IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, April 2024

Analysis of Damage Assessment Tweets During Disaster using Sentiment Analysis

Vardhan Reddy Dereddy[1], Chikoti Manisai[2], Bommagani Pavan[3], Mr. Nanda Kumar[4]
Students, Department of Electronics and Computer Engineering[1,2,3]
Asst. Professor and Internal Guide, Department of Electronics and Computer Engineering[4]
Sreenidhi Institute of Science and Technology, Hyderabad, India

Abstract: *This seems to be an abstract or summary of a paper on monitoring Twitter for damage assessments after a disaster. Using simple linear regression and Support Vector Regression methods for weighting and the random forest methodology for classification, the research provides a novel approach that makes use of low-level lexical characteristics, top-most frequency word features, and syntactic elements relevant to damage assessment. The accuracy of the suggested method for identifying damage assessment tweets is 94.62%, as measured across 14 typical disaster datasets for binary and multi-class categorization. Significant advancements were observed when comparing the proposed method to the state-of-the-art for both in-domain and cross-domain scenarios. The suggested method does not require labelled tweets or tweets of a specific disaster kind in order to be trained and implemented; instead, it can be trained on historical disaster datasets.*

Keywords: Twitter Disaster Damage Assessment Infrastructure damage social media

I. INTRODUCTION

This article explores how businesses and media outlets are increasingly turning to Twitter and other microblogging sites to gauge consumer sentiment. Several services now offer Twitter sentiment analysis, and it is rapidly becoming a useful tool in this context. However, little is known about how people's emotions come across in microblogging due to the informal nature of the language and the limited character count of each message. Microblogging poses new issues to the study of sentiment expression, which have been addressed in previous research on genres such as online reviews and news articles.

Sentiment analysis is discussed in this article; this is the method of analyzing written content to determine if the tone is positive, negative, or neutral. Businesses can utilize sentiment analysis for a variety of purposes, including gathering customer feedback, forecasting election results, and gleaning information from movie reviews. Traditional methods of sentiment analysis, such as the bag of words technique, incorrectly categorize phrases with negative because they ignore morphology and word order. Instead, we take into account the interplay between the set of words and employ a function to ascertain and aggregate the sentiment of each phrase. Naive Bayes, Maximum Entropy, Long Short-Term Memory (LSTM), and TensorFlow are only some of the sentiment analysis frameworks mentioned in the article. Sentiment analysis and text classification techniques are discussed in the literature review section.

Sentiment analysis, which is also known as Opinion Mining, is a subfield of Natural Language Processing (NLP) that focuses on analyzing text in order to detect the feelings, attitudes, and opinions that have been stated about a specific entity. This entity could be a person, an event, or a subject. Sentiment analysis is also sometimes referred to as opinion mining. Although a significant amount of research has been conducted in this area, the majority of the studies that have been conducted have focused mostly on the analysis of more formal and longer texts, such as reviews.

This work uses binary and multi-class classification methods to examine the issue of recognizing tweets about damage assessment during catastrophes, both for human and infrastructure damage. The authors present a new method based on weighted features, created with linear regression and Support Vector Regression (SVR), to capture the various categorical information connected to damage during disasters. High-frequency terms and more fundamental lexical and syntactic characteristics are included here. The authors test several many classifiers to find the one that works best with their specified features. They test their method on several disaster datasets and employ several criteria to gauge its

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, April 2024

efficacy.

1. This research presents a novel approach to mining Twitter during catastrophes for damage assessment-related tweets. Low-level lexical and grammatical elements, as well as the most frequently used words in such tweets, form the basis of this methodology. To quantify the importance of these characteristics, the authors employ linear regression weighting techniques and Support Vector Regression (SVR). Even when the model is trained on datasets from different disasters, the suggested approach is vocabulary-independent, allowing it to properly detect tweets relating to damage assessment in a variety of circumstances.
2. In this study, we compare our proposed method to other state-of-the-art methods by looking at their performance on several benchmark datasets. The outcomes demonstrate that their method is superior to the state-of-the-art solutions, with an increase in accuracy of up to 37.12%. When applied to in-domain and cross-domain damage assessment tweets, their proposed technique achieves the highest performance on binary and multi-class classification tasks.

II. PROBLEM DISCUSSION

This paper acknowledges that sentiment analysis of microblogging is a nascent field with room for more research. Due to its140-character constraint, Twitter poses a particular difficulty for sentiment analysis. User evaluations, papers, web blogs and articles, and general phrase-level sentiment analysis have been intensively explored. LSTM and TensorFlow have shown promise in sentiment categorization, however, manual tagging is costly. Unsupervised and semi-supervised algorithms have improved, but they still need work. The authors suggest comparing novel features and classification approaches to baseline performance.

III. EXISTING METHOD

The In [1] showed emergency social media use and computational methods for processing social media data. Most crisis-data categorization methods use uni-grams and POS tags to extract catastrophe situational awareness information. Several studies have detected earthquake and other disaster tweets using n-gram features, keyword position, and surrounding phrases. Most of these strategies are vocabulary-dependent and not cross-domain.

In 2018, a vocabulary-independent strategy was presented to categorize tweets as situational or non-situational usinglow-level lexical and syntactic data. The authors compared their method to the Bag-Of-Words (BOW) model for English and Hindi tweets using different disaster datasets. The random forest classifier identified eyewitness accounts from textual and domain-expert data. The authors investigated direct, indirect, and susceptible direct eyewitness reports on earthquakes, floods, storms, and forest fires. They detected urgent tweets using - low-supervision and transfer learning and tested their method on Nepal, Macedonia, and Kerala earthquake datasets. With fewer labeled data, their strategy outperformed baseline methods. These disaster- related tweet studies did notcategorize subcategories such infrastructure damage, human harm, resource demand, and availability.

Disaster-related tweet subcategories. A few research have derived catastrophe tweet subgroups. In one investigation, decision trees outperformed SVM, Random Forest, and Adaboost classifiers in identifying catastrophe help requests using context and content data. Recently, another study examined resource availability and demand utilizing communication, location, and infrastructure damage parameters. They extracted resource demands and availabilityfrom disaster tweets using word and character embeddings and re-ranking feature selection. The scientists created a system to analyze tweets about resource availability and need, including location, and related them by resource similarity and geography.

One study classified photos into three damage severity categories using pre-trained models using VGG-16 architecture. They tested their strategy on disaster datasets from Nepal and Typhoon Ruby. Deep neural networks estimated Hurricane Dorian damage from social media photos in another study. The DANN and VGG-19 model identified picture degradation, which they evaluated on datasets from Nepal, Ecuador, Ruby, and Matthew hurricanes

Information retrieval classified English damage assessment tweets for text data damage assessment. Semi automatically selecting seed phrases, they outperformed state-of-the-art methods on Nepal, Italy, and Indonesia Tsunami datasets. Their keyword selection technique required human annotators and was not tested on other disaster datasets.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, April 2024

Another study used fresh features weighted by linear regression and SVR algorithm to recognize disaster tweets automatically. They performed error analysis on in-domain and cross-domain heterogeneous disaster datasets to improve their method.

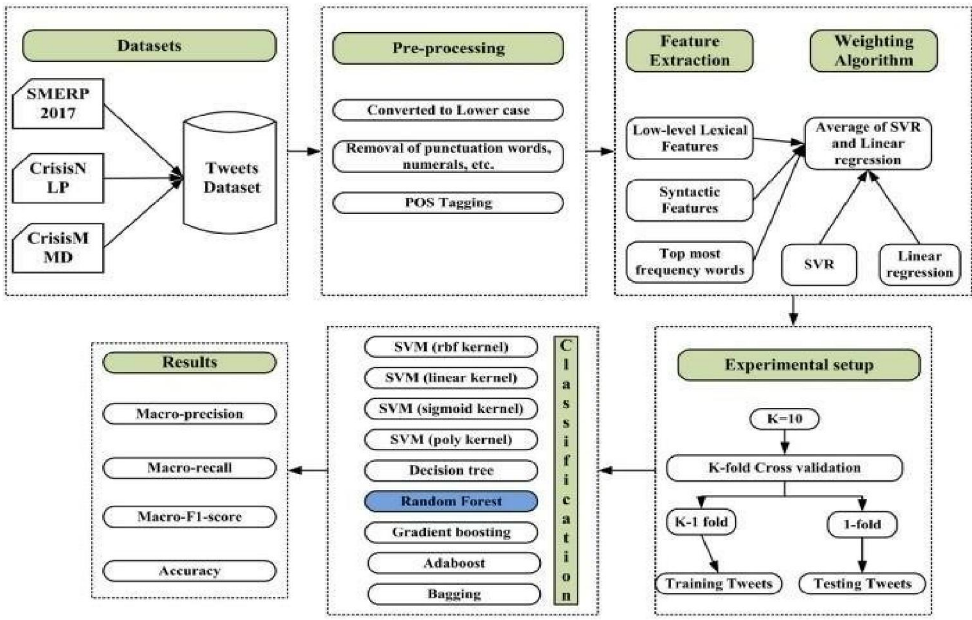

FIG.1. BLOCK DIAGRAM OF EXISTING METHOD

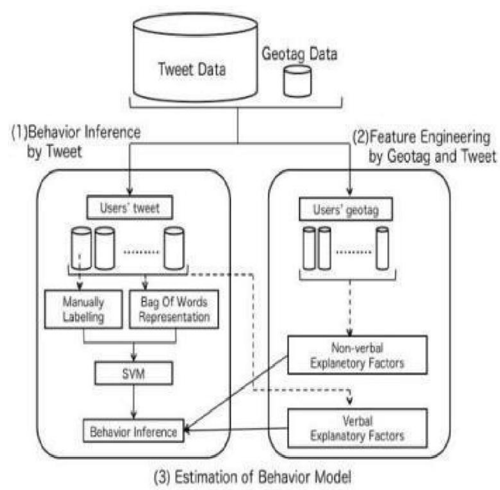

FIG.2. EXISTING FRAMEWORK

IV. PROPOSED METHOD

Long Short-Term Memory

The implementation of artificial recurrent neural networks (RNNs) with Long-Short-Term Memory (LSTM) units for the purpose of text production and tweet classification. LSTM units are the essential building blocks of RNNs, and they managethe flow of information by utilising gates such as input, output, and forget gates. LSTM units are the fundamental building blocks of RNNs. When using the LSTM method, the methodology chooses the term that is most likely to apply to each category depending on the tweet that is being categorized. In order to convert words into vectors, word embedding and count vectorization techniques utilising WordNet are utilized. These techniques then give an explicit representation of the context for the neural network to work with. The output of the word embedding is sent into the neural network, which, by keeping continuity while reducing the dimensionality of the input space, produces a better result. LSTM units each contain a cell that is accountable for remembering values over a period of time that can be completely arbitrary. A cell, an input gate, an outputgate, and a forget gate are the constituent parts of the LSTM unit, which is depicted in Figure 3.

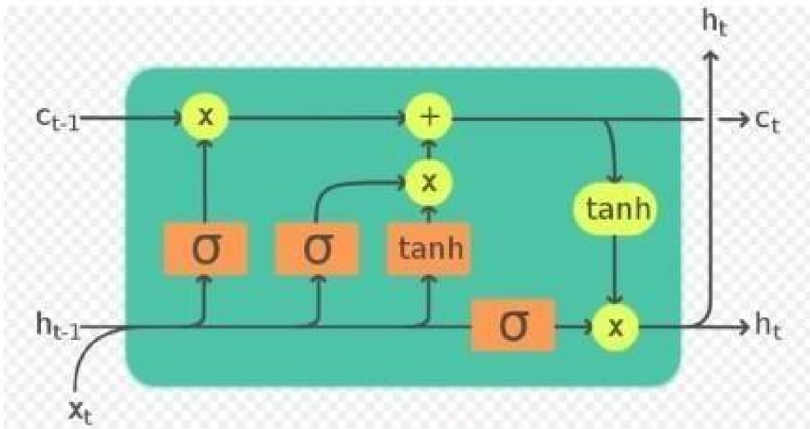

FIG. 3. LSTM ARCHITECTURE

Vectorization

Word vectorization, the act of translating text into numerical form, can be accomplished with the help of libraries like Sklearnand Keras.

Tokenization

In the first stage, called tokenization, the text is broken down into individual words. Tokens can be anything from individualletters to entire words.

Padding

Padding is another method for meeting the condition of neural network models that all inputs be of the same size and shape.

Stemming

Stemming is the process of breaking down words into their simplest form, or root, so that related terms can be comparedmore easily.

Word Embedding

In word embedding, words with similar meanings are assigned the same vectors in a learnt representation of the text. This is accomplished by teaching a neural network to represent words effectively by learning the values of their vectors in a lower-dimensional space.

TensorFlow

Python and TensorFlow remove stop words and punctuation, stem, lemmatize, and vectorize the deep learning dataset usingbag of words, count vectorization, and TF-IDF vectorization. Count Vectorization and TF-IDF vectorization partition the dataset using bigrams (n=2). Unigrams and n-grams with n > 2 are more accurate. WordNet word embeddings pre-train eachcatastrophic event sample and add a little Gaussian noise to balance the dataset.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, April 2024

The embedding layer breaks all sentences into words in Recurrent Neural Networks. The output layer connects word embedding and neural network layers such Long Short-Term Memory, Bi-directional Gated Recurrent Unit, and 3-layer fully connected. Fig. 3 shows layers-trained neural network accuracy after 5 iterations. GRU and LSTM generate reliable global predictions with testing set distributions that match training sets. A randomized input dataset depicting 26 calamities including the 2012 Colorado wildfires and 2012 Costa Rica earthquake batches tweets into the neural network. After 5 iterations, bi-directional GRU with 3-layers outperforms LSTM on tweets.

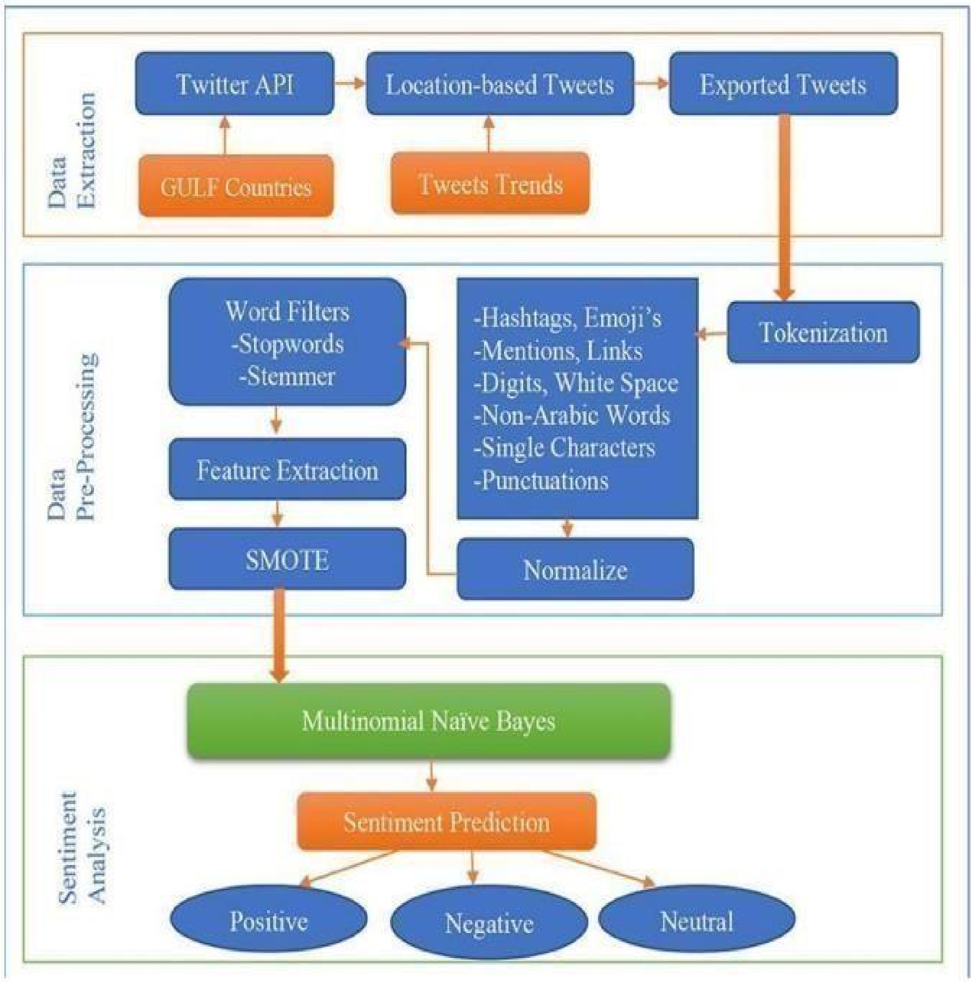

FIG.4. BLOCK DIGARAM OF PROPOSED METHOD

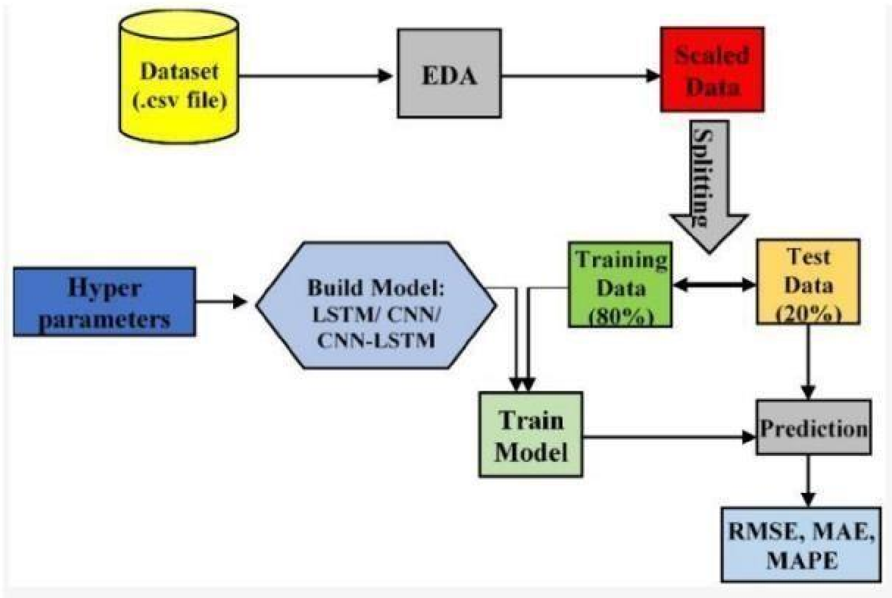

FIG.5. WORKFLOW OF PROPOSED METHOD

Dataset	Accuracy	Macro-precision	Macro-recall	Macro-f1-score
Italy Earthquake	93.27	92.81	92.79	92.81
Chile Earthquake	**94.62**	**94.86**	**94.62**	**94.61**
Nepal Earthquake	78.16	78.02	78.05	77.87
Pakistan Earthquake	82.26	82.74	82.25	82.19
Iraq-Iran Earthquake	79.58	78.73	78.86	78.48
Mexico Earthquake	79.37	79.90	79.37	79.26
Cyclone	86.92	87.26	86.92	86.89
Typhoon Hagupit	85.22	86.58	85.22	85.06
Hurricane Irma	77.23	77.45	77.23	77.18
Hurricane Harvey	77.32	76.14	76.72	76.28
Hurricane Maria	79.14	79.41	79.14	79.10
California Wildfires	**73.31**	**73.78**	**73.45**	**73.24**
Sri Lanka Floods	86.92	87.26	86.92	86.89
India Floods	91.46	91.48	91.48	91.43

TABLE.1. RESULTS OF PROPOSED METHOD ON VARIOUS DATASETS.

VI. COMPARATIVE STUDY

This section will detail the approaches that are now in use and compare the findings to the method that will be presented for detecting tweets related to damage assessment.

Classifiers	Macro-precision	Macro-recall	Macro-f1-score	Accuracy
SVM (kernel='rbf')	89.01	86.42	86.18	86.41
SVM (kernel='sigmoid')	89.01	86.42	86.18	86.41
SVM (kernel='linear')	91.69	90.38	90.30	90.38
SVM (kernel='poly')	29.40	50.14	33.68	50.43
Decision tree classifier	92.67	92.59	92.58	92.59
Random forest	**93.27**	**92.81**	**92.79**	**92.81**
Gradient boosting	92.51	92.16	92.15	92.16
Adaboost	91.31	91.07	91.06	91.07
Bagging	91.07	90.73	90.71	90.73

TABLE.2. COMPARISION BETWEEN EXISTING AND PROPOSED METHODS

V. FUTURE SCOPE

LSTM and TensorFlow sentiment analysis of disaster damage assessment tweets has several applications and research prospects. Its emotional impact analysis aids disaster response. Real-time sentiment analysis helps measure catastrophes' emotional impact and adapt response operations. Thus, disasters require real-time sentiment analysis techniques for social media data.

Sentiment analysis models are usually developed on English language data, hence they may not apply to non-English speaking catastrophe zones. Thus, future research could construct sentiment analysis models for other languages or adapt existing algorithms to multilingual data. This could improve sentiment analysis across areas.

LSTM and TensorFlow sentiment analysis of damage assessment tweets during catastrophes offers several uses and research opportunities. Sentiment analysis can reveal the emotional impact of disasters and improve disaster response with continuing innovation. For global catastrophe response, sentiment analysis models must be adaptive and relevant.

VI. CONCLUSION

In conclusion, LSTM and TensorFlow can improve the accuracy and nuance of insights gained from sentiment analysis of damage assessment tweets during catastrophes. This can help disaster relief workers better anticipate and address the psychological and emotional needs of individuals they serve. Sentiment analysis with LSTM and TensorFlow allows for the identification of not only the overarching sentiment of a tweet, but also the individual emotions and intensity of those emotions conveyed in the tweet.

REFERENCES

[1]. "Sentiment Analysis of Tweets During Hurricane Sandy" by J. Lu, S. S. Pan, and L. H. Yang (2014) 2. "Assessing Disaster Damage Using Social Media Analytics: A Study of Hurricane Harvey" by A. Agrawal,

[2]. H. Choudhury, and R. Bhattacharya (2018). Sentiment Analysis of Tweets During Disaster Events" by S. Sarkar, B. Saha, and S. Chakraborty (2018).

[3]. Assessing Disaster Damage through Twitter Sentiment Analysis" by K. Vaddadi and S. K. Bhattacharyya (2019).

[4]. Madichetty, S. and Sridevi, M., 2021. A novel method for identifying the damage assessment tweets during disaster. Future Generation Computer Systems, 116, pp.440-454.

[5]. Hara, Y., 2015. Behavior analysis using tweet data and geo-tag data in a natural disaster. Transportation Research Procedia, 11, pp.399-412.

[6]. Seddighi, H., Salmani, I. and Seddighi, S., 2020. Saving lives and changing minds with Twitter in disasters and pandemics: a literature review. Journalism and Media, 1(1), pp.59-77.

[7]. Muhammad Imran, Carlos Castillo, Fernando Diaz, Sarah Vieweg, Processing social media messages in mass emergency: A survey, ACM Comput.Surv. 47 (4) (2015) 67.

[8]. Kate Star bird, Leysia Palen, Amanda L. Hughes, Sarah Vieweg, Chatter on the red: what hazards threat reveals about the social life of microblogged information, in: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, ACM, 2010, pp. 241–250.

[9]. Madichetty Sreenivasulu, M. Sridevi, A survey on event detection methods on various social media, in: Recent Findings in Intelligent Computing Techniques, Springer, 2018