

Foresight Project

DRC

Data

1010
1010

- The data is all derived from open source data. The main data sources are the World Bank development indicators, ACLED, UCDP, EM DAT, UN agencies (UNHCR, WFP, FAO), IDMC, etc. In total, the system aggregates data from 18 sources, and contains 148 indicators
- Given that the data is taken from reputable data sources, the data is deemed to be highly reliable. The data has a few shortcomings. First, coverage is uneven across geographies and across dimensions. For instance, economic and labour statistics tend to have better availability compared to governance and violence statistics. Data from institutional providers can often have a delay. The most recent indicators can be a few years old. The data is collected globally. For training we limit the data from 1995 through till 2018, the latest data available for displacement. For cross validation, we use a 5-year period between 2010-2015. Following the standard cross-validation setup for time series data, models are trained on data for the years (1995, y) and predictions made for y+t, where y is in the 5-year time period.
- The system uses several methods to address data gaps. We distinguish between the missing data in the features (or indicators) and missing target variable (i.e. forced displacement). Data with missing target variables are simply excluded from training. For missing values in indicators, we employ two methods. To address data lag, we make indicator projections for each country using an auto-regressive model (i.e. AR(n) model). An auto-regressive model is a time series forecasting model where future values depend only on previous values of the variable. . The 'n' denotes the number of lag variables and is determined using a heuristic approach. For cases where data is insufficient, we simply treat it as missing which is better than projecting incorrectly. Intermediate missing values are computed by interpolation
- We follow a simple standardization scheme, intended to keep data ingestion tasks lightweight. A data transformer is implemented for each of the data sources to ensure that each indicator data point is associated with a country and year. The resulting dataset can be cross-referenced and serves as input to the model.

Machine Learning Model

>_

- **Forecast Modelling:** The machine learning model employed is an Ensemble. An ensemble model works by leveraging several constituent models to generate independent forecasts that are then aggregated. Here we employ two gradient boosted trees to generate the point forecasts. The model hyperparameters were determined by means of a grid search. Each year-ahead forecast has a separate model. In other words, we train a set of Ensemble models for $y(t+h) = f(x(t))$, where $h = 0, 1, 2, 3$. The associated confidence intervals were generated by empirical bootstrap method, where the source error distributions were generated on a retrospective analysis. Model training data was limited to data since 1995 and we studied performance for the two main countries: Afghanistan and Myanmar.
- **Scenario "what-if" analysis:** The system supports what-if analysis for selected dimensions of economy, conflict, natural environment, governance, and population. For a user-specified scenario, the projections are updated based on estimated elasticities for each cluster. The elasticities capture the change displacement stock for a unit change in the thematic cluster. The elasticities are determined using an ordinary least squares regression. To compute the scenario projections, we use the ensemble model and apply the elasticities from the simpler regression model. For Afghanistan and Myanmar some of the scenario elasticities are not statistically significant. In these cases, we rely on a basket of 25 countries with a history of displacement, to provide an estimate.
- **Bayesian Network analysis:** To better understand causal links between key dimensions of the scenarios being constructed, the system implements a Bayesian network model which establishes probabilistic links between various dimensions. For a particular country level case, analysis can perform any number of inference tasks by specifying constraints to key dimensions and seeing the probability displacement distributions change. The model is built on data from 28 countries with a history of displacement

Foresight Project

DRC

Sources



1. Estimate of Occurrence for Complex Disasters disaster group
2. Estimate of Total Deaths for Complex Disasters disaster group
3. Estimate of Injured for Complex Disasters disaster group
4. Estimate of Affected for Complex Disasters disaster group
5. Estimate of Homeless for Complex Disasters disaster group
6. Estimate of Total affected for Complex Disasters disaster group
7. Estimate of Total damage for Complex Disasters disaster group
8. Estimate of Occurrence for Natural disaster group
9. Estimate of Total Deaths for Natural disaster group
10. Estimate of Injured for Natural disaster group
11. Estimate of Affected for Natural disaster group
12. Estimate of Homeless for Natural disaster group
13. Estimate of Total affected for Natural disaster group
14. Estimate of Total damage for Natural disaster group
15. Estimate of Occurrence for Technological disaster group
16. Estimate of Total Deaths for Technological disaster group
17. Estimate of Injured for Technological disaster group
18. Estimate of Affected for Technological disaster group
19. Estimate of Homeless for Technological disaster group
20. Estimate of Total affected for Technological disaster group
21. Estimate of Total damage for Technological disaster group
22. Political Rights
23. Civil Liberties
24. Freedom of religion
25. Freedom from political killings
26. Human Rights Score Mean
27. Human Rights Score Standard Deviation
28. Corruption Perception Index
29. Electoral democracy index
30. Freedom of religion
31. Freedom from political killings
32. Exclusion by Political Group index
33. Exclusion by Social Group index
34. Access to electricity (% of population)
35. Age dependency ratio (% of working-age population)
36. Battle-related deaths (number of people)
37. CPIA property rights and rule-based governance rating (1=low to 6=high)
38. Fertility rate, total (births per woman)
39. Food production index (2004-2006 = 100)
40. GDP per capita, PPP (current international \$)
41. Individuals using the Internet (% of population)
42. Inflation, consumer prices (annual %)
43. Livestock production index (2004-2006 = 100)
44. PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)
45. Population ages 20-24, female (% of female population)
46. Population ages 20-24, male (% of male population)
47. Population ages 25-29, female (% of female population)
48. Population ages 25-29, male (% of male population)
49. Population density (people per sq. km of land area)
50. Population growth (annual %)
51. Population, female (% of total population)
52. Population, total
53. Rural population
54. Unemployment, female (% of female labor force) (modeled ILO estimate)
55. Unemployment, male (% of male labor force) (modeled ILO estimate)
56. Unemployment, total (% of total labor force) (modeled ILO estimate)
57. Urban population
58. Urban population growth (annual %)
59. Literacy rate, adult total (% of people ages 15 and above)
60. Renewable internal freshwater resources per capita (cubic meters)

Freedom House

HUMAN RIGHTS SCORES



V-Dem INSTITUTE













THE WORLD BANK

Foresight Project

DRC

Sources

- 
61. Number of conflict events per year
62. Number of state-based conflict events per year
63. Number of non-state conflict events per year
64. Number of one-sided conflict events per year
65. Fatalities from conflict events per year
66. Fatalities from state-based conflict events per year
- 
70. Total number of violent incidents annually
71. Total number of fatalities annually from violence
- 
72. Magnitude score of episode of warfare episode
73. Magnitude score of episode of international violence
74. Magnitude score of episode of international warfare
75. Magnitude score of episode of civil violence
76. Magnitude score of episode of civil warfare
77. Magnitude score of episode of ethnic violence
78. Magnitude score of episode of ethnic warfare
79. Length of conflict in years
80. Scaled failure of State authority
- 
90. Security Apparatus
91. Factionalized Elites
92. Group Grievance
93. Economy
94. Economic Inequality
95. Human Flight and Brain Drain
- 
102. Asylum-seekers
103. Others of concern
104. Refugees (incl. refugee-like situations)
- 
107. UNHCR total externally displaced person
108. IDMC Internally displaced persons
109. Total forced displacement
- 
110. Maternal mortality ratio (per 100 000 live births)
111. Number of maternal deaths
112. Tuberculosis treatment coverage
113. Number of incident tuberculosis cases
114. Tuberculosis - new and relapse cases
115. Number of under-five deaths (thousands)
- 
116. Number of infant deaths (thousands)
117. Number of neonatal deaths (thousands)
118. External health expenditure (EXT) as percentage of current health expenditure (CHE) (%)
119. Out-of-pocket expenditure as percentage of current health expenditure (CHE) (%)
- 
120. Price of wheat in retail
- 
121. Average dietary energy supply adequacy
67. Fatalities from non-state conflict events per year
68. Fatalities from one-sided conflict events per year
69. Civilian fatalities from conflict events per year
81. Scaled collapse of democratic institutions
82. Scaled violence associated with regime transition
83. Scaled number of rebel combatants or activists
84. Scaled annual number of fatalities related to fighting
85. Scaled portion of country affected by fighting
86. Scaled annual number of Deaths
87. Institutionalized Democracy
88. Institutionalized Autocracy
89. Revised Combined PolityScore
96. State Legitimacy
97. Public Services
98. Human Rights
99. Demographic Pressures
100. Refugees and IDPs
101. External Intervention
105. Returned IDPs
106. Returned refugees