Webology (ISSN: 1735-188X)
Volume 18, Number 5, 2021

Social Media Text Data Analysis Based On Unsupervised Machine Learning Algorithms

Praveen Sharma and Dr. Deepika Pathak

Department of Computer Application, Dr. A. P. J. Abdul Kalam University, Indore (M.P.) - 452016

Corresponding Author Praveen Sharma

Abstract—Social media is one of the source of news and information. A significant amount of users are consuming and contributing the information by sharing the identified events and information about their surroundings. Therefore, the social media is a valuable source of information for gaining the information related to disasters events i.e. natural as well as human made. In this context, this paper is motivated to analyze the social media text data in order to categorize the text using unsupervised learning algorithm. The main reason behind applying the unsupervised learning approach is that the social media based data is not labeled. Therefore, the supervised learning cannot be applied on such data for extracting the required facts form the social media unlabelled data. In this paper, we have first provides a review on existing contributions on identifying the disaster events using machine learning approach. Next, an overview of two popular clustering algorithms is provided. The overview includes the details about k-means and Fuzzy C means algorithm, and their utilization of extracting the social media data based disaster events are demonstrated. Finally the performance of the model has been evaluated based on a dataset obtained from Kaggle. The experimental results demonstrate the FCM clustering algorithm outperform as compared to k-means based algorithm.

Keywords—machine learning algorithm, supervised learning, smart grid applications, review, experimental study, performance comparison.

I. INTRODUCTION

The social media becomes popular in recent years. The people around the world are being connected with this platform. Therefore information sharing and communication is becomes easy and efficient. The shared information using social media is reachable to all around the world. In this context, the social media is become very important for informing the people during the disaster situations. During the disasters the social media platform becomes a tool to distribute the information about the disaster events. Additionally, the governmental agencies and emergency responders are also utilizing this circulated information for finding providing immediate relief to the disaster victims.

In this context, mining and organizing the social media data will be helpful during disaster recovery and management. However, the disaster management has three main phases:

1. **Pre-disaster:**The early stage detection and/or prediction of a disaster event are known as the pre-disaster stage. During this phase the social mediaplatform can be used to identify the location, and the key area of affect.

2. **During disaster:** Additionally, if the disaster event is detected in advance this platform will be used for providing the awareness and precautions to prevent the losses.
3. **Post disaster:** after disaster effect the social media platforms can be used to figure out the area of affect, location and requirements of people affected by the disasters.

Therefore, the social media is useful for all the stages of the disaster management and recovery process. In this context, this paper is motivated to explore the use of social media content analysis for automated discovery of the target disaster events. The automated content analysis requires the machine learning algorithms to employ on the data which accurately and efficiently work on the data. In this scenario the social media data is producing the unlabelled data and the supervised learning techniques are not properly work with the unlabelled data. Therefore we need an unsupervised learning algorithm to deal with this issue. The proposed work is aimed to explore the following in this presented study:

1. **Explore the recent techniques of disaster management:**The aim is to review of the recently contributed research articles based on machine learning for disaster management. Using this review we are trying to identify the source of data, and utilized machine learning algorithms.
2. **Provide an overview of unsupervised learning algorithms:**In this section we provide an overview of two popular unsupervised learning algorithms namely k-means clustering and Fuzzy C Means (FCM) clustering. Additionally the methodology is described to employ these algorithms to the text data analysis.
3. **Performance comparative performance study of unsupervised learning algorithm:**In this phase the performance of the clustering algorithms are measured and compared to find an appropriate method for future extension prospective.

In this section we provide an overview of the proposed work involved in this paper. The next section includes the review of recent literature about the disaster management using machine learning techniques. Further the machine learning method is discussed to analyze the text data. Further the performance of the implemented algorithms is analyzed and their comparison performed. Finally, the conclusion of the work has been discussed and future plan of study will e discussed.

II. RELATED WORK

In this section, we provide the available solutions for the rapid response in disaster management problems based on Machine Learning (ML)by recent researchers. The importance of disaster management is evident by the increasing number of natural and manmade disasters such as Irma and Manchester attacks. The estimated cost of the Irma hurricane is more than 80 billion USD; more than 40 lives have been lost and thousands were misplaced. Disaster management plays a key role in reducing the human and economic losses. **M. Aqib et al [1]** have developed a disaster management system that uses VANET, cloud computing, and simulations to evacuation strategies. They extend earlier work by using deep learning to predict urban traffic behavior. Moreover, use GPUs to deal with intensive nature of deep learning algorithms. They are first to apply deep learning in disaster management and use real-world road traffic through the UK Department for Transport. The results demonstrate the effectiveness of deep learning approach in disaster management and prediction of traffic behavior in emergency situations.

Recent years include the world's hottest year, besides the COVID-19, by climate-related disasters, based on data collected by the Emergency Events Database. Besides the human losses, disasters cause catastrophic socioeconomic impacts, including economic losses. Recent developments in artificial intelligence,ML and deep learning have been used to better cope with the severe and catastrophic impacts. **V. Linardos et al [2]** aims to provide an overview of the studies, presented since 2017, focusing on ML and DL methods for disaster management. In particular, focus has been given on the areas of disaster and hazard prediction, risk and vulnerability assessment, disaster detection, early warning systems, disaster monitoring, damage assessment and post-disaster response. Furthermore, recently developed ML and DL applications for disaster management have been analyzed.

Social media serves as an integral part of the crisis response following emergency (disaster) event. Regardless of the kind of disaster, whether it is a hurricane, a flood, an earthquake or a man-made disaster like a riot or a terrorist attack, social media platforms have proven to be a powerful facilitator of communication and coordination between victims and communities. Several research articles have been published on social media utilization for disaster response. Many of those discuss automated ML approaches to extract disaster indicating posts, useful for coordination. There is a scarcity of review of all the major research pertaining to the utilization of ML approaches for disaster response using social media. **L. Dwarakanath et al [3]** reviews domain and classifies them across three disaster phases – early warning and detection, post-disaster coordination and response, damage assessment. This review help in choosing further topics to automated approaches for actionable information classification and disaster coordination and help emergency teams to make informed decisions in disaster situations.

Flood control and disaster management deal with reduction or prevention of devastating effects. **H. S. Munawar et al [4]** is developed an understanding about the flood risks, explore the existing systems for managing the risks and flood management to overcome the gaps in the current. Floods as a disaster will be viewed in detail along with an analysis of the threats to community and economy. The identification of a problem and working towards a solution for managing the damage and destruction by flood is crucial. The use of technology for disaster response is also investigated with its limitations. Various advancements in the disaster management are explored and the gaps are identified. Finally, a ML and image processing based solution is proposed. The system is formulated so as to overcome the limitations of the current technologies and present a robust and reliable model.

R. Veloso et al [5] present datasets of Brazilian disasters from January 2003 to February 2021, through reports from government and institutes. The datasets include 9 types of disaster, and number of affected people during 18-year for 5,402 municipalities, totaling more than 65,000 observations. Data on geographical, demographic and socioeconomic aspects of the municipalities are provided. The data address a number of applications using supervised and unsupervised ML techniques. The dataset can be useful for analyses and allows the visualization. The data can be used to optimization related to logistics and/or disaster management. Describe two real-world cases for the location-allocation. They use the datasets and other information such as costs and distances.

Human activity is always a reflection of its external conditions. If a group of people is in emergency, then their activities and behavior will be different as compared to normal conditions. To detect an emergency, Human Activity Recognition (HAR) (such as shouting, running here and there, crying, searching for an exit door) can play an important role. By detecting the emergency and its degree, the Emergency Management System (EMS) can manage the situation. **S. Nanda et al [6]** use ML algorithms such as Random Forest (RF), IBK, Bagging, J48 and MLP on Smartphone and Smartwatch Activity and Biometric Dataset for human activity and RF is found the best algorithm with accuracy 87.1977%.

Current disaster management procedures suffer from a number of shortcomings like high temporal lags or limited temporal and spatial resolution. **B. Resch et al [7]** analyze social media posts to assess the footprint of damage caused by natural disasters through combining ML (Latent Dirichlet Allocation) for semantic information extraction with spatial and temporal analysisfor hot spot detection. The results demonstrate that earthquake footprints can be reliably and accurately identified. A number of relevant topics can be identified without a priori knowledge, revealing clearly differing temporal and spatial signatures. Furthermore, they are able to generate a damage map that indicates where losses have occurred. The validation results using statistical measures, complemented by the official earthquake footprint by US Geological Survey and the results of the HAZUS loss model, shows that approach produces valid and reliable outputs. Thus, this approach may improve current disaster management procedures.

When reference data is not available to assess the quality of OpenStreetMap (OSM) data, its metadata can be used. **A. Madubedube et al [8]** applied unsupervised ML for analyzing OSM history data to understand who contributed when and how. Even though no statements can be made about the quality of the data, the results provide insight into the quality. Most of the data in Mozambique was contributed by small groups (25%). Results revealed a new class: contributors who were new and attracted by HOT mapping events during disaster relief. More studies of the world would establish whether the patterns observed. Intrinsic methods cannot replace ground truth or extrinsic methods, but provide ways for gaining insight about quality, and can also be used to inform efforts to improve the quality. They provide suggestions for contributor-focused intrinsic quality assessments.

Social media platforms provide information for disaster response. ML could be used to identify such information. Supervised learning relies on labeled data, which is not readily available for disaster. While labeled data available for a prior source, supervised classifiers learned only from the source disaster not perform well. **H. Li et al [9]** propose a domain adaptation approach, which learns classifiers from unlabelled data, in addition to source labeled data. This approach uses the Naïve Bayes, with an iterative Self-Training strategy. Results on identifying tweets relevant to a disaster are show that the domain adaptation classifiers are better as compared to the supervised classifiers.

Prediction models of heavy rain damage using ML were developed for the Seoul Capital in the Korea. **C. Choi et al [10]** used data on the occurrence of heavy rain damage from 1994 to 2015 and weather big data. The model was developed by ML techniques such as decision trees, bagging, random forests, and boosting. Result of evaluating the prediction performance, the AUC value of the boosting model using meteorological data from the past 1 to 4 days was the highest at 95.87%. By using the prediction model to predict the heavy rain damage for each region, greatly reduce the damage through proactive management.

A common approach to multisource data fusion is to combine in a stacked vector. The classifiers are transformed by integrating the contextual information from neighboring pixels to improve the accuracy. The decision level fusion approach combines statistical and ML techniques. It treats data source separately using support vector machines (SVM) and uses a Probabilistic Neural Network (PNN) to fuse. This approach consists of pixel-level or feature-level fusion. The results of the disaster management are presented by **B. Gokaraju et al [11]**. For levee landslide, they used the multi-temporal datasets of air-borne synthetic aperture radar sensor (UAVSAR). For Tornado disaster, used multi-source and multi-temporal datasets synthetic aperture radar sensor (RADARSAT-2) and multispectral sensor (RapiEye) datasets. The results of fusion outperformed the non-data fusion techniques in both with kappa accuracies of 82.8% and 72%.

Coronavirus pandemic has manifested in the form of fear and panic, fueled by incomplete and inaccurate information. Therefore, it is a need to address and understandinformational crisis and gauge sentiment, so appropriate messaging and policy can be implemented. **J. Samuel et al [12]** identify public sentiment with the pandemic specific Tweets and R. They demonstrate insights into the progress of fear-sentiment over time as COVID-19 peaks, using textual analytics and visualizations. They provide an overview of two ML methods, and compare in classifying Tweets. They observe a classification accuracy of 91% with the Naïve Bayes and logistic regression accuracy of 74% with shorter Tweets, and both methods showed weaker performance for longer Tweets.

Social networks are used for communications and help related requests. During disaster, such requests need to be mined for timely help. The sentiment of people during and after the disaster determines the success of the disaster response and recovery. **J. R. Ragini et al [13]** propose an approach for disaster response through sentiment analysis. The model collects disaster data from social networks and categorizes them according to the needs of people. The categorized data are classified through ML algorithm. Features like, parts of speech and lexicon are analyzed to identify the best classification strategy. The results show that lexicon based approach is suitable for analyzing the needs. The practical implication of the methodology is the real-time categorization and classification of social media data for disaster response and recovery.

People use social media to share information during disasters and emergencies. Information on social media, in the early hours are valuable for responders and decision makers, helping them gain situational awareness and plan relief. Processing social media content to obtain such information involves solving challenges, including parsing brief and messages, handling overload, and prioritizing types of information. **M. Imran et al [14]**highlights challenges and presents techniques to deal with social media messages, focusing on crisis scenarios.

World Health Organization declared COVID-19 as pandemic. After that Ministry of Home Affairs, India decided to treat COVID-19 as a “notified disaster”, leading to a complete shutdown. This has affected all sectors of the country including the education. The near-total closure of schools, colleges, and universities has disrupted academic activities. **A. Khattar et al [15]** survey to understand the day to day living, activities, learning styles, and mental health of students during this crisis and assess how they are adapting to the e-learning and how they are managing social lives.

M. Rahman et al [16] proposes an approach by integrating statistical, ML, and multi-criteria decision analysis. Flood inventory and flood causative factors were prepared using remote sensing data and the Mike-11 hydrological model and data from different sources. The flood inventory was divided into training and testing data, where 334 locationswere used for training and the 141 locations for testing. Using the area under the receiver operating curve (AUROC), predictive power of the model was tested. The results revealed that LR model had the highest success rate (81.60%) and prediction rate (86.80%). Furthermore, different combinations of the models were evaluated and the best combination was used for generating a flood hazard map for Bangladesh. The performance of 11C integrated models was evaluated using the AUROC and found that LR-FR model had the highest predictive power with an AUROC of 88.10%.

Excessive withdrawal of groundwater, and other natural resources, has been introduced as land subsidence and the earth fissuring. Fissuring is turning into the disasters and responsible for economic, social, and environmental loss. Modeling the fissure hazard is important for water management, and enforces the groundwater recharge policies. Modeling the formation of fissures and prediction of the hazardous areas has been challenge. **B. Choubin et al [17]** proposing ML models for prediction of fissuring hazards. The Simulated annealing feature selectionwas used to identify features, and the generalized linear model (GLM), multivariate adaptive regression splines (MARS), classification and regression tree (CART), random forest (RF), and SVM have been used for prediction. Results indicated that all the models had good accuracy (> 86%) and precision (> 81%). GLM model had the lowest performance, while the RF model was the best model.

III. PROPOSED WORK

The required model for recovering the disaster content using the machine learning algorithm is provided in figure 1. The model includes the data set as the primary component of the demonstrated system. The dataset is obtained from the Kaggle [16], the dataset contains 5 attributes and 7503 instances of data. The attributes are ID, keyword, location, text and target. The ID is unique values defined for each instances there for we eliminate this attribute, additionally, the keyword and location has a lot of missing values therefore we also removed these two attributes also. Therefore after data preprocessing it only contains the sentiment text which indicates the disaster event and target or class attribute which is defined by 0 and 1 values. Here, 0 indicates the non disaster event and 1 defines the disaster event. This preprocessed dataset is again used with text based preprocessing technique.

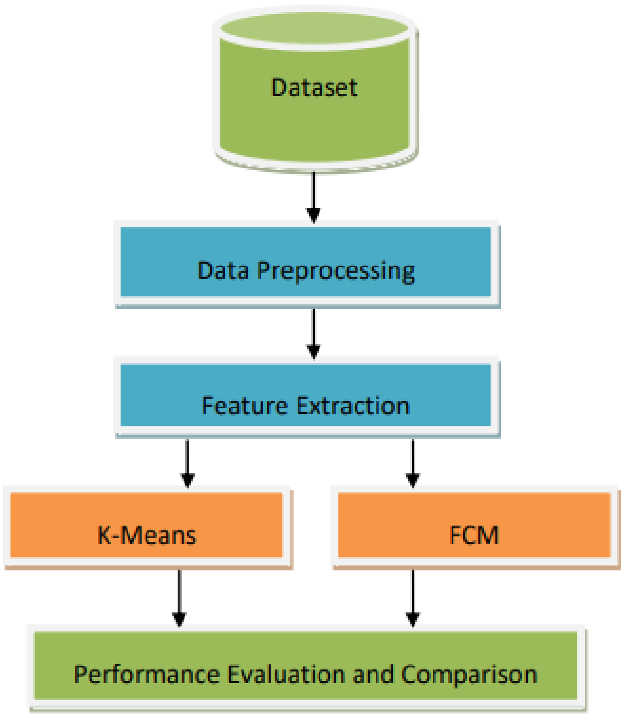


Figure 1: Proposed Model for comparing unsupervised learning technique on social media disaster data

The aim of text preprocessing is to reduce the noise of the social media contents such as abbreviations, stop words and special characters. Using the python based preprocessing library we eliminate the stop words, special characters and abbreviations. Further, in order to Vectorize the text for utilizing with the machine learning algorithms we utilize the Term Frequency and Inverse

Document Frequency (TF-IDF) concept. That is a classical method of feature selection in text processing techniques. The TF-IDF is defined using the following formula:

$$\mathrm{TF} = \frac{\text{count of a term in a document}}{\text{total term in document}}$$

And

$$\mathrm{IDF} = \log\left(\frac{\mathrm{N}}{\mathrm{df(t)}}\right)$$

Where $\mathrm{df(t)}$ is Document frequency of a term t, and N is Number of documents containing the term t. Finally for computing the features in normalized manner we need to calculate the weights for each term using the following formula:

$$\mathrm{w} = \mathrm{tf} * \mathrm{IDF}$$

The weights are used for extracting the features form the text and transformed the data into a two dimensional matrix. Now the k-means and FCM clustering is used for calculating the clusters of the data.

K-Means Clustering: The traditional k-means algorithm is implemented for cluster the entire text in dataset. The classical k-means algorithm is given as. Let there are N instance of data to be cluster $(x_i, x_2 \ \dots \ x_n)$, and the number of clusters k = 2.The algorithm return k clusters based on dissimilarity between each instance and its nearest cluster center. First select k instances as initial cluster centers$(m_1, m_2, \dots, m_k)$. Then calculate the distance between each object x_i and each cluster center, further assign each instance to the nearest cluster. The formula for calculating distance is given as:

$$d(x_i, m_i) = \sqrt{\sum_{j=1}^{d} \left(x_i - m_{j1}\right)^2}, i = 1 \dots N, j = 1 \dots k$$

$d(x_i, m_i)$is the distance between data i and cluster j.

Calculate the mean of objects in each cluster as the new cluster centers,

$$m_i = \frac{1}{N}\sum_{j-1}^{n_i} x_{ij}, i = 1,2, \dots, K$$

N_i is the number of samples of current cluster i;

Repeat these steps until the algorithm return$(m_1, m_2, \dots, m_k)$optimal clusters.

Fuzzy C Mean (FCM): FCM is working similar as K-means clustering algorithm. The principle distinction is that, rather than utilizing a hard choice for data grouping it can assign multiple classes to the data instance. This technique is first created by Dunn in 1973 and then enhanced by Bezdek in 1981. It depends on minimization of the objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m \parallel x_i - c_j \parallel, \qquad 1 \le m < \infty$$

Where m is any real number greater than 1, $u_{ij}$is the degree of membership of x_i in the cluster j, x_i is the i^th^ of d-dimensional measured data, c_j is the d-dimension center of the cluster, and ||*|| is expressing the similarity between data and cluster center.

Fuzzy clustering is carried out through an optimization process of the objective function, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left[\frac{\| x_i - c_j \|}{\| x_i - c_k \|}\right]^{\frac{2}{m-1}}}$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij} \,.\, x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

After applying the clustering algorithms the experiments are carried out and performance of both the models are calculated. The obtained performance is reported in the next section.

IV. RESULTS ANALYSIS

This section describes the conducted experiments and obtained evaluation results. In this context, the different training and validation sets are prepared for experiments. We have defined 70-30, 75-25 and 80-20 percentage of training and validation sets.

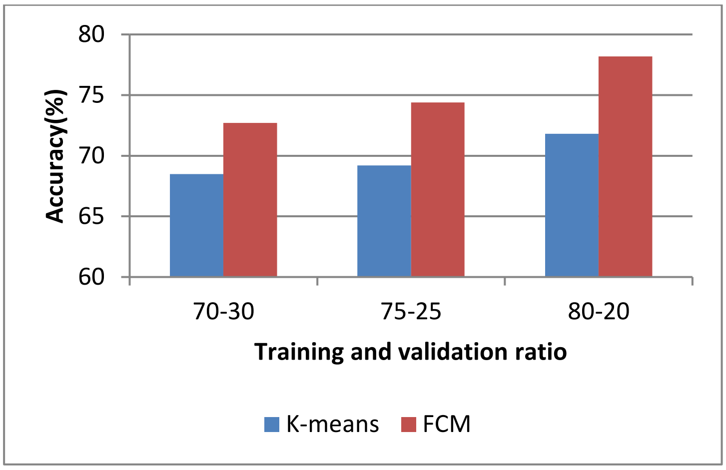
The accuracy of the models is calculated for the given training and validation sets. The accuracy of the model is calculated as the ratio of correctly predicted targets and total instances given for prediction. That can be calculated using:

$$\text{accuracy} = \frac{\text{correctly predicted samples}}{\text{total samples}} \text{X100}$$

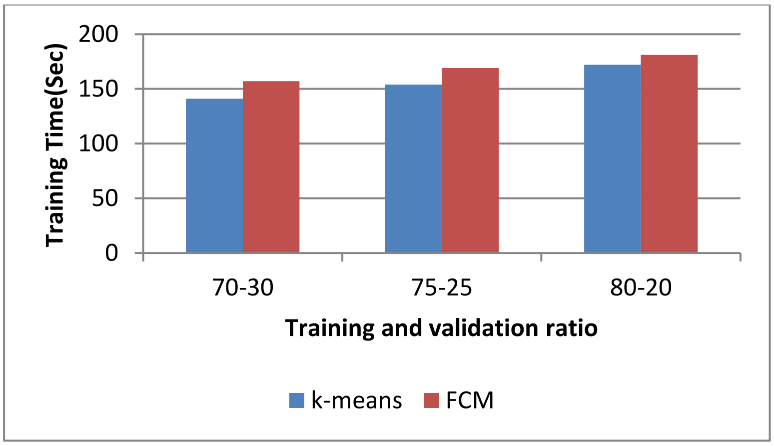
Table 1: Performance of Clustering Algorithms

		Accuracy (%)		Training time (Sec)	
S. No.	**Training and validation ratio**	**k-means**	**FCM**	**k-means**	**FCM**
1	70-30	68.5	72.7	141	157
2	75-25	69.2	74.4	154	169
3	80-20	71.8	78.2	172	181

The figure 2(A) demonstrates the accuracy of both the clustering algorithms for identifying the disaster events using the social media data. In figure X axis demonstrate the training and validation ratio for experiments and Y axis shows the obtained accuracy in terms of percentage (%). According to the obtained results it is found that the increasing amount of training sample will improve the categorization performance.



(a)



(b)

Figure 2 shows the performance in terms of (A) Accuracy and (B) Training Time

Next evaluation parameter is training time of the algorithms, which is calculated using the following formula:

$$\text{training time} = \text{end time} - \text{start time}$$

Figure 2(B) demonstrates the training time of both the clustering algorithms. The X axis demonstrates the training and validation ratio used for experiments, additionally Y axis shows the training time in terms of seconds (Sec). According to the obtained results we found that the increasing amount of training time will increase the training time. Based on the performance measured using both the parameters we found that the FCM is more accurate then the k-means clustering algorithm, but the time requirement of the FCM algorithm ishigher as compared to k-means algorithm.

V. CONCLUSION

The social media platform is a tool for distributing information worldwide. This platform is used for sharing information to the target users as well as to all the social media users. Therefore it becomes very beneficial for various applications. In this paper we utilize the social media data in order to identifying the disaster events. In this context the unsupervised learning technique is used for classifying the twits into the disaster and non disaster events. The FCM and k-means clustering is employed on the text data for identifying the disaster events. In this context the entire work is divided in three main parts, first a review is conducted, next the model for processing the text is described and then the performance of the implemented model is measured. Based on the obtained performance FCM is found more accurate then k-means clustering algorithm, but the FCM is slower than the k-means algorithm. However, FCM is expensive algorithm but provide more accurate results therefore in near future the FCM algorithm is modified for obtaining more accurate information extraction. Additionally for future application the process is implemented to automatic data extraction and processing from the social media for more effective system development.

REFERENCES

1. M. Aqib, R. Mehmood, A. Albeshri, A. Alzahrani, "Disaster Management in Smart Cities by Forecasting Traffic Plan Using Deep Learning and GPUs", Institute for Computer Sciences, Social Informatics and Telecomm.Engineering 2018, LNICST 224, pp. 139–154, 2018.
2. V. Linardos, M. Drakaki, P. Tzionas, Y. L. Karnavas, "Machine Learning in Disaster Management: Recent Developments in Methods and Applications", Mach. Learn. Knowl. Extr. 2022, 4, 446–473
3. L. Dwarakanath, A. Kamsin, R. A. Rasheed, A. Anandhan, L. Shuib, "Automated Machine Learning Approaches for Emergency Response and Coordination via Social Media in the Aftermath of a Disaster: A Review", IEEE Access, VOLUME 9, 2021
4. H. S. Munawar, "Flood Disaster Management: Risks, Technologies, and Future Directions", Machine Vision Inspection Systems (Vol. 1): Image Processing, Concepts, Methodologies and Applications, (115–146) © 2020 Scrivener Publishing LLC
5. R. Veloso, J. Cespedes, A. Caunhye, D. Alem, "Brazilian disaster datasets and real-world instances for optimization and machine learning", Data in Brief 42 (2022) 108012
6. S. Nanda, C. R. Panigrahi, B. Pati, A. Mishra, "A Novel Approach to Detect Emergency Using Machine Learning", Progress in Advanced Computing and Intelligent Engineering, Advances in Intelligent Systems and Computing 1199, 2021
7. B. Resch, F. Usländer, C. Havas, "Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment", Cartography and Geographic Information Science, 2018, Vol 45, No 4, 362–376
8. A. Madubedube, S. Coetzee, V. Rautenbach, "A Contributor-Focused Intrinsic Quality Assessment of OpenStreetMap in Mozambique Using Unsupervised Machine Learning", ISPRS Int. J. Geo-Inf. 2021, 10, 156
9. H. Li, D. Caragea, C. Caragea, N. Herndon, "Disaster response aided by tweet classification with a domain adaptation approach", J Contingencies and Crisis Management. 2018;26:16–27
10. C. Choi, J. Kim, J. Kim, D. Kim, Y. Bae, H. S. Kim, "Development of Heavy Rain Damage Prediction Model Using Machine Learning Based on Big Data", Hindawi Advances in Meteorology Volume 2018, Article ID 5024930, 11 pages
11. B. Gokaraju, R. A. A. Nóbrega, D. A. Doss, A. C. Turlapaty, R. C. Tesiero, "Data Fusion of Multi-Source Satellite Data Sets For Cost Effective Disaster Management Studies", 978-1-4673-9558-8/15/$31.00 ©2015 IEEE
12. J. Samuel, G. G. Md. N. Ali, M. M. Rahman, E. Esawi, Y. Samuel, "COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification", Information 2020, 11, 314;
13. J. R. Ragini, P. M. R. Anand, V. Bhaskar, "Big data analytics for disaster response and recovery through sentiment analysis", International Journal of Information Management, 42, 13–24, 2018

14. M. Imran, C. Castillo, F. Diaz, S. Vieweg, “Processing Social Media Messages in Mass Emergency: Survey Summary”, International World Wide Web Conference Committee, WWW ’18 Companion, April 23–27, 2018, Lyon, France
15. A. Khattar, P. R. Jain, S. M. K. Quadri, “Effects of the Disastrous Pandemic COVID 19 on Learning Styles, Activities and Mental Health of Young Indian Students - A Machine Learning Approach”, Proceedings of the International Conference on Intelligent Computing and Control Systems, IEEE, 2020
16. M. Rahman, C. Ningsheng, M. M. Islam, A. Dewan, J. Iqbal, R. M. A. Washakh, T. Shufeng, “Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis”, Earth Systems and Environment, 2019
17. B. Choubin, A. Mosavi, E. H. Alamdarlood, F. S. Hosseinid, S. Shamshirband, K. Dashtekiang, P. Ghamisih, “Earth fissure hazard prediction using machine learning models”, Environmental Research 179 (2019) 108770