of crisis maps.[12]

[13] provide an overview of existing and proposed methods and systems to retrieve information about emergencies from social media, such as CrisisTracker and

TweetTracker. Furthermore, suggest a platform to collect human annotations in order to maintain automatic supervised classifiers for social media messages and describe automatic methods for extracting brief, self-contained information items from social media, which are relevant to disaster response.

[3] have presented a practical system that can extract disaster-relevant information from tweets and describe the methodology they have followed to filter out messages that do not contribute to valuable information into two main classes "Personal" & "Informative". If the message is not related to the disaster they have categorized it as "Other". Further they have differentiated the messages belong to informative class as direct, i.e., written by a person who is a direct eyewitness of what is taking place or indirect, when the message repeats information reported by other sources.

Tweet labeling procedure of Integrating Social Media Communications into the Rapid Assessment of Sudden Onset Disasters of [14] was under

- Infrastructure Damage: Information about destruction and/or damage of roads, bridges, buildings; disruptions to basic services, e.g. hospitals.
- Community Needs: Information about shelters, food, and location of missing persons, water, and hygiene.
- Humanitarian Support: Information about deployment of aid, recovery services, and in-kind donations and contributions of goods and services.

To obtain a set of tweets they have performed a keyword search using Twitter's Streaming API.

1.4 Problem in Brief

All above research describe different approaches used to assist decision making in disaster management internationally but we don't find any framework designed for Sri Lanka. I would like to suggest a model or framework which can be used with social media text mining to assist victims, build preparedness among citizens near the disaster, provide awareness of actual situation to the community as well as to combine helping hands with needy citizens.

To collect information we can use Facebook and Twitter because of their popularity in Sri Lanka. Analyzing the unstructured data is a challenge since the information posted by social media users consists with text as well as symbols, links, images etc. Through sentiment analysis we can extract the real picture of the text.

1.5 Aim and objectives

Aim: To provide the decision making support through social media text mining during natural disasters. Objectives of the research are,

1. Identifying the techniques and tools can be used to extract important information through text mining during mass emergencies posted in social media
2. Identify methods to be followed to clean data extracted from social media
3. Discover features of text mining such as text analytics, text processing, sentiment analysis and knowledge discovery
4. Identify changes to be done in data extracted through social media in order to provide an effective decision support
5. Discover knowledge in text mining

Develop a model or framework to use mined data in disaster management decision making

1.6 Proposed solution

An automatic system for disaster related information extraction requires two components: Classification of posts, tweets and Extraction from posts, tweets. Identifying informative messages and then extracting information (e.g. damages, donation offers, support requests) is required out of relevant messages contribute to situational awareness because the messages generated during disaster situations vary greatly in value. The final system output should consist with accurate, brief, self-contained pieces of information most likely to augment situational awareness. [3]

1.7 Summary

In this chapter it is identified the importance of content posted in Social Media during natural disasters and how we can use those unstructured data to derive meaningful

information about the situation. Further what others have done in this disaster management area is presented in this chapter in brief and in detail information will be given in the next chapter.

Chapter 2

Literature Review on Social Media Text mining & Decision making

2.1 Introduction

In chapter 1, we describe what social media is and what the usages of social media in general public are and how social media is used during natural disaster situations to build awareness among public about the disaster including motivation for the study as well as research objectives. Further there was a brief introduction to the problem of research and importance of using social media as an information source to collect data during natural disaster situations. This chapter explains how social media has been contributed to disaster response and decision making during natural disaster situations according to the related work done by other researches in the discipline.

An initial literature review was conducted to discover the instances of social media usage for decision making during natural disaster situations and we could find that large amount of researches done with data collected through social media such as Twitter and Facebook. Most of the researches have been conducted subjected to a particular natural disaster happened in the past and to prove the validity and similarity of information collected using social media with actual information collected through reliable sources such as disaster management authorities and news sites. Therefore we can trust the information posted in social media during natural disaster situations as a reliable source to assist decision making and reveal information about the real picture of disaster quicker than other humanitarian and disaster management authorities and organizations.

Therefore the researchers conducted in this area of study can be categorized as

- Usage of social media during disaster situations
- Assessment of disasters using information posted in social media
- Collecting information from social media
- Processing social media messages in mass emergencies and
- Building decision support systems using social media data posted during disaster situations.

Therefore the chapter 2 describes in detail review of literature related to above categories and finally the research problem will be highlighted with limitations in related work and importance of having a proposed system to assist decision making in Sri Lanka.

2.2 Use of social media during disaster situations

[15] Describes possibilities of using social media in natural disaster management. In the paper they have presented analysis of communication types in between participants in natural disaster events as well as guidelines for organizing information exchange by social media. They have identified that social media can be used in three ways according to many researches and those are

- Preparing for a natural disaster
- Responding during and immediately after the natural disaster
- Recovering from the natural disaster

Further the research reveals social media has been used to warn people help in coordination of response and recovery in recent disasters by emergency managers, people affected and seeking information about the disaster. Team of emergency managers and volunteers around the world who join together using social media and provide services during disaster situations are known as Virtual Operations Support Teams.

For example, researchers found that in the half hour leading up to a potential fatal storm hitting a festival in Belgium, the public published more than 2,000 related tweets. That number soared to more than 80,000 tweets during the first four hours of the disaster. Also, the first reports of the 2008 earthquake in China came from Twitter, not the government. [6] Further the paper says that during disaster situations people are using social media not only to seek information but also expects emergency managers to monitor and respond to their posts.

A result accomplished through Twitter and Facebook by American Red Cross shows that practicing public relations through social media is effective and necessary in the emerging digital age. [16] Further in responding to disasters, including the 2010 Haiti earthquake, the 2012 Sandy super storm and the 2013 Boston Marathon bombings, social media was used for relaying information, one and two-way communication,

offering/requesting assistance and organizing disaster response according to the research by Anson and all. [4]

2.3 Assessment of disasters, Processing social media messages and assist decision making in mass emergencies

Micro blogging platforms like twitter has attracted significant public and research interest due to their potential in use of disaster situations. Twitter is a platform which allows its users to post 140-charater messages and to follow messages from any other registered users, therefore that openness place Twitter somewhere in between a purely social network and purely informational network [7]. An assessment done with relates to one of costliest disasters in US history, Hurricane Sandy 2012 to show relationship between proximity to Sandy's path and hurricane-related social media activities. Further they have demonstrated that per-capita Twitter activities strongly correlate with the per-capita economic damage inflicted by the hurricane [7].

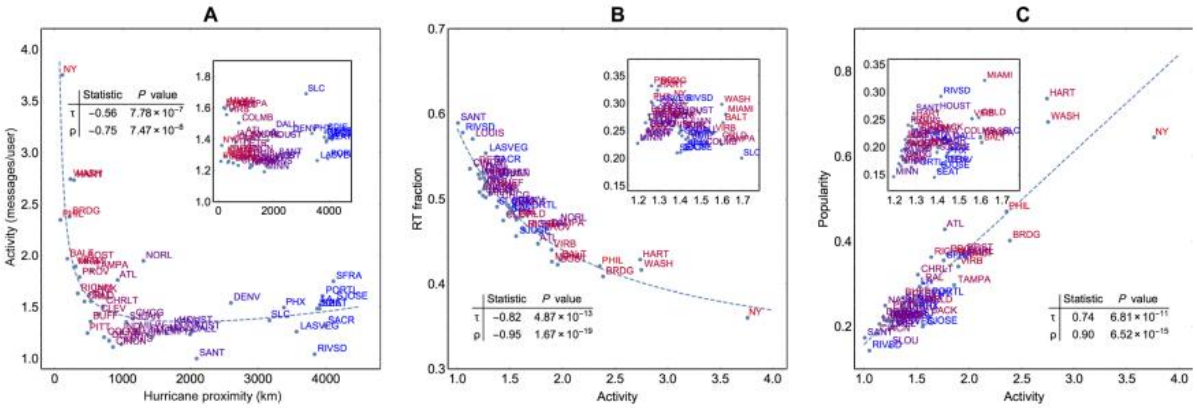


Figure 2.1: Characteristic features of Twitter activity across locations (labeled by color according to hurricane proximity; blue, farther from the disaster; red, closer to the disaster)

Figure 2.1 shows the results of assessment across locations during Hurricane Sandy, US. It reflects that per-capita Twitter activity strongly correlates with the per-capita economic damage inflicted by the hurricane.

Paper titled Mining Social Media to Inform Peatland Fire and Haze Disaster Management presents the potential of social media to assist disaster management by comparing satellite imagery and almost all GPS-stamped tweets from Sumatra Island,

Indonesia posted during Peatland fires and haze events [12]. They have analyzed temporal correlations between the datasets and their geo-spatial interdependence and demonstrates that social media are a valuable source of complementary and supplementary information for haze disaster management. They have created four rich taxonomies for identification of different topics related to haze and Peatland fires. Using them they have presented correlations between the public discourse on Twitter and peat fire hotspots.

According to [13] active communication channels during mass emergencies such as first responders, decision makers and public use of social media gains insight into situation. Parsing brief and informal messages, handling information overload, and prioritizing different types of information found in messages are important in processing social media messages to obtain information. Most large social media platforms provide programmatic access to their content through an Application Programming Interface (API) [13]. There are two types of APIs as search APIs and streaming or filtering APIs. Search APIs allows query past messages and stream or filtering APIs allows data collectors to subscribe to real time data.

2.4 Opportunities and barriers of using social media for disaster preparedness

Due to network connectivity during disaster situations, general public access to social media data gets limited which is a serious obstacle to research in this space [13].

Data preprocessing is done by most researchers using available techniques according to type of data they are having and goal of analysis. Natural Language Processing is used since all the data consisting of unstructured textual form in social media [22]. By using NLP toolkit tokenizing, part-of-speech tagging (POS), semantic role labeling, dependency parsing, named entity recognition, and entity linking can be performed according to the publication by M. Imran, C. Castillo, F. Diaz, and S. Vieweg [13].

Twitter messages are brief, informal, noisy, unstructured and often contains misspellings and grammatical mistakes [23]. Due to 140 character limitation twitter users intentionally shorten words by using abbreviations, acronyms, slangs, and sometimes words without spaces [23]. Therefore we have to pay attention in order to improve the accuracy of Natural Language Processing due to this informal nature.

Opportunities of using social media during disaster situations include Disaster Risk Reduction and preparedness such as by examining and managing the causal factors of disasters, including reducing exposure to hazards, reducing the vulnerability of people and property and increased preparedness for disaster events using eyewitness accounts and images or videos of the impacted area on social media posted by citizens [4].
Online social networks allow the establishment of global relationships that are domain related or can be based on some need shared by the participants and emergency service agencies are utilizing the power of social media and SMS to instantly broadcast and amplify emergency warnings to the public [15]. Further paper emphasizes following critical tasks that can be implemented by social media.

- Prepare citizens in areas likely to be affected by a disaster;
- Broadcast real-time information both for affected areas and interested people;
- Receive real-time data from affected areas;
- Mobilize and coordinating immediate relief efforts; and Optimize recovery activities

2.5 Practical Extraction of relevant information

Messages posted during disaster situations are extremely varied therefore filtering messages related to disaster as well as messages not contributing to generate valuable information is important. Imran and all in their paper titled Practical Extraction of Disaster-Relevant Information from Social Media describes that they have started their process by classifying into two main classes [3]:

- Personal: if a message is only of interest to its author and her immediate circle of family/friends and does not convey any useful information to people who do not know its author.
- Informative: if the message is informative (of interest to other people beyond the author's immediate circle).
- Other: if the message is not related to the disaster.

Classification is done by a set of multi-label classifiers trained to automatically classify a tweet into one or more of the above classes. Naive Bayesian classifiers are

used as implemented in Weka suing a rich set of features including word unigrams, bigrams, Partof-Speech (POS) tags and others [3].

Multiclass categorization of twitter messages has been done by using well known algorithms Naive Bayes (NB), Support Vector Machines (SVM), and Random Forest (RF) in paper published by Imran and all [23]. In the paper they have mentioned that training all classifiers has been done by preprocessed data. Following are the preprocessing steps they have followed.

- First, removing stop-words, URLs, and user mentions from the Twitter messages
- then Stemming using the Lovins stemmer
- Using Unigrams and bi-grams as features
- Using information gain, a well-known feature selection method to select top 1k features.

Labeled data used in this task has been annotated by the paid workers [23].

2.6 Summary

This chapter describes in detail how social media is used during disaster situations and how the information posted can be used to assist decision making. Chapter has been organized to present literature of use of social media during disaster situations, opportunities and barriers and practical extraction of relevant information. There are lots of researches conducted in this area and the widely used social media is Twitter considering its openness to extract real data. Therefore in this work also information extracted though Twitter will be used considering the amount of literature done related to disaster risk reduction.

Chapter 3

Technology Adapted in Social Media Text mining & Decision Making

3.1 Introduction

Chapter 2 described similar research work in details with related to usage of social media in natural disaster situations, assessment based on social media messages about the disaster, data collection from social media, processing extracted data to build situation awareness and building decision support systems to support humanitarian activities and risk reduction. Further chapter 2 described approaches followed by other researchers including tools and algorithms used to collect data, monitor data and analyze data to support decision making. Therefore at the end of chapter 2 we can conclude that social media becomes rich with important and novel information when there are natural disasters due to rapid use by public to share their information, look for family members and friends as well as the messages to extend support hands toward affected people. Using that content we can propose a system to build situational awareness and to support decision making using social media text mining. Therefore the technologies to address proposed system will be elaborated in this chapter with related to Facebook and Twitter.

3.2 Technologies to collect data in social media

Due to popularity in Sri Lanka Facebook and Twitter messages will be considered here for data collection. There are number of tools available for Twitter and Facebook but limitation are there to use those tools such as not available free of charge, number of tweets or post can be read is limited and special permission or qualification required to use the tool (such as only available for PhD students, post Docs and Professors)

Application Programming Interfaces are available for both Twitter and Facebook and all the tools extract data through these APIs. Facebook is less open than Twitter when it comes to sharing information. Privacy is more of a concern because people share more personal data on Facebook. Tools available to extract data can be divided into

two groups such as Applications and modules & libraries which require programming knowledge to extract data. Following table presents some of the tools available for Facebook and Twitter.

Social Media	Data Collections Tools	Modules & Libraries
Twitter	DD-CSS, Discovertext, DMI-TCAT, BU-TCAT, Flocker, iScience MapsNaoyun, Netlytic, NodeXL, Nvivo/Ncapture, SocioViz, Sodato	140dev, Hosebird, Pattern, poll.emic, Python-Twitter, Social Feed Manager, SocialMediaMineR, streamR, T, tStreamingArchiver, twarc, Twecoll, tweepy, Twitter, Stream Downloader, Twitter-Tap, TwitterGoggles, TWurl, twutil, Twython, yourTwapperKeeper
	Tweet Archivist, Chorus-TweetCatcherDesktop, Twitter Demand Collector and Analyzer	
	Twitonomy, Webometrics	
Facebook	Digitalfootprints, Discovertext, Infoextractor, Netvizz, NodeXL (with Social Network Importer), Nvivo/Ncapture, Sodato	Facebook Python SDK, Facepager, fb_scrape_public, RFacebook, SocialMediaMineR

Table 3.1: Applications, Modules and Libraries to collect data from Facebook and Twitter

Other than above mentioned tools, there are Social data vendors such as Brandwatch, Crimson Hexagon, Datasift, Gnip, Plus one social, Pulsar, SocialPeeks, Sysomos, Texifter, Twitris, Awario etc. Some of these tools allow data analysis and visualization within their own platforms or independently and some tools are Internet marketing tools which monitor every corner of the Web for mentions of given keywords in real time.

3.2.1 Twitter API

Twitter developer platform [3] allows application developers to use its APIs with key areas such as ‘Ads API’ to programmatically create and manage ad campaigns, ‘Search Tweets’ to Use the Search API to gather Tweets, ‘Filter real time Tweets’ to Get only the Tweets required in real time and ‘Direct Message API’ to build personalized customer experiences with twitter's Direct Message platform

3.2.2 Facebook Graph API

The Graph API is the primary way to get data out of, and put data into, Facebook's platform. It's a low-level HTTP-based API that can be used to programmatically query data, post new stories, manage ads, upload photos, and perform a variety of other tasks that an app might implement.

3.3 Text Mining

Text mining is the discovery and extraction of interesting, non-trivial knowledge from free or unstructured text.[17]

In the general framework of knowledge discovery, Data Mining techniques are usually dedicated to information extraction from structured databases. Text Mining techniques, on the other hand, are dedicated to information extraction from unstructured textual data and Natural Language Processing (NLP) can then be seen as an interesting tool for the enhancement of information extraction procedures.[18]

Further text mining is described as the automated process of detecting and revealing new, uncovered knowledge and inter-relationships and patterns in unstructured textual data resources. Target of text mining is to discover knowledge in huge amounts of texts. Figure 5 show the text mining process in a high level view. [19]

According to the paper the first step of text mining process is collecting a set of un-structured text documents. Then, the pre-processing for the documents is performed to remove noise and commonly used words, stop words, stemming. This process produces a structured representation of the documents known as Term-document matrix, in which, every column represents a document and

[3] https://developer.twitter.com/

every row represents a term occurrence throughout the document. The final step is applying data mining techniques such as clustering, classification, association rules to discover term associations and patterns in the text and then, finally, visualizing these patterns using tools such as word-cloud or tag-cloud.

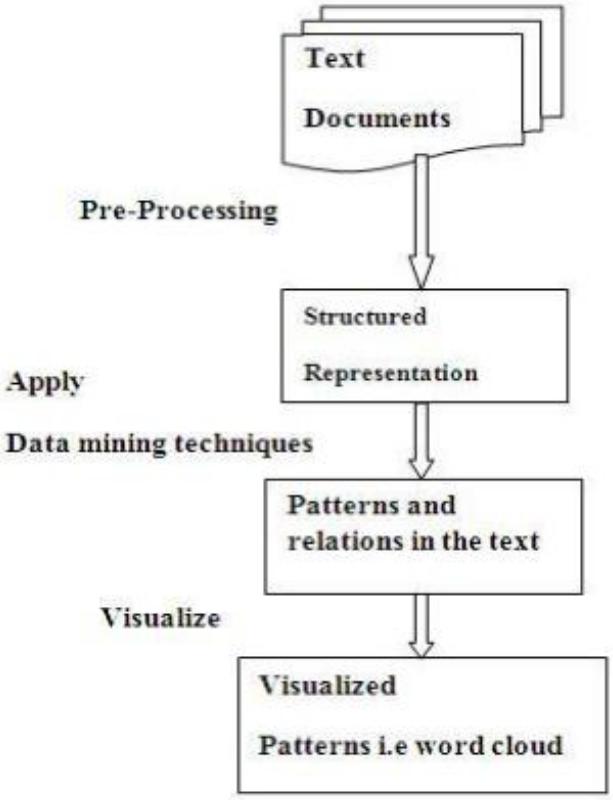


Figure 3.1: Text mining process

3.4 Sentiment analysis

Sentiment analysis deals with the computational detection and extraction of opinions, beliefs and emotions in written text. It combines theories and methodologies from a diverse set of scientific domains, such as psychology, natural language processing and machine learning.[20]

Sentiment analysis, the process of automatically distilling sentiment from text, provides little insight regarding the language granularities beyond the use of positive and negative words.[21] Sentiment minig is also known as opinion mining and subjectivity analysis [19] that attempts to make automatic systems to determine human opinion from text written in natural language.

[5] In their paper, they propose a technique of extracting situation awareness information using concepts of semi-supervised machine learning along with creating interactive map to locate the vulnerable areas during a disaster.

3.5 Content analysis for decision making

Content analysis can be identified as natural language processing (NLP) since our data set consisting with tweet contents related to a natural disaster. (NLP) is a field of computer science, artificial intelligence and linguistics concerned with the interactions between computers and human (natural) languages. [22]

Therefore Natural language processing classifies can be used to classify the data into different categories such as tweets building situational awareness, factual information etc. NLP classifiers require "training" data in the form of annotated or coded text, which they use to "learn" how to distinguish between different types of discourse.

Automated sentiment analysis of digital texts uses elements from machine learning such as latent semantic analysis, support vector machines, bag-of-words model and semantic orientation.

- Machine learning - a system capable of the autonomous acquisition and integration of knowledge learnt from experience, analytical observation, etc.

Machine learning can be further subdivided into Supervised and Unsupervised learning.

  - Supervised learning - such as Regression Trees, Discriminant Function Analysis, Support Vector Machines.
  - Unsupervised learning - such as Self-Organizing Maps (SOM), K-Means.

Machine Learning aims to solve the problem of having huge amounts of data with many variables and is commonly used in areas such as pattern recognition (speech, images), financial algorithms (credit scoring, algorithmic trading) energy forecasting (load, price) and biology (tumor detection, drug discovery). [22]

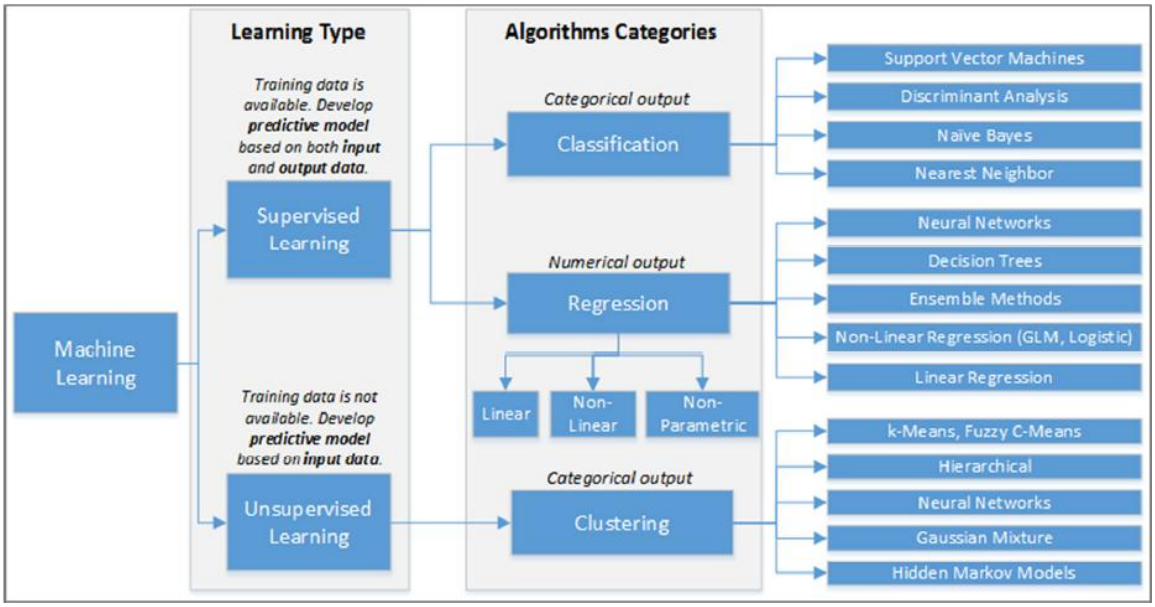


Figure 3.2: Machine Learning Overview

3.6 Scientific programming tools

- **R:** used for statistical programming
- **MATLAB**: used for numeric scientific programming. MATHLAB is significantly faster than the traditional programming languages and can be used for a wide range of applications.
- **Mathematica**: used for symbolic scientific programming (computer algebra).

Python: can be used for (natural) language detection, title and content extraction, query matching and, when used in conjunction with a module such as scikit-learn, it can be trained to perform sentiment analysis, e.g., using a Naï¨veBayes classifier.

3.7 Business toolkits

- **RapidMiner:** provides data mining and machine learning procedures including: data loading and transformation (Extract, Transform, Load, a.k.a. ETL), data preprocessing and visualization, modeling, evaluation, and deployment. RapidMiner is written in Java and uses learning schemes and attribute evaluators from the Weka machine learning environment and statistical modeling schemes from the R project.[22]

3.8 Text Mining capabilities of RapidMiner

RAPIDMINER is the most popular open source software in the world for data mining,and strongly supports text mining and other data mining techniques that are applied in combination with text mining. The common practice in text mining is the analysis of the information extracted through text processing to form new facts and new hypotheses, that can be explored further with other data mining algorithms. Text mining applications typically deal with large and complex data sets of textual documents that contain significant amount of irrelevant and noisy information.

The power and flexibility of RAPIDMINER is due to the GUI-based IDE (integrated development environment) it provides for rapid prototyping and development of data mining models, as well as its strong support for scripting based on XML (extensible mark-up language). The visual modeling in the RAPIDMINER IDE is based on the defining of the data mining process in terms of operators and the flow of process through these operators.Many packages are available for RAPIDMINER, such as text processing, Weka extension, parallel processing, web mining, reporting extension, series processing, PMML, community, and R extension packages. The package that is needed and used for text mining is the Text Processing package, which can be installed and updated through the Update RapidMiner menu item under the Help menu.

3.9 Summary

This chapter describes the technologies available to read and extract social media data and then to process data. The main concern in processing data is being them unstructured natural language. Therefore Natural Language Processing technologies especially text mining capabilities and software toolkits available in the industry has been presented here.

Chapter 4

Approach for Social Media Text Mining in Disaster Situations

4.1 Introduction

Chapter 3 describes how the technology can be adopted to text mining, sentiment analysis which particularly emphasis on data mining and analysis. This chapter describes our approach to computer aided decision making support system, during disaster situations using Social media data posted by citizens and present our hypothesis, input, output, process, users, and features of the decision support system.

4.2 Hypothesis

We hypothesize that the issue of unavailability of a framework for decision support during disaster situations and introducing text mining approach using Social media data including text mining and semantic analysis. This hypothesis has been inspired by the amount of data posted in Social media during disaster situations by citizens with lot of information about the disaster, damages, existing state, nature, and support parties as well as victims.

4.3 Input

Social media data (text) posted by individuals during disaster situations will be the input to this research work. It includes features such as identity, time, and location of the post. These inputs are limited to textual posts shared in social media such as Facebook and twitter of English language.

4.4 Output

Visualization of present accurate information about the disaster in categories such as content building situational awareness, factual subjectivity, action-oriented, supporting decision-making and contribute to an emotion-oriented segment.

4.5 Process

Process of converting inputs into outputs is presented using a diagram in figure 4.1 below.

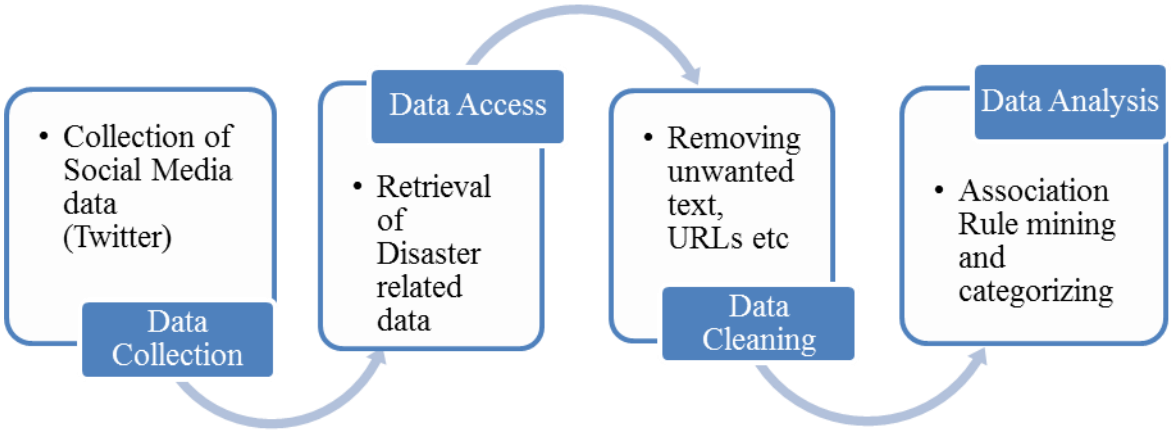


Figure 4.1: Approach to analyze data

4.6 Introduction to Design

In chapter 4 the approach for the research is discussed with the technologies can be used to collect, clean, analyze and visualize data for decision making in natural disaster management using publicly available social media content. Design chapter will discuss architectural design and analysis of the research in a more specific manner. Analysis includes data collection and technology identification and the design will present the components of the proposed system which will present the knowledge discovery at the end of content analysis.

4.7 Top Level Design of the System

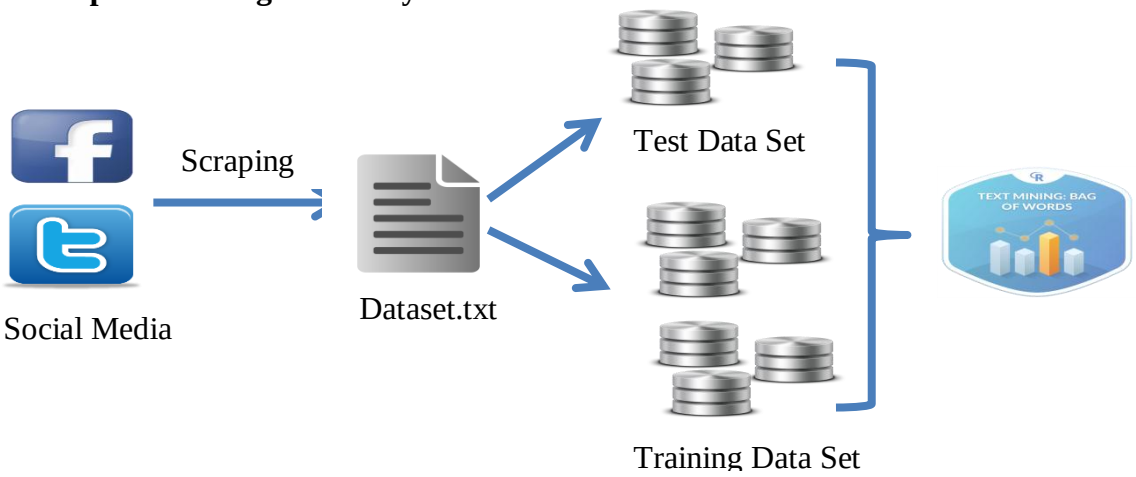


Figure 4.2: Top Level Design of the System

4.8 Summary

This chapter elaborates hypothesis along with input, output, process as well as top level design of the system. System is proposed to design to deliver an output of meaningful information extracted out of social media content available during disaster situations.

Chapter 5

Analysis and Design of Proposed Solution

5.1 Introduction

In the previous chapter, we briefly discussed the approach we have taken to solve the identified problem. This chapter describes the system design which includes two sub systems of Data preprocessing and Association rule mining and comparing rules to identify new data category.

5.2 System Design

Major components of research has been designed as data collection, data preprocessing, tokenizing, mining association rules for categories of decision making and finally detecting the relevant category when new data is entered. Text processing up to association rule generation is proposed to do over RapidMiner and the rest of categorization is done using a software module designed with .Net technologies.

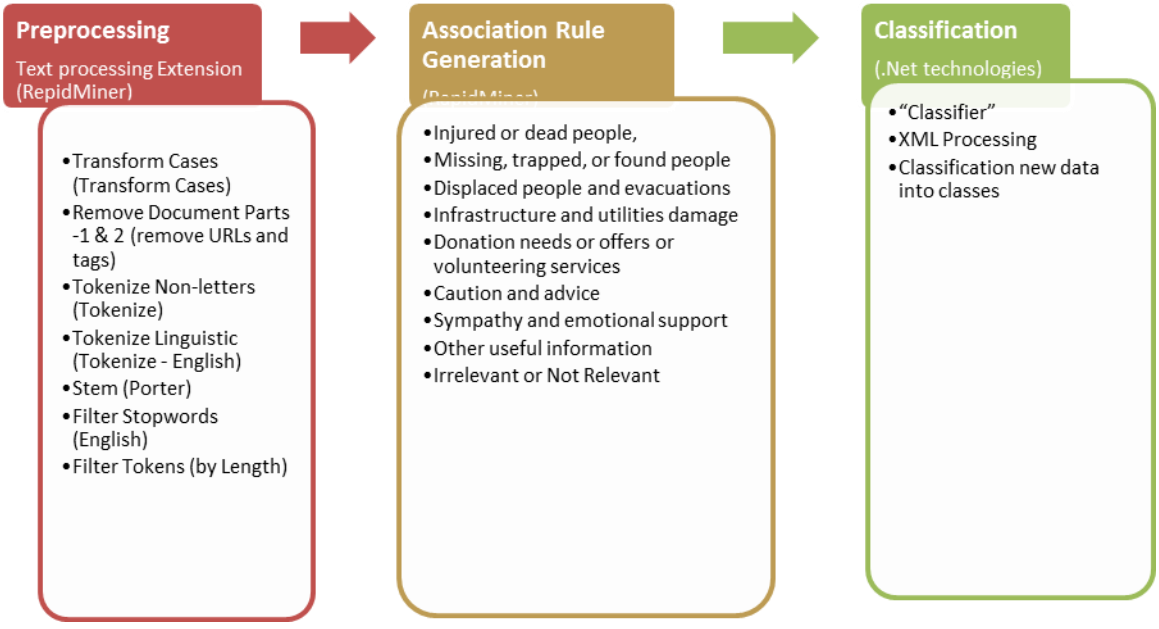


Figure 5.1: Design of Propose System

5.3 RapidMiner Process Model

Preprocessing and association rule generation part will be done using RapidMiner and for text processing we need to install Text Processing Extension. Using that we can preprocess data.

5.3.1 Text Processing

Text mining (also referred to as text data mining or knowledge discovery from textual databases), refers to the process of discovering interesting and non-trivial knowledge from text documents. The common practice in text mining is the analysis of the information extracted through text processing to form new facts and new hypotheses, that can be explored further with other data mining algorithms. Text mining applications typically deal with large and complex data sets of textual documents that contain significant amount of irrelevant and noisy information. Feature selection aims to remove this irrelevant and noisy information by focusing only on relevant and informative data for use in text mining. Some of the topics within text mining include feature extraction, text categorization, clustering, trends analysis, association mining and visualization.

5.3.2 RapidMiner Text Processing Extension Package

Many packages are available for RAPIDMINER, such as text processing, Weka extension, parallel processing, web mining, reporting extension, series processing, PMML, community, and R extension packages. The package that is needed and used for text mining is the Text Processing package, which can be installed and updated through the Update RapidMiner menu item under the Help menu.

5.4 Classifier - .Net Technology based tool

In order to classify data into meaningful categories a tool is developed using .Net technologies. It has been developed using C# in Visual Studio Environment. Microsoft Visual Studio is an integrated development environment (IDE) from Microsoft. It is used to develop computer programs, as well as web sites, web apps, web services and mobile apps. Visual Studio uses Microsoft software development platforms such as Windows API, Windows Forms, Windows Presentation Foundation, Windows Store and Microsoft Silverlight. It can produce both native code and managed code.

5.4.1 C#

C# is a multi-paradigm programming language encompassing strong typing, imperative, declarative, functional, generic, object-oriented (class-based), and component-oriented programming disciplines. It was developed by Microsoft within its .NET initiative and later approved as a standard by Ecma (ECMA-334) and ISO (ISO/IEC 23270:2006). C# is one of the programming languages designed for the Common Language Infrastructure.

C# is a general-purpose, object-oriented programming language. Its development team is led by Anders Hejlsberg. The most recent version is C# 7.2, which was released in 2017 along with Visual Studio 2017 version

5.4.2 Classifier Design

Input : 1. Set of Association rules belong to 8 classes (in XML format)

2. Incoming twitter data

Process: Classifying twitter data into 9 classes

(8 defined classes and1 irrelevant class)

Output: Twitter messages categorized in to 9 classes

5.6 Summary

In detail analysis on system design is presented here. After generating association rules from RapidMiner using its text mining extension a classifier will be developed to classify tweets into meaningful classes. To develop the classifier .Net technologies will be used.

Chapter 6

Implementation

6.1 Introduction

In this chapter implementation of the proposed solution will be elaborated. In order to derive situational awareness during disaster situations it is very important to have a real dataset therefore we will be using few of datasets among human-annotated Twitter corpora collected during 19 different crises that took place between 2013 and 2015

6.2 Challenges in proposed system implementation

The very first challenge faced was to find a suitable data set and for that we could build a link with one of key researchers in the area Muhammad Imran, Qatar Computing Research Institute (HBKU), Doha, Qatar who has contributed to disaster related social media text mining and developed many tools and mainly the real time twitter data collection application: AIDR (Artificial Intelligence for Disaster Response). With availability of crisis-related posts collected from Twitter, human-labeled tweets, dictionaries of out-of-vocabulary (OOV) words, word2vec embeddings, and other related tools[4] we were able to extract real world data set consisting of tweets from Twitter. The resource consisting with human labeled data annotated by paired workers, annotated by volunteers, Word2vec embeddings trained using crisis-related tweets and Out-Of-Vocabular (OOV) words and their meanings. Further actual disaster related data available from 19 crises from 2013 to 2015 categorized into crisis types with the countries as

- Earthquake: Nepal,Chile, USA, Pakistan
- Typhoon: Vanuatu, Phillippines, Mexico
- Volcano: Iceland, Floods: Pakistan,India
- War & Conflicts: Palestine & Israel, Pakistan
- Biological: Middle East Respiratory Syndrome (MERS) Worldwide, Ebola Virus Outbreak Worldwide

[4] http://crisisnlp.qcri.org/

- Landslide: Worldwide and
- Airline Accident: Malaysia.

These tweets were in languages English, Spanish and French and considering the applicability of the crises in Sri Lankan context we have selected India and Pakistan flood (2014) dataset since flood is the most common natural disaster type occurs in Sri Lanka frequently affecting people catastrophically.

Two datasets[5] were consisting of 1,236,610 and 5,259,681 tweets in English collected during 2014-09-06 to 2014-09-06: Pakistan Floods 2014 and 2014-08-10 to 2014-09-03: India Floods 2014. Datasets were consisting of tweet-ids, user-ids only therefore Imran, Mitra & all have published a tool to download full tweets content from Twitter. Tweets downloader tool has been written in Java and the tool can make 180 API calls per 15 minutes, which each API call allows getting up to 100 tweets. (i.e. it can download up to 72,000 tweets per hour) [23]

6.3 Downloading tweets using Tweets Retrieval Tool (V1.2)

To download tweets using tool it requires four tokens "consumer.key", "consumer.secret", "access.token", "access.token.secret" from a twitter developer app. Therefore the first step was to create twitter developer app and the apps.twitter.com application interface offers the ability to generate an OAuth access token for the owner of the application to make requests on behalf of a single user establishing a connection to the Streaming API.

Once logged in, using button to create a new application we can create an application to generate tokens. These token are used to authenticate requests to the Twitter Platform.

Once the tokens are collected following steps can be followed to proceed with downloading tweets.

1. Put the tweets ids in a text file, one per line.

[5] https://en.wikipedia.org/wiki/2014_India%E2%80%93Pakistan_floods

2. Get following four tokens from a Twitter app. Once obtained, put them into the twitter.properties file available with the tool.

   "consumer.key", "consumer.secret", "access.token", "access.token.secret"

3. Run the JAR file as shown in the following command. The command needs two parameters. The first parameter is the file containing tweets-ids. And, the second parameter is the path and name of output file where the tool should store the downloaded tweets.

   "java -classpath TweetsRetrieval-1.2-jar-with-dependencies.jar qa.qcri.tweetsretrieval.TweetsRetrievalTool sample_tweet_ids.txt output.txt"

Following figure 3 shows how the data downloaded through Twitter Streaming API using the Tweets Retrieval Tool v1.2

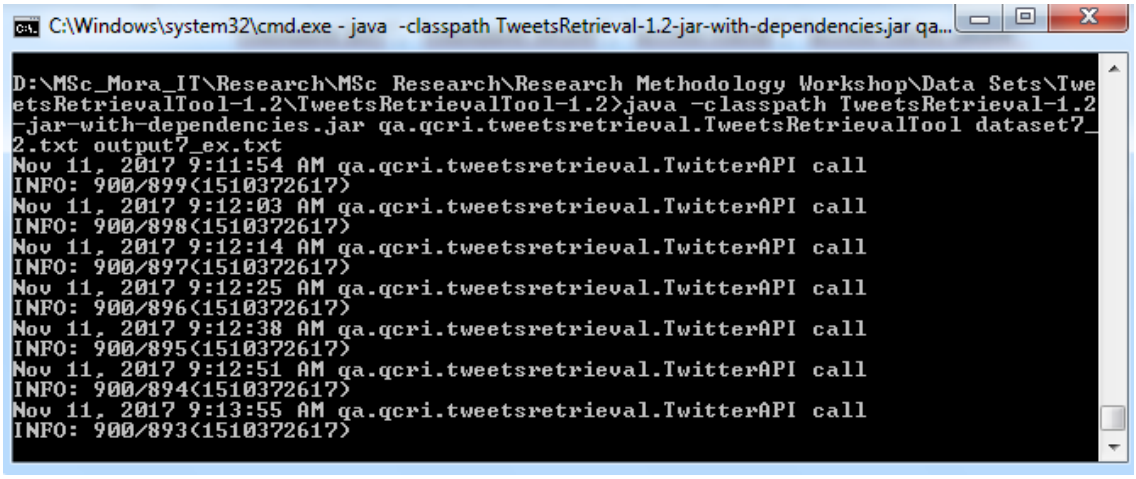


Figure 6.1: shows how the data downloaded through Twitter Streaming API using the Tweets Retrieval Tool v1.2

Each raw of tweets collected consisting with date, time, tweet-id, text (tweet) and many more other information. Following figure shows some of tweets collected using the tool.

Figure 6.2: Sample of full texts of tweets downloaded

The data is in JSON format. JSON stands for JavaScript Object Notation. This format makes it easy to humans to read the data, and for machines to parse it.

6.4 Labeled Data

Annotation Scheme used in this work has been taken from the previous research published by Muhammad Imaran and team in their paper titled Twitter as a Lifeline: Human-nnotated Twitter Copora for NLP of crisis-related messages. They have taken these annotation schemes using input taken from formal crisis response agencies such as United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA).

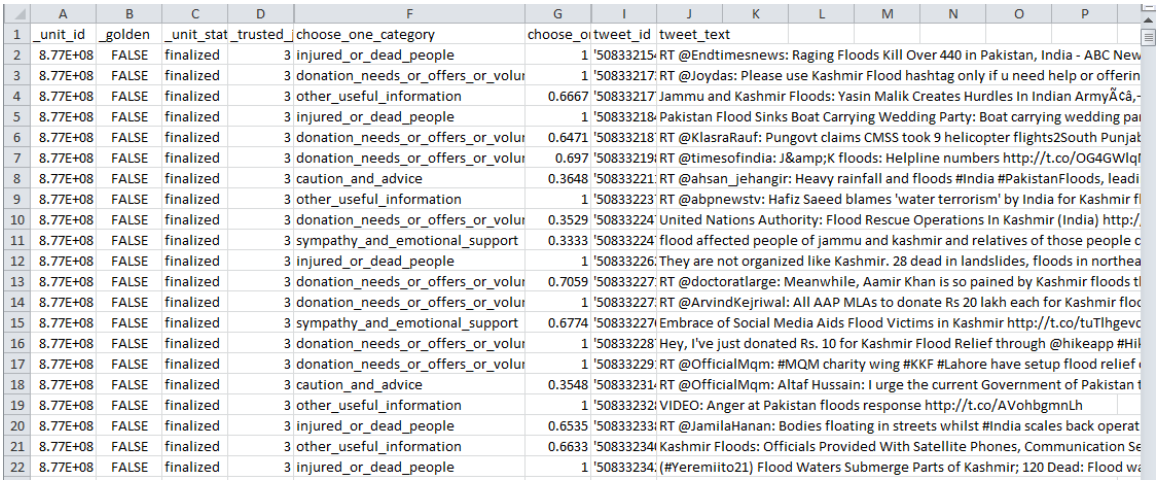
6.5 Annotation Scheme

Categorizing messages by information types into 9 categories

- **Injured or dead people**: Reports of casualties and/or injured people due to the crisis
- **Missing, trapped, or found people**: Reports and/or questions about missing or found people
- **Displaced people and evacuations**: People who have relocated due to the crisis, even for a short time (includes evacuations)
- **Infrastructure and utilities damage**: Reports of damaged buildings, roads, bridges, or utilities/services interrupted or restored

- **Donation needs or offers or volunteering services**: Reports of urgent needs or donations of shelter and/or supplies such as food, water, clothing, money, medical supplies or blood; and volunteering services
- **Caution and advice**: Reports of warnings issued or lifted, guidance and tips
- **Sympathy and emotional support**: Prayers, thoughts, and emotional support
- **Other useful information**: Other useful information that helps understand the situation
- **Not related or irrelevant**: Unrelated to the situation or irrelevant

The 9 category types (including two catch-all classes: “Other Useful Information” and “Irrelevant”) used by the UN OCHA



	A	B	C	D	F	G	I	J	K	L	M	N	O	P
1	_unit_id	_golden	_unit_stat	_trusted_	choose_one_category	choose_o	tweet_id	tweet_text						
2	8.77E+08	FALSE	finalized	3	injured_or_dead_people	1	'50833215	RT @Endtimesnews: Raging Floods Kill Over 440 in Pakistan, India - ABC New						
3	8.77E+08	FALSE	finalized	3	donation_needs_or_offers_or_volur	1	'50833217	RT @Joydas: Please use Kashmir Flood hashtag only if u need help or offerin						
4	8.77E+08	FALSE	finalized	3	other_useful_information	0.6667	'50833217	Jammu and Kashmir Floods: Yasin Malik Creates Hurdles In Indian ArmyÃ¢â‚¬						
5	8.77E+08	FALSE	finalized	3	injured_or_dead_people	1	'50833218	Pakistan Flood Sinks Boat Carrying Wedding Party: Boat carrying wedding pa						
6	8.77E+08	FALSE	finalized	3	donation_needs_or_offers_or_volur	0.6471	'50833218	RT @KlasraRauf: Pungovt claims CMSS took 9 helicopter flights2South Punjat						
7	8.77E+08	FALSE	finalized	3	donation_needs_or_offers_or_volur	0.697	'50833219	RT @timesofindia: J&K floods: Helpline numbers http://t.co/OG4GWlqI						
8	8.77E+08	FALSE	finalized	3	caution_and_advice	0.3648	'50833221	RT @ahsan_jehangir: Heavy rainfall and floods #India #PakistanFloods, leadi						
9	8.77E+08	FALSE	finalized	3	other_useful_information	1	'50833223	RT @abpnewstv: Hafiz Saeed blames 'water terrorism' by India for Kashmir fl						
10	8.77E+08	FALSE	finalized	3	donation_needs_or_offers_or_volur	0.3529	'50833224	United Nations Authority: Flood Rescue Operations In Kashmir (India) http:/						
11	8.77E+08	FALSE	finalized	3	sympathy_and_emotional_support	0.3333	'50833224	flood affected people of jammu and kashmir and relatives of those people c						
12	8.77E+08	FALSE	finalized	3	injured_or_dead_people	1	'50833226	They are not organized like Kashmir. 28 dead in landslides, floods in northea						
13	8.77E+08	FALSE	finalized	3	donation_needs_or_offers_or_volur	0.7059	'50833227	RT @doctoratlarge: Meanwhile, Aamir Khan is so pained by Kashmir floods tl						
14	8.77E+08	FALSE	finalized	3	donation_needs_or_offers_or_volur	1	'50833227	RT @ArvindKejriwal: All AAP MLAs to donate Rs 20 lakh each for Kashmir floc						
15	8.77E+08	FALSE	finalized	3	sympathy_and_emotional_support	0.6774	'50833227	Embrace of Social Media Aids Flood Victims in Kashmir http://t.co/tuTlhgevc						
16	8.77E+08	FALSE	finalized	3	donation_needs_or_offers_or_volur	1	'50833228	Hey, I've just donated Rs. 10 for Kashmir Flood Relief through @hikeapp #Hil						
17	8.77E+08	FALSE	finalized	3	donation_needs_or_offers_or_volur	1	'50833229	RT @OfficialMqm: #MQM charity wing #KKF #Lahore have setup flood relief						
18	8.77E+08	FALSE	finalized	3	caution_and_advice	0.3548	'50833231	RT @OfficialMqm: Altaf Hussain: I urge the current Government of Pakistan t						
19	8.77E+08	FALSE	finalized	3	other_useful_information	1	'50833232	VIDEO: Anger at Pakistan floods response http://t.co/AVohbgmnLh						
20	8.77E+08	FALSE	finalized	3	injured_or_dead_people	0.6535	'50833233	RT @JamilaHanan: Bodies floating in streets whilst #India scales back operat						
21	8.77E+08	FALSE	finalized	3	other_useful_information	0.6633	'50833234	Kashmir Floods: Officials Provided With Satellite Phones, Communication Se						
22	8.77E+08	FALSE	finalized	3	injured_or_dead_people	1	'50833234	(#Yeremiito21) Flood Waters Submerge Parts of Kashmir; 120 Dead: Flood wa						

Figure 6.3: Annotated tweets by paid workers from the Crowdflower crowdsourcing platform for labeling

Categorizing tweets in to above nine categories has been done by paid workers from the Crowdflower crowdsourcing platform for labeling. It is done as at least three different workers were required to agree on a label before a task is finalized. No worker has been allowed to perform more than 200 tasks in total 9 categories were used in this task, as described above.

6.6 Association Rule Mining

Using RapidMiner association rules for each human annotated category were generated after preprocessing twitter content.

Row ...	category	tweet_text
1	displaced_people_and_evacuations	RT @SAMAATV: #PakistanFloods affect more than one million http://t.co/OzlQ6TKSgU http://t.co/V67Rep6FdZ
2	displaced_people_and_evacuations	RT @NorthAndrew: Srinagar residents on makeshift bridge +10 days since flood #KashmirFloods #IndiaPakistanFlood...
3	displaced_people_and_evacuations	RT @dna: Jammu and Kashmir: Indian Army rescues 11,000 people from floods, 100 columns deployed http://t.co/VgB...
4	displaced_people_and_evacuations	RT @RadioPakistan: Over 90 thousand people rescued in flood hit areas of #Multan http://t.co/alteDtxyL2
5	displaced_people_and_evacuations	JKLF chief Yasin Malik 'hijacks' rescue boat in flood-hit Jammu and Kashmir http://t.co/bUwN5G33C1
6	displaced_people_and_evacuations	Over 1,84,000 People Rescued in Flood-Hit Jammu and Kashmir So Far: Government: More than 1,84,000 people have...
7	displaced_people_and_evacuations	Jammu & Kashmir floods: Rajnath Singh reviews situation, massive rescue Ã¢â‚¬Â¦ Ã¢â‚¬â€œ Times ofÃ‚Â India ...
8	displaced_people_and_evacuations	#USAHeadlines Evacuation ordered as flood death toll rises in Pakistan and India http://t.co/nDJQWegDIO

Figure 6.4: Annotated data after loading to RapidMiner with Category and Tweet-text attributes

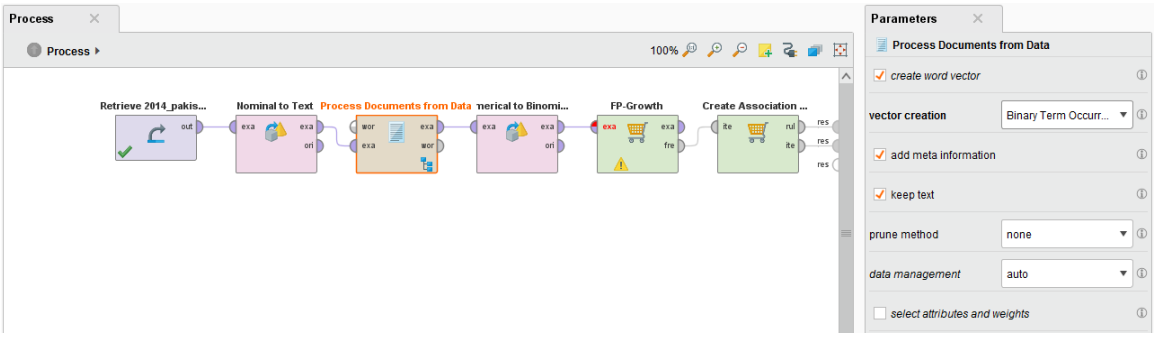


Figure 6.5: Operators for Process and the Parameters for Process Documents from Files operator

Process Documents from Files operator performs text processing which involves preparing the text data for the application of conventional data mining techniques. Process Documents from Files operator reads data from a collection of text files and manipulates this data using text processing algorithms. This is a nested operator, meaning that it can contain a sub-process consisting of a multitude of operators. Indeed, in this Process, this nested operator contains other operators inside. This sub-process consists of eight operators that are serially linked (Figure 12):

- Transform Cases (Transform Cases)
- Remove Document Parts -1 & 2 (remove URLs and tags)

- Tokenize Non-letters (Tokenize)
- Tokenize Linguistic (Tokenize - English)
- Stem (Porter)
- Filter Stopwords (English)
- Filter Tokens (by Length)

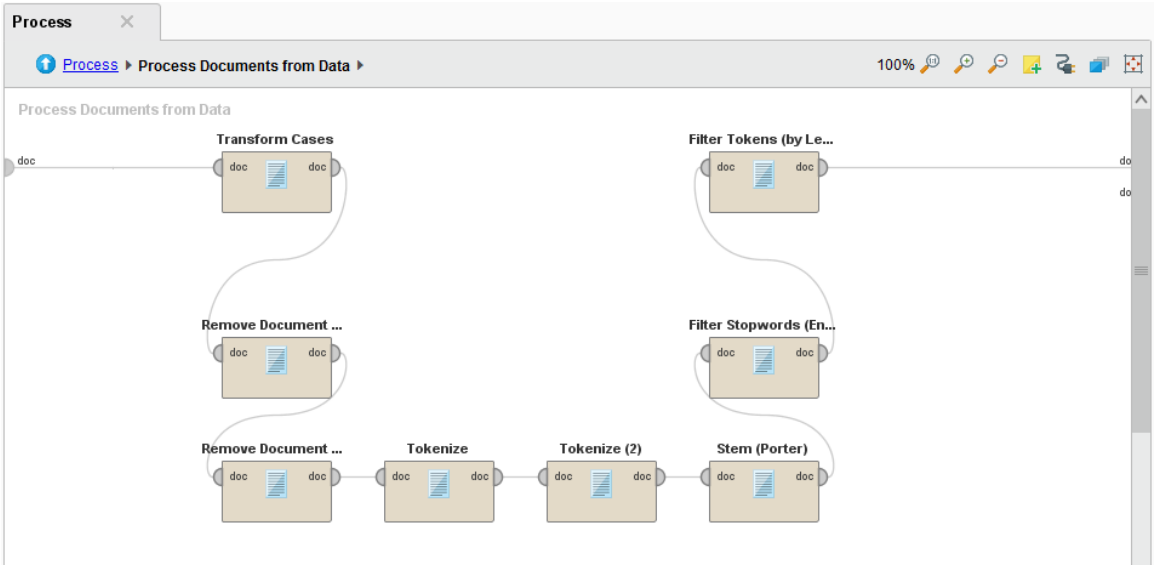
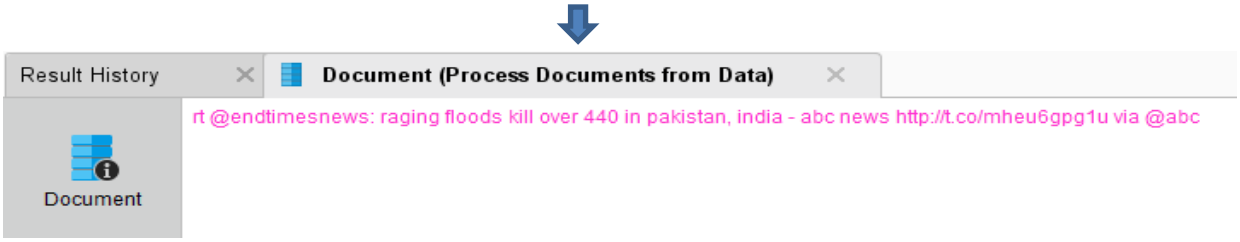


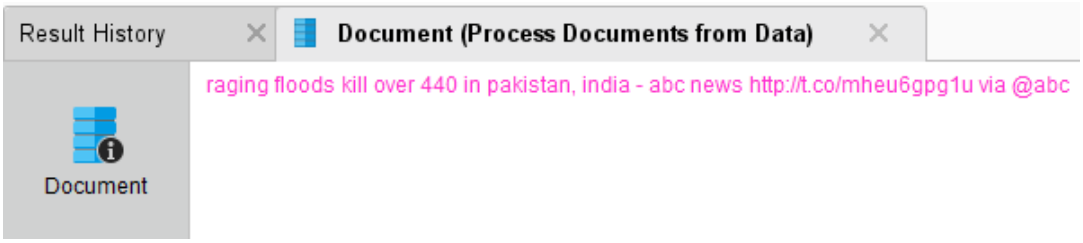
Figure 6.6: Operators within the Process Documents from Files nested operator

6.6.1 Text processing step by step

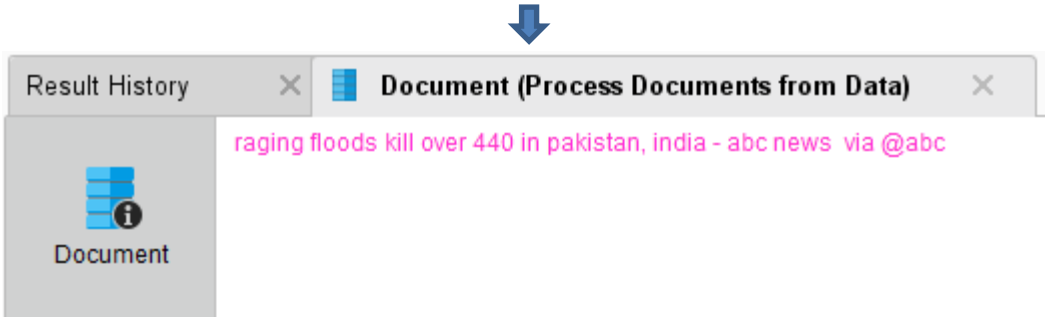
1. Transform cases



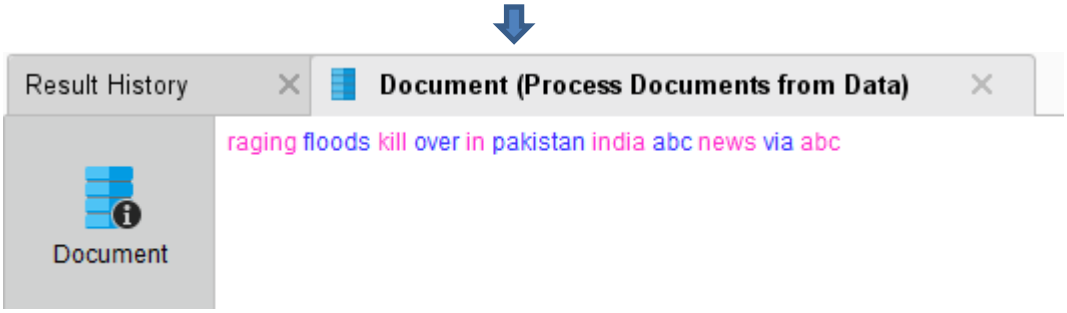
2. Remove Document Parts -1 (Tags are removed)



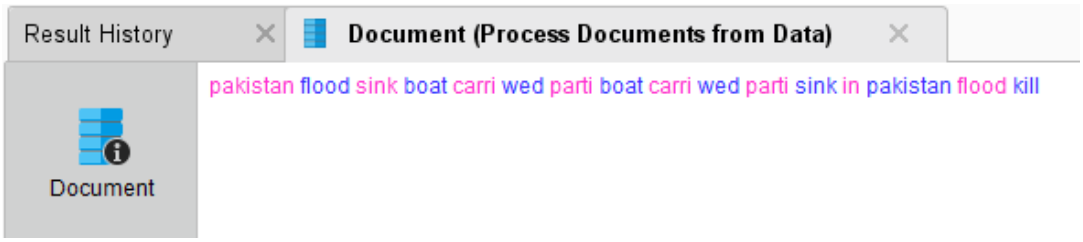
1. Remove Document Parts -2 (URLs are removed)



4. Tokenize Non-letters & Linguistic (Tokenize - English)



5. Stem (Porter)



6. Filter Stopwords (English)

7. Filter Tokens (by Length)

Figure 6.7: Process Documents from Files sub-process output

The sub-process basically transforms the text data into a format that can be easily analyzed using conventional data mining techniques such as association mining and cluster modeling.

In this sub-process, the Tokenize Non-letters (Tokenize) and Tokenize Linguistic (Tokenize) operators are both created by selecting the Tokenize operator, but with different parameter selections. The former operator tokenizes based on non-letters whereas the latter operator tokenizes based on the linguistic sentences within the English language. The Filter Stop words (English) operator removes the stop words in the English language from the text data set. The Filter Tokens (by Length) operator removes all the words composed of less than min chars characters and more than max chars characters. In this example, words that have less than 2 characters or more than 25 characters are removed from the data set. The Stem (Porter) operator performs stemming and the Transform Cases(2) (Transform Cases) operator transforms all the characters in the text into lower case.

The Numerical to Binomial operator transforms the data into binominal form. This means that each row represents a tweet, a few columns provide metadata about that tweet and the remaining columns represent the words appearing in all the tweets, with the cell contents telling (true or false) whether that word exist in that tweet or not. FP-Growth algorithm is used for identifying the frequent item sets. In this example, the min support parameter is 0.9, meaning that the operator generates a list of the frequent sets of words (itemsets) that appear in at least 90% of the tweets.

The final operator in this process, namely Create Association Rules, receives the list of frequent word sets from the FP-Growth operator, and computes the rules that satisfy the specified constraints on selected association mining criteria. In this

example, the association rules are computed according to the the criterion of confidence, as well as gain theta and laplace k. The specified minimal values for these 3 criteria are 0.9, 2.0 and 1.0, respectively.

The final result is the set of association rules. In the Table View, table grid presents the generated association rules, with one rule in each row. For example, the first row states “IF kashmir THEN flood” with a Support level of 0.321 and Confidence level of 1.00. This rule means that in 32 of the 100 documents, words with stem kashmir and flood appear together. Furthermore in 100% of the documents where a word derived from the stem kashmir appears, at least one word derived from the stem flood is observed. On the left hand side of the grid we can select and Show rules matching a particular set of words.

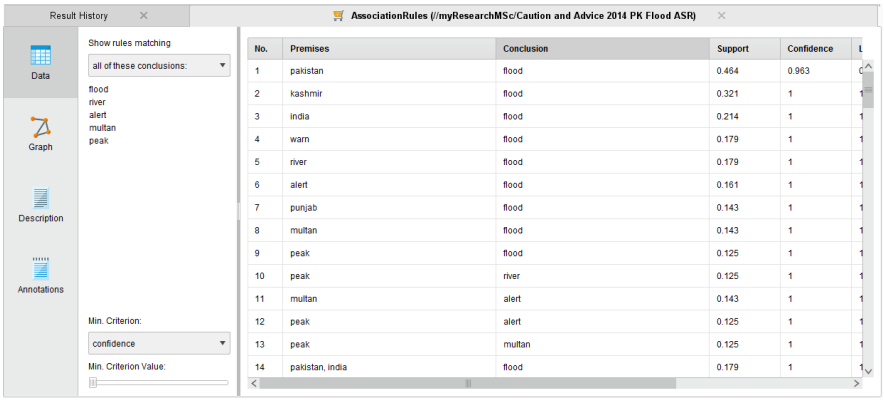


Figure 6.8: Table View for the Association Rules generated by the Process

6.7 Clustering tweets using Association Rules

After generating association rules for each category, as the second module of this research, a tool is created to classify tweets automatically using Association Rules for each category name “Classifier” using .Net technologies. The development language used is C# and to input association rules to the classifier XML format has been used. Using the write operator in RapidMiner all the association rules generated was stored in a XML.

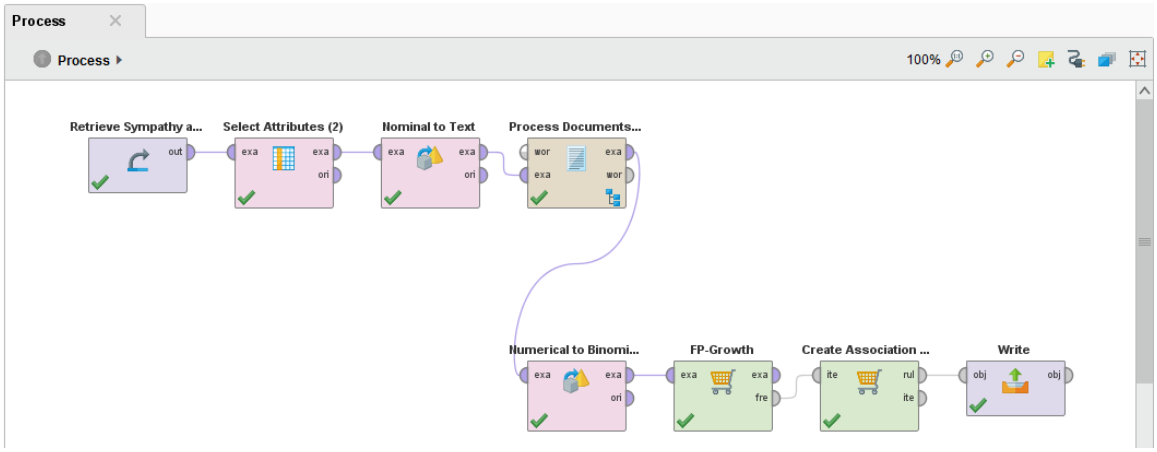


Figure 6.9: Using write operator to write association rules into XML

```
<object-stream>
  <AssociationRules id="1" serialization="custom">
    <com.rapidminer.operator.AbstractIOObject>
      <default>
        <source>Create Association Rules</source>
      </default>
    </com.rapidminer.operator.AbstractIOObject>
    <com.rapidminer.operator.ResultObjectAdapter>
      <default>
        <annotations id="2">
          <keyValueMap id="3"/>
        </annotations>
      </default>
    </com.rapidminer.operator.ResultObjectAdapter>
    <AssociationRules>
      <default>
        <associationRules id="4">
          <com.rapidminer.operator.learner.associations.AssociationRule id="5">
            <confidence>0.9629629629629629</confidence>
            <totalSupport>0.4642857142857143</totalSupport>
            <lift>0.9804713804713805</lift>
            <laplace>0.9879518072289157</laplace>
            <gain>-0.5</gain>
```

Figure 6.10: XML file of Association Rules

6.8 Classifier

An algorithm is written to retrieve association rules of all 9 categories and then to remove common rules for all categories since those rules cannot be used for classification. To uniquely identify the category of a new incoming tweet, the tweet is processed to remove stop words after tokenization and then stemming is done with porter stemmer. Then the key words of the tweet compared with all the association rules in all 9 categories and the rule having highest confidence out of all matching

rules identified and the category of that rule is considered as the category of new tweet.

Any tweet not matching with any of 9 classes will be categorized as irrelevant for the subjective disaster. Such data is not considered and once the data has been automatically categorized into 9 classes, those sub sets of data can be used to relief related or other humanitarian efforts related to disaster.

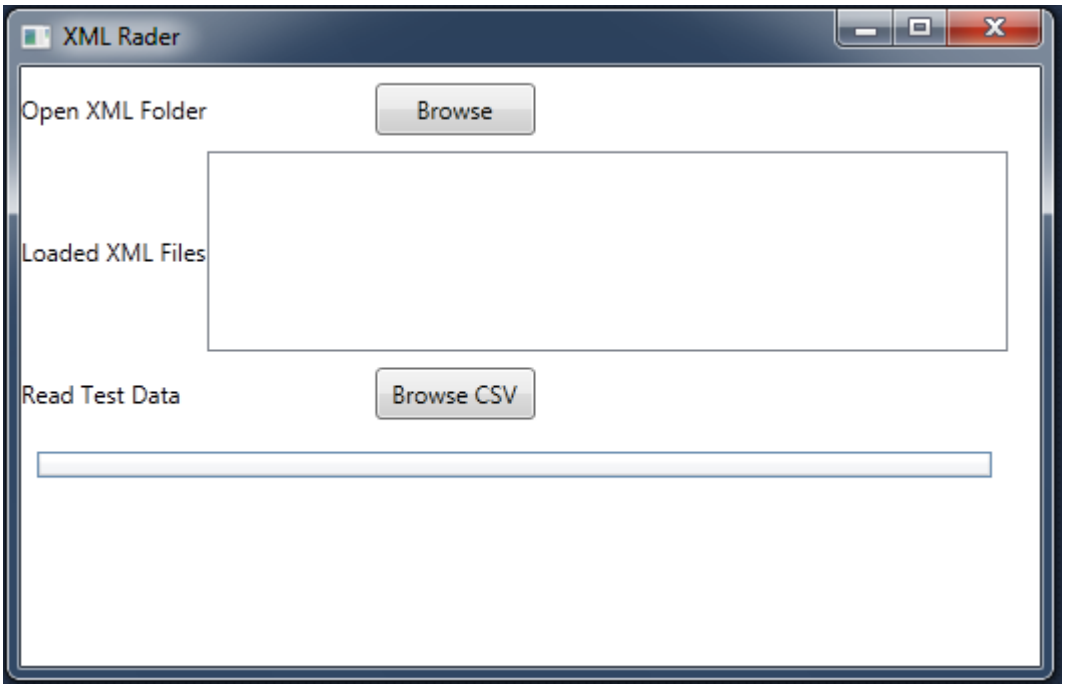


Figure 6.11 Main interface of Classifier

All the XML files of association rules should be stored in a folder and that folder can be uploaded using browse button of the application. Then the data set can be uploaded in CSV format and after that algorithm will be applied and final result will be displayed as follows. Result will be displayed as a summary of number of tweets categorized in to each category. In order to evaluate the algorithm, first a labeled data set is given and following output is given by the classifier.

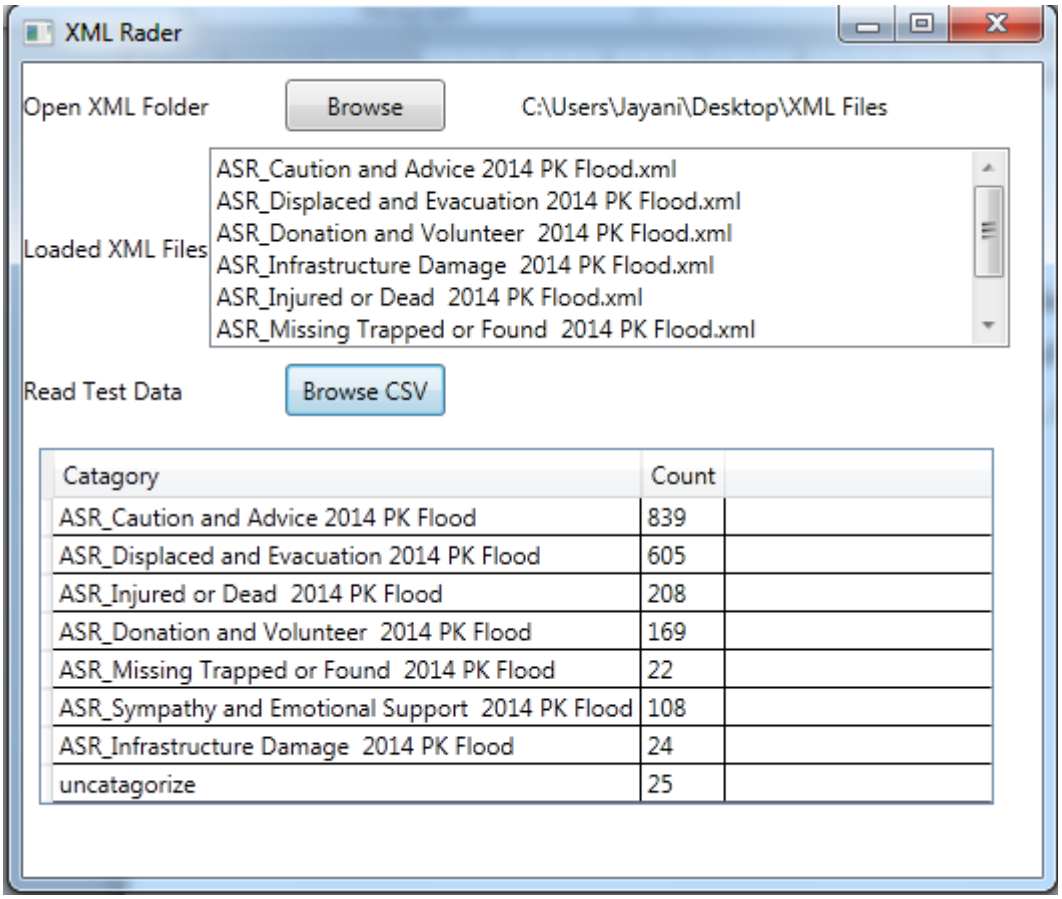


Figure 6.12: Categorized tweets

6.9 Summary

This chapter describes how the proposed solution has been implemented from the starting point of downloading tweets through twitter API using the Id's available at CrisisNLP platform [23]. Human annotated data as well as real tweets posted during actual natural disasters has been taken to demonstrate the classification of tweets to generate situational awareness to assist decision making. Next chapter describes the results of classification.

Chapter 7

Evaluation

7.1 Introduction

The previous chapter discussed the details on implementation of all the modules of the proposed solution. This chapter justifies and evaluates the results and the models generated in each module.

7.2 Evaluation of Association Rule Mining

In this work association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true. The frequent if/then patterns are mined using the operators like the FP-Growth operator. The Create Association Rules operator takes these frequent itemsets and generates association rules.

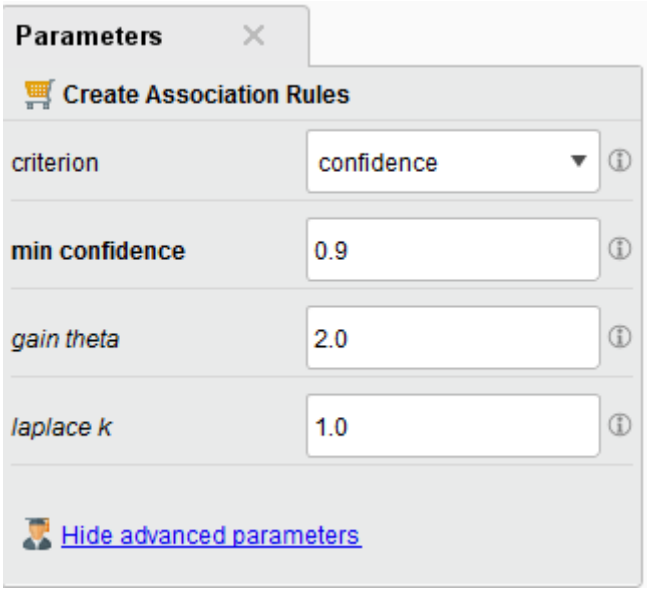
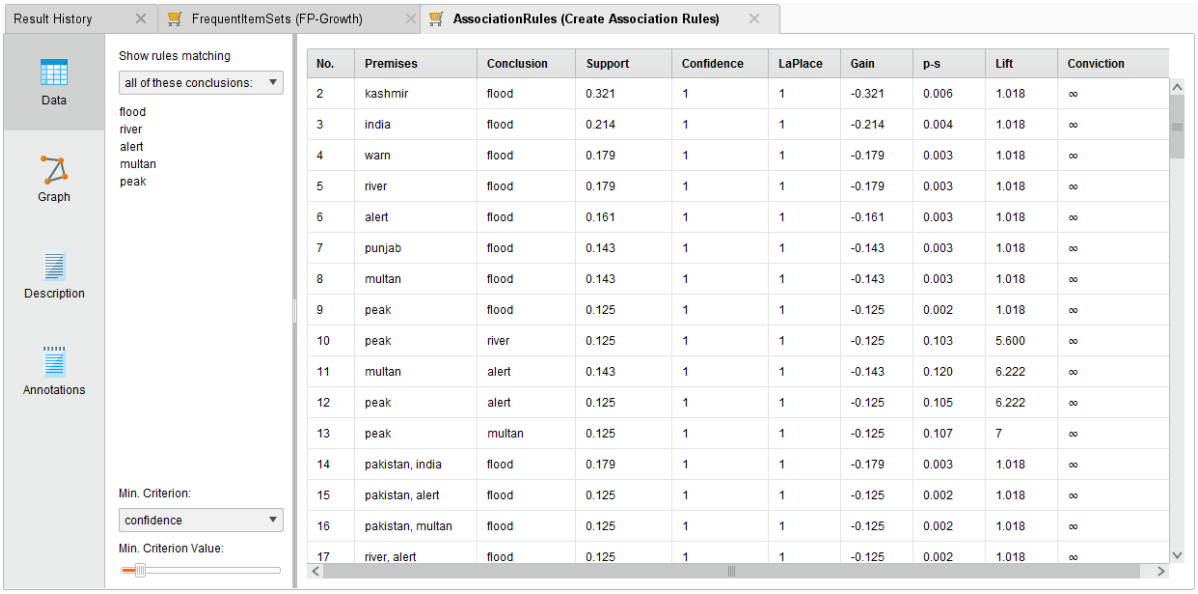


Figure 7.1: Parameters of Create Association Rules Operator



No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Conviction
2	kashmir	flood	0.321	1	1	-0.321	0.006	1.018	∞
3	india	flood	0.214	1	1	-0.214	0.004	1.018	∞
4	warn	flood	0.179	1	1	-0.179	0.003	1.018	∞
5	river	flood	0.179	1	1	-0.179	0.003	1.018	∞
6	alert	flood	0.161	1	1	-0.161	0.003	1.018	∞
7	punjab	flood	0.143	1	1	-0.143	0.003	1.018	∞
8	multan	flood	0.143	1	1	-0.143	0.003	1.018	∞
9	peak	flood	0.125	1	1	-0.125	0.002	1.018	∞
10	peak	river	0.125	1	1	-0.125	0.103	5.600	∞
11	multan	alert	0.143	1	1	-0.143	0.120	6.222	∞
12	peak	alert	0.125	1	1	-0.125	0.105	6.222	∞
13	peak	multan	0.125	1	1	-0.125	0.107	7	∞
14	pakistan, india	flood	0.179	1	1	-0.179	0.003	1.018	∞
15	pakistan, alert	flood	0.125	1	1	-0.125	0.002	1.018	∞
16	pakistan, multan	flood	0.125	1	1	-0.125	0.002	1.018	∞
17	river, alert	flood	0.125	1	1	-0.125	0.002	1.018	∞

Figure 7.2 : Data (Tabular) view Association Rules

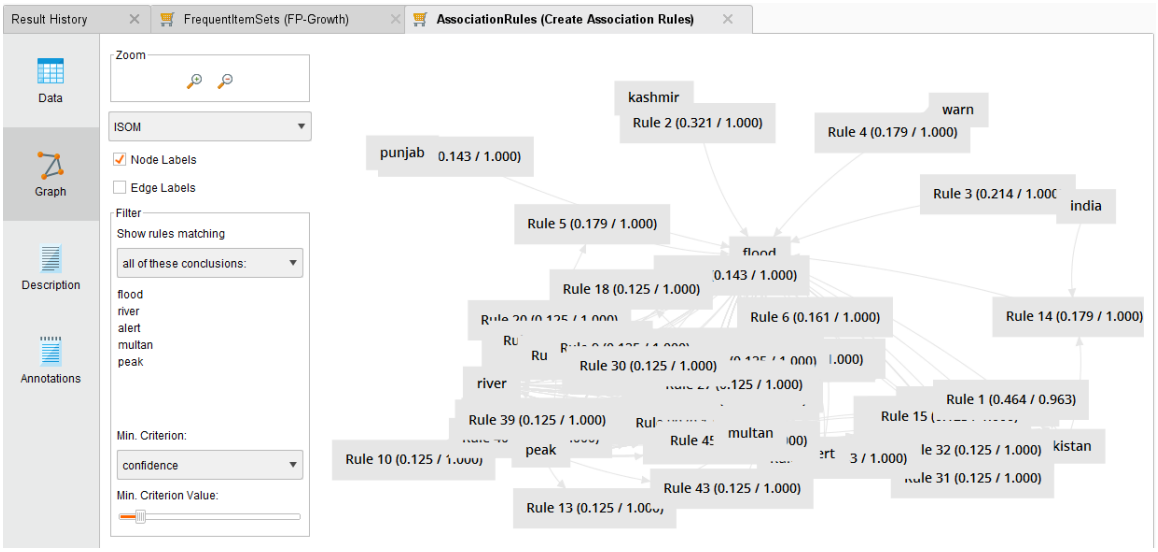


Figure 7.3: Graph view of Association Rules

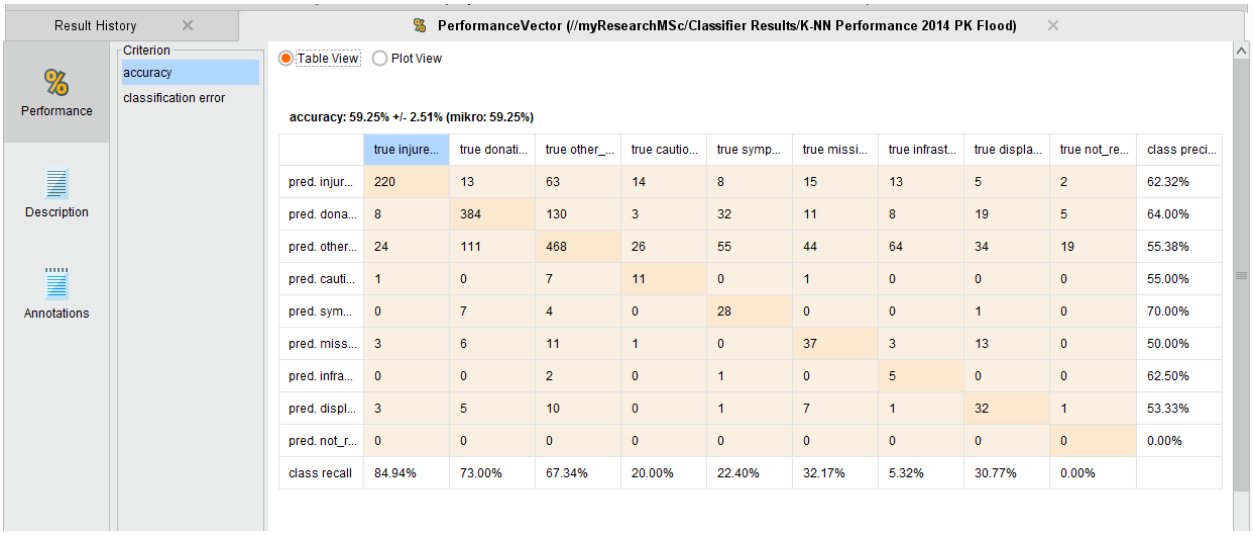
AssociationRules

```
Association Rules
[pakistan] --> [flood] (confidence: 0.963)
[kashmir] --> [flood] (confidence: 1.000)
[india] --> [flood] (confidence: 1.000)
[warn] --> [flood] (confidence: 1.000)
[river] --> [flood] (confidence: 1.000)
[alert] --> [flood] (confidence: 1.000)
[punjab] --> [flood] (confidence: 1.000)
[multan] --> [flood] (confidence: 1.000)
[peak] --> [flood] (confidence: 1.000)
[peak] --> [river] (confidence: 1.000)
[multan] --> [alert] (confidence: 1.000)
[peak] --> [alert] (confidence: 1.000)
[peak] --> [multan] (confidence: 1.000)
[pakistan, india] --> [flood] (confidence: 1.000)
[pakistan, alert] --> [flood] (confidence: 1.000)
[pakistan, multan] --> [flood] (confidence: 1.000)
[river, alert] --> [flood] (confidence: 1.000)
[river, multan] --> [flood] (confidence: 1.000)
[peak] --> [flood, river] (confidence: 1.000)
[flood, peak] --> [river] (confidence: 1.000)
[river, peak] --> [flood] (confidence: 1.000)
[multan] --> [flood, alert] (confidence: 1.000)
[flood, multan] --> [alert] (confidence: 1.000)
[alert, multan] --> [flood] (confidence: 1.000)
[peak] --> [flood, alert] (confidence: 1.000)
[flood, peak] --> [alert] (confidence: 1.000)
[alert, peak] --> [flood] (confidence: 1.000)
[peak] --> [flood, multan] (confidence: 1.000)
[flood, peak] --> [multan] (confidence: 1.000)
```

Figure 7.4: Association Rules of Caution and Advice Category

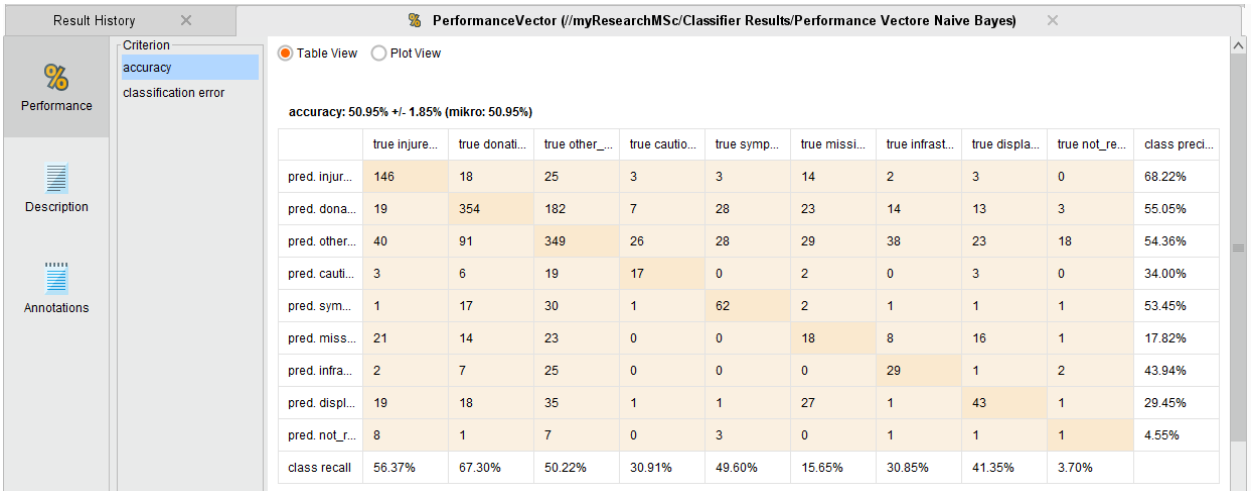
7.3 Evaluation of different classifiers

According to Literature, to classify text well-known learning algorithms have been used such as Naive Bayes (NB), Support Vector Machines(SVM), and Random Forest (RF). When running our dataset with classifiers such as K-NN and Naïve Bayes following outputs were received.



	true injure...	true donati...	true other_...	true cautio...	true symp...	true missi...	true infrast...	true displa...	true not_re...	class preci...
pred. injur...	220	13	63	14	8	15	13	5	2	62.32%
pred. dona...	8	384	130	3	32	11	8	19	5	64.00%
pred. other...	24	111	468	26	55	44	64	34	19	55.38%
pred. cauti...	1	0	7	11	0	1	0	0	0	55.00%
pred. sym...	0	7	4	0	28	0	0	1	0	70.00%
pred. miss...	3	6	11	1	0	37	3	13	0	50.00%
pred. infra...	0	0	2	0	1	0	5	0	0	62.50%
pred. displ...	3	5	10	0	1	7	1	32	1	53.33%
pred. not_r...	0	0	0	0	0	0	0	0	0	0.00%
class recall	84.94%	73.00%	67.34%	20.00%	22.40%	32.17%	5.32%	30.77%	0.00%	

Figure 7.5 – Performance of K-NN algorithm



	true injure...	true donati...	true other_...	true cautio...	true symp...	true missi...	true infrast...	true displa...	true not_re...	class preci...
pred. injur...	146	18	25	3	3	14	2	3	0	68.22%
pred. dona...	19	354	182	7	28	23	14	13	3	55.05%
pred. other...	40	91	349	26	28	29	38	23	18	54.36%
pred. cauti...	3	6	19	17	0	2	0	3	0	34.00%
pred. sym...	1	17	30	1	62	2	1	1	1	53.45%
pred. miss...	21	14	23	0	0	18	8	16	1	17.82%
pred. infra...	2	7	25	0	0	0	29	1	2	43.94%
pred. displ...	19	18	35	1	1	27	1	43	1	29.45%
pred. not_r...	8	1	7	0	3	0	1	1	1	4.55%
class recall	56.37%	67.30%	50.22%	30.91%	49.60%	15.65%	30.85%	41.35%	3.70%	

Figure 7.6 – Performance of Naïve Bayes algorithm

7.4 Summary

Following table shows the comparison of tweet categorization to 9 pre-defined classes by two standard classifiers as well as the new association rule based classifier introduced in this work. Final output shows that the classifier developed in this work classifies tweets into some classes with higher accuracy and others with average level of accuracy. Therefore the algorithm must be reviewed for optimization as a future work of this study.

Category	K-NN	Naïve Bayes	*Classifier
Injured or dead people	62.32%	68.22%	95.85%
Missing, trapped, or found people	50.00%	17.82%	62.86%
Displaced people and evacuations	53.33%	29.45%	70.42%
Infrastructure and utilities damage	62.50%	43.94%	26.97%
Donation needs or offers or volunteering services	64.00%	55.05%	72.11%
Caution and advice	55.00%	34.00%	27.45%
Sympathy and emotional support	70.00%	53.45%	91.53%
Other useful information	55.38%	54.36%	82.88%
Not related or irrelevant:	0.00%	4.22%	92.59%

Table 7.1 Accuracy of different classifiers

Chapter 8

Discussion

8.1 Introduction

All previous chapters discussed the problem identified and the proposed solution as well as its implementation. This chapter discusses some limitations and future work, improvements which can be proposed for other researchers.

Social media data can be taken for numerous activities and in this research work it is given focus to text mining during natural disasters to assist decision making for relief authorities. With the massive use of social media during disaster situations incoming information filtering is very much important for situational awareness. Therefore in this work twitter data has been used to cluster disaster related information into 9 classes as

1. Injured or dead people,
2. Missing, trapped, or found people
3. Displaced people and evacuations
4. Infrastructure and utilities damage
5. Donation needs or offers or volunteering services
6. Caution and advice
7. Sympathy and emotional support
8. Other useful information and finally Irrelevant or Not Relevant class to group all other tweets not belong to any of above disaster related categories.

Data categorization is done automatically using a tool developed, named “Classifier” and in this tool a novel approach is used other than traditional classification algorithms such as Support Vector Machines (SVM), Naive Bayes (NB), and Random Forest (RF) etc. which were used in related work of literature.

The algorithm is written with a novel approach using association rules already generated using human – annotated disaster data set and searching the best category

by giving priority to the highest confidence of matching rules for key words of new data.

Since we are using a real dataset consisting with tweets posted during natural disasters such as Pakistan Flood 2014 we had to preprocess a data to remove unnecessary words, urls, tags etc and text processing techniques were used using RapidMiner Text Processing extention. And then to generate association rules, FP- Growth algorithm is used.

We have given focus to text only of tweets though they contains lot more features such as images, videos, geo-tags, user-tags, etc. In this work text posted in a tweet is considered and preprocessing is done to extract key words. Therefore not like other text mining related researches, we ended up with a very small set of words to generate patterns.

And this approach is designed to automatically classify tweets into informative classes in a known disaster in a known location. By removing geographical keywords, we can apply the same model to Sri Lanka during Flood situations because disaster related keywords will remain same since we have considered confidence of 90% for English words.

8.2 Limitations

Some of the limitations of this research are we are using English as the language and the disaster is known. Whenever there is a similar natural disaster we can run this model to automatically classify data into important categories which can be used to assist decision making efficiently. Further we are considering only text portion of tweet and we are not going to consider images, videos, geo-tags, user-tags for classification. Another key point is we are considering twitter data though there are many more social media available.

8.3 Future Work

As future extensions of this study we can proposed to expand the data collection using Facebook like social media since majority of Sri Lankans are using. But then we have to pay attention to language as well since most of data coming in Sinhala or Tamil

other than English from general public. And to provide the situational awareness rich content analysis can be considered with images, videos, geo-tags since sometimes geographical information is very much important in decision making during disaster situations. Further in this work two modules are distinct and can be merged to assist real time text categorization with a web extension.

8.4 Summary

By making use of available disaster related data and text mining techniques this study aims to categorize disaster related data into 9 meaningful categories automatically. In order to practically apply this approach combining two modules is important, which are generating association rules and classifying new data. By removing geographical keywords we can apply the same model in Sri Lanka in disaster situations.

References

[1] J. A. Obar and S. S. Wildman, “Social Media Definition and the Governance Challenge: An Introduction to the Special Issue by Jonathan A. Obar, Steven S. Wildman :: SSRN.”

[2] P. Kallas, “Top 15 Most Popular Social Networking Sites and Apps [August 2017],” *DreamGrow*, 02-Aug-2017. [Online]. Available: https://www.dreamgrow.com/top-15-most-popular-social-networking-sites/. [Accessed: 14-Aug-2017].

[3] M. Imran, S. Elbassuoni, C. Castillo, F. Diaz, and P. Meier, “Practical extraction of disaster-relevant information from social media,” 2013, pp. 1021–1024.

[4] S. Anson, H. Watson, K. Wadhwa, and K. Metz, “Analysing social media data for disaster preparedness: Understanding the opportunities and barriers faced by humanitarian actors,” *Int. J. Disaster Risk Reduct.*, vol. 21, pp. 131–139, Mar. 2017.

[5] N. Pandey and S. Natarajan, “How social media can contribute during disaster events? Case study of Chennai floods 2015,” in *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2016, pp. 1352–1356.

[6] J. D. Fraustino, B. Liu, and Y. Jin, “Social media use during disasters: a review of the knowledge base and gaps.,” 2012.

[7] Y. Kryvasheyeu *et al.*, “Rapid assessment of disaster damage using social media activity,” *Sci. Adv.*, vol. 2, no. 3, pp. e1500779–e1500779, Mar. 2016.

[8] Z. Li, C. Wang, C. T. Emrich, and D. Guo, “A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods,” *Cartogr. Geogr. Inf. Sci.*, vol. 0, no. 0, pp. 1–14, Feb. 2017.

[9] “Importance of Disaster Management.” [Online]. Available: https://targetstudy.com/articles/importance-of-disaster-management.html. [Accessed: 08-Jul-2017].

[10] “What is social media? - Definition from WhatIs.com.” [Online]. Available: http://whatis.techtarget.com/definition/social-media. [Accessed: 08-Jul-2017].

[11] A. Kaur and D. Chopra, “Comparison of text mining tools,” in *2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2016, pp. 186–192.

[12] M. Kibanov, G. Stumme, I. Amin, and J. G. Lee, “Mining Social Media to Inform Peatland Fire and Haze Disaster Management,” *Soc. Netw. Anal. Min.*, vol. 7, no. 1, Dec. 2017.

[13] M. Imran, C. Castillo, F. Diaz, and S. Vieweg, “Processing Social Media Messages in Mass Emergency: A Survey,” *ACM Comput. Surv.*, vol. 47, no. 4, pp. 1–38, Jun. 2015.

[14] S. Vieweg, C. Castillo, and M. Imran, “Integrating social media communications into the rapid assessment of sudden onset disasters,” in *International Conference on Social Informatics*, 2014, pp. 444–461.

[15] D. Velev and P. Zlateva, “Use of social media in natural disaster management,” *Intl Proc Econ. Dev. Res.*, vol. 39, pp. 41–45, 2012.

[16] R. L. Briones, B. Kuch, B. F. Liu, and Y. Jin, “Keeping up with the digital age: How the American Red Cross uses social media to build relationships,” *Public Relat. Rev.*, vol. 37, no. 1, pp. 37–43, Mar. 2011.

[17] A. Kao and S. R. Poteet, *Natural Language Processing and Text Mining*. Springer Science & Business Media, 2007.

[18] M. Rajman and R. Besançon, “Text Mining: Natural Language techniques and Text Mining applications,” *SpringerLink*, pp. 50–64, 1998.

[19] E. Younis, *Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study*. 2015.

[20] “Sensing Social Media: A Range of Approaches for Sentiment Analysis | SpringerLink.” [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-43639-5_6. [Accessed: 08-Jul-2017].

[21] F. Villarroel Ordenes, S. Ludwig, K. de Ruyter, D. Grewal, and M. Wetzels, “Unveiling What Is Written in the Stars: Analyzing Explicit, Implicit, and Discourse Patterns of Sentiment in Social Media,” *J. Consum. Res.*, vol. 43, no. 6, pp. 875–894, Apr. 2017.

[22] B. Batrinca and P. C. Treleaven, “Social media analytics: a survey of techniques, tools and platforms,” *AI Soc.*, vol. 30, no. 1, pp. 89–116, Feb. 2015.

[23] M. Imran, P. Mitra, and C. Castillo, “Twitter as a lifeline: Human-annotated twitter corpora for NLP of crisis-related messages,” *ArXiv Prepr. ArXiv160505894*, 2016.

Appendix A – Association Rules

Displaced and Evacuation

[pakistan] --> [flood]
[kashmir] --> [flood]
[india] --> [flood]
[warn] --> [flood]
[river] --> [flood]
[alert] --> [flood]
[punjab] --> [flood]
[multan] --> [flood]
[peak] --> [flood]
[peak] --> [river]
[multan] --> [alert]
[peak] --> [alert]
[peak] --> [multan]
[pakistan, india] --> [flood]
[pakistan, alert] --> [flood]
[pakistan, multan] --> [flood]
[river, alert] --> [flood]
[river, multan] --> [flood]
[peak] --> [flood, river]
[flood, peak] --> [river]
[river, peak] --> [flood]
[multan] --> [flood, alert]
[flood, multan] --> [alert]
[alert, multan] --> [flood]
[peak] --> [flood, alert]
[flood, peak] --> [alert]
[alert, peak] --> [flood]
[peak] --> [flood, multan]
[flood, peak] --> [multan]
[multan, peak] --> [flood]
[pakistan, alert] --> [multan]
[river, peak] --> [flood, multan]
[flood, river, peak] --> [multan]
[multan, peak] --> [flood, river]
[flood, multan, peak] --> [river]
[river, multan, peak] --> [flood]
[pakistan, multan] --> [alert]
[river, alert] --> [multan]
[river, multan] --> [alert]
[river, alert] --> [peak]
[peak] --> [river, alert]
[river, peak] --> [alert]
[alert, peak] --> [river]
[river, multan] --> [peak]
[peak] --> [river, multan]
[river, peak] --> [multan]
[multan, peak] --> [river]
[peak] --> [alert, multan]
[alert, peak] --> [multan]
[multan, peak] --> [alert]
[pakistan, alert] --> [flood, multan]
[flood, pakistan, alert] --> [multan]
[pakistan, multan] --> [flood, alert]
[flood, pakistan, multan] --> [alert]
[pakistan, alert, multan] --> [flood]
[river, alert] --> [flood, multan]
[flood, river, alert] --> [multan]
[river, multan] --> [flood, alert]
[flood, river, multan] --> [alert]
[river, alert, multan] --> [flood]
[river, alert] --> [flood, peak]
[flood, river, alert] --> [peak]
[peak] --> [flood, river, alert]
[flood, peak] --> [river, alert]
[river, peak] --> [flood, alert]
[flood, river, peak] --> [alert]
[alert, peak] --> [flood, river]
[river, peak] --> [alert, multan]
[alert, peak] --> [river, multan]
[river, alert, peak] --> [multan]
[multan, peak] --> [river, alert]
[river, multan, peak] --> [alert]

[peak] --> [flood, alert, multan]
[flood, peak] --> [alert, multan]
[alert, peak] --> [flood, multan]
[flood, alert, peak] --> [multan]
[multan, peak] --> [flood, alert]
[flood, multan, peak] --> [alert]
[alert, multan, peak] --> [flood]
[river, alert] --> [multan, peak]
[river, multan] --> [alert, peak]
[river, alert, multan] --> [peak]
[peak] --> [river, alert, multan]
[alert, multan, peak] --> [river]
[river, alert] --> [flood, multan, peak]
[flood, river, alert] --> [multan, peak]
[river, multan] --> [flood, alert, peak]
[flood, river, multan] --> [alert, peak]
[river, alert, multan] --> [flood, peak]
[flood, river, alert, multan] --> [peak]
[peak] --> [flood, river, alert, multan]
[flood, peak] --> [river, alert, multan]
[river, peak] --> [flood, alert, multan]
[flood, river, peak] --> [alert, multan]
[alert, peak] --> [flood, river, multan]

Caution and Advice

[armi] --> [flood]
[hit] --> [flood]
[peopl] --> [flood]
[kashmir] --> [flood]
[evacu] --> [flood]
[india] --> [flood]
[kashmir, rescu] --> [flood]
[rescu] --> [flood]
[jammu] --> [kashmir]
[jammu] --> [flood, kashmir]
[flood, jammu] --> [kashmir]
[pakistan] --> [flood]
[thousand] --> [flood]
[jammu] --> [flood]
[srinagar] --> [flood]
[amp] --> [flood]
[kashmir, jammu] --> [flood]
[kashmir, peopl] --> [flood]
[pakistan, thousand] --> [flood]
[pakistan, india] --> [flood]
[rescu, hit] --> [flood]

Donation and Volunteer

[flood] --> [kashmir]
[relief] --> [flood, kashmir]
[relief] --> [kashmir]
[flood, relief] --> [kashmir]
[donat] --> [kashmir]
[donat] --> [flood, kashmir]
[flood, donat] --> [kashmir]
[victim] --> [flood]
[help] --> [flood]
[donat] --> [flood]
[jammu] --> [kashmir]
[jammu] --> [flood, kashmir]
[flood, jammu] --> [kashmir]
[relief] --> [flood]
[pakistan] --> [flood]
[kashmir] --> [flood]
[kashmir, relief] --> [flood]
[amp] --> [flood]
[jammu] --> [flood]
[punjab] --> [flood]
[hit] --> [flood]
[kashmir, jammu] --> [flood]
[kashmir, donat] --> [flood]

Infrastructure Damage

[flood] --> [kashmir]
[jammu] --> [kashmir]
[jammu] --> [flood, kashmir]
[flood, jammu] --> [kashmir]
[kashmirflood] --> [flood]
[srinagar] --> [kashmir]
[srinagar] --> [flood, kashmir]
[flood, srinagar] --> [kashmir]
[kashmir] --> [flood]
[pakistan] --> [flood]
[jammu] --> [flood]
[srinagar] --> [flood]
[vallei] --> [flood]
[hit] --> [flood]
[amp] --> [flood]
[damag] --> [flood]
[vallei] --> [kashmir]
[kashmir, jammu] --> [flood]
[kashmir, srinagar] --> [flood]
[vallei] --> [flood, kashmir]
[flood, vallei] --> [kashmir]
[kashmir, vallei] --> [flood]

Injured or Dead

[india] --> [flood, pakistan]
[india] --> [pakistan]
[flood, india] --> [pakistan]
[india, death] --> [pakistan, toll]
[india, death] --> [flood, pakistan, toll]
[flood, india, death] --> [pakistan, toll]
[death] --> [toll]
[death] --> [flood, toll]
[flood, death] --> [toll]
[pakistan, death] --> [india, toll]
[pakistan, death] --> [flood, india, toll]
[flood, pakistan, death] --> [india, toll]
[dead] --> [flood]
[pakistan, india, death] --> [toll]
[pakistan, india, death] --> [flood, toll]
[flood, pakistan, india, death] --> [toll]
[pakistan, death] --> [toll]
[pakistan, death] --> [flood, toll]
[flood, pakistan, death] --> [toll]
[india, death] --> [toll]
[india, death] --> [flood, toll]
[flood, india, death] --> [toll]
[india, toll, death] --> [pakistan]
[india, toll, death] --> [flood, pakistan]
[flood, india, toll, death] --> [pakistan]
[india, death] --> [pakistan]
[india, death] --> [flood, pakistan]
[flood, india, death] --> [pakistan]
[india, toll] --> [pakistan]
[india, toll] --> [flood, pakistan]
[flood, india, toll] --> [pakistan]
[india] --> [flood]
[kill] --> [flood]
[pakistan, toll, death] --> [india]
[pakistan, toll, death] --> [flood, india]
[flood, pakistan, toll, death] --> [india]
[pakistan, death] --> [india]
[pakistan, death] --> [flood, india]
[flood, pakistan, death] --> [india]
[pakistan, toll] --> [india]
[pakistan, toll] --> [flood, india]
[flood, pakistan, toll] --> [india]
[pakistan] --> [flood]
[kashmir] --> [flood]
[pakistan, india] --> [flood]
[toll] --> [flood]
[death] --> [flood]
[reach] --> [flood]
[reach] --> [death]
[pakistan, toll] --> [flood]
[pakistan, death] --> [flood]
[pakistan, kill] --> [flood]
[india, toll] --> [flood]
[india, death] --> [flood]
[india, kill] --> [flood]
[toll, death] --> [flood]
[reach] --> [flood, death]
[flood, reach] --> [death]
[death, reach] --> [flood]
[pakistan, india, toll] --> [flood]
[pakistan, india, death] --> [flood]
[pakistan, toll, death] --> [flood]
[india, toll, death] --> [flood]
[pakistan, india, toll, death] --> [flood]

Missing Trapped or Found	Other Useful Information	Sympathy and Emotional Support
[flood, rescu] --> [kashmir]	[flood] --> [kashmir]	[prai] --> [kashmir]
[flood] --> [kashmir]	[jammu] --> [flood, kashmir]	[prai] --> [flood, kashmir]
[strand] --> [kashmir]	[flood, jammu] --> [kashmir]	[flood, prai] --> [kashmir]
[strand] --> [flood, kashmir]	[jammu] --> [kashmir]	[flood] --> [kashmir]
[flood, strand] --> [kashmir]	[india] --> [flood]	[peopl] --> [kashmir]
[trap] --> [kashmir]	[amp] --> [flood]	[peopl] --> [flood, kashmir]
[trap] --> [flood, kashmir]	[pakistan] --> [flood]	[victim] --> [kashmir]
[peopl] --> [kashmir]	[kashmir, jammu] --> [flood]	[victim] --> [flood, kashmir]
[peopl] --> [flood, kashmir]	[jammu] --> [flood]	[flood, victim] --> [kashmir]
[flood, peopl] --> [kashmir]	[kashmir] --> [flood]	[flood, peopl] --> [kashmir]
[flood, trap] --> [kashmir]		[peopl] --> [flood]
[thousand] --> [kashmir]		[amp] --> [flood]
[thousand] --> [flood, kashmir]		[pakistan] --> [flood]
[flood, thousand] --> [kashmir]		[prai] --> [flood]
[jammu] --> [kashmir]		[kashmir] --> [flood]
[jammu] --> [flood, kashmir]		[victim] --> [flood]
[flood, jammu] --> [kashmir]		[prayer] --> [flood]
[kashmirflood] --> [flood]		[jammu] --> [flood]
[srinagar] --> [flood]		[affect] --> [flood]
[rescu] --> [flood]		[jammu] --> [kashmir]
[trap] --> [flood]		[kashmir, prai] --> [flood]
[strand] --> [flood]		[kashmir, victim] --> [flood]
[kashmir] --> [flood]		[kashmir, peopl] --> [flood]
[peopl] --> [flood]		[jammu] --> [flood, kashmir]
[thousand] --> [flood]		[flood, jammu] --> [kashmir]
[jammu] --> [flood]		[kashmir, jammu] --> [flood]
[pakistan] --> [flood]		
[india] --> [flood]		
[amp] --> [flood]		
[kashmir, rescu] --> [flood]		
[kashmir, strand] --> [flood]		
[kashmir, trap] --> [flood]		
[kashmir, peopl] --> [flood]		
[kashmir, thousand] --> [flood]		
[kashmir, jammu] --> [flood]		