CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

LR-TED: Employing Linguistic Rules on Tweets' Content to classify Eyewitness Messages for Disaster Events

by

Sajjad Haider

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

Faculty of Computing
Department of Computer Science

2022

LR-TED: Employing Linguistic Rules on Tweets' Content to classify Eyewitness Messages for Disaster Events

By

Sajjad Haider

(DCS153005)

Dr. Raheel Nawaz, Professor
Manchester Metropolitan University, UK
(Foreign Evaluator 1)

Dr. Ding Xiaofeng, Professor
Huazhong University of Science and Technology, China
(Foreign Evaluator 2)

Dr. Azhar Mahmood
(Thesis Supervisor)

Dr. Abdul Basit Siddiqui
(Head, Department of Computer Science)

Dr. Muhammad Abdul Qadir
(Dean, Faculty of Computing)

DEPARTMENT OF COMPUTER SCIENCE
CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY
ISLAMABAD
2022

To my parents and family.

CAPITAL UNIVERSITY OF SCIENCE & TECHNOLOGY
ISLAMABAD
Expressway, Kahuta Road, Zone-V, Islamabad
Phone:+92-51-111-555-666 Fax: +92-51-4486705
Email: info@cust.edu.pk Website: https://www.cust.edu.pk

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in the thesis, entitled **"LR-TED: Employing Linguistic Rules on Tweets' Content to Classify Eyewitness Messages for Disaster Events"** was conducted under the supervision of **Dr. Azhar Mahmood**. No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the **Department of Computer Science, Capital University of Science and Technology** in partial fulfillment of the requirements for the degree of Doctor in Philosophy in the field of **Computer Science.** The open defence of the thesis was conducted on **November 02, 2022**.

Student Name : Sajjad Haider (DCS153005) ____________

The Examination Committee unanimously agrees to award PhD degree in the mentioned field.

Examination Committee :

(a) External Examiner 1: Dr. Hammad Afzal
Associate Professor
NUST, Islamabad ____________

(b) External Examiner 2: Dr. Kifayat Ullah Khan
Associate Professor
FAST-NUCES, Islamabad ____________

(c) Internal Examiner : Dr. Nayyer Masood
Professor
CUST, Islamabad ____________

Supervisor Name : Dr. Azhar Mahmood
Associate Professor
CUST, Islamabad ____________

Name of HoD : Dr. Abdul Basit Siddiqui
Associate Professor
CUST, Islamabad ____________

Name of Dean : Dr. Muhammad Abdul Qadir
Professor
CUST, Islamabad ____________

AUTHOR'S DECLARATION

I, **Sajjad Haider (Registration No. DCS-153005)**, hereby state that my PhD thesis titled, **'LR-TED: Employing Linguistic Rules on Tweets' Content to Classify Eyewitness Messages for Disaster Events'** is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/ world.

At any time, if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my PhD Degree.

(Sajjad Haider)

Dated: 02 November, 2022 Registration No: DCS153005

PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in the thesis titled "**LR-TED: Employing Linguistic Rules on Tweets' Content to Classify Eyewitness Messages for Disaster Events**" is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/ cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of PhD Degree, the University reserves the right to withdraw/ revoke my PhD degree and that HEC and the University have the right to publish my name on the HEC/ University Website on which names of students are placed who submitted plagiarized thesis.

(Sajjad Haider)

Dated: 02/ November, 2022

Registration No: DCS153005

List of Publications

It is certified that following publication(s) have been made out of the research work that has been carried out for this thesis:-

1. **S. Haider** and M. T. Afzal, "Autonomous Eyewitness Identification by Employing Linguistic Rules for Disaster Events," *CMC-Computers Materials & Continua*, 66(1), 481498, 2021.

2. **S. Haider**, M. Azhar, S. Khatoon, A. Majed, and M. T. Afzal, "Automatic Classification of Eyewitness Messages for Disaster Events Using Linguistic Rules and ML/AI Approaches." *MDPI Journal: Applied Sciences*, 12 (19), 9953, 2022.

3. **S. Haider**, M. T. Afzal, M. Asif, H. Maurer, A. Ahmad, A. Abuarqoub, "Impact analysis of adverbs for sentiment classification on Twitter product reviews." *Concurrency and Computation: Practice and Experience*, First published: 29 August 2018, Online version: https://doi.org/10.1002/cpe.4956, 2018.

Sajjad Haider

(DCS153005)

Acknowledgement

First of all praise be to Allah, The most gracious, Who blessed me with the opportunity, capability and resources to pursue the doctoral program, and it's all due to His grace that I saw it through.

I dedicate my dissertation work to my PhD supervisor Dr. Azhar Mahmood for the patient guidance, encouragement and advice he has provided throughout my time as his student. I have been extremely lucky to have a supervisor who cared so much about my work, and who responded to my questions and queries so promptly. Whenever I felt lost and demotivated, he pulled me out of the disbelief and guide me the to the right way. There just aren't enough words to show my gratitude to him. Thank you Sir, for all your help, support and guidance.

I am grateful to Dr. Muhammad Abdul Qadir, Head of the Center for Distributed and Semantic Computing (CDSC), who motivated the entire group to work hard. I am also thankful to Dr. Muhammad Tanvir Afzal and other members of CDSC group, whose discussion and constructive criticism maintained an environment that was conducive for research.

I am indebted to all the internal, external and foreign examiners for taking time to read and evaluate my synopsis and thesis, for identifying the shortcomings, and for providing their learned opinion on how to improve upon these.

There are numerous other people at C.U.S.T. who helped me in pursuit of my PhD in one way or the other including the faculty members at Department of Computer Science, managerial and support staff, and the librarian to mention a few. Thank you all.

I take this opportunity to sincerely acknowledge my senior colleges Mr. M. M. Khan, Mr. Mazahir Alvi and Mr. Mathar Abdullah, for providing moral support which helped me complete my work comfortably in parallel to my office duties.

In the end but most importantly, I would like to thank my parents, my wife and kids, although I understand that any amount of gratitude shown to them is woefully inadequate. My family's unconditional support is largely the reason that this PhD is completed. I owe a lot to my parents, who encouraged and helped me at every stage of my personal and academic life, and longed to see this achievement come true. I am very much thankful to my family, my wife, son and daughters, who supported me in every possible way to see the completion of this work. May Allah reward you for your sacrifice and selflessness.

Abstract

Social networking platforms provide a vital source for disseminating information across the globe, particularly in case of disaster. These platforms are a great means to find out the real account of the disaster. Twitter is an example of such a platform, which has been extensively utilized by the scientific community due to its unidirectional model. It is considered a challenging task to identify eyewitness tweets about the incident from the millions of tweets shared by twitter users. The research community has proposed diverse sets of techniques to identify eyewitness tweets from content. A recent state-of-the-art approach (Zahra et al.) has proposed a comprehensive set of features to identify eyewitness tweets. However, this approach suffers some limitations. Firstly, automatically extracting the feature-words remains a perplexing task against each feature identified by the approach. The approach lacks the strength of a generic approach as it needs a separate dictionary for different event types. Secondly, all identified features were not incorporated in the implementation, due to implementation complexity. This work has utilized the english language structure & patterns, linguistics features, and words relationship in a sentence, to achieve automatic extraction of feature-words by creating grammar rules and proposed a generalized approach to cover different types of disasters such as earthquakes, floods, hurricanes, and wildfires. Additionally, all identified features were implemented which were left out by state-of-the-art technique. The proposed approach (LR-TED: Linguistic Rule-based approach for Twitter Eyewitness Detection) is evaluated for all disaster types, including earthquakes, floods, hurricanes, and wildfire events. Based on the use of a static dictionary, the manual features-based approach was able to produce the maximum F-Score value of 0.92 for Eyewitness identification for the earthquake category, whereas the LR-TED approach, using entirely automatic grammar rules, secured a maximum F-Score value of 0.93 in the same category. The implication of the LR-TED can be realized when processing millions of tweets in real-time using automated processes rather than involving domain experts in building the static dictionary. Using linguistic features, language patterns, and

words relationship in a sentence for feature extraction and eyewitness identification is the novel contribution of the LR-TED approach. LR-TED is adaptable for diverse events and unseen content, whereas the manual features-based approach requires human involvement in creating dictionaries of related words, for all the identified features, for the new disaster type. The estimation of efforts required by manual features-based and proposed LR-TED approaches are discussed for new disaster types and the LR-TED outperforms the manual features-based approach in terms of required time, cost, and human resources. LR-TED can be evaluated on different social media platforms for the identification of eyewitness reports for various disaster types in the future.

Contents

List of Figures

List of Tables

Abbreviations

AUC	Area Under the Curve
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
HIT	Human Intelligence Tasks
LR-TED	Linguistic Rule-based approach for Twitter Eyewitness Detection
ME	Maximum Entropy
NB	Naive Bayes
NER	Named Entity Recognition
NLP	Natural Language Processing
POS	Part-Of-Speech
RDMS	Relational Database Management System
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristics
SDSU	San Diego State University
SVM	Support Vector Machines
TED	Twitter Earthquake Detector
USGS	U.S. Geological Survey

Symbols

Ew	Eyewitness
$F1$	F1-Score
NEw	Non-Eyewitness
Un	Unknown

Chapter 1

Introduction

This chapter covers the background and basic approaches for the identification of eyewitness, followed by the research motivation. The comprehensive study of literature led us to form the problem statement and research objectives that are discussed after the research motivation. The scope of the research is explained after the research objectives. Finally, the chapter concludes with the adopted methodology to conduct this research with applications of the proposed idea at the end of this chapter. The thesis outline is also presented at the end to escort through the formation of upcoming chapters.

1.1 Background

In this digital era, the social media platforms like Instagram, Twitter, and Facebook are extensively used for daily activities. People around the world are connected with these platforms to share reviews, options, ideas, and information regarding numerous things and topics [1] [2]. Among these platforms, Twitter is the most commonly used social media platform due to its distinctive unidirectional model of the relationship between users. Over 130 million active users post or re-tweet 500+ million tweets on daily basis. This volume is believed to be increased by 30% a year[1].

The implicit and explicit information from Twitter is explored by the researchers of diversified domains like; location extraction, event detection, recommendation systems, disaster management systems, sentiment analysis, in the education sector, influential user identification, and financial market predictions. Researchers of these domains have proposed diversified approaches to help the community in selecting the appropriate information required.

Over 130 million active users on Twitter are consuming and disseminating messages resulting in the collection of massive information. With this massive amount of information available, the user can get an overwhelming response. The true importance of recommendation systems can be realized when suitable information is extracted from the huge data shared by the users [3]. Traditionally the recommendation systems are of two types such as personalized and non-personalized. In non-personalized recommendation systems, the user's preferences are not utilized, For example, "top 5 movies of the month". But the personalized recommendation system utilizes the user preferences, its profile, and item characteristics. for the recommendation. Content-based recommendation and Collaborative filtering approaches are used for personalized recommendation systems [4]. Research of this domain has also exploited the twitter's concept of "follow", and based on the idea the topology-based approaches for followee recommendation system are proposed [5][6][7][8][9].

The connection between education and technology is stimulated as a fundamental component in the transformation of the education sector over the past few years. The researchers have exploited Twitter for its applications in the educational sector since its launch in 2006 and there are several discussions on how to actively use it for educational purposes [10][11]. A recent study by Tian et al. in 2019, investigated the participation of students in Twitter supported activities [12]. A

[1] `https://www.internetlivestats.com/twitter-statistics/`

comprehensive survey was conducted by Jeffrey et al. and find out how Twitter can be utilized educationally: communication, classroom activities, and professional development (PD) [10]. The usage of twitter as an educational application is commonly discussed for higher education rather than the first-grades [13], but few recent studies have exploited its contributions for undergraduate and school studies [14][15][16]. The potential uses of hash-tags used by professors for different courses, the students of universities are engaged for access to information related to course material [17].

Sharing one's thoughts, opinions, and reviews are the key features provided by social media platforms, and the availability of huge data from these platforms can affect a user's decision making. The researchers of this domain have proposed the technique of sentiment analysis to answer this issue. Sentiment analysis is a process of mining the opinions, emotions, views, and attitudes from the content using Natural Language Processing NLP) and classifying the text into "positive", "negative" and "neutral" categories [18]. A recent study in 2020, has systemically reviewed the literature of sentiment analysis techniques on Twitter [19]. Various proposed techniques of recommender systems adopted the sentiment analysis approach for the categorization of opinions. A recent study has proposed a fine-grained recommender system using a sentiment analysis based approach and evaluated it on twitter's data [20]. Support vector machines (SVM), Maximum Entropy (ME), Naive Bayes (NB) are machine learning approaches that achieved great success in sentiment analysis [21][22][23]. The study discussed the importance of sentiment analysis and demonstrated the high results accuracy for machine learning methods like Nave Bayes (NB) and Support vector machines (SVM). The role of human sentiment, emotions, and mood in making the financial decision is studied and comprehensively expressed in the survey by Huina et al. [24].

The capacity of Twitter in the field of event detection is also studied by researchers of the domain, and various techniques are proposed to detect events from tweets [25]. These event detection techniques are categorized into feature-pivot techniques and document-pivot techniques. The feature-pivot techniques are discussed and applied by researchers and analyze the feature distribution for the detection of events [26] [27]. The document-pivot techniques cluster the document based on textual similarity like TF-IDF from the large noisy data [28][29][30][31]. A recent study in 2019, has proposed an online algorithm that group together the tweets into clusters by effectively discovering the interest based on their temporal and textual features [32]. A major challenge with these event detection systems is high computational cost. Hassan et al. in 2019 have proposed an incremental clustering-based technique to provide a solution with low computational cost and named it TwitterNews+ [33].

Twitter is exploited by researchers for location detection to determine the position using Natural Language Processing (NLP) and to respond accordingly either it's an emergency or a help call [34]. Bothorel et al. presented a comprehensive study of location detection and recommendation from social media platforms [35]. Location identification is a challenging task when the user names and home locations of the users are not reliable. A recent study in 2019 by Kumar et al. [36] presented a Convolutional Neural Network (CNN) based model for location extraction.

Analyzing the data is essential for event detection especially in emergency conditions due to natural disasters such as flooding, hurricane, fire, or earthquake [1][32]. The studies have proven the importance of credible information required for responding to such emergency conditions, and the eyewitness of the event can be a useful resource for information relating to the event [37].

1.2 Eyewitness

The worldwide acceptance of Twitter attracted the research community to exploit and discover implicit & explicit information from it. Researchers of the domain have proposed various systems based on twitters information. A recent study on Twitter has discussed its role for targeted news recommendations, advertising, response systems, disaster, and emergency alerts systems [1]. Twitter is considered a potential source of breaking news and it has proved itself as a news breaker.

Studies have proved that 85% of the trending topics on Twitter are news headlines [38]. The question ascends, how Twitter does that? The answer to the question is that it uses the tweets posted by the eyewitness user. Following are some events that prove Twitter's capacity of breaking the news are outlined below:

1. Emergency landing of Delta Aircraft flight on the remote island of Alaska, due to its potential engine failure[2]. The incident was tweeted by a passenger of the flight and that tweet was found by the news agency.
2. California Earthquake[3]. Zone of tweets about the earthquake were available on twitter a minute before USGS[4] recorded time.
3. Westgate Shopping Mall attack in Nairobi, Kenya [5]. The eyewitnesses tweeted about the event about thirty-three minutes before the break of the news by any news channel.
4. The bombing incident of Boston[6]. Eyewitness tweets about the incident were available well before the coverage by any news channel.
5. The New York airplane crash Hudson Bay[7]. An eyewitness from Hudson Bay tweeted the event and the news became the headline of the Daily Telegraph.

1.3 Research Motivation

In today's fast age, the information related to a new event gets broadcasted in no time. The immense freedom-of-speech and liberty of expression allow everyone to bounce opinion on any event. With this liberty, any information can be spread. This information is available on electronic media and other social media platforms

[2] http://channelnewsasia.com/news/world/delta-flight-middle-of-the-ocean-seattle-beijing-emergency-land-11062706
[3] https://latimesblogs.latimes.com/technology/2008/07/twitter-earthqu.html
[4] https://www.usgs.gov/
[5] https://en.wikipedia.org/wiki/Westgate_shopping_mall_attack
[6] https://en.wikipedia.org/wiki/Boston_Marathon_bombing
[7] https://www.telegraph.co.uk/technology/twitter/4269765/New-York-plane-crash-Twitter-breaks-the-news-again.html

like twitter. On Twitter, this information is shared, re-tweeted, and replied multiple times. An eyewitness can't be identified by just finding the disaster-related keywords. The keywords may have come as a reference in the past, replied, or re-tweeted, or it can be a hash-tag.The comprehensive study of the literature shows that the information shared by an eyewitness is potentially useful in understanding the true state and the severity of the event and suggests that the identification of eyewitnesses' is a vital task.

The one who sees an occurrence of an event or an object is called the eyewitness, or "especially: one who reports on what he or she has seen". The eyewitness, in Cambridge Dictionary, is defined as "a person who saw something happen, for example, a crime or an accident".

Furthermore, the emergency services organizations and agencies responsible to respond in the event of any disastrous situation rely on the information shared by the eyewitness sources. These emergency responding services have to take swift and effective measures to control the situation, therefore the availability of credible information about the event is vital.

Twitter remained a potential platform where millions of tweets are shared by the community. Among them there could be tweets highlighting disaster events. From such tweets, it remained an open challenge to find Eyewitness tweets. Many different researchers have proposed various approaches and have employed different features to identify the eyewitness tweets. For example, accounts of users and connected network was exploited by Truelove et al. [39]. During pre-processing, they have utilized some obvious keywords to marks tweet messages as candidate eyewitness messages. However, they have not exploited the content of the tweet messages to mark a tweet messages as eyewitness messages. Another research by Doggett et al. [40] have categorized eyewitness messages into two broad categories such as: eyewitness and non-eyewitness using language constructs. A hybrid approach by adopting the linguistic features and meta-features (e.g. type of application used), was presented by Fang et al. [41], for the identification of eyewitness reports. Apart from language constructs, they have also used stylistics characteristics as

TABLE 1.1: Characteristics for Eyewitness Identification (manual features-based approach [37]).

Sr.#	Eyewitness Feature	Examples
1	Reporting small details of surroundings	"window shaking", "water in basement"
2	Words indicating perceptual senses	"seeing", "hearing", "feeling"
3	Reporting impact of disaster	"raining", "school canceled", "flight delayed"
4	Words indicating intensity of disaster	"intense", "strong", "dangerous", "big"
5	First person pronouns and adjectives	"i", "we", "me"
6	Personalized location markers	"my office", "our area"
7	Exclamation and question marks	"!", "?"
8	Expletives	"wtf", "omg", "s**t"
9	Mention of a routine activity	"sleeping", "watching a movie"
10	Time indicating words	"now", "at the moment", "just"
11	Short tweet length	"one or two words"
12	Caution and advice for others	"watch out", "be careful"
13	Mention of disaster locations	"area and street name", "directions"

well to find eyewitness message. Subsequently, Tanev et al. [42] also presented a hybrid approach which uses the combination of lexical, stylistic, and semantics features with meta data attached with each message, to categorize the tweet as eyewitness or non-eyewitness.

From the comprehensive study of the literature, it is evident that the researchers of the domain has adopted diversified techniques and feature sets for the identification of the eyewitness tweets from a large pool of text. Recently, a study by Zahra et al. [37] has identified a comprehensive feature list for the identification of the eyewitness tweets. The manual features-based approach has applied these features to a dataset of 6000 tweets for the training of the approach and used the dataset of 8000 tweets for testing. The set of thirteen features was carefully identified by the domain-experts for eyewitness identification as described in Table-1.1.

In this work the terms "State-of-the-art" and "the manual features-based approach" are interchangeably used to refer the state-of-the-art [37].

Figure-1.1 shows a tweet posted by an eyewitness user. The tweet words are tagged with relevant features identified by the domain-experts [37]. The identified

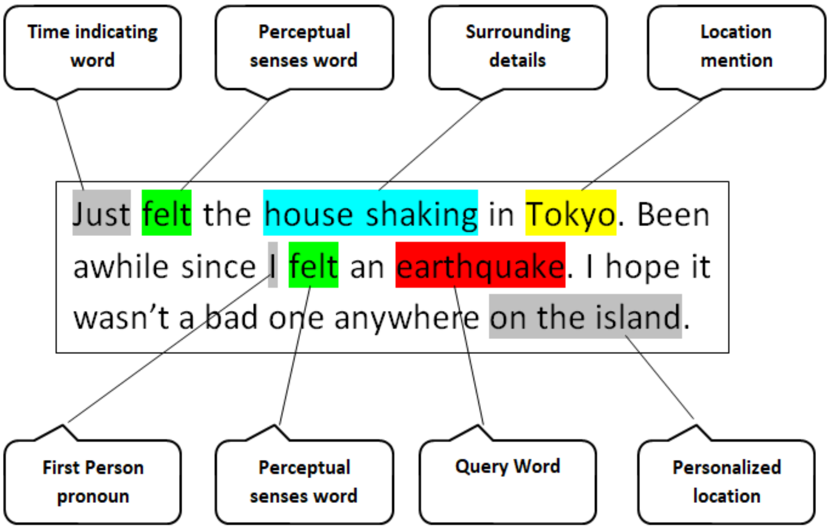

FIGURE 1.1: Demonstration of Features Identification, Example-1 (Earthquake).

feature-words are marked using different colors and they are tagged with the corresponding labels in Figure-1.1. The starting word, "Just" from the tweet specifies the moment in time as "Time indicating words" identified in Table-1.1. Subsequently, sample tweet contains "house shaking", "felt", "I", "Tokyo", and "On Island" which indicates the characteristics of "Surrounding details", "Perceptual senses words", "First-person noun", "Location mention" and "Personalized location" correspondingly. By identifying these features from the content, the tweet can be marked as an eyewitness.

Figure-1.2 indicates the appropriate feature tags for each word that are identified manually from the tweet content. The feature-words identified are "Big", "aftershocks", and "now", and tagged to their corresponding labels of "Intensity of disaster", "Surrounding details", and "Time indicating words" respectively, from the Table-1.1.

First, two examples are shown in Figures-1.1 and 1.2, are taken from earthquake events, a flood event related tweet by an eyewitness is shown in Figure-1.3 and the feature-words are identified with different colors and tagged with appropriate identified feature by the manual features-based approach. The content of the tweet is "I almost died driving home from work because it started to downpour and flood

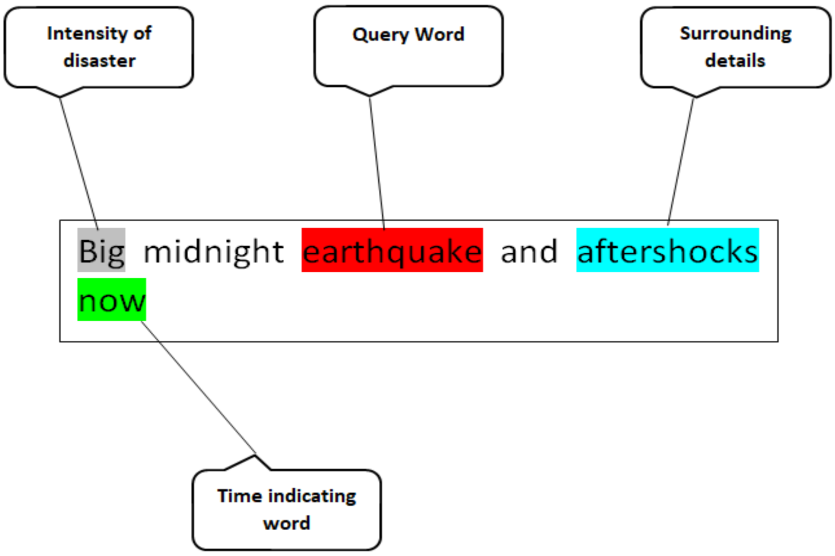

FIGURE 1.2: Demonstration of Features Identification, Example-2 (Earthquake).

on the freeway and lightning and its 99 f**king degrees out". From the content we have "I" as "First person pronouns and adjectives", then is "driving home" as "Mention of a routine activity" and after that, we have "started downpour", "flood", "freeway", "lightning" and "f**king" as "Surrounding Details", "Query Word", "Disaster Location", "Impact of Disaster", and "Expletives" respectively. The features found are noticeable in diverse colors with suitable labels in the Figure-1.3.

The manual features-based [37] adopted the manually created dictionaries for the identification of eyewitnesses. Three identified domain expert features, "Mention of disaster locations", "Reporting small details of surroundings", and "Personalized location markers" were not implemented in the manual features-based approach due to their implementation complexity. The identified features such as; "First-person pronouns and adjectives", "Words indicating perceptual senses", "Expletives", and "Exclamation and question marks" has the limitation of a predefined list of words. So, their implementation was not possible without utilizing the relevant predefined lists. For the remaining identified features from Table-1.1, the manual features-based approach [37] adopted the same technique of using a manually created dictionary, to identify the feature words from the content.

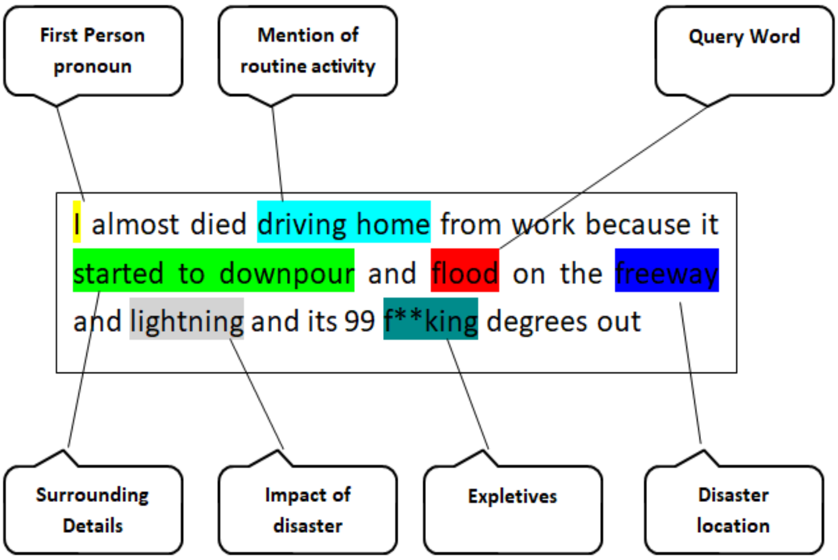

FIGURE 1.3: Demonstration of Features Identification, Example-3 (Flood).

By investigating the manual features-based approach, two main issues have been identified. Firstly, creating a dictionary of words is not adaptable for diverse events and unseen tweet content. The idea of using the static dictionary is not generic to cover different disaster types like earthquakes, floods, hurricanes, and wildfires. Secondly, the implementation of the features that were dropped in the manual features-based approach. Each identified feature of manual features-based approach has its importance, and based on its importance these features are selected after detailed study, so dropping any such feature effects the results for eyewitness identification. The process of extracting feature-words by utilizing the static dictionary negates the concept of a fully automatic approach. This leaves an open research gap to propose an automated technique of extracting the feature-words from the content, corresponding to the features identified in Table-1.1.

1.4 Problem Statement

The investigation and observations discussed in the research motivation section we formulate the problem statement as; the State-of-the-art approach is not generic and requires human interaction for creating dictionaries if the approach is adopted for new events or disaster types, some of the features had been dropped where each feature has its significance for the identification of eyewitness tweets, also the effectiveness of the approach is evaluated for one machine learning model whereas the other machine-learning and deep learning models could be more effective for eyewitness identification.

1.5 Research Questions

The problem statements identified after detailed study, investigation and the observations are the research objectives that are addressed in this research. In addition, we have to the following questions:

1. Can we develop a generic approach which might be capable of extracting feature words automatically instead of using the static dictionaries that requires manual intervention in case of any new disaster type?
   See Section-3.6, Section-5.5.
2. Can any additional or dropped feature from the state-of-the-art approach have a further impact to improve eyewitness identification?
   See Tables-5.6, 5.7 & 5.8, in Section-5.3.1
3. How far the advanced deep learning models are suitable to classify more accurately for eyewitness identification based on identified features?
   See Section-5.4.3

1.6 Research Objectives

The first objective of this research work is to automatically extract the feature-words using language structural analysis and by creating the Grammar rules and to propose a generic approach that cover different disaster types by discouraging the use of a static dictionary of different disaster type.

The second objective is to construct the language rules for those identified features that were dropped in the manual features-based approach, due to their implementation complexity with the proposed model of Zahra et al. [37].

Our third objective in this research is to evaluate the automatically extracted features using different machine learning models from literature in this domain and the advance deep-learning models used for textual data and study which model perform best for classification of eyewitness tweets.

1.7 Scope of the research

This research intends to contribute to (i) creating of Grammar rules, using language structural analysis, to automatically extract the feature-words without any human interaction, (ii) constructing grammar rules for those identified features that are dropped in the manual features-based approach, and (iii) proposal of a generic framework to cope the identified limitation of the manual features-based approach for different disaster types. The scope of this research is limited to four disaster events; earthquake, flood, hurricane, and wildfire.

1.8 Research Methodology

In this research, we have exploited the linguistic features, language structure, and existing relationship between the words of a sentence which explains the context of the sentence. We identified that this relationship among the words in a sentence can be exploited to automatically extract the feature-words, and to achieve this we have to define grammar rules. This research proposed a set of grammar rules to automatically extract the feature-words from the tweet content. Grammar rules are created for the identified features, where either the manual features-based approach [37] adopted the static dictionary or dropped the feature due to its implementation complexity.

The activities carried out for the research work has been divided into phases, levels, and tasks as bellow;

- **Phase-1: Decide "What to research?"**;
  - **Level-1: Research Idea Formulation,** includes following tasks;
    1. In-depth Study of the Literature
    2. Identification of Research Gap
    3. Formulation of Research Problem
- **Phase-2: Research Planning**
  - **Level-2: Proposed Approach and Architecture,** based on findings from first phase an innovative approach, using Grammar Base-Rules for eyewitness identification, is proposed using the following tasks;
    1. Architecture of the proposed idea.
    2. Parsing of Tweets
    3. Language Structure Analysis
    4. Feature Extraction Rules
    5. Feature Evaluation
  - **Level-3: Dataset,** at level-3 the task is "Collection/Selection of Dataset" for evaluation of proposed approach.
  - **Level-4: Benchmark Selection,** Task at Level-4 is to select a Benchmark dataset to compare the results of the proposed idea with the manual features-based approach.
- **Phase-3: Implementation of Proposed-Approach** Third phase is to implement the Research Idea proposed in Phasae-2.
  - **Level-5: Pre-processing of the Dataset,** the dataset selected for experiments using proposed approach requites pre-processing activities.
  - **Level-6: Evaluation and Results,** the proposed approach is evaluated over data from Level-5 and following tasks are performed;

1. Implementation and Evaluation
2. Comparison with a state-of-the-art approach
3. Results; Discussions and Interpretation
4. Effort Estimation

The "Design Science" is a research methodology for Information Technology domain, which has the guidelines for the iterative processes for development and evaluation of the research projects. A customized "Design Science Methodology" of the proposed solution is shown in Figure-1.4.

1.9 Applications of Proposed Approach

The implication of the proposed research work can be exploited in various applications. Some of the potential applications are highlighted below:

1. Disaster Management System
2. Disaster or Emergency Alert System
3. Emergency Response for Institutes and Agencies

1.10 Thesis Outline

The thesis document contains five chapters. Chapters 1 provided the introduction to the topic, its background, and its motivations. Chapter 2 discusses in detail, the current state-of-the-art methods and strategies for the detection of eyewitnesses. In Chapter 3, the architecture of the LR-TED is explained. Chapter 4, discusses the proposed Grammar-Based Rules for the extraction of eyewitness features. In Chapter 5, the results are evaluated and compared with the benchmark dataset of the manual features-based approach. The last chapter is the conclusion that summarizes the proposed approach.

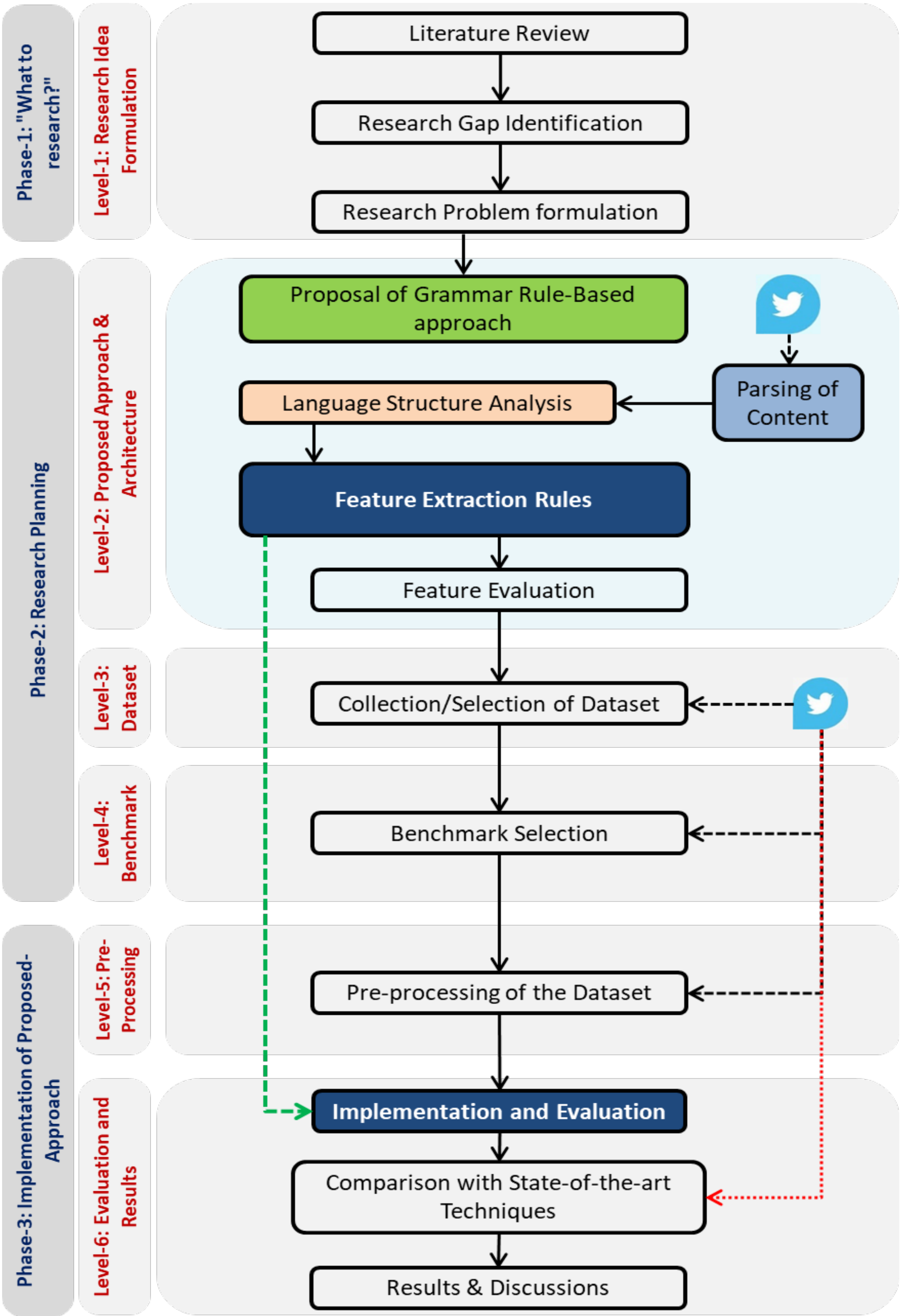

FIGURE 1.4: Customized Design Science Methodology

Chapter 2

Literature Review

In this chapter, the scope and importance of the eyewitness and different approaches for its identification, proposed in the literature are debated. The core objective of the literature survey is to get an overview of the efforts made by the research community in this domain. The detailed literature review and the current state-of-the-art approaches in this field of study are discussed in this section.

The rapid growth of online social networks and micro-blogging services is witnessed in recent decades ([43] [44]). Facebook, LinkedIn, Flickr, Twitter, Foursqure are among the popular services. These platforms serve different purposes for socializing, for instance, Flickr provides a platform to the users for sharing their photos, Twitter and Facebook are general purposes social networking platforms, LinkedIn is domain-specific and Foursqure is a location-based platform [45]. These social networking platforms such as Twitter facilitate the users to share information, reviews, opinions, and experiences about different topics, events, and news, of their interest [46] [47]. These platforms are commonly used in emergencies and natural disasters to inform others about the event. These social networks like Twitter, become the first place for many people to look for the updated news and its real-time content [48] [49] [50].

The statistics[1] of Twitter express for its fame as it has 134 million active users per day with 500+ million tweets per day, and its data volume is increasing 30% per year. Kwak et al. recognize that 85% trending topic of twitter is the news headlines [38]. On many occasions, Twitter has proved its ability to break news to the world before it comes-up from any other channel.

The formation of sections in this chapter is that; it started by discussing the importance of social media platforms in today's daily life. The following sections are formulated as we will discuss the role of social media platform TWITTER and its implications in various real-life applications and domains such as sentiment analysis, recommender systems for financial and educational sectors, and for the events, location, and news detection. After the general section of a discussion on Twitter's role, we discuss in detail its role in the field of disaster events, the strategies used for the extraction of such information. The importance of eyewitness reports and the extraction of features for identification of eyewitness are discussed in detail. The chapter is concluded by presenting a summary of recent techniques, by discussing the feature extraction approaches in this domain.

2.1 Role of Twitter

The social media platforms are exploited by the researchers and community for different goals such as fund-raising, education purposes, financial market predictions, alerts systems, location extraction, event detection, recommendation systems, disaster management systems, sentiment analysis, influential user identification. A recent study by Auter and Fine finds out that Twitter and Facebook are the most visited social media platforms in the U.S. and around the globe [51]. Due to its unidirectional model of relationship and worldwide coverage with more than 300 million monthly active users, Twitter is exploited by researchers. Researchers of these domains have proposed approaches to help the community in selecting the appropriate required information using the Twitter platform.

[1] https://www.internetlivestats.com/twitter-statistics/

The following subsections discuss some of the key areas that are studied by researchers over the last few years and the advancement achieved in those domains.

2.1.1 Sentiment Analysis

In today's age, the freedom-of-speech and liberty of expression, allow everyone to share their thoughts or opinion on any point. This sharing of thoughts, opinions, and reviews are the key features provided by social media platforms. Sentiment analysis is undertaken as a classification job where the user of the approach chooses which classification algorithm to use [52]. It is a process of mining the opinions, emotions, views, and classifying them into "positive", "negative" and "neutral" categories [18]. The extraction of useful information from the huge data available on these platforms always remains a challenge for researchers of this domain.

Sentiment classification can be done using three levels of extractions that are document level, sentence level, and feature level [53]. A comparative study performs by Bhavitha et al., divided the sentiment analysis studies into three types that are (1) lexicon-based, (2) machine-learning-based, and (3) hybrid approaches [54].

Lexicon-based: The lexicon-based approaches are divided into two classes that are dictionary-based and corpus-based. Initially, the sentiment classification was performed using dictionary-based approaches where the static dictionaries of key terms like; WordNet and SentiWordNet, were used to performed by matching terms from content with the dictionary terms [53]. The corpus-based approaches adopt the technique of using the content of the documents in the collection. The commonly used techniques are conditional random field (CRF) [55], k-nearest neighbors (k-NN) [56], and hidden Markov models (HMM) [57]. A recent study by Haider et al. [58] analyzed the impact of adverbs as feature for sentiment classification for product reviews on Twitter.

Machine-learning-based: The Machine-learning-based approaches includes the naive Bayes classifier [59], maximum entropy classifier [60], or support vector machines (SVM) [61]. These are called the traditional models and they use lexical features, part-of-speech, sentiment lexicon-based features, and the accuracy of the system depends on the selection of features [53]. In addition to these models that are categorized as traditional models we have deep-learning models category that includes, convolutional neural networks (CNN) [62], deep neural network (DNN) [63], and recursive neural network (RNN) [64]. These models produce better results in comparison with the traditional model, using the document level, sentence level, and feature level approach.

Hybrid: The hybrid approaches use the properties of both Lexicon-based and Machine-Learning based techniques. The most commonly used features in hybrid approaches are the sentiment-lexicon [65]. A recent study by Kumar et al. [19], has systemically reviewed the literature of sentiment analysis techniques on Twitter.

2.1.2 Recommender Systems

In recent years the recommender systems are significantly evolved and become the essential part of the WWW. From the product recommender system to the article recommendation system, this domain is exploited by researchers to help the community taking their decisions from mobile app selection to the financial decisions [20].

In general, the recommendation systems can be categorized as, personalized and non-personalized. In non-personalized recommendation systems, the user's preferences as an attribute to approach are not utilized, e.g. "top 5 movies of the month". In a personalized recommendation system, user preferences such as its profile, and item characteristics are utilized by the approaches for recommendation purposes.

Collaborative filtering based: The Collaborative filtering based approaches adopt either user-to-user based collaborative filtering or item-to-item based collaborative filtering. The user-collaborative filtering approach that is, if a person P1 has the same opinion as a person P2 on an issue A1, then A1 is more likely to have the same P2's opinion on issue A2 than a randomly chosen person. The recommender system adopting this technique recommend the results based on other people's review having the same preferences, likes, dislikes [4]. These types of recommender systems are called personalized recommender systems. The item-to-item collaborative filtering adopts the same concept of collaborative filtering as adopted in user-to-user based, where items X and Y are highly similar and if the maximum number of the users who purchase these items X also buy item Y. Then, the recommendation is given to a user over an unrated item Y is predicted based on the user's rated item X [4].

Content based: The Content-based recommendation systems use features and characteristics to find similar items. The user's history of likeness to purchase a similar option is used as a parameter for recommendation [4][3].

Other Approaches in Social Media: The researches of this domain have adopted the idea of using the user's profile, social-media information, reviews, comments, and tags for the recommendation of items. There are different recommender systems proposed for various tasks that utilities these social-media platforms for recommendation purposes[4][3][5][6].

Followee Recommendation: This concept was firstly adopted by Twitter, called the unidirectional model. In this model, a user may follow a user who shares interesting trending topics but they are unknown to each other. These users can be treated as information sources for the user. Researchers of this domain have also exploited the twitter's concept of "follow", and based on the idea the topology-based approaches for followee recommendation system are proposed [5][6][7][8][9].

On Twitter, there is an option called "who to follow", on the main page of a user, that shows the recommendation based on existing follow-ship and followees of his followees.

2.1.3 Financial Decision Making

The role of human sentiment, emotions, and mood in making the financial decision, is comprehensively studied [66][67][24]. For financial predictions, the online data sources are categorized into four types that are *News, Media, Web Search (query), and Social Media feeds*. Daniel et al. [66], adopted sentiment analysis to detect the popularity of events using tweets posted by the financial community. The existence of the financial community was proved and discussed by Steve et al. [68]. The author proves with pieces of evidence the existence and influence of this community on the financial market. The sentiment analysis approach is used by the author [68] with a centrality based network approach for the identification of influential nodes.

Trust plays an important role in everyone's daily life, as it helps people in making their decisions. The financial communities over social media platforms work using the concept of trust, and finding trustful information from large twitter data is not an easy task. Trust networks are created by financial communities. The people are not directly connected to these networks but they use the relationship of Friend-Of-A-Friend (FOAF)[67]. By using data of stock-related tweets, a trust management framework was proposed by Yefeng et al. [67], it is called the trust modeling technique. The approach builds a user-to-user network using the technique of the trust management system.

It is proved by various researches that the prediction of future stock and returns are taken from social media sites. The concept of investor's opinion and price discovery process is influenced by the existing financial communities over the social media platforms [69]. A similar study by Ruiz et al. [70] proved the correlation between activity predicted by social media platforms and the stock-market events. The author also discussed the real-life application of the study in developing profitable approaches.

Renault [71] adopted the technique of field-specific lexicons rather than using the standard dictionary. Renault utilized the dataset of messages from *StockTwits.com* and created a lexicon of words used by investors while sharing their opinions. He proposed the use of the lexicon dictionary instead of the standard dictionary used for work matching. The results of the approach concluded that the change in investor sentiment of the first half-hour helps forecast the last half-hour of the market return. Renault comprehensively discussed the effect of online investor's sentiment using; *"investor sentiment indicators", "Predictive regressions", and "Exploring investor base heterogeneity"*.

The researchers of this domain have proposed language modeling based techniques for financial predictions using social media platforms. Gro-Klumann et al. [72] discussed the importance of social media in financial decision making and comprehensively exploited the role of sentiment analysis in this domain. The author proposed the concept of statistical language modeling technique and builds a directional sentiment metrics. The results are then linked to aggregate the returns of the stock index. The role of experts from the financial community is the people behind the wheel of financial markets and Twitter sentiment [72]. The study has proven its claim by getting useful results from the proposed idea.

2.1.4 Education Sector

The connection between education and technology is stimulated as a fundamental component in the transformation of the education sector over the past few years. The researchers have exploited Twitter for its applications in the educational sector since its launch in 2006 and there are several blogs on how to actively use it for educational purposes [10][11].

A comprehensive survey was conducted by Jeffrey et al. and find out how Twitter can be utilized educationally: communication, classroom activities, and professional development (PD) [10]. Researchers of this domain have exploited Twitter as a potential tool as PD that allows others to respond to a teachers question using just in time and on the spot approach, and the teacher can inform the students about the latest updates on any topic [73]. Traditionally the just-in-time PD setup needs to involve knowledge brokers and flexible structure to implement the idea and the teacher was not driving all the arrangements. But for the teacher, Twitter makes just in time study possible using the educational tags and its driven by the teacher without the involvement of any other party [73]. The dataset of various educational tags and 3598 users to support teachers in their discussion and delivery.

A recent article by Rosell-Aguilar studied the role of educational tag *"#MFLtwitterati"* on Twitter [16]. The study has proven that the teacher who used this tag, get engaged in collaborative practices. The research work of Rosell-Aguilar contributes to promoting the use of Twitter and Mobile Learning as a Personal Learning Network, and encourage the community for Continuous Professional Development.

A recent study by Tian et al. investigated the participation of students in Twitter supported activities [12]. The experiment of the techniques proves Twitter as a useful resource in keeping the learners engaged in live discussions and questions for productive participation. The participation of the class student in class discussion activities resulted that those students posted more tweets than other students with low or not participation. The results outperform the traditional Blackboard discussion technique by sharing reviews and personal reflections.

The usage of twitter as an educational application is commonly discussed for higher education rather than the first-grades [13], but few recent studies have exploited its contributions for undergraduate and school studies [14][15][16]. The potential use of hashtags by professors for different courses keeps the students engage for access to information related to course material [17], and to keep them up to date on all topics discussed.

Recently, a study is performed by Jeffrey et al. [74] which evaluated 16 of educational hashtags for 13 months data of 2.6 million tweets. It was observed that the tweets related to these tags were less original tweets but re-tweets and link sharing. The studies in this domain have limitations like the language, as the studies in the literature have been exploited using the English language only.

2.1.5 Event Detection

The capacity of twitter in the field of event detection is also studied by researchers of the domain, and various techniques are proposed to detect events from tweets [25]. These event detection techniques are categorized into feature-pivot techniques and document-pivot techniques.

The feature-pivot techniques are discussed and applied by researchers and analyze the feature distribution for the detection of events [26] [27]. The document-pivot techniques cluster the document based on textual similarity like TF-IDF from the large noisy data [28][29][30][31]. Based on the available information the event detection techniques are divided into two categories; Unspecified events and Specified events [25]. Where the specified events are those for whom we have no prior information and specified events are those events on which we have prior information or features available.

A recent study has proposed an online algorithm that group together the tweets into clusters by effectively discovering the interest based on their temporal and textual features [32]. The approach evaluated the temporal and textual features for the identification of groups of interest. The approach is called the bursty approach as it creates the groups of tweets of the same interest in almost real-time. This work utilizes the concept of the time window and active/inactive clusters for consideration of historical data within the approach [32].

A most recent study conducted by Subramaniyaswamy et al., reported that approximately 75% of the world's population is using social media platforms [75]. In every country of the world, Law enforcement agencies take all sorts of measures to safely conduct public events. For that, the information-sharing platforms are also analyzed for security purposes using sentiment analysis and lexicon-based techniques. Any sense of threat from tweets is alerted as a threat to an event [75].

The researchers, Gonzalo et al. [52], exploited the concept of adopting sentiment analysis with Bayesian networks for the identification of critical events. This work exploited the classifier on the Spanish language using Bayesian networks to perform sentiment analysis. The results of the idea produced better results in comparison to the Support Vector Machine (SVM) and the random forest techniques. The relation among words is also exploited as a feature to the model [52].

Event detection systems have high computational cost, which is a major challenge with these systems. Hassan et al. have proposed an incremental clustering-based technique to provide a solution with low computational cost and named it TwitterNews+ [33].

Kumar et al. [36] proposed the idea of using Convolutional Neural Network (CNN) for the identification of location extraction words from tweet content. The proposed technique achieved the accuracy of the result of 92% for extraction of the location words, of three to four-words long locations.

2.1.6 Location Detection

With every passing second, the bulk of twitter messages are generated by millions of active users on twitter [43] [44]. The users share their locations either explicitly or implicitly by mentioning the location names or discussing its attached characteristics [76]. This shared location mentioned in these messages is utilized by various applications for text mining such as news, emergency events [45] [77]. The data attached with Twitter includes (i) short tweet text of maximum 140 characters that are posted by a user, (ii) User network information (a limited version of it), and (iii) The meta-data information of the user and tweet like its time, and time-zone [45] [78]. Based on these data types discussed, the approaches and techniques are divided into three categories. 1) Tweet Content, 2) Tweet Context, and 3) Twitter Network.

Based on the available information on the twitter network and with the tweet, the location predictions can be categorized into three major types called user home location, mentioned location, and tweet location. The following paragraphs discuss each of these location types with approaches proposed in these categories.

Home Location The term "Home Location" refers to the user's home address. The information regarding home location is collected from users' profile that is declared by the user himself at the time of account creation or updated later at any stage [45] [78]. Moreover, this home location can be obtained using different techniques proposed by various researchers of this domain. The following are three techniques, commonly used to present the home location. (1) Geographical grid: that presents the earth map as a grid of equal-sized cells. The users' home location cell is marked for user identification. (2) Administrative region: in this technique, the home locations are presented using the country name, its state, and city. (3) Geographical coordinates: in this presentation type, the home locations are identified with the help of latitudes and longitudes coordinate information received from the user when using GPRS enabled devices. For better results and coverage the user-provided information and geotags can be used in combination as a hybrid approach.

Tweet Location Tweet location refers to the location from where the tweet has been posted [44] [45], and its a challenging task with big noisy data shared by users belonging to different communities in an unstructured format. Such location information is very useful for real-time applications related to emergency response as it shows the current location of the user [78]. This type of location information is generally obtained from geolocation information attached to the tweet and labeled as geotags that came from GPRS enabled devices IP based location techniques. These techniques have their pros and cons. Another feasible approach for identification of tweet location is the point-of-interest (POI) and the coordinates attached to these POIs. The tweet location identification is a challenging task and an active research area in this domain.

Mentioned Location Both categories of location predictions discussed above require information about twitter network location information attached to the tweet. Such information is generally not shared by users due to obvious privacy reasons.

An alternative solution to this problem can be the use of user mentioned location names in the tweet content [44] [45]. These mentioned locations can be useful for location predictions and to better understand the context of the tweet. The mention of location by a user can be an explicit location or the implicit location in the tweet content [78]. The researchers of this domain have proposed various techniques for extraction of these mentioned locations from the content that includes two sub-tasks called recognition and disambiguation of location. These mentioned location prediction approaches involve tasks of pre-processing and pre-defined location databases such as Foursquare[2].

The location detection for all three categories is a vast field of study and researchers have extensively exploited this domain and proposed hundreds of techniques in this domain. Each category has a large number of techniques for prediction of location either from tweet content, context, user profile, network information.

2.2 Role of Twitter for Disaster Events

Twitter, with its popularity and real-time information sharing about natural disasters, has motivated many people to look forward to the latest news [50]. The authors Kryvasheyeu et al. [46] presented an in-depth study of Twitter activities about disaster events before, during, and after the event. For evaluation, the author studied the Sandy-Hurricane event, by analyzing the tweets from 50 major cities of the U.S. The study concluded with findings and results that the social media platforms play an important role to analyze the destruction caused by the event of the large-scale catastrophe.

The researchers have proposed diversified information extraction approaches to help institutions in responding to disasters. The role of Evidence Aid[3], a dedicated website with social media and other people works for disaster risk reduction, pre-planning, reactions, reclamation, restoration, and flexibility [47]. Similarly,

[2] https://foursquare.com/.

[3] www.EvidenceAid.org

Amaratunga promoted the idea of CBR, the community-based researches, as a tool for planning against the risks of a disaster event in remote, rural, and coastal areas in Canada [48]. It was a pilot project called "Virtual Community Of Practice" (VCOP). The base idea of VCOP was to gather or extract the information and disseminate it among the communities in remote areas.

Twitter is exploited during the social crisis, and its emergence and domination are discussed among other social reporting platforms [79]. The author exploited the issues of information processing from social reporting content for extracting the useful information by avoiding the rumors and gossips or the noisy data. Similarly, Haworth et al. [80] highlighted the importance of volunteered geographic information (VGI) for all phases of disaster management from prevention to recovery. The citizens generated geospatial data practices that are utilized by intelligent observers of the domain. The key idea is to utilize the data contributed by the users to take preemptive measures for any unseen event.

The disaster management or emergency response system needs precise, credible, and accurate information to respond accordingly. Meier [81] exploited the Haiti earthquake of 2010 using a live crisis maps approach to study the humanitarian responses and impact of social media technologies for the event.

2.3 Information Extraction from Twitter

The extraction of useful information from big data volume of tweets is a vital task particularly in the situation of a catastrophe event. Due to their nature, the disaster management system requires accurate and precise information to react accordingly. Based on crowd-sourcing; the research community have proposed diversified techniques for classification of user tweets such techniques include; a technique called Figureeight[4] based on the idea to provide help (formerly known as Crowdflower [82]), a category labeling technique of Standby Task-Force[5] [83],

[4] https://www.figure-eight.com/

[5] http://www.standbytaskforce.org/about-us/our-history/

Amazon Mechanical Turk (AMT)[6] [84] a technique for removal of biasedness by involving “experts” and “non-expert” volunteers. To facilitate the disaster relief institutions, the research community has developed various real-time crawler-based applications; such as AIDR [83], Twitcident [85], TweetTracker [86], situational awareness based ScatterBlogs [87], and application for Cross-language aspects on Twitter [88]. Another crawler-based application for Super Typhoon Yolanda (Typhoon Haiyan) in the Philippines was developed by Takahashi et al. [89]. The credibility of data generated from crowd-sourced based techniques suffers from noise and quality issues for reliability and objectivity of information as input to real-life disaster management systems [90].

In the event of an emergency or a disaster initially, the information processing tasks were performed by using known-networks of volunteers based crowd-sourcing platforms, such as; CrisisCamp[7] and Ushahidi[8] [91]. Such IT platforms are analyzed and studied by Zook et al. [91] and comprehensively discussed their importance and role for relief works either for an individual or for agencies in the Haiti relief efforts. Similarly, Ostermann et al. [92] exploited the networks of volunteers by adopting a semi-automatic approach of extracting geographic information from social media. To evaluate the proposed idea author used the forest fires events to study the role and impact of social media in crisis events.

The effectiveness of a disaster management system relies on reliable information as input to the system. The research community has discussed the importance and effectiveness of information input to a disaster management system and its responses for relief and related recommendations [80] [93]. One of the practical examples is the Twitter Earthquake Detector (TED) program, funded by the American Recovery and Reinvestment Act. It gathers real-time earthquake-related messages and applies location, time, and keyword filtering to track the earthquakes, to generate

[6] https://www.mturk.com/
[7] https://crisiscommons.org/crisiscamp/
[8] https://www.ushahidi.com/

alerts and reports, like USGS[9][10]. Similarly, San Diego State University has developed an Emergency warning system (SDSU[11]), based on social media platforms for broadcasting emergency warnings to San Diego citizens.

2.4 Twitter Eyewitness Reports

Alecia Swasy is Chair in Business Journalism at Washington and Lee University, published a book titled: "How Journalists Use Twitter: The Changing Landscape of U.S. Newsrooms"[94] discussed the influence of social media on daily life as well as on newspapers. In his study, the author has shown how journalists from major newspapers are promoting Twitter as a news tool. A year after this book, the author published an article, critically discussing the impact of social media on journalists and news organizations [95].

Mengdie et al. [96] conducted an in-depth study on Twitter to understand the factors behind the news broke. The authors of the paper identified the three opinion leader groups, that play a vital role to spread the tweet. The author discussed the unidirectional relationship model of Twitter and the significance of the elite users like journalists and politicians and their impact on spreading the news. The author concluded the study by suggesting Twitter as a potential tool for news.

A recent incident happened when a Boeing 767-300ER of Delta Aircraft[12] flight took off from Beijing made an emergency landing on a remote island in Alaska, due to its potential engine lost. The incident was shared by an eyewitness who posted the tweet with a picture of the crashed plane, and the tweet became the news source of a news agency. Tweet with text is shown in the Figure-2.1.

[9] https://gcn.com/articles/2009/12/21/usgs-earthquake-twitter-tweets.aspx,
[10] https://www.usgs.gov/connect/social-media
[11] https://phys.org/news/2014-10-viral-messaging-twitter-based-emergency.html
[12] http://channelnewsasia.com/news/world/delta-flight-middle-of-the-ocean-seattle-beijing-emergency-land-11062706

FIGURE 2.1: Eyewitness Tweet; Emergency Landing of Delta Aircraft

A US Airways flight[13] with 155 passengers and crew members, crashed into Hudson River, New York. There were dozens of tweets posted by Twitter users, around 15 minutes before the crash, about the possible plane crash. After the plane crash, the first tweet came with crash information after 4 minutes of recorded crash-time, with a picture telling the whole story. Later the news of the plane crash became the headline of The Daily Telegraph. The picture was shared through the iPhone device using the "TwitPic" photo uploading software for Twitter at that time, shown in Figure-2.2. Antony Mayfield [97] in his book discussed the event and share his thoughts that the journalists and other users of Twitter will look for eyewitness reports of the breaking news.

[13] https://www.telegraph.co.uk/technology/twitter/4269765/New-York-plane-crash-Twitter-breaks-the-news-again.html

FIGURE 2.2: Eyewitness Tweet; Aircraft Crash of Hudson River

In the event of operation against Osama Bin Laden in Abbottabad[14], the event was tweeted by a young IT consultant Mr. Shoaib Athar. He tweeted about the Halicaptor movement over the city at midnight (1 AM PST), without even knowing whats going on. The original tweet he wrote was; "Helicopter hovering above Abbottabad at 1 AM (is a rare event),". After the announcement by the officials, he again tweeted; "Uh oh, now I'm the guy who liveblogged the Osama raid without knowing it".

In Nairobi Kenya, four masked gunmen attacked the Westgate Shopping Mall. It was a massive attached that resulted in 71 deaths and 200+ causalities. The information about the incident was broken by an eyewitness who tweeted it 33 minutes before any TV Channel report [15]. A similar type of incident happened during the annual Marathon in Boston city. The homemade bombs took the lives

[14] https://www.poynter.org/reporting-editing/2015/today-in-media-history-in-2011-twitter-broke-the-news-of-osama-bin-ladens-death/

[15] https://en.wikipedia.org/wiki/Westgate_shopping_mall_attack

of three and injured several spectators near the finishing line. The information about the incident was available on Twitter well before any news channel's reports [16].

The popularity and importance of Twitter are discussed for breaking the news of natural disasters, as it happens. Twitter is considered to be a go-to source for the majority of people for news. Southern California sees an earthquake of 5.4-magnitude. Multiple tweets about California Earthquake were available on Twitter a minute earlier than the U.S. Geological Survey (USGS)[17], recorded time. The first tweet that breaks the news was: *Nicholas Hawkins; "holy **** earthquake in so cal"*, the text with asterisks shows the presence of expletive words.

2.5 Eyewitness Feature Extraction

The institutions, responsible to respond and provide relief services in event of a disaster, need credible, precise, and accurate information to actively respond to the event [98]. The person who is witnessing an event (eyewitness) can honestly provide details such as; precise location, the intensity of the event, the estimated number of fatalities, and other related pieces of information regarding the event. The extraction of eyewitness reports from a large amount of Twitter data is a challenging task and still an open research area of this domain. Diakopoulos et al. [99] presented a technique of identifying the eyewitness reports from millions of tweets associated with journalism. Olteanu et al. [100] presented the study of eyewitness identification on criminal justice as well as for natural disasters. The study by Kumar et al. [101] in 2013, exploited the users location information to identify the local and remote users for a disaster event. Morstatter et al. [102] presented the technique to identify a feature set to automatically classify the disaster events using the tweet content, by identifying the semantic patterns. The studies by Kumar et al. [101] and Morstatter et al. [102], used the language

[16]https://en.wikipedia.org/wiki/Boston_Marathon_bombing
[17]https://latimesblogs.latimes.com/technology/2008/07/twitter-earthqu.html

information and the location of the tweet to capture the source account, although identification of an eyewitness account is not taken into consideration.

Processing millions of tweets to identify the eyewitness reports remains a perplexing task and an open research area. Truelove et al. [39] in their study, presented and evaluated the theoretical model, for the identification of witness account and related impacting accounts, for various events. The model exploited the different account categories of witnesses like; Witness Account (WA) that provides the direct event's observation, Impact Account (IA) that is not the direct observer but potentially impacted, Relay Account (RA) are accounts that relays on other IA and WA accounts, and Not Witness Impact or Relay Accounts (NWIRA) the accounts such as on-topic individual and original (OIO), but not come in the category of WA, IA, RA. The results of the study by Truelove et al. describe the impact of various factors that directly affect witnessing characteristics. The events of Shark Sighting, Concert, Protest, and Cyclones were tested and evaluated.

Doggett et al. [40] presented the set of linguistic features to categorize the disaster event-related tweets into two categories of "eyewitness" and "non-eyewitness". The technique by Doggett et al. [40] apply the linguistic features like filters to calculate the semantic similarity for the document. For the eyewitness category, the author has identified the features namely; "First person", "Immediate temporal markers", "Locative markers", "Exclamative or emotive punctuation", and "Lexical exclamations and expletives". For the non-eyewitness category, four features identified are; "Jokes, memes, and incongruous emotion or sentiment", "Wrong part of speech, mood, or tense", "Popular culture references", and "Temporal markers". The identified features demonstrated their importance in surfing the breaking news from social media platforms.

The importance of linguistics features is further exploited by researchers and Fang et al. [41] presented an approach for eyewitness identification by combining linguistics with meta-features. The author presented the hybrid approach and constructed the feature set including the three Meta features (like; Mentions/hashtags, client application used, length of tweet) and three linguistic features (including;

Temporal information, Impersonal or Personal Expressions, Crisis sensitive feature). For topic classification, the authors adopted the word dictionary of LIWC[18] and OpenCalais[19] API. Fang et al. [41] evaluated the proposed technique by employing the static dictionary-based technique.

Tanev et al. [42], in 2017, present a set of eyewitness features using metadata, stylistic, lexical, and semantics dimensions. The list of stylistic dimension includes; "Personal", "All caps". The metadata list of features includes; "hashtags", "user mentions", "URLs", "Retweet counts", "application type". The lexical dimensions include; "uni-grams", "n-grams", "number of uni-grams". The semantics dimensions cater the "category definitions words", e.g. "Shelter Needed".

From the literature, it is identified that "None of the above" approaches has presented the concept of employing the expert-driven features for disaster events, to identify the reliable eyewitness sources. The state-of-the-art [37] approach, presented the methodology to categorizing the tweet whether it is posted by an eyewitness or not. The author developed a feature set for the identification of eyewitness reports by using an expert-driven based approach to recurrent natural catastrophes. The author identified a feature set to identify the eyewitness tweet, those identified features are called the domain-expert features. Zahra et al. evaluated the technique by using the word-based (bag-of-words) features in combination with the identified features to classify eyewitness tweets, using different datasets of tweets.

A recent approach, presented in 2019 by Zahra et al. [37], proposed the methodology to categorize the tweet by using expert-driven engineering for disaster events. The authors identified a comprehensive feature list called domain-expert features. Zahra et al. [37] evaluated the proposed technique using word-based features (bag-of-words) and the identified domain-expert features to classify the tweets. Figure-2.3 depicts the overall methodology adopted by Zahra et al. [37] to classify the eyewitness messages.

[18] http://www.liwc.net/
[19] http://www.opencalais.com/opencalais-api/

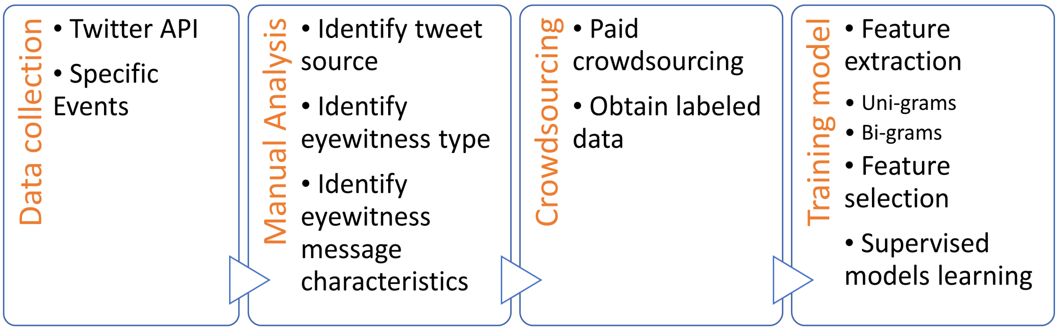

FIGURE 2.3: Eyewitness Identification Approach, Zahra et al. [37]

2.6 Summary

We have critically investigated all the literature in the above sections which led us to classify the significant problem discussed in this paper. The approaches presented, before Zahra et al. [37], aim for the identification of potential location information from tweet content and has not fully exploited the dimensions of identifying that either a source is an eyewitness or not. The technique proposed by Zahra et al. [37] explored the concept for the identification of a trustworthy eyewitness source. The summary of eyewitness feature-based techniques for the identification of eyewitness reports is demonstrated in Table-2.6. The Table-2.6 demonstrates the methodology, results, and shortcomings of Zahra et al [11] in tweet identification. Based on this literature review, we have identified some of the gaps and presented in Chapter-1, and to overcome those identified gaps we have proposed appropriate methodology discussed in Chapter-3 and 4.

TABLE 2.1: Summary of Eyewitness/Source Identification Techniques.

State-of-the-art	Concept	Approach	Findings	Shortcomings
Diakopoulos et al. [99]	This work presented the approach of finding the trustworthy sources of information.	The author exploited the role of technology during the crisis events. Author adopted a human centered design approach and developed the SRSR system to extract the useful information from the tweets by journalists.	SRSR approach achieved the high accuracy and secured the precision of 89% and recall of 32%.	Adopted the principle dictionary approach for eyewitness detector. No features to identify the eyewitness.
Olteanu et al. [100]	In this work, author has exploited the role of social-media and its related data during various types of disaster situations.	The author presented the idea of tweet broadcast recommendation, using the temporal information, based on the information seek by a user. The idea is tested for 26 different crisis.	Secured the accuracy between 0-54% for eyewitness accounts during the natural hazards.	Require twitter users and network information to recommend. No features identified.
Morstatter et al. [102]	The author aimed to identify the non-geotag tweets from crisis region and classify the tweets.	Author evaluated the techniques by (i) studying the difference between language of the tweets, (ii) identifying the linguistic patterns, and (iii) propose the automatic approach of identifying the non-geotag tweet from crisis region in real-time.	Using Naive Bayes classifier, achieved the F-Score of 0.882 for Hurricane Sandy and 0.831 for Boston Bombing.	Only focused the distinction adopted was the language of the tweet. No additional features are identified.

...Continued on next page

Table 2.1 – continued from previous page

State-of-the-art	Concept	Approach	Findings	Shortcomings
Kumar et al. [101]	This work aims to identify the active users subset during the disaster events, for the tweets coming from 5 countries of "Arab Spring".	Author of the approach adopted two dimensions to identify the relevant users during crises, (i) Users Location, (ii) User's affinity for a discussion topic. For topic of discussion detection, author adopted the Latent Dirichlet Allocation (LDA) and for geo-locality based profile location adopted the OpenStreetMaps Services.	Author claim to provide more and quality information for user identification.	Require twitter user profile and network information to recommend. No features identified.
Truelove et al. [39]	This work explored the idea for the identification of the eyewitness accounts and the related accounts to it.	The author presented the model for identification of the witness, related and relayed impact accounts. The tweets of 18th, 19th February 2013 of bushfire event at Hume Freeway road, connecting Melbourne and Sydney, were utilized to evaluate the proposed approach.	Secured 77% accuracy of classification of smoke events. The discussed event's effect is not truly an eyewitness, because of witnessing distance from disaster location.	Require manual pre-processing, Metadata, Location and Network info. No features listed for the identification of the eyewitness.

...Continued on next page

Table 2.1 – continued from previous page

State-of-the-art	Concept	Approach	Findings	Shortcomings
Doggett et al. [40]	Presented a filtration-based technique to identify posts from the eye-witnesses using predefined list of keywords.	The author presented the linguistic features and applied them as filters for identification of eyewitness reports.	Avg. accuracy of 62% is achieved.	It require Twitter user's Geo-location info. Eyewitness events were identified.
Fang et al. [41]	Presented a hybrid approach to automatically identify the eyewitness in the event of an emergency	A defined set of meta-data and linguistic features to identify the eyewitness, using the LIWC dictionary for identification of keywords related to an event, and then used OpenCalais for event labeling.	Achieved the average F1 score of 89.7%.	Used a static dictionary of terms. The technique requires language and location information for identification.
Tanev et al. [42]	The author presented a set of syntax and linguistic-based features for event detection.	To calculate the effects of a disaster event, adopted an unsupervised approach. The approach is then evaluated on news articles to detect the events.	The proposed approach achieved 42% Precision and 66% Recall, using an unsupervised approach.	Used the domain-specific data of news to detect the events.

...Continued on next page

Table 2.1 – continued from previous page

State-of-the-art	Concept	Approach	Findings	Shortcomings
Zahra et al. [37]	The authors manually investigated the tweet sources related to identify the eyewitnesses and classify them into thirteen identified features by the domain-experts.	Experts of related domains manually analyzed the tweets to identify the proposed features from the tweet content. Feature-words were extracted for all thirteen features.	F-Score of 0.917	Manual implementation. Failed to Implement all characteristics.

Table-2.6 presents an in-depth study of the state-of-the-art techniques in the domain of eyewitness identification. Table-2.6 demonstrated the concept, approach, and findings concluded by the authors. Subsequently, the shortcomings of each technique are discussed.

Table-2.2 shows the comprehensive list of features, adopted by authors of the state-of-the-art. A recent study by Zahra et al. [37], utilized most of the identified features from literature. Zahra et al. [37] in the study, discussed and evaluated the existing features, and also presented new features identified by domain experts. Table-2.2 presents the complete list of features identified from the literature and a corresponding value of "Y" in the relevant column of the state-of-the-art approach adopting that feature.

The Figure-2.4 depicts the taxonomy of various approaches from literature for eyewitness identification. It also identify the different Machine Learning approaches adopted by researchers of this domain to evaluate the effectiveness of these approaches.

Table 2.2: Feature-List Identified from Literature

Identified Features	**Doggett et al.** [40]	**Fang et al.** [41]	**Tanev et al.** [42]	**Zahra et al.** [37]
Reporting small details of surroundings	-	Y	-	Y
Words indicating perceptual senses	-	Y	-	Y
Reporting impact of disaster	-	-	Y	Y
Words indicating intensity of disaster	-	-	-	Y
First person pronouns and adjectives	Y	Y	Y	Y
Personalized location markers	Y	-	-	Y
Exclamation and question marks	Y	-	-	Y
Expletives	Y	-	-	Y
Mention of a routine activity	-	-	-	Y
Time indicating words	Y	Y	-	Y
Short tweet length	-	Y	Y	Y
Caution and advice for others	-	-	-	Y
Mention of disaster locations	-	-	-	Y
All Caps.	-	-	Y	-
Contains media	-	Y	-	-
URL	-	Y	Y	-
Journalist account involved	-	Y	-	-
Type of client application used	-	Y	Y	-
Mentions or Hash-tags	-	Y	-	-
Word Embedding (Semantic word relationship)	-	Y	-	-

We have recognized two main problems that have been ignored by the state-of-the-art approach [37] for identification of eyewitness tweets. Firstly, the study has used a predefined list of keywords or dictionary, which isn't extendable for diverse events and unobserved reports. Secondly, the practical implementation of proposed techniques is not helpful when we have millions of tweets to be treated in no-time. Manually identifying the feature-words using domain experts support

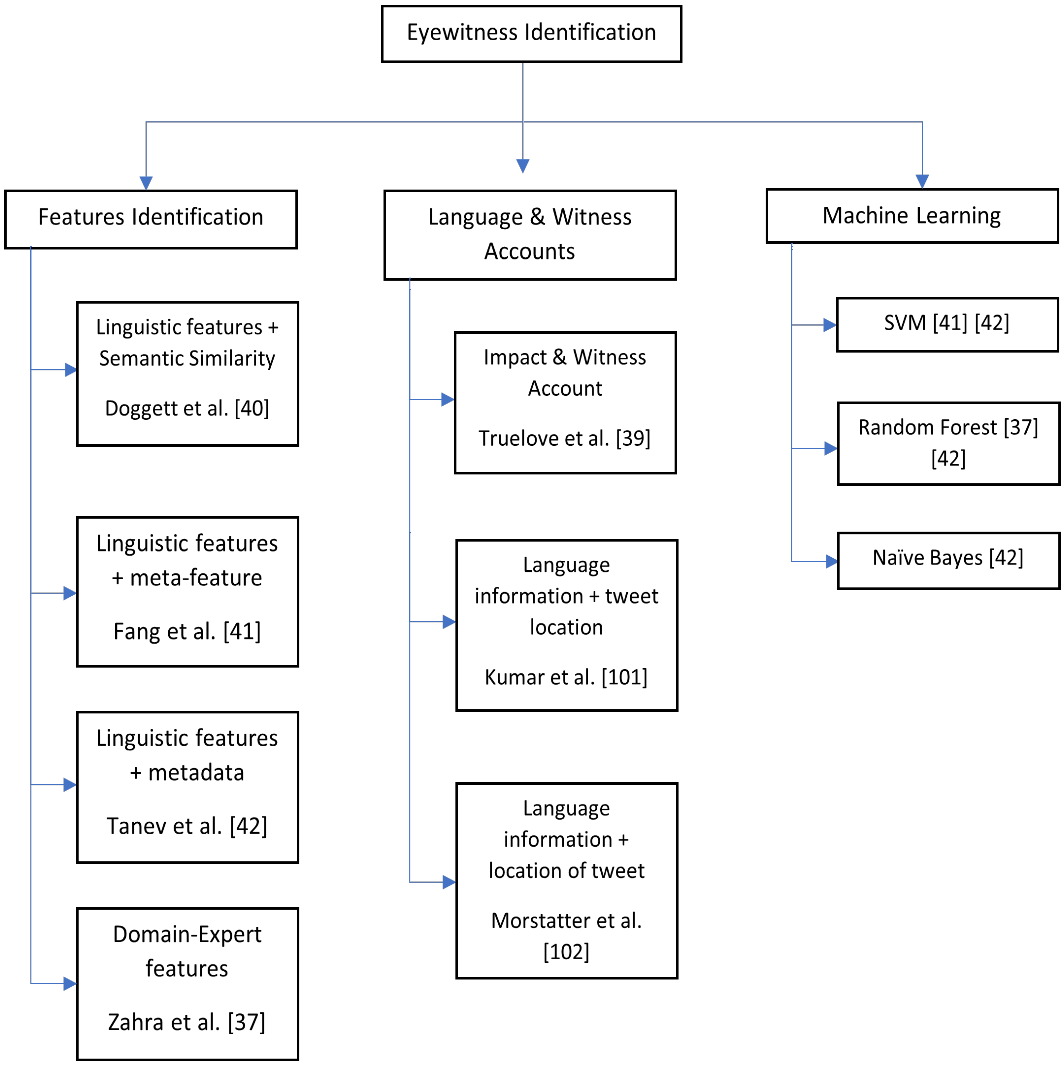

FIGURE 2.4: Approaches Used for Eyewitness Identification Using Content

and maintaining the predefined lists is time-consuming and updating new dataset in real-time is cumbersome.

In the light of the limitations discussed above, this work presented an automatic eyewitness tweet identification solution. It can perceptively process millions of tweets without needing the static dictionaries created by domain experts.

Chapter 3

Linguistic Rule-based Approach for Classification of Eyewitness Messages

This chapter illustrate the comprehensive discussion of the proposed LR-TED approach. The following sections discuss the data collection, gold-standard dataset, the importance of part-of-speech tagging and various required features, available options, and selection of appropriate grammar toolkit for implementation. The pre-processing methodology, the evaluation parameters, and strategies. The complete approach is explained in Figure-3.1.

The LR-TED model describes the overall approach for Employing Linguistic Rules on Tweets' Content to classify Eyewitness Messages for Disaster Events, and provides the "Rule identification for feature extraction" as an individual module that is comprehensively discussed in Chapter-4.

The dataset used to implement the proposed approach contains only the tweets that containing the query words such as; "earthquake", "hurricane", "flood", and "wildfire". The same dataset has also been used and publicly released by the manual features-based approach [37]. The dataset is acquired from CrisisNLP repository: `https://crisisnlp.qcri.org/`.

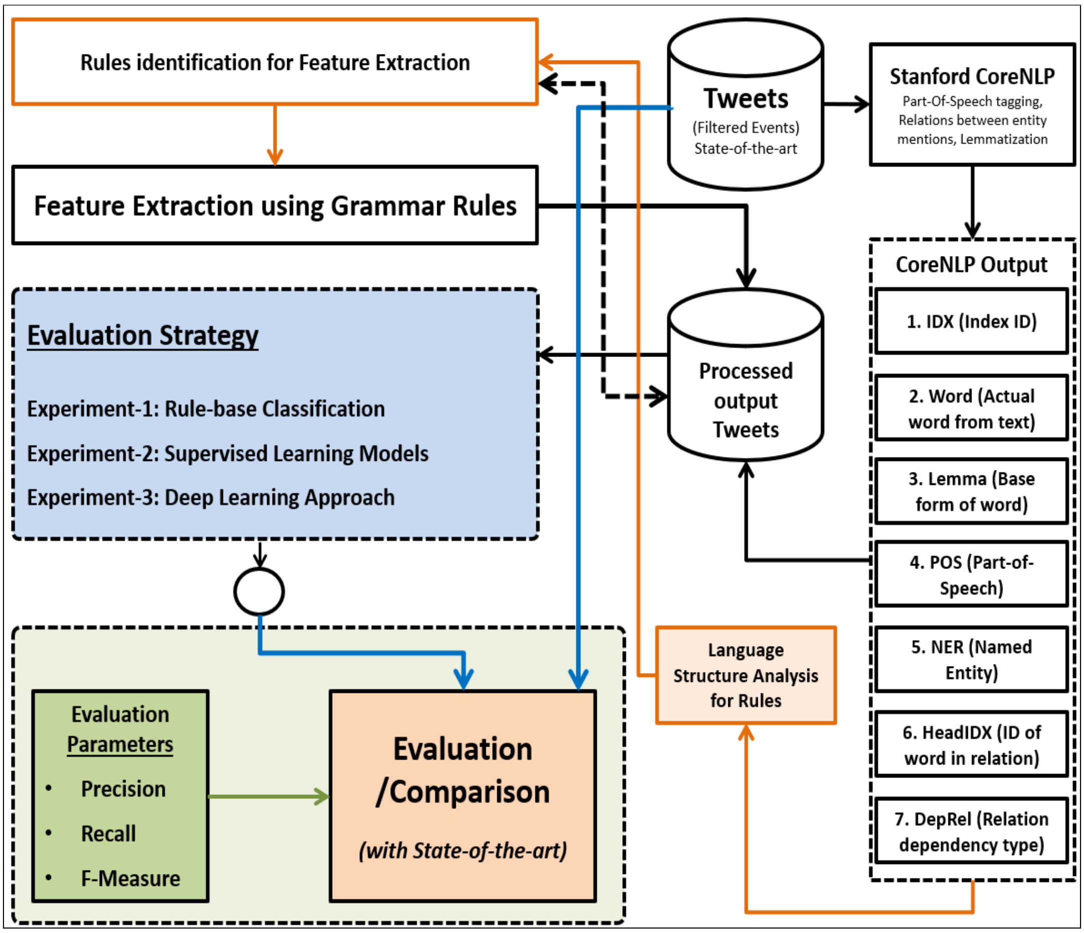

FIGURE 3.1: Proposed "**L**inguistic **R**ules to classify **T**witter **E**yewitness messages for **D**isaster events (LR-TED)" Research Approach

As the first step, the tweets are pre-processed to eliminate the noise, if any. After the pre-processing step, the tweets are then parsed and annotated by Part-of-Speech (POS) tagging tool. The selected tagging tool also provides the functions of NER, lemmatization, and entity-relationship details. In this research, we have used the CoreNLP[1] tool, that labels every word of the tweet to the corresponding Part Of Speech (POS) Tag, and also recognizes the relationship among the words within the sentence. CoreNLP generated output is then manually studied and evaluated by critically analyzing the language structure to identify the grammar rules for identification of all thirteen features identified by the manual features-based approach [37].

[1] https://stanfordnlp.github.io/CoreNLP/

The manually identified grammar rules are then converted into automatic functions. These automatic functions are then applied to the data for each identified feature, and the results are compiled for further steps. Now, the next step is to identify the class of the tweets into “eyewitness”, “unknown”, or “non-eyewitness” source, based on the extracted features. For this, we perform three experiments; (i) by using the manual classification approach (rule-base) where we have searched for those tweets that contain at least two features from the specified set, the threshold is set by evaluating the labeled “eyewitness” results having a minimum number of identified features. If the tweet contains no identified feature-word, it is then classified as “unknown”. If the tweet contains two or more features then it is classified as “eyewitness”, and if the tweet only has one identified feature, it is classified as “non-eyewitness”. (ii) by adopting the machine learning approaches, and (iii) by adopting the deep-learning approaches. The results of the proposed technique are stored in the dataset for further evaluations and experiments. The following subsections briefly describe each process, from data set selection to results evaluation, is illustrated in the approach shown in Figure-3.1.

3.1 Gold-Standard Dataset

The dataset employed in this research, is the same as used by Zahra et. al. [37] for evaluation of their technique. Initially, the dataset contained 25 million tweets from July 2016 to May 2018, collected by using the Twitter Streaming API (Zahra et. al. [103]). The author then refined the dataset by selecting the tweets that contain the focused keywords such as; “earthquake”, “aftershock”, “foreshock”, “hurricane”, “flood”, “forest fire”, and “fire”.

Taining Dataset: From the dataset of 25 million tweets, 2000 tweets as a sample for each of three disaster types, hurricane, flood, and earthquake are taken as the first dataset from 1st to 28th August 2017, for manual analysis by state-of-the-art authors. The details of sample tweets taken for training and analysis of the LR-TED approach are given in table-3.1.

TABLE 3.1: Dataset Statistics (Training)

Category	**Flood**	**Earthquake**	**Hurricane**	**Total**
Eyewitness	148	367	296	811
Non-Eyewitness	113	321	100	534
Unknown	1739	1312	1604	4655
Total Sample	**2000**	**2000**	**2000**	**6000**

Testing Dataset: The second dataset comprises 8000 tweets, excluding those used in the first dataset, picked at random from the collection of 25 million, and annotated by using the state-of-the-art crowd-sourcing methodology. The dataset contains 2000 random tweets for each disaster type; earthquake, flood, hurricane, and wildfire. The detail of the second dataset is described in table-5.1.

The datasets of 6000 and 8000 tweets, contains the annotated data by domain-experts and by the paid crowd-sourcing techniques adopted by state-of-the-art [37]. The training and testing datasets are separately available by state-of-the-art.

In this research we use the first dataset of 6000 tweets, as described in table-3.1, for manual analysis for language structure, linguistics, and the second dataset of 8000 tweets, as described in Chapter-5 table-5.1, is utilized for evaluation of the LR-TED approach.

The dataset also includes the annotated data gathered by the manual analysis and crowd-sourcing. So, for comparison of the results generated by the proposed technique, we use the same dataset as the gold-standard.

3.2 Pre-Processing

The language used in tweets generally does not follow the grammar rules of writing. To remove the noise from the text of the tweet, we have pre-processed the selected dataset. Generally, the pre-processing task of tweets includes, the removal of hash-tags, HTML tags, extra white spaces, and special symbols. After performing the pre-processing function, the noise-free tweets are forwarded to the parser.

3.3 Language Processing

Let us consider the tweet text *"Just in case y'all haven't heard California is on fire. Even in Oakland we're literally inhaling and driving through clouds of smoke..."*. From this tweet we can manually identify different features as discussed in table-1.1, such as “California” and “Oakland” are the locations, “driving” and “inhaling” are routine activities, “heard” show the property of perceptual senses. In this sentence, we have these feature-words as verbs, adjectives, and nouns. The motivation is to extract these verbs, adjectives, nouns, and adverbs for the extraction of the feature-words.

For identification of the feature-words, we need to perform grammatical tagging or part-of-speech tagging. For language processing, Part-of-speech (POS) tagging is one of the most important tasks. Jurafsky et al. [104] have comprehensively discussed the Part-of-speech (POS) tagging and different tag-sets and the related approaches adopting those tag-sets. The Stanford, Lancaster Oslo/Bergen (LOB), CLAWS, VOLSUNGA, and POS Tagger, are briefly introduced and Stanford NER achieved the accuracy of 97.4% among others [104]. The approach of Named Entity Recognition (NER) has been applied to various text types and for diverse entities types such as blogs, tweets, newspapers, locations, fiction, persons, and chemicals.

Each tweet from the dataset will be tagged with the appropriate part-of-speech tag. In addition to part-of-speech tagging, we wanted to exploit the language structure by investigating the relationship among words in a sentence either by identifying the patterns or the dependency relations of words.

The data from social media platforms is not structured and potentially contain a lot of noise in the text. The words might be used in wrong forms and to cover this issue by using base words for more accurate matching. There are two types of approaches used for extracting the base-word that are; Stemming and Lemmatization. The approach of lemmatization has an advantage over the stemming process.

Tokens:	Tokenization of the tweet content
POS:	Part-of-speech for tagging of each word
NER:	Named Entity Recognition for entity identification
LEMMA:	Lemmatization for getting the base word
DepRel:	Relation Dependency among words

TABLE 3.2: Standard CoNLL Output-Format

Field Number	**Field Name**	**Description**
1	ID (idx)	Token Counter, starting at 1 for each new sentence.
2	FORM (word)	Word form or punctuation symbol.
3	LEMMA (lemma)	Lemma of word form.
4	POSTAG (pos)	Fine-grained part-of-speech tag.
5	NER (ner)	Named Entity tag.
6	HEAD (headidx)	Head of the current token, which is either a value of ID or zero ('0').
7	DEPREL (deprel)	Dependency relation to the HEAD.

The list of all requirements with part-of-speech tagging, to understand the structure and linguistics features from content, includes the followings;

The input to the proposed system is the text of the tweet, and for any tagging system, the first step is to tokenize the text. After tokenization rest of the tagging rules are implemented. The worth of Stanford is known in the domain of natural language processing [105] [106]. The accuracy of the Stanford NER tool and POS tagging, it is exploited by researchers [106] of this domain with high accuracy around 97.4%. It also has the edge over other similar approaches that it has a complete package of different grammar rules.

The Stanford CoreNLP is a Natural Language Processing (NLP) toolkit that includes various grammatical analysis tools including; the parser, named entity recognizer (NER), part-of-speech (POS) tagger, sentiment analysis, and the open information extraction tools.

There are numbers of major languages that are available as an online service with Stanford CoreNLP (see `http://corenlp.run/`), or it can be implemented into modern programming languages as an API.

(F)	**Rule**			**Example**		
	SSES	→	SS	caresses	→	caress
	IES	→	I	ponies	→	poni
	SS	→	SS	caress	→	caress
	S	→		cats	→	cat

FIGURE 3.2: Porter's Algorithm

Stanford CoreNLP provides all the features required in the above list. The Stanford CoreNLP, output in CoNLL format. The author of the standard CoNLL output format is Gabor Angeli [2]. The table-3.2 is taken from the author Gabor Angeli.

IDX: The first field in table-3.2 is "IDX", the token counter or ID index of a word/token in a sentence from the text, and it starts from 1 after for each sentence.

Word: The second field is the "Word" field that contains the token word extracted directly from the text.

Lemma: The third field in the table is titled as "Lemma" and it has the base token word from tweet text. In Natural Language Processing (NLP) there are two ways of finding the base word; (i) Stemming of words, which is defined as *"Stemming, usually refers to a crude heuristic process that chops off the ends of words in the hope of achieving this goal correctly most of the time, and often includes the removal of derivational affixes"* [107]. The most commonly used stemming algorithm is called the *Porter's Algorithm*, which works with 5 phases for word reduction as shown in Figure-3.2. (ii) Lemmatization is defined by Stanford NLP as *Lemmatization usually refers to doing things properly with the use of vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only and to return the base or dictionary form of a word, which is known as the lemma* [107].

POS: The fourth field is "POS", that contains the corresponding fine-grained Part-of-speech (POS) tag of the token word, that are defined using Penn-Treebank-Tagset (see `https://www.ling.upenn.edu/courses/Fall_2003/ling001/pen`

[2] `https://nlp.stanford.edu/nlp/javadoc/javanlp/edu/stanford/nlp/pipeline/CoNLLOutputter.html`

n_treebank_pos.html). The POS tags with examples are described in Figure-3.3.

NER: The fifth field of the table is titled as “NER” that holds the Named-Entity Recognition (NER) tag.

The sixth and seventh fields, titled “HeadIDX” and “DepRel” respectively, contain the dependency type and index of the dependent word.

HeadIDX: The sixth field of “HeadIDX” holds the ID index of the dependent word or zero when no relation (it identifies the word in relationship).

HeadIDX: The seventh field, hold the type of relationship dependency to the second word. Here Stanford has provided a complete list of dependency in a manual [108], available at https://nlp.stanford.edu/software/dependencies_manual.pdf.

Figure-3.3 show the Penn Treebank Part-Of-Speech tags with their description and examples, available at https://www.cis.upenn.edu/~bies/manuals/tagguide.pdf. The first column contains tags, issues by Penn Treebank POS-tagger to words in a sentence. The second column describes the meaning of the tag, e.g. “JJ” is an “adjective” and the third column shows some example words like “yellow” is an adjective. Figure-3.3 shows 45 tags that the Penn Treebank POS-tagger can assign to a word in a sentence. The column on the left and right sides of the lines are the same and used to present the table into one image that can fit in on a single page. The use of these tags can be observed in Chapter-4 where based on these POS tags different rules are proposed for different features.

3.4 Evaluation Strategy

The solutions to the research objectives are (i) classification-based approach, (ii) pattern recognition, or template matching technique. Pattern matching techniques have wide application areas for last few decades from medical to businesses [109]

Tag	Description	Example	Tag	Description	Example
CC	coordin. conjunction	*and, but, or*	SYM	symbol	+,%, &
CD	cardinal number	*one, two*	TO	“to”	*to*
DT	determiner	*a, the*	UH	interjection	*ah, oops*
EX	existential ‘there’	*there*	VB	verb base form	*eat*
FW	foreign word	*mea culpa*	VBD	verb past tense	*ate*
IN	preposition/sub-conj	*of, in, by*	VBG	verb gerund	*eating*
JJ	adjective	*yellow*	VBN	verb past participle	*eaten*
JJR	adj., comparative	*bigger*	VBP	verb non-3sg pres	*eat*
JJS	adj., superlative	*wildest*	VBZ	verb 3sg pres	*eats*
LS	list item marker	*1, 2, One*	WDT	wh-determiner	*which, that*
MD	modal	*can, should*	WP	wh-pronoun	*what, who*
NN	noun, sing. or mass	*llama*	WP$	possessive wh-	*whose*
NNS	noun, plural	*llamas*	WRB	wh-adverb	*how, where*
NNP	proper noun, sing.	*IBM*	$	dollar sign	*$*
NNPS	proper noun, plural	*Carolinas*	#	pound sign	*#*
PDT	predeterminer	*all, both*	“	left quote	‘ or “
POS	possessive ending	*’s*	”	right quote	’ or ”
PRP	personal pronoun	*I, you, he*	(	left parenthesis	[, (, {, <
PRP$	possessive pronoun	*your, one’s*	)	right parenthesis	],), }, >
RB	adverb	*quickly, never*	,	comma	,
RBR	adverb, comparative	*faster*	.	sentence-final punc	. ! ?
RBS	adverb, superlative	*fastest*	:	mid-sentence punc	: ; ... – -
RP	particle	*up, off*			

Figure 3.3: Penn Treebank POS

[110] [111]. A comprehensive review of pattern recognition categorizes these techniques into six categories namely; Structural techniques, Statistical techniques, Neural Network Approach, Template Matching, Fuzzy Model, and Hybrid Models [112]. For noisy and unstructured data, statistical model-based techniques perform best. For recognition of the patterns and their relationships among each other from the text structure, is obtained by using the structural models [112]. The performance of the statistical models depends upon the feature selection, and statistical methods generate ambiguous results due to the feature extraction approaches. The structural models' based approaches explore the structure of the sentence and attempt to recognize the general patterns. The author [112] discussed those structural approaches that need large training datasets; furthermore, they do not work effectively on the dataset that does not contain any structure

such as the tweets available on Twitter. The study by Seema and Rajeshwar [106], recommends the use of a statistical model for complex patterns. Both structural and statistical models have their pros and cons and it is evident from the literature [112] that the results of these models depend on how good are the extracted features for pattern recognition in the statistical approach.

Our task in this research is to extract the feature-words, and for feature extraction, the language processing techniques are applied. Based on task requirements and the findings discussed above, we are motivated to adopt a grammar rule-based approach for identification and dissemination of the feature-words to its identified features of the manual features-based approach [37] for identification of eyewitness.

The reports posted on twitter can be classified into three classes of eyewitness reports.

1. **”Dont know/Unknown” cases and noise:** In a tweet, a user may use disaster-related keywords as metaphors, for example as press cuttings, naughty headlines. These messages were categorized as noise or “unknown” class. In our dataset, the tweets that were too ambiguous were classified as “unknown”.

2. **Eyewitness reports:** A tweet that contains first-hand information about the event is considered as an eyewitness report. Information in the tweet can be helpful for response teams in emergency conditions.

3. **Non-eyewitness reports:** The tweets with no explicit eyewitness characteristics were categorized as non-eyewitness reports.

To implement the strategy adopted for this research, we critically analyzed the training dataset of 2000 tweets for three disaster types; earthquake, flood, and hurricane, for identification of patterns or templates that can potentially identify the feature-words. The discussion sessions were conducted with language experts for their useful inputs on language analysis, structural rules, and linguistics, that were potentially helpful in the identification of feature-words.

The training dataset of tweets contains free-vocabulary and informally written text and results in details are not explained by the authors of the manual features-based approach [37]. We used the same dataset as a training dataset for the LR-TED as was used as a training dataset by the state-of-the-art ([37]) with which we will be comparing our results. Based on detailed and critical analysis of language structure, linguistics features, and exiting relational dependencies of words with each other, we proposed the grammar rules for the extraction of feature-words for all identified features of the manual features-based approach [37].

Each tweet of the dataset is annotated and the results are saved in a separate dataset, for detailed study and evaluation. From a comprehensive study of the annotated dataset of tweets, it was identified that the structure of the language and the dependency relationship of words can be exploited, to automatically extract the feature-words from the content of the tweet. We have proposed language structure-based grammar rules for every identified feature, where the state-of-the-art adopted the approach of manually created static dictionary for feature extraction. The grammar-rule for each identified feature has been briefly discussed in Chapter-4 with examples.

The defined rules are applied to each annotated tweet from the dataset, which is saved separately from the benchmark dataset. The state-of-the-art [37] has not discussed any approach or methodology that explains the process of marking a tweet, with identified feature-word, as an eyewitness, non-eyewitness, or unknown. For example, if a tweet contains a feature word that explains the surrounding details, and no other feature is identified even with the human eye, then what is the threshold to accept that tweet as an eyewitness? To answer this problem we conducted three different experiments, namely "Manual Classification", "Supervised Learning Models", and "Deep Learning Approach", for the evaluation of our LR-TED approach. The strategy for each experiment is explained in the following sub-sections, and the results generated through these experiments are described in Chapter-5.

3.4.1 Experiment-1: Manual Classification

To classify the tweet into defined categories of “eyewitness”, “non-eyewitness” and “unknown”, we manually defined some parameters as bellow;

1. The tweet containing no feature-word is marked as “Unknown”. We have discussed this class in detail at the start of this section.

2. The tweet containing only one feature-word is marked as “Non-eyewitness”. A tweet may contain a location name in the tweet with the disaster query word, but there it can be a reference or a URL. For example, *"The 2005 Earthquake in Pakistan took the lives of thousands, and many cities were abandoned"*. In this text we the query word and feature from the identified list which is the location name Pakistan. But this report can not be classified as an eyewitness report.

3. The tweet containing two or more feature-words is marked as an “Eyewitness” tweet. We have an exceptional case for the feature of “Exclamation and question marks”. If we have a question mark or an exclamation mark with the disaster word, the tweet will be considered as “eyewitness”. For example, *"Earthquake!"*.

We have set the high standards for our approach, as it is observed by manual analysis of the tweets that the annotated results of the gold-standard dataset have tweets that are marked as “eyewitness” based on the single identified feature-word.

3.4.2 Experiment-2: Supervised Learning Models

In our second experiment, we decided to expand our experiment-1 to the next level by exploiting the result using various evaluation strategies used in the literature. We have critically analyzed, what model they have used and what are the configuration and evaluation parameters. After comprehensive study we have identified that “Random Forest”, “Support Vector Machine”, and “Naive Bayes” are the top machine learning models belong to the classification category and commonly used by researchers in this domain.

The manual features-based approach adopted the Random Forest model of classification. We employed the same model and settings as adopted by the state-of-the-art [37], which is the Random Forest. For textual data based classification, Random Forest is considered as the best choice [113]. The Random Forest model of classification is trained on tweets dataset for each disaster type; earthquake, flood, hurricane, and wildfire. The 10-fold cross-validation technique with standard input parameters is used to evaluate the model's performance. The evaluation parameters, as used by the state-of-the-art are precision, recall, and F-score. The results of these parameters are calculated for the evaluation and comparison of results and briefly discussed in Section-5.3.2.

The results of all three, “Random Forest”, “Support Vector Machine”, and “Naive Bayes” are discussed in detail in Section-5.3.2.

3.4.3 Experiment-3: Deep Learning Approaches

After evaluating the supervised learning models in our third experiment, we decided to further investigate the deep learning algorithm as experiment number three in this research and compare its results with the results achieved by all previous experiments. A recent survey has comprehensively studied the deep learning models adopted in this domain for NLP applications. For the classification of textual data or documents, three models are used in deep learning that are Artificial Neural Network (ANN), Recurrent Neural Network (RNN), and Convolutional Neural Networks (CNN) [114]. To study the deep learning-based classification approach in our research, we adopted these three models in our third experiment.

ANN systems are computing systems that work on nonlinear data like a neural network of animal brains [115]. The results from the ANN approach are calculated, evaluated, and comprehensively discussed in Section-5.3.3 for each disaster type.

RNN is also a neural network architecture based model used for classification and text mining by assigning weights to the data points [114]. LSTM is the implementation of RNN model that we have adopted in this research for experiment number three. Detailed results are discussed in Section-5.3.3.

Convolutional Neural Networks (CNN) is another type of neural network based model, initially designed for image processing and effectively adopted for text classification [114]. CNN use convolution layers to generate the results. We have evaluated our approach on CNN and the detailed results are discussed in Section-5.3.3.

In the domain of eyewitness identification, feature extraction is performed manually as done by the recent state-of-the-art approach, there are no automatic approaches exist, or no benchmark dataset are available for comparison of our results. The gold-standard dataset of 8000 tweets, used by stat-of-the-art [37] contains the annotated results that are utilized as a benchmark for comparison of results.

The results are saved in the dataset for comparison with annotated results of the gold-standard dataset. The results and comparison with state-of-the-art are briefly discussed in Chapter-5.

3.5 Evaluation Parameters

From the comprehensive study of the existing approaches it is identified that, for maximum number of approaches, the researchers has adopted precision, recall, and f-measure as evaluation parameters for the evaluation of the approach. To evaluate the effectiveness of LR-TED, we adopted the precision, recall, and f-measure. The state-of-the-art approach also exploited these evaluation measures. Therefore, we also selected these measures of evaluation.

- Precision

  - Precision is normally described as "the fraction of relevant retrieved against the total retrieved". To calculate the value of precision, formula in this study would be "the fraction of the number of truly identified eyewitness tweets against the total number of tweets retried" using the below given formula.

$$Precision = RelevantRetrieved/TotalRetrieved \tag{3.1}$$

- Recall
  - The recall is generally described as "the fraction of relevant retrieved against the total relevant". To calculate the Recall value, the formula could be "the fraction of the number of truly identified eyewitness tweets against the total number of relevant tweets" using the formula given bellow.

$$Recall = RelevantRetrieved/TotalActualRelevant \tag{3.2}$$

- F-measure/F1
  - The F-measure is defined as the "Harmonic Mean of Precision and Recall". F-measure is also named as "F-Score" or "F1". Using the subsequent formula in equation number-3.3.

$$F1 = (2 * (Precision * Recall))/(Precision + Recall) \tag{3.3}$$

- Effort Estimation
  - The effort estimation is defined and adopted to evaluate the effectiveness of Proposed and State-of-the-art approaches. The efforts are estimated in-terms of "Time", "Human Resource", and "Cost".

The LR-TED approach is evaluated on the benchmark dataset, the results are then compared with the manual features-based approach [37], using the evaluation parameters of precision, recall, and f-measure.

3.6 Summary

In this chapter, we have tried to achieve the second part of the first research objective, *"Propose a generic approach that can automatically identify features from the content of any disaster type by discouraging the use of a static dictionary of each disaster type."* , discussed in Section-1.6. We critically analyzed the literature and the available approaches in this domain of feature identification, like statistical and structural model-based approaches. After a comprehensive study of the existing approaches, it was identified that their implication for this research work is not feasible. Based on these findings we adopted the technique of using language structure, linguistic rules, language constructs, and words dependency relationship for identification of feature-words. We created grammar rules that can identify feature-words from the tweet content without using any static dictionary as used by the recent state-of-the-art approach. Usage of grammar rules rather than using a manually created list of keywords for a specific domain makes this approach a generic approach that can be implemented for various disaster event types. The approach is trained on a dataset of 6000 tweets. The LR-TED is not dependent on a dataset of specific disaster types and can be adapted for different disaster events. The results and comparison of LR-TED approach are evident in the next chapters when it is applied to the testing dataset of different disaster types like hurricane, earthquake, flood, and fire events, for all categories like an eyewitness, non-eyewitness, and unknown category, on randomly selected 8000 tweets.

Chapter 4

Grammar Rule-Based Feature Extraction

In the previous chapters, we have introduced the idea of defining linguistic-based grammar rules to automatically extract the feature-words, and reviewed the state-of-the-art techniques and approaches for identification of eyewitness and extraction of feature-words in detail and evaluated the existing systems. Recent approaches for identification of eyewitness adopted the technique of features identification and comprehensively discussed and exploited by researchers of this domain [37], [42], [41]. Based on the idea of identifying eyewitness by features from content, a recent approach proposed a diversified set of features for the identification of eyewitness from the text. The manual features-based approach [37] adopted the approach of static dictionary for relevant feature-words of a disaster event and the same are matched for identification of feature-words. This use of manually created static dictionaries is not scale-able for new disaster event types and involvement of domain-experts for identification of new feature-words is not possible to process millions of disaster tweets in real-time and the few identified features were not implemented by the manual features-based approach due to their implementation complexity. The author of the the state-of-the-art approach [37] stated that these characteristics are not implemented because they "proved too abstract to operationalize and were not implemented". Learning and applying the concept of

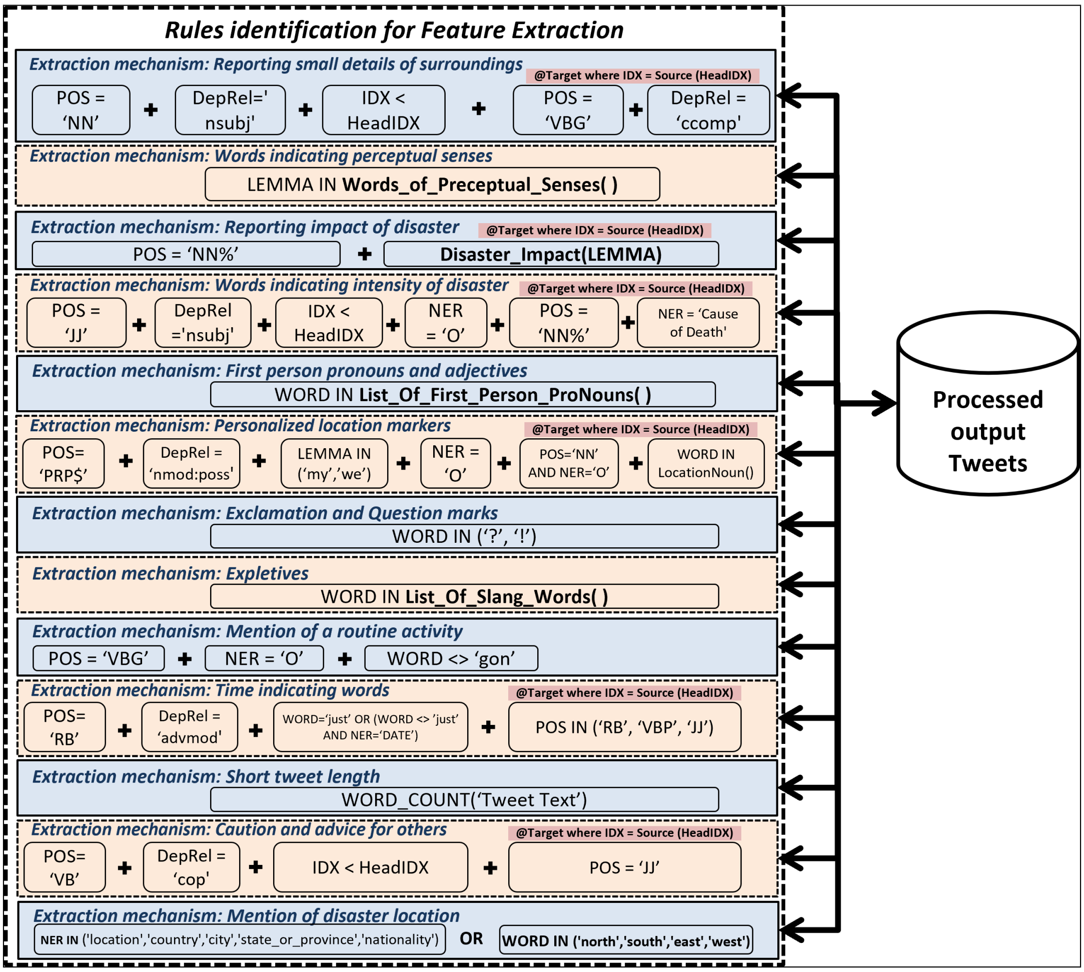

FIGURE 4.1: Feature Extraction Rules

automatically extracting the feature-words using grammar rules for identification of eyewitness reports, without human interaction by the proposed technique is a novel contribution of this research.

Figure-4.1 briefly describe the module of "Rules identification for feature extraction" from Figure-3.1. The figure clearly illustrate the extraction mechanism required for automatic extraction of the feature-words from the contents of the tweets. The processed-output tweets dataset in the Figure-4.1 is the dataset that contains the parsed tweets by CoreNLP tool, having values for all fields discussed in Section-3.3. The parsed dataset is available for all 13 features, to be executed in parallel or one after another. The generated rules will be applied to the dataset and the results are saved into the same dataset for further evaluation and comparison of the results. Figure-4.1 in sub-part of Figure-3.1.

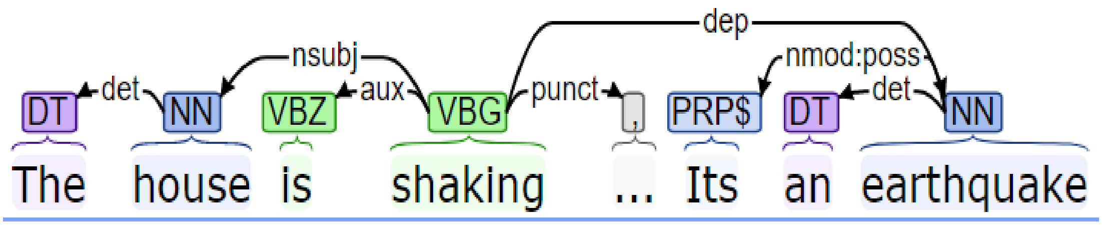

FIGURE 4.2: Word Dependencies

4.1 Linguistic Features and Words Relationship

In the literature review, we have discussed in detail the approaches adopted by the researchers for identification of eyewitness using content such as linguistic feature set, stylistics features, and we identified that the structure of the language and relationship of words in a sentence is not studied in this domain. From an in-depth study, we found that the language structure, linguistics rules, and the relationship among the words, in a sentence, can be explored to automatically extract the feature words from tweet's content by defining the grammar rules.

To explain the concept let's see the following example where we have tweet text "The house is shaking...Its an earthquake". Figure-4.2, explains all the relationship among different words or the basic dependencies.

From Figure-4.2 we see that each word has a dependency relation with other words to explain its meaning in the sentence. In this example the noun "house" alone does not explains its existence until the verb "shaking" is not related to it and it became "house shaking" that explains the details of surroundings.

4.2 Parsing of Tweets (Stanford CoreNLP)

The grammatical tagging or the Part-Of-Speech (POS) tagging, is the tagging technique where each word in a sentence is tagged, based on both, its definition and its context, with its corresponding particular part of speech. The NER (Named Entity Recognition) tool by Stanford has been proven its usefulness by achieving

Table 4.1: Stanford CoreNLP (Standard Output)

IDX	Word	Lemma	POS	NER	HeadIDX	DepRel
1	The	the	DT	O	2	det
2	house	house	NN	O	4	nsubj
3	is	be	VBZ	O	4	aux
4	shaking	shake	VBG	O	0	ROOT
5	...	...	:	O	4	punct
6	Its	its	PRP$	O	8	nmod:poss
7	an	a	DT	O	8	det
8	earthquake	earthquake	NN	CAUSE OF DEATH	6	ccomp

high F-measure scores in modern-day approaches [105] [106]. Thus, we have also adopted the Stanford CoreNLP tool for our approach.

In this research, we adopted a Java suite of Core NLP tools (see http://stanfordnlp.github.io/CoreNLP/). The annotated results of the tweets are stored in the same database. The database contains all the tweets to be parsed data in order to evaluate and carryout the comparison. For example, we have the tweet text as explained above; *"The house is shaking...Its an earthquake"*. The CoreNLP output in CoNLL format[1], the standard output of Stanford CoreNLP is shown in Table-4.1.

In Table-4.1, the first column named "IDX" is the Indexed ID of the word in the sentence. Column-2 contains the actual Word field that contains the word extracted directly from the text. The third column titled "Lemma" has the base word. Column-4 "POS" holds the corresponding fine-grained Part-of-speech (POS) tag of the token word, that are defined using Penn-Treebank-Tagset (see https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html). Column-5 is "NER", that contains the Named-Entity tag. Columns 6 and 7, titled "HeadIDX" and "DepRel". Column-6 of "HeadIDX" holds the Index ID of the word in relation or zero if no relation, and column 7 contains the dependency relation type, the details are available on https://nlp.stanford.edu/software/dependencies_manual.pdf.

[1] https://nlp.stanford.edu/nlp/javadoc/javanlp-3.5.0/edu/stanford/nlp/pipeline/CoNLLOutputter.html

4.3 Feature Extraction (Grammar Rule-Based)

A comprehensive study and critical analysis of the tagged dataset of tweets, we have recognized that the linguistic features, structure of the language, and the association among the words in a sentence could be exploited for automatic extraction of the feature-words from content of the tweet. Unlike the manually created static dictionaries or lists as used by the manual features-based approach [37], in this research we have presented grammar rules for all thirteen features. Additionally, in our approach we have created rules for those identified features, that were not implemented by the manual features-based approach due to the complexities lies in implementation.

This research has presented the grammar-rules for maximum number of the identified features. During the critical analysis, it is identified that some identified features such as "First person pronouns and adjectives", "Words indicating perceptual senses", "Expletives", and "Exclamation and question marks" are in some way restricted to a pre-defined list of words. For example, for the identified feature such as; "Exclamation and question marks" we only have to look for "?" and "!", and there exists no other character on the list. So for its implementation, we cannot devise the grammar rules however use a static list of two characters (?, !) and search if they exist in the content. Similarly, for the identified feature of expletive words, no grammar rules are need to be created from language structure due to its scope. Again we have to limit the implementation to a static dictionary. Another feature that does not require a grammar rule is "Short length of tweet". To implement this feature we only have to get the word count of the tweet content.

In this research, we have fourteen features, thirteen identified features by the manual features-based approach [37] and one new proposed feature. From this list, we have to adopt the static dictionary based implementation technique for four features namely; "First person pronouns and adjectives", "Words indicating perceptual senses", "Expletives", and "Exclamation and question marks". A straightforward approach of word count for one identified feature of "Short length

Table 4.2: Stanford CoreNLP (Standard Output), Example

IDX	Word	Lemma	POS	NER	HeadIDX	DepRel
1	The	the	DT	O	2	det
2	house	house	NN	O	4	nsubj
3	is	be	VBZ	O	4	aux
4	shaking	shake	VBG	O	0	ROOT
5	...	...	:	O	4	punct
6	Its	its	PRP$	O	8	nmod:poss
7	an	a	DT	O	8	det
8	earthquake	earthquake	NN	CAUSE OF DEATH	6	ccomp

of tweet". With these findings, we are left with eight identified features, and for those, we have proposed the grammar rules, as explained in the following sub-sections.

In the following sub-sections, we have briefly discussed all identified feature either with its proposed Grammar Rule or by using the count formula and predefined lists.

4.3.1 Feature-1: "Reporting small details of surroundings"

The first feature in the list of features identified by the manual features-based approach [37] is "small surrounding details". The feature was dropped by state-of-the-art approach [37] during the implementation phase, and stated that "proved too abstract to be implemented". From the perspective of human reading, its an important feature, and to completely understand the tweet's context and intentions of the user. The significance of such a feature is also discussed by Fang et al. [41]. To automatically extract the feature-words of such type, reported by Twitter users that are potentially explaining the surrounding details, we have to define grammar rules.

The working of the extraction process using the grammar rule is briefly discussed using the Table -4.2. The same data, as described in Table-4.1, is used to demonstrate the working.

The tagged word's relationship with other words in a sentence is shown in the Table-4.2. From the tweet content, "house shaking" is the candidate feature-word that is explaining the surrounding details. It is impossible to extract this feature-word by using a bag-of-words technique with uni-gram and bi-gram, as adopted by the manual features-based approach.

The working of feature-word extraction process, using the proposed grammar rule, is briefly explained in this sub-section. The process remains the same for all features with their respective grammar rule. To clearly understand the extraction process, we have adopted some terminologies like; the term "token" is used to identify the a row in the table, used the term "OnTarget()" which means the rule within the braces applies on target token.

Using the Table-4.2, we have demonstrated the successful extraction of the feature-word, by applying the proposed grammar rule. From Table-4.2, the token number three as (IDX = 2), have a noun value for POS(POS = "NN") and its dependency relationship has type subjective as DepRel (DepRel = "nsubj") with token number eight (IDX = 4) linked by using HeadIDX value (HeadIDX = 4). On the target side, the token has the value of verb as POS (POS = "VBG") and it fulfill the required dependency relational condition (DepRel = "nsubj"). By applying this grammar rule we have successfully identified the feature-word of "house shaking" from the tweet content. The proposed grammar rule for feature number one is as follows;

```
RULE-1;

   POS in ('NN','NNS') and DepRel in ('nsubj','dobj')
   and IDX < HeadIDX and
   NER<>('CAUSE_OF_DEATH','URL','NUMBER','TIME','MONEY') and
   OnTarget(POS in ('VBG','VBD') and DepRel in ('ccomp','dobj'))
```

By defining the grammar rule for the identification of feature-words that are "reporting small details of surroundings", we achieved our second research object discussed in Section-1.6.

4.3.2 Feature-2: "Words indicating perceptual senses"

The words, pertaining to this feature, are connected to a limited list such as; "hearing, seeing". To implement the proposed grammar rule we adopted the technique of matching the "Lemma" field of each token of a tweet, with a static dictionary or list of related lemma-based words for "hearing, seeing". A such predefined list of related words is online available at http://liwc.wpengine.com/. The manual features-based approach adopted the same list for the implementation of this feature. The proposed grammar rule for this feature is not fully a grammar rule but an hybrid approach used by combining the Bag-of-Words technique with a small grammatical condition as follows;

```
RULE-2;
    POS like 'VB%' and
    LEMMA in PreceptualSensesWordList()
```

Here we need to clarify that the "%" symbol used with "VB" (verb), denotes that it can be any form of the verb, verb past tense, verb gerund, verb past participle, as described in Figure-3.3. The "%" symbol is used in Microsoft SQL Server query formation to denote anything in-place of "%". So "VB%" means anything like "VB", "VBD", "VBG", "VBN", and "VBP".

4.3.3 Feature-3: "Reporting impact of disaster"

In the event of disaster or emergency, a common practice of twitter users is to report the impact of the disaster event. Such useful information shared by users helps evaluate the impact size and severity. Example of feature-words of this category contains the terms; "school canceled, flight delayed". To automatically extract such terms from the content, we have proposed to implement the grammar rule in combination with the list of impact-words such as; "Cancel, Delay, Suspen, lost, postpone, defer, reschedule, rearrange". The list of impact-words is not predefined but generated from a common list of feature-words falling in this category. The implementation of proposed grammar rule is formulated as bellow;

```
RULE-3;

    POS like 'NN%' and NER<>'URL' and
    OnTarget(NER<>'URL' and DisasterImpactWord(LEMMA))
```

Here POS can be a "NN" (noun) or any form of the noun, plural noun, proper noun singular or plural, as described in Figure-3.3. So "NN%" means anything after "NN" like "NNS", "NNP", "NNPS".

To demonstrate the implementation of the proposed rule let's see the parsed tweet in Table-4.3.

In Table-4.3 it is evident that at token 15 (IDX = 15), a noun ("NN") and it has a relationship dependency (DepRel="dobj") with token 13 (IDX = 13). By applying this grammar rule we have extracted the feature-word "flights cancelling", which is reporting the impact of the disaster. The base words can also be utilized as extracted feature words like in this case it will be "flight cancel".

4.3.4 Feature-4: "Words indicating the intensity of the disaster"

The words, explaining the intensity and severity of the catastrophic event, such as; strong, big, dangerous, and intense, lies in this category of identified features. Identification of such words in a disaster event-related tweet could be helpful to

Table 4.3: Feature-3: Stanford CoreNLP Example

IDX	Word	Lemma	POS	NER	HeadIDX	DepRel
1	Thanks	thanks	NNS	O	0	ROOT
2	to	to	TO	O	4	case
3	Cyclone	Cyclone	NNP	O	4	compound
4	Debbie	Debbie	NNP	O	1	nmod
5	for	for	IN	O	8	mark
6	not	not	RB	O	7	neg
7	only	only	RB	O	8	cc:preconj
8	destroying	destroy	VBG	O	1	acl
9	my	my	PRP$	O	10	nmod:poss
10	hometown	hometown	NN	O	8	dobj
11	but	but	CC	O	8	cc
12	also	also	RB	O	13	advmod
13	cancelling	cancel	VBG	O	8	conj
14	all	all	DT	O	15	det
15	flights	flight	NNS	O	13	dobj
16	making	make	VBG	O	15	acl
17	me	I	PRP	O	18	nsubj
18	miss	miss	VB	O	16	ccomp
19	the	the	DT	O	21	det
20	dixiechicks	dixiechick	NNS	O	21	compound
21	concert	concert	NN	O	18	dobj
22	!	!	.	O	1	punct

the disaster response agencies. The decision making authorities need such details to respond/react accordingly, in such situations. To implement the proposed grammar rule, we have formulated the grammar rule as bellow;

```
RULE-4;

   POS = 'JJ' and IDX < HeadIDX
   and NER='O' and
   OnTarget (POS is like 'NN%' and
           NER = 'CAUSE_OF_DEATH')
```

The proposed rule is applied to a tagged tweet in Table-4.4, which also explains how the relationship of words is exploited for the identification of this feature.

From Table-4.4 we can see an adjective (JJ) at row number 3 and the target row number is 5 that contains the noun (NN) word “earthquake” which is categorized as *"CAUSE_OF_DEATH"* in Named Entry Recognition (NER) tagging category.

TABLE 4.4: Feature-4: Stanford CoreNLP, Example

IDX	Word	Lemma	POS	NER	HeadIDX	DepRel
1	Um	um	NN	O	5	compound
2	pretty	pretty	RB	O	3	advmod
3	strong	strong	JJ	O	5	amod
4	#	#	#	O	5	dep
5	earthquake	earthquake	NN	CAUSE OF DEATH	0	ROOT
6	just	just	RB	O	5	dep
7	now	now	RB	DATE	5	dep
8	#	#	#	O	5	dep
9	BayArea	BayArea	NNP	O	5	dep
10	??	??	CD	NUMBER	9	nummod

So using this rule we can identify the disaster intensity feature-word "strong earthquake".

4.3.5 Feature-5: "First person pronouns and adjectives"

The "first person" as a term can refer to the speaker himself or a group of people that includes the speaker (i.e.," I," "me," "we," and "us"). This category of feature has a static list of words to follow, and the implementation is pretty upfront. The simple technique of looking into a predefined list is adopted by the manual features-based approach [37]. In this research, we have proposed to adopt the same approach for its implementation. We have formulated the proposed function as bellow;

```
RULE-5;

    LEMMA in FirstPersonPNouns()
```

4.3.6 Feature-6: "Personalized location markers"

This is another important feature that is explaining the intensity and severity of the catastrophic event by extracting the user's personalized location indicators. This category contains feature-words such as; my home, our area, my office, and our city. The feature was dropped by state-of-the-art approach [37] during the implementation phase, and stated that "overlapping with feature number five". Here we negate this author's statement from the manual features-based approach, while they are not truly overlapping. We analyzed that such feature-words contain pronoun based word as its part, like "My School". The real difficulty in developing the grammar rule for this feature is to differentiate the location nouns from other nouns identified in the tweet content. To address this situation we adopted a list of location nouns, available freely at https://www.wordhippo.com/what-is/the-noun-for/locate.html. The formed grammar rule for extraction of this feature is formulated as bellow;

```
RULE-6;

    POS='PRP$' and IDX < HeadIDX
    and NER='O' and LEMMA in ('my','we')
    and DepRel = 'nmod:poss' and
    OnTarget(LocativeNoun(WORD)
     and NER='O' and POS='NN')
```

Different languages and regions have different language structures and the features that depend on possessive pronouns need adjustment for implementation purposes [37]. That's the reason for adjusting these personalized location markers by using a manual list of locations. The list of location nouns includes locations such as area, backyard, apartment, porch, roof, bedroom, boutique, building, company, and workplace. The demonstration of the rule is shown from Table-4.5.

By defining the grammar rule for the identification of feature-words that indicates the "personalized location markers", we achieved our second research object discussed in Section-1.6.

From Table-4.5 we can see a possessive pronoun (PRP$) at row number 13 with required relational dependency (DepRel) value of "nmod:poss" that indicates the possession modifier relation at and the target, that is row number 15 that contains

Table 4.5: Feature-6: Stanford CoreNLP, Example

IDX	Word	Lemma	POS	NER	HeadIDX	DepRel
1	Either	either	CC	ORGANIZATION	4	dep
2	Berkeley	Berkeley	NNP	ORGANIZATION	4	nsubj
3	just	just	RB	O	4	advmod
4	had	have	VBD	O	0	ROOT
5	a	a	DT	O	7	det
6	minor	minor	JJ	O	7	amod
7	earthquake	earthquake	NN	CAUSE OF DEATH	4	dobj
8	or	or	CC	O	7	cc
9	a	a	DT	O	10	det
10	giant	giant	JJ	O	7	conj
11	bumped	bump	VBN	O	10	acl
12	into	into	IN	O	15	case
13	my	my	PRP$	O	15	nmod:poss
14	whole	whole	JJ	O	15	amod
15	house	house	NN	O	11	nmod

the location noun (NN) word "house". So by applying this rule we extracted the personalized location markers feature-word "my house".

4.3.7 Feature-7: "Exclamation and question marks"

This feature identified in the manual features-based approach can be easily acquired using a straightforward known list of characters, that includes "!" and "?". The grammar rule for the extraction of this feature is formulated as follows:

```
RULE-7;

  WORD like ('%?%','%!%')
```

If the text of the tweet contains these characters then the tweet is marked as true for feature number 7 that is the presence of an exclamation or a question mark. The role of this identified feature is evident from the tweets like "earthquake?", that implicitly explain the feeling of a person who just felt an earthquake and wanted to confirm from others.

4.3.8 Feature-8: "Expletives"

The expletives are the words or phrases that are not required to explain the basic meaning of the sentence are called the expletives e.g. omg, wtf, s**t. To implement this identified feature, the Wiktionary & Slangs List, as adopted by state-of-the-art [37], are employed in this research. Such types of words, at this time, are not possible to identify by using the linguistic rule. The static list of expletive words used in this research work is available online at https://www.speakconfidentenglish.com/english-internet-slang/.

```
RULE-8;

  WORD in SlangWordsList()
```

The collection of slang words varies from one region to another even within a country. So the commonly available list of expletive words that are utilized by researchers in literature.

4.3.9 Feature-9: "Mention of a routine activity"

This identified feature contains such words which explains the daily routine activities such as; running, watching a movie, sleeping. The manual features-based approach implemented the proposed feature by adopting the list of words available online at https://www.vocabulary.cl/Lists/Daily_Routines.html. In this research, we critically analyzed the data and proposed the following grammar rule;

```
RULE-9;

  POS = 'VBG'
  and NER = 'O'
  and WORD <> 'GON'
```

Table 4.6: Feature-9: Stanford CoreNLP, Example

IDX	Word	Lemma	POS	NER	HeadIDX	DepRel
1	I	I	PRP	O	4	nsubjpass
2	am	be	VBP	O	4	auxpass
3	literally	literally	RB	O	4	advmod
4	stuck	stick	VBN	O	0	ROOT
5	in	in	IN	O	9	case
6	chase	chase	NN	O	9	nmod:poss
7	's	's	POS	O	6	case
8	apartment	apartment	NN	O	9	compound
9	complex	complex	NN	O	4	nmod
10	and	and	CC	O	4	cc
11	watching	watch	VBG	O	4	conj
12	my	my	PRP$	O	15	nmod:poss
13	car	car	NN	O	15	compound
14	flood	flood	NN	CAUSE OF DEATH	15	compound
15	url	url	NN	URL	11	dobj

By using this grammar rule we can identify the daily routine activities directly from the text. A demonstration of one such tweet is shown in Table-4.6.

From Table-4.6, we can see in row number 11 we have a routine activity of "watching" that is retrieved by word matching, but it is retrieved because it's a verb gerund (VBG) that it exhibits ordinary verbal properties and word has no NER value as proposed in the rule. So by implementing the proposed rule we can extract the word "watching" as a feature word.

4.3.10 Feature-10: "Time indicating words"

The "Time indicating words" are considered as an essential feature to comprehend and respond accordingly in the event of a disaster or emergency situation e.g. "just", "at the moment", "now". To implement the proposed grammar rule of the feature, we have formulated the proposed grammar rule as bellow;

```
RULE-10;

  POS = 'RB' and DepRel = 'advmod'
  and lemma not in ('!!','|','a.') and
```

TABLE 4.7: Feature-10: Stanford CoreNLP, Example

IDX	Word	Lemma	POS	NER	HeadIDX	DepRel
1	Good	good	JJ	O	4	amod
2	little	little	JJ	O	4	amod
3	#	#	#	O	4	dep
4	earthquake	earthquake	NN	CAUSE OF DEATH	0	ROOT
5	just	just	RB	O	6	advmod
6	now	now	RB	DATE	9	advmod
7	in	in	IN	O	9	case
8	#	#	#	O	9	dep
9	SanFrancisco	SanFrancisco	NNP	LOCATION	4	acl:relcl

```
((WORD = 'just') or (WORD <> 'just'
 and NER = 'DATE'))
and OnTarget(POS is in ('RB', 'CD', 'VBP', 'JJ'))
```

The above-defined rule looks for an adverb (RB) word that has a dependency relation value of an adverb modifier (DepRel = 'advmod') with some limitation to avoid the noise in selection. On the target side, the word should be POS tagged with any of the values either an adverb (RB), a verb (VBP), or an adjective (JJ). Table-4.7 demonstrates the working of the rule from the content of a real tweet.

The words in row number 5 and 6 of Table-4.7 demonstrate the implication of defined rule. By using this rule we have extracted the feature-word "just now" which the time indicating word.

4.3.11 Feature-11: "Short tweet length"

The identification of this feature is very straightforward and its role dependents on other features to further explain the real concept of its presence. For example, if a tweet has only a single word "earthquake", is a valid response by search query as the content of the tweet has the query word. The feature of short-length tweets, identified by the manual features-based approach [37] in the current scenario, the existence of the feature is true, but it does not express the true meaning. Although the tweet with the content "earthquake!" changes the meaning and context of the tweet. The manual features-based approach has not discussed any standard for the number of words count for this feature or from the literature therefore the implementation of this feature is complex. We set the tweet length to be below nine words for implementation purposes. This, the value of word count, remains an open research area as future work in this domain. To implement this feature in our words we only need a word count which is formulated as follows;

```
RULE-11;

  NER<>'URL' and
  TweetWordsCount([Tweet-Content])
```

4.3.12 Feature-12: "Caution and advice for others"

The identified feature contains the words such as; "be careful", "watch out". Such words are used to warn others to be aware of an event. It is implemented by using static dictionaries by the manual features-based approach [37]. We have presented the language-based grammar rules for the extraction of such words from the tweet content. To implement the purpose feature, we have formulated the proposed grammar rule as bellow;

```
RULE-12;

  POS = 'VB' and IDX < HeadIDX
  and DepRel = 'cop'
  and OnTarget(POS = 'JJ')
```

In daily routines, people become attentive when they are alerted about any event. The same idea is adopted by the manual features-based approach as a feature to intimate the reader about an event. The demonstration of the feature is given in Table-4.8.

Table 4.8: Feature-12: Stanford CoreNLP, Example

IDX	**Word**	**Lemma**	**POS**	**NER**	**HeadIDX**	**DepRel**
1	#	#	#	O	2	dep
2	Myriad2017	myriad2017	NN	O	8	nsubj
3	is	be	VBZ	O	8	cop
4	still	still	RB	O	8	advmod
5	on	on	IN	O	8	case
6	:	:	:	O	8	punct
7	be	be	VB	O	8	cop
8	careful	careful	JJ	O	0	ROOT
9	in	in	IN	O	10	mark
10	getting	get	VBG	O	8	advcl
11	here	here	RB	O	10	advmod
12	-LRB-	-lrb-	-LRB-	O	15	punct
13	following	follow	VBG	O	15	case
14	heavy	heavy	JJ	O	15	amod
15	traffic	traffic	NN	O	10	dep
16	-RRB-	-rrb-	-RRB-	O	15	punct
17	!	!	.	O	8	punct

From row numbers 7 and 8 of the Table-4.8, we can see the implication of the proposed rule and we have successfully extracted the term “be careful”. Similarly, we can extract the cautionary and advisory words from the content by applying this rule.

4.3.13 Feature-13: “Mention of disaster locations”

This identified feature by the manual features-based approach [37] was not implemented by the authors, due to its implementation complexity. Author stated that *”tweets are often too short and informal at times and location identification, as well as pronouns from social media data, is another aspect of extensive research”*. We exploited the features of the CoreNLP tool, to automatically extract the location, mentioned in the tweet text. To implement the purpose feature, we have formulated the proposed grammar rule as follows;

```
RULE-13;

  LocationType(NER)
  or WORD is in ('east', 'west', 'north', 'south')
```

The rule is very straightforward and needs no additional checks for its implementation. But to make sure that it covers aspects of location words an extension to the basic rule is done as shown in the above rule. CoreNLP covers the various NER values for implementation such as 'COUNTRY', 'CITY', 'LOCATION', 'STATE OR PROVINCE', 'NATIONALITY'.

We have achieved our second research object discussed in Section-1.6, by defining the grammar rule for the identification of feature-words that "mention of disaster locations".

4.4 New Feature (Proposed)

In the above sub-sections, we have proposed all thirteen-rules for identified features of the manual features-based approach. From the detailed study and literature review of the state-of-the-art approaches we have identified that the existence of URLs in the content or are considered as re-tweet or indirect message, Fang et al. [41]. The addition of URLs in tweet content is a type of linking the content to the previously published content and such information depicts that the information is not the first-hand information, Tanev et al. [42]. The idea leads us to a distinguishing feature where we check for the presence of any URL within the tweet content and we named the feature as "NO-URL" feature. This feature can either have "Yes" or "NO" as value for each tweet.

To implement this new proposed feature in our research we formulated the rule as follows;

```
RULE-14 (New Rule);

  IF (URL Found in content)
   THEN ''NO''
   ELSE ''YES''
  END
```

The implementation of the rule is very straightforward, and we have to look for the existence of URL in the tweet content. If the system found the URL in the tweet the value of the rule is set to "False" and if URL is not found the value is set to "True".

4.5 Summary

The proposed methodology is briefly explained in Chapter-3, and as mentioned in Methodology (Chapter-3) and Figure-3.1 contains a module of "Rules identification of feature-extraction" that was not discussed in the previous chapter. This chapter describes the rule identification module for feature extraction, where each of the fourteen features is discussed in detail with a comprehensive example and related discussion about the feature. The chapter discusses the Stanford CoreNLP tool which is a combination of multiple tools that generates results for all the required features as discussed in Chapter-3. The grammar rules are created for maximum features identified by the manual features-based approach [37], by automatically extracting the feature-words from tweet content and achieved our first research objective discussed in Section-1.6. The second research objective, to construct the language rules for those identified features which were dropped during implementation by the state-of-the-art approach because of to their implementation complexity with the methodology adopted by the manual features-based approach [37], from Section-1.6. It is also achieved by the defined grammar rules for feature namely 1-"Reporting small details of surroundings", 6-"Personalized location markers", and 13-"Mention of disaster locations", that were dropped by state-of-the-art approach [37] due to their implementation complexity. In addition to the thirteen features as identified in state-of-the-art, we in this research work has proposed a new feature to further improve the results.

Chapter 5

Experiments and Results

In this chapter, we present the overall evaluation results of LR-TED. In the following subsections, we have briefly explained the results generated at each step of our proposed LR-TED approach, described in Figures 3.1 and 4.1.

5.1 Dataset and Pre-Processing of Tweets

The LR-TED approach is assessed using a dataset of 8,000 tweets, containing the tweets with the query words of earthquake, hurricane, flood, and wildfire. Each tweet in the dataset also contains the corresponding annotated result provided by the manual features-based approach [37]. The statistics of each category; "Eyewitness", "unknown", and "Non-Eyewitness" are illustrated in the Table-5.1.

Table 5.1: Statistics of Evaluation Dataset

Category	**Earthquake Tweets**	**Flood Tweets**	**Hurricanes Tweet**	**Wildfire Tweets**	**Total**
Eyewitness	1600	627	465	189	2881
Non-Eyewitness	200	822	336	432	1790
Unknown	200	551	1199	1379	3329
Total	**2,000**	**2,000**	**2,000**	**2,000**	**8,000**

Since the selected dataset already have been evaluated and updated by the manual features-based approach [37]. The dataset contained few tweets which required the pre-processing steps. It was observed that, in 2942 tweets, their exist a " ' " or comma was found. Such tweets are updated in the phase of pre-processing. The dataset also contains the annotated results, calculated by the authors of the manual features-based approach, against each tweet. In this work, we used the same dataset as the benchmark to compare the results of LR-TED with the results of the manual features-based approach [37]. Since the dataset is shared by the author of the state-of-the-art approach [37] and is available in MS Excel file format, we proposed to adopt the simple "Replace" functions as shown in equation-5.1 & 5.2.

$$Replace('' \text{``}, '' \text{``}). \tag{5.1}$$

$$Replace('' \text{``}, '' \text{``}). \tag{5.2}$$

The formula shown in the equation-5.1, is executed first as it may add extra white-spaces in the content. After execution of replace function mentioned in equation-5.1, we executed the equation-5.2 for removal of extra white-spaces. The function shown in equation-5.2 is iteratively executed until all extra spaces were removed.

As discussed, that each tweet from the dataset also includes the annotations performed by the manual features-based approach [37]. The annotated data is used as a gold-standard dataset to compare the results of both, the LR-TED approach and the the manual features-based approach [37].

5.2 Parsing and Feature Extraction

The next step for implementation of the proposed methodology is, to process each tweet of the focused dataset, for Tokenization, Part-of-speech (POS) tagging, Named Entity Recognition (NER), Lemmatization, and Dependency Relation. The Stanfored CoreNLP tool is adopted for this task. Each tweet from the dataset of 8000 tweets, is processed using the the adopted tools, that finally generated the resultant 22,932 rows for all essential features as discussed in the Section-4.2.

The resultant data of the Stanford CoreNLP tool is critically analyzed to identify the language structure and patterns. The analysis revealed the importance of relationship among different words of a sentence. By studying the word relationships and patterns, the grammar rules were generated for extraction of feature-words, as comprehensively debated in the Section-4.3.

By implementing the proposed grammar rules, the feature-words for each category are extracted. In order to validate the correctness of the extracted feature-words it is identified that the manual features-based approach has not discussed or shared any such count for the identified features. In this work, we decided to manually validate the extracted feature-words from each tweet. To implement the idea, the dataset of extracted feature-words along with list of identified features and sample data was shared with an expert of "Natural Language Processing (NLP)", having brilliant language skills. The expert is a Ph.D. Scholar in semantic computing domain and has published various research articles in the domain of "information extraction" and "semantics analysis".

The methodology adopted for manually extracting the feature-words from each tweet, a brief session with the language expert was conducted. A clear understanding of the feature-words and their mapping to identified features used in the proposed idea was given to experts. The list features (see Table-1.1), with the state-of-the-art paper by the manual features-based. [37], and the testing dataset of 2000 tweets to be annotated was shared with the expert. The required output

TweetID	Tweet Text	Feature #1	Feature #2	Feature #3	Feature #4	Feature #5	Feature #6	Feature #7	Feature #8	Feature #9	Feature #10	Feature #11	Feature #12	Feature #13
2407	Anyone feel that little earthquake?		feel	little				?						
2423	I just felt a (very minor) earthquake in San Francisco.		felt	minor		i					just			San Francisco
2426	#earthquake bay area?							?						Bay area

FIGURE 5.1: Manual Feature Extraction (Example)

is discussed with examples and sample data. The output file format includes the structure as shown in Figure-5.1.

For example, we have short tweet text *I just felt a (very minor) earthquake in San Francisco.*, that has at-least five feature-words that indicate different features. They are I, just, felt, minor, and San Francisco, of First Person pronoun, Time indicating word, Perceptual senses word, words indicating intensity of disaster, and Mentioned Location features respectively from Table-1.1. As shown in the example the expert is requested to put relevant words from the text into the relevant field under the feature column. The feature of the short tweet was not involved in this exercise as its implementation is straight-forward and need no feature extraction.

The expert provided the results into the desired output (into an excel file) for further evaluation and matching with the features extracted by LR-TED approach are shown in Table-5.2.

The count of feature words, manually extracted by the language expert, against each identified feature, for 2,000 tweets are shown in Table-5.2. Then results by language experts are then compared with the results generated by LR-TED approach of grammar rules. Both results are then evaluated using evaluation parameters of precision, recall, and f-measures. The comparative results of manually extracted feature-words by the language expert and automatically extracted feature-words by LR-TED approach of grammar-rules are shown in Table-5.3.

From Table-5.3, it is evident that LR-TED approach of automatically extracting the feature-words using linguistic grammar rules proposed in Section-4.3 achieved

Table 5.2: Feature-Words Count (Manual Extraction by Language-Expert, State-of-the-art approach)

Feature No.	**Identified Feature**	**Automatically Extracted Words**	**Manually Extracted Words**
1	"Reporting small details of surroundings"	107	110
2	"Words indicating perceptual senses"	481	496
3	"Reporting impact of disaster"	2	3
4	"Words indicating intensity of disaster"	269	299
5	"First person pronouns and adjectives"	592	595
6	"Personalized location markers"	33	38
7	"Exclamation and question marks"	0	0
8	"Expletives"	438	450
9	"Mention of a routine activity"	234	297
10	"Time indicating words"	127	155
11	"Short tweet length"	0	0
12	"Caution and advice for others"	7	11
13	"Mention of disaster locations"	357	593
14	"NO-URL"	0	0
	Total	**2647**	**3047**

Table 5.3: Feature Words Extraction

	Automatically		**Manually**	**Evaluation**		
Feature No.	**Total Retrieved**	**Relevant Retrieved**	**Identified**	**Precision**	**Recall**	**F-Score**
Feature-1	108	107	110	0.99	0.97	0.98
Feature-2	512	481	496	0.94	0.97	0.95
Feature-3	2	2	3	1.00	0.67	0.80
Feature-4	292	269	299	0.92	0.90	0.91
Feature-5	794	592	595	0.75	0.99	0.85
Feature-6	33	33	38	1.00	0.87	0.93
Feature-7	0	0	0	0.00	0.00	0.00
Feature-8	484	438	450	0.90	0.97	0.94
Feature-9	275	234	297	0.85	0.79	0.82
Feature-10	134	127	155	0.95	0.82	0.88
Feature-11	0	0	0	0.00	0.00	0.00
Feature-12	7	7	11	1.00	0.64	0.78
Feature-13	690	357	593	0.52	0.60	0.56
Feature-14	0	0	0	0.00	0.00	0.00
Total	**3331**	**2647**	**3047**			

TABLE 5.4: Feature-Words Counts (Automatically Extracted by Grammar-Rules)

Feature #	Identified Feature	Feature Words
1	"Reporting small details of surroundings"	628
2	"Words indicating perceptual senses"	761
3	"Reporting impact of disaster"	96
4	"Words indicating intensity of disaster"	411
5	"First person pronouns and adjectives"	2044
6	"Personalized location markers"	118
7	"Exclamation and question marks"	1442
8	"Expletives"	739
9	"Mention of a routine activity"	1565
10	"Time indicating words"	250
11	"Short tweet length"	1444
12	"Caution and advice for others"	63
13	"Mention of disaster locations"	2879
14	"No-URL"	3338

good f-score for each feature type against the manually extracted feature-words by the language expert.

In the Table-5.3 we can see that Feature-11, which is "Short tweet length", has no result because this feature is not covered by a language expert and it has nothing to do with extraction of words. Similarly for Feature-7 titled as "Exclamation and question marks" and new Feature-14 titled as "NO-URL", are identifiable by the human eye and can be identified easily using an automatic approach. For the rest of the features, the automatic approach has achieved comparative results in comparison to manual results by language experts.

The grammar-rules proposed in this work, as described in Section-4.3 are implemented to the selected dataset of 8000 tweets for all disaster types; earthquake, flood, hurricane and wildfire events. The feature-words were automatically extracted for each proposed grammar-rule. Table-5.4 illustrate the feature-words mined from dataset of processed tweets for all identified features.

The Table-5.4 shows that our automatic approach successfully extracted the feature-words against each identified feature of the manual features-based approach [37].

TABLE 5.5: Experiment-1: Manual Classification(Counts)

Disaster Type	**Earthquake**			**Flood**			**Hurricane**			**Wildfire**		
Category	**Ew**	**NEw**	**Un**	**Ew**	**NEw**	**Un**	**Ew**	**NEw**	**Un**	**Ew**	**NEw**	**Un**
Eyewitness (Ew)	1523	73	4	462	127	38	347	92	26	131	45	13
Non-Eyewitness (NEw)	82	93	25	210	229	112	320	563	316	523	704	152
Unknown (Un)	157	32	11	494	237	91	154	136	46	203	168	61
TOTAL	**1762**	**198**	**40**	**1166**	**593**	**241**	**821**	**791**	**388**	**857**	**917**	**226**

5.3 Results of LR-TED Approach

In Section-3.4 (Evaluation Strategy) we briefly discussed three different types of experiments that we proposed to conduct for the evaluation of our LR-TED approach. The LR-TED approach with only the 13 features as identified by the manual features-based approach, achieve the maximum score of 0.91 for earthquake eyewitness while the addition of 14th feature, improves the results to 0.93 and similar improvement is results is observed for all categories of all disaster events. In the following sub-sections, results and discussion of each experiment are explained.

5.3.1 Experiment-1: Manual Classification

In this experiment, the classification of each tweet into the “Eyewitness”, “Non-eyewitness”, and “Unknown” classes is done in using defined rules as discussed in Section-3.4.1.

After applying the proposed grammar rules described in Section-4.3, the results of 8000 tweets are then used to classify each tweet using defined rules of Section-3.4.1. The results are generated for all thirteen features using LR-TED for each category of “eyewitness”, “non-eyewitness”, and “unknown”, and compared with the benchmark dataset. The comparison of results by LR-TED and benchmark dataset, are illustrated in Table-5.5.

Table 5.6: Experiment-1: Manual Classification (F-Scores) 10-Features

Disaster Type	**Earthquake**			**Flood**			**Hurricane**			**Wildfire**		
Category	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**
Eyewitness	0.92	0.7	0.79	0.46	0.43	0.44	0.51	0.48	0.5	0.23	0.39	0.29
Non-Eyewitness	0.11	0.3	0.16	0.24	0.33	0.27	0.53	0.29	0.37	0.6	0.28	0.38
Unknown	0.16	0.19	0.17	0.34	0.27	0.3	0.09	0.25	0.14	0.11	0.26	0.15

Table 5.7: Experiment-1: Manual Classification (F-Scores) 13-Features

Disaster Type	**Earthquake**			**Flood**			**Hurricane**			**Wildfire**		
Category	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**
Eyewitness	0.88	0.76	0.82	0.38	0.57	0.45	0.4	0.63	0.49	0.15	0.63	0.24
Non-Eyewitness	0.19	0.48	0.27	0.33	0.42	0.37	0.71	0.47	0.56	0.78	0.53	0.63
Unknown	0.23	0.13	0.17	0.44	0.2	0.27	0.12	0.17	0.14	0.28	0.17	0.21

Table 5.8: Experiment-1: Manual Classification (F-Scores) with New-Feature

Disaster Type	**Earthquake**			**Flood**			**Hurricane**			**Wildfire**		
Category	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**
Eyewitness	0.86	0.95	0.91	0.40	0.74	0.52	0.42	0.75	0.54	0.15	0.69	0.25
Non-Eyewitness	0.47	0.47	0.47	0.39	0.42	0.40	0.71	0.47	0.57	0.77	0.51	0.61
Unknown	0.28	0.06	0.09	0.38	0.11	0.17	0.12	0.14	0.13	0.27	0.14	0.19

Table-5.5 describes the result counts generated for each disaster type. For the alignment and formatting of the table, the category names are denoted as; “Ew” for eyewitness, “Un” for unknown, and “NEw” for non-eyewitness categories.

Table-5.6 describes the results generated using the features implemented in the state-of-the-art approach’s model, for all disaster types for the selected evaluation parameters of precision, recall, and f-score. The features dropped by the manual features-based approach are not included in the results shown in Table-5.6. The evaluation parameter names are denoted as; “Pr” for Precision, “Re” for Recall, and “F1” for F-Score parameters.

Table-5.7 describes the results generated, using all thirteen features as identified by the domain experts in the manual features-based approach, for all disaster types and for each evaluation parameter of precision, recall, and f-score. The evaluation parameter names are denoted as; “Pr” for Precision, “Re” for Recall, and “F1” for F-Score parameters.

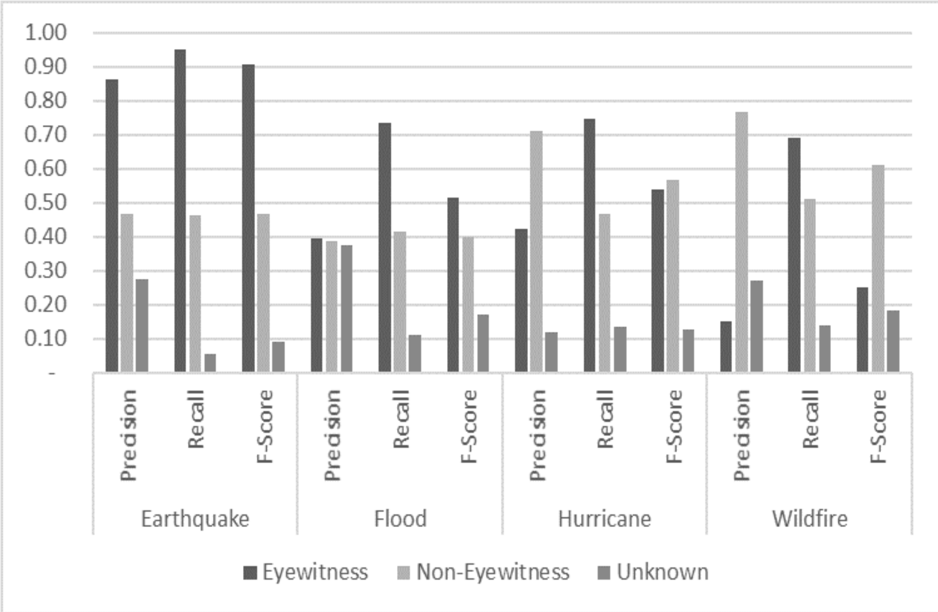

FIGURE 5.2: Experiment-1: Manual Classification (Results)

Tables-5.6, 5.7 & 5.8 shows impact the features dropped by the manual features-based approach, and addition of newly proposed feature to the existing feature list from the manual features-based approach. It is evident that if we compare the results of eyewitness category of earthquake event type we can see the major performance shift from F-Score of 0.79 (for 10-Features), 0.82 (for all 13-Features) to 0.91 (with addition of new-Feature). In the remaining chapter we have used all 14-features for our experiments.

The values for evaluation parameters; precision, recall, and f-measure are calculated, by applying the formulas explained in Section-3.5. The results for each disaster type are illustrated using bar-chart for each evaluation parameter (precision, recall, and f-score) using Table-5.7. The Figure-5.2 depicts graphical representation of Table-5.7.

The generic findings from the results are that, (i) the approach performs best for eyewitness category and identification of eyewitness remains successful and accurate for all disaster types, (ii) the results for the non-eyewitness category are also significant but the accuracy is for this category is low as compare to the eyewitness category. The reason would be that the evaluation dataset contains a low number of tweets for the non-eyewitness and unknown categories. It is observed that the

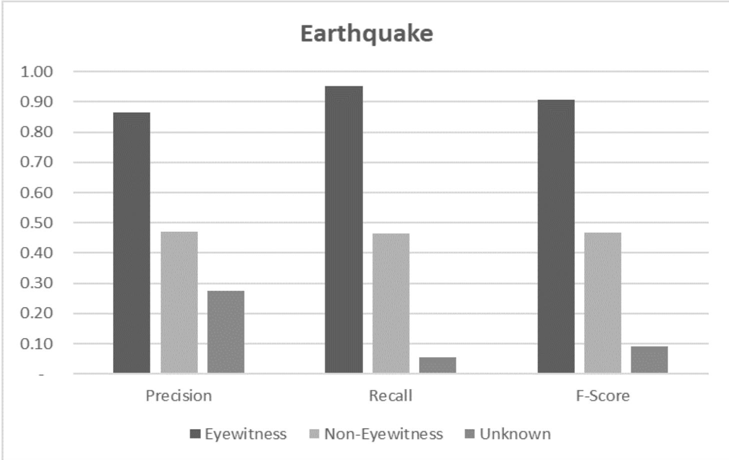

FIGURE 5.3: Manual Classification Results (Earthquake Events)

"Earthquake" disaster category achieved the highest result values for precision, recall, and f-score for eyewitness and non-eyewitness categories. Remaining disaster categories secured the average results for all parameters. Similar proportion of results is observed by the manual features-based approach. The reason for the difference in the results of the disaster types is the nature of these disasters, and characteristics attached to these disaster types. The proposed LR-TED approach achieved the same scores as the manual features-based approach for eyewitness category of earthquake event. The results of each disaster type are briefly discussed separately in the following sub-sections.

5.3.1.1 Earthquake Events

The results for earthquake disaster type events are shown in bar-chart in Figure-5.3, for all the parameters including precision, recall, and f-score by using Table-5.7.

From Figure-5.3, we can see high scores for the eyewitness category for each evaluation parameter (precision, recall, f-score). The author of the manual features-based approach [37] identified that the features related to the non-eyewitness category are overlapping with the eyewitness category. When comparing the results of LR-TED approach for different disaster types we have identified that highest results are generated for earthquake disaster events as such events covers the maximum

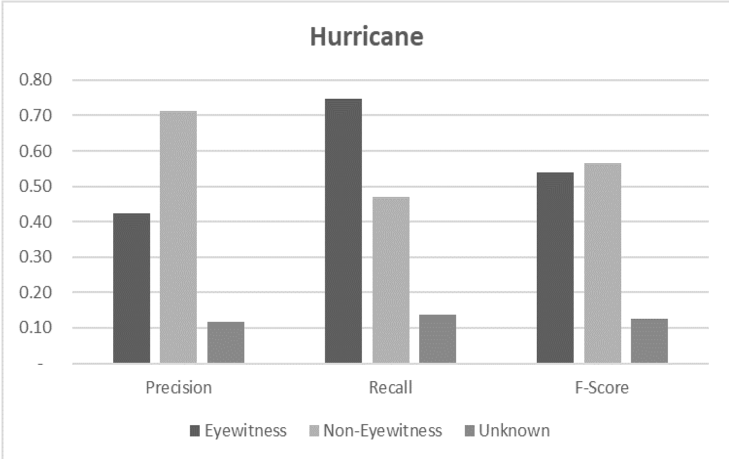

FIGURE 5.4: Manual Classification Results (Hurricane Events)

number of features identified. During manually analyzing the results of disaster types "flood", "wildfire", "earthquake" and "Hurricane" it was observed that the nature of the "earthquake" disaster covers all fourteen features (see Table-1.1) to identify the tweet as an eyewitness.

5.3.1.2 Hurricane Events

The results for disaster events of hurricane types are shown in bar-chart in Figure-5.4. The results for precision, recall, and f-score from Table-5.7 are graphically represented in Figure-5.4.

From Figure-5.4, we can see that LR-TED achieved good results for eyewitness and non-eyewitness categories for each parameter; like precision, recall, and f-score. The results of hurricane events secure good f-score in comparison to the flood events but cannot achieve comparative score with earthquake disaster type. For the hurricane disaster type, the features of "caution and advice for others", "disaster intensity", "words of perceptual senses", and "time indicating words" remained useful features in the identification of eyewitness.

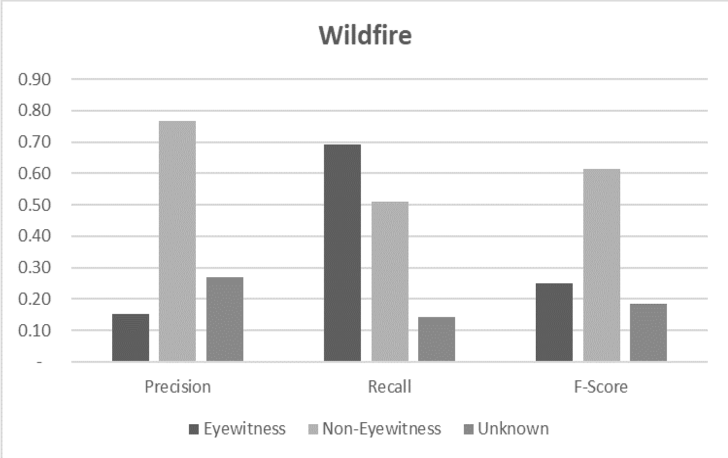

FIGURE 5.5: Manual Classification Results (Wildfire Events)

5.3.1.3 Wildfire Events

The results for disaster type events of wildfire/fire are shown in bar-chart in Figure-5.5.

The results of Figure-5.5 shows the results for the evaluation parameters of precision, recall, and f-score, taken from Table-5.7. In the Figure-5.5, we can see the high recall value for eyewitness category, but low values for precision and f-score. When comparing the results of LR-TED for different disaster type we have identified that, in the event of wildfire the features like "words indicating the intensity", "personalized location markers", "reporting the impact of disasters", and "mention of locations", were the useful features in the identification of eyewitness for such catastrophic events.

5.3.1.4 Flood Events

The results for flood events as disaster type are shown in bar-chart in Figure-5.6, for all parameters of precision, recall, and f-score by using the Table-5.7.

From the Figure-5.6, we can see the good results for each evaluation parameters such as; precision, recall, and f-score, for eyewitness category. The results for non-eyewitness and unknown categories are low in comparison to the eyewitness category. When comparing the results of the flood as disaster type by LR-TED

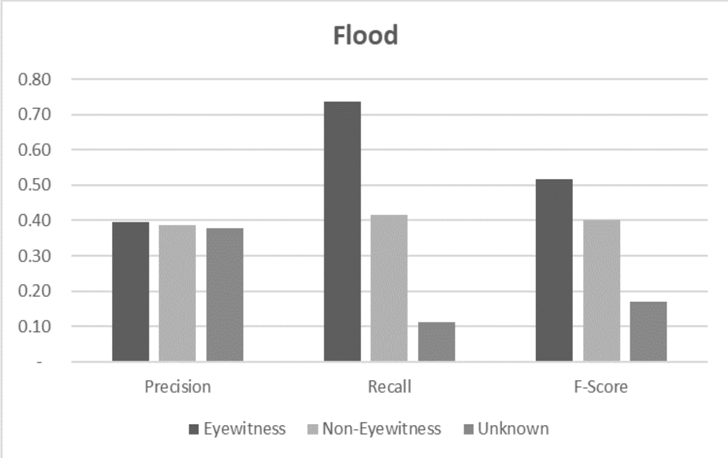

FIGURE 5.6: Manual Classification Results (Flood Events)

with other disaster types we have identified that the features like "words indicating the intensity", "reporting impact of disasters", "personalized location markers", and "mention of locations", were the useful features for such disaster type events.

One of the objectives of this research is to propose a generic approach for all disaster types. To achieve this objective, we applied the same evaluation strategy for all disaster types rather than selecting different features for each disaster type. The findings from the results are shared by discussing a list of features that are more useful than others, in terms of feature extraction for different disaster types. The same behavior of the manual features-based approach is witnessed as the same proportion of results is observed for different disaster types, that we discuss in detail in the comparison section.

5.3.2 Experiment-2: Supervised Learning Models

For our second experiment, we decided to train the supervised learning model, as discussed in Section-3.4.2. In this experiment, we are experiencing various supervised learning models as discussed in Section-3.4.2. The machine learning models that we have identified, other then Random Forest, from the literature are Naive Bayes, Support Vector Machine (SVM), that are commonly used as supervised machine learning models for classification tasks.

Table 5.9: Experiment-2: Supervised Learning Model (Random Forest)

Disaster Type	**Earthquake**			**Flood**			**Hurricane**			**Wildfire**		
Category	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**
Eyewitness	0.88	0.97	0.92	0.52	0.35	0.42	0.57	0.59	0.58	0.39	0.17	0.24
Non-Eyewitness	0.76	0.74	0.75	0.6	0.74	0.66	0.77	0.83	0.8	0.8	0.92	0.85
Unknown	0.22	0.05	0.08	0.53	0.58	0.56	0.49	0.33	0.39	0.67	0.49	0.56

5.3.2.1 Random Forest

The manual features-based approach [37] has considered the Random Forest model as a machine learning model. We used the Random Forest model of classification on our dataset of tweets for each disaster type; earthquake, flood, hurricane, and wildfire. The 10-fold cross-validation strategy was used to evaluate the results as used by the manual features-based approach [37]. The standard input parameters for Random Forest configurations available in the tool was used for performance evaluation. The evaluation parameters; precision, recall, and F-score are calculated and shown in Table-5.9.

The Table-5.9 shows values of precision, recall, and f-score for different disaster types obtained by applying the Random Forest classification technique using WEKA Tool [1]. WEKA is a GUI (graphical user interface) based open source machine learning software that is exploited by researchers of this domain [116] [117] [118]. The results are shown in Table-5.9 are further evaluated and discussed in the comparison section.

5.3.2.2 Naive Bayes

From our proposed methodology, we have results in vector format where each column denotes the identified feature and we have one such vector for each tweet. This is a problem of supervised classification and Naive Bayes is one of the top supervised classification models for such data. The Naive Bayes models are commonly used by researchers in the domain of data mining for classification and

[1] https://www.cs.waikato.ac.nz/ml/weka/

Table 5.10: Experiment-2: Supervised Learning Model (Naive Bayes)

Disaster Type	**Earthquake**			**Flood**			**Hurricane**			**Wildfire**		
Category	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**
Eyewitness	0.89	0.95	0.92	0.52	0.36	0.43	0.58	0.6	0.59	0.33	0.28	0.30
Non-Eyewitness	0.70	0.76	0.73	0.58	0.73	0.64	0.76	0.84	0.80	0.81	0.89	0.85
Unknown	0.26	0.10	0.14	0.54	0.56	0.55	0.53	0.30	0.39	0.65	0.48	0.55

recommendation task [119]. The Naive Bayes approach uses Bayes theorem which classifies by combining the previous knowledge and the new knowledge [120]. In this experiment, we used the WEKA Tool for evaluating our results using 10-fold cross-validation and standard input parameters for performance evaluation. The results for evaluation parameters; precision, recall, and F-score are using Naive Bayes are given in Table-5.10.

From Table-5.10 we can see that the Naive Bayes has produced good results for the eyewitness category of earthquake disaster type. The Naive Bayes model performed very similarly to the random forest model, the comparison of results is discussed in comparison Section-5.4.

5.3.2.3 Support Vector Machine (SVM)

As discussed in the above subsection, the results of LR-TED are available in the form of vectors. For vector data, the SVM model is considered as a top classification model for machine learning applications. The SVM algorithm is commonly used by researchers for classification and prediction based task [121] [122]. WEKA tool has an implementation of the SVM model. Using standard input parameters and 10-fold cross-validation we used the SVM model from the WEKA tool for evaluation. The values of precision, recall, and f-score using SVM models are shown in Table-5.11.

From Table-5.11 we can see that the SVM model produced good results for eyewitness categories for earthquake and hurricane disaster types. But fail to perform for eyewitness category for disaster type wildfire and unknown category for earthquake

Table 5.11: Experiment-2: Supervised Learning Model (Support Vector Machine)

Disaster Type	**Earthquake**			**Flood**			**Hurricane**			**Wildfire**		
Category	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**
Eyewitness	0.88	0.98	0.93	0.53	0.32	0.40	0.60	0.53	0.57	0.00	0.00	0.00
Non-Eyewitness	0.71	0.77	0.74	0.66	0.48	0.56	0.72	0.90	0.80	0.77	0.97	0.86
Unknown	0.00	0.00	0.00	0.50	0.74	0.60	0.64	0.17	0.27	0.71	0.42	0.53

Table 5.12: Deep Learning Approach (Artificial Neural Network - ANN)

Disaster Type	**Earthquake**			**Flood**			**Hurricane**			**Wildfire**		
Category	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**
Eyewitness	0.88	0.98	0.93	0.49	0.42	0.45	0.58	0.62	0.60	0.48	0.18	0.26
Non-Eyewitness	0.76	0.74	0.75	0.58	0.74	0.65	0.76	0.85	0.80	0.79	0.93	0.86
Unknown	0.32	0.04	0.07	0.53	0.49	0.51	0.54	0.27	0.36	0.67	0.47	0.55

events. The results will be further discussed and compared with other models in the comparison Section-5.4.

5.3.3 Experiment-3: Deep Learning Approaches

As discussed in Section-3.4.3, we decided to further investigate the deep learning algorithm in this research work and adopted the Artificial Neural Network (ANN), Recurrent Neural Network (RNN), and Convolutional Neural Networks (CNN) for this task.

5.3.3.1 Artificial Neural Network (ANN)

We used the Artificial Neural Network (ANN) model of classification on our dataset of tweets for each disaster type; earthquake, flood, hurricane, and wildfire, using 10-fold cross-validation with standard input parameters for performance evaluation of ANN model. In the WEKA tool, ANN is available with a different name "Multilayer Perceptron" under the functions tree of classifiers. The evaluation parameters; precision, recall, and F-score are calculated and shown in Table-5.12.

From Table-5.12 it is evident that ANN produced high values for the eyewitness category for earthquake, flood, and hurricane disaster type events. The model

Table 5.13: Deep Learning Approach (Recurrent Neural Network - RNN)

Disaster Type	**Earthquake**			**Flood**			**Hurricane**			**Wildfire**		
Category	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**
Eyewitness	0.8	1	0.89	0.46	0.04	0.07	0.66	0.5	0.57	0	0	0
Non-Eyewitness	0	0	0	0.47	0.83	0.6	0.69	0.94	0.8	0.77	0.97	0.86
Unknown	0	0	0	0.5	0.59	0.54	0	0	0	0.69	0.44	0.53

performed well for the non-eyewitness category for all disaster types. The results are further evaluated and compared in the comparison Section-5.4.

5.3.3.2 Recurrent Neural Network (RNN)

RNN is a powerful model for classification of textual data. LSTM is the implementation of RNN and work for text classification. To analyze different Deep Learning models for our proposed methodology, we also adopted LSTM model of RNN implementation on our dataset of tweets. For its experiment, WEKA has an external project called "wekaDeepLearning4j". Using this project we added a LSTM layer with one output value and used 10-fold cross-validation for performance evaluation of RNN model. The evaluation parameters; precision, recall, and F-score are calculated and shown in Table-5.13.

From Table-5.13, the LSTM model did not produced the results for various categories and different disaster types. For example it did not produced results for eyewitness category of Wildfire disaster type. Similarly no results for non-eyewitness category of earthquake, unknown category of earthquake and hurricane disaster type. The results are further discussed, evaluated and compared in the comparison Section-5.4.

5.3.3.3 Convolutional Neural Networks (CNN)

The CNN model was originally introduced for image processing. The model is then used for classification of textual data. It use multiple Convolutional layers for classification. In the category of evaluation of deep-learning approaches we also

TABLE 5.14: Deep Learning Approach (Convolutional Neural Network - CNN)

Disaster Type	**Earthquake**			**Flood**			**Hurricane**			**Wildfire**		
Category	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**
Eyewitness	0.88	0.99	0.93	0.54	0.33	0.41	0.58	0.58	0.58	0.47	0.13	0.21
Non-Eyewitness	0.76	0.76	0.74	0.62	0.67	0.64	0.75	0.87	0.81	0.79	0.93	0.86
Unknown	0.5	0.01	0.01	0.52	0.64	0.57	0.52	0.23	0.32	0.64	0.48	0.54

adopted CNN for our experiments. WEKA tool using the "wekaDeepLearning4j" project has the option to implement the CNN model. We implemented the technique by adding multiple concolutional layers and used 10-fold cross-validation for performance evaluation of CNN. The resultant evaluation parameters are shown in Table-5.14.

From Table-5.14, it is evident that CNN produces best results for each category for all disaster types. The CNN produced better results than RNN and produced the highest F-Score for "Earthquake" and "Hurricane" disaster types for the eyewitness category. The results are further discussed and compared in the comparison Section-5.4.

5.4 Comparison of Proposed and State-of-the-art Approache

In the first sub-section of the comparison section, we will discuss in detail the Experiment-1 of manual classification and discuss the results for each disaster type. In the second sub-section, we will compare and discuss the results achieved from all experiments to identify the best performing model.

5.4.1 Experiment-1: Comparison of Proposed LR-TED(Manual) Approach with State-of-the-art Approach

Each identified feature by the manual features-based approach [37] is independently evaluated and the results are matched with state-of-the-art approach [37],

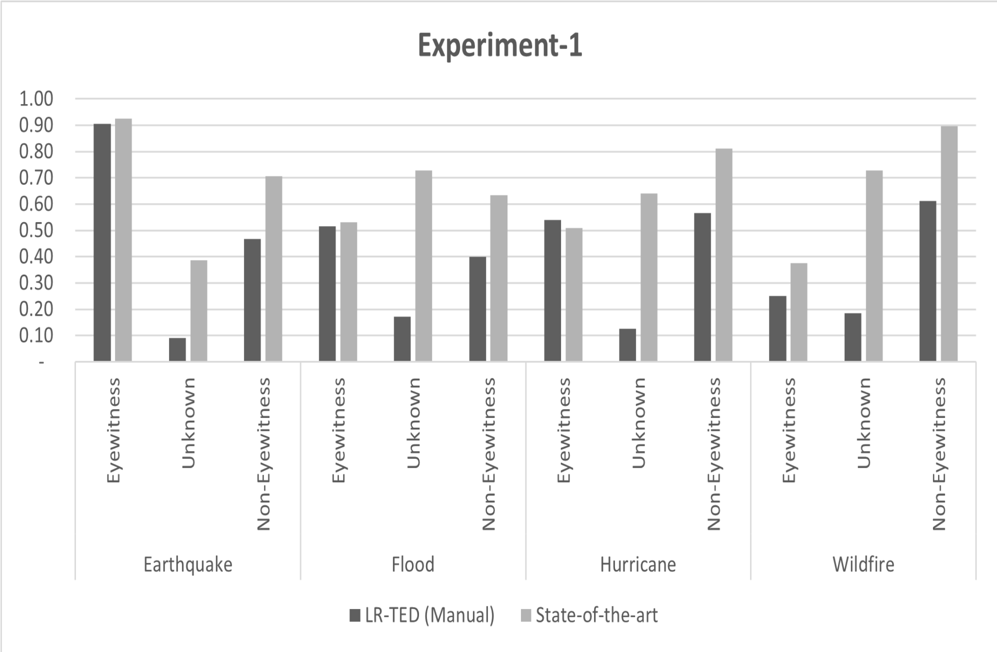

FIGURE 5.7: Comparison of LR-TED (Manual) and State-of-the-art Approach

for each disaster type and category. The F-Scores comparison of LR-TED (Manual) approach and the manual features-based approach are shown as a graph in the Figure-5.7.

To comprehensively compare all disaster events type for all categories or classes of the eyewitness, non-eyewitness, and unknown categories we have discuss the results for each event type in the following sub-sections.

5.4.1.1 Earthquake Events

The performance of LR-TED is best for earthquake events, we have the same observation about the manual features-based approach where the identified features perform best for earthquake events rather than floods, hurricanes, and wildfire events. In Table-5.15 the values of all evaluation features (precision, recall, and f-score) are shown for LR-TED and state-of-the-art approach.

When the f-scores are compared for both approaches, the comparison is shown in Figure-5.8.

TABLE 5.15: Comparison of LR-TED (Manual) and State-of-the-art Approach (Earthquake)

	LR-TED (Manual)			State-of-the-art Approach		
Category	**Precision**	**Recall**	**F-Score**	**Precision**	**Recall**	**F-Score**
Eyewitness	0.86	0.95	0.91	0.87	0.97	0.92
Non-Eyewitness	0.47	0.47	0.47	0.79	0.65	0.71
Unknown	0.28	0.06	0.09	0.33	0.10	0.15

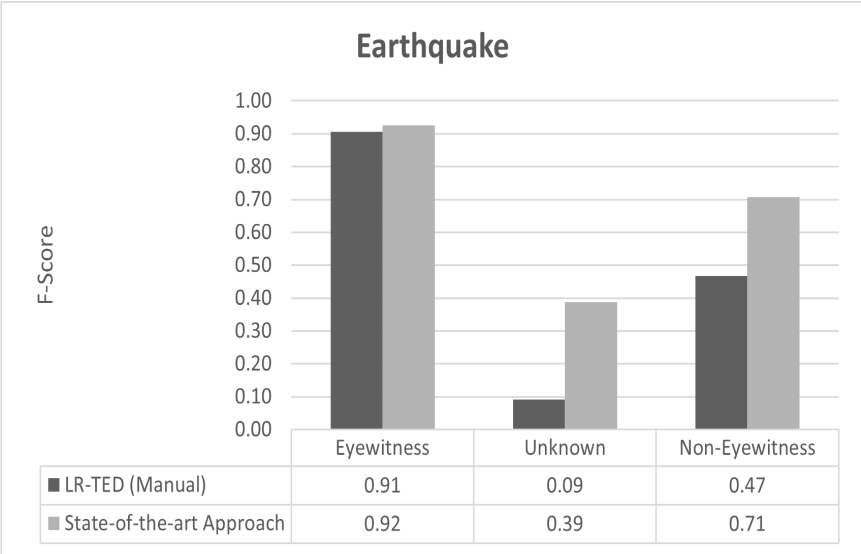

FIGURE 5.8: F-Score Comparison of LR-TED (Manual) and State-of-the-art Approach (Earthquake)

From Figure-5.8, the manual features-based approach, using manually created dictionary of words, achieved best f-score of "0.92", for earthquake disaster events. The LR-TED approach, using the grammar rules, for extraction of feature words secured the best f-score of "0.91" for the earthquake disaster events.

LR-TED produced the highest results for eyewitness earthquake events. By manual analysis of the results, it is identified with the nature of the disaster type the list of important features from the list of fourteen is different. For example, feature number seven "Exclamation and question marks" is an important feature for earthquake disaster type. But its importance for other disaster types is not considered. Similarly, another feature that performs best for earthquake events is the "short tweet length".

The manual features-based approach adopted the results of Doggett et al. [40]

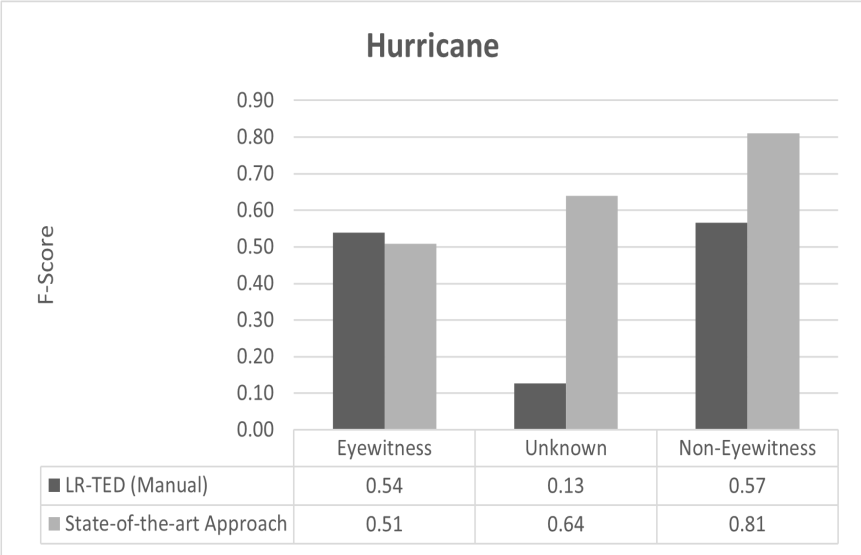

FIGURE 5.9: F-Score Comparison of LR-TED (Manual) and State-of-the-art Approach (Hurricane)

techniques for identification of non-eyewitness and the result details are not explained [37]. Although the focus of this research is to eventually identify the eyewitness reports from tweet content.

The proposed, LR-TED approach produced the same results and can be considered as a potential approach as; i) no static dictionary management for each disaster type, ii) processing, possibly millions of tweets for unseen events, in real-time, iii) a generalized approach that works for various disaster types without any human effort.

5.4.1.2 Hurricane Events

The results for hurricane events are better results for eyewitness class in comparison to the flood events. The details of the comparison with the manual features-based approach are shown in Figure-5.9.

The results from Figure-5.9 show that LR-TED achieved comparative results in comparison with a manual dictionary-based approach. The results of the non-eyewitness category are very high for the manual features-based approach and they are adopted by using the approach of Doggett et al. [40]. Since the focus is mostly

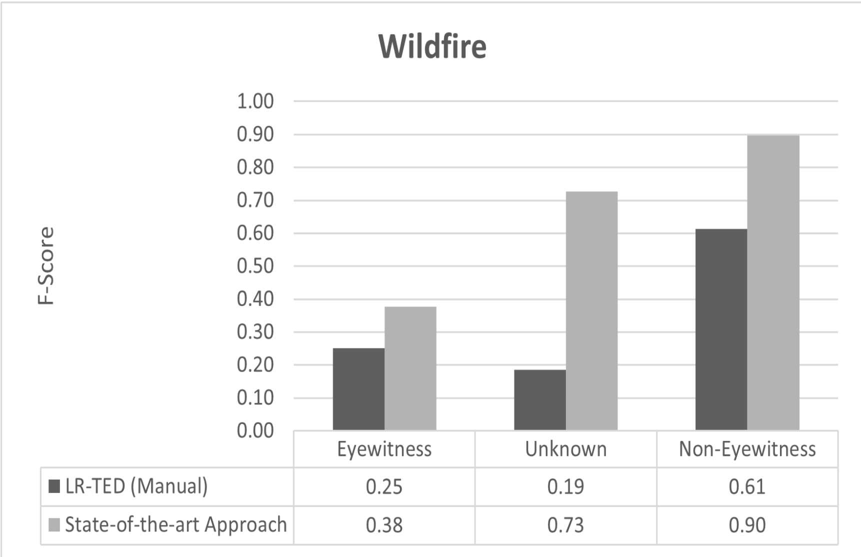

FIGURE 5.10: F-Score Comparison of LR-TED (Manual) and State-of-the-art Approach (Wildfire)

on eyewitness identification, so the classes of non-eyewitness and unknown are not discussed. For hurricane events, the important features from the list of fourteen features are, caution and advice for others, disaster intensity, words of perceptual senses, and time indicating words. Securing the f-score of 0.47 in comparison to 0.51 by applying the automatic approach of feature extraction without the use of manual dictionaries is a considerable effort.

5.4.1.3 Wildfire Events

The f-score of LR-TED and the manual features-based approach for wildfire events are shown in Figure-5.10 .

The proposed LR-TED approach produced lowest for wildfire events in comparison to the other disaster types discussed in this research. Even for the eyewitness category, the results are very low. This proportion of low results for these disaster types is also low for the manual features-based approach.

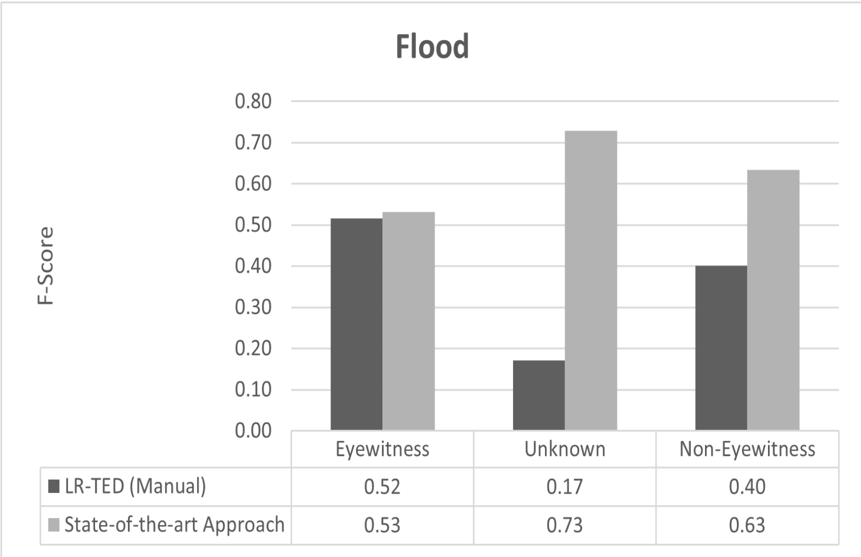

Figure 5.11: F-Score Comparison of LR-TED (Manual) and State-of-the-art Approach (Flood)

5.4.1.4 Flood Events

In Figure-5.11, the f-score for flood disaster type are shown for both State-of-the-art approach and LR-TED.

From Figure-5.11 it evident that the LR-TED approach generated comparative results for the eyewitness category, but did not compete the results of the manual features-based approach. One of the reasons for these results could be the training dataset where it contains a small number of eyewitness tweets to train the model. These variations in the performance of LR-TED approach are acceptable to implement the concept of a generalized approach. For example, most earthquake-related classes score the high results in terms of the f-score and the produced low results for flood events. Based on the nature of the event, the importance of the features varies as the features like words indicating the intensity, reporting impact of disasters, personalized location markers, and mention of locations, were the useful features for flood events.

The state-of-the-art approach, by using the manually created static dictionary, secured the F-Score of “0.92” for identification of “Eyewitness” tweets under the earthquake category. While for the same category, the grammar rules-based LR-TED approach, achieved the F-Score of "0.91". For the non-eyewitness category, the manual features-based approach adopted the Doggett et al. [40] technique and secured the F-Score of 0.71, and the LR-TED secured 0.47. As discussed earlier, the focus of this research is for identification of eyewitness, and for non-eyewitness identification, we can adopt the Doggett et al. [40] as adopted by the manual features-based approach. For the unknown category, the manual features-based approach secured the F-Score of "0.15", while the LR-TED approach secured the F-Score of "0.09". The results for other disaster types remain in the same proportion in comparison to the manual features-based approach. LR-TED approach outperform the manual features-based approach by securing the F-Score of "0.54" where the manual features-based approach secured the F-Score of "0.51". The proposed LR-TED technique for "wildfire" disaster type produce comparative but not greater results in contrast to the results generated for other disaster types. The proposed LR-TED generated comparative results, considering its an automatic approach that avoid human interaction for dictionary management.

The comparison discussed in this section only discusses the results of experiment number one, where the results of manual classification are compared with the manual features-based approach. The detailed comparison of each experiment performed in this word is discussed in detail in the following sections.

5.4.2 Experiment-2: Comparison of Proposed LR-TED (Supervised Learning Models) Approach with State-of-the-art Approach

We evaluated all thirteen features identified by the manual features-based approach [37] and the newly added feature, independently and the results are compared with the state-of-the-art approach [37]. Comparison is done for each disaster type and category. The F-Scores comparison of LR-TED approach by adopting the ANN

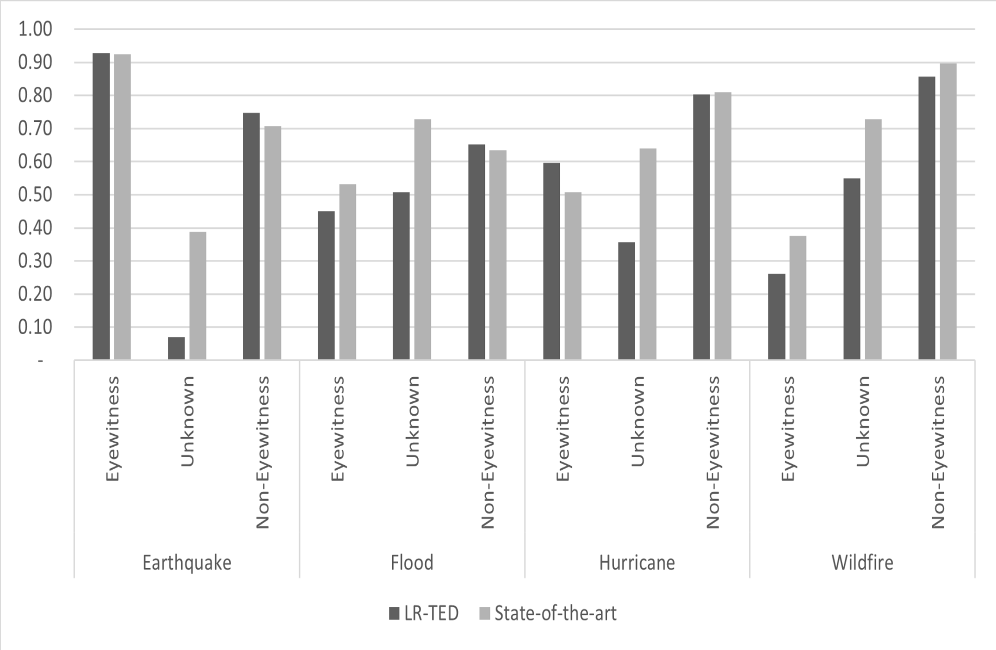

Figure 5.12: Comparison of LR-TED and State-of-the-art Approach

deep-learning model for classification and F-Scores of the manual features-based approach are shown as a graph in the Figure-5.12.

For the comparison of all disaster events type for all categories namely the eyewitness, non-eyewitness, and unknown categories we have discuss the results for each event type in the following sub-sections.

5.4.2.1 Earthquake Events

The performance of LR-TED with ANN model produced best results for earthquake events, same observation are found for the manual features-based approach where the identified features perform best for earthquake events. In Table-5.16, the results for all evaluation features (precision, recall, and f-score) are shown for ANN model based LR-TED and the state-of-the-art approach.

The comparison of F-Scores for both approaches is shown in Figure-5.13. From Figure-5.13, the manual features-based approach, by adopting the manually created dictionary of words, achieved the best f-score of “0.92”, for earthquake disaster events. The LR-TED approach, using the grammar rules secured the best

TABLE 5.16: Comparison of LR-TED and State-of-the-art Approach (Earthquake)

	LR-TED			**State-of-the-art Approach**		
Category	**Precision**	**Recall**	**F-Score**	**Precision**	**Recall**	**F-Score**
Eyewitness	0.88	0.98	0.93	0.87	0.97	0.92
Non-Eyewitness	0.76	0.74	0.75	0.79	0.65	0.71
Unknown	0.32	0.04	0.07	0.33	0.10	0.15

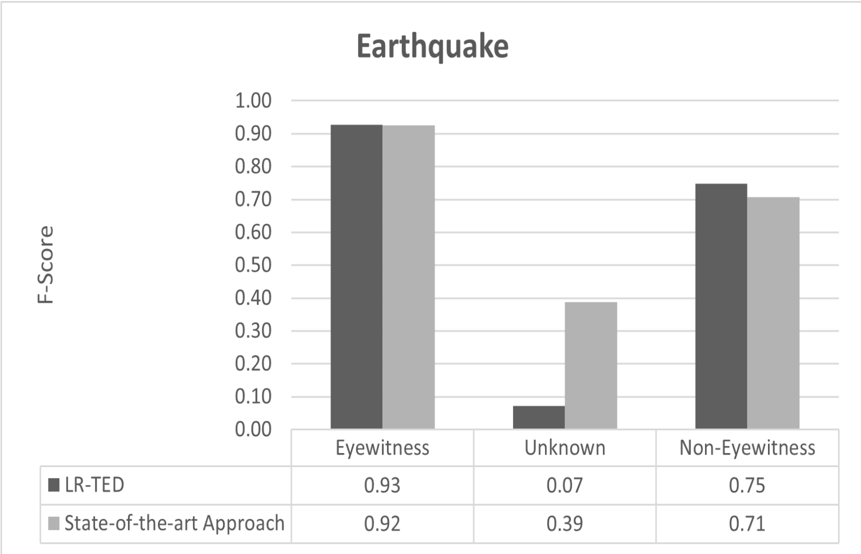

FIGURE 5.13: F-Score Comparison of LR-TED and State-of-the-art Approach (Earthquake)

f-score of "0.93" for the earthquake disaster events.

LR-TED produced the highest results for eyewitness earthquake events. The results for both eyewitness and non-eywitness classes fo earthquake events are better than the results produced by the manual features-based approach. Although the focus of this research is to eventually identify the eyewitness reports from tweet content.

The proposed, LR-TED approach produced better results for eyewitness category and can be considered as a potential approach as it is not dependent on static dictionary management without any human effort.

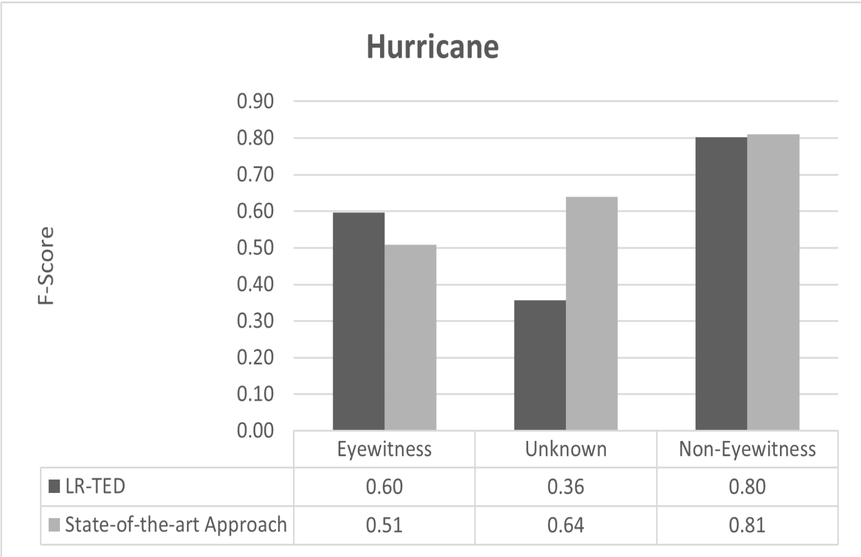

FIGURE 5.14: F-Score Comparison of LR-TED and State-of-the-art Approach (Hurricane)

5.4.2.2 Hurricane Events

The comparison for hurricane disaster events shows better results for eyewitness class for LR-TED approach. The details of the comparison with state-of-the-art approach are shown in Figure-5.14.

The results from Figure-5.14 show that LR-TED achieved best results in comparison to the manual dictionary-based approach by the manual features-based approach [37]. Proposed LR-TED approach secured the F-Score of "0.60" in comparison to F-Score by the manual features-based approach of "0.51" by adopting the ANN model of classification.

5.4.2.3 Wildfire Events

The comparative results in terms of F-Scores for both LR-TED approach and state-of-the-art approach for the wildfire disaster events are shown in Figure-5.15 .

The proposed LR-TED approach was not able to produce best results for wildfire disaster events. Similar, low results for wildfire disaster types are observed for

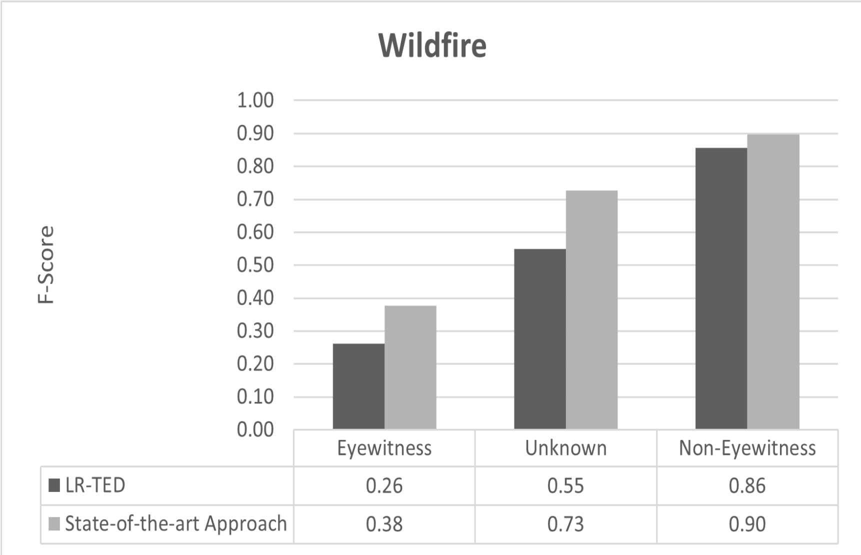

FIGURE 5.15: F-Score Comparison of LR-TED and State-of-the-art Approach (Wildfire)

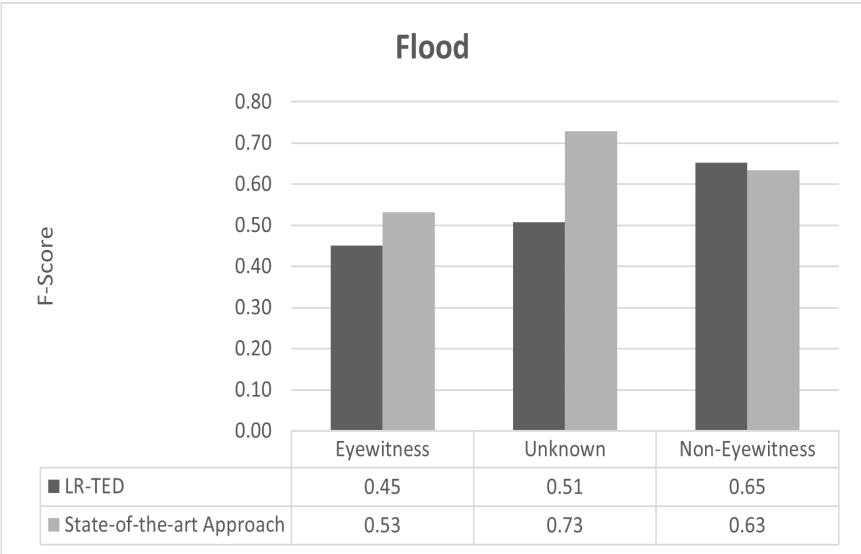

FIGURE 5.16: F-Score Comparison of LR-TED and State-of-the-art Approach (Flood)

the manual features-based approach. The results shows that LR-TED produced comparative results without using manually created dictionaries and list.

5.4.2.4 Flood Events

Figure-5.16 shows the resultant F-Scores for flood disasters type for automatic LR-TED and manual dictionary based state-of-the-art approach.

From Figure-5.16 it evident that the LR-TED approach generated comparative results for the eyewitness category, but did not beat the results secured by the manual features-based approach. The reasons for such results could be the training dataset where it contains a small number of eyewitness tweets to train the model. To implement the concept of a generalized approach, such variations in the performance of LR-TED approach are acceptable.

The state-of-the-art approach, by using the manually created static dictionary, secured the F-Score of "0.92" for identification of "Eyewitness" tweets under the earthquake category. While for the same category, the grammar rules-based LR-TED approach, achieved the F-Score of "0.93". Similarly, for hurricane event type, the proposed LR-TED approach secure "0.60" in comparison to the "0.51" by the manual features-based approach. The comparative results for flood and wildfire disaster types are produced by LR-TED approach without needing any human involvement.

5.4.3 Experiment-3: Comparison of Proposed LR-TED Approach with State-of-the-art Approach (Manual, Supervised Learning Models, and Deep Learning Approaches)

In this research, we have independently evaluated each feature type for each experiment, and results are compared with the manual features-based approach [37]. The comparison of LR-TED approach with state-of-the-art approach is shown in Section-5.4.1, in terms of selected evaluation parameters of precision, recall, and F-Scores. The manual features-based approach use static dictionaries that are manually created for evaluation of an event. The approach is not scale-able for new events. For example, consider the subsequent scenarios:

1. The words in a static dictionary might change, when the domain changes

Table 5.17: Comparison of Results - Eyewitness (All Experiments)

MODEL Disaster Type	**Manual (F1)**	**R-F (F1)**	**N-B (F1)**	**SVM (F1)**	**LR-TED (F1)**	**RNN (F1)**	**CNN (F1)**	**SA [37] (F1)**
Earthquake	0.91	0.92	0.92	0.93	0.93	0.89	0.93	0.92
Flood	0.52	0.42	0.43	0.4	0.45	0.07	0.41	0.53
Hurricane	0.54	0.58	0.59	0.57	0.6	0.57	0.58	0.51
Wildfire	0.25	0.24	0.3	0	0.26	0	0.21	0.38

2. The undetected tweets may have diverse vocabulary or words that directly affect the performance of the manual features-based approach.

In a catastrophic event, it would be a cumbersome effort for processing of potentially millions of tweets for the identification of eyewitness. Such a task may require thousands of domain-experts to be involved in the process of updating the dictionary. To cover such gap, we have studied the linguistic features, language structure, word relations, and language patterns; have been critically reviewed for automatic identification of the eyewitness tweets. Subsequently, we adopted the idea of creating grammar rules, to automatically extract the feature-words. The grammar rules were created to extract the feature-words for all thirteen features identified by the manual features-based approach [37], without consuming the static dictionary, with reasonable accuracy.

In this section, the results achieved from each experiment in Section-5.3, using different models and approaches are compared to find out which model performs better and for which category and disaster types. In Table-5.17, the f-score achieved using different models applied in experiments to simplify the comparison is shown for the "eyewitness" class for each model and disaster type. Similarly, the Table-5.18 for "non-eyewitness" class and Table-5.19 for "unknown" class. The state-of-the-art approach [37] is labeled as "SA [37]" in these tables.

In this section, we are comparing results achieved from each experiment. This research focuses on the identification of the "eyewitness" class of users from tweet content. Therefore in this section, we have compared our results at the class level. The following sub-sections discuss the results acquired from all experiments for identified classes of the eyewitness, non-eyewitness, unknown.

TABLE 5.18: Comparison of Results - Non-Eyewitness (All Experiments)

MODEL **Disaster Type**	**Manual (F1)**	**R-F (F1)**	**N-B (F1)**	**SVM (F1)**	**LR-TED (F1)**	**RNN (F1)**	**CNN (F1)**	**SA [37] (F1)**
Earthquake	0.47	0.75	0.73	0.74	0.75	0	0.74	0.71
Flood	0.4	0.66	0.64	0.56	0.65	0.6	0.64	0.63
Hurricane	0.57	0.8	0.8	0.8	0.8	0.8	0.81	0.81
Wildfire	0.61	0.85	0.85	0.86	0.86	0.86	0.86	0.9

TABLE 5.19: Comparison of Results - Unknown (All Experiments)

MODEL **Disaster Type**	**Manual (F1)**	**R-F (F1)**	**N-B (F1)**	**SVM (F1)**	**LR-TED (F1)**	**RNN (F1)**	**CNN (F1)**	**SA [37] (F1)**
Earthquake	0.09	0.08	0.14	0	0.07	0	0.01	0.39
Flood	0.17	0.56	0.55	0.6	0.51	0.54	0.57	0.73
Hurricane	0.13	0.39	0.39	0.27	0.36	0	0.32	0.64
Wildfire	0.19	0.56	0.55	0.53	0.55	0.53	0.54	0.73

5.4.3.1 Eyewitness

Table-5.17 shows results from all experiments for eyewitness class using manual supervised, and deep learning classification models for different disaster events. The columns labeled as "R-F", "N-B", and "State-of-the-art" means the "Random Forest", "Naive Bayes", and "State-of-the-art approach" respectively. From Table-5.17 for comparison of different experiments of eyewitness class, the maximum f-scores achieved are 0.93, 0.52, 0.60, and 0.26 for earthquake, flood, hurricane, and wildfire events respectively. For earthquake events, the maximum scores are achieved using supervised and deep learning models. Similarly for hurricane and wildfire event types, supervised learning approaches secured maximum scores. For flood events, the manual approach outperforms the supervised learning models. The LR-TED approach, as an individual approach for eyewitness class, has produced better results, for three out of four disaster type, and comparative results for flood disaster types.

From Table-5.17, the LR-TED was then compared with the manual features-based approach for each disaster types such as earthquakes, hurricanes, floods, and wildfire. By adopting the the static dictionaries, the state-of-the-art approach [37] achieve the F-Score of "0.92", for the eyewitness category for earthquake events, 0.55 for flood, 0.60 for hurricane and 0.46 for wildfire categories. The LR-TED approach achieved maximum F-Score values of 0.93, 0.45, 0.60, and 0.26 in the respective categories. For earthquake and hurricane disaster types, the LR-TED outperform the state-of-the-art approach. For remaining two disaster types, LR-TED produced comparative score. These scores can be considered as significant scores by an automated process without using a static dictionary.

5.4.3.2 Non-Eyewitness

The Table-5.18 depicts the results for comparison of different experiments of the non-eyewitness class using different models, the maximum f-scores achieved are 0.75, 0.65, 0.80, and 0.86 for earthquake, flood, hurricane, and wildfire events respectively. These maximum scores are achieved by using supervised and deep learning models for all disaster-types. From Table-5.18, for non-eyewitness identification the state-of-the-art adopted the approach of Doggett et al. [40] to secure f-score of 0.71, 0.65, 0.83, and 0.90 for earthquake, flood, hurricane and wildfire events respectively. For evaluation of the non-eyewitness class, the manual features-based approach adopted Doggett et al. [40] approach, that can be adopted in the proposed methodology.

In comparison to the manual features-based approach, the proposed LR-TED approach secure a maximum f-score of 0.75, 0.65, 0.80, and 0.86 for respective disaster types for the non-eyewitness category. The Random-Forest algorithm and LR-TED as an individual models has performed better and achieved significant scores for each disaster type without using a static dictionary.

5.4.3.3 Unknown

From Table-5.19 for comparison of different experiments of unknown class, the maximum f-scores achieved from experiments performed for earthquake, flood, hurricane, and wildfire events are 0.14, 0.60, 0.39, and 0.56 respectively. For earthquake events, the maximum score for the unknown class is 0.14, achieved by using the SVM model used in Experiment-2. For flood, hurricane, and wildfire event types, supervised learning approaches secured maximum scores. The results of LR-TED was then compared with the manual features-based approach for all disaster-types. Using manually created static dictionaries, the manual features-basedapproach [37], was able to secure the F-Score of "0.71" for unknown class for the wildfire category, 0.69 for flood, 0.55 for hurricane and 0.15 for earthquake disaster category. LR-TED approach achieved the comparative F-Score values of 0.07, 0.51, 0.36, and 0.55 for LR-TED in the respective categories. These scores can be considered as a significant score for respective events without using manually created static dictionaries. For unknown class, the random forest and naive bayes models produced comparative results for all disaster types.

From the above experiments we have observed that LR-TED, Naive Bayes, SVM, and Random Forest algorithms performed best for eyewitness, non-eyewitness, and unknown categories respectively. LR-TED scores remained best among other models for each disaster type. While comparing the results with the manual features-based approach we have demonstrated that the algorithms used by the proposed LR-TED approach have achieved comparative results considering that the proposed LR-TED approach is a fully automatic approach without using the static dictionary as used by the manual features-based approach.

By adopting the ANN model for classification in LR-TED, we achieved the best results for "Earthquake" and "Hurricane" disaster types. ANN model based LR-TED also produced comparative results for "Flood" and "Wildfire" disaster types. After comprehensive study and experiments we finally selected the ANN model as part of LR-TED based proposed solution for identification of eyewitness features and classification of the tweets.

As the proposed LR-TED technique automatically extracts the feature-words without involving any human interaction to manually create and maintain the dictionary of feature-words, we are in position to claim that the grammar rule base LR-TED approach could be deemed as a potential approach for identification of eyewitness tweets, particularly in the scenarios where potentially millions of tweets needs to be processed in quick time and for the unknown events.

5.5 Effort Estimation

In this section, we critically estimated the efforts required by each approach for the identification of eyewitness tweets. In the following subsections, we have initially presented the required steps to identify the eyewitness tweets for an unseen event and explain the different procedures adopted by State-of-the-art approach and the proposed approach of LR-TED.

TABLE 5.20: Methodology Steps for State-of-the-art Approach and Proposed LR-TED.

Steps	**State-of-the-art Approach**	**Proposed LR-TED Approach**
Step-1: Disaster-related data collection	Twitter Streaming API to collect data from July 2016 to May 2018 using a methodology described in this work (Zahra et al., 2017)	The dataset employed in this research is the same as used by state-of-the-art for the evaluation of their technique.
Step-2: Pre-processing of Data	Generally, the pre-processing task of tweets includes, the removal of hashtags, HTML tags, extra white spaces, and special symbols.	The dataset employed in this research is the same as used by Zahra et. al. for the evaluation of their technique. The employed tool has the capability of avoiding the data noise while processing the POS tagging.
Step-3: Manual Analysis to determine eyewitness types and characteristics	Manually analyzed data from all three disaster types for eyewitness identification. Created feature sets, or lists of feature words for each characteristic.	Manual analysis is not required by LT-TED. The proposed system can automatically identify the eyewitness feature words based on the grammar rules created by analyzing the language structure for each eyewitness identification characteristic.

...Continued on next page

Table 5.20 – continued from previous page

Steps	State-of-the-art Approach	Proposed LR-TED Approach
Step-4: Crowd sourcing to obtain labeled data	Performed crowdsourced labeling on our second data sample for primarily two reasons: First, to acquire more training data so we can use machine learning algorithms to automatically classify the reports. Second, to validate our own eyewitness reports taxonomy developed from the manual analysis with our first data sample.	The dataset employed in this research is the same as used by the manual features-based approach for the evaluation of their technique. When the approach is adopted for a new event type, it does not require the event-specific words and data for training and evaluation of the approach.
Step-5: Automatic feature extraction	(i) Unigrams and (ii) bi-grams and compute their TF-IDF scores, by adopting the manually created list of feature words for each identified characteristic for eyewitness identification. Each unigram or bi-gram word from the content is then matched with the manually created or maintained lists or dictionaries.	The proposed LR-TED adopted the Stanford CoreNLP tool to process each tweet of the focused dataset, for Tokenization, Part-of-speech (POS) tagging, Named Entity Recognition (NER), Lemmatization, and Dependency Relation. Using tools data. Here, one of the most important things to be noticed is that the predefined lists and dictionaries adopted by the manual features-based approach are never the same for all event types.
Step-6: Automatic labeling	The labeling for the manual approach is done using crowd sourcing in step 4.	For the LR-TED approach, the labeling of tweets is done after the extraction of feature words using Rules, Supervised Learning models, and Deep learning models.

Table 5.21: Characteristics implementation by State-of-the-art & Proposed LR-TED Approach

Characteristics	**State-of-the-art Approach**	**Proposed LR-TED**
"Reporting small details of surroundings"	Dropped	Rules
"Words indicating perceptual senses"	List (LIWC)	List of Lemma Words
"Reporting impact of disaster"	List	Rules
"Words indicating intensity of disaster"	List	Rules
"First person pronouns and adjectives"	List	List (required)
"Personalized location markers"	Dropped	Rules & List
"Exclamation and question marks"	List	List (required)
"Expletives"	List	List
"Mention of a routine activity"	List	Rules
"Time indicating words"	List	Rules
"Short tweet length"	Word Count	Word Count
"Caution and advice for others"	List	Rules
"Mention of disaster locations"	Dropped	Rules

In this research, we already have adopted the technique of validating the automatically extracted features with the ones manually extracted by the experts in the field as discussed in Section-5.2. Unfortunately, the focused dataset does not contain any information, list, or table containing the feature words extracted from each tweet. This process of identifying the words by the way of explanation in the sentence is called the annotation.

For the maximum number of identified characteristics, the proposed LR-TED approach has defined grammar rules, for a limited number of characteristics the use of lists is obligatory. The following Table-5.21 provides the characteristics-based comparison of state-of-the-art approach and the LR-TED approach.

The proposed LR-TED approach has the limitation of adopting pre-defined Words indicating perceptual senses, First-person pronouns and adjectives, and Exclamation and question marks where we have a limited number of words, and the list is static whatever the event types are. For Expletives characteristics we have the limitation that we must adopt the similar approach as employed by state-of-the-art. The vocabulary of slang/expletive words can change with the passage of time. So, the effort estimation for the Expletives characteristic is like state-of-the-art approach. While comparing both approaches in terms of effort estimation, the proposed LR-TED approach will outperform the manual features-based approach in approximately 10 characteristics, whereas the manual features-based approach requires a manually created list for token comparison to identify the feature words.

5.5.1 Annotation Efforts Estimation, in Literature

In 2015, Chen et al. [123], discussed the annotation process as a time-consuming effort. The author exploited the clinical domain, where some tasks require domain experts (e.g. physicians or nurses) to annotate text; thus the cost of annotation could be very high.

The creation of new lists or dictionaries for a new domain (in our case the events), is an expensive task in terms of time and financial cost [124]. The human effort and crowd effort are used alternatively in the literature. Effort estimation is measured for the identification of cheater detection by Eickhoff and de Vries [125]. In literature, the effort is estimated as the amount of time required to complete a Human Intelligence Tasks (HIT). The effectiveness of the effort estimation is studied by Jain et al. [126] by improving the user experience.

In 2014, Hirth et al. [127] presented their study by dividing the required efforts into the dimensions of Reading time and Answering time. The reading time reflects the efforts required to read the input information, while the answering time is evaluated in terms of effort required to notify the output, based on the defined rules.

The annotation of large-scale data is a tedious job and requires time and effort. Estimation of time lines in advance for such a task is always important. Recently, Gomes et al. in 2020 [128] presented their work for estimation of human efforts for named entity annotation.

The prediction of effort is based on multiple features, that includes the followings.

- Number of categories of named entity

$$JobEffort = \sum_{t=0}^{T} (\frac{NumberTokens_t}{PredictedSoT_t} * Redund)$$

- Input length in number of tokens
- Number of sentences
- Number of punctuation tokens
- Number of stop words
- Average word length
- Average word length without stop-words

Gomes et al. [128] assigned Human Intelligence Tasks (HITs) to non-expert contributors and labeled the task as Speed-on-task, which is measured in words-per-minute (wpm). The mathematical transformation is as follows.

```
Speed-on-task (SoT) = NumberofTokens/Time-on-task
```

The author employed the predicted efforts at the HIT level to compute the job effort. To calculate the effort on the job, Gomes et al. [128] formulate the equation as follows.

Where HITs are presented as T, the number of tokens is represented as *NumberTokens* for a task (in our case the tweet), the predicted mean value for speed-on-task is represented as *PredictedSoT* and required redundancy is represented as *Redund.*

5.5.2 Effort Estimation for the Proposed LR-TED and State-of-the-art Approach

To extract such data, the tweets with the list of characteristics were shared with the three experts in the fields, having their expertise in the fields of information extraction and sentiment analysis. They put their efforts for 20-days into carefully reading each tweet and identifying the feature words against all identified features. On average, each expert invested 120 hours to complete the given task of 2000 tweets.

The ideal scenario should be, to have an actual human annotation cost, but it requires multiple user studies. For the same data, the annotation cost may vary for different users [129]. If we do not consider the correlation among the results provided by the experts, each tweet requires a minimum effort of approximately 2.5 minutes to carefully read and identify the features words. Notifying each feature word under each feature type took around 1 minute. The time division is discussed by Hirth et al. who have segregated the effort time into two categories *reading time* and *answering time*. The estimated time also considers the average time required by a person to read a limited number of words per minute. In 2016, Rayner et al. [130] presented the study of human reading capabilities and stated that an average college graduate adult can read 200 to 400 words per minute (wpm). The average time to identify and report the feature words from each tweet took 3.5 minutes.

The estimated values are calculated based on the efforts invested by experts in annotating the provided data for 2000 tweets. Here we must keep in mind the baseline concepts integrated with the provided time lines are that the tweet length cannot exceed the limit of 144 characters, and the URLs, hashtags were removed during the pre-processing task.

To formulate the effort, let us consider the following.

```
 t: the time to identify the feature words from a tweet
 N: number of tweets to be processed
 FRT: Feature-words Reporting Time in minutes
So,
 FRT = t x N,(here in this example, t = 3.5 minutes)
```

Generally, the efforts are estimated in terms of time, and from the literature review, the time is calculated in minutes [128].

In the event of a disaster, millions of relevant tweets are received, that need to be processed quickly. While the manual approach requires human interaction and several people to quickly perform this task. The compilation of data from each human into one single dataset is also a time-consuming task. The proposed approach does not require defined dictionaries or lists for feature identification. Human efforts are not required to create and maintain the manual dictionaries or lists for the identified features.

With the passage of time the language evolves, and the vocabulary increases or changes with time. The tweets may contain diverse vocabulary, in that case, such words need to be evaluated time-to-time, by the language experts to maintain the dictionaries to entertain the relevant features.

The proposed LR-TED approach utilized the Stanford CoreNLP tool and over the top, we have defined a set of grammar rules that reduce human efforts. The prototype system is created by adopting the modular approach where each required task is implemented with the best available approach. For example, the major task in the implementation of the proposed approach is the implementation of Stanford CoreNLP which is available in Java code. The output is then loaded to an RDMS for implementation of the grammar rules. The effort in terms of time is only required for the processing of tweet content through the Stanford CoreNLP tool which took a small amount of time (in a few minutes) to generate the results. The rules are written in the form of SQL queries that also required 2-3 minutes each to formulate the results. The estimated effort in terms of time for the proposed LR-TED is in minutes for 2000 tweets which is very less in comparison to the time required by several experts to manually annotate the tweets.

5.5.2.1 Effort (Speed-on-task) Estimation using HIT

After the analysis of the data, it is identified that on average a tweet contains 25 tokens, and the time to complete the annotation of one tweet is 3.5 minutes. The mathematical transformation to calculate the speed-on-task is as follows.

```
 NumberofTokens (per tweet) = 25
 Time-on-task = 3.5 minutes
So,
 Speed-on-task (SoT) =  NumberofTokens/Time-on-task
    = 25/3.5
    = 7.14
```

The speed of the annotation task is 7.14 tokens per minute by using the estimation of per tweet processing time as 3.5 minutes, where each tweet contains 25 tokens on average.

5.5.3 Required Time - Effort Estimation

The Table-5.22 illustrates the time taken by both approaches at each required step. Here we made some assumptions, and they are the same for both the approaches, such as.

- The dataset adopted for both approaches is a publicly available dataset provided by state-of-the-art [103] and requires only the integration/import time within our approach and tool used.
- Since the data is already used for experiments, it is gone through the pre-processing task and requires no extensive pre-processing for the manual features-based approach. But for the LR-TED we must perform some replacement functions for the character within the data if required.
- In the above section, we have already estimated the time required by each expert to complete 2000 tweets. We assumed that the same time is required to manually process the tweets each step, considering the time requires for Training, Preparation of UI/Printouts, and processing through the participants.

TABLE 5.22: Processing Steps for State-of-the-art Approach with Estimated Time

Processing Steps	**Sub-Steps**	**Estimated Time (Hrs.)**
Step-1	Data Collection	0.5
Step-2	Pre-processing	0
Step-3	Event Type identification	25
Step-4	Labeling the data (Crowd-sourcing)	95
Step-5	Feature Extraction	120
Step-6	Tweet Labeling	0

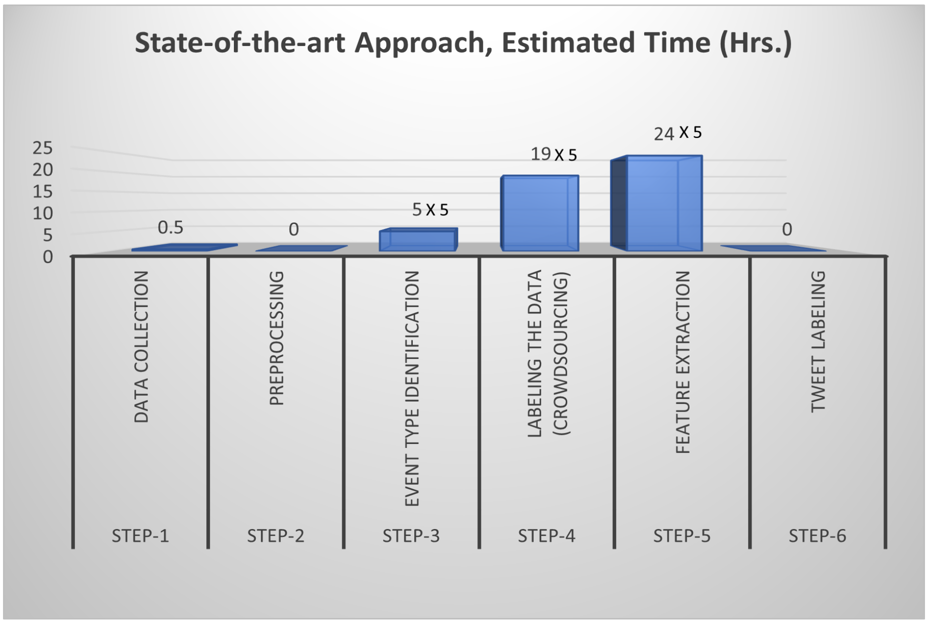

FIGURE 5.17: State-of-the-art Approach - Estimated Time (Hrs.)

In the following sub-sections, we will reflect the estimated time for the manual features-based approach, at each defined step. The details of the estimated time are given in the following Table-5.22.

The effort estimation in-terms of required time are presented in Hours. From the above table it is evident that most of the time consumption relates to steps 3, 4, & 5, naming Event type identification, labeling of data using crowd-sourcing, and feature extraction. The data is illustrated in the Figure-5.17.

For the proposed LR-TED approach, we further break down the required time estimation, wherever required. The details of the estimated time required by the

TABLE 5.23: Processing Steps for Proposed LR-TED Approach with Estimated Time

Processing Steps	**Sub-Steps**	**Estimated Time (Hrs.)**
Step-1: Data Collection	Download/Access the dataset	0.25
Step-1: Data Collection	Import/Conversion of data to the required format	0.25
Step-2: Pre-processing	Review/Replacement for data cleaning	0.5
Step-3: Event Type Analysis	Not Required	0
Step-4: Labeling the data (Crowd-sourcing)	Not Required	0
Step-5: Feature Extraction	Processing through Stanford CoreNLP	0.5
Step-5: Feature Extraction	ETL activities to shift the CoreNLP output to SQL Server	0.2
Step-5: Feature Extraction	Feature Extraction Queries Execution	0.75
Step-5: Feature Extraction	Result formation from SQL Server	0.05
Step-6: Tweet Labeling	Execution of Labeling Queries and Tools	0.25

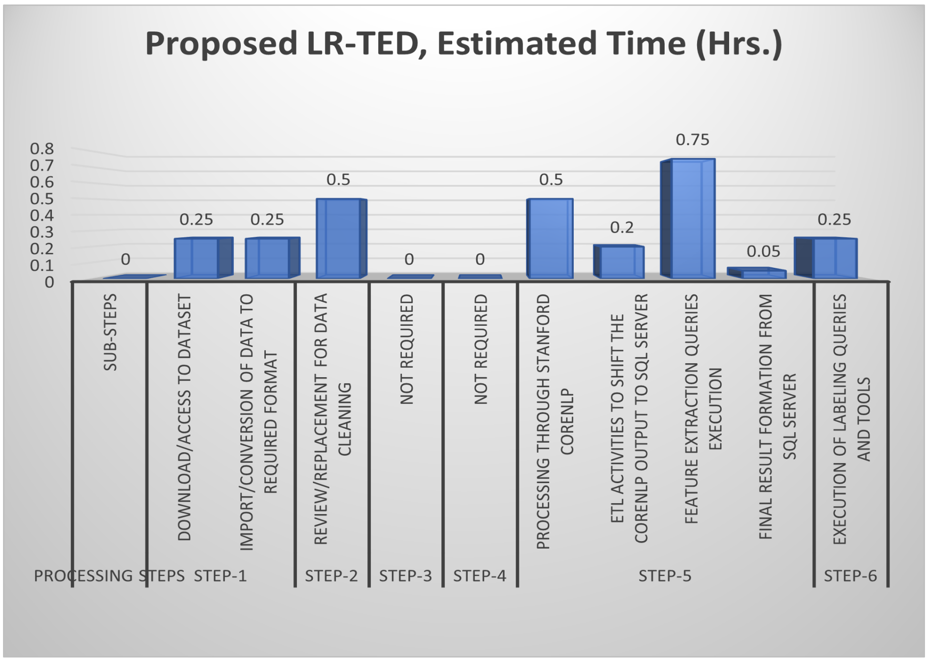

FIGURE 5.18: LR-TED Approach - Estimated Time (Hrs.)

proposed LR-TED approach at each step with details of all sub-task are given in Table-5.23. The data is illustrated in the Figure-5.18.

The time required to collect the data using twitter API is not included for both the approaches. The data downloading time is calculated only for the downloading

TABLE 5.24: Comparison of LR-TED & State-of-the-art Approach, Effort Estimation - Required Time

Processing Steps	**State-of-the-art Approach (Hrs.)**	**LR-TED (Hrs.)**
Step-1: Data Collection	0.5	0.5
Step-2: Pre-processing	0	0.5
Step-3: Event Type Analysis	25	0
Step-4: Labeling the data (Crowd-sourcing)	95	0
Step-5: Feature Extraction	120	1.5
Step-6: Tweet Labeling	0	0.25

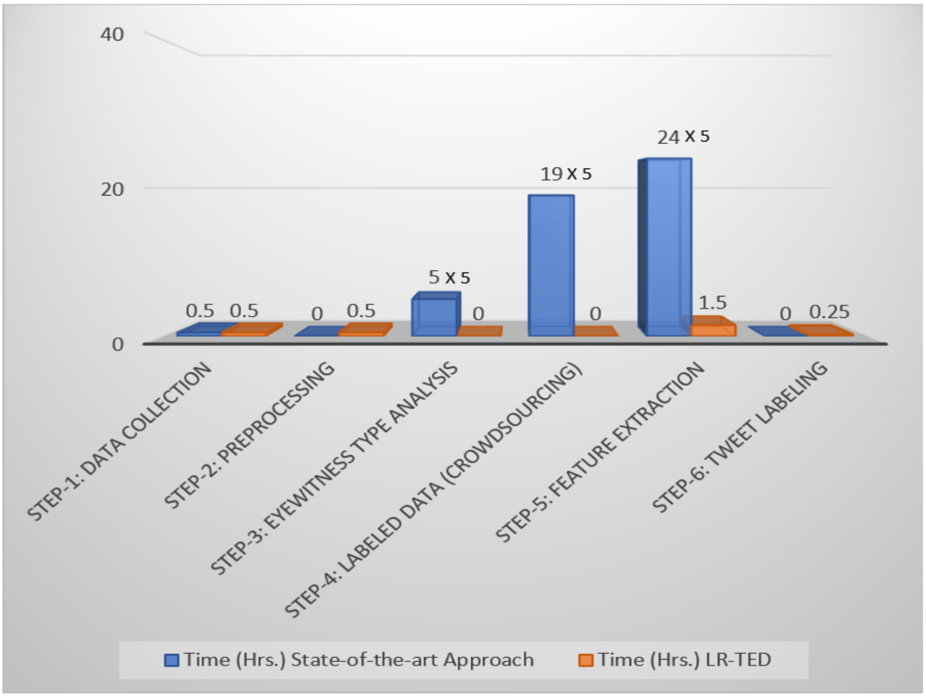

FIGURE 5.19: Comparison of Effort Estimation - Required Time (Hrs.)

of selected dataset as used by the manual features-based approach, considering some ETL tasks to be performed on the downloaded data prior to be utilized by LR-TED or state-of-the-art approach.

To comparison of both, State-of-the-art, and LR-TED approaches for effort estimation in terms of required time are presented in the following Table-5.24.

The comparison is further shown as a graph to visually present the true comparison between both approaches in Figure-5.19.

Table 5.25: Comparison of LR-TED & State-of-the-art Approach, Effort Estimation - Required Resources

Resource Type	**State-of-the-art Approach**	**LR-TED**
IT Person (Data Extraction) Level-2	1	1
Trainer Level-2	1	1
Domain Experts Level-4	3	0
Data Analysis Team Level-3	2	0
Resources to Annotate/Label Level-2	50	0

The above graph illustrates the comparison of both approaches. It is evident that our proposed LR-TED approach outperforms the manual features-based approach in terms of time-based effort estimation.

5.5.4 Human Resource & Cost - Effort Estimation

The measures of effort estimation to complete a task includes time, human resource, and cost. In the above section, we already have discussed the efforts in terms of time. In this section, we will discuss the required efforts in terms of HR and other related costs. To implement the manual features-based approach, more HR efforts are required in comparison to the proposed LR-TED.

The minimum cost to hire a resource for the content-related task is USD 100/day[2]. Amazon crowd-sourcing cost 243.36 USD per person for a dataset object. As estimated above, to manually process 2000 tweets we require 120 hours/person. Approx. 15 days (8 Hrs. each), and for crowd-sourcing, we require multiple resources to work on the same task.

For the proposed LR-TED approach we only require one IT-Person who can finish the task in 1.5 hours. Other than the data extraction when involved. The following Table-5.25 depicts the estimated number of resources required by both approaches.

We have the following assumptions for the above table for resource levels. Level-1 = Basic, Level-2 = Intermediate, Level-3 = Professional, Level-4 = Expert. We

[2] https://contently.net/rates-database/rates/

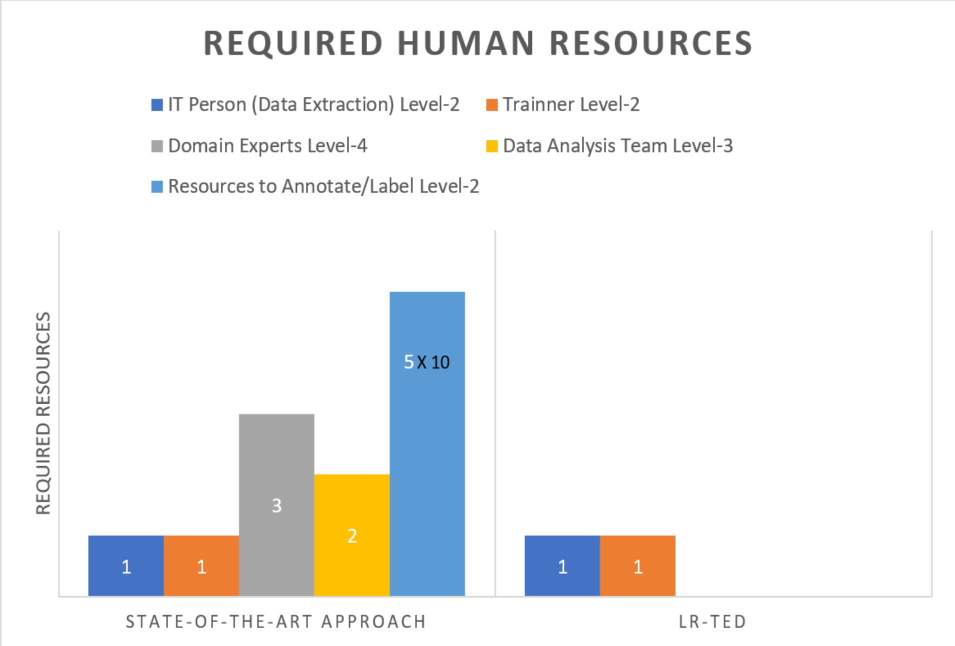

FIGURE 5.20: Comparison of Effort Estimation - Required Resources

TABLE 5.26: Rate-Card - Required Resources

Task/Resource Type	**Cost Unit**	**Price/Rate**
Text Classification (Expert)	Per Record Cost	0.08
Named Entity Recognition (NER)	Per Record Cost	0.08
IT Person	Per Hr. Cost	1.23
Trainer	Per Hr. Cost	0.75
Data Analyst	Per Hr. Cost	1.75

can have disagreements on the level definition and rates, even though the required resources can clearly define the difference. The comparison is illustrated in the Figure-5.20.

The tentative rate-card for crowd-sourcing and other required services as shown in Table-5.26. The estimates are taken from Amazon and Contently.

5.6 Summary

This chapter presented overall results and the adopted methodology by LR-TED. The results generated by the proposed LR-TED approach for various experiments were then compared with the state-of-the-art approach [37], using the annotated gold-standard dataset. The chapter comprehensively discussed the available dataset and the implementation of tagging and parsing techniques as discussed in Chapter-3 and 4. The grammar rules defined in Chapter-4 are applied to the focused dataset of 8,000 tweets and classified the tweets into eyewitness, non-eyewitness, and unknown classes and described in Chapter-3 and 4. In this chapter, the results are generated for each experiment discussed in chapter-3 for the proposed LR-TED approach. The results are briefly discussed for disaster events such as; earthquakes, floods, hurricanes, and wildfire. The results achieved by each model are compared among other selected models. The results are then further compared and comprehensively discussed with the results available in the gold-standard dataset of the manual features-based approach [37]. The findings and observations for each experiment are comprehensively discussed in this chapter. From the experiments, it is evident that the supervised models and deep learning approaches achieved good results. The results showed that LR-TED achieved the best scores for eyewitness, non-eyewitness, and unknown classes respectively. Both, LR-TED and State-of-the-art approaches are then comprehensively discussed in terms of required efforts to be adopted for new event types. It is evident that LR-TED approach is cost-effective in-terms of required time, human resource, and financial. The chapter has discussed that this research work achieved comparative results considering that LR-TED is fully automatic and discourages the use of the manual dictionary as used by the manual features-based approach.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Twitter is deliberated as a potential news breaker social media platform. Users across the world share the information with their followers concerning the incidents they observe, particularly disastrous events such as earthquakes, floods, wildfires, and hurricanes. The identification of eyewitness tweets about the incident from potentially millions of tweets shared by Twitter users is considered a challenging task. However, on the other hand, to understand the true intensity of the event, the information shared by an eyewitness is vital. The emergency response services can use such credible information for disaster management systems.

We critically analyzed more than one hundred recent papers from the literature and investigated that the researchers in this domain have studied Twitter for advertising, disaster and emergency alerts and response systems, targeting news recommendations. The huge importance of the eyewitnesses attracted the research community to identify them from the tweets. Several efforts in this regard were made to identify the eyewitnesses such as the twitter user accounts, and their relational user networks. Truelove et al. presented the technique of using a predefined list of query words, to identify the target tweet. Author check for the presense of query words in the tweet content during the pre-processing phase. For identification of eyewitness tweet, no characteristics were presented. In comparison to Truelove et al., Doggett et al. presented the list of linguistic features to identify the eyewitness tweets. Similarly, Fang et al. exploited the hybrid technique by combining the metadata features, such as "type of application used for reporting", with linguistic features for the identification of eyewitness tweets. Fang et al. adopted the linguistic and stylistic features to identify the eyewitness tweets. The use of the hybrid approach is also exploited by Tanev et al. The author adopted stylistic and linguistic features with Twitter's metadata to identify the eyewitness.

The recent study presented by Zahra et al. [37] (the manual feature-based approach), identified thirteen characteristics from the tweet content to identify the eyewitness tweet. However, they employed the manually created static dictionary in their methodology. A critical analysis of the studies of the domain, reveals two major limitations of the manual feature-based approach of using the manually created static dictionaries. Firstly, lack of automation for the extraction of feature-words and involvement of domain-experts is required to update the static dictionary. The approach lacks the strength of a generic approach as it needs a separate dictionary for different event types. Secondly, the author dropped the implementation of some identified features.

This research has utilized the language structure, linguistics, and word relation to achieve; (i) automatic extraction of feature-words by creating of grammar rules and propose a generic approach that can automatically identify features from the content of any disaster type by discouraging the use of a static dictionary of each disaster type; and (ii) implementing all identified features that were dropped by the manual feature-based approach.

Identifying the eyewitnesses from Twitter is useful to many real life applications. It is an active research area, and diverse methods are proposed in the literature for this task. However, the manual feature-based approach requires manual intervention for extracting the feature set, which limits the applicability of these approaches on large data streams. Therefore, this research proposed a method that utilizes the language structure, linguistics, and word relation to automatically extract the feature set. This research proposed a generic approach that covers different disaster types. The proposed LR-TED approach is evaluated on the benchmark dataset, and results are compared with the manual feature-based approach in terms of the evaluation parameters (precision, recall, and f-measure). The evaluation results show that the LR-TED achieved an f-score of 0.93 for the earthquake disaster events category, which is better than the manual feature-based approach. The implication of the proposed technique can be recognized when we have to process potentially millions of tweets to respond to the event on earliest. We critically estimated the efforts required by LR-TED and the state-of-the-art approach for eyewitness tweets identification. From effort estimation, it is evident that the proposed LR-TED approach is effective in-terms of required time, human resource, and financial costs.

6.2 Implication

Reviewing and employing the idea of automatic identification of eyewitness messages by using linguistic features, language patterns & structure, and the words dependency relationship in a sentence on twitter, without human involvement, to identify eyewitness by the proposed technique, is a novel contribution of this research. The implication of this research work can be a potential contribution to various applications in this domain such as Disaster Management Systems, Disaster or Emergency Alert Systems, Emergency Response for Institutes and Agencies.

6.3 Limitation

This research proposed grammar-based rules for ten out of fourteen identified features. The identified features such as; "First-person pronouns and adjectives", "Words indicating perceptual senses", "Expletives", and "Exclamation and question marks", are constrained in terms of implementation by only using the predefined list of words. For example, for "Expletives" words no grammar rules can be developed, because of its scope, and we have to limit our implementation methodology to static dictionaries rather than adopting the grammar rules.

6.4 Future Work

This research has opened a debate of how useful these grammar-rules can be exploited for the extraction of feature term for identification of eyewitness, from the social media platforms. In the future, the use of manually created static dictionaries is not required for the identification of the important feature terms from the text for eyewitness identification. Furthermore, the proposed LR-TED approach can be tested for new disaster types, and needs to be evaluated on different social media platforms, other than Twitter and for various disaster types.

Bibliography

[1] M. Imran, C. Castillo, F. Diaz, and S. Vieweg, "Processing social media messages in mass emergency: A survey," *ACM Computing Surveys (CSUR)*, vol. 47, no. 4, pp. 1–38, 2015.

[2] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, "Microblogging during two natural hazards events: what twitter may contribute to situational awareness," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2010, pp. 1079–1088.

[3] K. Kaur and K. S. Dhindsa, "Study of existing design architectures and frameworks in twitter," in *2019 International Conference on Automation, Computational and Technology Management (ICACTM).* IEEE, 2019, pp. 334–337.

[4] J. Horrigan, *Online shopping: Internet users like the convenience but worry about the security of their financial information.* Pew Internet & American Life Project Washington, DC, 2008.

[5] S. Banihashemi, J. Li, and A. Abhari, "Scalable machine learning algorithms for a twitter followee recommender system," in *2019 Spring Simulation Conference (SpringSim).* IEEE, 2019, pp. 1–8.

[6] J. S. Fu, "Unpacking the influence of informational, organizational, and structural factors on the longitudinal change of the npo follower-followee network on twitter," *International Journal of Communication*, vol. 13, p. 24, 2019.

[7] M. G. Armentano, D. Godoy, and A. Amandi, "Towards a followee recommender system for information seeking users in twitter," in *Proceedings of the Workshop on Semantic Adaptive Social Web (SASWeb 2011). CEUR Workshop Proceedings*, vol. 730, 2011, pp. 27–38.

[8] M. G. Armentano, D. Godoy, and A. A. Amandi, "A topology-based approach for followees recommendation in twitter." in *ITWP@ IJCAI*, 2011.

[9] M. G. Armentano, D. Godoy, and A. Amandi, "Recommending information sources to information seekers in twitter," in *International workshop on social web mining*, 2011.

[10] J. P. Carpenter and D. G. Krutka, "How and why educators use twitter: A survey of the field," *Journal of research on technology in education*, vol. 46, no. 4, pp. 414–434, 2014.

[11] E. Brown, "Teachers take to twitter to improve craft and commiserate," *Washington Post*, vol. 22, 2012.

[12] T. Luo, S. J. Shah, and H. Cromptom, "Using twitter to support reflective learning in an asynchronous online course," *Australasian Journal of Educational Technology*, vol. 35, no. 3, 2019.

[13] D. P. Domizi, "Microblogging to foster connections and community in a weekly graduate seminar course," *TechTrends*, vol. 57, no. 1, pp. 43–51, 2013.

[14] V. Abella-García, V. Delgado-Benito, V. Ausín-Villaverde, and D. Hortigüela-Alcalá, "To tweet or not to tweet: Student perceptions of the use of twitter on an undergraduate degree course," *Innovations in Education and Teaching International*, vol. 56, no. 4, pp. 402–411, 2019.

[15] R. Kimmons, J. P. Carpenter, G. Veletsianos, and D. G. Krutka, "Mining social media divides: an analysis of k-12 us school uses of twitter," *Learning, media and technology*, vol. 43, no. 3, pp. 307–325, 2018.

[16] F. Rosell-Aguilar, "Twitter: A professional development and community of practice tool for teachers," *Journal of Interactive Media in Education*, vol. 1, 2018.

[17] D. Krutka, "Social media as a catalyst for convergence culture: Immersing pre-service social studies teachers in the social media terrain," *Digital social studies*, pp. 271–302, 2014.

[18] V. Kharde, P. Sonawane *et al.*, "Sentiment analysis of twitter data: a survey of techniques," *arXiv preprint arXiv:1601.06971*, 2016.

[19] A. Kumar and A. Jaiswal, "Systematic literature review of sentiment analysis on twitter using soft computing techniques," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 1, p. e5107, 2020.

[20] M. Aivazoglou, A. O. Roussos, D. Margaris, C. Vassilakis, S. Ioannidis, J. Polakis, and D. Spiliotopoulos, "A fine-grained social network recommender system," *Social Network Analysis and Mining*, vol. 10, no. 1, p. 8, 2020.

[21] P. Gamallo and M. Garcia, "Citius: A naivebayes strategy for sentiment analysis on english tweets," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014*. Citeseer, 2014.

[22] S. Liu, F. Li, F. Li, X. Cheng, and H. Shen, "Adaptive co-training svm for sentiment classification on tweets," in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 2013, pp. 2079–2088.

[23] R. Xia, C. Zong, and S. Li, "Ensemble of feature sets and classification algorithms for sentiment classification," *Information sciences*, vol. 181, no. 6, pp. 1138–1152, 2011.

[24] H. Mao, S. Counts, and J. Bollen, "Predicting financial markets: Comparing survey, news, twitter and search engine data," *arXiv preprint arXiv:1112.1051*, 2011.

[25] F. Atefeh and W. Khreich, "A survey of techniques for event detection in twitter," *Computational Intelligence*, vol. 31, no. 1, pp. 132–164, 2015.

[26] X. Wang, M. S. Gerber, and D. E. Brown, "Automatic crime prediction using events extracted from twitter posts," in *International conference on social computing, behavioral-cultural modeling, and prediction.* Springer, 2012, pp. 231–238.

[27] S. Goorha and L. Ungar, "Discovery of significant emerging trends," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2010, pp. 57–64.

[28] H. Becker, M. Naaman, and L. Gravano, "Selecting quality twitter content for events," in *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[29] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*, 2011, pp. 675–684.

[30] J. Hurlock and M. L. Wilson, "Searching twitter: Separating the tweet from the chaff," in *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[31] K. Lee, B. D. Eoff, and J. Caverlee, "Seven months with the devils: A long-term study of content polluters on twitter," in *Fifth international AAAI conference on weblogs and social media*, 2011.

[32] C. Comito, A. Forestiero, and C. Pizzuti, "Bursty event detection in twitter streams," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 13, no. 4, pp. 1–28, 2019.

[33] M. Hasan, M. A. Orgun, and R. Schwitter, "Real-time event detection from the twitter data stream using the twitternews+ framework," *Information Processing & Management*, vol. 56, no. 3, pp. 1146–1165, 2019.

[34] S. Malmasi and M. Dras, "Location mention detection in tweets and microblogs," in *Conference of the Pacific Association for Computational Linguistics*. Springer, 2015, pp. 123–134.

[35] C. Bothorel, N. Lathia, R. Picot-Clemente, and A. Noulas, "Location recommendation with social media data," in *Social Information Access*. Springer, 2018, pp. 624–653.

[36] A. Kumar and J. P. Singh, "Location reference identification from tweets during emergencies: A deep learning approach," *International journal of disaster risk reduction*, vol. 33, pp. 365–375, 2019.

[37] K. Zahra, M. Imran, and F. O. Ostermann, "Automatic identification of eyewitness messages on twitter during disasters," *Information processing & management*, vol. 57, no. 1, p. 102107, 2020.

[38] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 591–600.

[39] M. Truelove, M. Vasardani, and S. Winter, "Testing a model of witness accounts in social media," in *Proceedings of the 8th workshop on geographic information retrieval*, 2014, pp. 1–8.

[40] E. Doggett and A. Cantarero, "Identifying eyewitness news-worthy events on twitter," in *Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media*, 2016, pp. 7–13.

[41] R. Fang, A. Nourbakhsh, X. Liu, S. Shah, and Q. Li, "Witness identification in twitter," in *Proceedings of the Fourth International Workshop on Natural Language Processing for Social Media*, 2016, pp. 65–73.

[42] H. Tanev, V. Zavarella, and J. Steinberger, "Monitoring disaster impact: detecting micro-events and eyewitness reports in mainstream and social media." in *ISCRAM*, 2017.

[43] C. Li and A. Sun, "Extracting fine-grained location with temporal awareness in tweets: A two-stage approach," *Journal of the Association for Information Science and Technology*, vol. 68, no. 7, pp. 1652–1670, 2017.

[44] D. Inkpen, J. Liu, A. Farzindar, F. Kazemi, and D. Ghazi, "Location detection and disambiguation from twitter messages," *Journal of Intelligent Information Systems*, vol. 49, no. 2, pp. 237–253, 2017.

[45] X. Zheng, J. Han, and A. Sun, "A survey of location prediction on twitter," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 9, pp. 1652–1671, 2018.

[46] Y. Kryvasheyeu, H. Chen, N. Obradovich, E. Moro, P. Van Hentenryck, J. Fowler, and M. Cebrian, "Rapid assessment of disaster damage using social media activity," *Science advances*, vol. 2, no. 3, p. e1500779, 2016.

[47] C. Allen, "A resource for those preparing for and responding to natural disasters, humanitarian crises, and major healthcare emergencies," *Journal of evidence-based medicine*, vol. 7, no. 4, pp. 234–237, 2014.

[48] C. A. Amaratunga, "Building community disaster resilience through a virtual community of practice (vcop)," *International Journal of Disaster Resilience in the Built Environment*, 2014.

[49] E. Schnebele and G. Cervone, "Improving remote sensing flood assessment using volunteered geographical data," *Natural Hazards Earth System Science*, vol. 13, pp. 669–677, 03 2013.

[50] J. Teevan, D. Ramage, and M. R. Morris, "# twittersearch: a comparison of microblog search and web search," in *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011, pp. 35–44.

[51] Z. J. Auter and J. A. Fine, "Social media campaigning: Mobilization and fundraising on facebook," *Social Science Quarterly*, vol. 99, no. 1, pp. 185–200, 2018.

[52] G. A. Ruz, P. A. Henríquez, and A. Mascareño, "Sentiment analysis of twitter data during critical events through bayesian networks classifiers," *Future Generation Computer Systems*, vol. 106, pp. 92–104, 2020.

[53] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electronics*, vol. 9, no. 3, p. 483, 2020.

[54] B. Bhavitha, A. P. Rodrigues, and N. N. Chiplunkar, "Comparative study of machine learning techniques in sentimental analysis," in *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT).* IEEE, 2017, pp. 216–221.

[55] D. Pinto, A. McCallum, X. Wei, and W. B. Croft, "Table extraction using conditional random fields," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, 2003, pp. 235–242.

[56] M. R. Huq, A. Ali, and A. Rahman, "Sentiment analysis on twitter data using knn and svm," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 8, no. 6, pp. 19–25, 2017.

[57] S. Soni and A. Sharaff, "Sentiment analysis of customer reviews based on hidden markov model," in *Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015)*, 2015, pp. 1–5.

[58] S. Haider, M. Tanvir Afzal, M. Asif, H. Maurer, A. Ahmad, and A. Abuarqoub, "Impact analysis of adverbs for sentiment classification on twitter product reviews," *Concurrency and Computation: Practice and Experience*, 2018.

[59] V. Malik and A. Kumar, "Sentiment analysis of twitter data using naive bayes algorithm," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 6, no. 4, pp. 120–125, 2018.

[60] H. Wu, J. Li, and J. Xie, "Maximum entropy-based sentiment analysis of online product reviews in chinese," in *Automotive, Mechanical and Electrical Engineering*. CRC Press, 2017, pp. 559–562.

[61] A. L. Firmino Alves, C. D. S. Baptista, A. A. Firmino, M. G. d. Oliveira, and A. C. d. Paiva, "A comparison of svm versus naive-bayes techniques for sentiment analysis in tweets: a case study with the 2013 fifa confederations cup," in *Proceedings of the 20th Brazilian Symposium on Multimedia and the Web*, 2014, pp. 123–130.

[62] M. A. Nielsen, *Neural networks and deep learning*. Determination press San Francisco, CA, USA:, 2015, vol. 2018.

[63] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1253, 2018.

[64] D. Britz, "Recurrent neural networks tutorial, part 1–introduction to rnns," *online] www. wildml. com, Available at: http://www. wildml. com/2015/09/recurrent-neural-networkstutorial-part-1-introduction-to-rnns/.[Accessed 02 Jun. 2020]*, 2015.

[65] A. C. Pandey, D. S. Rajpoot, and M. Saraswat, "Twitter sentiment analysis using hybrid cuckoo search method," *Information Processing & Management*, vol. 53, no. 4, pp. 764–779, 2017.

[66] M. Daniel, R. F. Neves, and N. Horta, "Company event popularity for financial markets using twitter and sentiment analysis," *Expert Systems with Applications*, vol. 71, pp. 111–124, 2017.

[67] Y. Ruan, A. Durresi, and L. Alfantoukh, "Using twitter trust network for stock market analysis," *Knowledge-Based Systems*, vol. 145, pp. 207–218, 2018.

[68] S. Y. Yang, S. Y. K. Mo, and A. Liu, "Twitter financial community sentiment and its predictive relationship to stock market movement," *Quantitative Finance*, vol. 15, no. 10, pp. 1637–1656, 2015.

[69] H. Chen, P. De, Y. J. Hu, and B.-H. Hwang, "Wisdom of crowds: The value of stock opinions transmitted through social media," *The Review of Financial Studies*, vol. 27, no. 5, pp. 1367–1403, 2014.

[70] E. J. Ruiz, V. Hristidis, C. Castillo, A. Gionis, and A. Jaimes, "Correlating financial time series with micro-blogging activity," in *Proceedings of the fifth ACM international conference on Web search and data mining*, 2012, pp. 513–522.

[71] T. Renault, "Intraday online investor sentiment and return patterns in the us stock market," *Journal of Banking & Finance*, vol. 84, pp. 25–40, 2017.

[72] A. Groß-Klußmann, S. König, and M. Ebner, "Buzzwords build momentum: Global financial twitter sentiment and the aggregate stock market," *Expert Systems with Applications*, vol. 136, pp. 171–186, 2019.

[73] S. P. Greenhalgh and M. J. Koehler, "28 days later: Twitter hashtags as just in time teacher professional development," *TechTrends*, vol. 61, no. 3, pp. 273–281, 2017.

[74] J. Carpenter, T. Tani, S. Morrison, and J. Keane, "Exploring the landscape of educator professional activity on twitter: an analysis of 16 education-related twitter hashtags," *Professional Development in Education*, pp. 1–22, 2020.

[75] V. Subramaniyaswamy, R. Logesh, M. Abejith, S. Umasankar, and A. Umamakeswari, "Sentiment analysis of tweets for estimating criticality and security of events," in *Improving the Safety and Efficiency of Emergency Services: Emerging Tools and Technologies for First Responders*. IGI Global, 2020, pp. 293–319.

[76] C. Li and A. Sun, "Fine-grained location extraction from tweets with temporal awareness," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 2014, pp. 43–52.

[77] B. Cao, F. Chen, D. Joshi, and S. Y. Philip, "Inferring crowd-sourced venues for tweets," in *2015 IEEE International Conference on Big Data (Big Data)*. IEEE, 2015, pp. 639–648.

[78] O. Ajao, J. Hong, and W. Liu, "A survey of location inference techniques on twitter," *Journal of Information Science*, vol. 41, no. 6, pp. 855–864, 2015.

[79] O. Oh, M. Agrawal, and H. R. Rao, "Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises," *Mis Quarterly*, pp. 407–426, 2013.

[80] B. Haworth and E. Bruce, "A review of volunteered geographic information for disaster management," *Geography Compass*, vol. 9, no. 5, pp. 237–250, 2015.

[81] P. Meier, "Crisis mapping in action: How open source software and global volunteer networks are changing the world, one map at a time," *Journal of Map & Geography Libraries*, vol. 8, no. 2, pp. 89–100, 2012.

[82] H. Purohit, C. Castillo, F. Diaz, A. Sheth, and P. Meier, "Emergency-relief coordination on social media: Automatically matching resource requests and offers," *First Monday*, vol. 19, no. 1, 2014.

[83] M. Imran, C. Castillo, J. Lucas, P. Meier, and S. Vieweg, "Aidr: Artificial intelligence for disaster response," in *Proceedings of the 23rd International Conference on World Wide Web*, 2014, pp. 159–162.

[84] R. Snow, B. Oconnor, D. Jurafsky, and A. Y. Ng, "Cheap and fast–but is it good? evaluating non-expert annotations for natural language tasks," in *Proceedings of the 2008 conference on empirical methods in natural language processing*, 2008, pp. 254–263.

[85] F. Abel, C. Hauff, G.-J. Houben, R. Stronkman, and K. Tao, "Twitcident: fighting fire with information from social web streams," in *Proceedings of the 21st International Conference on World Wide Web*, 2012, pp. 305–308.

[86] S. Kumar, G. Barbier, M. A. Abbasi, and H. Liu, "Tweettracker: An analysis tool for humanitarian and disaster relief," in *Fifth international AAAI conference on weblogs and social media*, 2011.

[87] D. Thom, R. Krüger, T. Ertl, U. Bechstedt, A. Platz, J. Zisgen, and B. Volland, "Can twitter really save your life? a case study of visual social media analytics for situation awareness," in *2015 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 2015, pp. 183–190.

[88] M. Imran, P. Mitra, and J. Srivastava, "Cross-language domain adaptation for classifying crisis-related short messages," *arXiv preprint arXiv:1602.05388*, 2016.

[89] B. Takahashi, E. C. Tandoc Jr, and C. Carmichael, "Communicating on twitter during a disaster: An analysis of tweets during typhoon haiyan in the philippines," *Computers in Human Behavior*, vol. 50, pp. 392–398, 2015.

[90] M. F. Goodchild and L. Li, "Assuring the quality of volunteered geographic information," *Spatial statistics*, vol. 1, pp. 110–120, 2012.

[91] M. Zook, M. Graham, T. Shelton, and S. Gorman, "Volunteered geographic information and crowdsourcing disaster relief: a case study of the haitian earthquake," *World Medical & Health Policy*, vol. 2, no. 2, pp. 7–33, 2010.

[92] F. Ostermann and L. Spinsanti, "Context analysis of volunteered geographic information from social media networks to support disaster management: A case study on forest fires," *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, vol. 4, no. 4, pp. 16–37, 2012.

[93] P. M. Landwehr and K. M. Carley, "Social media in disaster relief," in *Data mining and knowledge discovery for big data*. Springer, 2014, pp. 225–257.

[94] A. Swasy, *How journalists use Twitter: The changing landscape of US newsrooms.* Lexington Books, 2016.

[95] A. Swasey, "I studied how journalists used twitter for two years. heres what i learned," *Poynter. org*, vol. 22, 2017.

[96] M. Hu, S. Liu, F. Wei, Y. Wu, J. Stasko, and K.-L. Ma, "Breaking news on twitter," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 2751–2754.

[97] A. Mayfield, *Me and my web shadow: How to manage your reputation online.* A&C Black, 2010.

[98] M. Truelove, M. Vasardani, and S. Winter, "Towards credibility of microblogs: characterising witness accounts," *GeoJournal*, vol. 80, no. 3, pp. 339–359, 2015.

[99] N. Diakopoulos, M. De Choudhury, and M. Naaman, "Finding and assessing social media information sources in the context of journalism," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012, pp. 2451–2460.

[100] A. Olteanu, S. Vieweg, and C. Castillo, "What to expect when the unexpected happens: Social media communications across crises," in *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, 2015, pp. 994–1009.

[101] S. Kumar, F. Morstatter, R. Zafarani, and H. Liu, "Whom should i follow? identifying relevant users during crises," in *Proceedings of the 24th ACM conference on hypertext and social media*, 2013, pp. 139–147.

[102] F. Morstatter, N. Lubold, H. Pon-Barry, J. Pfeffer, and H. Liu, "Finding eyewitness tweets during crises," *arXiv preprint arXiv:1403.1773*, 2014.

[103] K. Zahra, F. O. Ostermann, and R. S. Purves, "Geographic variability of twitter usage characteristics during disaster events," *Geo-spatial information science*, vol. 20, no. 3, pp. 231–240, 2017.

[104] D. Jurafsky and J. H. Martin, "Speech and language processing (draft). october 2019," *URL https://web. stanford. edu/˜ jurafsky/slp3*, 2019.

[105] J. R. Finkel, T. Grenager, and C. Manning, "Incorporating non-local information into information extraction systems by gibbs sampling," in *Proceedings of the 43rd annual meeting on association for computational linguistics*. Association for Computational Linguistics, 2005, pp. 363–370.

[106] T. Gui, Q. Zhang, H. Huang, M. Peng, and X.-J. Huang, "Part-of-speech tagging for twitter with adversarial neural networks," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 2411–2420.

[107] I. C. on Strategic Innovative Marketing, *Strategic Innovative Marketing: 5th IC-SIM, Athens, Greece 2016*. Springer, 2017.

[108] M.-C. De Marneffe and C. D. Manning, "Stanford typed dependencies manual," Technical report, Stanford University, Tech. Rep., 2008.

[109] C. B. Rjeily, G. Badr, A. H. El Hassani, and E. Andres, "Medical data mining for heart diseases and the future of sequential mining in medical field," in *Machine Learning Paradigms*. Springer, 2019, pp. 71–99.

[110] S. Passova, A. Ladizginsky, and S. Passov, "Systems and methods for business process modelling," Oct. 8 2019, uS Patent 10,438,166.

[111] D. B. Renita and C. S. Christopher, "Novel real time content based medical image retrieval scheme with gwo-svm," *Multimedia Tools and Applications*, pp. 1–17, 2020.

[112] S. Asht and R. Dass, "Pattern recognition techniques: A review," *International Journal of Computer Science and Telecommunications*, vol. 3, no. 8, pp. 25–29, 2012.

[113] B. Xu, X. Guo, Y. Ye, and J. Cheng, "An improved random forest classifier for text categorization," *JCP*, vol. 7, pp. 2913–2920, 2012.

[114] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Information*, vol. 10, no. 4, p. 150, 2019.

[115] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, and S. Iyengar, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1–36, 2018.

[116] A. Desai and S. Rai, "Analysis of machine learning algorithms using weka," 02 2013.

[117] K. Barua, P. Chakrabarti, A. Panwar, and A. Ghosh, "A predictive analytical model in education scenario based on critical thinking using weka," *International Journal of Technology Research and Management*, vol. 5, no. 3, 2018.

[118] P. Sharma, "Comparative analysis of various clustering algorithms using weka," *International Research Journal of Engineering and Technology (IRJET)*, vol. 2, no. 04, 2015.

[119] P. Valdiviezo-Diaz, F. Ortega, E. Cobos, and R. Lara-Cabrera, "A collaborative filtering approach based on naïve bayes classifier," *IEEE Access*, vol. 7, pp. 108 581–108 592, 2019.

[120] A. P. Wibawa, A. C. Kurniawan, R. P. Adiperkasa, S. M. Putra, S. A. Kurniawan, Y. R. Nugraha *et al.*, "Naïve bayes classifier for journal quartile classification," *International Journal of Recent Contributions from Engineering, Science & IT (iJES)*, vol. 7, no. 2, pp. 91–99, 2019.

[121] M. Sivaram, E. L. Lydia, I. V. Pustokhina, D. A. Pustokhin, M. Elhoseny, G. P. Joshi, and K. Shankar, "An optimal least square support vector machine based earnings prediction of blockchain financial products," *IEEE Access*, vol. 8, pp. 120 321–120 330, 2020.

[122] F. Karimi, S. Sultana, A. S. Babakan, and S. Suthaharan, "An enhanced support vector machine model for urban expansion prediction," *Computers, Environment and Urban Systems*, vol. 75, pp. 61–75, 2019.

[123] Y. Chen, T. A. Lasko, Q. Mei, J. C. Denny, and H. Xu, "A study of active learning methods for named entity recognition in clinical text," *Journal of biomedical informatics*, vol. 58, pp. 11–18, 2015.

[124] M. Marrero, J. Urbano, S. Sánchez-Cuadrado, J. Morato, and J. M. Gómez-Berbís, "Named entity recognition: fallacies, challenges and opportunities," *Computer Standards & Interfaces*, vol. 35, no. 5, pp. 482–489, 2013.

[125] C. Eickhoff and A. P. de Vries, "Increasing cheat robustness of crowdsourcing tasks," *Information retrieval*, vol. 16, no. 2, pp. 121–137, 2013.

[126] A. Jain, A. D. Sarma, A. Parameswaran, and J. Widom, "Understanding workers, developing effective tasks, and enhancing marketplace dynamics: A study of a large crowdsourcing marketplace," *arXiv preprint arXiv:1701.06207*, 2017.

[127] M. Hirth, S. Scheuring, T. Hoßfeld, C. Schwartz, and P. Tran-Gia, "Predicting result quality in crowdsourcing using application layer monitoring," in *2014 IEEE Fifth International Conference on Communications and Electronics (ICCE)*. IEEE, 2014, pp. 510–515.

[128] I. Gomes, R. Correia, J. Ribeiro, and J. Freitas, "Effort estimation in named entity tagging tasks," in *Proceedings of the 12th Language Resources and Evaluation Conference*, 2020, pp. 298–306.

[129] B. Settles and M. Craven, "An analysis of active learning strategies for sequence labeling tasks," in *proceedings of the 2008 conference on empirical methods in natural language processing*, 2008, pp. 1070–1079.

[130] K. Rayner, E. R. Schotter, M. E. Masson, M. C. Potter, and R. Treiman, "So much to read, so little time: How do we read, and can speed reading

help?" *Psychological Science in the Public Interest*, vol. 17, no. 1, pp. 4–34, 2016.