Geo-semantic surveillance and clustering of social risk hotspots using print media reports

Toyib Ogunremi1
Olubayo Adekanbi1
Anthony Soronnadi1
David Akanji1

1Data Scientists Network

Introduction

Kidnapping is a significant social risk in Nigeria which often lack adequate intervention due to underreporting of cases by the public due to fear of retaliation from suspected perpetrators or involvement of security operatives.

This area is also under-explored by researchers due to the unavailability of open-source local crime data. In response, we have developed a data-driven solution by generating a reliable dataset of crime locations and entities in Nigeria. We have therefore generated insights into the spatio-temporal dynamics of kidnapping occurrences nationwide.

Methodology

<table>
<thead>
<tr>
<th>Data Acquisition (Online Newspaper Scraping)</th>
<th>Data Classification (Filtering)</th>
<th>Rescue Operation Reports, Public commentary, and Editorial on Kidnapping</th>
<th>Geoparsing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Retrieval (IR) Victims Kidnapper</td>
<td>1. Toponym recognition</td>
<td>2. Toponym resolution</td>
<td>Data Preprocessing (NER)</td>
</tr>
<tr>
<td>Exploratory Data Analysis</td>
<td>Cluster of risk hotspots</td>
<td>Crime entities aggregation</td>
<td>Visualization of resolved locations</td>
</tr>
</tbody>
</table>

Entity Recognition & Semantic Analysis

<table>
<thead>
<tr>
<th>publish_date</th>
<th>headline</th>
<th>victims</th>
<th>kidnapper</th>
<th>address</th>
<th>Latitude, Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/9/2023</td>
<td>Bandits kidnap 40 Kaduna worshippers, nine airlft’s children</td>
<td>Bandits</td>
<td>Chikun, Kaduna, Nigeria</td>
<td>10.278389, 7.179026</td>
<td></td>
</tr>
<tr>
<td>4/7/2023</td>
<td>Gunmen kidnap Nasarawa ex-deputy governor, Gve-Wado</td>
<td>Gunmen</td>
<td>960134, Wamba, Nasarawa, Nigeria</td>
<td>8.943208, 8.6023254</td>
<td></td>
</tr>
<tr>
<td>3/12/2023</td>
<td>Gunmen kidnap nine in Abuja estate</td>
<td>nine</td>
<td>Gve-Wado</td>
<td>Grow Home Estate, Bwata Area council, 901101, FCT, Nigeria</td>
<td>9.1921367, 3.8858955</td>
</tr>
</tbody>
</table>

Table 1. Sample generated dataset

Results

Figure 1. Nigeria kidnapping Crisis (Al Jazeera)

Figure 2. Methodology workflow

Figure 3. Newspaper headline on Kidnapping

Figure 4. Process of Entity recognition and Semantic Analysis

Figure 5. Periodic hotspots clusters

Figure 6. Regional distribution of cases over the last 7 years.

Figure 7: Potential application in Surveillance systems and Dashboarding

Conclusion

Our analysis revealed the spatio-temporal analysis of kidnapping cases in Nigeria, 2019 has most states affected, this can be attributed to the COVID-19 induced hardships. We observed that some state’s high kidnapping rates could be attributed to factors such as overpopulation, politics, Religious intolerance and high cost of living, as noted in high-risk states like Kaduna, Lagos, Rivers, and Federal Capital Territory FCT.

This study gathered data from online newspaper report from Punch Newspaper, which has a cutoff of articles beyond 2016 on their website for the subject matter.

Future work will focus on improving feature extraction from newspaper articles beyond the syntactic entities to circumstances of crime events and other crime types like theft, rape, riot, murder etc. while also extending this surveillance technique to other major African languages like Hausa, Igbo and Yoruba which can finally be augmented with a speech to text pipeline to handle audio reported crimes.

References