GEO-SEMANTICS ANALYSIS OF ENVIRONMENTAL DISASTERS IN NIGERIA USING NATIONAL PRINT MEDIA DATA FOR DISASTER MANAGEMENT

Benedict Ajanaku
Data Science Nigeria
benedict@datasciencenigeria.ai

Rashidat Sikiru
Data Science Nigeria
rasheedat@datasciencenigeria.ai

Anthony Soronnadi
Data Science Nigeria
anthony@datasciencenigeria.ai

Ife Adebara
Data Science Nigeria
ife@datasciencenigeria.ai

Olubayo Adekanmbi
Data Science Nigeria
olubayo@datasciencenigeria.ai

1 ABSTRACT

In recent years, Nigeria has experienced various natural and environmental disasters, including floods, food insecurity, fire outbreaks, oil spills, and banditry. These events have caused extensive damage, disrupted lives and properties, and displaced many humans, with emergency response efforts often hindered by the lack of accurate and up-to-date disaster location information. In addressing this gap, we investigated the application of geo-semantics analysis on environmental disasters in Nigeria using media data to map locations and to improve emergency response planning. We developed a disaster location-based (DLB) NER model by fine-tuning three Named Entity Recognition (NER) models—spaCy, BERT, and DistilBERT—with an extensive compiled dataset of disaster-related Nigerian news articles. Each model was evaluated using three metrics, with BERT achieving the highest performance: precision of 0.99331, recall of 0.99349, and f1-score of 0.99297, followed by DistilBERT with precision of 0.99236, recall of 0.99297, and f1-score of 0.99240, and spaCy with precision of 0.95, recall of 0.77, and f1-score of 0.85. The model was used to recognize toponyms and extract location details. Using Nominatim, we resolved the toponyms into coordinates and visualized disaster hotspots. These results show that the fine-tuned NER models can be used in providing precise, real-time mapping, and improving situational awareness for focused interventions. Our approach provides a transformative framework for incorporating print media data into emergency response strategies and informing humanitarian assistance efforts more effectively.

2 INTRODUCTION

Disasters often prompt immediate information sharing, with location being a critical aspect Lingad et al. (2013). Timely identification of affected areas is essential for assessing impact and coordinating response efforts. However, in Nigeria, disaster management remains a challenge due to limited resources and response capacity. While the country experiences fewer severe natural disasters compared to other regions Orhewere (2012), improving preparedness and response mechanisms is crucial for mitigating risks.

Artificial Intelligence (AI) has demonstrated significant potential in disaster management, particularly through natural language processing (NLP), machine learning, and predictive analytics Abdul et al. (2024); Olorunsogo et al. (2024). These technologies enhance real-time decision-making, risk assessment, and resource allocation.

This paper focuses on the use of Geo-NLP (using NLP to extract geographical information), with a focus on using information from news articles to help disaster responses and humanitarian efforts Idakwo et al. (2023). The aim is to demonstrate the usefulness of NLP to contribute to the ongoing discussion to improve emergency response capabilities.

3 RELATED WORK

The extraction of location information from text has been widely studied, with approaches evolving from traditional machine learning to deep learning methods. Early research by Li et al. (2007) applied Support Vector Machines (SVM) to Chinese texts, achieving high precision (93.82%), recall (86.69%), and f1-score (90.12%) using linguistic features.

With the rise of social media as a key information source, Dhavase and Bagade (2014) explored location extraction from Twitter, evaluating classifiers like Naïve Bayes, ZeroR, and MultiScheme, with Naïve Bayes achieving the highest accuracy (93.44%). Further advancing social media-based location extraction, Suleman et al. (2023) fine-tuned transformer models (BERT, RoBERTa, DistilBERT) for disaster-related tweets, with f1-scores of 0.6256, 0.6744, and 0.6723, respectively. Lingad et al. (2013) compared NER tools and found Stanford NER most effective for disaster-related location extraction.

Beyond social media, Idakwo et al. (2023) applied NLP and geo-parsing to extract crash locations from news articles using a SpaCy-based model with an f1-score of 93.62%. Chen et al. (2020) proposed a Bidirectional LSTM-CRF model to improve extraction from informal Twitter messages, addressing challenges related to non-standard place names.

Efforts have also been made to enhance NER in underrepresented languages. Sugiartaa and ERa (2021) developed a rule-based NER system for Balinese texts, achieving an f1-score of 0.92, highlighting the effectiveness of rule-based approaches where annotated data is scarce.

Overall, these studies illustrate the progression of location-based NER, showing how NLP advancements enhance geographic information extraction across diverse domains, particularly for disaster response and emergency planning.

4 METHODOLOGY

The proposed framework comprises five (5) stages as shown in figure1: Data Collection, Data Preprocessing, Custom NER Model Training, Geo-parsing, and Geo-Visualization. This methodology was adopted from Idakwo et al. (2023) paper.

Figure 1: Methodological Framework

4.1 Data Acquisition and Preprocessing

This study utilized disaster-related news articles sourced from Tribune Nigeria through a systematic query approach, yielding a total of 3,722 articles covering fire outbreaks, floods, and oil spills. Data extraction was performed using the Beautiful Soup Python package, and the extracted information was systematically organized into three key columns: headline, article link, and article text. The resulting dataset was converted into a comma-separated values (CSV) format for further analysis. To construct a robust dataset for Named Entity Recognition (NER) model development, the corpus was then refined using a RoBERTa-based text classification model, selected for its superior performance compared to BERT and other state-of-the-art models, to filter out non-relevant articles. Subsequent sentence tokenization using the NLTK Python library was performed to segment the corpus into discrete sentences, thereby enhancing the granularity and accuracy of further analysis Li et al. (2007).

For NER, annotation was conducted using the SpaCy NER Annotator Idakwo et al. (2023), as shown in figure 2. The annotation process focused on key geographical entities relevant to disaster-prone areas, including Village, Market, Town/City, Centre, State, LGA, River, Landmark, Road, and Community. This tool provided a user-friendly web interface for manually labeling and tagging entities—a critical step in ensuring that the NER models would be trained on high-quality, context-specific data. Following preprocessing and annotation, a total of 1,286 annotated datasets were obtained. This dataset was then stratified into training (80%), development (10%), and test (10%) subsets to ensure a balanced evaluation of the model's performance. Table 1 shows the word distribution of the datasets.

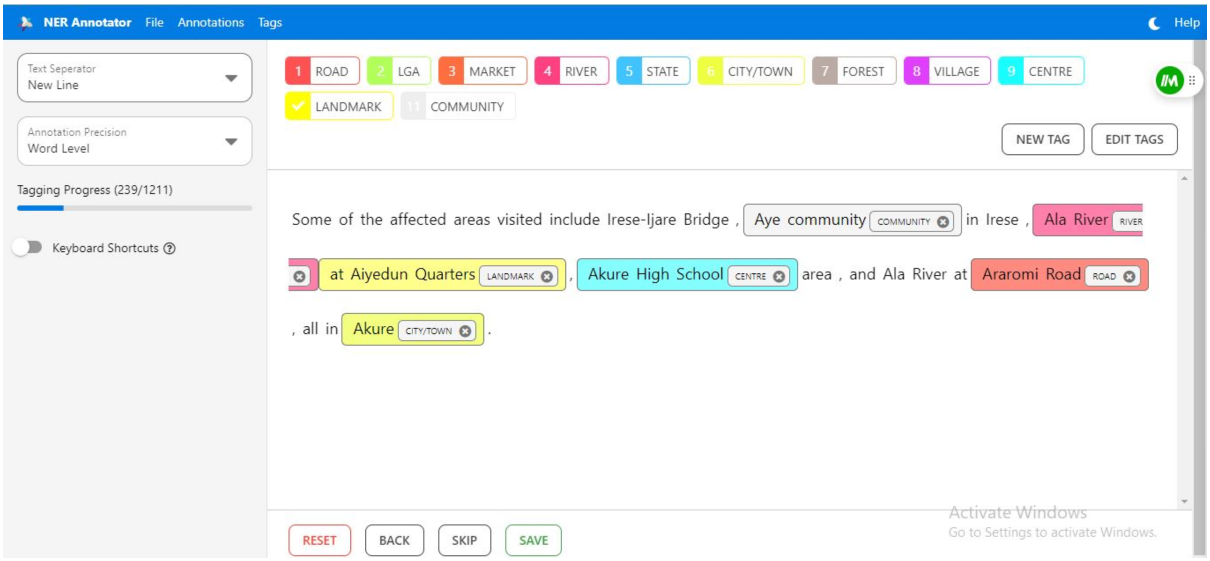

Figure 2: SpaCy NER Annotator Tool

Table 1: Dataset Word Distribution

Entity	Train Count	Train Percent (%)	Test Count	Test Percent (%)
Village	17	1	8	1
Market	140	7	70	10
Centre	65	3	80	12
State	536	28	145	22
LGA	350	18	103	15
City/Town	350	18	107	17
Landmark	130	7	71	11
Road	82	4	28	4
Community	192	10	47	7
River	76	4	5	1

4.2 Experimental Setup

This phase involves fine-tuning multiple Named Entity Recognition (NER) models—spaCy, BERT, and DistilBERT—to develop a Disaster-Location-Based (DLB) NER model. These models were trained using the annotated dataset to establish and improve location entity recognition Idakwo et al. (2023).

For SpaCy model was trained for 1,200 steps on an NVIDIA T4 GPU in Google Colab using the Stochastic Gradient Descent (SGD) optimizer with a batch size of 1,000, a dropout rate of 0.1, and a learning rate of 0.001. Model evaluation occurred every 200 steps to monitor performance and convergence. Similarly, BERT and DistilBERT were fine-tuned for 5 epochs each, using a learning rate of 0.00002, a batch size of 16 for memory efficiency, and a weight decay of 0.01 to prevent overfitting. Training was carried out on an NVIDIA T4 GPU in Google Colab under consistent experimental conditions. This structured fine-tuning process optimized the DLB NER model across SpaCy, BERT, and DistilBERT architectures, enhancing its capacity to accurately identify disaster-prone locations effectively.

The overall process, from data collection and preprocessing through manual annotation and model experimentation, required a significant investment of time and resources. The systematic scraping, cleaning, and refinement of over 3,700 articles, coupled with the meticulous manual labeling process, spanned several weeks. This effort was critical to ensure that the final dataset was both comprehensive and of quality, thereby enabling robust model training.

4.3 Geo-parsing & Visualization

This phase encompassed toponym recognition and resolution, key aspects of Geographic Information Retrieval (GIR). The DLB NER model was employed for toponym identification, while Nominatim, an OpenStreetMap-based tool, facilitated toponym resolution by linking place names to geographic coordinates. The identified toponyms were then visualized using Folium within Jupyter Notebook, following the methodology of Ivan (2018), enabling an interactive spatial representation of disaster-prone locations.

5 RESULTS

5.1 Performance Evaluation of the Models

The performance of the fine-tuned Disaster-Location-Based (DLB) NER models was assessed using precision, recall, and f1-score metrics. As shown in Table 2, the fine-tuned models exhibit variations in performance, highlighting the impact of domain-specific adaptation. BERT achieved the highest performance with a precision of 0.99331, recall of 0.99349, and f1-score of 0.99297, closely followed by DistilBERT with a precision of 0.99236, recall of 0.99297, and f1-score of 0.99240. In contrast, SpaCy exhibited lower performance with a precision of 0.95, recall of 0.77, and f1-score of 0.85. These results confirm the effectiveness of transformer-based models for NER tasks. The superior precision and recall of BERT and DistilBERT highlight their robustness in identi- fying disaster-related entities. Their f1-scores indicate suitability for real-world applications, ensuring accurate extraction of critical disaster information for emergency response teams.

Table 2: Comparison of DLB NER of the various models

Models	**precision**	**recall**	**f1-score**
spaCy	0.95	0.77	0.85
BERT	0.99331	0.99349	0.99297
DistilBERT	0.99236	0.99297	0.99240

5.2 Toponyms Parsing and Visuallization

Geo-parsing toponyms from text generally results in point-based representations located at the centers of towns, cities, villages, or other polygon/line geospatial features.

In this phase, toponyms were extracted, and their corresponding latitude and longitude coordinates are presented in table 3. These locations were then visualized on an interactive map, as shown in figure 3, providing a clear representation of the spatial distribution of the identified toponyms.

Location	**Entity**	**Latitude**	**Longitude**
Awe	LGA	8.111812	9.146371
Nasarawa	STATE	8.438787	8.238285
Lokoja	CITY/TOWN	7.802355	6.743033
Ogbaru Market	MARKET	5.921286	6.760997

Table 3: Latitude and Longitude of Some Extracted Toponyms

The interactive visualization (figure 3) employs color coding—blue for floods, red for fires, and green for oil spills—to differentiate disaster types. Findings indicate that oil spills are concentrated in Nigeria's South-South and South-East regions, while flooding is the most widespread disaster nationwide.

6 CONCLUSION & FUTURE WORK

The paper presents a comprehensive methodological framework for extracting geographical locations from disaster-related articles using Named Entity Recognition (NER) models. Our approach demonstrates strong overall performance in identifying and geo-visualizing locations associated with disasters. The ability to accurately extract and locate disaster-affected areas is vital for enhancing situational awareness and improving the efficiency of emergency response operations. By precisely pinpointing affected regions, our method facilitates the more effective allocation of resources, enabling timely assistance to impacted populations. The findings highlight the critical role of advanced NER models in disaster management, illustrating their capacity to transform textual data into actionable geographic insights. This technological advancement not only aids immediate response efforts but also contributes to long-term disaster planning and preparedness strategies.

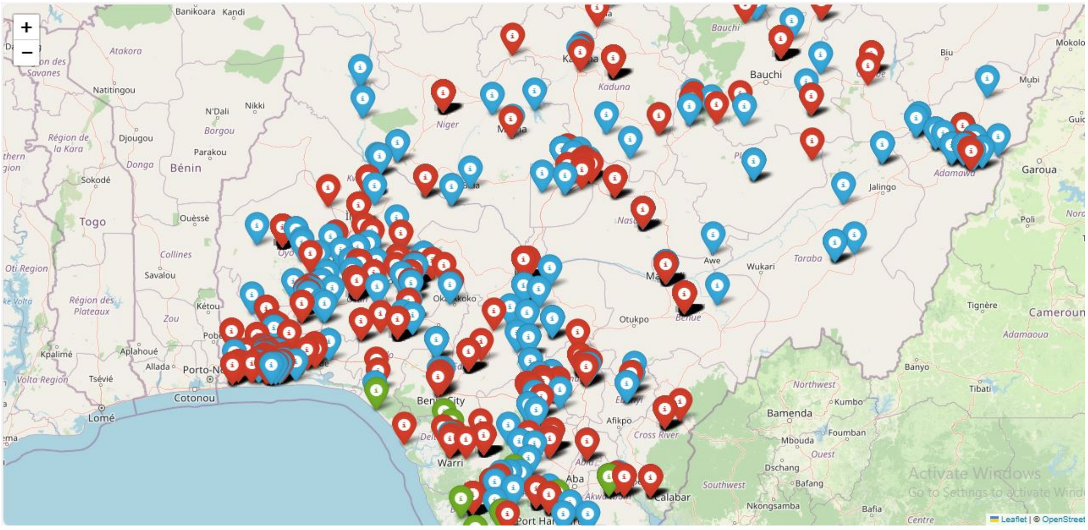

Figure 3: Map Showing the Disaster Locations

In addition, we recognize that the current framework primarily utilizes news articles, which, while valuable, may not capture the full immediacy of rapidly unfolding events. Datasets sourced from platforms like Twitter or Bluesky, which offer robust and timely data streams, could further enhance the framework's effectiveness by reducing disaster response times.

Future research will focus on further optimizing the NER model by fine-tuning hyperparameters, using more diverse datasets—including those from social media—and exploring advanced model architectures to enhance performance. Evaluating the framework on social media posts will be a key step in determining its potential to create even more significant improvements in disaster response times, ultimately leading to a more robust and dynamic system for emergency management.

References

Samira Abdul, Ehizogie Paul Adeghe, Bisola Oluwafadekemi Adegoke, Adebukola Adejumoke Adegoke, and Emem Henry Udedeh. Ai-enhanced healthcare management during natural disasters: conceptual insights. *Engineering Science & Technology Journal*, 5(5):1794–1816, 2024.

Zi Chen, Badal Pokharel, Bingnan Li, and Samsung Lim. Location extraction from twitter messages using bidirectional long short-term memory model. In *GISTAM*, pages 45–50, 2020.

Nikhil Dhavase and AM Bagade. Location identification for crime & disaster events by geoparsing twitter. In *International Conference for Convergence for Technology-2014*, pages 1–3. IEEE, 2014.

Patricia Ojonoka Idakwo, Olubayo Adekanmbi, Anthony Soronnadi, and Amos David. Geo-parsing and geo-visualization of road traffic crash incident locations from print media for emergency response and planning. In *5th Workshop on African Natural Language Processing*, 2023.

Anthony Ivan. Spatial visualizations and analysis in python with folium. *Towards Data Science*, 2018.

Lishuang Li, Degen Huang, and Chunrong Chen. Identification of location names from chinese texts based on support vector machine. *JOURNAL-DALIAN UNIVERSITY OF TECHNOLOGY*, 47(3): 433, 2007.

John Lingad, Sarvnaz Karimi, and Jie Yin. Location extraction from disaster-related microblogs. In *Proceedings of the 22nd international conference on world wide web*, pages 1017–1020, 2013.

Tolulope O Olorunsogo, Anthony Anyanwu, Temitayo Oluwaseun Abrahams, Temidayo Olorunsogo, Benedicta Ehimuan, Oluwatosin Reis, et al. Emerging technologies in public health campaigns: Artificial intelligence and big data. *International Journal of Science and Research Archive*, 11(1):478–487, 2024.

M. Orhewere. Nigeria's brand of disaster management. *Daily Times*, 2012. URL `http://www.dailytimes.com.ng/opinion/nigeria%E2%80%99s-brand-disaster-management`.

Ni Putu Ayu Sherly Anggita Sugiartaa and Ngurah Agus Sanjaya ERa. Location named-entity recognition using rule-based approach for balinese texts. *Jurnal Elektronik Ilmu Komputer Udayana p-ISSN*, 2301: 5373, 2021.

Muhammad Suleman, Muhammad Asif, Tayyab Zamir, Ayaz Mehmood, Jebran Khan, Nasir Ahmad, and Kashif Ahmad. Floods relevancy and identification of location from twitter posts using nlp techniques. *arXiv preprint arXiv:2301.00321*, 2023.