ResilientNet-Real-Time Disaster Responseand Verification System

A project report submitted in partial fulfillment ofthe
requirements for the degree of

Bachelor of EngineeringIn
Computer Engineering

by

Kris Corriea (9185)

Hisbaan Sayed (9432)

Omkar Satupe (9232)

Mathew Lobo (9204)

Under the guidance of

Ms. Supriya Kamoji

Assistant Professor

DEPARTMENT OF COMPUTER ENGINEERING

Fr. Conceicao Rodrigues College of Engineering, Bandra (W), Mumbai -400050

University of Mumbai

(2023-24)

This work is dedicated to my family.

I am very thankful for their motivation and support.

Internal Approval Sheet

CERTIFICATE

This is to certify that the project entitled **" ResilientNet-Real-Time Disaster Response and Verification System "** is a bonafide work of **Kris Corriea (Roll No. 9185), Hisbaan Sayed (9432), Omkar Satupe (9232) and Mathew Lobo (9204)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of Bachelor in **Computer Engineering.**

Ms. Supriya Kamoji
Assistant Professor

Dr. Sujata Deshmukh
Head of Computer Engineering

Dr. S. S. Rathod
Principal

Approval Sheet Project

Report Approval

This project report entitled by **ResilientNet-Real-Time Disaster Response andVerification System** by **Kris Corriea (Roll No. 9185) Hisbaan Sayed (9432), Omkar Satupe (9232) & Mathew Lobo (9204)** is approved for the degree of Bachelor of Engineering in Computer Engineering.

Examiner 1. ————————————

Examiner 2. ————————————

Date:

Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Kris Corriea (9185)

Hisbaan Sayed (9432)

Omkar Satupe (9232)

Mathew Lobo (9204)

Date:

Abstract

Responding to India's urgent need for effective disaster management, we introduce ResilientNet, an innovative system leveraging real-time big data processing and advanced AI technologies. ResilientNet gathers diverse multimedia content from a wide range of social media services, including Twitter, Instagram, Facebook, etc., and utilises the GEMINI API, enabling comprehensive analysis and verification. Data is stored in the NEO4J database and visually represented on a user-friendly website dashboard for easy accessibility and insights. This research explores the efficacy of crowdsourced fact-checking, contributing to a novel disaster-focused tweet verification system. ResilientNet's amalgamation of crowdsourcing and AI creates a comprehensive graph of critical metrics and trends, enabling authorities to counter misinformation and direct disaster response efforts efficiently.

Keywords:

Disaster Management, NEO4J Database, Tweet Verification, Tweet Classification, Knowledge Graph, BERT

Acknowledgments

We have great pleasure in presenting the report on **"ResilientNet-Real-Time Disaster Response and Verification System".** I take this opportunity to express my sincere thanks towards the guide Prof. Supriya Kamoji, C.R.C.E, Bandra (W), Mumbai, for providing the technical guidelines, and the suggestions regarding the line of this work. We enjoyed discussing the work progress with him/her during our visits to department.

We thank Dr. Sujata Deshmukh, Head of Computer Engineering department, Principal and the management of C.R.C.E., Mumbai for encouragement and providing necessary infrastructure for pursuing the project.

We also thank all non-teaching staff for their valuable support, to complete our project.

Kris Corriea (9185)

Hisbaan Sayed (9432)

Omkar Satupe (9232)

Mathew Lobo (9204)

Date:

Contents

List of Figures

List of Tables

Glossary

AI Artificial Intelligence. 1

GPT Generative Pre-trained Transformers. 2

API Application programming interface. 2

NER Named Entity Recognition. 3

CNN Convolutional Neural Network. 3

MANET Mobile Ad hoc Network. 4

LSTM Long short-term memory. 9

Bi-LSTM Bi-Directional Long Short-Term Memory. 9

BERT Bidirectional Encoder Representations from Transformers. 9

RoBERTa Robustly Optimized BERT Approach. 9

Chapter 1

Introduction

1.1 Introduction

Due to the complexity of disaster management caused by the rapid dissemination of misinformation, this research paper presents ResilientNet, an innovative framework that aims to redefine the disaster response capabilities of India. ResilientNet, which incorporates crowdsourced fact-checking, advanced AI technologies, and real-time big data processing, is a revolutionary solution that effectively tackles existing deficiencies. By prioritizing information gaps and inaccuracies, this framework aims to establish a cohesive and adaptable system that facilitates precise disaster assessment and coordinated response. Amidst India's ongoing efforts to enhance disaster management, ResilientNet emerges as an exemplar of ingenuity, presenting not only technological progressions but also a paradigmatic transformation in the realm of disaster resilience. To ensure effective disaster management in India, ResilientNet implements state-of-the-art artificial intelligence and real-time data processing. ResilientNet powers data visualization and real-time disaster assessment through the utilization of sophisticated fact-checking tools and social media. By providing authorities with actionable insights to counter misinformation, ResilientNet facilitates the coordination of disaster response and response efforts. Our artificial intelligence and crowdsourced fact-checking technologies transform disaster management. Our objective at ResilientNet is to usher in a new era of disaster preparedness and response in India, as well as to encourage timely and well-informed decision-making.

1.2 Motivation

ResilientNet responds to India's critical demand for efficient disaster management, employing state-of-the-art AI and real-time data processing. By leveraging social media content and advanced fact-checking tools, ResilientNet provides instant disaster assessment and data visualization. ResilientNet equips authorities with actionable insights to combat misinformation, enabling effective disaster response and coordination. Our unique blend of crowdsourced fact-checking and AI technology represents agroundbreaking shift in disaster management. Through ResilientNet, we aspire to drive timely and well-informed decision-making, ushering in a new era of disaster preparednessand response across India.

1.3 Objectives

- Develop the ResilientNet Framework: Create an integrated platform that utilizes real-time big data processing, advanced AI algorithms, and crowdsourced fact-checking to enhance disaster management.
- Data Aggregation and Verification: Efficiently collect and analyze multimedia content from Twitter, employing GEMINI APIs to verify and establish the credibility of disaster-related information.
- NEO4J Database Integration and Visualization: Utilize the NEO4J graph database to store and visually represent aggregated data, providing a user-friendly dashboard for authorities to access actionable insights and real-time trends.
- Accurate Location and Severity Assessment: Design algorithms for precise disaster location pinpointing and severity assessment, enabling authorities to make informed decisions during response efforts.
- Dynamic Querying in NEO4J: Enhance ResilientNet's functionality by enabling dynamic querying within the NEO4J database based on specific needs, allowing authorities to extract relevant information efficiently for tailored response strategies.

Chapter 2

Literature Review

Disaster management is an essential aspect of public safety and emergency response. In light of India's urgent need for effective disaster management, the introduction of ResilientNet, an innovative system leveraging real-time big data processing and advanced AI technologies, has the potential to revolutionize disaster response efforts. This literature review aims to explore the existing research findings related to big data processing, AI technologies, and IoT applications in disaster management. The review will also highlight the knowledge gaps and suggest potential future research directions to address the efficacy of crowdsourced fact-checking and tweet verification systems in disaster management.

2.1 Big Data Processing and AI Technologies

[22] covered big data and machine learning in environmental and water management methods, applications, and future directions. The study suggested big data and machine learning could solve environmental problems. [14] and [23] also discussed big data analytics and IoT's growing role in disaster management, including recent advances, taxonomy, and prospects. ResilientNet uses big data and AI to gather diverse multimedia content from social media platforms and analyze and verify it using the GPT 3.5 API. [1] A test F1-Score of 0.92 for DisKnow's CNN-based classifier was nearly 10% higher than the current state-of-the-art. The improvement is due to Twitter-specific embeddings, model architecture changes, and noise-reducing pre-processing. CNN-based event-type classification model from DisKnow had a test F1-Score of 0.882. It wasn't as effective as modern methods. Not enough fact-checking, real-time analytics, and data visualization. The trustworthiness filtering component of DisKnow was not evaluated in this paper. The DisKnow classification models and Named Entity Recognition (NER) component are ongoing and should improve, according to the paper. However, it does not specify areas for improvement or experimental results for these components. The paper suggests adding new node types to DisKnow's knowledge graph architecture to provide more information. The paper states that the knowledge graph becomes incomprehensible for human use as it grows. It does not propose or discuss automatic knowledge graph comprehension methods.

[7] says they used multiple algorithms. The document suggests using the Cosine Similarity of Tweet Vector Sum (CSTVS), Dot Product of Search Term Vector and Tweet Vector Sum (DP), Mean Cosine Similarity (MCS), and SCSSC for text analysis. The DP formula performed best. The document also mentions gradient descent-based decision trees in Random Forest (RF) and Gradient-Boosted (GB) models. Additionally, the study used non-parametric logistic regression (LR). Hurricane Irma was the only natural disaster examined in the paper.

2.2 IoT Applications in Disaster Management

[13] and [16] discussed the disaster management implications of IoT. They stressed how IoT and big data analytics can build disaster-resilient smart cities and regional environmental monitoring and management systems. [14] examined the City Geospatial Dashboard, which uses IoT and big data analytics for disaster management geospatial solutions. These findings support ResilientNet's use of IoT for real-time data analysis in disaster management. [2] demonstrate the system's ability to detect negative disaster-related social media posts, identify negativity factors, and pinpoint disaster locations with high accuracy. The performance evaluation, examples of integrating social media and historical event datasets, deep learning techniques, and clear advice and prediction of imminent disasters are not provided. Social media allows people to geolocate their observations and help disaster management, according to [9]. However, social media data's use in disaster management was still poorly understood, highlighting the need for more research.

2.3 Crowdsourced Fact-Checking and Tweet Verification

[21] and [22] indirectly support disaster management crowdsourced fact-checking and tweet verification. ResilientNet uses crowdsourcing and AI to create a graph of critical metrics and trends to help authorities fight misinformation and coordinate disaster response. Crowdsourced fact-checking in disaster management is not well-studied. Further research is needed to fill this knowledge gap.

Chapter 3

Problem Statement

3.1 Drawbacks of Existing System

- Delayed Data Integration: Conventional disaster management systems often encounter challenges in integrating real-time data swiftly, leading to delays in accurate disaster assessment.
- Misinformation Challenges: The prevalence of misinformation during crises poses a significant threat, and current systems may lack robust mechanisms topromptly verify and filter accurate information.
- Complex Data Representation: Traditional frameworks may struggle with providing a user-friendly and intuitive representation of complex relationships within disaster-related data, hindering efficient decision-making for authorities.

3.2 Solution To Above Problem

To develop ResilientNet, an innovative disaster management solution for India, integrating real-time big data processing, AI technologies, and crowdsourced fact-checking to enhance accurate disaster assessment, timely response, and informed decision-making.

Chapter 4

Project Description

4.1 Overview of the project

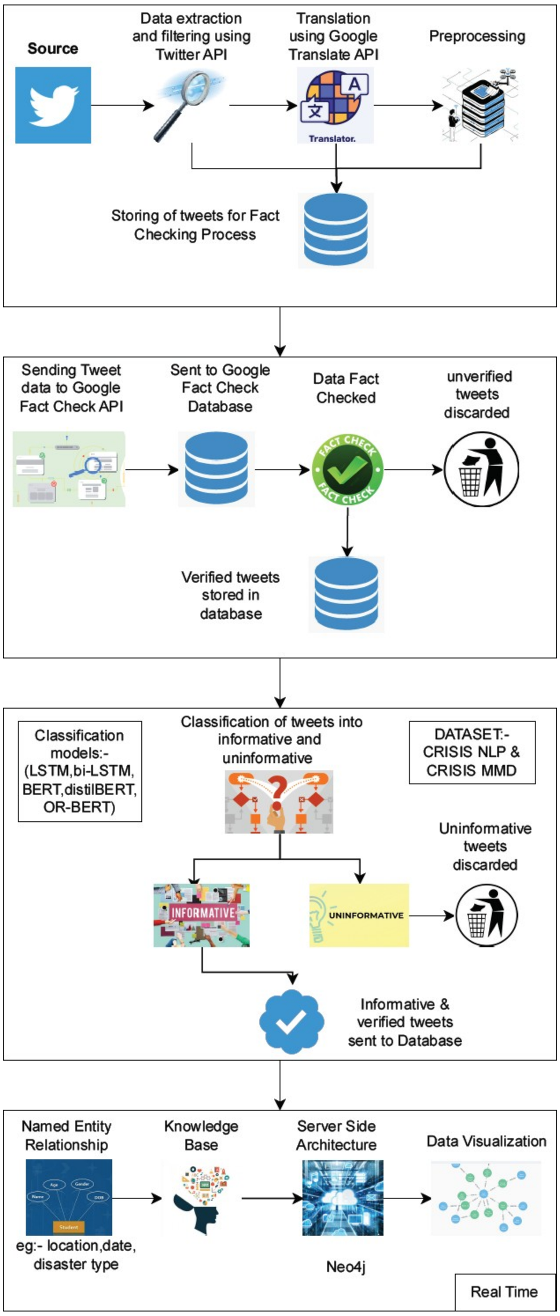

Figure 4.1: Architecture Model

The ResilientNet framework comprises several modules designed to enhance disaster management through advanced technologies:

1. Fact-Checking Module: Integrates GEMINI into a React application to verify user-provided statements from Twitter data. Transmits user input to GEMINI for analysis and evaluates responses to determine truth value. Contributes to a more trustworthy digital environment by providing fact-checking reports.

2. Preprocessing Module: Refines Twitter data by removing non-essential elements like hashtags and emoticons. Ensures subsequent analytical processes operate on the most pertinent content, minimizing noise in the dataset.

3. Classification Module: Categorizes tweets as disaster-related versus non-disaster-related and informative versus non-informative using the DistilBERT model. The distinction between informative and non-informative tweets ensures the prioritization of actionable disaster-related information.

4. Named-Entity-Relationship Module: Utilizes spaCy for Named Entity Recognition (NER) due to its speed, customization, and pre-trained models. Enables rapid and accurate extraction of relevant information from tweets, crucial for disaster management.

5. NEO4J Database and Visualization Module: Utilizes Neo4j as a graph database management system for representing complex, interconnected data in disaster management. Facilitates intuitive modeling of disaster-related entities and relationships, enabling informed decision-making and resource optimization. Enables real-time updates, historical data analysis, predictive analysis, and collaboration among disaster responders.

Highlights and Proposed Methodology

1. Extracting tweets from Twitter using the Twitter API. Creating a Twitter developer platform for this purpose. Using real-time tweets from Twitter.
2. After extraction, data cleaning processes are performed. Converting all the tweet data into English using Translate API.
3. After converting to English, Fact-checking the tweet data using GEMINI, using classification algorithms to classify tweets into informative and non-informative.
4. Extracting disaster-related informative data from the tweet.
5. Integrate it with neo4J.
6. Graphic visualization of the real-time data.

4.2 Module Description

4.2.1 Fact-Checking module

A React application module that verifies and validates user-supplied information via the Gemini API. The objective of this module is to bolster the credibility of the digital environment through the verification of information that is encountered.

The constituents:

1. User Input Gathering: Individuals provide assertions, data, or claims for the purpose of fact verification, with the possibility that they indicate Twitter as the originating platform. This input is collected by the React application prior to its transmission to the Gemini API.

2. Integration with Gemini API: The React application transmits the user's input to the Gemini API for the purpose of analysis. To establish a connection with Gemini, employ the Google AI-provided Gemini API.

3. Response Processing: Analyze and assess the Gemini responses that have been received. Extract pertinent information, claims, or expressions from the output of the model.

4. Fact-Checking Algorithm: Utilize the functionalities of Gemini to validate the data against reputable sources, which may include Twitter if the user so specifies. Construct a fact-checking algorithm that utilizes Gemini's analysis to evaluate the veracity of the user-supplied statement. Based on the analysis, ascertain the truth value (true, false, uncertain, informative/non-informative). Potential use of the database-stored results in training future fact-checking models.

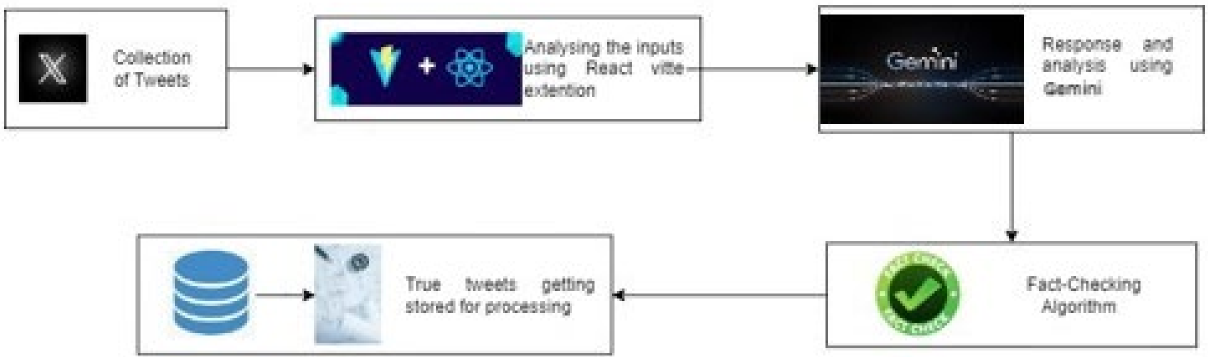

Figure 4.2: Dataflow Architecture

4.2.2 Preprocessing Module

The preprocessing module in ResilientNet is a fundamental element in our project, playing a crucial role in handling the large amount of Twitter data and efficiently extracting valuable insights relevant to disaster management. Essentially, the main purpose of this module is to improve the quality of tweets by removing unnecessary and irrelevant information that is commonly found in social media data streams. Our preprocessing module carefully eliminates unnecessary elements like hashtags, emoticons, and irrelevant text to ensure that subsequent analytical processes focus on the most relevant and meaningful content in the tweets.

The extensive coverage of disaster-related multimedia content on Twitter is one of the key factors that influenced our decision to choose CrisisMMD as the dataset for our preprocessing module. CrisisMMD covers a broad range of both natural and human-caused disasters, including floods, earthquakes, wildfires, accidents, conflicts, and industrial incidents. ResilientNet's extensive coverage allows it to analyse and respond to a wide range of disaster scenarios, making it more useful and applicable in different situations.

In addition, CrisisMMD offers highly valuable annotated data that plays a crucial role in training ResilientNet's algorithms using supervised learning methods. ResilientNet can utilise annotated data to accurately detect and differentiate disaster-related information within Twitter data streams. This improves its capacity to provide valuable insights and aid disaster management authorities in making prompt and well-informed decisions.

The preprocessing module of ResilientNet is closely connected to the main goal of our project, which is to utilise real-time social media data to enhance disaster management efforts. Our objective is to improve ResilientNet's ability to analyse large volumes of Twitter data and extract useful information by using CrisisMMD as our dataset. This will enable disaster management authorities to make informed decisions by providing them with actionable insights. Our goal is to enhance disaster management practices by utilising real-time social media data analytics.

4.2.3 Classification Module

The classification module in ResilientNet is crucial for categorising tweets based on two important criteria: whether they are related to disasters or not, and whether they provide useful information or not. Using the DistilBERT model, tweets are classified as either disaster-related (target: 1) or unrelated (target: 0) at the beginning. The choice of

DistilBERT is a deliberate decision based on its strategic advantages, as it is highly efficient and effective in maintaining a large portion of the performance of its larger counterpart, BERT, while significantly reducing the amount of computational resources needed. This ability to scale makes it feasible to handle large amounts of tweets in real-time situations where disaster response is required.

Moreover, DistilBERT's expertise in understanding contextual information in text is crucial for accurately categorising tweets related to disasters and distinguishing between informative and non-informative content. The binary categorization acts as the main filter for identifying tweets that are relevant to disaster situations. This helps in prioritising and validating data efficiently.

Afterwards, a more detailed categorization is performed within the group of tweets related to disasters to differentiate between informative tweets that provide practical information about disasters and non-informative tweets that contain metaphorical or irrelevant content. This more detailed categorization ensures that only relevant and useful disaster-related information is kept, improving the system's ability to verify and prioritise such data from Twitter's large dataset.

The importance of this module lies in its capacity to simplify the process of identifying and prioritising valuable information related to disasters. This, in turn, enhances the effectiveness of ResilientNet in assisting disaster management authorities with timely and actionable insights obtained from social media data. ResilientNet enables decision-makers to make well-informed decisions and allocate resources effectively during disaster response efforts by precisely categorising tweets according to their relevance and informativeness.

4.2.4 Theoretical Analysis of Algorithms

Algorithm	**Key Features**	**Advantages**	**Efficiency**
LSTM	Sequential data processing	Captures sequentialpatterns in text	Computationally intensive
Bi-LSTM	Bidirectional sequential processing	Considers context fromboth directions	Moderate
BERT	Pre-trained transformer-basedmodel	Captures contextual understanding	Relatively slower

RoBERTa	Enhanced version of BERT	Improved pre-training process	Slower than DistilBERT
DistilBERT	Lighter and fasterversion of BERT	Retains much of BERT's performance	Efficient andfaster

4.2.5 Algorithms Used

4.2.5.1 LSTM

Long Short-Term Memory (LSTM) systems speak to a significant headway within the domain of repetitive neural systems (RNNs), designed to address the confinements characteristic in conventional RNN designs.

At its center, LSTM is planned to delude the vanishing slope issue experienced in customary RNNs, which hinders the network's capacity to capture long-range conditions inside consecutive information. The design of an LSTM organize comprises specialized units, or cells, prepared with three entryways:

the input entryway, disregard gate, and yield entryway, alongside a cell state. Each door serves an unmistakable work in controlling the stream of data through the cell, empowering the arrangement to hold or dispose of data based on its pertinence.

The input door administers the degree to which modern data is joined into the cell state, leveraging sigmoid actuation to tweak input values. At the same time, the disregard door, too controlled by sigmoid actuation, decides the degree to which earlier data is held or eradicated from the cell state. This instrument empowers the arrange to specifically protect data considered related while disposing of insignificant or obsolete information, in this manner relieving the effect of the vanishing angle issue.

Vitally, LSTM systems are expanded by the cell state, acting as a transport belt to encourage the seamless engendering of data over progressive time steps. This coherence guarantees that germane data continues over delayed groupings, empowering the arrange to observe perplexing designs and conditions inalienable in consecutive information.

Besides, the yield door, controlled by sigmoid and tanh actuations, directs the dispersal of data from the cell state to the yield of the LSTM unit, typifying the network's capacity to produce precise expectations or classifications based on the prepared input grouping.

The fundamental operation of LSTM systems unfurls over discrete time steps, wherein input information is successively bolstered into the arrange, and the inside states of LSTM cells are iteratively overhauled. Through a handle of forward engendering and backpropagation, the organize refines its parameters to play down the dissimilarity between anticipated and genuine results, in this way improving its prescient ability over time.

4.2.5.2 Bi-LSTM

Bidirectional Long Short-Term Memory (Bi-LSTM) systems speak to an impressive advancement of conventional LSTM designs, invested with the capacity to gather experiences from both past and future settings inside consecutive information. This increase engages Bi-LSTM systems to capture a more comprehensive understanding of transient flow, in this manner improving their adequacy over a range of consecutive modeling assignments, from opinion investigation to discourse acknowledgment.

At its pith, a Bi-LSTM organize comprises two LSTM layers working in pairs, each handling the input grouping in contradicting transient headings: One layer forms the grouping from past to future (forward LSTM), whereas the other navigates it from future to past (in reverse LSTM). This bidirectional traversal empowers the organize to acclimatize data from both going before and succeeding time steps, encouraging an encompassing comprehension of the worldly setting basic the input grouping.

Inside each LSTM layer, the basic components stay practically equivalent to those found in conventional LSTM systems. Each layer is invested with specialized units, or cells, lodging input entryways, disregard entryways, yield entryways, and cell states. These components organize the stream of data through the organize, empowering it to specifically hold or dispose of germane information while circumventing the vanishing slope issue characteristic of customary RNN structures. Amid the forward pass, the forward LSTM layer forms the input arrangement consecutively, capturing conditions from past to future. At the same time, the in reverse LSTM layer attempts a parallel traversal of the input arrangement in turn around, observing designs from future to past. The yields of both LSTM layers are at that point concatenated, amalgamating experiences gathered from both worldly headings.

This bidirectional combination prepares Bi-LSTM systems with a nuanced understanding of the successive information, empowering them to perceive perplexing designs and relationships that might escape unidirectional partners. By leveraging data from both past and future settings, Bi-LSTM systems show improved versatility to worldly ambiguities and are proficient at modeling complex conditions inside successive information.

Besides, Bi-LSTM systems exceed expectations in assignments requiring an encompassing comprehension of consecutive information, such as common dialect preparation and signal recognition. Their capacity to capture bidirectional conditions renders them especially proficient at assignments requiring relevant understanding and nuanced induction.

4.2.5.3 BERT

Bidirectional Encoder Representations from Transformers (BERT) heralds a watershed moment in the domain of natural language processing (NLP), revolutionizing the landscape with its unparalleled ability to comprehend contextual nuances and infer semantic meaning from textual data. As a pre-trained language model, BERT epitomizes the convergence of cutting-edge techniques in deep learning and transformer architectures, culminating in a versatile framework capable of outperforming traditional NLP models across an array of tasks, including sentiment analysis, question answering, and named entity recognition.

At its core, BERT operates on the premise of unsupervised pre-training followed by task-specific fine-tuning, effectively leveraging vast corpora of unlabeled text to imbue the model with a rich understanding of language semantics. The architectural cornerstone of BERT resides in the transformer architecture, renowned for its efficacy in capturing long-range dependencies within sequential data while mitigating the computational inefficiencies associated with recurrent neural networks. BERT's transformer architecture comprises multiple layers of self-attention mechanisms and feedforward neural networks, orchestrated to encode contextual information from both preceding and succeeding tokens within a given input sequence. This bidirectional encoding imbues BERT with a nuanced understanding of the contextual nuances inherent in natural language, enabling it to discern intricate semantic relationships and disambiguate polysemous terms with unprecedented accuracy.

During the pre-training phase, BERT undergoes an extensive unsupervised learning regimen wherein it is exposed to copious volumes of text from diverse sources. Leveraging masked language modeling and next-sentence prediction objectives, BERT learns to predict masked tokens within input sequences and discern whether two consecutive sentences are contiguous or not. This pre-training paradigm equips BERT with a rich contextual understanding of language semantics, enabling it to encode text representations imbued with nuanced contextual information.

Following pre-training, BERT is fine-tuned on task-specific datasets through supervised learning, wherein the model's parameters are further refined to optimize performance on downstream NLP tasks. By fine-tuning task-specific data, BERT adapts its learned representations to the intricacies of the target task, exhibiting superior performance compared to models trained from scratch. Crucially, BERT's contextual embeddings encapsulate a wealth of semantic information, enabling the model to excel across diverse NLP tasks without necessitating task-specific architectural modifications. This versatility renders BERT a potent tool in the NLP practitioner's arsenal, offering unparalleled performance across a spectrum of tasks with minimal task-specific customization.

4.2.5.4 RoBERTa

RoBERTa, an extension of the BERT (Transformers Bidirectional Encoder Representations) architecture, represents a significant advance in the field of natural language processing (NLP) and is designed to overcome limitations and improve performance for a variety of language understanding tasks.It has been an evolution of the BERT model, RoBERTa builds on the foundations of its predecessor and introduces improvements and optimizations that significantly increase its effectiveness and versatility. The key innovation underlying RoBERTa is its training methodology. This is a departure from his BERT pre-training approach of leveraging larger datasets and exploring alternative training objectives.

In particular, RoBERTa applies a process of continuous unsupervised learning and exposes the model to a huge corpus of texts from different disciplines and genres.

This comprehensive training program enables RoBERTa to gain a more comprehensive understanding of the language's semantics and enrich its representation with a wider range of linguistic nuances and contextual subtleties. RoBERTa's architecture inherits BERT's transformer-based design and includes multiple layers of self-attention mechanisms and feedforward neural networks. These components work synergistically to encode contextual information from input sequences, facilitate the model's ability to recognize subtle semantic relationships, and infer meaning with unprecedented accuracy.

In the pre-training stage, RoBERTa goes through iterative training cycles to perform the task of predicting masked tokens in the input sequence and detecting whether two consecutive sentences are consecutive. However, unlike BERT, RoBERTa abandons the next goal of sentence prediction and instead focuses only on modeling masked speech. This change simplifies the pre-training task while allowing RoBERTa to use larger data sets more effectively, improving the model's robustness and generalization capabilities.

In addition, RoBERTa includes training hyperparameter improvements, such as increased batch size, lengthened training sequences, and dynamic masking strategies. These optimizations help the model capture extensive dependencies in text sequences and reduce the effects of noise and spurious correlations in large data sets. After pre-training, RoBERTa is fine-tuned through supervised learning on task-specific datasets to further tune the model's parameters to optimize the performance of her downstream NLP tasks. By fine-tuning task-specific data, RoBERTa adapts the learned representation to the complexity of the target task, achieving state-of-the-art performance across a variety of NLP benchmarks.

4.2.5.5 DistilBERT

DistilBERT, a refined form of the BERT (Bidirectional Encoder Representations from Transformers) model, speaks to a noteworthy breakthrough within the field of characteristic dialect handling (NLP), advertising a streamlined however profoundly effective system for dialect understanding assignments. Created with the point of lessening computational complexity and memory impression while protecting the prevalent execution of its forebear, DistilBERT accomplishes a commendable adjustment between demonstrate measure and computational assets, making it especially well-suited for sending in resource-constrained situations.

At its center, DistilBERT acquires the transformer-based design of BERT, comprising different layers of self-attention components and feedforward neural systems. In any case, not at all like its forerunner, DistilBERT utilizes a prepare of knowledge distillation to compress the first BERT demonstrate into a more compact frame. This distillation process includes preparing a smaller, understudy show (DistilBERT) to imitate the behavior and execution of a bigger, educator demonstrate (BERT) by leveraging the latter's pre-trained representations as supervision signals. Amid the refining handle, DistilBERT learns to inexact the representations and forecasts of BERT whereas following the limitations forced on show estimate and computational assets. This compression is accomplished through different procedures, counting parameter sharing, layer pruning, and information refining, wherein the understudy demonstrate is prepared to coordinate the logits (crude yield scores) and representations created by the instructor demonstrate.

By distilling the information typified inside BERT into a more compact shape, DistilBERT accomplishes critical decreases in show estimate and computational overhead without relinquishing execution. This compression empowers DistilBERT to be conveyed more productively over a range of NLP errands, from opinion investigation to content classification, while keeping up competitive levels of precision and strength. Besides, DistilBERT presents refinements in preparing flow and design plans to optimize execution and asset utilization. For occurrence, DistilBERT utilizes an altered consideration component known as "refined consideration," wherein consideration weights are scaled to energize sparsity and diminish computational overhead. Also, DistilBERT utilizes a smaller lexicon measure and token implanting measurement compared to BERT, advance decreasing memory requirements and deduction time.

Taking after the refining handle, DistilBERT experiences fine-tuning on task-specific datasets through administered learning, wherein the model's parameters are advanced and refined to optimize the execution of downstream NLP errands. By fine-tuning task-specific information, DistilBERT adjusts its learned representations to the complexities of the target errand, accomplishing competitive execution with altogether decreased computational assets.

4.2.6 Named-Entity-Relationship Module

SpaCy for Named Entity Recognition (NER) in ResilientNet is strategic for disaster management. Due to its customized NLP, SpaCy excels in tokenization, part-of-speech labeling, and named entity recognition. BERT is known for contextualized embedding, but its extensive architecture may be too much for NER, where spaCy excels. SpaCy's speed and lightweight implementation match emergency management's real-time needs. SpaCy's adaptability is ideal for processing a large number of tweets quickly. A well-documented API makes it user-friendly for researchers and developers, making it a practical choice for targeted NLP tasks. SpaCy's pre-trained named entity recognition models, which include date, time, and location, greatly reduce model training effort. This skill is ideal for disaster management, where fast and accurate information extraction is crucial. SpaCy's integration goes beyond NLP. ResilientNet stores and retrieves data in NEO4J, and SpaCy's seamless integration with other tools and databases makes it a better option than BERT in NLP pipelines. SpaCy is a wise choice for ResilientNet's Named Entity Recognition module. The combination of speed, user-friendliness, pre-trained models, and integration capabilities is ideal for disaster management.

```
Tweet: Flood in Mumbai on Tuesday at one pm, 1000 people affected
Word            POS             Tag             Label
Flood           NOUN            NN              NATURAL_DISASTER
in              ADP             IN              OTHER
Mumbai          PROPN           NNP             PLACE_AFFECTED
on              ADP             IN              OTHER
Tuesday         PROPN           NNP             DATE
at              ADP             IN              OTHER
one             NUM             CD              TIME
pm              NOUN            NN              TIME
,               PUNCT           ,               OTHER
1000            NUM             CD              NUM_PEOPLE_AFFECTED
people          NOUN            NNS             OTHER
affected                VERB            VBD             OTHER
```

Figure 4.2.6: Named-Entity-Relationship

4.2.7 NEO4J Database and Visualization Module

Neo4j, as a graph database management system, plays a crucial role in disaster management systems due to its unique capabilities, which precisely correlate with the complex and interconnected nature of disaster response and recovery efforts. Neo4j's capacity to represent data as a graph, with nodes representing entities and relationships representing their connections, is fundamental to its operation. This graph-based approach allows for the intuitive modeling of complex and interrelated components in disaster management, such as locations, resources, victims, and response teams. Neo4j excels at efficiently managing these relationships, allowing respondents to make well-informed decisions regarding resource allocation, coordination, and response strategies. In addition, Neo4j's geospatial capabilities make it ideally suited for disaster management because it can effectively manage spatial data, such as the locations of afflicted areas, resources, and infrastructure. This allows first responders to conduct geospatial analyses and make location-based decisions in real-time. Resource optimization is another crucial Neo4j function. By tracking the availability of resources and allocating them to areas in need, the system ensures an effective response. Real-time updates are managed without a hitch, enabling access to the most current and accurate information, a necessity in the dynamic context of disaster response. In addition, Neo4j allows for the storage and analysis of historical data pertaining to past disaster incidents, response efforts, and outcomes. By analyzing historical incidents, responders can learn from experience, identify best practices, and make data-driven enhancements to future response strategies. Neo4j's graph algorithms and query capabilities also enable predictive analysis, enabling responders to forecast the potential spread and impact of disasters and devise mitigation strategies. Neo4j's ability to represent communication channels and hierarchies within the graph facilitates collaboration and communication, thereby promoting effective coordination among various organizations and agencies involved in disaster management. Although Neo4j does not provide native role-based access control, application logic can be used to implement access control mechanisms, thereby enhancing data security and privacy. In conclusion, Neo4j's capabilities enable disaster responders to more effectively manage complex, interconnected data, thereby facilitating decision-making, resource optimization, and overall disaster management efforts.

4.2.8 Database

We are leveraging Neo4j, a powerful graph database, to store and display informative, fact-checked tweets in a graphical format. The process begins by applying named entity recognition (NER) to the classified tweets to extract important labels such as disaster name, location, date, and other relevant information. Once these labels are identified, they are mapped to the Neo4j database and displayed as nodes in the graph. Each node encapsulates a specific label, forming a comprehensive representation of the key entities and attributes associated with a tweet. Furthermore, Neo4j enables the representation of relationships between these labeled nodes, enhancing graphical representations with contextual connections. By visually representing the interconnectedness of these labels in graphs, Neo4j provides a comprehensive and intuitive way to explore and analyze the information contained in tweets, unlocking deeper insights and information thereby enabling informed decision-making.

4.2.9 Database Design

4.2.9.1 Nodes

1) Disaster Node: This node tells us the type of disaster.

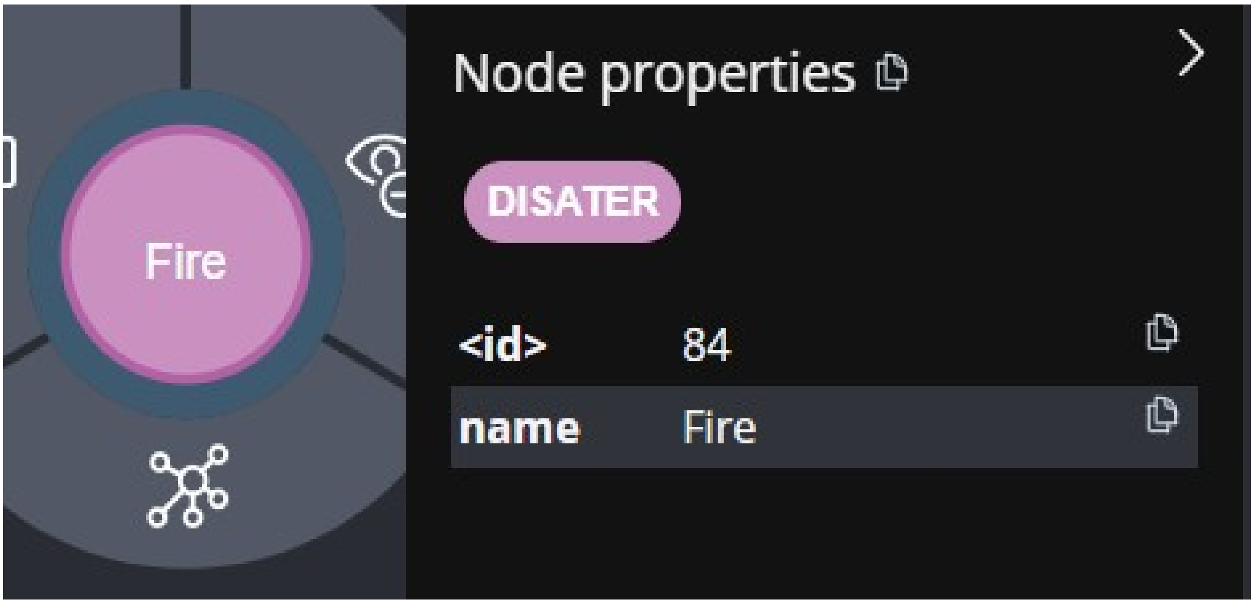

Figure 4.2.9.1.1: Disaster Node

2) Date Node: This node tells us the date when the disaster occurs.

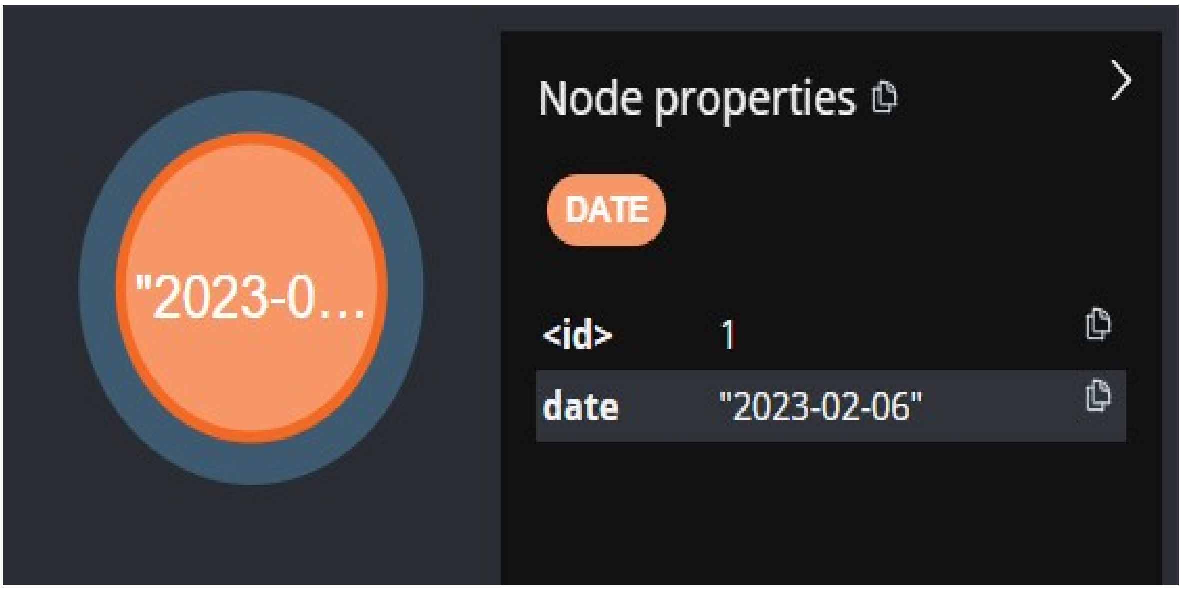

Figure 4.2.9.1.2: Date Node

3) Location Node: This node tells us the location of the disaster.

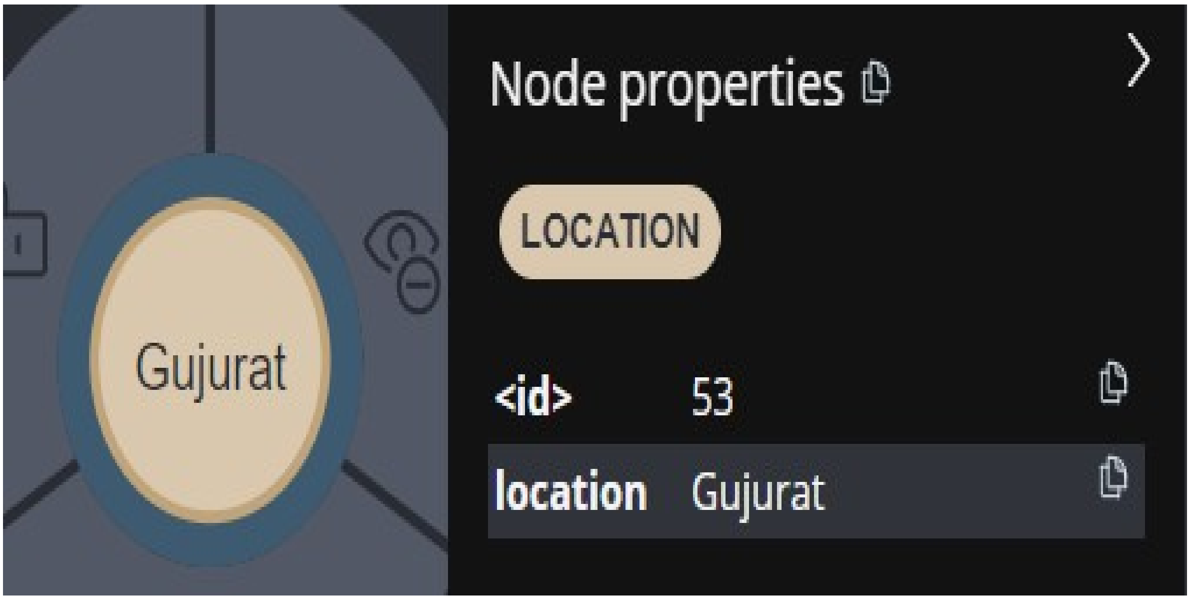

Figure 4.2.9.1.3: Location Node

4) Dead Node: This node tells us the number of people who have died during the disaster.

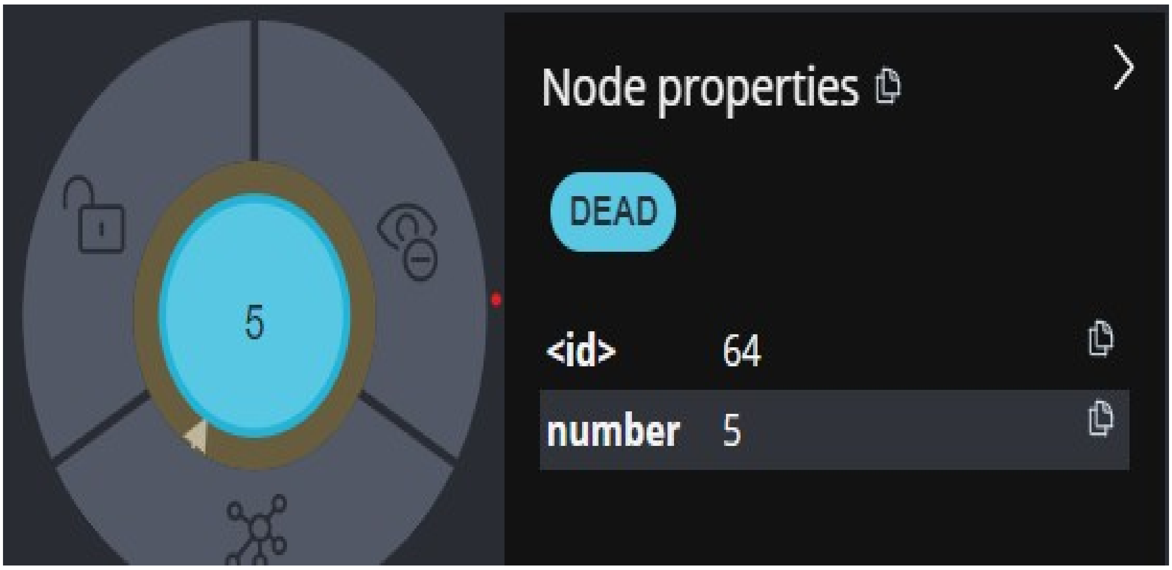

Figure 4.2.9.1.4: Dead Node

5) Injured Node: This node tells us the number of people who have been injured during the disaster.

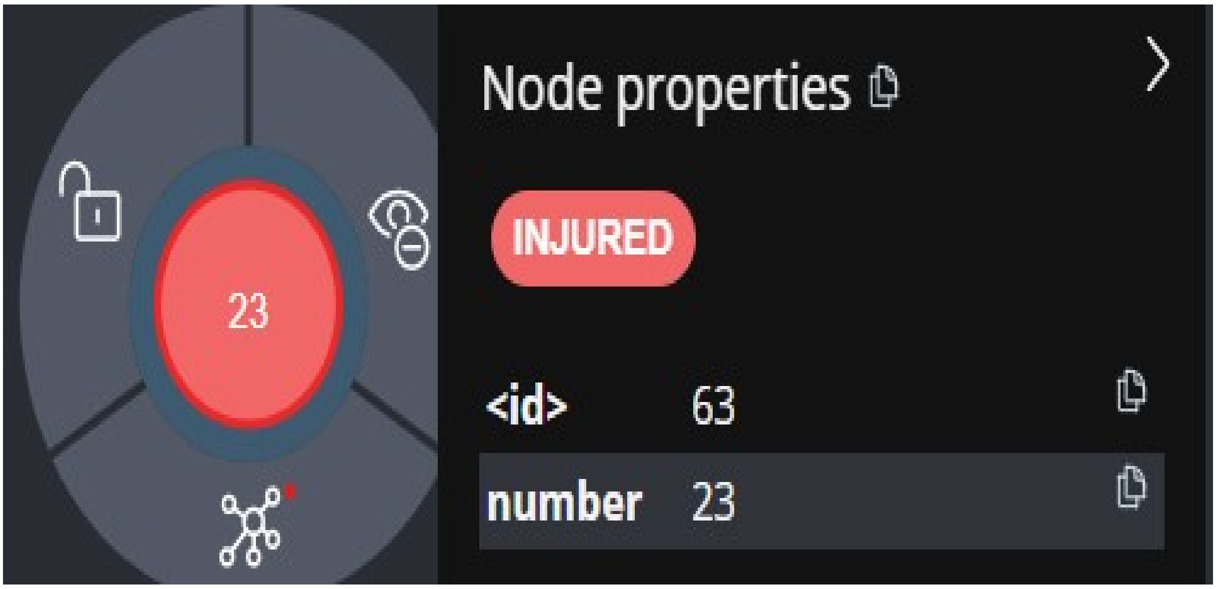

Figure 4.2.9.1.5: Injured Node

4.2.9.2 Relationships

1) AT: Gives a relationship between disaster and location.

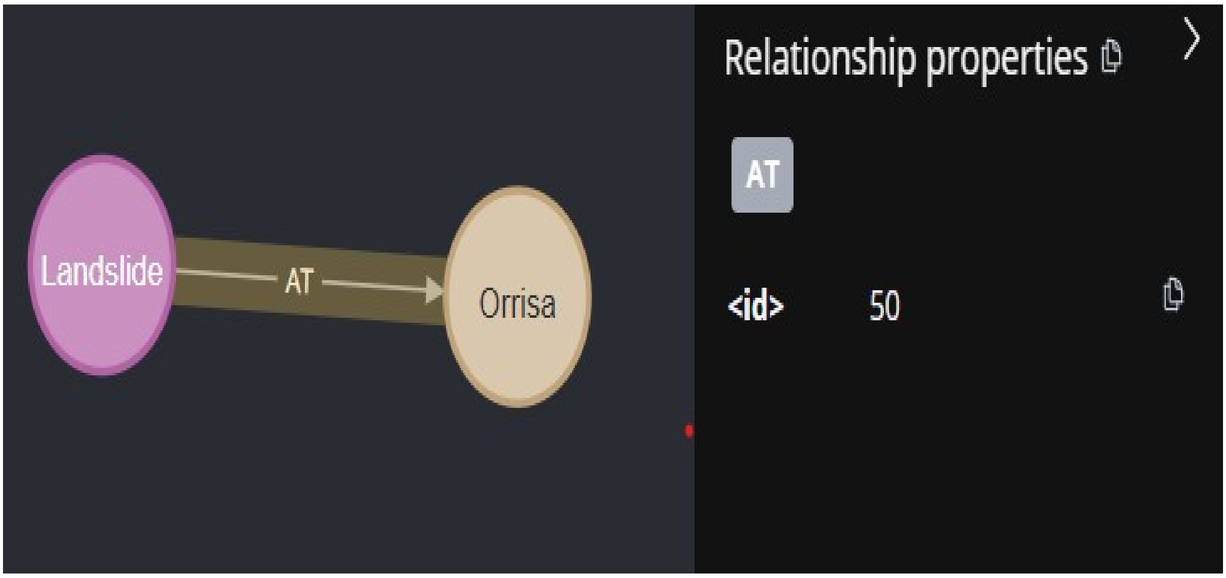

Figure 4.2.9.2.1: AT relationship

2) ON: Gives a relationship between Disaster and Date.

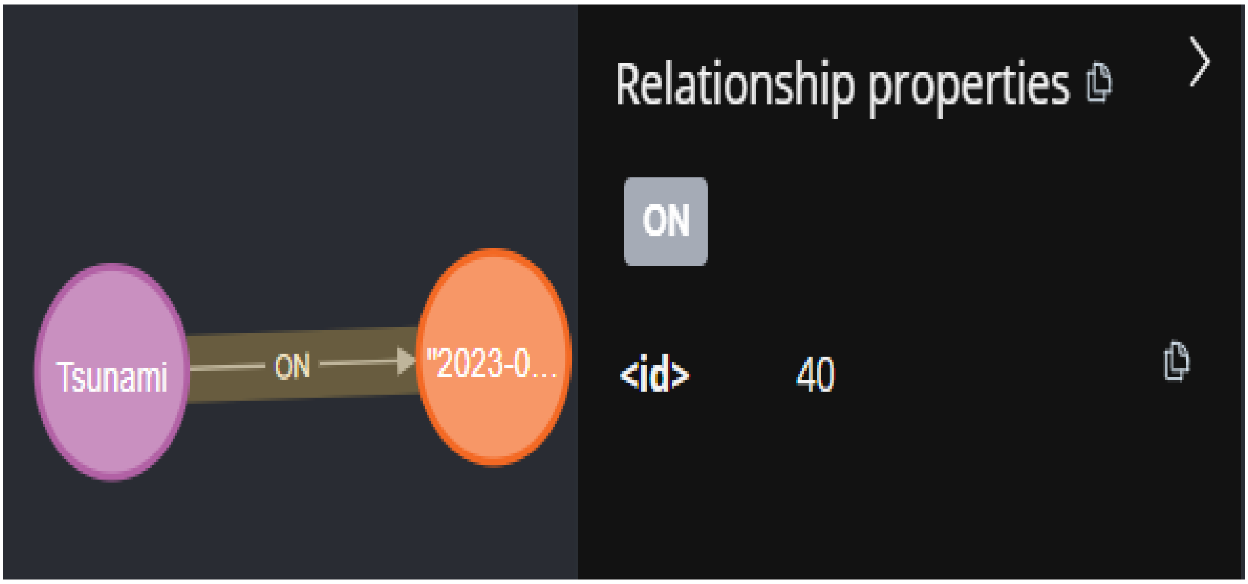

Figure 4.2.9.2.2: ON relationship

3) DEAD: Gives a relationship between Disaster and Dead.

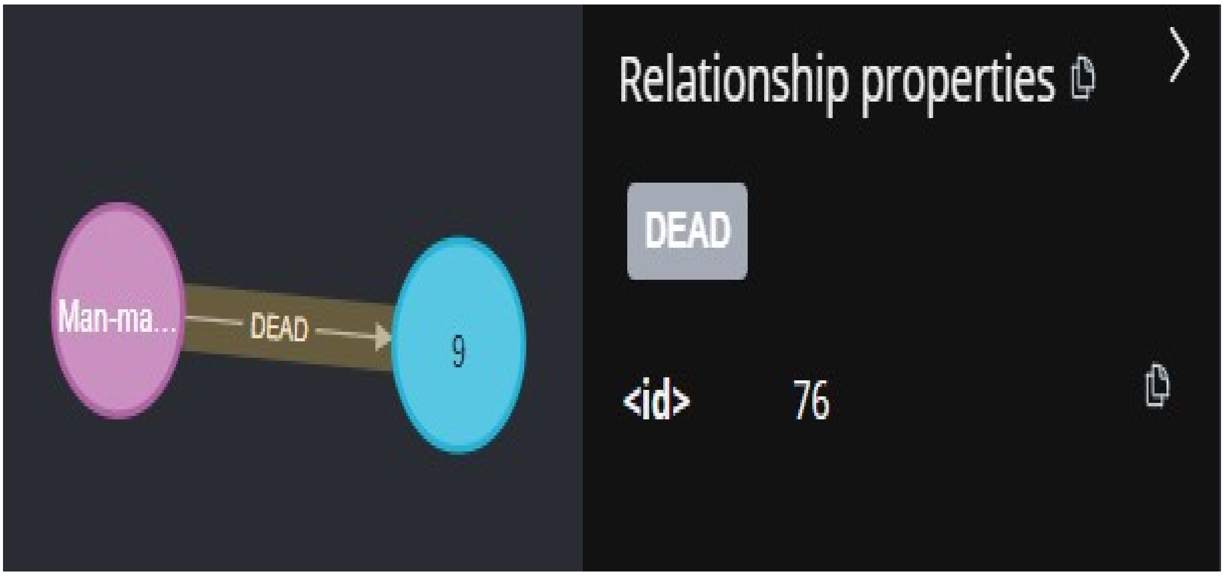

Figure 4.2.9.2.3: DEAD relationship

4) HURT: Gives a relationship between Disaster and Injured.

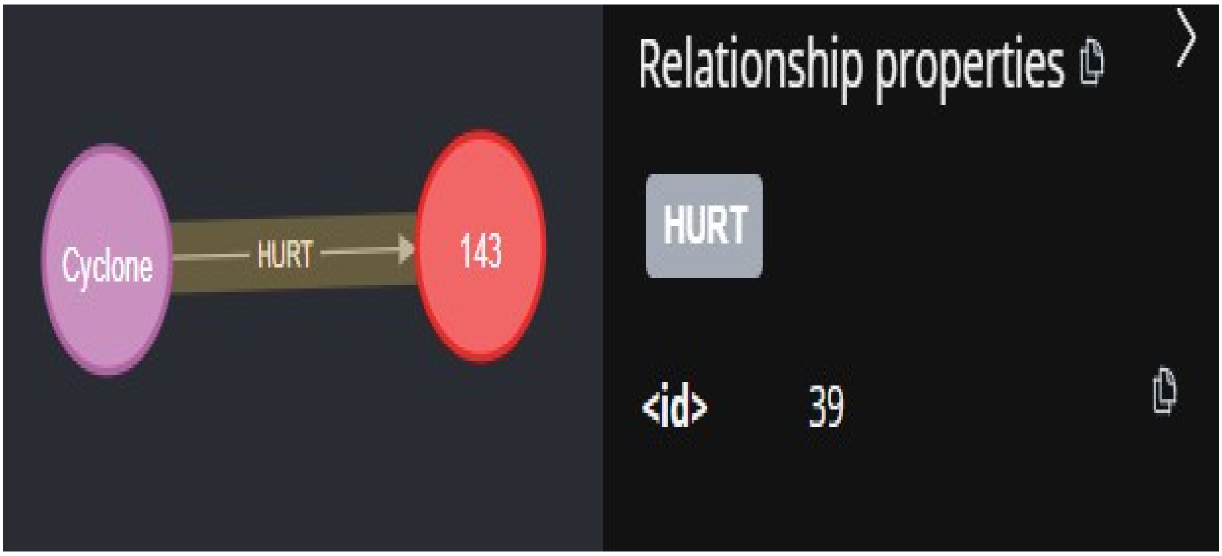

Figure 4.2.9.2.4: HURT relationship

4.2.10 Class Diagram

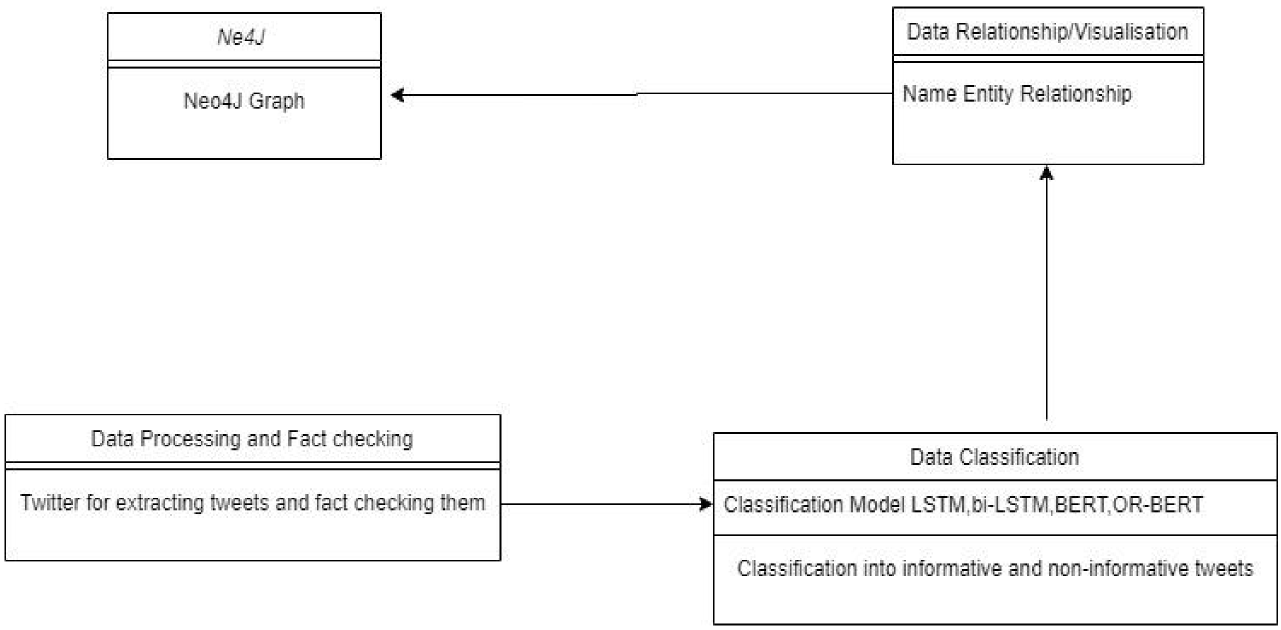

Figure 4.2.10: Class Diagram

4.2.11 Software And Hardware Used

We have made use of the following software to build our project:

- Programming Languages: Python, JavaScript, and relevant frameworks for AI algorithms, web development, and data analysis.

- Database Management: NEO4J or similar graph database system for storing and querying aggregated data.

- API Integration: Twitter API for accessing tweets, Gemini APIs for content verification and data enrichment. Google Translate API for converting all text/tweets to English.

Chapter 5

System Performance Evaluation

To evaluate the proposed system, disaster-related tweets are categorized, and flood-related information is extracted. By comparing disaster tweet classification algorithms to existing systems, model evaluation metrics are utilized. Extraction of flood knowledge requires the visualization of local and global system behavior.

5.1 Comparative results of Accuracy, Precision, Recall, and F1 Score

Sr. No	Algorithm	Accuracy	Precision	Recall	F1 score
1.	LSTM	0.85	0.82	0.87	0.84
2.	Bi-LSTM	0.87	0.85	0.88	0.86
3.	BERT	0.88	0.87	0.87	0.85
4.	RoBERTa	0.89	0.88	0.86	0.87
5.	**DistilBERT**	**0.91**	**0.92**	**0.91**	**0.89**

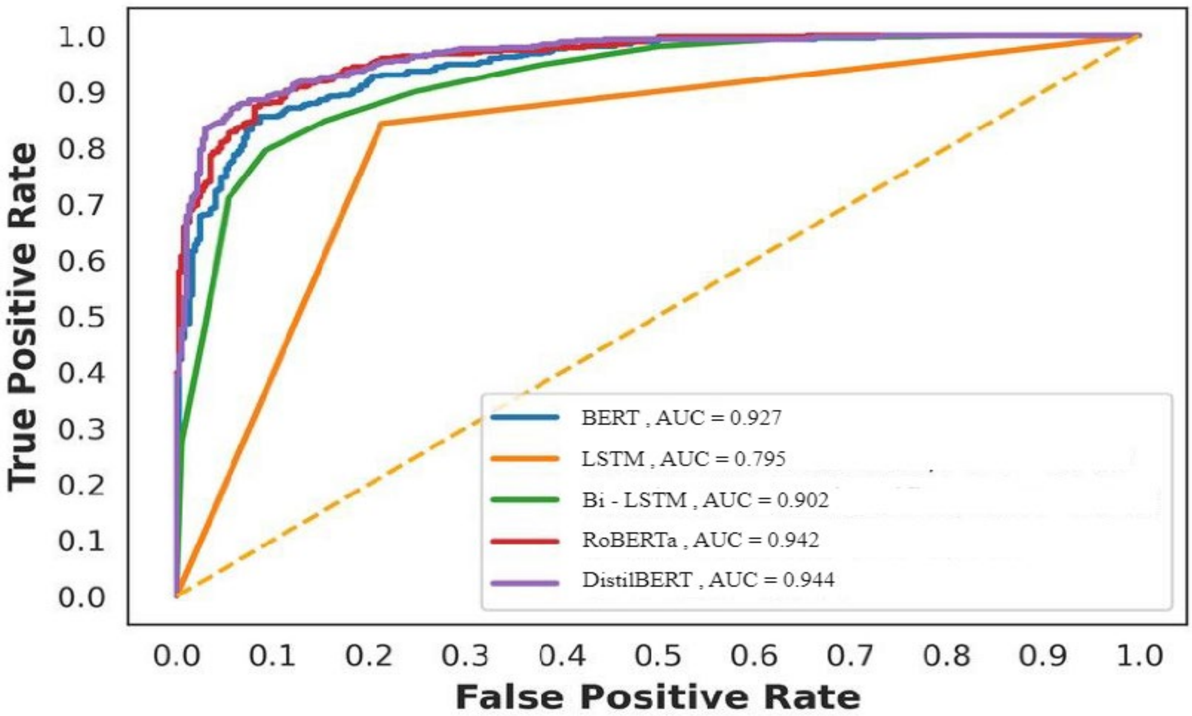

Figure 5.1: ROC / AUC for algorithms used

```
# Compile
classifier.compile(
    loss=keras.losses.SparseCategoricalCrossentropy(from_logits=True), #'binary_crossentropy',
    optimizer=keras.optimizers.Adam(1e-5),
    metrics= ["accuracy"]
)

# Fit
history = classifier.fit(x=X_train,
                         y=y_train,
                         batch_size=BATCH_SIZE,
                         epochs=EPOCHS,
                         validation_data=(X_val, y_val)
                        )
```

```
Epoch 1/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 848s 975ms/step - accuracy: 0.8405 - loss: 0.3439 - val_accuracy: 0.9061 - val_loss: 0.2241
Epoch 2/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 750s 935ms/step - accuracy: 0.9155 - loss: 0.1960 - val_accuracy: 0.9142 - val_loss: 0.2098
Epoch 3/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 771s 962ms/step - accuracy: 0.9262 - loss: 0.1725 - val_accuracy: 0.9140 - val_loss: 0.2116
Epoch 4/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 719s 896ms/step - accuracy: 0.9349 - loss: 0.1535 - val_accuracy: 0.9129 - val_loss: 0.2204
Epoch 5/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 720s 897ms/step - accuracy: 0.9486 - loss: 0.1298 - val_accuracy: 0.9148 - val_loss: 0.2283
Epoch 6/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 745s 929ms/step - accuracy: 0.9574 - loss: 0.1100 - val_accuracy: 0.9144 - val_loss: 0.2446
Epoch 7/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 800s 997ms/step - accuracy: 0.9657 - loss: 0.0908 - val_accuracy: 0.9143 - val_loss: 0.2599
Epoch 8/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 746s 930ms/step - accuracy: 0.9708 - loss: 0.0760 - val_accuracy: 0.9134 - val_loss: 0.2886
Epoch 9/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 746s 930ms/step - accuracy: 0.9772 - loss: 0.0621 - val_accuracy: 0.9119 - val_loss: 0.3254
Epoch 10/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 721s 899ms/step - accuracy: 0.9819 - loss: 0.0516 - val_accuracy: 0.9128 - val_loss: 0.3494
Epoch 11/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 764s 953ms/step - accuracy: 0.9832 - loss: 0.0444 - val_accuracy: 0.9126 - val_loss: 0.3556
Epoch 12/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 781s 974ms/step - accuracy: 0.9864 - loss: 0.0382 - val_accuracy: 0.9130 - val_loss: 0.3696
Epoch 13/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 745s 929ms/step - accuracy: 0.9890 - loss: 0.0312 - val_accuracy: 0.9118 - val_loss: 0.3899
Epoch 14/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 800s 997ms/step - accuracy: 0.9912 - loss: 0.0252 - val_accuracy: 0.9082 - val_loss: 0.4343
Epoch 15/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 744s 927ms/step - accuracy: 0.9906 - loss: 0.0251 - val_accuracy: 0.9124 - val_loss: 0.4614
Epoch 16/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 802s 1000ms/step - accuracy: 0.9922 - loss: 0.0212 - val_accuracy: 0.9128 - val_loss: 0.4477
Epoch 17/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 720s 898ms/step - accuracy: 0.9928 - loss: 0.0198 - val_accuracy: 0.9129 - val_loss: 0.4748
Epoch 18/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 746s 930ms/step - accuracy: 0.9943 - loss: 0.0164 - val_accuracy: 0.9136 - val_loss: 0.4580
Epoch 19/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 800s 997ms/step - accuracy: 0.9941 - loss: 0.0152 - val_accuracy: 0.9096 - val_loss: 0.4860
Epoch 20/20
802/802 ━━━━━━━━━━━━━━━━━━━━ 719s 896ms/step - accuracy: 0.9952 - loss: 0.0135 - val_accuracy: 0.9124 - val_loss: 0.4779
```

Figure 5.2: DistilBERT Training Accuracy

```
# Evaluate the model on the test set
_, test_accuracy = classifier.evaluate(X_val, y_val)

# Generate predictions on the test set
test_predictions = classifier.predict(X_test)
```

```
433/433 ━━━━━━━━━━━━━━━━━━━━ 60s 133ms/step - accuracy: 0.9154 - loss: 0.4703
104/104 ━━━━━━━━━━━━━━━━━━━━ 17s 142ms/step
```

Figure 5.3: DistilBERT Testing Accuracy

We received a testing accuracy of 91.54%

Chapter 6

Implementation Details

6.1 Fact-Checking

The output generated by the fact-checking module serves as a comprehensive report, meticulously evaluating the veracity of statements or claims provided by users. This evaluation process is facilitated through seamless integration with the Gemini model, which employs advanced algorithms to assess the accuracy of information against a multitude of reliable sources, including Twitter. By leveraging this integration, the module delivers insightful assessments of the information's truth value, categorizing statements as either true, false, or uncertain, based on rigorous evaluation criteria.

Furthermore, the output of the fact-checking module goes beyond simple categorization by providing users with detailed supporting evidence or sources to substantiate the determination made. Through this meticulous approach, users gain access to transparent and accountable assessments of the information they encounter online. By providing users with trustworthy information grounded in reliable sources, the output of the fact-checking module contributes to fostering a more credible and reliable digital environment, where misinformation is effectively countered with evidence-based insights.

By promoting transparency, accountability, and access to trustworthy information, this output serves as a cornerstone in the ongoing efforts to combat misinformation and cultivate a culture of informed discourse and critical thinking online. Through its meticulous evaluation process and provision of supporting evidence, the fact-checking module output stands as a beacon of reliability, guiding users toward a more credible and trustworthy online experience.

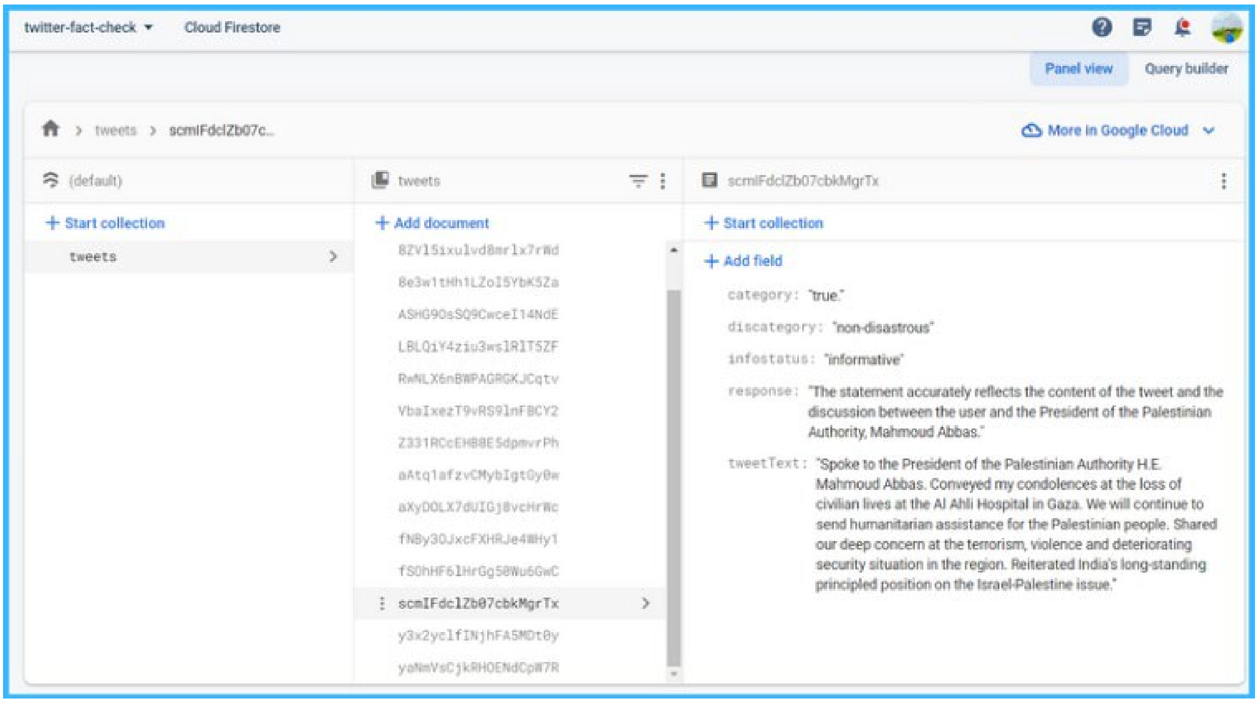

Figure 6.1.1: Firebase to Store Tweets

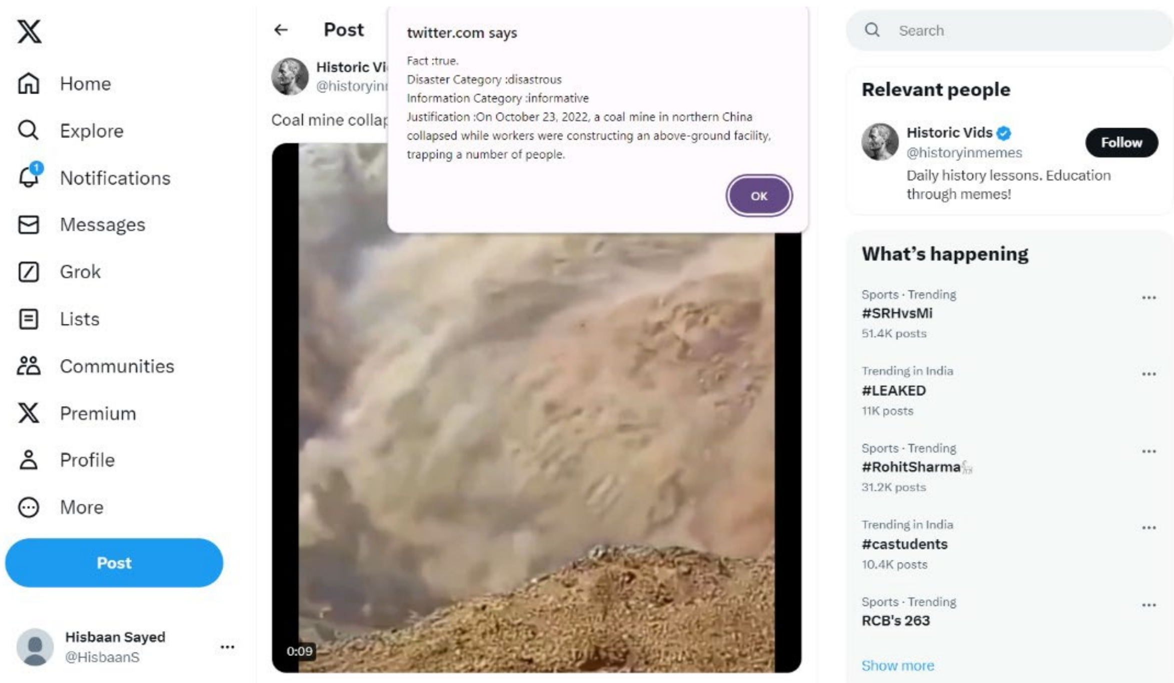

Figure 6.1.2: Fact Checking of Tweet

	A	B	C	D	E
1	infostatus	discategory	tweetText	response	category
2	non-informative	non-disastrous	Our PCT boys are real	This tweet assumes that	false:
3	informative	non-disastrous	Did Hamas Kill &	The conclusion is based c	FALSE
4	non-informative	non-disastrous	I think that most Arab	The statement makes ger	false.
5	informative	non-disastrous	Attention	The tweet states that the	FALSE
6	informative	disastrous	ANALYSIS ON THE	The tweet provides a logi	true.
7	informative	non-disastrous	EXPERT ANALYSIS:	As an AI fact-checking as	false:
8	informative	non-disastrous	Deeply shocked at the	This tweet expresses con	TRUE
9	informative	non-disastrous	BREAKING: massive	There is no evidence of a	false:
10	informative	disastrous	Israel just bombed a	There is no current credit	false.
11	informative	non-disastrous	BREAKING - US Nation	There is no evidence that	false.
12	non-informative	non-disastrous	Palestine	The tweet contains a per	false.
13	informative	non-disastrous	Spoke to the President	The statement accurately	true.
14	informative	disastrous	No way home. I've bee	While it is true that there	false.
15	non-informative,	non-disastrous	Who asked Pakistani	The tweet makes claims	FALSE

Figure 6.1.3: CSV File of Fact-Checking

6.2 Data Preprocessing and Cleaning

The F1-Score, a ubiquitous metric in text categorization, serves as the primary measure for comparing models. Combining precision and recall, this metric offers a comprehensive assessment of classification performance, particularly beneficial in information retrieval tasks where accurately identifying incorrectly classified cases is paramount. In our experiments, we employ CrisisMMD, a dataset renowned for its substantial size and diverse set of class labels, ranging from two to eleven categories. To ensure consistency across evaluations, classification experiments are conducted on the train, development, and test datasets. By leveraging CrisisMMD's varied label set and sizable corpus, our assessments provide robust insights into model performance across different categorization tasks.

Moreover, our experiments delve into the effectiveness of smaller datasets by training models on them and subsequently evaluating their performance on the combined test set. This approach allows us to discern the impact of dataset size on model generalization and classification accuracy. By systematically comparing results across datasets of varying sizes, our study sheds light on the efficacy of models in handling different data volumes and underscores the importance of dataset size in training and evaluating text categorization systems.

res

rname	Location	co-ordinates	Follower Count	Friends count	Protected	Verified	date	tweet	reply count	like count	retweet count	Unnamed: 12	Unnamed: 13	clean_tweet	target
endra_	Varanasi, India	NaN	39	179	FALSE	FALSE	2019-09-30 16:51:49+00:00	#BIHARfloods I request to all, pls don't take ...	0	0	0	NaN	not_humanitarian	BIHARfloods I request to all, pls don't take j...	1
endra_	Varanasi, India	NaN	39	179	FALSE	FALSE	2019-09-30 16:18:32+00:00	I request to all, pls don't take jibe at #Biha...	0	0	0	NaN	not_humanitarian	I request to all, pls don't take jibe at Bihar...	1
Azhars	Warangal	NaN	8155	243	FALSE	FALSE	2019-09-15 05:04:03+00:00	SIO volunteers continues its works in flood ef...	0	20	2	NaN	donation_and_volunteering	SIO volunteers continues its works in flood ef...	1
nohan	Moranhat, India	NaN	42	49	FALSE	FALSE	2019-09-09 07:41:19+00:00	Happy birthday padman @akshaykumar\n🎂 😍 You are...	0	1	0	NaN	not_humanitarian	Happy birthday padman You are God for assames...	1
tweets	Bengaluru, India	NaN	196	196	FALSE	FALSE	2019-08-12 12:29:36+00:00	Can someone please provide one-one POC's from ...	2	1	0	NaN	donation_and_volunteering	Can someone please provide one-one POC's from ...	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ajitnath	Bangalore (India)	NaN	26	94	FALSE	FALSE	2017-04-03 04:43:02+00:00	#HeavyRain from last one week in Barak Valley ...	0	0	0	NaN	not_humanitarian	HeavyRain from last one week in Barak Valley o...	1
aLintu	NaN	NaN	6	44	FALSE	FALSE	2017-04-02 10:07:18+00:00	This is raing time.Assam flood affected area.p...	0	0	0	NaN	not_humanitarian	This is raing time.Assam flood affected area.p...	1

Figure 6.2: Refined Tweets after Preprocessing

6.3 Named-Entity-Relationship

Tweet: Tuesday at one o'clock, there was a flood in Mumbai, and 1000 people were affected.

Word	POS	Tag	Label
Tuesday	PROPN	NNP	DATE
at	ADP	IN	OTHER
one	NUM	CD	TIME
o'clock	NOUN	NN	TIME
,	PUNCT	,	OTHER
there	PRON	EX	OTHER
was	VERB	VBD	OTHER
a	DET	DT	OTHER
flood	NOUN	NN	NATURAL_DISASTER
in	ADP	IN	OTHER
Mumbai	PROPN	NNP	PLACE_AFFECTED
,	PUNCT	,	OTHER
and	CCONJ	CC	OTHER
1000	NUM	CD	NUM_PEOPLE_AFFECTED
	SPACE	_SP	OTHER
people	NOUN	NNS	OTHER
were	AUX	VBD	OTHER
affected	VERB	VBN	OTHER
.	PUNCT	.	OTHER

Figure 6.3: Labelled Sentences

6.4 Knowledge graph

Imagine tapping into real-time insights from the continuous flow of tweets—an exemplary system showcasing this capability delineates the entire process from tweet acquisition and classification to data extraction and visualization via an extensive Knowledge Graph. Tailored for seamless operation in real-time settings, this system highlights its efficacy by simulating the process with 500 preloaded tweets. By leveraging Spacy's Named Entity Recognition (NER), the tweets undergo categorization based on informativeness, and crucial entities are extracted. Through this simulated scenario, the system effectively demonstrates its adaptability and efficiency in managing tweet volumes spanning from 15 to 500, smoothly expanding the Knowledge Graph in line with the escalating data influx. This scalability serves as a testament to its preparedness for handling the dynamic nature of real-world data streams, presenting a promising avenue for unearthing valuable insights within the expansive landscape of social media, tweet by tweet.

In envisioning the application of real-time insights from tweets, the system elucidates a journey encompassing tweet retrieval, classification, information extraction, and visualization within a comprehensive Knowledge Graph. Initially designed for live deployment, the system's capabilities are vividly illustrated through a simulation featuring 500 pre-loaded tweets. By harnessing Spacy's NER capabilities, the system efficiently categorizes tweet informativeness and extracts pertinent entities, showcasing its adaptability and scalability. As the tweet volume escalates from 15 to 500, the Knowledge Graph seamlessly expands, affirming its readiness for real-world scenarios and its potential to uncover invaluable insights from the dynamic realm of social media, tweet by tweet.

This iterative process of tweet acquisition, classification, and data extraction culminates in the visualization of insights within a comprehensive Knowledge Graph. While originally conceived for live deployment, the system's capabilities are exemplified through a simulation employing 500 pre-loaded tweets. Utilizing Spacy's NER functionality, the system adeptly categorizes tweets based on their informativeness and extracts essential entities, thereby demonstrating its flexibility and scalability. As the tweet volume escalates, from a mere 15 to a substantial 500, the Knowledge Graph expands seamlessly, affirming its adaptability to real-world scenarios and its potential to unveil valuable insights from the dynamic expanse of social media, tweet by tweet.

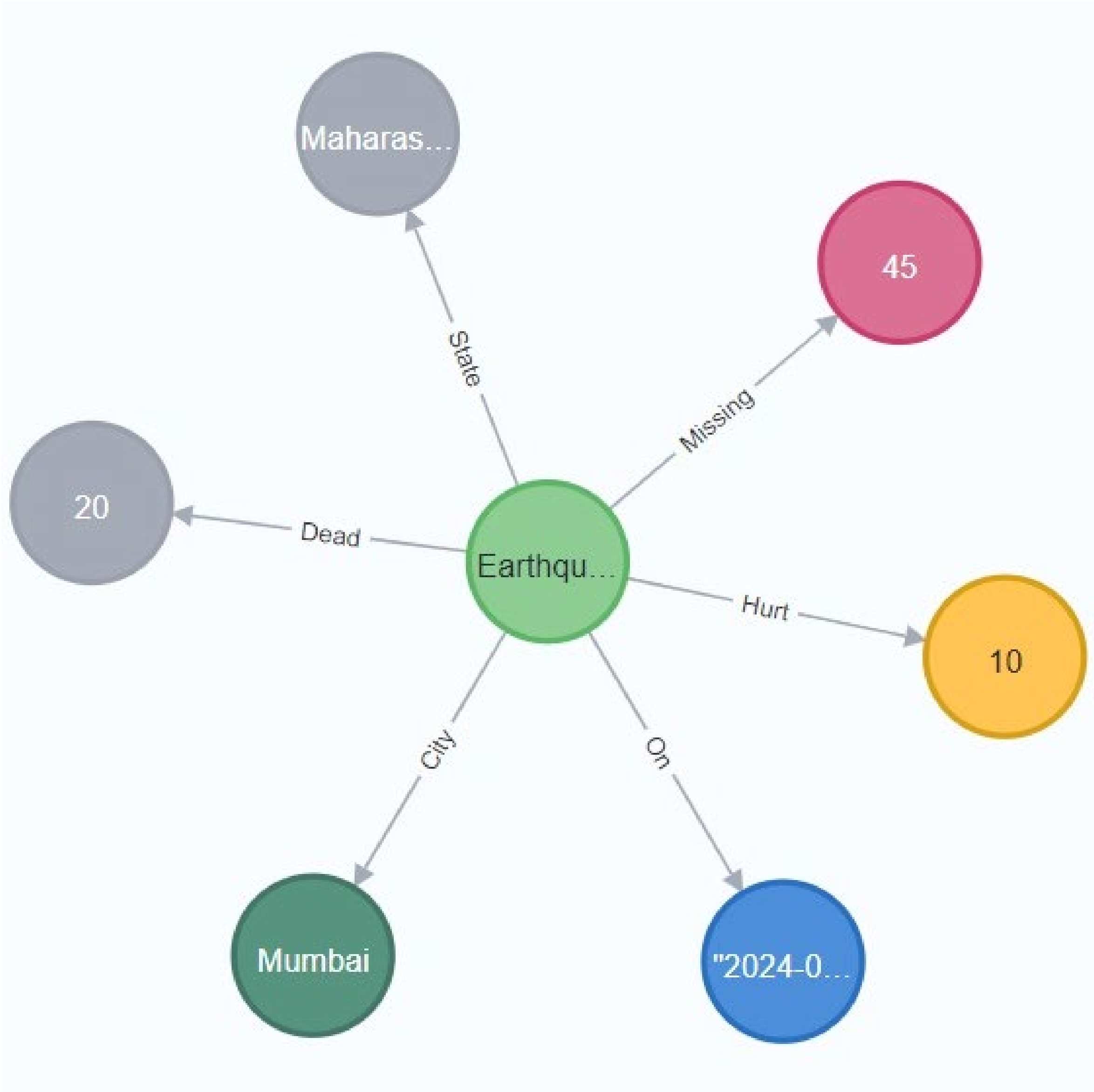

Figure 6.4.1: Knowledge graph after 15 tweets

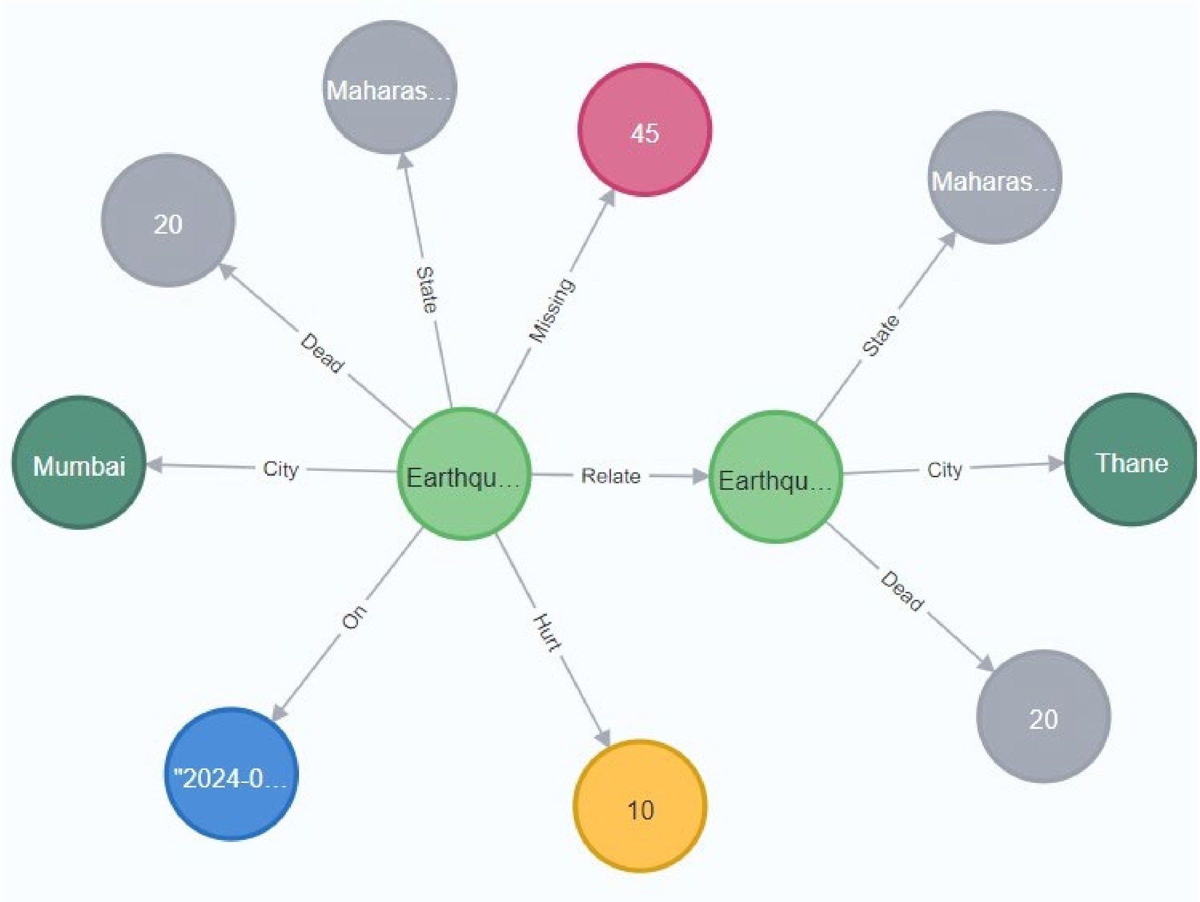

Figure 6.4.2: Knowledge graph after 50 tweets

The system kicks off with just 15 processed tweets, generating a basic Knowledge Graph (Figure 4.3). This graph features a central node representing the initial earthquake event and five connected nodes holding various attributes. Figure 4.4 reveals what these neighbors signify: state, city, date, injured, and missing people. This extracted information, thanks to a custom Spacy NER model, equips decision-makers with immediate awareness of the disaster, its location, date, and initial impact.

As more tweets roll in, existing nodes in the graph dynamically update based on new information. Figure 4.4 shows the impact of 50 tweets: the earthquake node sees its casualty numbers rise, and a new "deaths" node even emerges. Meanwhile, a completely separate disaster, also starting on the same date, pops up on the graph with its own location and impact details. This seamless evolution from 15 to 50 tweets showcases the system's impressive ability to handle live data feeds, continuously refining its picture of unfolding events in real time. Each tweet serves as a tiny brushstroke, contributing to a bigger, clearer understanding of the disaster landscape.

Each new disaster will have its own node and information about the disaster like location, injured, dead, etc. of each disaster will have their own node in the database.

The database will keep on updating as and when new tweets about different disasters arrive. With the help of the dashboard, we can perform various queries. For instance, disasters occurring at a particular location (state) for a particular time frame.

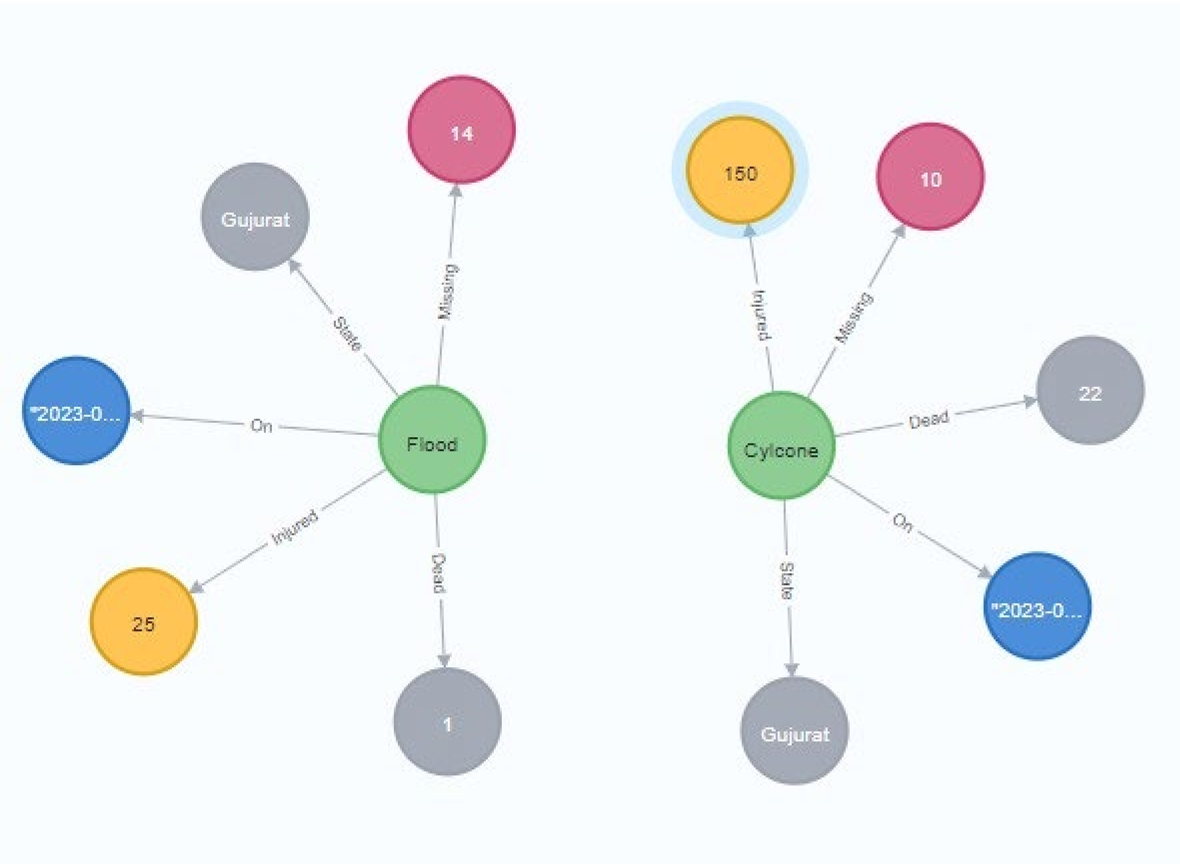

Figure 6.4.3: Disasters in Gujarat in 2023

Considering another state, Kerala, on 18th October 2023, a flash flood occurred there.

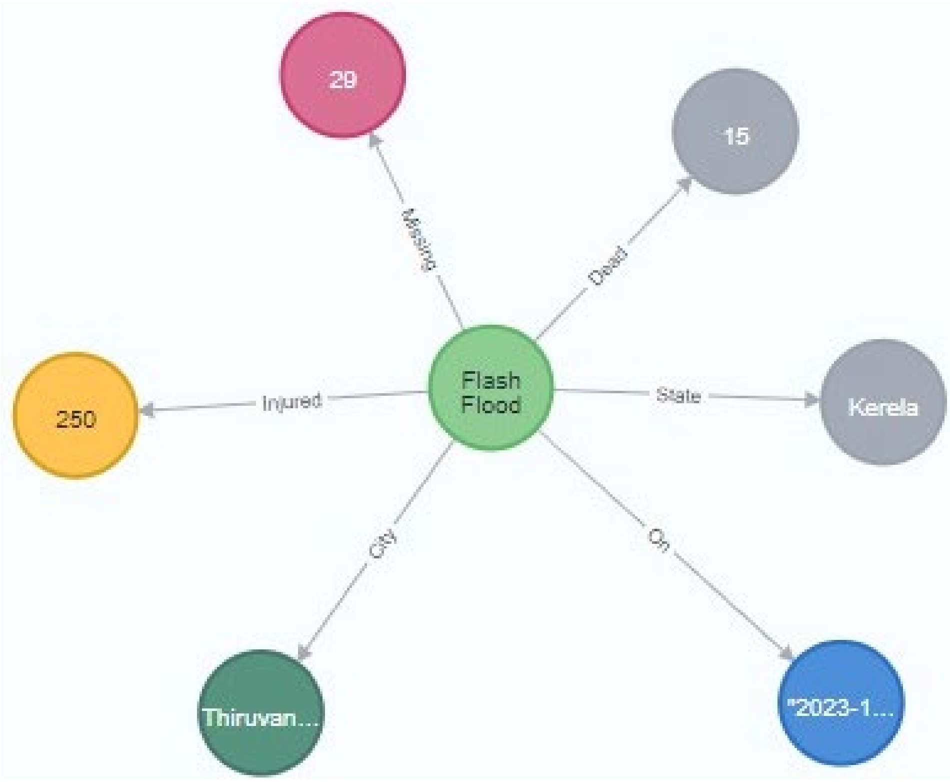

Figure 6.4.4: Disaster in Kerala on 18th October 2023

Furthermore, we can find data about the different disasters that have occurred in a particular state for a given timeline. For instance, the number of disasters occurring in Maharashtra for the year 2023.

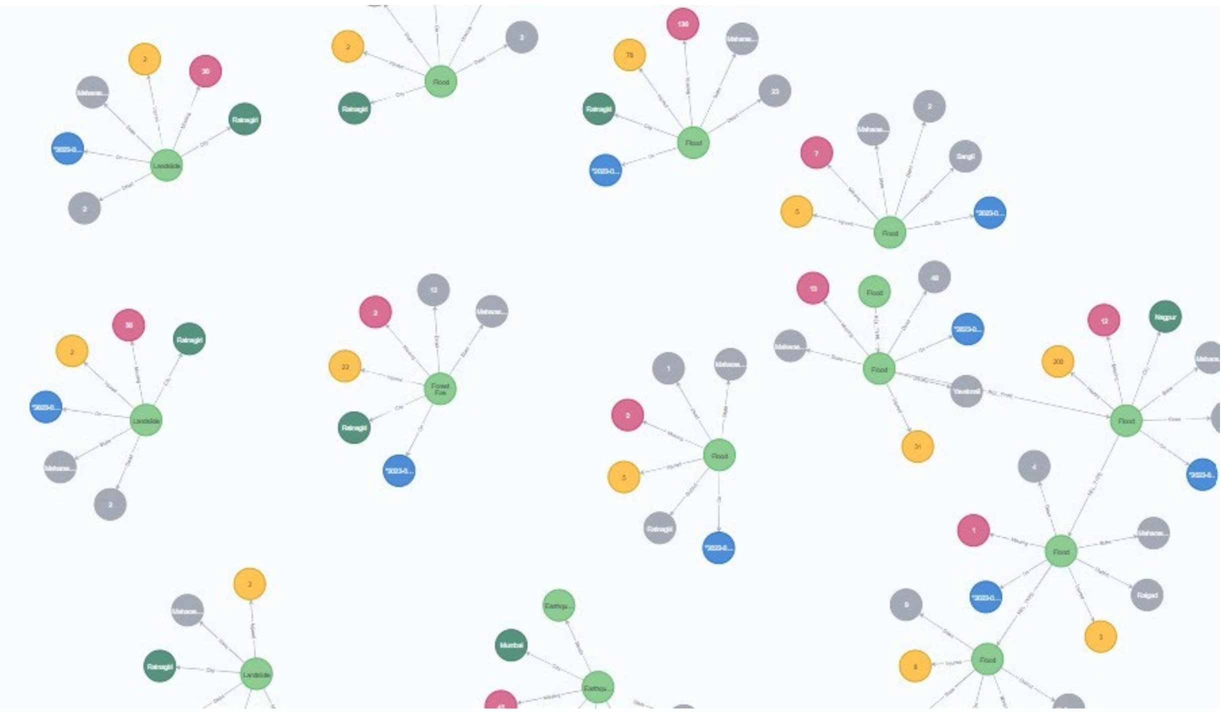
Figure 6.4.5: Disaster in Kerala on 18th October 2023

Using the dashboard, we can look for disaster information ranging from a few years. For instance, the horrible floods in Kashmir in 2014 and Uttarakhand in 2013.

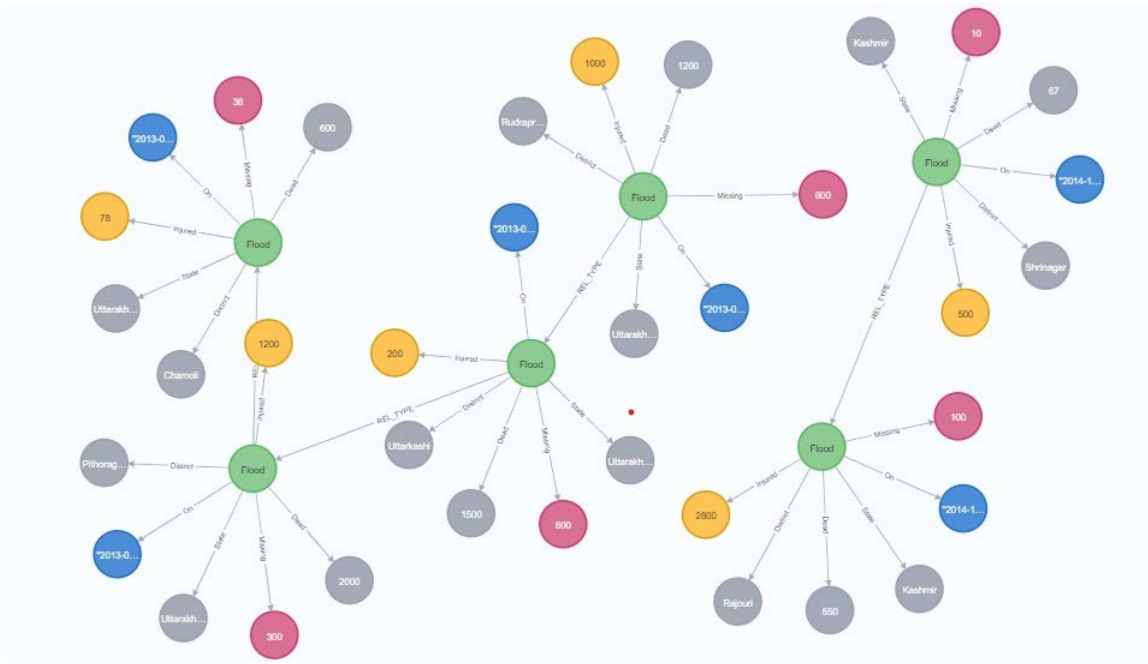

Figure 6.4.6: Disasters in Kashmir and Uttarakhand from 2004-2024

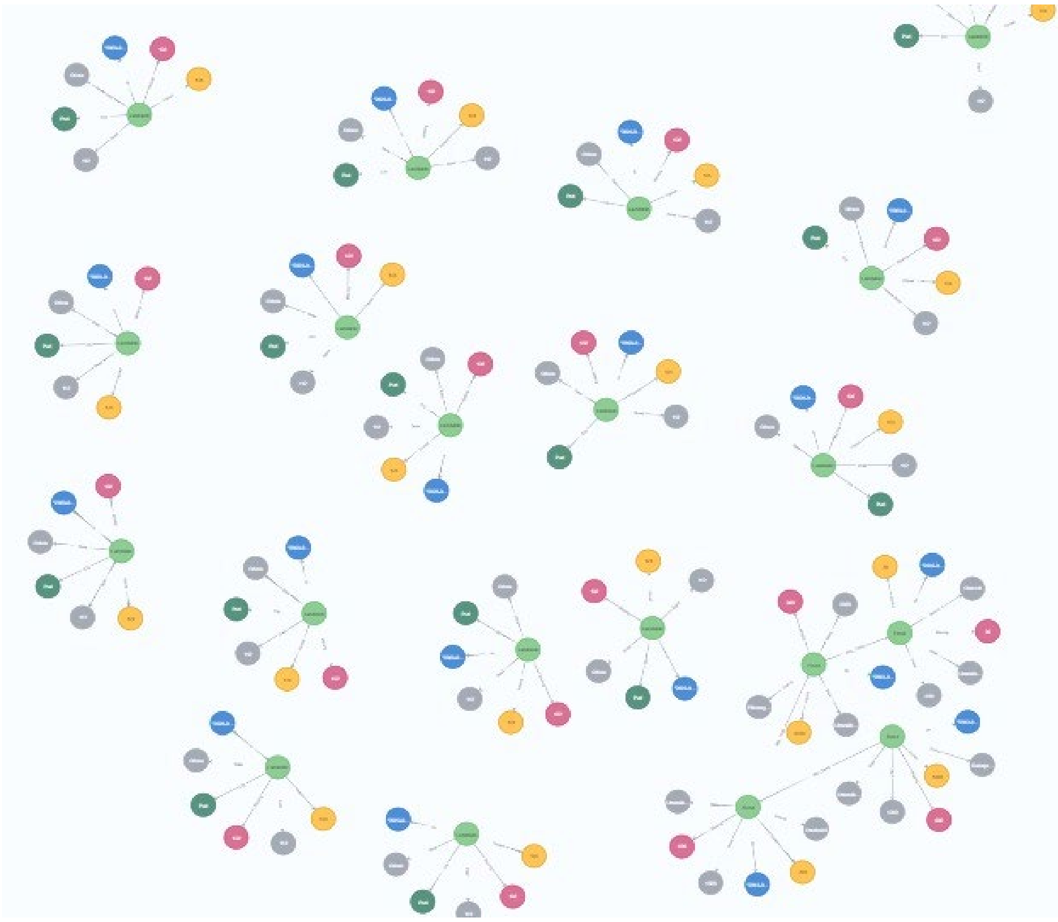

Figure 6.4.7: Overview of major disasters in India from 2004-2024

Similarly, on a National Scale, we can run the same query, to get all the disasters of India for the year 2024.

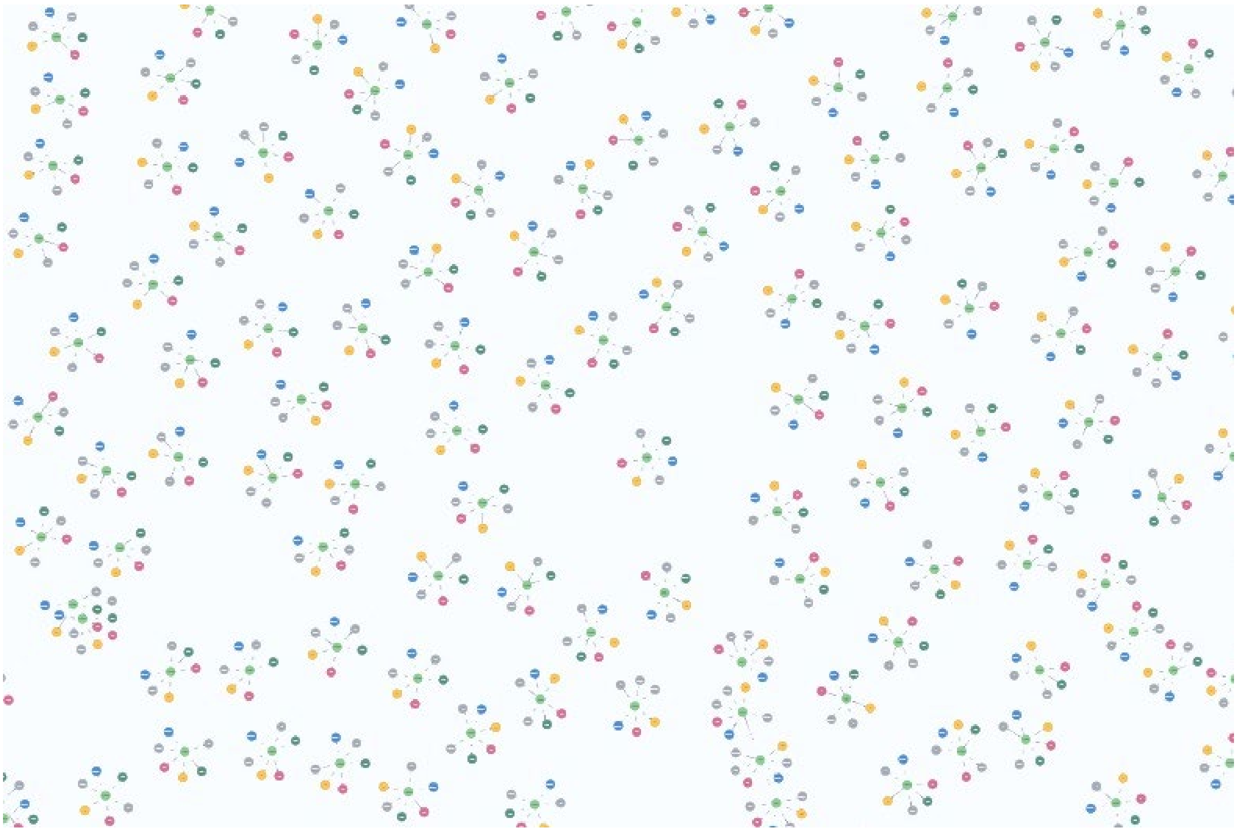

Figure 6.4.8: Disasters in India in 2023

Chapter 7

Conclusion And Future Enhancements

7.0.1 Result Analysis

- ResilientNet, a novel disaster management solution for India, demonstrated significant potential in merging real-time big data processing and crowdsourced fact-checking to enhance accurate disaster assessment, timely response, and informed decision-making.

- ResilientNet effectively pinpointed disaster locations and assessed severity by aggregating and analyzing diverse multimedia content from Twitter, powered by GPT3.5 Fact-checking and the NEO4J database.
- The user-friendly dashboard presented actionable insights, enabling authorities to promptly counter misinformation, understand real-time trends, and efficiently direct disaster response efforts.

- Research on the efficacy of crowdsourced fact-checking for disaster tweets informed a novel disaster-focused tweet verification system. This innovative amalgamation of crowdsourcing and AI technologies fact-checked disaster-related tweets, creating a comprehensive graph of critical metrics and trends.

7.0.2 Conclusion

The impact of disasters on our society is substantial, particularly in terms of the number of casualties. Disaster management focuses on the prevention and mitigation of the risks associated with catastrophic events that threaten humanity. Despite its relative newness, social media provides a way to utilize individuals as sensors to rapidly detect such hazards. The proposed system improves existing tools by incorporating a dynamic approach to extract and present geographical and temporal data. In addition, it grants immediate access to this information to decision-makers. When immediate decision-making is required, even the most minute additional information can exert a substantial influence. Our system's ability to extract and establish correlations among pertinent data enables it to exert a significant influence. The knowledge graph's ability to grow, incorporate fresh information, and revise pre-existing data contributes to its utility as a dependable and up-to-date repository of knowledge. This facilitates faster and more informed decision-making, particularly in time-sensitive situations. Incorporating a mechanism for screening trustworthy information should be considered for inclusion in future updates of the proposed system, alongside the extensive number of hazard detections. To enhance the extracted information of the system, it is necessary to enhance other existing components such as classification models and NER entities. Ultimately, it is proven that as a knowledge network expands, it becomes impractical for human use. Additionally, the system necessitates an automated method of notifying decision-makers, including both the general public and emergency personnel. Digital assistants are highly likely to be chosen for this job in the near future. The proposed model's contribution can be highly beneficial in providing a comprehensive summary of information in flood scenarios. With minor modifications tailored to specific scenarios, it can be utilized as a practical instrument in potential imminent crises.

7.0.3 Future Enhancements

Further research and development are warranted to refine ResilientNet's capabilities, enhance its scalability, and explore its potential integration with existing disaster management systems. Additionally, pilots with relevant stakeholders could beconducted to assess the feasibility and impact of ResilientNet in real-world settings.

References

[1] Fahim K.Sufi (2022). AI-SocialDisaster: An AI-based software for identifying and analyzing natural disasters from social media. https://doi.org/10.1016/j.simpa.2022.100319

[2] Khajwal, A., Cheng, C., & Noshadravan, A. (2022). Multi-view deep learning for reliable post-disaster damage classification. ArXiv, abs/2208.03419. https://doi.org/10.48550/arXiv.2208.03419.

[3] J. Son, C.-S. Lim, H.-S. Shim, and J.-S. Kang, "Development of Knowledge Graph for Data Management Related to Flooding Disasters Using Open Data," Future Internet, vol. 13, no. 5, p. 124, May 2021, doi: 10.3390/fi13050124.

[4] Sreenivasulu, M., Muthukumarasamy, S., & Jayadev, P. (2021). Multi-modal classification of Twitter data during disasters for humanitarian response. Journal of Ambient Intelligence and Humanized Computing, 12, 10223 - 10237. https://doi.org/10.1007/s12652-020-02791-5.

[5] Sreenivasulu, M., & Sridevi, M. (2021). A novel method for identifying the damage assessment tweets during disaster. Future Gener. Comput. Syst., 116, 440-454. https://doi.org/10.1016/j.future.2020.10.037.

[6] Mohanty, S., Biggers, B., SayedAhmed, S., Pourebrahim, N., Goldstein, E., Bunch, R., Chi, G., Sadri, F., McCoy, T., & Cosby, A. (2021). A multi-modal approach towards mining social media data during natural disasters - a case study of Hurricane Irma. International journal of disaster risk reduction : IJDRR, 54. https://doi.org/10.1016/j.ijdrr.2020.102032.

[7] Yunhao Zhang, Jun Zhu, Qing Zhu, Yakun Xie, Weilian Li, Lin Fu, Junxiao Zhang & Jianmei Tan (2020) The construction of personalized virtual landslide disaster environments based on knowledge graphs and deep neural networks, International Journal of Digital Earth, 13:12, 1637-1655, DOI: 10.1080/17538947.2020.1773950

[8] Kankanamge, N., Yigitcanlar, T., Goonetilleke, A., & Kamruzzaman, M. (2020). Determining disaster severity through social media analysis: Testing the methodology with South East Queensland Flood tweets. International journal of disaster risk reduction, 42, 101360. https://doi.org/10.1016/j.ijdrr.2019.101360.

[9] J. Boné, M. Dias, J. C. Ferreira, and R. Ribeiro, "DisKnow: A Social-Driven Disaster Support Knowledge Extraction System," Applied Sciences, vol. 10, no. 17, p. 6083, Sep. 2020, doi: 10.3390/app10176083.

[10] Priyanka, E.., & Thangavel, S.. (2020). Influence of Internet of Things (IoT) In Association of Data Mining Towards the Development Smart Cities-A Review Analysis. Journal of Engineering Science and Technology Review , 13 , 1-21 . http://doi.org/10.25103/jestr.134.01

[11] Ren, L.., Liu, Yuxin., Wang, Xiaokang., Lu, Jinhu., & Deen, M.. (2021). Cloud–Edge-Based Lightweight Temporal Convolutional Networks for Remaining Useful Life Prediction in IIoT. IEEE Internet of Things Journal , 8 , 12578-12587 . http://doi.org/10.1109/jiot.2020.3008170

[12] Sishodia, R.., Ray, R.., & Singh, Sudhir Kumar. (2020). Applications of Remote Sensing in Precision Agriculture: A Review. Remote. Sens. , 12 , 3136 . http://doi.org/10.3390/rs12193136

[13] Shah, Syed Attique., Seker, D.., Hameed, S.., & Draheim, D.. (2019). The Rising Role of Big Data Analytics and IoT in Disaster Management: Recent Advances, Taxonomy and Prospects. IEEE Access , 7 , 54595-54614 . http://doi.org/10.1109/ACCESS.2019.2913340

[14] Lwin, K.., Sekimoto, Y.., Takeuchi, W.., & Zettsu, K.. (2019). City Geospatial Dashboard: IoT and Big Data Analytics for Geospatial Solutions Provider in Disaster Management. 2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM) , 1-4 . http://doi.org/10.1109/ICT-DM47966.2019.9032921

[15] Muhammad, Khan., Rodrigues, J.., Kozlov, Sergey., Piccialli, F.., & Albuquerque, V.. (2020). Energy-Efficient Monitoring of Fire Scenes for Intelligent Networks. IEEE Network , 34 , 108-115 . http://doi.org/10.1109/MNET.011.1900257

[16] Fang, S.., Xu, Lida., Zhu, Yunqiang., Ahati, Jiaerheng., Pei, Huan., Yan, Jianwu., & Liu, Zhihui. (2014). An Integrated System for Regional Environmental Monitoring and Management Based on Internet of Things. IEEE Transactions on Industrial Informatics , 10 , 1596-1605 . http://doi.org/10.1109/TII.2014.2302638

[17] Ragaveena, S.., Edward, A. Shirly., & Surendran, U.. (2021). Smart controlled environment agriculture methods: a holistic review. Reviews in Environmental Science and Bio/Technology , 20 , 887 - 913 . http://doi.org/10.1007/s11157-021-09591-z

[18] Salh, A.., Audah, L.., Shah, N.., Alhammadi, A.., طربوش), Qazwan Abdullah (غزوان عبد الله محمد., Kim, Y.., Al-Gailani, S. A.., Hamzah, S.., Esmail, B.., & Almohammedi, Akram A.. (2021). A Survey on Deep Learning for Ultra-Reliable and Low-Latency Communications Challenges on 6G Wireless Systems. IEEE Access , 9 , 55098-55131 . http://doi.org/10.1109/ACCESS.2021.3069707

[19] Pathan, A.. (2020). Urgent health challenges for the next decade 2030: World Health Organization. , 113-117 . http://doi.org/10.37881/1.512

[20] Bisht, D.., Singh, Rajesh., Gehlot, A.., Akram, S.., Singh, Aman., Montero, Elisabeth Caro., Priyadarshi, Neeraj., & Twala, Bhekisipho. (2022). Imperative Role of Integrating Digitalization in the Firms Finance: A Technological Perspective. Electronics . http://doi.org/10.3390/electronics11193252

[21] Sun, A.., & Scanlon, B.. (2019). How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. Environmental Research Letters , 14 . http://doi.org/10.1088/1748-9326/ab1b7d

[22] Shah, Syed Attique., Seker, D.., Rathore, M.., Hameed, S.., Yahia, Sadok Ben., & Draheim, D.. (2019). Towards Disaster Resilient Smart Cities: Can Internet of Things and Big Data Analytics Be the Game Changers?. IEEE Access , 7 , 91885-91903 . http://doi.org/10.1109/ACCESS.2019.2928233

PAPER NAME

Final_Project_Report_ResilientNet.pdf

WORD COUNT	CHARACTER COUNT
8986 Words	59762 Characters
PAGE COUNT	FILE SIZE
53 Pages	1.8MB
SUBMISSION DATE	REPORT DATE
May 1, 2024 11:17 AM GMT+5:30	May 1, 2024 11:18 AM GMT+5:30

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 8% Internet database
- 2% Publications database
- Crossref database
- Crossref Posted Content database
- 7% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less then 8 words)