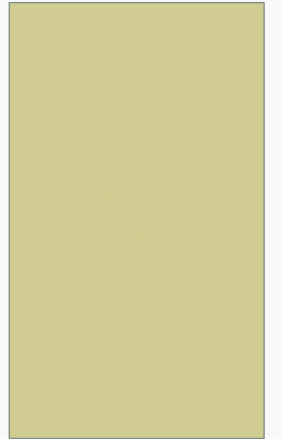# CMPT733 PROJECT MILESTONE
TOPIC : "GLOBAL DATA ANALYSIS FOR SOCIAL UNREST PREDICTION"

TEAM: RAKSHA HARISH, TAVLEEN SAHOTA, HARMAN JIT SINGH

PROBLEM STATEMENT

- How can we predict social unrest events (protests, riots, etc.) across the globe based on factors like historical event patterns, poverty, hunger, crime rate, natural calamities, etc.?

- What insights can be derived from the study of these factors and are there important relationships among these factors?

- What other approaches can be implemented to accurately predict social unrest?

- **Application domain : World governments, policy making, Social science, world hunger/disaster management.**

IDEAS

- Study global factors like poverty, hunger, crime rate, natural calamities and social biases in order to determine the factors that may lead to social unrest in the future.

- Analyse historical event to determine patterns useful for social unrest event prediction.

- Build competing prediction models and determine their accuracy.

- Suggest possible preventive strategies to avoid future social unrest.

DATASETS

- The GDELT Project (Global Data on Events, Location, and Tone) data – (over 2.5 TB), using Google BigQuery.
- 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

PROJECT DEVELOPMENT

1. EDA : Perform comprehensive analysis including descriptive statistics, temporal analysis, comparative analysis, etc. Use these findings to determine next analysis/course-of-action.
2. Visualize data patterns to observe previous trends that led to social unrest.
3. Integration of a variety of data sources based on relevance (for specific time period).
4. Feature Selection and Engineering from the available datasets and (possibly) real-time twitter feed, to predict social unrest.
5. Analysis of the twitter feed, and finding similarities in the available cleaned datasets.
6. Build competing models to predict how and when these global factors may lead to social unrest and fine-tune the best one.
7. Build an interactive UI allowing users to access real-time social unrest predictions and other insights.

TECHNOLOGIES

- Python with relevant packages for EDA + Visualization
- Data Storage – AWS Cloud S3 buckets
- Execution in Jupyter Notebook or JupyterLab (in AWS)
- Interactive Web-app using Flask
- SCM - GitHub

TENTATIVE TIMELINE

Task	Week	Status
Literature Review	Week 6-7	Completed
EDA + Twitter API	Week 7-9	In-progress
Modeling	Week 10-11	
EDA - iteration 2 (if needed)	Week 11-12	
Improving Model	Week 11-13	
UI	Week 10-12	
Integration + Debugging	Week 12-13	
Documentation	Week 13	

Note: This is a tentative timeline to track progress, not to be mistaken for Waterfall methodology.