Trends and Determinants of Domestic Terrorism

Data Science 1 Project Proposal

Mary Kryslette C. Bunyi

Problem Statement

Terrorist activity has heightened globally amid deepening polarization and the ease of access to electronic communication channels. This is a concern especially in the Philippines, where there are fears that rebel groups have become and could continue to become more emboldened and radicalized due to the influence of foreign terrorist fighters.

However, terrorism literature specific to the terrorism environment in the Philippines is scant. I hope to address this by analyzing historical and emerging terrorism trends in the Philippines. Since Philippine data is not granular enough for statistical learning models and given varying results noted in existing terrorism literature on the determinants of terrorist incidents [1-6], I intend to take a machine learning approach in determining which variables matter most in predicting country-level terrorist activities on an annual basis.

Data Collection Plan

Data wrangling, exploration, and visualization will primarily revolve around the Global Terrorism Database (GTD) published by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) [7]. The database provides almost 50 years' worth of data on more than 200,000 terrorist attacks [7].

To determine the relationship of socioeconomic, political, and geographical variables with terrorism, in line with literature [1-6], the project will use the following supplemental data sources: (1) World Bank DataBank for socioeconomic variables, Worldwide Governance Indicators, and World Development Indicators [8]; (2) Harvard Dataverse for Human Rights Protection Scores [9]; (3) The Correlates of War Project for The World Religion Dataset [10]; and (4) Our World in Data for data on the Terrain Ruggedness Index [11] and as an alternative source to extract data on the Human Development Index, military spending, and country-level terrorism indicators [12-14]. All data sources provide downloadable spreadsheets through their respective websites.

Methodology

Data wrangling will mostly involve the consolidation and reconciliation of the various data sources. This shall include the reshaping of data, group manipulations, joining of multiple datasets, and interpolation of missing data. Python packages such as NumPy [15], pandas [16], and country-converter [17] will be used.

Data visualizations will depend on the insights drawn from exploratory data analysis and may cover purely GTD data or aggregated GTD data in comparison with the chosen macro-level variables. These may take the form of bar graphs on terrorism incidents and casualties over time, a heat map showing terrorism hotspots, and a histogram of commonly used modi operandi or motivations, among others.

Given the breadth of analysis done on terrorism incidents at a global scale and the lack of analysis specific to the Philippines, the priority is to create visualizations of trends and a mapping of terrorist attacks in the Philippines through time. Further data processing to generate the graphs will require the use of NumPy [15] and pandas [16]. In the data visualization proper, packages such as Matplotlib [18], plotnine [19], and/or seaborn [20] will be used. As Philippine data is not sufficiently granular for machine learning methods, country-year data will be used for this purpose. Incident-level data from the GTD will be collapsed into annual data and matched against the supplemental data sources, which contain year-level variables.

I aim to predict the number of annual domestic terrorist incidents and victims/casualties using explanatory variables culled from the alternative data sources such as economic growth, GDP per capita, inflation, population size, poverty, inequality, education/literacy, deprivation (as measured by the Human Development Index), unemployment rate, religion, terrain ruggedness, military spending, political stability, government effectiveness, rule of law, and human rights protection score. These variables were chosen based on previous terrorism studies [1-6]. Since the determinants we have identified are not necessarily applicable to international terrorist incidents, these will be excluded from the scope of the analysis. The data will be divided into training and test sets, with the split based on chronology. The test set shall contain most recent data points. The training data will then be subject to machine learning and cross validation methods using scikit-learn [21]. These algorithms may include K-fold Cross Validation, Linear Regression, Decision Trees, and Random Forest.

Success Indicators

I would define success broadly as generating data and insights that would enhance Philippine law enforcement's knowledge of terrorism and understanding of its dynamics. Specific indicators would be: (1) the creation of a new country-year dataset consolidating variables from the GTD and the supplemental data sources; (2) putting together useful descriptive information such as locations that are becoming hotspots and emerging modi operandi of terrorist actors; and (3) building a predictive model incorporating country-level factors to determine the variables most important in predicting terrorism.

Works Cited

[1] Orlandrew E. Danzell, Yao-Yuan Yeh & Melia Pfannenstiel (2019) Determinants of Domestic Terrorism: An Examination of Ethnic Polarization and Economic Development, *Terrorism and Political Violence, 31:3, 536-558*, DOI:/underline%7B10.1080/09546553.2016.1258636}

[2] Muhammad Nasir, Amanat Ali & Faiz Ur Rehman (2011) Determinants of Terrorism: A Panel Data Analysis of Selected South Asian Countries, *The Singapore Economic Review, 56:2, 175-187*, DOI:/underline%7B10.1142/S0217590811004225}

[3] K. Peren Arin, Oliver Lorz, Otto F.M. Reich & Nicola Spagnolo (2011) Exploring the dynamics between terrorism and anti-terror spending: Theory and UK-evidence, *Journal of Economic Behavior and Organization, 77, 189-202*, DOI:/underline%7B10.1016/j.jebo.2010.10.007}

[4] Mohammad Nurunnabi & Asma Sghaier (2018) Socioeconomic Determinants of Terrorism, *Digest of Middle East Studies, 27:2, 278-302*, DOI:/underline%7B10.1111/dome.12139}

[5] Andreas Freytag, Jens J. Krüger, Daniel Meierrieks & Friedrich Schneider (2011) The origins of terrorism: Cross-country estimates of socio-economic determinants of terrorism, *European Journal of Political Economy, 27, S5–S16*, DOI:/underline%7B10.1016/j.ejpoleco.2011.06.009}

[6] Raul Caruso & Friedrich Schneider (2011) The socio-economic determinants of terrorism and political violence in Western Europe (1994–2007), *European Journal of Political Economy 27, S37–S49*, DOI:/underline%7B10.1016/j.ejpoleco.2011.02.003}

[7] National Consortium for the Study of Terrorism and Responses to Terrorism (START). (n.d.) *Global Terrorism Database (GTD)* https://www.start.umd.edu/gtd/about/

[8] World Bank. (n.d.) *DataBank | The World Bank.* https://databank.worldbank.org/home

[9] Fariss, Christopher, 2019, "Latent Human Rights Protection Scores Version 3", https://doi.org/10.7910/DVN/TADPGE, Harvard Dataverse, V1, UNF:6:0sWy9tSpQVpzz2xGoGLtkA== [fileUNF]. Retrieved from: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/TADPGE)

[10] Zeev Maoz and Errol A. Henderson. 2013. "The World Religion Dataset, 1945-2010: Logic, Estimates, and Trends." *International Interactions, 39: 265-291.* Retrieved from: https://correlatesofwar.org/data-sets/world-religion-data)

[11] Hannah Ritchie and Max Roser (2013) - "Terrain Ruggedness Index". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/grapher/terrain-ruggedness-index [Online Resource]

[12] Max Roser (2014) - "Human Development Index (HDI)". *Published online at OurWorldInData.org.* Retrieved from: 'https://ourworldindata.org/human-development-index' [Online Resource]

[13] Max Roser and Mohamed Nagdy (2013) - "Military Spending". *Published online at OurWorldInData.org.* Retrieved from: 'https://ourworldindata.org/military-spending' [Online Resource]

[14] Hannah Ritchie, Joe Hasell, Cameron Appel and Max Roser (2013) - "Terrorism". *Published online at OurWorldInData.org.* Retrieved from: 'https://ourworldindata.org/terrorism' [Online Resource]

[15] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke & Travis E. Oliphant. **Array programming with NumPy**, Nature, **585**, 357–362 (2020), DOI:10.1038/s41586-020-2649-2

[16] Wes McKinney. **Data Structures for Statistical Computing in Python**, Proceedings of the 9th Python in Science Conference, 51-56 (2010)

[17] Stadler, K. (2017). The country converter coco - a Python package for converting country names between different classification schemes. The Journal of Open Source Software. doi: 10.21105/joss.00332

[18] John D. Hunter. **Matplotlib: A 2D Graphics Environment**, Computing in Science & Engineering, **9**, 90-95 (2007), DOI:10.1109/MCSE.2007.55

[19] J. D. Hunter, "Matplotlib: A 2D Graphics Environment", Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007.

[20] Waskom, M., Botvinnik, Olga, O'Kane, Drew, Hobson, Paul, Lukauskas, Saulius, Gemperline, David C, … Qalieh, Adel. (2017). *mwaskom/seaborn: v0.8.1 (September 2017)*. Zenodo. https://doi.org/10.5281/zenodo.883859

[21] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.