Project Deliverable Pt1

GROUP – 20

Anol Kurian Vadakkeparampil 5626-8544

Aadithya Kandeth 6980-2791

Spatiotemporal Analysis and Prediction of Global Societal Unrest

Description:

This project aims to predict and analyze global societal unrest (unemployment rates/ geographic diffusion) caused by significant events like recessions/economic-sanctions/political disputes using Global News coupled with spatiotemporal indicators such as Education Statistics and Economic Prospects. Our inputs encompass extensive GDELT event data as our main source supported by micro-data from various sources like worldbank.org, with missing data filled using LLMs like OpenAI GPT Api. We plan to process this data to detect patterns, trends, and predict potential social unrest indicators using real time world news. The output comprises of a predictive model and insights that offer early warnings and forecasts of social unrest events globally.

Societal Applications:

By predicting and analyzing global social unrest, we enable early intervention and conflict prevention, potentially saving lives during times of crisis. For instance, our insights can aid humanitarian organizations in swiftly responding to and mitigating the impact of social unrest events, ensuring the timely delivery of aid to affected regions.

Our predictive models can inform policymakers about areas at risk, facilitating the allocation of resources to address underlying issues, such as economic disparities or political instability before they escalate into conflicts. This project strives to contribute to a safer, more stable world by harnessing data-driven insights to anticipate and manage societal challenges.

Primary Dataset:

Data is sourced from the Global Database of Events, Language, and Tone (GDELT) 2.0 Event Database

- Url: https://www.gdeltproject.org/data.html#rawdatafiles.
- 56 Columns/Attributes per record with 1.5+ billion event records.
  Key attributes include Latitude-Longitude, EventRootCode (e.g., "Protest"), Actor1/Actor2-Code (e.g., "Government"), Datetime, NumMentions, and ToneCode.
- For ground truth, labels can be derived from Cameo Code "14.3 Protest" classifying peaceful demonstrations. Goldstein Scale assigns numeric scores (-10 to +10) indicating event impact on country stability.
- Dataset Size:
  ~500 GB only GDELT(1979 – present) updated every 15 minutes.

Micro-Data mentioned on next page.

Micro-Data Sources:

https://databank.worldbank.org/databases	World Development Indicators Education Statistics - All Indicators Health Nutrition and Population Statistics Global Economic Prospects
https://ourworldindata.org/	Population and Demographic Change Health Energy and Environment Happiness Index Education and Knowledge Human Rights and Democracy Violence and War
https://acleddata.com/	Event deadliness Danger to civilians Geographic diffusion Armed group fragmentation
Potential Additional sources	https://dhsprogram.com/ -Birth Registration -Health Insurance -Maternal Mortality