

Final Project: Global Data Analysis for Social Unrest Prediction

Raksha Harish, Tavleen Sahota, Harman Jit Singh

1. List 3 questions that you intend to answer (1 point)

Question-1: How do factors such as poverty, hunger, natural calamities, and social biases influence social unrest (protests, demonstrations, riots, etc.) across the globe?

Question-2: What patterns are observed across social unrest events?

Question-3: Can these insights be used to predict social unrest and develop preventative strategies?

2. List all the datasets you intend to use (1 point)

- World food program (WFP) data
- International Relations and Human Rights data – Harvard University
- World Poverty data – The World Bank
- Hazards and Disaster Risk (Socioeconomic Data) – NASA, USA
- Crime rates data – UN Office on Drugs and Crime
- Net National Household Income per annum – The World Bank
- ACLED Data – The Armed Conflict Location and Event Data Project
- Civil Unrest Event Data (data.world) – Data from the Social, Political, Economic Event Database (SPEED) Project, University of Illinois Urbana-Champaign
- The GDELT Project (Global Data on Events, Location, and Tone) data

3. Give us a rough idea on how you plan to use the datasets to answer these questions. (2 points)

- Data Collection: The datasets mentioned in Part-2 are available for public download at their respective websites. Web scraping can also be used to collect additional data relevant to our analysis.
- Data Exploration: EDA will be carried out in order to understand the underlying nature of data as most of the data is collected from different sources.
- Data Cleaning: Data collected from different sources lacks uniform structure and certain aspects of the raw data are irrelevant to our work. Hence, we intend to clean the data by performing ETL tasks and filtering it to meet specific conditions (e.g., use data for a specific time period).

- Data Integration: Since we are using datasets from a variety of sources, we need to integrate them to observe the influence of certain factors for specific countries across the world. Data integration would also help in understanding the combined effect of factors that are actually leading to social unrest.
- Data Analysis: As we intend to analyze a number of factors and utilize a variety of data, we will conduct literature review to determine appropriate analysis methodology for each factor. Generally, we intend to perform statistical analyses such as predictive and causal analyses as well as use deep learning in the context of text analytics. We will primarily use evaluation metrics to evaluate our analysis results.
- Data Product: We intend to present the EDA results and visualizations in an interactive web-app. Based on our findings, we will select suitable features and also train a model to predict the probability of social unrest in the future.

4. Think about that once your project is complete, what impacts it can make. Pick up the greatest one and write it down. (1 point)

Conducting comprehensive analysis of publicly available global data, to determine factors affecting social unrest can allow us to anticipate these latent instabilities and take measures to handle or prevent them. Such insights may be of interest to governments, social scientists and other decision makers, as it encourages effective policy-making and prioritization of citizen grievances among other things.