iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52862

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Identification and Classification of Informative Tweets During Disasters

Sonali C. Patil[1], Rushikesh Kadam[2], Sumit Sapate[3], Tejas Khandejod[4], Prasad Bhosale[5]
[1, 2, 3, 4, 5]*Department of technology, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India*

Abstract: ***Twitter serves as a popular social media platform for sharing information during disasters. However, the task of distinguishing informative tweets from the overwhelming volume of content can be daunting. In this research, we propose an innovative approach that combines Natural Language Processing (NLP), Exploratory Data Analysis (EDA), Support Vector Machine (SVM), and the Xception algorithm. Our objective is to effectively categorize disaster-related tweets based on their textual and visual content. By leveraging TensorFlow and Keras, we aim to accurately classify tweets as either disaster-related or unrelated. Additionally, we integrate the Google Translate API to facilitate the translation of text data into multiple languages. To enhance the precision of tweet label prediction, we utilize the late fusion technique to consolidate the output of our text-based model.***

Keywords: ***Tweet Analysis, Artificial Intelligence, Support Vector Machine (SVM), Feature Extraction, Text Processing, Machine Learning, Google Translator, Xception Algorithm.***

I. INTRODUCTION

In recent times, social media platforms, especially Twitter, have emerged as valuable sources of real-time information during disaster situations. They have become crucial channels for receiving updates, connecting with others, and contributing to response efforts. The task of categorizing and classifying a large number of tweets as informative or non-informative is of utmost importance during disasters. Informative tweets play a vital role in providing critical updates on the situation, emergency services, evacuation instructions, and facilitating relief efforts. Conversely, non-informative or irrelevant tweets contribute to information overload and hinder the dissemination of accurate information.

Humanitarian organizations and rescue teams heavily rely on data regarding the number of casualties, collapsed buildings, and other relevant updates to take immediate action. Categorizing this crisis-related data enables efficient decision-making and task prioritization for these organizations. Affected individuals also depend on information about available resources, such as medical assistance, food, and shelter. However, these informative tweets often get mixed with non-informative tweets expressing personal sentiments, gratitude, emotions, or unrelated content. While non-informative tweets hold value for personal expression and social interaction, they do not contribute to rescue or relief efforts.

During disasters, both institutional and volunteer rescue efforts are essential in responding to and mitigating the impact of the event. However, individual volunteers face limitations in terms of time and resources. Additionally, verifying each tweet becomes practically impossible for rescue teams due to the overwhelming volume of data and the speed at which tweets are posted. This urgency necessitates the development of specific frameworks to extract useful information from the vast amount of social media content during disasters.

The automatic classification of tweets presents a challenging task, particularly due to character limitations, irregular abbreviations, and grammatical errors. In this research, we propose a strategy that utilizes Natural Language Processing (NLP) to focus on the textual aspect of tweets. NLP, powered by artificial intelligence, enables computers to understand natural language inputs, whether spoken or written, akin to human comprehension.

Besides textual content, images also hold potential for providing real-time information during disasters, especially regarding the extent of damage. Image recognition algorithms, such as the Xception Algorithm introduced in this study, prove effective in identifying relevant images.

To enhance the accuracy of our approach, we integrate the Google Translator API, which supports over 100 languages, to translate data from one language to another. However, it is crucial to note that translation quality may vary based on various factors.

©IJRASET: All Rights are Reserved

II. RELATED WORK

In this paper [1], the authors propose a novel approach for classifying tweets by adapting image features. Their approach, compared to state-of-the-art models, utilizes a CNN-ANN dual approach where CNN is used for feature extraction and ANN serves as the classifier.

In another paper [2], the authors present a deep learning approach that combines CNN and Bi-LSTM to categorize tweets. Their system consists of seven modules, including input layer, embedding layer, BLSTM layer, attention layer, auxiliary features input, convolution layer, and output layer. The system also extracts location information from tweets to assist rescue operations. The proposed method outperforms other classification methods based on recall, precision, and F1 score. They also develop an adaptive multi-task hybrid algorithm for efficient rescue operation scheduling based on different rescue priorities.

They add a In a distinct study [3], the authors put forward a semantic approach for classifying tweets using the Dual-CNN technique. semantic embedding layer to traditional CNN layers to efficiently capture the context of tweets. Pre-processed tweets are inputted for word vector and concept vector initialization. The proposed approach in the paper utilizes the Alchemy API for semantic extraction, which involves extracting named entities and mapping them with subtypes using bases such as DBpedia and Freebase.The proposed approach achieves an accuracy greater than 79%.

Another paper [4] presents a method for event extraction from newswires or social media content in Indian languages like Hindi, English, and Tamil. The authors employ CNN and Bi-LSTM for event identification and utilize a heuristic-based approach to link event arguments and triggers. The achieved F-scores for Hindi, Tamil, and English datasets are relatively low compared to other models.

In a multi-channel representation approach [5], the authors propose a CNN model for tweet classification. The model includes an input layer, convolution layer, embedding layer, fully-connected layer, pooling layer, and output layer. The use of multichannel distribution representation allows consideration of the context of words in vectors. The approach achieves an accuracy of 77%.

Another paper [6] proposes a CNN model for categorizing tweets related to emergency response phases. The pre-processed tweets undergo feature extraction, pooling layer, and convolutional layer to extract features and reduce dimension. T

he outputs are then concatenated and processed by the Fully Connected Layer (FCL) using the softmax function to obtain the probability for each class.

The model achieves high accuracy with small amounts of data but experiences a decrease in accuracy with subsequent amounts of data.

In a multi-label classification approach [7], the authors propose a CNN model with respect to a seven-class taxonomy. The model contains seven similar classifiers predicting binary output for each label's relevance. The deep CNN architecture achieves an accuracy of 88%. The paper also investigates the effect of Twitter-specific linguistic knowledge on the deep CNN.

A text classification method based on a hybrid CNN-LSTM model is proposed in another paper [8]. The algorithm utilizes the Skip-Gram and CBOW models in word2vec to represent words as vectors. Tests conducted on the Chinese news corpus of Sogou.com validate the algorithm's capability in enhancing the precision of text classification.

In an investigation of pre-trained language representations [9], the authors evaluate their effectiveness in determining the informativeness and information type of social media during disasters. Models built from pre-trained word embeddings, such as Word2Vec, GloVe, ELMo, and BERT, are used for performance evaluation. BERT and ELMo achieve competitive results for disaster tweet classification, similar to Word2Vec and GloVe.

Finally, a multi-modal system [10] utilizing tweet texts for identifying informative Twitter content is proposed. The system employs LSTM for tweet text classification and achieves high F1-scores for various disaster-related Twitter datasets.

In a novel interpretation of Inception modules [11], the authors propose a new deep convolutional neural network architecture inspired by Inception. Instead of Inception modules, the architecture employs depthwise separable convolutions, which are intermediate steps between regular convolution and depthwise separable convolution operations.

III. AIM

Twitter has emerged as a vital communication channel during emergencies, allowing people to share real-time observations of ongoing incidents using their smartphones.

This has sparked the interest of various agencies, such as disaster relief organizations and news agencies, in programmatically monitoring Twitter. The goal of this project is to propose a system capable of classifying and predicting whether a tweet is informative or non-informative during disasters, leveraging previous data.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

IV. SYSTEM ARCHITECTURE

A. Text

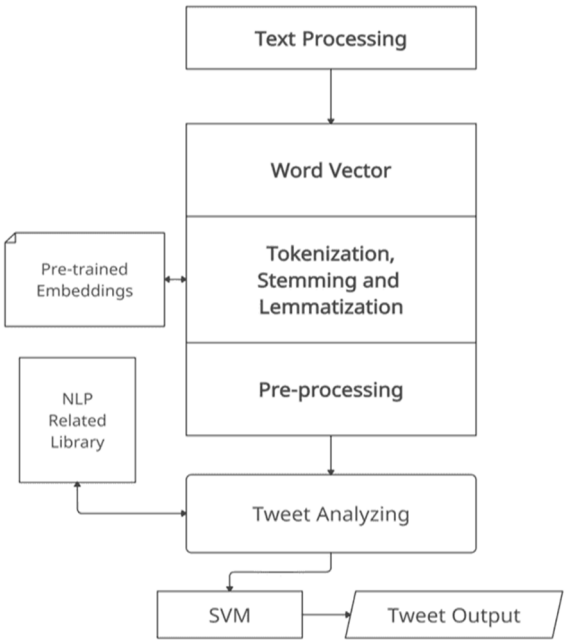


Fig 1. System architecture for Text Using SVM Model

B. Image

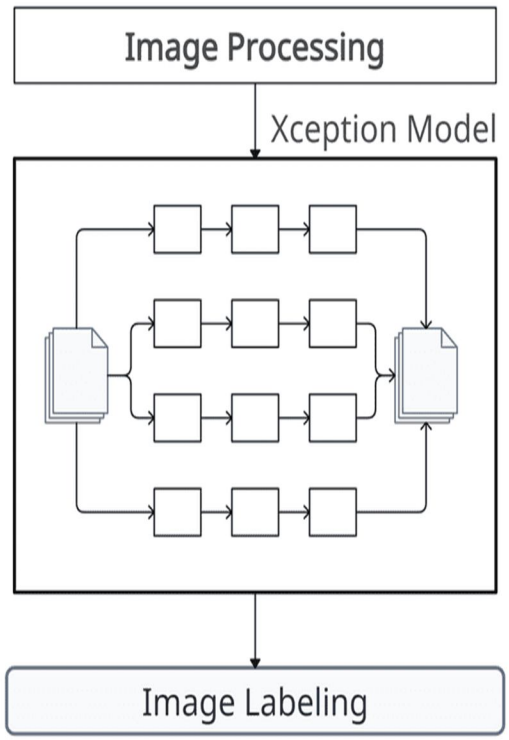


Fig 2. System architecture for Image Processing

We have referred to the architecture of the Xception Model for Image from the paper titled "Xception: Deep Learning with Depthwise Separable Convolutions" by Francois Chollet at Google, Inc.

©IJRASET: All Rights are Reserved

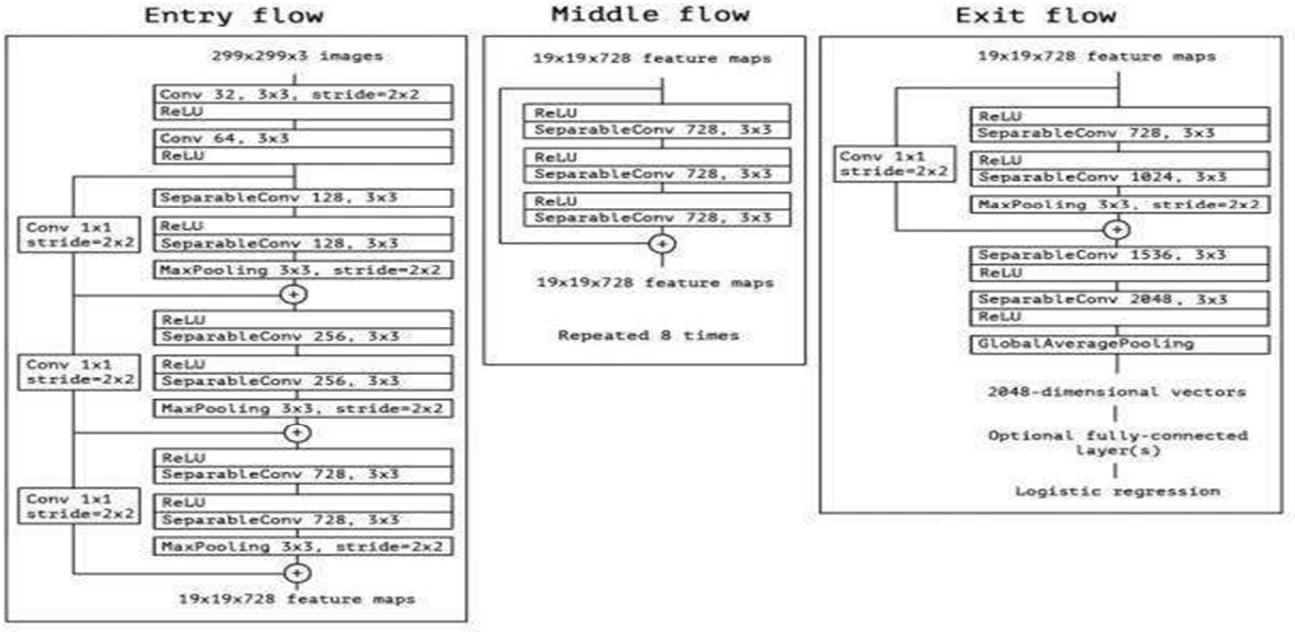


Fig 3. System Architecture of Xception Model for Image

In our study, we propose employing text processing techniques to analyze input text tweets. We will utilize pre-trained NLP libraries to extract essential keywords, which will then be processed using the SVM algorithm to differentiate the tweets. Additionally, we will utilize the Xception algorithm to analyze image inputs and classify them as either informative or non-informative in relation to disaster situations.

Different reason for using SVM & Xception algorithm as compared to other algorithms are as follows:

1) SVM Algorithm

SVM is a powerful algorithm that can handle both linear and non-linear classification tasks. Here's why SVM Algorithm can be suitable for tweet identification and classification during disasters

a) SVM is effective with high dimensional data. It can perform well with such data by finding an optimal hyperplane to separate different classes, even in high-dimensional spaces.

b) SVM algorithm have an ability to handle non-linear data . Disasters can generate a wide range of tweet patterns, including non-linear relationships between features SVM with non-linear kernels, such as the radial basis function (RBF) kernel, can capture complex patterns and improve classification accuracy.

c) Robustness to Outliers: In disasters situation, many tweets might contain irrelevant or misleading information. SVM is less sensitive to outliers, meaning it can focus on the majority of relevant tweets and still achieve good classification performance.

2) Xception Algorithm

Xception is a deep learning algorithm based on convolutional neural networks (CNNs) which is used to designed specifically for image classification tasks, but its architecture can also be adapted for tweet classification Here's why Xception algorithm can be suitable for tweet identification and classification during disasters.

1) Efficient Feature Extraction: Xception uses deep network architecture which allows it to learn hierarchical features from input data. In case of tweets, we can use techniques like word embeddings or TF-IDF to convert them into numerical representations. Xception can effectively extract features from these representations to classify tweets based on the extracted information.

2) Transfer Learning: Xception is often pre trained on large image datasets, such as ImageNet. However, the lower-level features learned during this pre training can be useful for other tasks, including tweet classification. By leveraging transfer learning, you can benefit from the knowledge acquired during training on vast image datasets and apply it to tweet classification during disasters.

3) Handling Visual Information: Although tweets are primarily text-based, they can sometimes contain images related to disasters. Xception, being originally designed for image classification, can handle the visual content in tweets effectively. By combining both textual and visual information, Xception can provide a comprehensive approach to tweet classification during disasters.

V. ALGORITHM

In this system, we employ text processing, data processing, and image processing to detect the content of tweets during disastrous situations and classify them into informative and non-informative categories.

A. Text processing steps

1) Enter as a tweet
2) Pre-processing the inputted data (resizing them and deleting any unwanted data).
3) Feature extraction process (to detect the features)
4) Using the training dataset and model developed, classify the processed data using the SVM algorithm.
5) Classified the input tweet as informative or not for disaster-related tweets and non-disaster-related tweets.

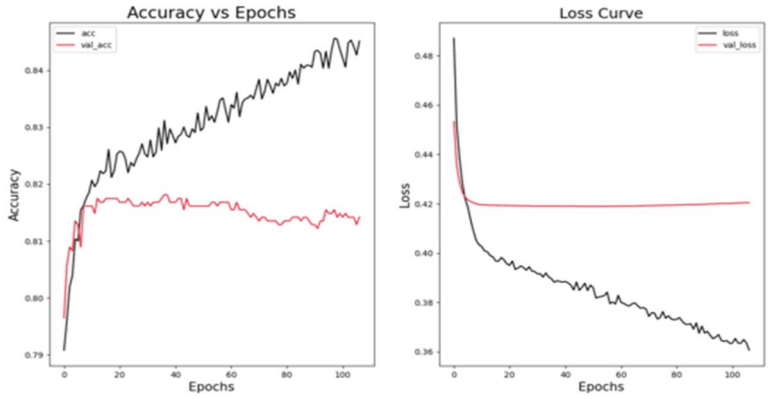


Fig 4. Accuracy and Loss Curves of developed model

B. Steps for Image Processing using Xception Algorithm

1) Input as image
2) Convolution: The input image is passed through a convolutional layer, which applies a set of filters to the image to extract relevant features.
3) Depthwise Separable Convolution: The output of the convolutional layer is processed through a sequence of depthwise separable convolutions, which decompose the convolution operation into a depthwise convolution and a pointwise convolution. This factorization helps reduce the number of parameters while preserving the model's representational capability.
4) Max Pooling: After each sequence of depthwise separable convolutions, a max pooling layer is applied to reduce the spatial dimensions of the feature maps and further extract relevant features.
5) Fully Connected Layers: The output of the final max pooling layer is flattened and processed by two fully connected layers to generate the final output.
6) Categorized the input image in one of the following categories: Earthquake, Infrastructure, Urban fire, Wild fire, Human damage, Drought, Land slide, non-Damage building, non-damage wild fire, Humans, Sea, Water disaster for disaster and non-disaster.

C. Steps for Google Translator API:

1) Collect Tweets: First collect Tweets related to disaster using Twitter API
2) Identify the language: Next Identify the language of each tweet.
3) Translate the Tweets: Using Google Translate API translate the tweets in desired language. For Example, if the user wants all tweets to be translated in English language, use Google Translate API to translate all Tweets into English.
4) Classify the Tweets: After translating the tweets use SVM algorithm to classify tweet as disaster or non-disaster.

VI. RESULTS

1) *Text Result*



Fig 5. Text Result

2) *Image Result:*



Fig 6. Image Result

3) Both text and Image Result



Fig 7. Figure showing both text and image result

Table No. I: Following table shows the Classification Performance between the different models. The Xception model gives us Top accuracy as compared to other models as shown in the table.

Table no. I

Model	Top1- accuracy	Top-5 accuracy
VGG-16	0.715	0.901
ResNet-152	0.770	0.933
Inception V3	0.782	0.941
Xception	0.790	0.945

VII. CONCLUSION

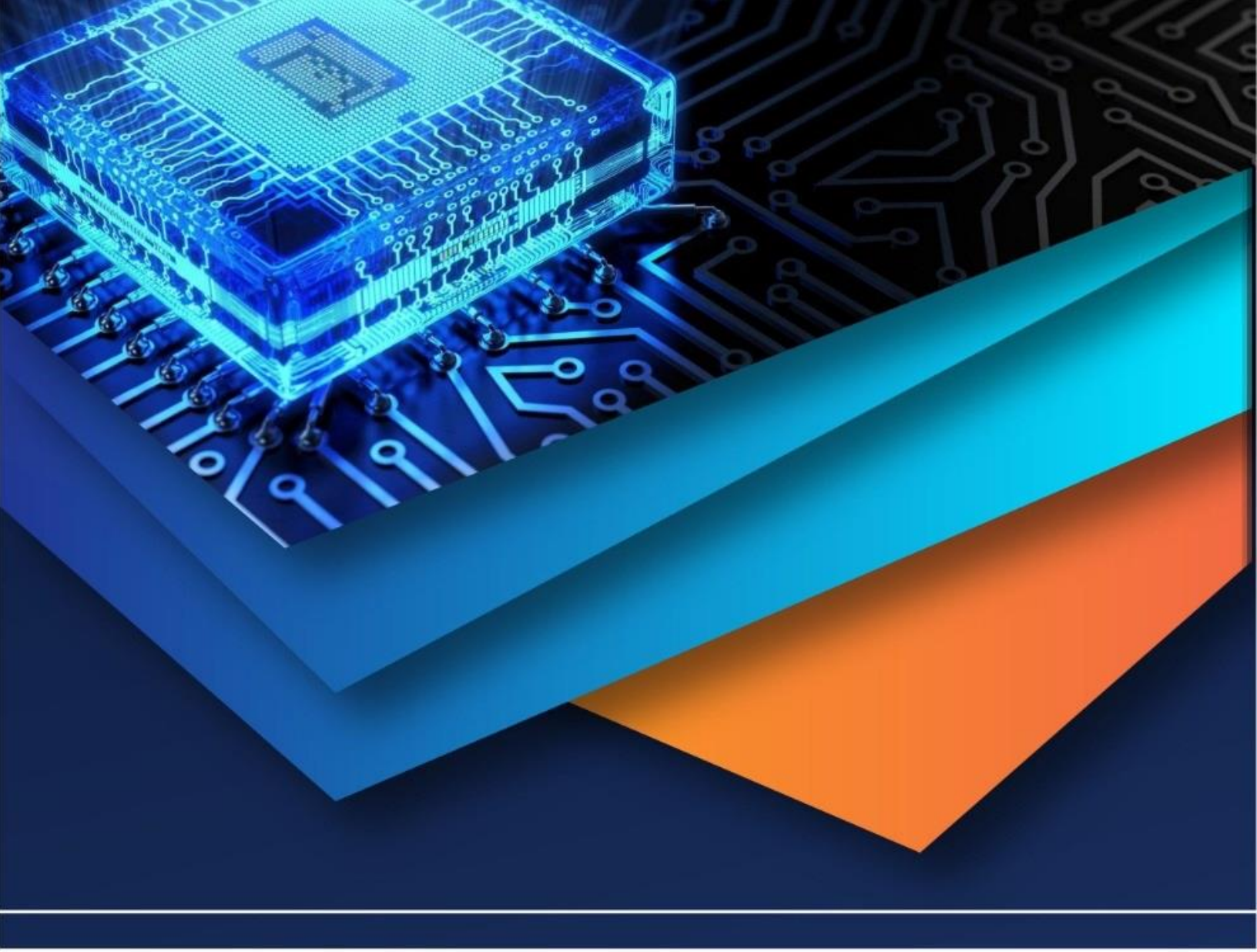
Through this research, we became familiar with various approaches to categorising tweets based on text and image. It categorises the information during catastrophes into informative and non-informative groups. After comparing our methodology to other ways for both text and image categorization, we came to the conclusion that it is more effective. This research study suggests a model that describes the NLP and SVM combined technique for text-based classification and the Xception Algorithm for picture classification. Therefore, the strategy will be useful to separate the crucial information from the tweets in times of crisis.

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.

[5] Journal of Software Engineering., agung triayudi.(2019). CONVOLUTIONALNEURALNETWORK FOR TEXT CLASSIFICATION ON TWITTER Journal of Software Engineering Intelligent SYstems, 4(3), 123– 131.
[6] Aipe, A., Ekbal, A., Sundararaman, M.N., Kurohashi, S. (2018). Deep Learning Approach towards Multi-label Classification of Crisis Related Tweets. ISCRAM.
[7] 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.
[8] Jain, Pallavi & Ross,Robert & Schoen-Phelan,Bianca. (2019). Estimating Distributed Representation Performance in DisasterRelated Social Media Classification. 10.1145/3341161.3343680.
[9] 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.
[10] Xception: Deep Learning with Depthwise Separable Convolutions, Francois Chollet Google, Inc.
[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.
[12] Nguyen, Dat & Mannai, Kamela & Joty, Shafiq & Sajjad, Hassan & Imran, Muhammad & Mitra, Prasenjit. (2016). Rapid Classification of Crisis-Related Data on Social Networks using Convolutional Neural Networks.
[13] Journal of Software Engineering, J., Systems, I., agung triayudi. (2019). CONVOLUTIONAL NEURAL NETWORK FOR TEXT CLASSIFICATION ON TWITTER. Journal of Software Engineering Intelligent Systems, 4(3), 123– 131
[14] Rushikesh Kadam, Sumit Sapate, Tejas Khandejod, Prasad Bhosale, Prof. Sonlai C. Patil. "Survey on Identification and Classification of Informative Tweets During Disasters" 2022, International Journal for Research in Applied Science & Engineering Technology (IJRASET).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)