



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: <https://doi.org/10.22214/ijraset.2022.47946>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Survey on Identification and Classification of Informative Tweets During Disasters

Prof. Sonlai C. Patil¹, Rushikesh Kadam², Sumit Sapate³, Tejas Khandejod⁴, Prasad Bhosale⁵

^{1, 2, 3, 4, 5}Department of Information technology, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India

Abstract: During the time of crisis, people often post countless instructive and informative tweets on various social media platforms like Twitter. Recognizing informative tweets could be a difficult errand during the fiasco from such an enormous pool of tweets. As a solution to the current issue of sorting out enlightening tweets, we present a technique to perceive the distinguishing calamity related informative tweets from the Twitter streams utilizing the textual content. Our objective is to construct a model by using Natural Language Processing(NLP), Exploratory Data Analysis(EDA) and Support Vector Machine(SVM) and Visual Geometry Group (VGG as Deep CNN) to categorize the textual and pictorial content for the tweets. In this kernel we'll explore classification of tweets as disaster or non- disaster, using TensorFlow and Keras. The output of the text-based model is consolidated using the late fusion technique to predict the tweet label.

Keywords: Tweet Analyse, Artificial Intelligence, Support Vector Machine (SVM), Feature Extraction, Text Processing, Machine Learning

I. INTRODUCTION

A natural disaster makes critical environmental dislocations taking broad endeavours from society to survive and manage with them. In similar cases of disasters, a substantial quantum of data is posted on social media platforms like Twitter. thus, the demand of the case is to apply certain tools which will help in analysing this huge quantum of tweets and can classify them among the instructional and non-informative orders for the rapid-fire early damage assessment during disasters.

The public authority or humanitarian association needs data on the circumstance updates of the number of people are wounded and died, the quantity of structures that are imploded, so forth, to take immediate action. The factual analysis and evaluation of these crisis related data can help humanitarian organizations and rescue squads in efficient decision-making and dynamic prioritization of tasks. Similarly, affected people need data like where clinical assets are accessible, food and shelter assets, etc. Tweets of this type are grouped under informative tweets. Among the large volume of tweets identified with a specific calamity, a number of them could also be simply expressing gratitude towards Twitter or neighborhood bunches for his or her assistance. Also some people post tweets which don't seem to be directly associated with the disaster, for instance, feelings, sentiments, emotions etc. These tweets don't seem to be helpful within the rescue work. Hence these sorts of tweets are named as non-informative tweets. Institutional and Volunteer rescue efforts save lots of individuals during a catastrophe. However, Individual volunteers have time limitations and lack of assets. Also it is not practically possible for the people from the rescue squad to reach out to every tweet on their own because of the huge volume and fast pace at which tweets are being posted. This generates an urgent need to construct certain frameworks that may consequently segregate useful information through an IDENTIFICATION AND CLASSIFICATION OF INFORMATIVE TWEETS DURING DISASTERS infinite volume of social media content. The auto-programmed classification of messages, particularly tweet messages, will be a difficult task because of their restriction in character length (280 characters), non-standard shortened forms, and linguistic blunders.

In this paper, we propose a method by highlighting textual nature of the tweets using NLP. Natural Language Processing (NLP) enables computers to understand natural languages as humans do. Whether the language is spoken or written, natural language processing uses artificial intelligence to take real-world inputs, process it, and make sense of it in a way a computer can understand. In this kernel we'll explore classification of tweets as disaster or non-disaster, using tensorflow and Keras.

II. RELATED WORK

In this paper [1] authors proposed a novel approach to classify tweets adapting image features. It used CNN-ANN dual approach to classify text. CNN is used for feature extraction. ANN is used as classifier. The images are classified using fine-tuned VGG-16 model. Almost 74% accuracy was achieved.

In this paper [2] authors proposed the deep learning approach combining CNN and Bi-LSTM to categorize the tweets. It consists of seven modules: Input layer, embedding layer, BLSTM layer, Attention layer, Auxiliary features input, Convolution layer, and Output layer. The system also extracts the location from the tweets data to assist in rescue operation held by rescue squad. The method outperformed many other classification methods based on Recall, Precision, F1 Score. They also developed an adaptive algorithm to perform efficient rescue operation scheduling according to the priorities.

In this paper [3] authors proposed a semantic approach for tweet classification by using Dual-CNN technique. A semantic embedding layer is added to the traditional CNN layers for capturing the context of the tweets efficiently. The pre-processed tweets are given as input to the word vector initialization as well as concept vector initialization. The semantic extraction used Alchemy API to extract named entities and then mapped them with the subtypes with multiple bases like DBpedia and Freebase. Accuracy greater than 79% is achieved.

In this paper [4] authors proposed a method to extract events from newswires or social media content for Indian Languages like Hindi, English, Tamil. Event identification is employed using CNN and Bi-LSTM together. Event Arguments and Event trigger are identified using the technique and these are linked correspondingly using heuristic based approach where every argument word is matched with nearest trigger word depending upon distance which is calculated by number of sentences between them. The system achieves an F-score of 39.71,37.42,39.91 for Hindi, Tamil and English datasets which is relatively low compared to other models.

In this paper [5] authors proposed an multi-channel representation along with CNN for tweet classification. The proposed model consists of 6 main structures, an input layer, convolution layer, embedding layer, fully-connected layer, pooling layer and output layer. The context of words in the vectors can be considered using multichannel distribution representation. In this representation, multi channels consisting of multiple values for each element in vector are represented. Each word has different vectors depending upon the context of word in tweets. Accuracy of 77% is achieved using this approach.

In this paper [6] authors proposed CNN model to categorize tweets related to emergency response phase. The pre-processed input of tweets is given to feature extraction phase to form word vectors which is further given to pooling layer and convolutional layer to extract features and reduce dimension. The output of CNN layers is given to flattening layers. The results of all these layers are concatenated into a single vector. The vector will be processed at FCL (Fully Connected Layer) using the softmax function to get the probability for each class. For small amounts of data, the model achieved 98% accuracy. However, the accuracy drops for subsequent amounts of data.

In this paper [7] authors proposed a multi-label classification CNN model with respect to seven class taxonomy and contains seven similar classifiers predicting the binary output specifying whether the label is relevant or not. Accuracy of 88% is achieved by the deep CNN architecture.

III. AIM

Twitter has become an important communication channel in times of emergency. They ubiquitousness of smartphones enables people to announce an emergency they are observing in real-time. Because of this, more agencies are interested in programmatically monitoring Twitter (i.e. disaster relief organizations and news agencies). The aim of this project is to propose a system which can be used to classify and predict the Tweet as Informative or Non-informative during disasters based on previous data.

IV. SYSTEM ARCHITECTURE

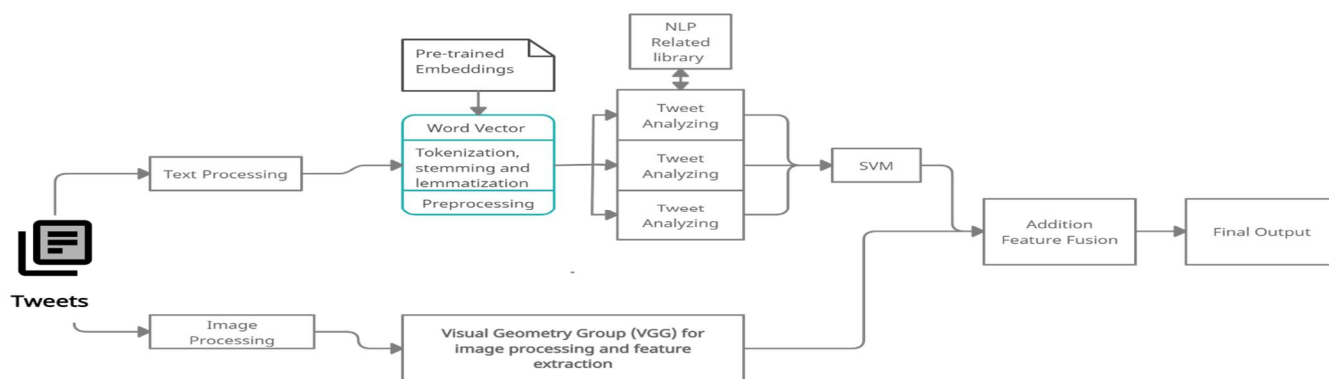


Fig . System Architecture

In this we will be using text processing to process input text tweet, here we will use pre-trained NLP libraries to process the data and find important keywords, then process those with SVM algorithm to differentiate the tweet. For Image input we will be using VGG and then using SVM we are classifying the result informative or not informative.

If the both text and image will be input to the system, the result of the system will depend on the result output by both the text and image processing and final result should be output.

V. ALGORITHM

In this system we are using text processing, data processing and image processing for this system to detect the Tweet label it contains during disastrous situations into informative and non-informative categories.

A. Steps For Text Processing

- 1) Input as tweet
- 2) Pre-processing the inputted data (removing unwanted data resize them)
- 3) Now Feature extraction process (to detect the features)
- 4) Classify the processed data using SVM algorithm using trained dataset created Model.
- 5) Categorized the input tweet is Informative or non-informative for disaster and non-disaster

B. Steps For Image Processing

- 1) Input as image
- 2) Processing the image using pre-trained model on VGG Model
- 3) Process the output of VGG model
- 4) Categorized the input image as disaster and non-disaster

VI. EXPECTED OUTPUT

In the Desktop application the user will input as text dataset of Tweets and trained and test dataset using SVM Algorithm and create best SVM Model.

Using these model To classify Tweets on the basis of text as it contains during disastrous situations into informative and non-informative categories.

VII. CONCLUSION

Through this research, we got acquainted with different perspectives to classify the tweets on the basis of text it contains during disastrous situations into informative and non-informative categories. After evaluation various of methods for Text Classification, we concluded that our approach has better efficiency compared to other approaches for text classification. This research paper proposes a model that specifies NLP and SVM combined approach for text based classification. Thus the approach will prove beneficial to segregate the important information from the tweets during crisis conditions.

REFERENCES

- [1] Madichetty, Sreenivasulu M, Sridevi. (2020). Classifying informative and non-informative tweets from the twitter by adapting image features during disaster. Multimedia Tools and Applications. 79. 10.1007/s11042-020-09343-1.
- [2] Md. Yasin Kabir, Sanjay Kumar Madria (2019). A Deep Learning Approach for Tweet Classification and Rescue Scheduling for Effective Disaster Management
- [3] Burel, G., Saif, H., Fern'andez, M., & Alani, Harith. (2017). On Semantics and Deep Learning for Event Detection in Crisis Situations.
- [4] Alapan Kuila, Bussa, S.C., Sarkar, S. (2018). A Neural Network based Event Extraction System for Indian Languages. FIRE.
- [5] Hashida, S., Tamura, K., Sakai, T. (2018). Classifying Tweets using Convolutional Neural Networks with Multi-Channel Distributed Representation.
- [6] Journal of Software Engineering., agung triayudi.(2019). CONVOLUTIONALNEURALNETWORK FOR TEXT CLASSIFICATION ON TWITTER Journal of Software Engineering Intelligent SYstems, 4(3), 123– 131.
- [7] Aipe, A., Ekbal, A., Sundararaman, M.N., Kurohashi, S. (2018). Deep Learning Approach towards Multi-label Classification of Crisis Related Tweets. ISCRAM.
- [8] X. She and D. Zhang, "Text Classification Based on Hybrid CNN-LSTM Hybrid Model," 2018 11th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 2018, pp. 185-189, doi: 10.1109/ISCID.2018.10144.
- [9] Jain, Pallavi & Ross, Robert & Schoen-Phelan, Bianca. (2019). Estimating Distributed Representation Performance in Disaster-Related Social Media Classification. 10.1145/3341161.3343680.
- [10] Abhinav Kumar, Jyoti Prakash Singh, Yogesh K. Dwivedi, Nripendra P. Rana (2020). A deep multi-modal neural network for informative Twitter content classification during emergencies.
- [11] Kumar, Abhinav Singh, Jyoti Dwivedi, Yogesh Rana, Nripendra. (2020). A deep multi-modal neural network for informative Twitter content classification during emergencies. Annals of Operations Research. 10.1007/s10479-020-03514-x.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)