Forecasting Conflict Fatalities using Sequential Neural Network

Hugo Troendle
University of Lausanne
MSc. in Management, Business Analytics
hugo.troendle@unil.ch

***Abstract*—This study aims to develop a predictive model for fatalities in violent conflicts to assist policymakers and peace-builders in anticipating and responding to emerging risks of conflict. Utilizing the Uppsala Conflict Data Program Dataset on Organized Violence per country-year, a Neural Network and a linear regression model with LASSO regularization were trained and their performances compared. The Neural Network demonstrated some predictive capability for low to very-low fatality conflict scenarios, but struggled with medium to high fatality conflicts. Conversely, the LASSO regression model exhibited robust generalization but favored zero-fatality predictions, limiting its precision for severe conflicts. Despite rigorous methodologies, both models failed to achieve the necessary accuracy for practical applications due to significant discrepancies in predicted fatalities. Future research should focus on incorporating additional features and increasing the number of observations by adopting a country-month approach. Moreover, developing separate models for low-fatality and high-fatality conflicts could enhance forecasting accuracy, though their practical application must be carefully considered.**

***Index Terms*—Conflict, Prediction, Machine learning, Fatalities**

I. Introduction

Ukraine, Ethiopia, Sudan, Gaza, Myanmar. Recent years have witnessed the highest number of violent conflicts worldwide since the end of the Second World War. As a result, hundreds of thousands of people, including countless civilians, are killed each year.

"Peace cannot be kept by force; it can only be achieved by understanding" - Albert Einstein

While this famous quote originally referred to empathy and human connection, the advent of data science can offer a new lens with which to approach it. By exploring deep underlying patterns within data using machine learning techniques, the root causes of fatalities in organized violence can be better understood. This research aims to develop a robust predictive model, providing policymakers and peacebuilders with a tool to better anticipate and respond to emerging risks of conflict. Using the Uppsala Conflict Data Program (UCDP) Dataset on Organized Violence per country-year, a Neural Network and a linear regression with LASSO regularization are trained and their performance compared. By offering a detailed analysis on conflict fatalities, this research seeks to contribute to the field of conflict studies and offer practical tools for decision-makers. The goal is to support efforts towards conflict prevention and peacebuilding, enabling proactive measures to save lives and promote stability in conflict-prone regions.

II. Literature Review

A. Conflict definition

Research on armed conflicts is plentiful. There exist many branches and sub-branches of research that aims to explore and understand different aspects of wars, terror, coups, and peace. In a literature review on conflict research published in 2006 [1], the authors display the different approaches used by different fields of academia: "economists are focused on game-theory and decision-making, psychologist explore inter-personal conflicts, sociologists take status and class conflicts as the focal point, while political science is centered on intra-national and international conflicts."
The definition itself of conflicts is debated, with different research program having each their own definition. Arbitrary boundaries on measurable metrics are necessary to categorize what constitutes a violent conflict. For example, the UCDP defines an armed conflict as "a contested incompatibility that concerns government and/or territory over which the use of armed force between two parties, of which at least one is the government of a state, has resulted in at least 25 battle-related deaths in one calendar year" [2]. The Correlates-of-War-Project (COW) [3] has a similar definition but sets the battle-related deaths lower bound to 100 and excludes civilian victims. Although their scope can differ, the number of fatalities resulting from the conflict is a common metric in most research and serves as a significant proxy indicator of conflict intensity.

B. Current research

A wide array of Data Science methods has been applied in the context of predicting the occurrence of armed conflicts through various lenses. Examples include leveraging geographical data to identify spatial instability and predict conflicts [4], automated machine learning and human-guided machine learning techniques to model and forecast armed conflicts [5], classification methods to predict the types of conflicts [6], or text mining of geo-located news articles to predict conflict outbreaks [7]. Earlier research also provides an overview of indicators significant for understanding the nature of conflict risk, such as the economic situation of a region [8] and its educational attainment [9]. However, while these studies focus on predicting, forecasting and modeling conflicts, they often fail to encompass the cost of human life associated with it. The VIEWS early warning system offers sub-national-level fatalities prediction for Africa and the Middle East, and global fatalities prediction at the country-level [10], but it excludes both non-state-based violence and one-sided violence. Fatalities caused by actors such as the Wagner Group and the Islamic State are therefore omitted unless directly in conflict with a state-based actor.

"War does not determine who is right – only who is left" - Bertrand Russel

C. Research question

How can data science be leveraged to predict fatalities in conflict and promote peace? In this context, my research does not aim to predict whether a conflict occurs, but rather aims to predict its intensity using fatalities per country-year as a proxy indicator. The rationale behind using these geographical and temporal boundaries is two-fold. First, the data availability was considered in regards to the project feasibility. Second, analyzing fatalities on a country-by-country basis allows for a clear identification of which governments may have failed to implement measures or policies that could have prevented these deaths. This research is intended as a tool for decision-makers to use before conflicts escalate. Similar to the effective approach used by the Conflict Forecast project team [7], additional steps after the model output, such as assigning a monetary value per loss of life, can serve as powerful policy drivers by displaying the results in a cost-benefit structure.

III. Methodology

A. Data selection and pre-processing

The first step in this project was to select, clean, and prepare relevant data to train the models. Three main difficulties were associated with this process. First, finding data on a global basis proved to be challenging. Even basic information such as Gross Domestic Product (GDP) is not openly available for every country. Second, independent state recognition is not equivalent in every data collection program. How the first two difficulties were addressed is discussed in the next section Dataset description.

The third difficulty resided in the fact that the data used to train the model must be available to decision-makers before the predicted events occurs. A few very interesting data sets, such as the "Number of political violence events by country-year" by the Armed Conflict Location and Events Data (ACLED) [11] therefore had to be disregarded in order to avoid introducing bias. The data used to train the model has to be predictable by any reasonable military or political analyst in a short to medium time range. As an example, given a conflict happens, the number of countries involved in a country-year can reasonably be guessed but the number of recorded explosions in a country-year can not be reasonably predicted.

B. Feature engineering and transformation

Two new features, *years_since_last_civilian_death* and *years_since_last_death*, were engineered using the existing features to carry information about the relative stability of the country-year in the model. The variable *years_since_last_civilian_death* is reset anytime a civilian is killed in the context of organized violence. As an example, the value of *years_since_last_civilian_death* in France in 2016 (the year after the November 13th terrorist attack in Paris) was reset to 1. The *years_since_last_death* variable behaves the same way but encompasses all victims of organized violence.

The features *country_cy* and *region_cy*, that represent respectively the country, and the region (continent) of the country-year were transformed into numerical categories. For model complexity reasons, the one-hot encoding technique, that maps categorical features into binary representation and prevents the model from interpreting numerical categories as ordinal, was not used. Indeed, the data set contains 196 unique countries, and one-hot encoding would have created 196 features, each corresponding to a country, dramatically increasing the model complexity.

C. Principal Component Analysis

A principal component analysis (PCA) was conducted on the education dataset. PCA is a popular statistical technique that transforms correlated variables into a set of linearly uncorrelated variables. This technique reduced the dimensionality of the education dataset from 2697 variables to 8 principal components that contained slightly more than 80% of the total variance in the initial dataset.

D. K-Means Clustering and Total Within Sum of Square

K-Means is a partitioning method for a given number of clusters. This technique randomly allocates each instance to a cluster, computes the clusters' centers, then reallocates the instances to the cluster with the closest centers until convergence. The number of clusters was determined using the Total within (cluster) sum of squares (TWSS). For each additional cluster, the TWSS diminishes. New clusters are created until an elbow can be observed on a graph plotting the number of clusters vs. the TWSS.

Using these techniques, each country was allocated to one of 5 clusters based on their Economic and Education level. This allowed for the creation of a new feature for the supervised model that accounts for additional information not present in the initial dataset. An additional PCA was conducted on the dataset containing the economic and education attributes for visualization purposes (see Figure 1).

E. Data splitting and shuffling

Data splitting is used to separate the data on which the model is trained, and the data on which its performance metrics are measured. A common strategy is to randomly split the training and test sets with 80-20 proportions, and then allocate a further 20% of the training set to a validation set. However, the application sought by this project calls for a different type of split. The model must predict future years based on historical and present data. To preserve the temporal nature of the data, the dataset was split by the *year_cy* feature. Data from 1989 to 2014 (approximately 75%) was used for training, while data from 2015 to 2022 was reserved for validation and testing. Both sets were then randomly shuffled. The data from 2015 to 2022 was subsequently divided into two equal parts, one for validation and one for testing.

F. Performance metrics

Root Mean Square Error (RMSE), Mean Squared Error (MSE) and Mean Absolute Error (MAE) were all considered as performance metrics for the different models tested. MAE was selected for two reasons. First, the MAE is robust to outliers, so the models will not overfit the few extremely bloody conflicts. Second, the results of this project are intended to be used by decision-makers, not necessarily familiar with data science. Therefore, it is important to use an intuitive metric, that is in the same unit as the target variable.

G. Neural Network and Hyperparameter Tuning

The main predictive model used was a Sequential Layer Dense Neural Network. Such model consists of an input layer, one or more hidden layers, and an output layer. The input layer feeds the features to the first hidden layer. Each neuron in a layer is connected to every neuron in the subsequent layer. The output layer has a neuron for each target variable. The fully connected nature of a dense layer allows the model to adapt to different output dimensionalities, and explore complex patterns and relationships within the data. The Neural Network was initially used to predict 3 output variables:

- *cumulative_total_deaths_parties_in_orgvio_cy* that represents the cumulative estimate of the fatalities for the parties involved in organized violence within the border of a country for a given country-year.
- *cumulative_total_deaths_civilians_in_orgvio_cy* that represents the cumulative estimate of the civilian fatalities in organized violence within the borders of a country for a given country-year.
- *cumulative_total_deaths_in_orgvio_best_cy* that represents the best estimate of the total fatalities in organized violence within the borders of a country for a given country-year.

After some elementary testing, the model showed better performance predicting only 1 output variable at a time, namely *cumulative_total_deaths_in_orgvio_best_cy*. The architecture and training were therefore adapted for a single neuron output.

The model was trained to minimize the Mean Absolute Error (MAE) loss function of the training set with varying learning rates, data subsets, batch sizes, dropout rates and architectures (number of layers, activation functions, output activation function). The Adam optimizer was chosen for its computational efficiency. Annex 1 summarizes the parameter selection process. To further optimize the parameters, hyper-parameter tuning using grid-search and 3-fold cross-validation was performed. However, given computational constraints, the matrix of parameters explored remained far from exhaustive.

H. Linear Regression model - LASSO

A simple linear regression model was built to compare its performance with the complex Neural Network model. A Least Absolute Shrinkage and Selection Operator (LASSO) regularization technique was also implemented to prevent multicollinearity issues (for instance, features *os_govt_inv_cy* that represent the involvement of the main Government in one-sided violence in a given country-year and *os_any_govt_inv_cy* that represents the involvement of any Government in one sided violence in a given country-year were unsurprisingly extremely correlated).

I. Results comparison

First, the performance of a Naïve model, that always predicts the previous year's fatalities, was computed. Then, all the models were tested on the validation set and their performance metrics were compared. The first comparison was made between the training set MAE and the validation set MAE to detect any signs of overfitting.
Subsequently, the top-performing models were further tested on two separate test sets. First, the MAE of each model was calculated for the entire test set, covering the years 2015 to 2022. Additionally, the MAE was computed for a subset of the test set consisting solely of the year 2015, to assess whether the models performed better in short-term predictions. Finally, the test set MAE was compared to the Naïve model.

IV. Dataset description

A. Main Dataset

The main dataset used in this project is the UCDP Country-Year Dataset on Organized Violence within Country Borders Version 23.1 [12]. It aggregates information on all conflicts globally between 1989 and 2022 in *.xlsx* format and required a negligible amount of cleaning. This data set contains our target variables (fatalities of parties involved, civilian, and total best estimate) and features such as *sb_intrastate_govt_inv_incomp_cy* (dummy variable equal to 1 if the government of the country is involved in intrastate violence within the country within the year), or *sb_interstate_dyad_names_cy* (the names of the state-based actors engaging in interstate violence within the country within the year), among many others.

The information contained in the main data set was enriched by other UCDP data sets. The categorical feature *incompatibility*, that represents the reason behind a state-based conflict, was imported from the UCDP Battle-related Deaths Dataset Version 23.1 [2] via the dyad ID. Similarly, the categorical feature *org*, that represent the organizational level of non-state actors, was imported from the UCDP Non-state Conflict Dataset Version 23.1 [13] via the dyad ID.

B. External Dataset

The first external dataset is the Education Statistics from the World Bank Public Data Catalog [14]. It contains over 4000 features that describe various aspects about education attainment across all countries but contains enormous amounts of missing values. To clean the data set, rows and columns containing mostly missing values were dropped, then, all the variables that contained the word "Total" were removed to only keep the disaggregated data. After pivoting the data set to have each indicator for each country per rows, columns containing more than 50% missing values were dropped again.

The second external dataset is the GDP data from the World Bank open data bank [15]. It contains the GDP in current USD for all countries. Given the typical growth pattern of GDP, any missing values were backfilled. Countries that were not included in the initial data set were added, and the value of their GDP was set as a copy of the country with the closest historical GDP [16]. North Korea was a special case; the values were manually entered and backfilled based on TradingEconomics estimations [17]. The data was then log-transformed to counteract heavy imbalances.

C. Resulting Dataset

After all the information were merged, the resulting dataset is initially of shape 6419x19, namely 6419 observations of 16 features and 3 target variables. However, after removing all rows in which no type of organized violence took place within the country-year (for example, Switzerland 2021), the resulting data set is of size 1895x19.

V. Implementation

A. Data Importation and Cleaning

The implementation of the project faced some difficulties during data importation and cleaning. The Education dataset had to be kept in .zip file format given its voluminous nature. The *zipfile* package was used to access and import the data. The *fuzzywuzzy* package was used to match country names across the different datasets, with an initial 70% similarity threshold. Subsequent matches were done manually through dictionaries, taking the UCDP main data set as the basis.

B. PCA, Clustering, and Data split

The PCA, Clustering and training-validation-test sets separation were made using the *sklearn* package. The validation set and test set were split with random state parameter set to 42 for reproducibility.

C. Modelling, Tuning and Testing

The Neural network was built using *Keras*' open-source library through *Tensorflow*. From *Keras*, the Adam optimizer was also used. The *sklearn* package was used for the LASSO regression and for the GridSearchCV function. The initial Hyperparameter Tuning, estimating just under 10'000 different models, revealed to be too computationally intensive. The Hyperparameter tuned model is therefore not fully optimized.

VI. Results

A. Clustering

The first results to analyze are the clustering of economical and educational attributes for each country.

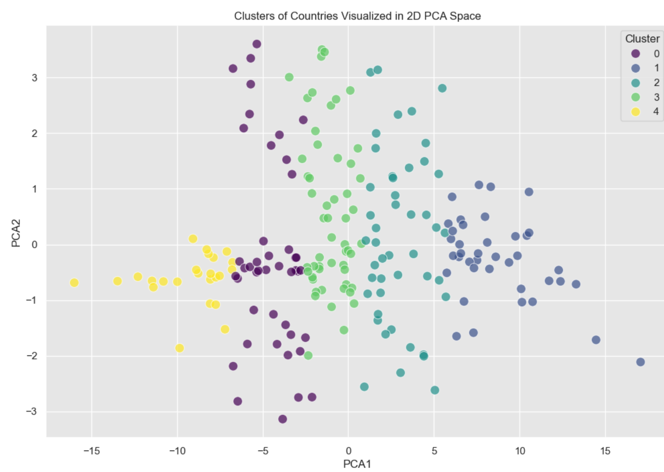


Fig. 1: 2D Visualization of clusters

1) Cluster 0 contains countries such as Liechtenstein, Djibouti, Monaco or Togo. It seems to be characterized by low GDP countries, with varying levels of educational attainment.
2) Cluster 1 contains countries such as Japan, Switzerland or the United States. It seems to be characterized by high GDP countries with high levels of educational attainment.
3) Cluster 2 contains countries such as Egypt, Ukraine or New Zealand. It seems to be characterized by medium to high GDP with medium to high levels of educational attainment.
4) Cluster 3 contains countries such as Haiti, Botswana of Bolivia. It seems to characterized by low GDP and low levels of educational attainment.
5) Cluster 4 contains micro-islands such as Tuvalu, the Comoros or Saint Kitts and Nevis.

B. Predictive models

In this section, only the results of the best performing models are discussed. A summary of the main models tested is available in Annex 1. However, please note that the analysis of the performance metrics displayed in Annex 1 should be done carefully, given the different architectures, data subsets, loss functions, and scaling applied.

The three models compared are:

1) The Sequential Neural Network with manually set parameters
2) The Hyperparameter-tuned Sequential Neural Network
3) The LASSO Regression

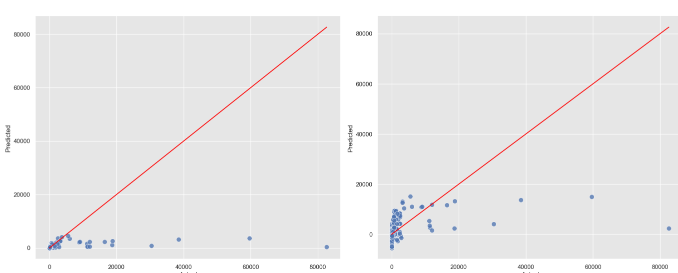
All three models were trained on the years 1989 to 2014, and their performance predicting either 2015-2022, or only 2015 was assessed. Note that the test set for 2015 is remarkably small, which can bias the results. Table 1 summarizes the performance of each model.

TABLE I: Model performance - Mean Absolute Error

Models Naive MAE = 1117	Training MAE 1984-2014	Test MAE 2015-2022	Test MAE 2015
Hand-tuned NN	1506	1643	1119
Hyperparameter-tuned NN	1512	1652	1121
LASSO Regression	2377	2377	1007

The first observation is that both Neural Network models have extremely similar despite their different architectures. Both Neural Network may demonstrate signs of overfitting, indicated by the lower Mean Absolute Error (MAE) on the training data compared to the 2015-2022 test data. In contrast, the LASSO Regression model demonstrates robust generalization, evidenced by consistent MAE scores across both training and test datasets. All 3 models perform better on the 2015 test set in comparison to both the training and 2015-2022 test sets, with the LASSO Regression model achieving the overall lowest MAE. While the significant differences in MAE between the 2015-2022 test set and the 2015 test set might hint at improved short-term forecasting accuracy, it is more probable that this gap is due to the limited size of the 2015 test set, that fails to generalize the target variable distribution in the data.

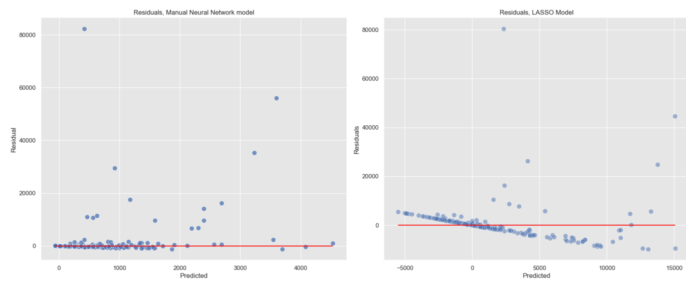
C. Predictive models



(a) Neural Network (b) LASSO Regression

Fig. 2: Actual vs. Predicted

Looking at Figure 2a, the hand-tuned Neural Network seems to have found precise patterns in low to very-low fatalities, with a considerable number of Predicted values landing on the 45-degree line between Actual and Predicted values. However, the model was not capable of generalizing its understanding for high-fatalities conflicts, constantly predicting low fatalities even for conflicts causing over 80'000 fatalities in a country-year. On the other hand, the LASSO model in Figure 2b tends to favor predictions of 0 fatalities, and finds a weak positive trend towards higher values.



(a) Neural Network (b) LASSO Regression

Fig. 3: Residuals

Looking at the Residuals of both models, the patterns observed in the previous paragraph are further confirmed. The hand-tuned Neural Network model in Figure 3a only predicts in 0-4'600 range, which causes considerable residuals in conflicts with higher death-tolls. The LASSO regression in Figure 3b predicts in the -5'000-15'000 range, with a slope indicating a poor understanding of the data.

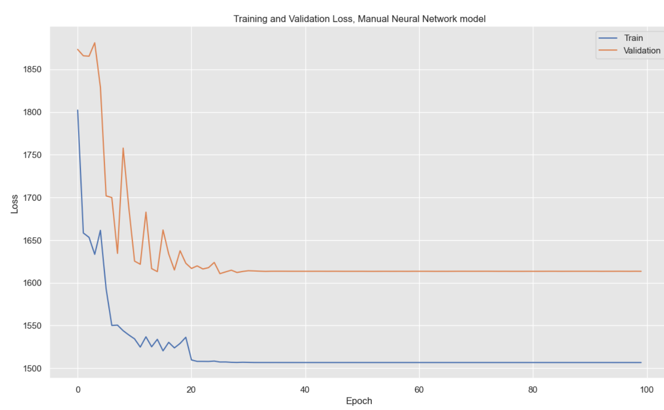


Fig. 4: Training and Validation Loss, Neural Network

Figure 4 displays the Training and Validation loss of the hand-tuned Neural Network. The training loss indicates rapid learning during the first 10 epochs, followed by 10 epochs of slower learnings. The validation loss sees an unstable phase between epochs 10 and 20. This can be linked to the overfitting issues highlighted previously, or could hint at disparities between the training and validation sets.

In the context of this research, the findings outlined in this section fail to offer practical real-world applications.Even the lowest Mean Absolute Error recorded (MAE = 1007), only performs about 10% better than the Naïve model (MAE = 1117). The magnitude of this discrepancy in anticipated fatalities renders it ineffective for informing meaningful decision-making by policymakers and peacebuilders. The results are however promising, and future, more in-depth research should be encouraged.

VII. Conclusion

This study set out to develop a predictive model for fatalities in violent conflicts to provide policymakers and peacebuilders with a tool to better anticipate and respond to emerging threats, ultimately contributing to conflict prevention and peacebuilding efforts. The UCDP Dataset on Organized Violence per country-year was used to train a Neural Network and a LASSO regression model. While the former showed some capability in predicting fatalities for low to very-low conflict scenarios despite some signs of overfitting, it struggled to generalize its understanding for medium to high fatality conflicts. Conversely, the LASSO regression model exhibited robust generalization across the training and test datasets but tended to favor predictions of zero fatalities, thereby lacking precision for more severe conflicts. Despite the rigorous methodology, the resulting models therefore failed to achieve the level of accuracy necessary for practical applications because of the substantial discrepancies in anticipated fatalities.

Moving forward, further research should focus on incorporating additional features and increasing the number of observations fed to the model, for example, by adopting a country-month approach rather than a country-year one. Moreover, developing separate models for low-fatality conflicts and high-fatality conflicts could significantly enhance forecasting accuracy. However, it is necessary to consider if the application of these models still makes sense in the context of real-world decision-making and their intended use for conflict prevention and peacebuilding.

References

[1] Schwarz, Oliver, Axt, Heinz-Jürgen, Milososki, Antonio. (2006). Conflict – a literature review.

[2] Davies, Shawn, Therese Pettersson, Magnus Öberg (2023). Organized violence 1989-2022, and the return of conflict between states. Journal of Peace Research 60(4). Therese (2023) UCDP Battle-related Deaths Dataset Codebook v 23.1 (https://ucdp.uu.se/downloads/).

[3] Sarkees, Meredith Reid and Frank Wayman (2010). Resort to War: 1816 – 2007. Washington DC: CQ Press. (https://correlatesofwar.org/data-sets/cow-war/)

[4] Weisi Guo. Predicting conflict - a year in advance (https://www.turing.ac.uk/about-us/impact/predicting-conflict-year-advance)

[5] V. D'Orazio, J. Honaker, R. Prasady and M. Shoemate, "Modeling and Forecasting Armed Conflict: AutoML with Human-Guided Machine Learning," 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 4714-4723, doi: 10.1109/BigData47090.2019.9005963.

[6] Podojil, Brandon Michael. The fog of war: predicting the incidence of global conflict. Diss. 2020. (https://repositories.lib.utexas.edu/items/a997e396-d9c7-4c67-81dc-a25158139c94)

[7] Mueller, Hannes and Christopher Rauh (2022), The Hard Problem of Prediction for Conflict Prevention. Journal of the European Economic Association, Volume 20, Issue 6, December, pp. 2440–246. Mueller, Hannes and Christopher Rauh (2022), Using Past Violence and Current News to Predict Changes in Violence. International Interactions, Volume 48, Issue 4, pp. 579-596. Mueller, Hannes and Christopher Rauh (2018), Reading Between the Lines: Prediction of Political Violence Using Newspaper Text. American Political Science Review, Volume 112, Issue 2, May, pp. 358 - 375. (https://conflictforecast.org/about)

[8] Fearon, James D, and David D Laitin. 2003. "Ethnicity, Insurgency, and Civil War." The American Political Science Review. Vol. 97.

[9] Barakat, Bilal, and Henrik Urdal. 2009. "Breaking the Waves? Does Education Mediate the Relationship Between Youth Bulges and Political Violence?" 5114. Policy Research Working Paper. https://doi.org/doi:10.1596/1813-9450-5114.

[10] Hegre, H., Allansson, M., Basedau, M., Colaresi, M., Croicu, M., Fjelde, H., Hoyles, F., Hultman, L., Högbladh, S., Jansen, R., Mouhleb, N., Muhammad, S. A., Nilsson, D., Nygård, H. M., Olafsdottir, G., Petrova, K., Randahl, D., Rød, E. G., Schneider, G., Von Uexkull, N., Vestby, J. (2019). "ViEWS: A political violence early-warning system". Journal of Peace Research, 56(2), 155-174. https://doi.org/10.1177/0022343319823860 (https://viewsforecasting.org/about/early-warning-system/)

[11] Raleigh, C., Kishi, R. Linke, A. Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices. Humanit Soc Sci Commun 10, 74 (2023). https://doi.org/10.1057/s41599-023-01559-4

[12] • Davies, Shawn, Therese Pettersson Magnus Öberg (2023). Organized violence 1989-2022 and the return of conflicts between states?. Journal of Peace Research 60(4). • Sundberg, Ralph and Erik Melander (2013) Introducing the UCDP Georeferenced Event Dataset. Journal of Peace Research 50(4).

[13] Sundberg, Ralph, Kristine Eck Joakim Kreutz (2012). Introducing the UCDP Non-State Conflict Dataset. Journal of Peace Research 49(2):351-362. Davies, Shawn, Therese Pettersson Magnus Öberg (2023). Organized violence 1989- 2022, and the return of conflict between states. Journal of Peace Research 60(4). Codebook: Pettersson, Therese (2023) UCDP Non-state Conflict Codebook v 23.1 (https://ucdp.uu.se/downloads/).

[14] "Education Statistics — Data Catalog," datacatalog.worldbank.org. https://datacatalog.worldbank.org/search/dataset/0038480/education-statistics

[15] World Bank, "GDP (current USD)," The World Bank, 2023. https://data.worldbank.org/indicator/NY.GDP.MKTP.CD

[16] Wikipedia Contributors, "List of countries by GDP (nominal)," Wikipedia, Jan. 28, 2019.

[17] Trading Economics, "North Korea GDP," Tradingeconomics.com, Sep. 11, 2019. https://tradingeconomics.com/north-korea/gdp

Appendix

The following tools were used for the elaboration of this report:

- OpenAI Chat-GPT: For the coding process and proof-reading
- GitHub Copilot: For the coding process
- GitHub Repository: For version control

Attached to this report:

- Annex 1: *Models summary.xlsx* - Overview of the main models tested