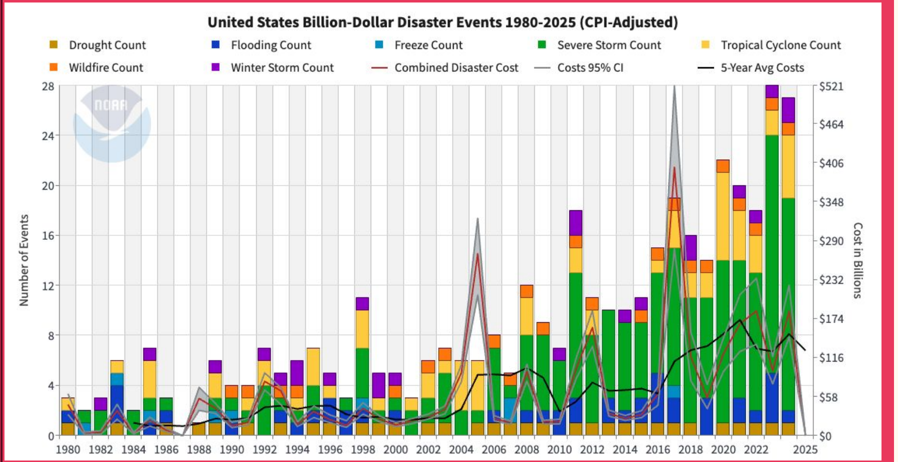
NATURE & PROTESTS

Irem Altan, Guga Gogia, Felicia Setiono, Ewerton Belchior
Erdos Data Science SP 2025

BACKGROUND

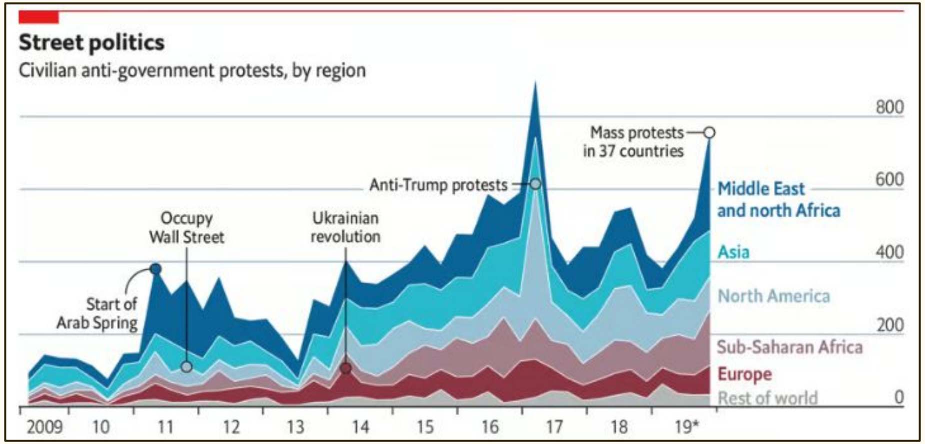
2010-2020: A Decade of Disruption

Rising Crises:
Increase in frequency, intensity & cost from extreme weather events



Source: *NOAA National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters (2025)*

Rising Voices:
The largest protest movements in recent history across the globe



Source: *CSI, The Age of Mass Protests: Understanding an Escalating Global Trend*

BACKGROUND

- Proposed by a paper titled "Economic Shocks and Civil Conflicts" by Miguel, Satyanath, and Sergenti (2004)

- For sub-Saharan Africa, rainfall fluctuations associated with GDP and GDP is associated with civil conflicts

GOALS:

1. Replicate two-step models from paper using more recent data, expanding to other countries.

2. Determine if including both climate and economic factors improve predicting protests.

DATASETS

1 Rainfall

Climate variables:

- Average rainfall
- Average surface temperature

Source: University of East Anglia, CRU TS datasets

2 Economy

Economic variables:

- GDP and %change
- Unemployment rate and % change
- % of oil share in GDP
- Democracy polity
- Ethnic Fractionation Index

Source: Our world in data, European University Institute Research Repository for EFI

3 Protests

Protest variables:

- # of distinct protests
- Whether protest happened or not

Source:Harvard Dataverse, Mass Mobilization Protest Data

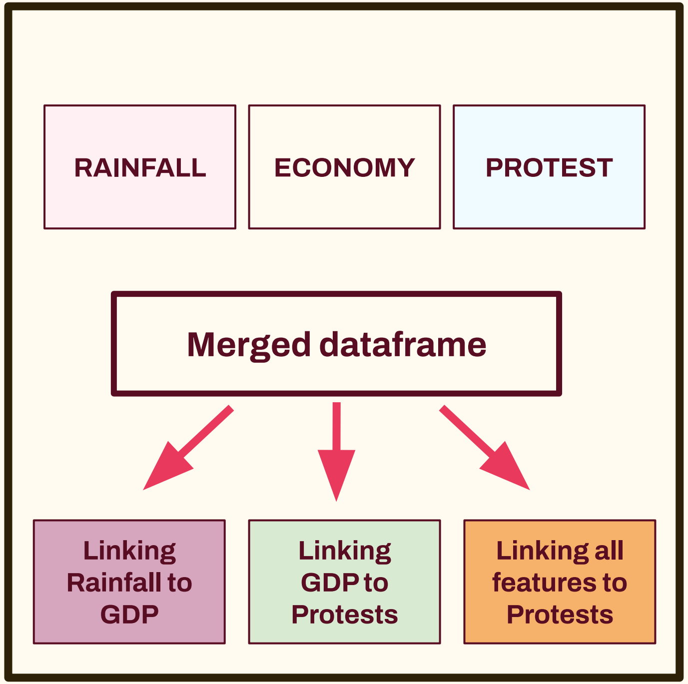
DATA PRE-PROCESSING

Cleaned and aggregated by country, by year.
Rainfall and economy (1960-2023)
Protest (1990-2020)

Merged by country names, year:

- 33 features
- 136 countries

Used the merged data to run three analyses

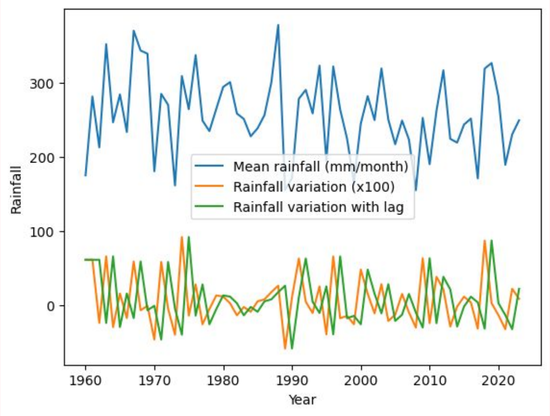


LINKING RAINFALL TO GDP VARIATION

BASELINE/LINEAR

Predictors:

- Rainfall variation ΔR(i,t): (Rainfall(i,t)-Rainfall(i,t-1))/Rainfall(i,t-1)
- Rainfall variation with lag ΔR(i,t-1):
- (Rainfall(i,t-1)-Rainfall(i,t-2))/Rainfall(i,t-2)
- Year(i)



Target variable: GDP variation
ΔGDP(i) = (GDP(i,t) - GDP(i,t-1))/GDP(i,t-1)

Countries: Sub-Saharan African countries
Years: 1960-2023

R^2=0.13
RMSE= 0.06

"Important" features with the original dataset (1980-2000):

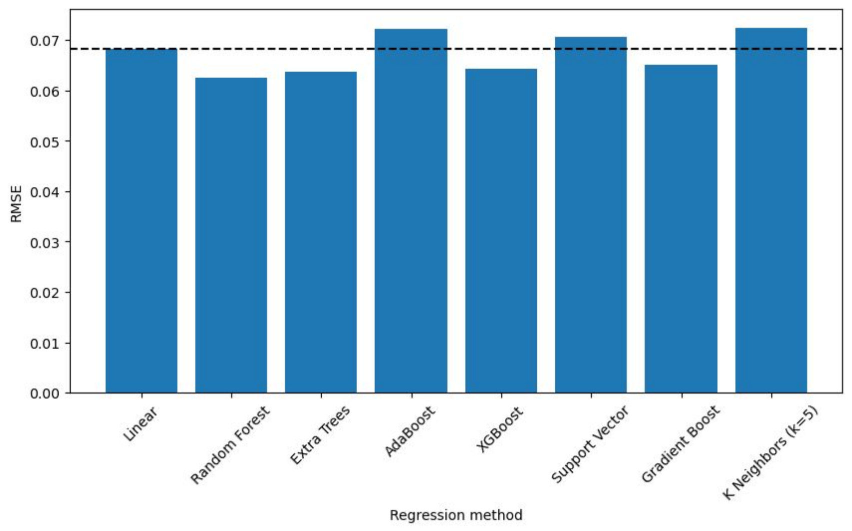
ΔR(i,t) & ΔR(i,t-1)

Bootstrapping with Kernel Ridge Regression on simulated data based on Linear Regression shows the data generating process is not linear (p=0.008).

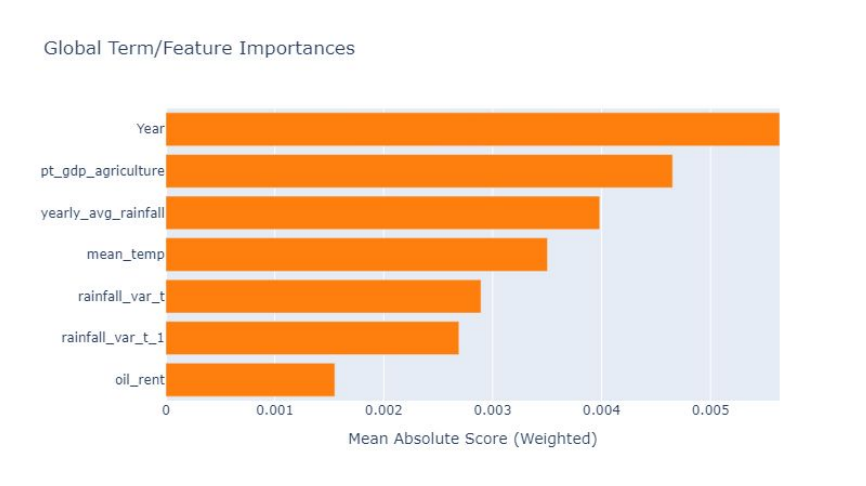
LINKING RAINFALL TO GDP VARIATION

Ensemble models and adding more predictors

- Percent GDP from agriculture
- Mean annual surface temperature
- Oil rent



Explainable Boosting Regressor



PREDICTING DEMONSTRATIONS

Target variable: Number of protest events

Predictors:
- GDP
- ΔGDP
- Unemployment rate
- ΔUnemployment rate
- % of oil share in GDP
- Democracy polity
- Ethnic Fractionation Index

+

Predictors:
- % of agriculture share in GDP
- Yearly average rainfall + variation + lag
- Cumulative rainfall difference
- Rainfall difference from average
- Number of years with rainfall below mean
- Average surface temperature + variation + lag

Models	RMSE	R2
Lin Regression	3.80	0.01
Ridge	3.80	0.01
Lasso	3.86	-0.01
Random Forest	3.57	0.10
XGboost	**3.53**	**0.12**

Feature Importances

- Cumulative rainfall difference
- % of agriculture share in GDP
- Average surface temperature
- Yearly average rainfall + variation + lag

FURTHER EXPLORATION

Adding other models:

Models	RMSE	R2
Random Forest	5.57	0.08
XGBoost	5.86	-0.05
CatBoost	5.44	0.13
Gradient Boosting	6.11	-0.13
LGBM	**5.31**	**0.17**

Hyperparameter Tuning for LGBM:

Final RMSE = 5.28

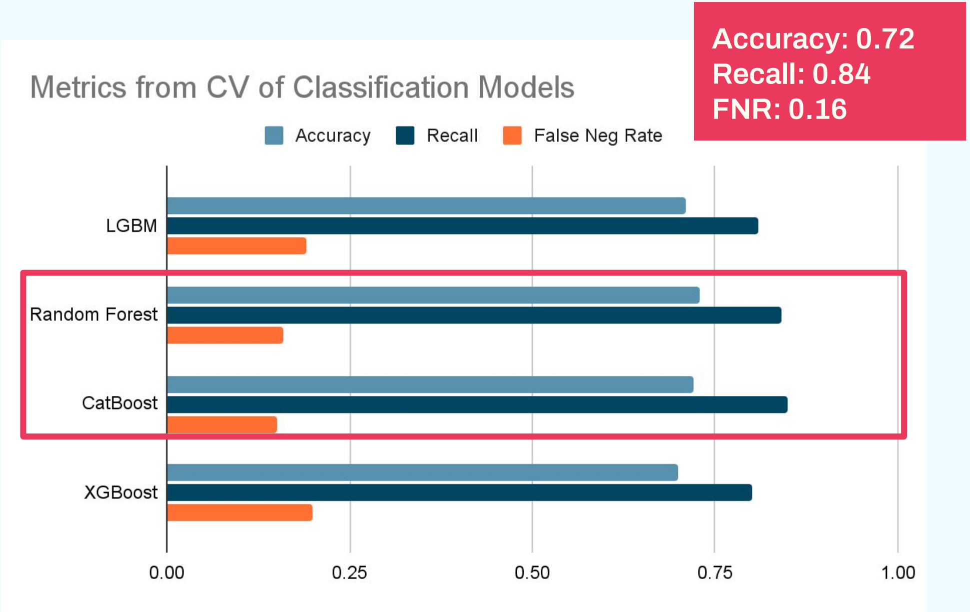
Final R^2 = 0.19

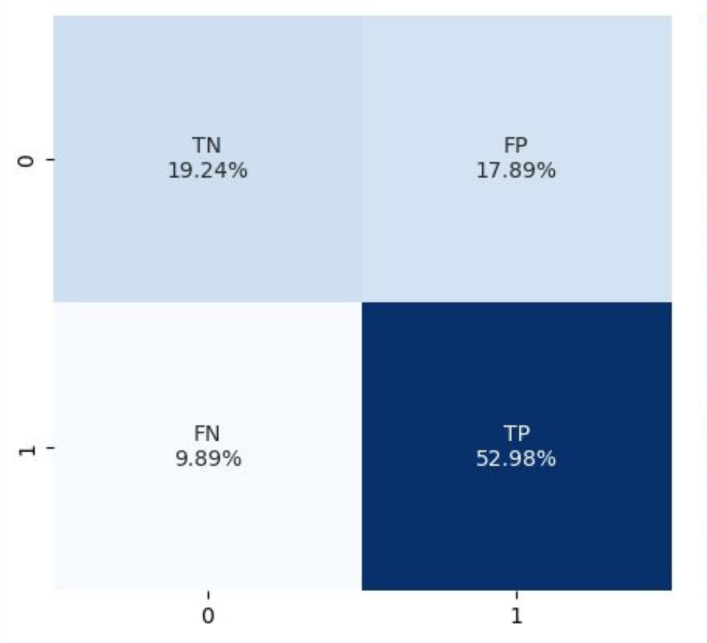
CLASSIFICATION MODELS

- **Target** : Boolean "did protest happen?"
- **Features:** Climate + Economic
- **Metrics:** Accuracy, recall, and false negative rate

Consideration: Stratified K-Fold cross validation with SMOTE resampling to deal with Imbalance data

CLASSIFICATION MODELS





FUTURE WORK

Nonlinear models after separating countries based on reliance on agricultural production.

Factoring in global food trade: does a natural disaster in a country's agricultural trade partner affect the country itself?

Factoring in natural disasters or other impactful events that do not directly arise from climate change.

Using recurrent neural networks, potentially with LSTM, to capture correlations across the years.

THANK YOU