Grenze International Journal of Engineering and Technology, June Issue

Disaster Response Classification Pipeline using NLP and Machine Learning

Lekha Shri LG[1], Sanmuga Priya M[2] and Barakkath Nisha U[3]
[1-3]Sri Krishna College of Engineering and Technology/M. Tech CSE, Coimbatore, India
Email: lekhashri369@gmail.com, sanmugapriyam, barakkathnisha}@skcet.ac.in

***Abstract*— Disasters, whether natural or human-made, generate vast amounts of text-based communications on social media and other digital platforms. Effectively classifying these messages is crucial for timely responses from emergency services and humanitarian organizations. This paper presents a Disaster Response Classification Pipeline using Natural Language Processing (NLP) and Machine Learning (ML) to categorize messages into 36 predefined disaster-related categories. The project follows a structured ETL (Extract, Transform, Load) pipeline to process raw messages, an ML pipeline to train a classification model, and a Flask-based web application for real-time message classification. The dataset, sourced from Figure Eight, contains multilingual messages from past disasters. Experimental results demonstrate the effectiveness of machine learning models in classifying messages with high accuracy, making this system valuable for real-world disaster management efforts.**

***Index Terms*— Disaster Response, Natural Language Processing (NLP), Machine Learning, Text Classification, Social Media Analysis, Emergency Communication, ETL Pipeline, Flask Web Application.**

I. INTRODUCTION

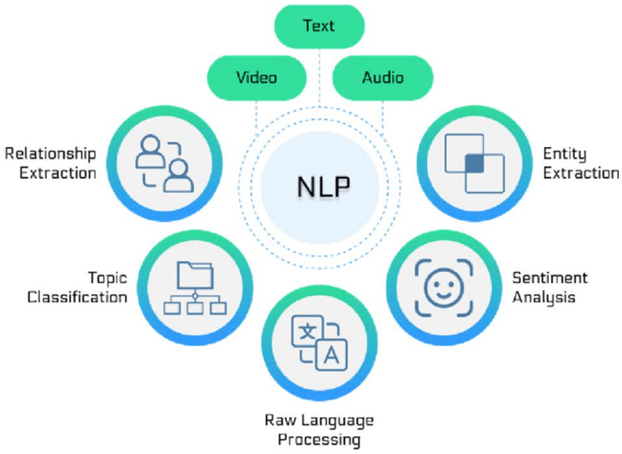

Fig 1

Grenze ID: 01.GIJET.11.2.1085_22

In the wake of increasing natural and human-made disasters, the ability to efficiently process and respond to crisis-related communications is crucial for effective disaster management. Social media and digital platforms serve as primary channels where affected individuals, relief organizations, and government agencies share urgent updates, requests for aid, and situational reports. However, the sheer volume of messages makes it difficult to filter critical information and respond promptly. Conventional methods of disaster response rely on manual processing, which is inefficient, slow, and prone to misclassification. This lack of automation and accuracy can delay life-saving interventions and hinder resource allocation.

Natural Language Processing (NLP) is a subfield of artificial intelligence (AI) that enables computers to understand, interpret, and process human language. NLP combines computational linguistics with machine learning to analyse and extract meaningful insights from large volumes of text. It is widely used in applications such as text classification, sentiment analysis, speech recognition, and translation systems. In the context of disaster response, NLP plays a crucial role in filtering and categorizing disaster-related messages, ensuring that emergency responders can quickly identify critical information from social media, news reports, and other text-based sources.

Machine Learning (ML) is a branch of AI that enables systems to learn from data and make predictions or decisions without being explicitly programmed. ML models identify patterns in data and use statistical techniques to improve their accuracy over time. Supervised learning, a common ML technique, is used in disaster classification models where labelled datasets train algorithms to recognize different types of disaster-related messages. By leveraging ML, the system can automatically classify messages, detect anomalies, and prioritize critical alerts, making disaster response faster and more efficient.

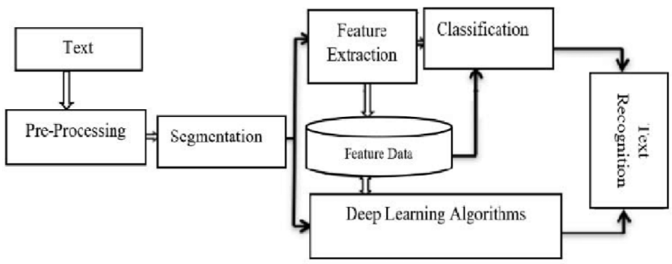

Fig 2

By integrating NLP and ML, this project creates a robust, automated disaster classification system that enhances situational awareness, reduces manual workload, and facilitates quicker decision-making in emergency situations.

This paper presents a Disaster Response Classification Pipeline that leverages NLP and Machine Learning to automate the categorization of disaster-related messages, enhancing the efficiency and accuracy of crisis communication analysis. By integrating NLP techniques for text preprocessing and feature extraction with ML algorithms for classification, the system enables real-time filtering and categorization of messages, ensuring that critical information reaches emergency responders promptly. Additionally, the model continuously learns from new data, leading to improved accuracy and adaptability, which helps refine predictions and reduce misclassification over time. Furthermore, the automation of disaster message classification facilitates faster decision-making and resource allocation, allowing emergency teams to identify and prioritize urgent cases effectively. Thus, the proposed NLP and ML-based pipeline provides a scalable, automated, and efficient solution for managing crisis communication, ultimately helping emergency teams respond faster and more effectively during disasters.

II. Literature Review

Viyom Mittal, Hongmiao Yu, and K. K. Ramakrishnan highlight the critical role of timely information delivery to first responders during disasters, emphasizing the growing reliance on social media platforms like Twitter and Facebook for crisis communication. Their study leverages Natural Language Processing (NLP) and Machine Learning (ML) to classify and forward relevant disaster-related messages to the appropriate emergency teams. The goal is to retrieve and deliver only actionable information in real-time, improving emergency response. Their work examines the entire pipeline—from retrieving tweets to classification and dissemination—and introduces improvements in data retrieval, relevance prediction, and prioritization through NLP and ML techniques.

S. Deepa Lakshmi and T. Velmurugan explored NLP-based methods for classifying disaster-related tweets, addressing the challenges of unstructured social media text. Their research utilized preprocessing techniques like tokenization, stop-word removal, lemmatization, and TF-IDF transformation to refine text before classification. They applied machine learning classifiers, including SVM, MLP, Ada-boost, and Multinomial Naïve Bayes (NB), identifying the best-performing model for disaster message classification. Their study highlights the importance of automated classification in disaster response, aiding emergency teams in making timely decisions. Our proposed Disaster Response Classification Pipeline builds upon this work by integrating a full NLP and ML pipeline, categorizing messages into 36 predefined disaster response categories, and deploying the system in real-time via a Flaskbased web application.

Akshat Anand and D. Rajeswari highlighted the role of social media, particularly Twitter, in crisis communication. Their research proposed an application that retrieves georeferenced tweets using web scraping and Twitter's API to provide real-time disaster-related information. By leveraging CNN-1D, LSTM, and Glove-100D models, their system ensures accurate classification of disaster-related tweets, aiding emergency responders. Social media facilitates disaster preparedness, real-time alerts, and recovery by offering geotargeted data. While addressing concerns about misinformation and privacy, the study aligns with our Disaster Response Classification Pipeline, which also utilizes NLP and ML for tweet classification. Our work enhances theirs by structuring classifications into predefined categories and deploying a realtime web-based system for efficient emergency response.

Qixuan Hou, Meng Han, and Zhipeng Cai [3] explored the role of social media as a vast information source with real-time updates and diverse data. Their survey examined existing studies and applications, outlining a standard pipeline for social mediabased analytics. Key techniques discussed include topic analysis, time series analysis, sentiment analysis, and network analysis, with applications in disaster management, healthcare, and business. The study also highlighted challenges such as data privacy, 5G integration, and multilingual support. Their research aligns with our Disaster Response Classification Pipeline, which also utilizes social media data for disaster response, focusing on NLP and ML for efficient tweet classification and crisis communication.

The literature emphasizes social media's role in disaster management, leveraging NLP and ML to extract actionable insights from platforms like Twitter. Key techniques include text preprocessing, feature extraction (TF-IDF, Glove), and classification models (SVM, MLP, AdaBoost, Naïve Bayes, CNN-LSTM). Real-time tweet retrieval via web scraping and Twitter APIs enhances response efficiency. Addressing challenges like data privacy, misinformation, and multilingual support, emerging solutions such as blockchain-based smart contracts and 5G integration improve reliability. These studies inform our Disaster Response Classification Pipeline, which enhances structured categorization, real-time deployment, and scalability for emergency response teams.

III. Proposed Methodology

Natural Language Processing (NLP) and Machine Learning (ML) provide a transformative solution to these challenges. By leveraging NLP techniques, disaster-related messages can be automatically classified into predefined categories such as medical assistance, food shortages, infrastructure damage, and rescue requests. Machine learning models can process largescale textual data in real-time, ensuring swift and accurate categorization of messages. This automated approach significantly enhances disaster response efforts by enabling authorities and humanitarian organizations to identify urgent needs and deploy resources accordingly.

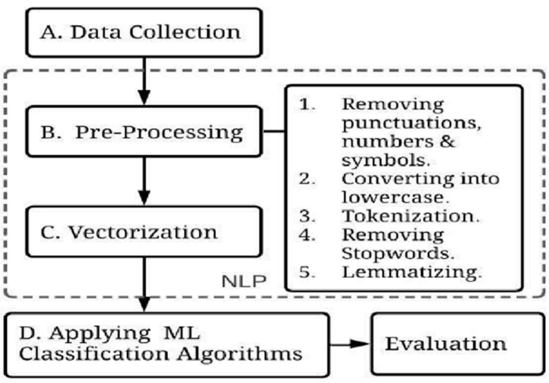

Fig 3

This paper introduces a Disaster Response Classification System designed to streamline and enhance disaster-related communication analysis. The system employs an end-to-end pipeline that extracts, cleans, and processes disaster messages, transforming them into structured data for efficient classification. Unlike conventional systems, it integrates realtime message filtering, classification, and prioritization, ensuring that emergency responders receive only the most relevant and actionable information. The system features a userfriendly interface, allowing both technical and non-technical users to input messages and receive instant categorizations into predefined disaster categories such as medical help, shelter, food, security, transportation, and weather-related events. By leveraging Natural Language Processing (NLP) and Machine Learning (ML), the system improves the accuracy and speed of disaster message classification, facilitating timely and effective crisis response.

In addition to classification, the system incorporates advanced analytics modules to enhance disaster response efficiency. The Message Classification Analytics Module analyzes categorized messages to identify trends in disaster requests, helping prioritize urgent responses. The Real-Time Disaster Trend Analytics Module tracks disaster patterns over time, enabling proactive resource planning. The Geospatial Disaster Impact Analytics Module maps crisis events, providing heatmaps for efficient aid distribution. The Response Effectiveness Analytics Module evaluates relief operations by tracking response times and impact assessments. Additionally, the Misinformation and Data Quality Analytics Module detects unreliable messages and filters out misleading content, ensuring accurate disaster response. By integrating these analytical capabilities, the system enhances situational awareness, optimizes resource allocation, and accelerates emergency response efforts, making it an essential tool for disaster management organizations and government agencies.

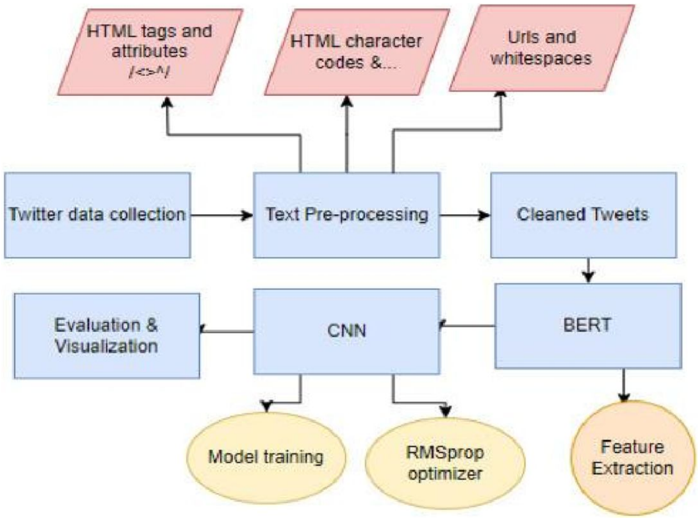

Fig 4

Unlike traditional rule-based categorization, which often fails to handle variations in language, slang, and multilingual content, the proposed system learns patterns from data and continuously improves its classification accuracy. Furthermore, it supports multi-category labelling, allowing a single message to be assigned to multiple relevant categories for a more precise and contextual understanding. The system is designed to integrate seamlessly with existing disaster response workflows, making it a scalable and adaptable solution for organizations worldwide.

By leveraging the power of NLP and ML, this system ensures faster decision-making, improved situational awareness, and efficient allocation of disaster relief resources. The automated classification of messages significantly reduces the burden on human analysts, allowing them to focus on highpriority tasks and strategic planning during emergencies. The adoption of such intelligent disaster response systems is a crucial step toward building more resilient and responsive disaster management frameworks.

A. User Dashboard Module

Description: Provides users with a intuitive interface to classify disaster response messages. The dashboard prompts users to "Enter a message to classify", which is then analyzed by the NLP and Machine Learning pipeline. Upon clicking the "Classify Message" button, the system categorizes the message into relevant disaster response categories.

Significance: This module enhances disaster response efforts by enabling rapid and accurate classification of messages. This streamlines the response process, allowing for timely and effective resource allocation. The dashboard's user-friendly interface and instant classification capabilities improve overall efficiency and decision-making.

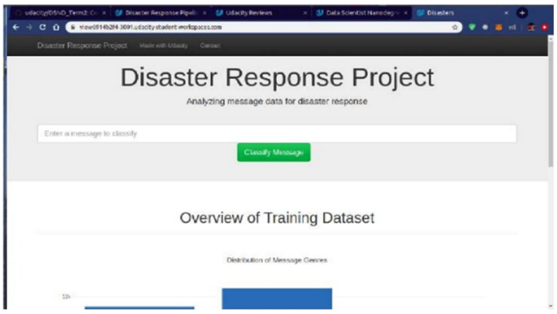

Fig 5

B. Distribution of Message Genres Module

Description: The Disaster Response Classification Pipeline analyzes and categorizes disaster-related messages from sources like social media, news, and direct reports. It classifies messages into predefined genres—Direct, News, Social, and Others—and visualizes their distribution using a graph (X-axis: message genres, Y-axis: message count). Using NLP and machine learning, the system preprocesses text, extracts features, and ensures accurate classification.
Significance: This module enhances situational awareness by organizing messages for efficient emergency response. It helps authorities detect trends, prioritize critical reports, and allocate resources effectively. By distinguishing between direct reports, news updates, and social media discussions, it ensures that urgent messages receive immediate attention, reducing manual filtering efforts.

C. Distribution of Disaster module

Description: The Disaster Response Classification Pipeline Using NLP and Machine Learning is an advanced system that categorizes disaster-related messages into specific disaster types to aid emergency response teams. After processing incoming messages using Natural Language Processing (NLP) and Machine Learning (ML), the system classifies them into predefined disaster categories such as Request, Medical Help, Security, Military, Child Help, Water, Food, Shelter, Money, Missing People, Death, Transport, Electricity, Shops, Weather Related, Floods, Storm, Fire, Earthquake, and Others. The classification results are visualized through a graph, where the X-axis represents the various disaster categories and the Y-axis denotes the number of messages classified under each type.
Significance: This module provides critical insights into disaster trends, helping emergency teams prioritize responses based on the most reported disaster types. By visualizing the distribution of messages, authorities can identify urgent needs, allocate resources effectively, and streamline disaster management efforts. The system ensures real-time classification, reducing the time required for manual filtering and enabling faster decision-making. It helps organizations detect large-scale emergencies (e.g., earthquakes, floods) versus localized crises (e.g., medical help, missing persons), ensuring targeted intervention and better coordination among response teams. The automated classification and visualization improve situational awareness, enhancing preparedness and disaster relief efficiency.

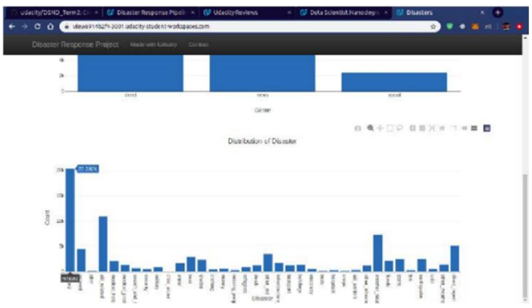

Fig 6

D. Message Classification Module

Description: In the Disaster Response Classification Pipeline Using NLP and Machine Learning, when a user inputs a message such as "Please health care equipments and medicines", the system processes the text using NLP techniques to classify it into predefined disaster-related categories. After clicking the "Classify Message" button, the model analyses the text, applies feature extraction (like TF-IDF or word embeddings), and predicts relevant categories. The output highlights specific categories from request, medical-help, security, military, child help, water, food, shelter, money, missing people, death, transport, electricity, shops, weather-related, floods, storm, fire, and earthquake. In this case, the system incorrectly classifies the message under "Weather-Related" and "Storm", instead of the expected "Medical-Help", indicating a possible misclassification due to keyword overlap or insufficient training data.
Significance: This classification system is crucial in disaster response as it helps emergency teams quickly identify and prioritize urgent messages, ensuring faster assistance. Incorrect classifications can mislead responders, delaying critical help. Continuous improvement of the model, through better data training, enhanced NLP techniques, and real-time feedback integration, will ensure higher accuracy and reliability in crisis situations.

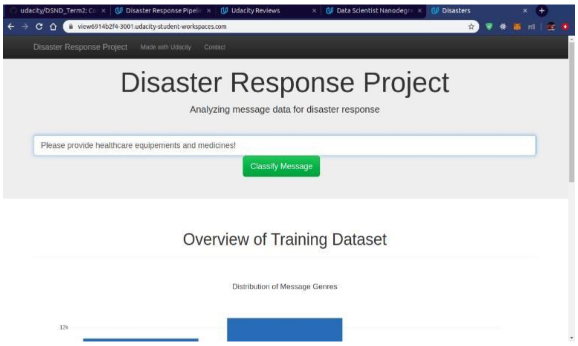

Fig 7

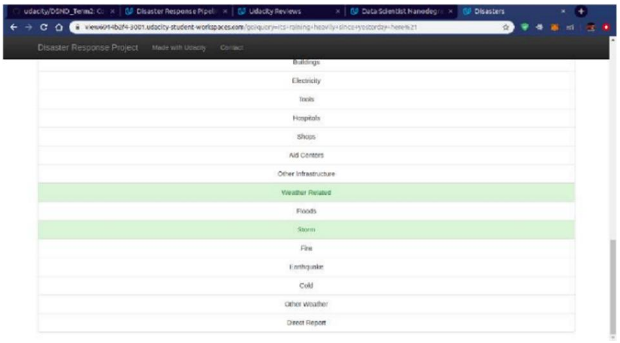

Fig 8

E. Message Classification Module

Description: This module processes incoming disasterrelated messages and categorizes them into predefined classes such as medical help, food, shelter, security, and transportation. Using NLP and machine learning models like CNN-LSTM, it ensures accurate classification of messages for effective emergency response.
Significance: Enables automatic sorting of critical messages, reducing manual effort and ensuring that relevant authorities receive the right information quickly for timely intervention.

F. Real-Time Disaster Trend Analytics Module

Description: Utilizes web scraping and API integration to gather real-time disaster-related messages from social media platforms like Twitter. It extracts geo-referenced data and detects emerging disasters or crises based on message trends.
Significance: Enhances situational awareness by providing real-time insights, allowing authorities to respond proactively and allocate resources efficiently during emergencies.

G. Disaster Impact Visualization Module

Description: Displays graphical representations of classified disaster messages, showing the distribution of different disaster types and their frequency. The visualizations include bar graphs and heatmaps to highlight affected regions and severity levels.
Significance: Helps decision-makers and emergency teams quickly assess the scale of disasters, prioritize response efforts, and allocate resources based on data-driven insights.

H. Emergency Resource Allocation Module

Description: Based on classified messages, this module suggests appropriate resource allocation strategies by identifying the most urgent needs in different disaster zones. It can prioritize aid distribution based on message trends and severity levels.
Significance: Optimizes resource deployment, ensuring that essential aid such as medical supplies, food, and shelter reaches the most affected areas first, reducing delays and inefficiencies.

I. Misinformation Detection and Data Validation Module

Description: Detects and filters out false or misleading disaster-related information using NLP techniques and credibility assessment models. It cross-checks data sources and flags unreliable content before classification.
Significance: Prevents the spread of misinformation during crises, ensuring that only verified and relevant information is used for disaster response, thereby improving the reliability of decision-making.

J. Response Effectiveness Analytics Module

Description: Evaluates the efficiency of disaster response efforts by analyzing message resolution times, response rates, and trends in public sentiment regarding relief operations.
Significance: Helps organizations improve emergency response strategies by identifying bottlenecks, assessing relief impact, and optimizing resource distribution.

K. Geospatial Disaster Impact Analytics Module

Description: Ensures that all financial transactions and Description: Maps classified disaster messages to geographic locations, generating heatmaps and regional impact reports. It identifies disaster-prone areas and monitors the spread of crisis events.
Significance: Enables authorities to allocate resources effectively by identifying high-risk zones and directing emergency aid to the most affected regions.

IV. FUTURE SCOPE

The Disaster Response Classification Pipeline can be enhanced by integrating deep learning models like transformerbased architectures (BERT, RoBERTa) for improved text understanding and classification accuracy. Expanding multilingual support will allow the system to process tweets in various languages, improving global disaster response. Incorporating real-time streaming analytics and predictive modeling can help identify potential disasters before they escalate. Additionally, integrating blockchain for secure data handling and 5G for faster communication will enhance reliability and efficiency.

V. CONCLUSION

The Disaster Response Classification Pipeline effectively utilizes NLP and ML to extract, categorize, and deliver disasterrelated information from social media in real time. By structuring and filtering vast amounts of unstructured data, the system enhances decision-making for emergency responders. The integration of real-time tweet processing, classification, and deployment makes it a valuable tool for crisis management. With further advancements in deep learning, real-time analytics, and multilingual support, this system has the potential to revolutionize disaster response strategies, ensuring faster and more effective aid during emergencies.

REFERENCES

[1] SD Lakshmi, T Velmurugan(March 2023) - Deriving Pipeline for Emergency Services Using Natural Language Processing Techniques, Research gate. https://www.researchgate.net/publication/369407856

[2] Viyom Mittal, Hongmiao Yu, KK Ramakrishnan(2022). FUSED: Fusing Social Media Stream Classification Techniques for Effective Disaster Response, IEEE Access. https://ieeexplore.ieee.org/document/9805386

[3] A Anand, D Rajeswari - Classification of Disaster Tweets Using Natural Language Processing Pipeline, Research gate. https://www.researchgate.net/profile/RajeswariDevarajan/publication/369407856_Deriving_Pipeline_for_Emergency_Services_Using_Natural_Language_Processing_Techniques/links/6423cbe5a1b72772e434cc46/Deriving-Pipelinefor-Emergency-Services-Using-Natural-Language-Processing-Techniques.pdf

[4] Kusumasari B, Prabowo NPA (2020) Scraping social media data for disaster communication: how the pattern of Twitter users affects disasters in Asia and the Pacific. Nat Hazards 103(3):3415–3435. https://doi.org/10.1007/s11069-020-04136-z

[5] Qixuan Hou, Meng Han, Zhipeng Cai(November 2020). Survey on data analysis in social media: A practical application aspect, IEEE Access. https:// ieeexplore.ieee.org/document/9259192

[6] Anand A, Patel R, Rajeswari D (2022) A comprehensive synchrzation by deriving fluentpipeline and web scraping through social media for emergency services. In: 2022 International conference on advances in computing, communication and applied informatics (ACCAI), pp 1–8. https://doi.org/10.1109/ACCAI53970.2022.9752629

[7] Seddighi H, Seddighi S, Salmani I, Sharifi Sedeh M (2020) Public-private-people partnerships(4P) for improving the response to COVID-19 in Iran. Disaster Med Public Health Prep 1–6. https://doi.org/10.1017/dmp.2020.202

[8] Khattak AM., et al.(2020) - Tweets classification and sentiment analysis for personalized tweets recommendation". Complexity (2020).

[9] Seddighi H, Salmani I (2020) Saving lives and changing minds with twitter in disasters andpandemics: a literature review, pp 59–77

[10] Lamsal R and Kumar TV. “Twitter-based disaster response using recurrent nets”. In Research Anthology on Managing Crisis and Risk Communications (2023): 613-632.

[11] Gulati K., et al.(2022) - Comparative analysis of machine learning based classification models using sentiment classification of tweets related to COVID-19 pandemic”. Materials Today: Proceedings 51. https://www.sciencedirect.com/science/article/pii/S221478532 1032843

[12] Balogun AL., et al.(2020) - “Assessing the potentials of digitalization as a tool for climate change adaptation and sustainable development in urban centres”. https://www.sciencedirect.com/science/article/abs/pii/S221067 0719307267

[13] Didi Yosra., et al.(2022) - COVID-19 Tweets Classification Based on a Hybrid Word Embedding Method”. Big Data and Cognitive Computing 6.2

[14] Rustam Furqan., et al. “Tweets classification on the base of sentiments for US airline companies”. Entropy 21.11 (2019). https://www.mdpi.com/1099-4300/21/11/1078

[15] Rastenis J., et al.(2021) - Multi-language spam/phishing classification by email body text: Toward automated security incident investigation. Electronics 10.6. https://www.researchgate.net/publication/350025483_MultiLanguage_SpamPhishing_Classification_by_Email_Body_Text_ Toward_Automated_Security_Incident_Investigation

[16] Eyal N (2012) Repeat triage in disaster relief: questions from Haiti. PLoS Curr 115(12):700–701. https://doi.org/10.1371/currents

[17] Feldman D et al (2021) - Communicating flood risk: Looking back and forward at traditionaland social media outlets. Int J Disaster Risk Reduct 15:43–51. https://doi.org/10.1016/j.ijdrr. 2015.12.004

[18] Yin J, Karimi S, Robinson B, Cameron M (2021) ESA: emergency situation awareness viamicrobloggers. In: Proceedings of the 21st ACM international conference on information and knowledge management, pp 2701–2703. https://doi.org/10.1145/2396761.2398732

[19] Kumar S, Barbier G, Ali Abbasi MA, Liu H (2019) TweetTracker: an analysis tool for humanitarian and disaster relief. In: Fifth international AAAI conference on weblogs and social media, no. April 2014, pp. 661–662. https://doi.org/10.1145/1935826.1935854

[20] Aaron Smith, Monica Anderson(2018) - Social Media Use in 2018. Research Gate https://www.pewresearch.org/internet/2018/03/01/socialmedia-use-in