The Report Committee for Brandon Michael Podojil
Certifies that this is the approved version of the following Report:

The Fog of War:

Predicting the Incidence of Global Conflict

APPROVED BY
SUPERVISING COMMITTEE:

Kenneth S Flamm, Supervisor

Michael G Findley

The Fog of War:

Predicting the Incidence of Global Conflict

by

Brandon Michael Podojil

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Public Affairs

The University of Texas at Austin

May 2020

Dedication

I dedicate this work to my daughter, Emerson. That she may live a world less afflicted by the horrors of war.

Acknowledgments

This research project has taken a considerable amount of my attention, particularly as I balance my responsibilities as a father and husband. For bearing far more than her share of the burden, I would like to acknowledge my wife, Kristin, for her unwavering support and devotion to our family.

Additionally, I would like to acknowledge my faculty advisors Dr. Kenneth Flamm and Dr. Michael Findley, for the significant effort they took to shape my initial research efforts into the more presentable report before you. Thanks also go to other members of the faculty at the LBJ School at UT-Austin, including Dr. Varun Rai for initially exposing me to the potential of machine learning and Dr. Ji Ma for his advice on coding and data management for this project.

Finally, I would like to thank the group of colleagues with whom I initially explored this line of research in support of a class project: Laura Sigelmann, Jaynish Vaghela, Sai Vijay Deeraj Krishna Kaki, and Achmad Z Khomaini. This report builds on the strong foundation and vision for the project we established together.

Abstract

The Fog of War:

Predicting the Incidence of Global Conflict

Brandon Michael Podojil, MPAff

The University of Texas at Austin, 2020

Supervisor: Kenneth S Flamm

Machine learning has revolutionized approaches to predicting the outcomes of various phenomena. The following examines the applicability of several machine learning techniques for forecasting conflict around the globe. The models are assessed against multiple specifications of variables, in their ability to predict specific forms of conflict, and their performance over a span of ten years. In the end, some of the produced models achieve relatively high levels of accuracy and recall, though none appear particularly useful for decision-makers as they seek to effectively and efficiently allocate scarce conflict prevention and mitigation resources.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

The true causes of armed conflict have long eluded researchers. In contrast, the consequences of war are stark, visceral, and real. Whereas the majority of conflict research has sought to determine what factors cause conflict, the present research contributes to a smaller, but growing field committed to predicting the outbreak of future conflict. While the former line of research is undoubtedly critical to enabling policymakers and social scientists to understand the best policy options to prevent conflict, the latter is just as critical to these leaders and thinkers as it has the potential to allow them to most efficiently and effectively allocate the scarce resources available to prevent and mitigate the effects of conflict.

This dichotomy between inferential (or causal) research approaches and a predictive approach to this field of research stems from the characteristics, strengths, and weakness of two larger statistical approaches. Traditional econometric techniques, such as those based on linear regression, are better suited to inferential studies. These approaches generally focus on producing interpretable coefficient estimates, ideally with clear causal linkages, that allow readers to understand the relationship between an independent and dependent variable clearly. The weakness of linear regression and other inferential techniques lies in their inflexibility. This relative inflexibility stems from the fact that researchers are imposing a function (generally a linear or logarithmic function) on the model. This restriction produces models with weaker predictive performance on new data than many machine learning algorithms, including those employed for this report.

On the other hand, predictive research introduces more flexible parametric or even non-parametric algorithms to build models that improve model performance against previously "unseen" data. The cost of this higher performance is degraded interpretability of any particular coefficient. In the same way that the results of a logistic regression are more challenging to communicate to readers than a simple linear regression, highly flexible and non-parametric approaches will not provide readers with clear causal linkages or clearly interpretable coefficients.

This current research effort attempts to answer the following question: What machine learning methods produce models that are most effective at predicting the incidence of conflict?[1] To address this question, the report first provides an overview of the relevant literature for both the inferential and predictive study of conflict. Next, it reviews the data and methodologies employed in this report. The report then describes the incremental progression employed to create and refine the models. First, I examined their ability to predict conflict one year in the future. I then reframed the dataset to evaluate model performance in predicting three distinct types of conflict – state-based, nonstate, and one-sided. Finally, I explored the models' ability to predict conflict at subsequently further points in time. The report also identifies several areas that I either did not address or only briefly considered, which deserve more attention in future research. Finally, it concludes with the predictions the model makes for conflict out to 2028 and considers if and how these models can help world leaders to prevent and mitigate future conflict.

---

[1] This research effort substantially builds upon I conducted with my colleagues, Laura Sigelmann, Jaynish Vaghela, Sai Vijay Deeraj Krishna Kaki, and Achmad Z Khomaini in the spring of 2019. This initial work found promising results which I have continued to refine. The original research employs a similar empirical strategy as that of this report including the approach for identifying relevant variables, the data imputation method, and the employment of algorithms with cross-validation. In this revisit, I have included new variables for consideration and removed some that raised concerns upon reevaluation. I have also expanded the number of algorithms employed against the data, included an evaluation by three specific types of conflict, and examined algorithm performance across longer forecasting timelines. The original report can be accessed by contacting the author at brandon.podojil@utexas.edu.

LITERATURE REVIEW

The following literature review provides a brief introduction to the two bodies of conflict literature. The first section examines the wide array of inferential research. From this first section, I arrived at a simple model describing the general indicators of potential conflict, which I used to inform my variable collection strategy. The second section surveys the literature that addresses previous conflict prediction efforts. In this section, I attempted to place the current report within the greater corpus and highlight trends observed by scholars in the field.

Indicators of Conflict:[2]

The study of conflict has resulted in an expansive body of work examining its causes and consequences. This work includes efforts to both build interpretable models aimed at describing the conditions which lead to conflict and to predict its outbreak. While predictive work often takes an inductive approach rather than deductive one, logic suggests that the most efficient and effective predictive variables are likely those that theory relates to the outcome of interest. With this consideration in mind, a literature review led me to a hypothesis that one can best predict conflict by examining governance, economic, demographic, and geographic variables as well as considering the effect of previous conflict. The simplified linear model below captures the overarching trends discovered in the literature and which drove both our original and my refined data collection strategy.

$$Conflict = \beta_i * Econ_i + \beta_j * Gov_{+} \beta_k * Geo_i + \beta_l * Demo_l + \beta_m * ConLag_m$$

[2] The paragraphs in the literature review section on governance, geographic, economic, demographic, and predictive indicators draw heavily from previous work by my colleague Laura Sigelmann for our research in the spring of 2019.

In this model, conflict is a function of economic (Econ), governance (Gov), geographic (Geo), demographic (Demo), and previous, or lagged, conflict trends (ConLag). The following provides a brief overview of the literature that led us to our general model as well as a survey of some researchers' efforts to predict conflict previously.

The nature of a country's governance likely plays a determinant role in the level of peace enjoyed there. Some authors argue that weak institutions contribute to a higher incidence of conflict as the state cannot provide sufficient means to assuage or deter a dissatisfied citizenry(Fearon and Laitin 2003). This hypothesis aligns with a USAID finding that state fragility is correlated with heightened insecurity (J. W. Busby et al. 2018). Specifically, USAID identifies the competitiveness of participation, quality of public service, and citizen participation as critical to strong institutions. Political violence scholars also argue that political efficacy, militant political activism, and political repression likely contribute to the incidence of conflict (Post et al. 2016). Additionally, democratic peace theorists have long asserted that democratic governments are less likely to engage in conflict (Hayes 2011).

Geographic factors have been studied substantially as contributing factors to conflict, but the history and validity of these hypotheses are highly controversial. While theories such as environmental determinism have largely been dismissed, certain geographic factors, in context, are likely contributors to conflict. The presence or lack of natural wealth is often cited as a contributor to potential conflict. This phenomenon, known as the resource curse, is commonly studied in the literature, though the findings present mixed results (O'Brochta 2019). Michael Ross found that of all-natural resources, large endowments of petroleum led to a higher incidence of conflict (Ross 2015). Conversely, Percival and Homer-Dixon find that environmental scarcity, rather than abundance, leads to violent conflict (Percival and Homer-Dixon 1996). The field of environmental security is still contentious, and general agreement has yet to be reached in the field (J. Busby 2018). Nonetheless, factors such as rugged terrain and dense tree cover could lead to tactical advantages or centralized security challenges and are therefore considered in this research.

Economic variables are also generally considered pivotal in understanding conflict. Fearon and Laitin found that countries that relied on primary commodities for their economies are under-bureaucratized for their GDP level, leading to weaker state institutions since rulers require fewer means to acquire revenue from the population (Fearon and Laitin 2003). Collier, Hoeffler, and Söderbom argue that this connection is due to the increased ability to finance a rebellion with the use of commodities (Collier, Hoeffler, and Söderbom 2004). Some studies have argued that poverty and poor economic prospects reduce the opportunity cost of joining a rebellion and facilitate rebel recruitment (Post et al. 2016). There appears to be a strong link between urban food prices and riots, highlighting the critical role of economics in civil unrest (Arezki, Rabah Brückner 2011). Finally, Collier found that countries with a high percentage of GDP composed of exports are more likely to experience conflict (Collier 2000). Whether this is due to the volatility of the economy or its correlation to low per capita incomes is unclear.

Demographic factors such as educational attainment and employment rate are apparently crucial in determining conflict risk. Hegre et al. used primarily demographic variables in their conflict model, largely neglecting governance and geographic variables (Hegre et al. 2011). Raleigh and Hegre found that population size increases conflict risks, but others found it does not affect the conflict's duration (Raleigh and Hegre 2009; Collier, Hoeffler, and Söderbom 2004). Education also seems to play a key factor. Barakat and Urdal found that higher levels of primary education and higher male literacy were associated with lower conflict incidence (Barakat and Urdal 2009). Various proxies for development have also been used to predict conflict. One such measure, infant mortality, was found to be strongly associated with the onset of armed conflict (Urdal 2005; Abouharb and Kimball 2007). Hegre et al. found that poverty was a significant contributor to conflict predictions (Hegre et al. 2011). Ethnic cleavages are a contentious variable, and research findings are inconclusive (Fearon and Laitin 2003; Collier and Hoeffler 2000). Regardless, data on country-level ethnicity or ethnic cleavages proved to be nonexistent, and we chose to leave this variable out.

In addition to examining the inferential studies on conflict, the review of relevant literature included a study of several predictive conflict assessments. The vast majority employ logistic regressions and find performance ratings that are on par with those in this report. The most compelling research uncovered was that of Hegre et al. which predicted that the current declining conflict trends would continue through 2010-2050 (Hegre et al. 2011). They predict that internal armed conflict will continue to decline to around 7% in 2050, particularly in Western Asia and North Africa. Conflict will increasingly be concentrated in East, Central, and Southern Africa, and East and South Asia. Interestingly, some studies considered the effect of previous conflict on future conflict. Hegre et al. considered previous conflict to be an endogenous variable, since past conflict is likely to affect other conflict risk factors, especially average income. This "conflict trap" may not be due to the past conflict itself, but rather to the effects conflict has on other predictors.

Conflict Prediction Literature:

A growing body of literature has started the process of examining the potential for machine learning algorithms to predict the incidence of conflict. The general trend over time has been an increasing use of more sophisticated techniques and more granular datasets to develop more specific predictions. In his 2017 article, *Conflict Forecasting and Its Limits*, Thomas Chadefaux provides an overview of predictive literature and its trends.

He observes the fact that the majority of predictive research has relied on relatively simple algorithms, like logistic regression, and data typically aggregated at the country-year level (Chadefaux 2017, 4). The author identifies that efforts to increase temporal specificity have shown promising results, while spatial specificity has been less successful (Chadefaux 2017, 5). The author concludes his review by exploring the theoretical limits to predicting conflict, given its uncertain nature.

A more recent article, by Valeria Helle, Andra-Stefania Negus, and Jakob Nyberg, takes an approach that most closely resembles that taken here. Their efforts focus on improving and refining a wide array of machine learning algorithms at a subnational-monthly level for the continent of Africa. Their approach finds that the random forest algorithm works best on their data and significantly improve model performance over traditionally used algorithms like logistic regression (Helle, Negus, and Nyberg 2018, 27).

One final article worthy of consideration, by Cederman and Weidmann, reviews the various efforts researchers have taken to predict conflict (Cederman and Weidmann 2017). Following this review, the authors consider the various limitations of this approach to research and ways to improve studies such as improving data management and employing baseline or naïve models to assess performance (both of which are implemented here). Like Chadefaux, the authors conclude that there is likely some limit to the potential for conflict prediction models due to the inherent nature of conflict and the fact that "historical 'accidents' often make a mockery of decontextualized out-of-sample extrapolation." (Cederman and Weidmann 2017, 2)

As mentioned previously, my research most closely resembles that of Helle and her co-authors, though several key points of differentiation remain. Whereas Helle et al. consider data at the subnational and monthly level, I remain at the country-year level. The key reason for this difference is that I intend to consider the entirety of global conflict rather than focusing on a particular continent or region. This scope better aligns with my intent for the research to inform a global response effort. Furthermore, Helle et al. look do not differentiate between types of conflict, nor do they look beyond a one-year time horizon. On the other hand, I explore model performance across both of these characteristics. I argue that given the different areas that their paper and mine focus on, the combination of the two papers provides a more thorough examination for the potential applicability of machine learning models to predicting conflict than any one alone.

DATA DESCRIPTION AND SUMMARY STATISTICS

The survey of relevant literature led me to conduct a broad search for data in line with the broad theoretical framework outlined in the previous section. Our literature review identified the five societal aspects that may contribute to conflict, economics, governance, geography, demographics, and previous conflict trends. Informed by this, I restricted variable selection to data that are relevant to these aspects but erred on the side of more inclusion than exclusion for variables within those aspects. This approach resulted in the inclusion of a large number of variables in the specification.

Data for this report span from 1989 to 2018 and were gathered from four different sources, the Uppsala Conflict Data Program, the Polity IV Project from the Center for Systemic Peace, the World Development Indicators by the World Bank, and the Country Ruggedness and Other Geographic Characteristics of Countries Dataset hosted by Diego Puga. Table 1 of the appendix outlines each variable name, description, and source.

The compiled data results in a panel dataset with observations at the country-year level. As referenced in the literature review, some papers have considered data at a more granular level, either by smaller periods of time or smaller geographic areas. Given that this research intends to determine if machine learning models can help leaders allocate resources to mitigate total global conflict, it follows that only variables that assess the entirety (or at least majority) of the world should be included. Therefore, this report maintains an observational unit of country-year since data for measurements that cover the world are almost exclusively reported in country-year terms. To be sure, a potential avenue for future research includes exploring the possibility of creating models for regions with rich datasets at smaller geographic scales or with more frequency, but that is beyond the scope of the current effort.

The initial number of observations in the panel dataset is 6363, representing data from 216 countries tracked by the World Bank for up to 30 years (based on when the World Bank identifies the establishment of each country). The total number of independent variables initially collected is 89.

From these 89, I selected 41 variables for inclusion in the final specification based on three criteria. First, I removed all variables missing more than 80% of the values. Second, I removed variables that had perfect collinearity. For example, I start with both urban and rural population percentage. I removed the rural population percentage as it is equal to 100-urban population percentage and thus perfectly collinear. Finally, I examined collinearity and removed one of any pairs variables with large correlations. I chose the variable to remove from these pairs based on which variable had more missing values in the dataset. A visualization of the missing data for all of the variables and for the final set of variables included in the base specification can be found in Figures 1 and 3 of the appendix, respectively. The correlation matrices for all variables and those included in the base specification are illustrated respectively in Figures 2 and 4 of the appendix.

Uppsala Conflict Data Program Variables:

Conflict data come from the Uppsala Conflict Data Program's Georeferenced Event Dataset (GED) Global Version 19.1 (Stina 2019). This dataset spans from 1989 to 2018 and served as the binding constraint on how far back in time the overall dataset could reach. UCDP distinguishes between three types of conflict: state-based, non-state, and one-sided. UCDP defines state-based conflict as "a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in a calendar year" (Stina 2019, 28). UCDP defines non-state conflict as "the use of armed force between two organized armed groups, neither of which is the government of a state, which results in at least 25 battle-related deaths in a year" (Stina 2019, 29). Finally, UCDP defines one-sided conflict as "the use of armed force by the government of a state or by a formally organized group against civilians which results in at least 25 deaths. Extrajudicial killings in custody are excluded" (Stina 2019, 30).

The source dataset is designed with each unique conflict event serving as the observational unit with relevant information about the event listed as variables. Importantly, this includes the year, country, and conflict type for each event. For this report's purposes, I aggregated the event data to create eight variables. The first four variables are indicators for conflict at the country-year level, one variable for any instance of conflict, and one variable for each of the three types of conflict. The second four variables calculate the sum of total conflict events listed by Uppsala for each particular country-year. Given the nature of the dataset, country-year combinations that did not have conflict present are unlisted. Therefore, country-year observations which are not found in the source dataset considered these to be zero, or no conflict, in the constructed variables.

The conflict incidence indicator represents whether or not conflict occurred in a given country-year, and the sum of conflict events provides a sense of the conflict's intensity. In order to gain a sense for the duration of a conflict, I also created a variable, *con_3last,* which sums the incidence of conflict from the observed year and the two previous resulting in a data range from 0 to 3 where the former represents three years without conflict, and the latter represents three years with conflict.

Finally, all of the dependent variables come from the Uppsala dataset as well. To achieve this, I led the conflict incidence variable by one year (i.e., each observation includes the next year's conflict incidence value). Thus, in each row of the dataset, the independent variables represent the values in that particular year and the led dependent variable represents the incidence of conflict in the next year. To create dependent variables to build models that predict each form of conflict, I applied a similar method to lead the conflict incidence variable for each of the three types of conflict. Finally, in order to assess model performance over time, I built data sets in which the conflict incidence variable is led by two, three, five, and ten years.

One somewhat distracting omission from the Uppsala dataset is data on Syria. Instead of including Syria in the global dataset, Uppsala has created a separate dataset covering the Syrian conflict, but this data only spans 2016-2019. In this same time period, very little reliable data from the World Bank or Polity IV project are available. For that reason, I decided to drop Syria from this research, though future efforts should include seeking ways to incorporate reliable data on this crucial missing country.

World Development Indicators:

I gathered the vast majority of the data for this research from the World Development Indicators (The World Bank 2019). For an overview of which specific variables came from this source, refer to Table 1 of the appendix. Data from the WDI span as far back in some cases as 1960, but given the constraints of the UCDP dataset, I restricted my collection of World Bank data to span from 1989 to 2018. Data from this dataset are reported at the country-year level and required no aggregation or munging for inclusion in the overall dataset. As an additional note, in some specifications, I included fixed effects. This adjustment incorporated indicator variables for each country, region, and year represented in the dataset according to the World Development Indicators.

Furthermore, some disparity exists regarding which countries were included in the different datasets. Omissions primarily consist of small island nations, territories with varying levels of sovereignty, and city-states. For consistency, I used the countries listed in the World Development Indicators as the basis for determining countries for inclusion and, thus each instance of a country-year. Countries listed in the World Bank data, but not in other tables are missing those relevant values. Countries not listed in the World Bank data, but that are listed in other tables, are not considered in this report. Table 2 of the appendix illustrates which datasets included which countries.

The decision to use the World Bank list of countries as the standard list stemmed from two factors, the relative prominence of the World Bank as an institution compared to my other data sources and the fact that the majority of my data comes from the World Bank. I argue that this decision lends more authority to my dataset and minimizes the amount of missing data compared to using other sources' country lists as the standard list.

Polity IV Project Variables:

The purpose of the Polity IV project is to assess the "authority characteristics of states in the world system for purposes of comparative, quantitative analysis" (Marshall, Gurr, and Jaggers 2019, 1). For this research project, I used the Polity IV Project dataset to assess various governance features not available from World Bank Estimates. Data for the Polity IV project are organized based on country-year already, and I gathered five variables for evaluation: *polity 2, durable, parcomp, corrupt, and rule_law*. The *polity 2* variable estimates the relative level of autocratic and democratic institutions in a particular country-year observation. I selected *polity 2* over the original polity variable as it rectifies some coding issues, namely the use of -66, -77, and -88 to identify special conditions in specific entries. -66 entries are "cases of foreign interruption" and recoded to missing values (Marshall, Gurr, and Jaggers 2019, 17). -77 entries are "cases of interregnum or anarchy" which they code as zeros (Marshall, Gurr, and Jaggers 2019, 17). -88 entries are "cases of transition" which are corrected based on a prorating of the transition values across the span of the transition (Marshall, Gurr, and Jaggers 2019, 17). These same qualitative codes are not corrected for in the variable *parcomp*, which measures the degree to which alternative parties or stakeholders can pursue their objectives in the government. To correct for this, I employed the same criteria in re-classifying those entries manually after data retrieval. *Durable*, which measures regime longevity in years, required no editing for inclusion, nor did *corruption* or *rule_law*.

Terrain Ruggedness and Other Geographic Data:

The dataset offered by Diego Puga presents the best data I could find to assess geographic features at the country level (Puga and Nunn 2012). The dataset includes three measures of terrain ruggedness, total land area, percent coverage of fertile soil, desert, and tropical climate, measures of how much of a country's terrain is near a coast, and a measure of diamond extraction from 1958 to 2000. A limitation of this dataset is that it is in a cross-sectional format rather than a panel one. It does not provide yearly measures to assess changes to these variables across time. Therefore, each country-year annotation for a given country will share the same values for the variables from this dataset. While measures such as terrain ruggedness are likely to have changed little from 1989 to 2018, other measures, such as a county's desert coverage, have great potential to have changed. A more applicable dataset for the future would include more granular measures of terrain characteristics at the country-year level, especially given the potential for climate change to impact the incidence of conflict.

Summary Statistics

The summary statistics for each variable are presented in the appendix in two tables. First, Table 3 provides the summary statistics for all of the collected variables before imputation. Table 4 provides the summary statistics for all of the collected variables after imputation. A notable observation from the summary statistics is the confirmation that each count of the variables is complete at 6363 total observations per variable in the following table.

Additionally, an examination of the mean for each of the conflict incidence variables provides us with an understanding of how frequently conflict occurs in the data. Overall, conflict occurs in roughly 25% of observations. State-based and one-sided conflict occurs in 18% of observations, and nonstate conflict only occurs in 10% of observations. These prevalence measures suggest significant class imbalances between observations with and without conflict that I address in the research approach section of this report.

A final observation demonstrates that the imputation process resulted in several instances where impossible values were input for missing data points. These values were all negative when the minimum value is actually zero. Variables affected include population percentage with cell phone access, population percentage with internet access, and the prevalence of severe wasting in children under five. Notably, the prevalence of severe wasting minimum value was only -0.01. Therefore, though the value is impossible actually to achieve, it is likely that the regression estimate for this value is not statistically different from zero. The other two values, access to the internet and to cell phones, represent relatively new phenomena and may highlight a weakness with my imputation method at estimating values that are significantly beyond the timespan of data used for the estimations. In the end, none of these variables were used in the final specification of variables. Nonetheless, these errors highlight the challenge of implementing an effective imputation technique. I discuss the imputation methodology used in this paper and make recommendations for future research in the research approach section of this report.

RESEARCH APPROACH

Besides understanding the data used to build the models in this project, one must also understand the general research approach employed. The following section outlines several decisions made and methodologies used, including the imputation methodology for missing values, which evaluation metrics I used to assess the models, naïve assessments I created to establish performance baselines, my selection of algorithms, and how I addressed dimension reduction, class imbalance, and overfitting.

Imputation Methodology

One of the most pressing challenges facing this project is the presence of a significant number of missing values. As mentioned previously, Figures 1 and 3 of the appendix illustrate the number of missing values from the total collected dataset and the variables included in the base specification, respectively. As one can see, a significant number of the data points are missing in the dataset.

A complete dataset is required to run most machine learning algorithms. Ensuring a complete dataset can be achieved via two methods, one can either drop variables or observations with missing values, or one can impute the missing values. The former option is the most viable in inferential statistics, where researchers are concerned with creating accurate and precise coefficients without the use of cross-validation. However, if this methodology were employed here, we would reduce the dataset to such a small size as to render it useless. The second option, imputation is less problematic for prediction research so long as all of the dependent variable outcomes are known since cross-validation will evaluate our model’s ability to predict the outcomes in a test set of the data. To be sure, the quality of our imputation method can improve or degrade our models’ performances, but those are captured in our evaluation metrics without issue.

Various python packages are available for the imputation of missing values (Pedregosa et al. 2020, chap. 6.4). Despite this, I opt to employ an original approach to imputing the missing values in my dataset for two reasons. First, many of the imputation methods are overly simplistic, such as merely inputting the mean or median value of the dataset for missing values. This approach is troublesome because of the panel nature of the data. The mean or median values would fail to capture the country and time trends in the data. Second, many of the more sophisticated packages assume that values are missing at random. Given the nature of development data, I do not assess this assumption to hold. Missing values are likely to follow one of several patterns. Older observations could be missing more values as measurement techniques improve over time. Otherwise, observations from countries in conflict could be more likely to be missing due to degraded access to these areas. Finally, observations from the developed world could be more likely to be missing as the majority of time and resources are spent on developing countries where the need for understanding is most urgent. For all of these potential reasons, I am skeptical of assuming that data are missing at random.

An outline of the methodology is presented in Figure 5 of the appendix. The methodology began by subsetting a particular indicator by country. At this point, three possibilities existed: no missing values, partial missing values, or missing all values. If no values were missing, no imputation was required, and the algorithm moved to the next country. If part of the country's data was missing, the algorithm replaced the missing values with the mean of a linear regression and a decision tree regression.[3] For indicators missing an entire country's data points, the algorithm expanded the subset to include the entire

[3] My justification for using these two estimation techniques is that the former presents a simple parametric approach, and the latter is a simple non-parametric approach. The inclusion of more estimation techniques or more sophisticated techniques increases the computational price for this algorithm. Future research should certainly consider this decision and whether other options result in higher performing models.

region to which the country belongs and imputed missing value with the mean of the linear and decision tree regressions. Finally, in the case of an entire region missing data for a particular indicator, the algorithm imputed the missing value as 60% of the global mean.

The decision to employ the weighted global mean for values in which the entire region is missing is based on an untested assumption that these regions were more likely to be underdeveloped. Upon further evaluation, this assumption does not seem to hold. The only region that has any missing values for an entire variable is North America, which by the World Bank’s classification includes Bermuda, Canada, and the United States. This omission supports the previous assessment that more developed regions may receive less focus than developing regions for many of these metrics. Unfortunately, this discovery does not support the assumption underlying the algorithm, as it is unlikely that values from these regions would be lower than the global mean. The variables for which North America is missing all values are literacy rates, population percentage living in slums, poverty headcount, and internally displaced persons due to conflict. None of these variables was included in the final specification as they were missing significant data points throughout the entire dataset. Therefore, though using the 60% weighted global mean to impute missing values for an entire region appears problematic, it did not impact the data used to create any of the models.

As noted in the data description section, this imputation methodology resulted in some distracting values that would not be possible (negative values where the lowest possible value is zero). Inaccurate values are an inherent challenge for any imputation methodology. As previously stated, future research should consider other approaches to imputation and evaluate if the algorithm performance metrics are robust to these other methods.

Evaluation Metrics

A standard metric used for machine learning model evaluation is the accuracy of the model. Accuracy is calculated as:

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Observations}$$

Scores for this metric (as well as the others presented later in this section) range between 1, in the case of a perfectly accurate model, and 0, for a model that incorrectly classifies each observation. This metric is not the most appropriate for this research due to the significant class imbalance between the observations with conflict and those without it. Additionally, conflicted areas often remain in conflict for an extended period. As I will demonstrate with naïve assessments, simply predicting that a country will be in conflict in a given year based on whether or not it was in conflict in the previous year yields an accuracy rating of 0.92. This high accuracy rating occurs despite failing to identify the first year for each new conflict, arguably one of the most important purposes of this project.

The issue with the accuracy metric is that it equally weights false positives and false negatives. In this case, a false positive is an instance where the model predicts a conflict, yet it does not occur. A false negative is an instance where the model does not predict a conflict, but it occurs regardless.

There are two reasons why this report focuses on reducing false negatives, both of which stem from the motivation to empower decision-makers to deploy resources to prevent conflict. First, one would hope that interventions reduce the chance that a conflict actually occurs, therefore as a practical matter, one would expect that more effective interventions would yield a higher number of false positives (or circumvented conflicts). Second, given the consequences of false positives (inefficiently deployed resources) and those of false negatives (greater human loss and suffering), a focus on the latter seems like a higher moral imperative.

Several alternative metrics to accuracy exist for one to evaluate machine learning models. The three that are relevant to this report are recall, precision, and the f1 score. Recall is calculated as:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

In this case, the numerator is the total number of conflicts that the algorithm correctly predicted, and the denominator is all actual instances of conflict. Therefore, the higher the number of false negatives, the lower the recall score. Given its focus on the correct prediction rate for actual instances of conflict, I used the recall score as the primary measure of performance for the models.

One concern with using recall is that a high rate of false positives does not reduce the model's recall score. A model that readily predicts conflict may have a high recall score but also result in a high false-positive rate. For example, if the model predicted that every country would be in conflict, it would have a perfect recall score, but such a high false positive rate as to render it useless. Therefore, though we stated that our primary focus is the model that minimizes false negatives, if too many false positives result, then the model may prove unusable.

The other two metrics mentioned above, precision and f1 score, provide a means to counter the secondary concern of too high a false positive rate. Precision provides an assessment of the rate at which the model predicts true positives compared to all positive predictions. It is calculated as:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

This measure is valuable in assessing false positives but has the inverse issue of recall in that it does not consider false negatives in its calculation. Here the f1 score proves its utility by calculating the harmonic mean of the precision and recall scores. The f1 score is calculated as:

$$f1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

As one can confirm from the formula, the f1 score equally considers the importance of reducing false positives and false negatives. Including this measure provides a useful hedge for my model assessment to ensure against too high a false positive rate.

To summarize, I primarily focused on each model's recall score for my evaluation in order to reduce the total number of false negatives or instances where conflict is not predicted yet occurs. As a secondary measure, I considered the f1 score (which incorporates model precision) to ensure that a high recall score does not come at the expense of too many false positives, which would render the model useless to decision-makers. Finally, I presented the accuracy of each model for the interested reader and to allow for comparison to other research, though I do not consider it when evaluating the models.

Naïve Assessments

In addition to identifying the metrics for model evaluation, naïve models can help researchers to establish baselines against which they can compare models. For this research, I selected two naïve methods to establish these baselines.

The first method simply predicted that no conflict would occur at all. Given the distribution of observations with and without conflict, the accuracy of this assessment is high at 0.71. This result established a clear floor for my models' performance, but also highlights the issue of using accuracy as the primary performance measure. The other two metrics from this naïve model are not very useful since there are no instances of true positives. The recall score is 0 (as is the precision score) since the number of true positives forms the numerator of the score, and the f1 score is undefined since the denominator is equal to the sum of the precision and recall scores.

The second naïve method to predict conflict incidence simply predicts the incidence of conflict in the current year based on the incidence of conflict in the previous year. (i.e., if a given country were in conflict in 1990, then this method would predict that it will be in conflict in 1991). This model achieves an accuracy rating of 0.92, again demonstrating the issue of using accuracy as the primary performance metric. The recall and f1 scores are both 0.83. Since the recall score is the target metric for this research, I consider the 0.83 score to be the floor for model performance in this metric. The f1 score of 0.83 is not necessarily a floor since models that perform better on recall may do so at the expense of precision. Instead, I only consider the 0.83 f1 score as context for evaluating the models. F1 scores lower than 0.83 become even more distracting, especially as recall scores increase, since this requires a corresponding drop in precision to achieve.

Previewing the results, I found recall rates for some models in the low 0.9's with an f1 score in the low 0.8's or even high 0.7's. Achieving this spread between the two scores means that the false positive rate must be in the low 0.6's or even the high 0.5's. Models with such poor precision, but high recall scores are likely just predicting large numbers of countries in conflict, rendering the model useless.

Algorithms Employed

In keeping with the recommendations identified in the literature review, I included a wide array of algorithms for consideration. The selected algorithms span an array of machine learning approaches as well as relative complexity. The strategy of testing a variety of algorithms allows researchers to observe which models tend to perform better for a given set of variables. Different models may perform better on a particular problem due to characteristics of the data that one approach may be more effective at identifying than other methods. The following section briefly describes each algorithm used.

The first algorithm employed was logistic regression (LR) (Pedregosa et al. 2020, chap. 1.1.11). This method is a standard econometric approach for evaluating a binomial dependent variable that fits a logistic function to the data in a way that reduces the sum of squared error.

The following algorithm used was K nearest neighbors (KNN) (Pedregosa et al. 2020, chap. 1.6). This algorithm predicts the class of an observation by examining the “k” number of the points nearest to it. In deciding which classification to predict, the algorithm selects the majority case from the “k” neighboring observations.

Third, I used a linear discriminant analysis algorithm (LDA) (Pedregosa et al. 2020, chap. 1.2). LDA seeks to create a linear hyperplane that best divides the various observations in order to minimize the proportion of each class distribution that is on the wrong side of the hyperplane. New observations are then classified based on where they lie in relation to the established hyperplane.

Similar to LDA, support vector classifiers (SVC) also seek to create a hyperplane in the data. In this case, though, the algorithm looks specifically at the observations which lie closest to the seam between the two classes. The SVC model attempts to create a hyperplane, which is the greatest maximum distance from these marginal observations.

Finally, this report considers three variations of classification trees: decision trees (DTREE) (Pedregosa et al. 2020, chap. 1.10), random forest trees (RF) (Pedregosa et al. 2020, chap. 3.2.4.3.1), and extreme gradient boosted random forest (XTREE) (Chen 2020). Each of these trees builds upon the previous. Decision trees use binary splits in the variables of the data in order to arrive at a classification. Random forests reduce the likelihood of overfitting on a particular training set by creating a large number of decision trees based on a bootstrapped subset of the training data. New observations are then classified based on the modal classification it receives from the numerous models. Finally, XTREE introduces the concept of regularization. Unlike random forest, which creates many independent trees, XTREE creates one tree at a time. Each iteration of the tree builds upon and refines the previous one in order to improve classification performance. The rate of this gradual build-up is controlled by the regularization term. Smaller regularization terms slow the rate of adjustment per iteration, reducing the change of overfitting while larger regularization terms allow the algorithm to more rapidly adjust based on new observations.

Dimension Reduction

To mitigate against the potential for issues with high dimensionality, I employed principal components analysis (PCA) (Pedregosa et al. 2020, chap. 2.5.1) as a means to reduce the number of dimensions in the various specifications. This methodology creates a series of orthogonal vectors, each of which explains the greatest possible share of the variance in a dataset as a replacement for the actual vector of independent variables. For this research, I used an 80% of the explained variance threshold to determine the number of principal components for all of the PCA iterations. This corresponded with roughly 20 components for the base specification and 190 components for the specifications with year, country, and regional fixed effects.

Class Imbalance

In the summary statistics, I found that conflict is a relatively rare occurrence accounting for roughly 25% of the observations. The numbers were even smaller when I specified each type of conflict. State-based and one-sided conflicts occurred in 18% of observations, and non-state conflict occurred in 10% of observations. Researchers should address class imbalances as they can introduce bias and degrade the performance of machine learning algorithms (Abraham and Elrahman 2013).

To correct for this, I apply the synthetic minority over-sampling technique or SMOTE (Lemaitre, Nogueira, and Aridas 2017). This technique oversamples from the minority class (in this case, observations with conflict) by creating synthetic observations of the minority class (Chawla et al. 2002). By oversampling, I achieve a better balance of the two classes, conflict and no conflict.

Overfitting

Overfitting represents an ever-present concern for any machine learning model. Overfitting results when the algorithm fits very well to the presented data, but poorly predicts new observations. To reduce this possibility, researchers employ cross-validation where the data are randomly split into a training set and a test set. The model is fit to the training set only while the test set is used to evaluate model performance. While some researchers opt to use only one train/test split, others use a technique called k-fold cross-validation, where the data is split into 'k' number of sections. In k-fold cross-validation, one of these sections at a time is used as the test set, and the other 'k'-1 sections are used as the training set. The algorithm iterates the model through 'k' times using each separate section as the test set in one iteration. Performance metrics are the average measures for each of these iterations (Pedregosa et al. 2020, chap. 3.1). For this research, I selected to use k-fold cross-validation with a 'k' value of five. As with all other hyperparameters, future research should include a sensitivity analysis to determine if this number is optimal or if other k values result in better-performing models.

MODEL PROGRESSION

The process of tuning a machine learning model balances the line between art and science. Though cross-validation helps to ensure the reliability of performance metrics, outcomes can be highly contingent on the decisions and assumptions one makes in creating the models. In this report, I considered a series of potential areas where the models could be improved and incrementally introduced them to evaluate changes in performance. My progression begins by creating the baseline models which predict any incidence of conflict one year out. I next examined model performance on the three specific types of conflict, and finally evaluating the models' performance over greater temporal distance. I also briefly considered the potential for neural networks and hyperparameter optimization. Each of these steps and the resulting performance are outlined in greater detail through the rest of this section.

Creation of the Base Specification

I first created models that predicted conflict one year in advance. I trained models against both a standard specification as well as one with a PCA of the variables. Additionally, I add two more variations to the base specification by including country, region, and year fixed effects to the variables and conducting a PCA analysis of the base specification with fixed effects included. In total, I created 28 created models. The results for these models can be found in Table 5 of the appendix.

For each model, the recall score improved with the introduction of fixed effects. The average model recall score without fixed effects was 0.852 and average performance with fixed effects was 0.875. Considering the variable sets that included fixed effects, the model with the best average recall score was KNN (0.940), but this was accompanied by the second-worst f1 score average (0.776). The next best average recall score came from the SVC models (0.886). These models were also accompanied by the best average f1 score (0.831) of the base specifications. Therefore, I concluded that the best performance came from the SVC model of the variable sets that included fixed effects.

Prediction Based on Type of Conflict

As mentioned in the data section, the Uppsala dataset classifies each entry into one of three types of conflict: state-based, nonstate, and one-sided. I consider the potential for the algorithms to predict the incidence of each type of conflict in the same manner that I considered overall conflict. Subdividing conflicts by type exacerbates the already existing class imbalance associated. Therefore, I hypothesized that predicting each type of conflict would be less effective than predicting overall conflict since fewer observations exist of each particular conflict type than conflict in general.

Tables 6, 7, 87 of the appendix report the evaluation metrics for each algorithm by conflict type. Generally, recall scores remained relatively high. Recall scores average 0.876, 0.856, and 0.864 for state-based, nonstate, and one-sided conflict, respectively. However, f1 scores reduced significantly to an average of 0.769, 0.627, and 0.741 for state-based, nonstate, and one-sided, respectively. This disparity suggests that these models are resulting in more false positives than the general conflict models. Furthermore, no model stands out as a definite high performer compared to the others, and none of these models appears to be particularly applicable for decisionmakers given the number of false positives that they produce.

Prediction Over Time

In addition to understanding how the models perform against the various forms of conflict, one may also wonder how well the models perform at predicting conflict further into the future. To assess algorithm performance over time, I employed conflict led variables in the base specification dataset to create models that predict conflict 1, 2, 3, 5, and 10 years out. Tables 9, 10, 11, and 12 and Figures 6, 7, 8, and 9 of the appendix illustrate the results. Prior to running the models, I expected a significant degredation in performance over time. My logic was that events and conditions from the previous year are far more likely to influence current political conditions (including conflict) than those from a decade ago. While model performance generally degrades over time, it does not do so by as much as one might expect. I argue that this is primarily due to the persistent nature of conflict. Countries that are in conflict are generally more disposed to remaining in conflict in the future, as suggested by the performance discovered in the second naïve assessment previously.

As with the other examinations, here I discovered that the average performance of the models with fixed effects was greater than those without across the four. At ten years out, models with fixed effects achieved an average recall score of 0.867 compared to an average recall score of 0.836 for the models without fixed effects. Again, as with the other examinations of model performance, KNN results in the highest recall rate, but a poor f1 score and SVC provides the next best recall score and the highest f1 score. Therefore, I concluded that, as with the base iteration, the SVC models on the variable specifications with fixed effects strike the best balance of relatively high recall and f1 scores over time.

Introduction of Neural networks

In addition to the seven algorithms that I used to create the models for this report, I decided to briefly consider an emerging technique for machine learning that has recently garnered significant attention, the neural network. This methodology creates a series of nodes and connections resembling the neurons of a brain, hence the algorithm’s name.

Given the enthusiasm for the potential of neural networks, I wanted to include a brief exploration of this method. I used the Keras python package to implement a relatively simple neural network against the base dataset in order to assess its performance (Chollet 2015). The neural network that I created consists of dense connections between four layers: input, output, and two hidden layers consisting of 20 nodes each. An illustration of the neural network can be seen in Figure 10 of the appendix. The resulting recall, f1, and accuracy scores are listed in Table 13 of the appendix. The average recall score across the four variable variants was 0.815, well below the average performance of the other models at 0.863.

Admittedly, my approach to developing the neural network represents only an initial analysis of its potential. Given the relative performance of the algorithm, future research and optimization seem warranted to assess if higher-performing models can be developed that may improve the predictive power of a neural network.

Hyperparameter Optimization

This research focused on the robustness of model performance in a range of ways, including the variables used in the dataset, across the types of conflict, and over time. One area of potential weakness in this research, though, is the fact that I kept the default hyperparameter (or tuning parameter) settings for each model throughout the assessment. As stated by the authors Philipp Probst, Anne-Laure Boulesteix and Bernd Bischl, "In contrast to direct, first-level model parameters, which are determined during training, these second-level tuning parameters often have to be carefully optimized to achieve maximal performance (Probst, Boulesteix, and Bischl 2019, 1).

To address this issue, I conducted some limited hyperparameter optimization exploration for the support vector machine algorithm in the base specification with fixed effects. This consisting of testing several tuning parameters across a range of values individually while holding the other tuning parameters at their default values. While non-default values demonstrated some initial potential, none of the models with 'optimized' hyperparameters performed better than the default models when cross-validated across the entire dataset. Admittedly, this approach to optimization was simplistic. The most thorough technique would be to consider every possible combination of each hyperparameter in order to identify the optimum value, but with current technology, this task is computationally expensive, if not impossible. That said, more sophisticated techniques exist to address some of the challenges associated with hyperparameter optimization practically (Claesen and De Moor 2015), and I leave that to future research efforts.

Areas for Future Refinement

In addition to the areas and features I considered in revisiting and refining my conflict prediction models, several other aspects are ripe for future consideration. As I have mentioned previously, I conducted only limited exploration of hyperparameter optimization with the current set of models and found no performance improvement. Further efforts could be undertaken to implement more sophisticated methods of searching for optimized hyperparameters to seek higher performing models. This attention should especially focus on the SVC and KNN algorithms, given their consistently high recall performance. Additionally, due to the relatively unsophisticated manner in which I considered them here, neural network algorithms are also deserving of more examination.

Areas that I did not at all address in this report include evaluating the models' sensitivity to the use of other imputation methods. Additionally, our model employed SMOTE to address oversampling. Several other techniques exist for addressing class imbalance and future research could examine model sensitivity here as well (Lemaitre, Nogueira, and Aridas 2017, chap. 2).

Application of Models

As a final exercise, I wanted to examine the potential for practical application of the models. To do so, I created datasets that used all available data points to train and fit models that predicted conflict beyond the years in which data are available. In line with the time horizons I explored previously, the datasets include data up to 2018 and predict conflict in 2019, 2020, 2021, 2023, and 2028 (1,2,3,5, and 10 years into the future).

I selected the variable specification with fixed effects, but without PCA as it performed best of all of the specifications when I examined performance over time with an average recall score of 0.870 at ten years into the future. Additionally, given its relatively strong performance throughout the entire process, I selected the SVC algorithm to make my projections.

The results for this forecast are illustrated in Figures 11, 12, 13, 14, and 15 of the appendix. Overall, the model's predictions seem generally plausible (perhaps except for Canada and Western Europe). Despite this, I am concerned neither the maps, nor the algorithms that created them, offer a substantial improvement over current practices to allocate resources and mitigate conflict. Merely asking a scholar, or even a follower of current events, would likely yield a similar set of predictions. With large swathes of the developing world consistently predicted to be in conflict, the models do not appear to be precise enough to assist decision-makers with fine-tuning the allocation strategy for their scarce resources. In order to increase their utility, model refinement must continue to focus on maximizing recall scores, but further attention likely must address the significant presence of false positives and the correspondingly low f1 scores.

Conclusion

The costs exacted by conflict upon society are sobering and daunting. The current research sought to mitigate this reality by providing decision-makers and analysts with highly effective prediction models, which would allow them to allocate conflict prevention and mitigation resources effectively. First, I evaluated a baseline set of models that predicted conflict one year in advance. I further explored the potential for these models by considering how the models performed at predicting three subsets of conflict: state-based, nonstate, and one-sided, and how the models performed over time. Based on this progression, I concluded that the most effective model was built by using the support vector classifier algorithm on the specification of variables I created with the inclusion of fixed effects for country, region, and year. I briefly explored neural networks and a simple hyperparameter optimization technique, but neither effort yielded improved results. I also highlighted several areas for future refinement that I did not consider in this current effort.

Finally, I applied the model to the most recent data available to create model predictions for the years 2019, 2020, 2021, 2023, and 2028. Projecting these predictions onto a map suggested that though the models can highlight areas that are persistently under threat of conflict, the predictions do not provide enough specificity to achieve the goal of empowering decision-makers to allocate resources more efficiently and effectively.

Cederman and Weidmann describe the potential ceiling for our ability to predict future conflict, suggesting that "perhaps the most pernicious problem pertains to the common failure to fully appreciate the fundamental complexity surrounding processes of peace and conflict" (Cederman and Weidmann 2017, 2). Inherently, conflict is among the most complex of human activities. It would seem that this research supports the conclusion that there exist inherent limits to predicting conflict. Despite this, I walk away inspired to continue the search.

APPENDIX

Table 1: Variables Evaluated and Included in the Final Specification.

Name	Description (Unit of Measure)	Source	Included in Final
index	Concatenation of country code and year to create unique entries	Constructed	✓
ccode	Country code according to World Bank	World Bank	✓
year	Year of observation	World Bank	✓
country	Country of observation	World Bank	✓
region	Region according to World Bank	World Bank	✓
con_ev_1	Number of state based conflict events (#)	Uppsala	
con_ev_2	Number of nonstate conflict events (#)	Uppsala	
con_ev_3	Number of one-sided conflict events (#)	Uppsala	
total_ev	Total Conflict Events (#)	Uppsala	✓
con_in_1	Incidence of state based conflict (Binary)	Uppsala	✓
con_in_2	Incidence of nonstate conflict (Binary)	Uppsala	✓
con_in_3	Incidence of one-sided conflict (Binary)	Uppsala	✓
con_in	Incidence of any conflict (Binary)	Uppsala	✓
con_3last	Incidence of any conflict in the last three years (Possible answers are 0,1,2,3)	Uppsala	✓
net_gs	Net trade in goods and services (BoP, current US$)	World Bank	✓
imp_gs	Imports of goods and services (% of GDP)	World Bank	
gdp_cap	GDP per capita (current US$)	World Bank	✓
gdp_cap_gr	GDP per capita growth (annual %)	World Bank	✓
exp_gs	Exports of goods and services (% of GDP)	World Bank	
inf	Inflation (%)	World Bank	✓
out_sch_a	Adolescents out of school (% of lower secondary school age)	World Bank	✓
out_sch_af	Adolescents out of school, female (% of female lower secondary school age)	World Bank	
out_sch_am	Adolescents out of school, male (% of male lower secondary school age)	World Bank	
out_sch_c	Children out of school (% of primary school age)	World Bank	✓
out_sch_cf	Children out of school, female (% of primary school age)	World Bank	
out_sch_cm	Children out of school, male (% of male primary school age)	World Bank	
hs_grad_f	Educational attainment, at least completed upper secondary, population 25+, female (%)	World Bank	
hs_grad_m	Educational attainment, at least completed upper secondary, population 25+, male (%)	World Bank	
hs_grad	Educational attainment, at least completed upper secondary, population 25+, total (%)	World Bank	
gov_ed	Government expenditure on education, total (% of GDP)	World Bank	✓
lit_r_f	Literacy rate, adult female (% of females ages 15 and above)	World Bank	
lit_r_m	Literacy rate, adult male (% of males ages 15 and above)	World Bank	
lit_r	Literacy rate, adult total (% of people ages 15 and above)	World Bank	
acc_tech	Access to clean fuels and technologies for cooking (% of population)	World Bank	
acc_ele	Access to electricity (% of population)	World Bank	✓
acc_ele_r	Access to electricity, rural (% of rural population)	World Bank	
acc_ele_u	Access to electricity, urban (% of urban population)	World Bank	
ag_pct	Agricultural land (% of land area)	World Bank	✓
wtr_wth	Annual freshwater withdrawals, total (% of internal resources)	World Bank	
arable_pct	Arable land (% of land area)	World Bank	✓
forest_pct	Forest area (% of land area)	World Bank	✓
low_land_pct	Land area where elevation is below 5 meters (% of total land area)	World Bank	
wtr_str	Level of water stress: freshwater withdrawal as a proportion of available freshwater resources	World Bank	
min_rent	Mineral rents (% of GDP)	World Bank	✓
nat_gas_rent	Natural gas rents (% of GDP)	World Bank	✓
oil_rent	Oil rents (% of GDP)	World Bank	✓
pop_den	Population density (people per sq. km of land area)	World Bank	✓
pop_den_low	Population living in areas where elevation is below 5 meters (% of total population)	World Bank	
pop_slum	Population living in slums (% of urban population)	World Bank	
pop_rural	Rural population (% of total population)	World Bank	
pop_urb	Urban population (% of total population)	World Bank	✓

Table 1 Continued

Name	Description (Unit of Measure)	Source	Included in Final
death_nonc	Cause of death, by non-communicable diseases (% of total)	World Bank	
death_c	Cause of death, by communicable diseases and maternal, prenatal and nutrition conditions (% of total)	World Bank	
sev_wast	Prevalence of severe wasting, weight for height (% of children under 5)	World Bank	
stunt	Prevalence of stunting, height for age (% of children under 5)	World Bank	
under_n	Prevalence of undernourishment (% of population)	World Bank	✓
under_w	Prevalence of underweight, weight for age (% of children under 5)	World Bank	
pop_grow	Population growth (annual %)	World Bank	✓
price_diesel	Pump price for diesel fuel (US$ per liter)	World Bank	
price_gas	Pump price for gasoline (US$ per liter)	World Bank	✓
phone_fixed	Fixed telephone subscriptions (per 100 people)	World Bank	✓
pop_internet	Individuals using the Internet (% of population)	World Bank	
phone_cell	Mobile cellular subscriptions (per 100 people)	World Bank	
mig_net	Net migration	World Bank	
unemp	Unemployment, total (% of total labor force) (modeled ILO estimate)	World Bank	✓
unemp_m	Unemployment, youth male (% of male labor force ages 15-24)	World Bank	
vul_emp_m	Vulnerable employment, male (% of male employment)	World Bank	✓
lf_rt_f	Female labor force participation rate (% of female labor force)	World Bank	✓
lf_rt_m	Male labor force participation rate (% of male labor force)	World Bank	✓
lf_rt_tot	Labor force participation rate (% of total labor force)	World Bank	
pop_tot	Population, total	World Bank	✓
gini	GINI index (World Bank estimate)	World Bank	✓
pov_hc	Poverty headcount ratio at national poverty lines (% of population)	World Bank	
mil_exp	Military expenditure (% of GDP)	World Bank	✓
idp_conflict	Internally displaced persons, new displacement associated with conflict and violence (number of cases)	World Bank	
idp_disaster	Internally displaced persons, new displacement associated with disasters (number of cases)	World Bank	
af_pers	Armed forces personnel (% of total labor force)	World Bank	✓
rugged	Constructed ruggedness index	Diego Puga	✓
rugged_popw	Terrain ruggedness weighted for population	Diego Puga	
rugged_pc	Terrain ruggedness (% of land that is moderate to highly rugged)	Diego Puga	
land_area	Total land area (1000 Ha)	Diego Puga	✓
soil	Fertile soil (%)	Diego Puga	✓
desert	Desert Coverage (%)	Diego Puga	✓
tropical	Tropical Climate (%)	Diego Puga	✓
near_coast	Portion of country withing 100 Km of ice-free coast (%)	Diego Puga	✓
gemstones	Diamond extraction from 1958-2000 (1000s of carats)	Diego Puga	✓
polity2	Measure of Autocracy/Democracy	Polity IV	✓
durable	Years current government has been in power	Polity IV	✓
parcomp	Measure of competitiveness of election process	Polity IV	✓
corrupt	Measure of corruption	Polity IV	
rule_law	Measure of rule of law	Polity IV	

Figure 1: Missing Values- All Variables

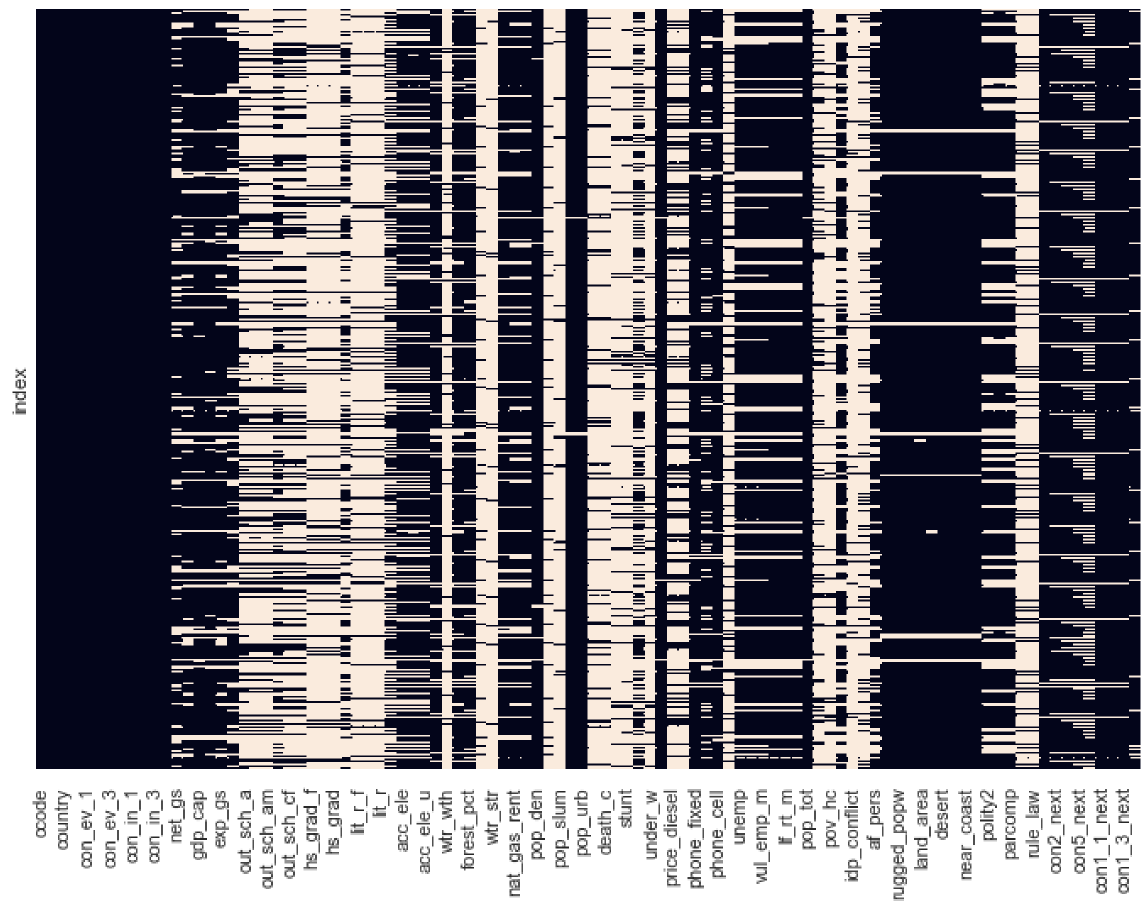


Figure 2: Correlation Matrix- All Variables

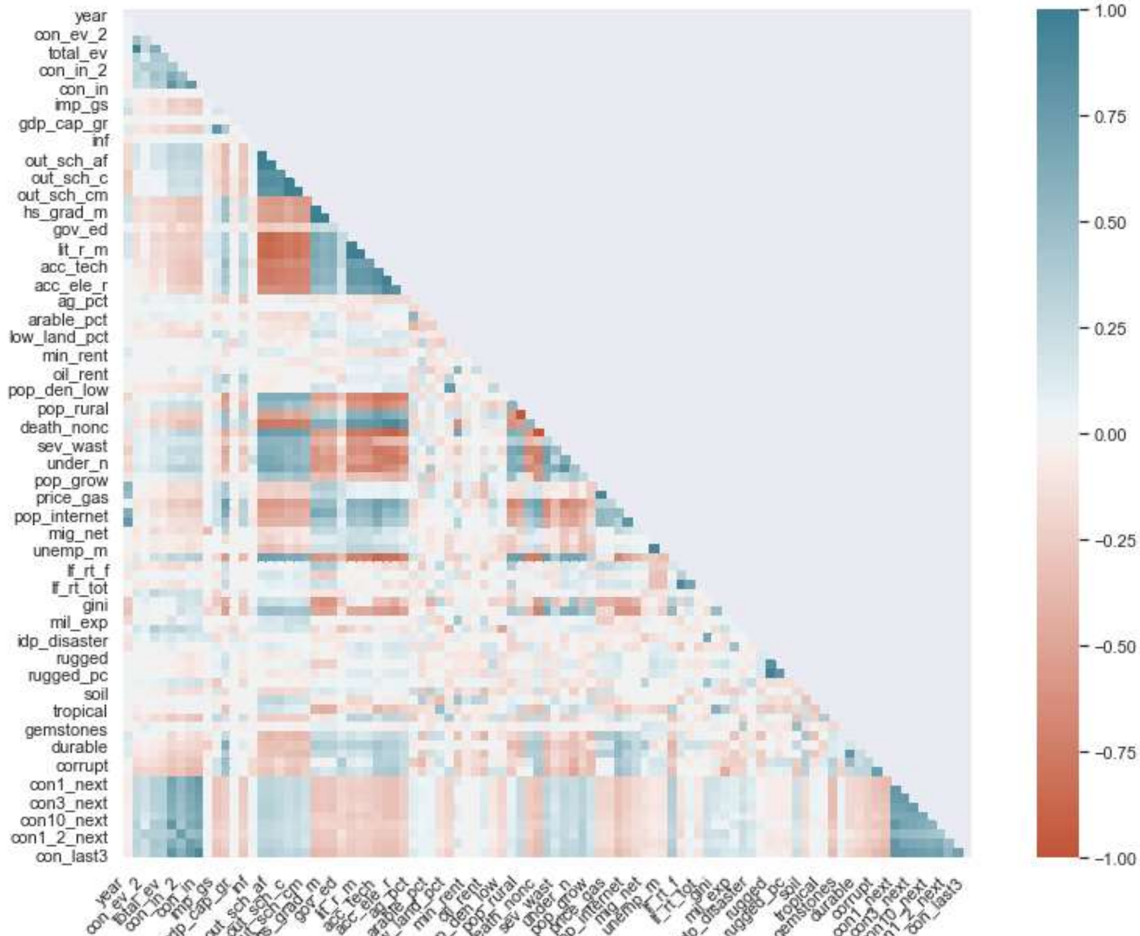


Figure 3: Missing Values- Included Variables Only

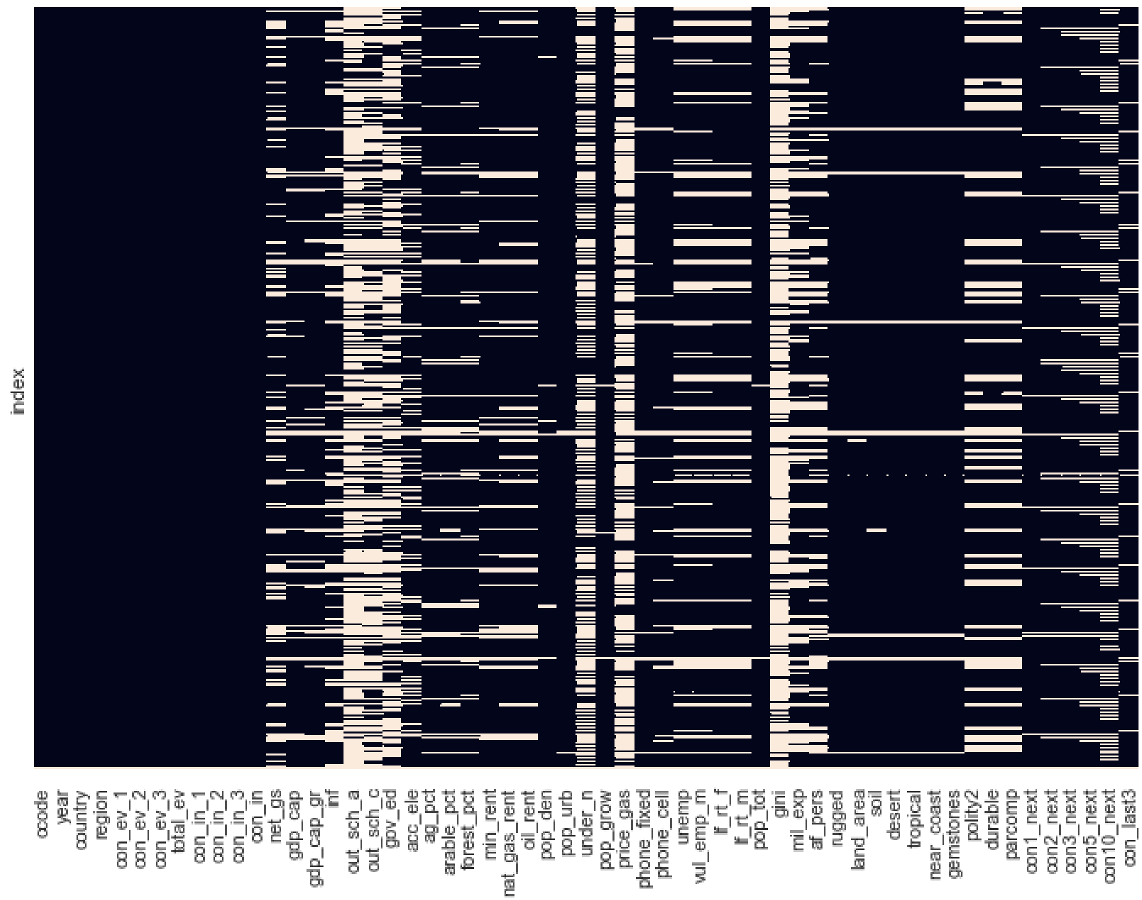


Figure 4: Correlation Matrix- Included Variables Only

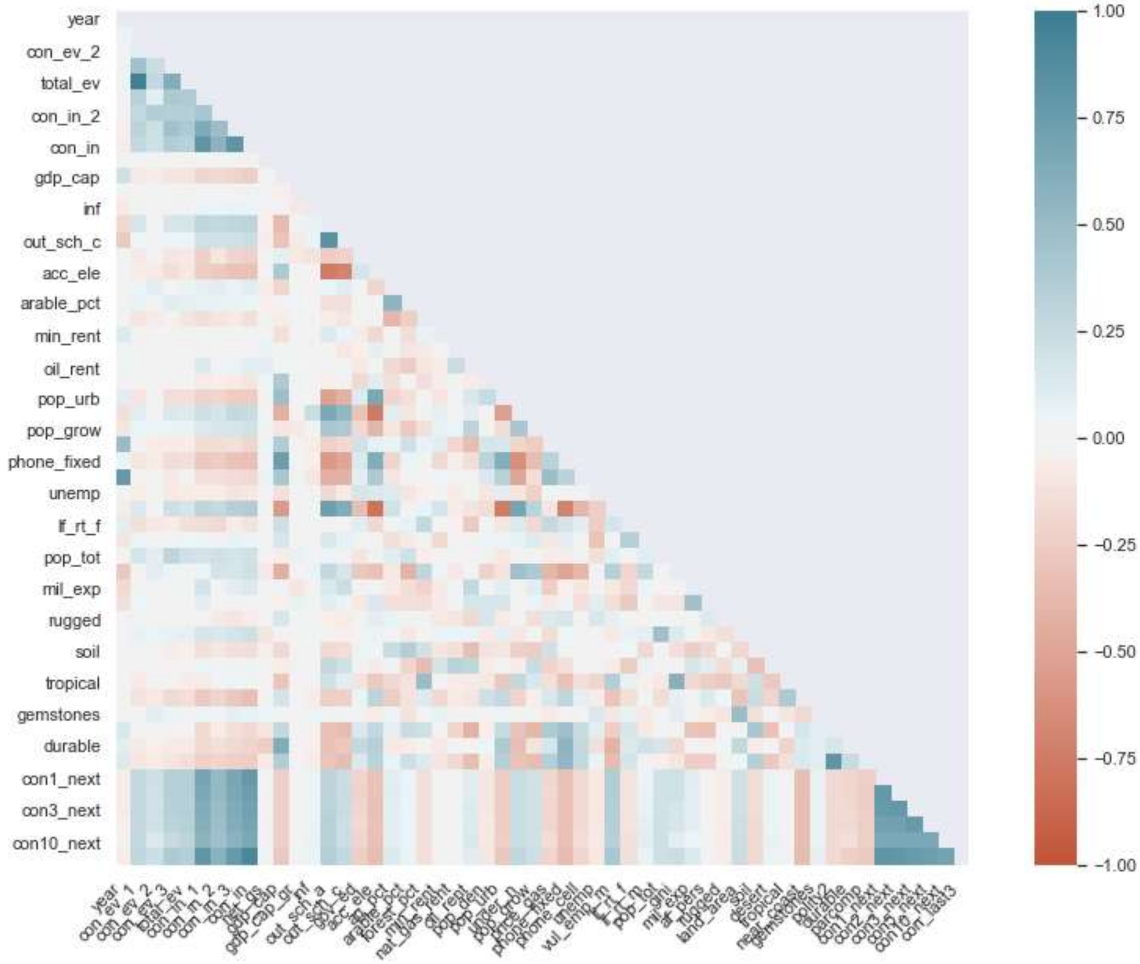


Table 2: Countries Included in Each Dataset

Country	W Bank	Uppsala	Polity IV	Puga
Afghanistan	✓	✓	✓	✓
Albania	✓	✓	✓	✓
Algeria	✓	✓	✓	✓
American Samoa	✓			✓
Andorra	✓			✓
Angola	✓	✓	✓	✓
Anguilla	✓		✓	✓
Argentina	✓	✓	✓	✓
Armenia	✓	✓	✓	✓
Aruba	✓			✓
Australia	✓	✓	✓	✓
Austria	✓		✓	✓
Azerbaijan	✓	✓	✓	✓
Bahamas	✓			✓
Bahrain	✓	✓	✓	✓
Bangladesh	✓	✓	✓	✓
Barbados	✓			✓
Belarus	✓		✓	✓
Belgium	✓	✓	✓	✓
Belize	✓			✓
Benin	✓		✓	✓
Bermuda	✓			✓
Bhutan	✓	✓	✓	✓
Bolivia	✓	✓	✓	✓
Bosnia and Herzegovina	✓	✓	✓	✓
Botswana	✓	✓	✓	✓
Brazil	✓	✓	✓	✓
British Indian Ocean Territory				✓
British Virgin Islands	✓			✓
Brunei Darussalam	✓			✓
Bulgaria	✓		✓	✓
Burkina Faso	✓	✓	✓	✓
Burundi	✓	✓	✓	✓
Cabo Verde	✓		✓	✓
Cambodia	✓	✓	✓	✓
Cameroon	✓	✓	✓	✓
Canada	✓	✓	✓	✓
Cayman Islands	✓			✓
Central African Republic	✓	✓	✓	✓
Chad	✓	✓	✓	✓
Channel Islands	✓			
Chile	✓		✓	✓
China	✓	✓	✓	✓
Christmas Island				✓
Cocos (Keeling) Islands				✓
Colombia	✓	✓	✓	✓
Comoros	✓	✓	✓	✓
Congo, Dem. Rep.	✓	✓	✓	✓
Congo, Rep.	✓	✓	✓	✓
Cook Islands				✓
Costa Rica	✓		✓	✓
Cote d'Ivoire	✓	✓	✓	✓
Croatia	✓	✓	✓	✓
Cuba	✓		✓	✓
Curacao	✓			
Cyprus	✓		✓	✓
Czech Republic	✓		✓	✓
Czechoslovakia			✓	
Denmark	✓		✓	✓
Djibouti	✓	✓	✓	✓
Dominica	✓			✓
Dominican Republic	✓		✓	✓
Ecuador	✓	✓	✓	✓
Egypt, Arab Rep.	✓	✓	✓	✓
El Salvador	✓	✓	✓	✓
Equatorial Guinea	✓		✓	✓
Eritrea	✓	✓	✓	✓
Estonia	✓		✓	✓
Eswatini	✓	✓	✓	✓
Ethiopia	✓	✓	✓	✓
Falkland Islands (Malvinas)				✓
Faroe Islands	✓			✓
Fiji	✓		✓	✓
Finland	✓		✓	✓
France	✓	✓	✓	✓
French Guiana				✓
French Polynesia	✓			✓
Gabon	✓		✓	✓
Gambia, The	✓	✓	✓	✓
Georgia	✓	✓	✓	✓
Germany	✓	✓	✓	✓
Germany West			✓	
Ghana	✓	✓	✓	✓
Gibraltar	✓			✓
Greece	✓		✓	✓
Greenland	✓			✓
Grenada	✓			✓
Guadeloupe				✓
Guam	✓			✓
Guatemala	✓	✓	✓	✓
Guinea	✓	✓	✓	✓
Guinea-Bissau	✓	✓	✓	✓
Guyana	✓	✓	✓	✓
Haiti	✓	✓	✓	✓
Holy See				✓
Honduras	✓	✓	✓	✓
Hong Kong SAR, China	✓			✓
Hungary	✓		✓	✓
Iceland	✓			✓
India	✓	✓	✓	✓
Indonesia	✓	✓	✓	✓
Iran, Islamic Rep.	✓	✓	✓	✓
Iraq	✓	✓	✓	✓
Ireland	✓		✓	✓

*Missing Uppsala entries are assumed to represent countries in which no conflict events have occurred during assessed time period.

Table 2 Continued

Country	W Bank	Uppsala	Polity IV	Puga	Country	W Bank	Uppsala	Polity IV	Puga
Isle of Man	✓				Nicaragua	✓	✓	✓	✓
Israel	✓	✓	✓	✓	Niger	✓	✓	✓	✓
Italy	✓	✓	✓	✓	Nigeria	✓	✓	✓	✓
Jamaica	✓	✓	✓	✓	Niue				✓
Japan	✓		✓	✓	Norfolk Island				✓
Jordan	✓	✓	✓	✓	North Macedonia	✓	✓	✓	✓
Kazakhstan	✓		✓	✓	Northern Mariana Islands	✓			✓
Kenya	✓	✓	✓	✓	Norway	✓		✓	✓
Kiribati	✓			✓	Occupied Palestinian Territory				✓
Korea, Dem. People’s Rep.	✓		✓	✓	Oman	✓		✓	✓
Korea, Rep.	✓		✓	✓	Pakistan	✓	✓	✓	✓
Kosovo	✓		✓		Palau	✓			✓
Kuwait	✓	✓	✓	✓	Panama	✓	✓	✓	✓
Kyrgyz Republic	✓	✓	✓	✓	Papua New Guinea	✓	✓	✓	✓
Lao PDR	✓	✓	✓	✓	Paraguay	✓	✓	✓	✓
Latvia	✓		✓	✓	Peru	✓	✓	✓	✓
Lebanon	✓	✓	✓	✓	Philippines	✓	✓	✓	✓
Lesotho	✓	✓	✓	✓	Pitcairn				✓
Liberia	✓	✓	✓	✓	Poland	✓		✓	✓
Libya	✓	✓	✓	✓	Portugal	✓		✓	✓
Liechtenstein	✓			✓	Puerto Rico	✓			✓
Lithuania	✓		✓	✓	Qatar	✓	✓	✓	✓
Luxembourg	✓		✓	✓	Republic of Moldova				✓
Macao SAR, China	✓			✓	Réunion				✓
Madagascar	✓	✓	✓	✓	Romania	✓	✓	✓	✓
Malawi	✓		✓	✓	Russian Federation	✓	✓	✓	✓
Malaysia	✓	✓	✓	✓	Rwanda	✓	✓	✓	✓
Maldives	✓			✓	St Helena				✓
Mali	✓	✓	✓	✓	St Pierre and Miquelon				✓
Malta	✓	✓		✓	Samoa	✓			✓
Marshall Islands	✓			✓	San Marino	✓			✓
Martinique				✓	Sao Tome and Principe	✓			✓
Mauritania	✓	✓	✓	✓	Saudi Arabia	✓	✓	✓	✓
Mauritius	✓		✓	✓	Senegal	✓	✓	✓	✓
Mayotte				✓	Serbia	✓	✓	✓	
Mexico	✓	✓	✓	✓	Serbia and Montenegro			✓	✓
Micronesia, Fed. Sts.	✓			✓	Seychelles	✓			✓
Moldova	✓	✓	✓		Sierra Leone	✓	✓	✓	✓
Monaco	✓			✓	Singapore	✓		✓	✓
Mongolia	✓		✓	✓	Sint Maarten (Dutch part)	✓			
Montenegro	✓		✓		Slovak Republic	✓		✓	✓
Montserrat				✓	Slovenia	✓		✓	✓
Morocco	✓	✓	✓	✓	Solomon Islands	✓	✓	✓	✓
Mozambique	✓	✓	✓	✓	Somalia	✓	✓	✓	✓
Myanmar	✓	✓	✓	✓	South Africa	✓	✓	✓	✓
Namibia	✓	✓	✓	✓	South Sudan	✓	✓	✓	
Nauru	✓			✓	Spain	✓	✓	✓	✓
Nepal	✓	✓	✓	✓	Sri Lanka	✓	✓	✓	✓
Netherlands	✓	✓	✓	✓	St. Kitts and Nevis	✓			✓
Netherlands Antilles				✓	St. Lucia	✓			✓
New Caledonia	✓			✓	St. Martin (French part)	✓			
New Zealand	✓		✓	✓	St. Vincent and the Grenadines	✓			✓

*Missing Uppsala entries are assumed to represent countries in which no conflict events have occurred during assessed time period.

Table 2 Continued

Country	W Bank	Uppsala	Polity IV	Puga
Sudan	✓	✓	✓	✓
Suriname	✓		✓	✓
Svalbard and Jan Mayen Islands				✓
Sweden	✓	✓	✓	✓
Switzerland	✓		✓	✓
Syrian Arab Republic	✓		✓	✓
Taiwan				✓
Tajikistan	✓	✓	✓	✓
Tanzania	✓	✓	✓	
Thailand	✓	✓	✓	✓
Timor-Leste	✓		✓	✓
Togo	✓	✓	✓	✓
Tokelau				✓
Tonga	✓			✓
Trinidad and Tobago	✓	✓	✓	✓
Tunisia	✓	✓	✓	✓
Turkey	✓	✓	✓	✓
Turkmenistan	✓		✓	✓
Turks and Caicos Islands	✓			✓
Tuvalu	✓			✓
Uganda	✓	✓	✓	✓
Ukraine	✓	✓	✓	✓
United Arab Emirates	✓	✓	✓	✓
United Kingdom	✓	✓	✓	✓
United Republic of Tanzania				✓
United States	✓	✓	✓	✓
United States Minor Outlying Islands				✓
United States Virgin Islands				✓
Uruguay	✓		✓	✓
USSR			✓	
Uzbekistan	✓	✓	✓	✓
Vanuatu	✓			✓
Venezuela	✓	✓	✓	✓
Vietnam	✓		✓	✓
Virgin Islands (U.S.)	✓			
Wallis and Futuna Islands				✓
West Bank and Gaza	✓			
Western Sahara				✓
Yemen	✓	✓	✓	✓
Yugoslavia			✓	
Zambia	✓	✓	✓	✓
Zimbabwe	✓	✓	✓	✓

*Missing Uppsala entries are assumed to represent countries in which no conflict events have occurred during assessed time period.

Figure 5: Outline of Imputation Process

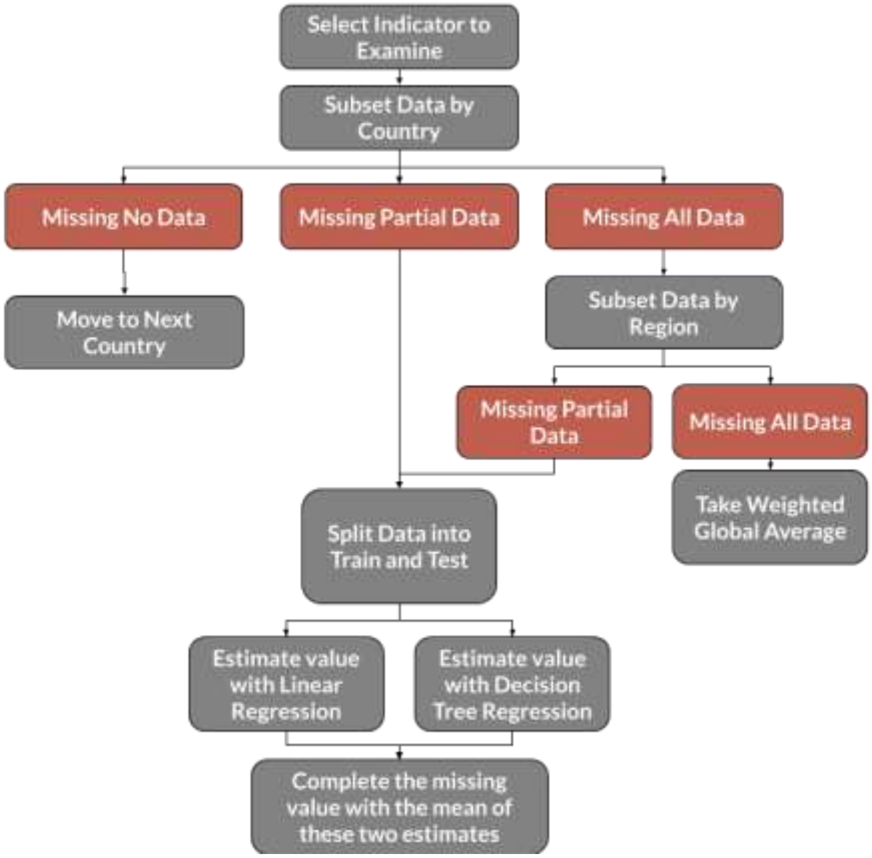


Table 3: Summary Statistics Prior to Imputation

Variable	Preimputation							
	count	mean	std	min	25%	50%	75%	max
year	6363	2003.67	8.61	1989	1996	2004	2011	2018
con_ev_1	6363	16.85	108.63	0	0	0	0	3567
con_ev_2	6363	2.23	18.47	0	0	0	0	639
con_ev_3	6363	4.89	23.29	0	0	0	0	524
total_ev	6363	23.98	124.71	0	0	0	1	3722
con_in_1	6363	0.18	0.39	0	0	0	0	1
con_in_2	6363	0.10	0.30	0	0	0	0	1
con_in_3	6363	0.18	0.38	0	0	0	0	1
con_in	6363	0.25	0.43	0	0	0	1	1
net_gs	4825	9.97E+08	4.30E+10	-7.62E+11	-1.59E+09	-2.11E+08	6.09E+08	3.58E+11
imp_gs	5277	46.86	28.43	0	28.48	41.06	58.07	427.58
gdp_cap	5809	12036.50	19863.55	94.56	991.61	3606.07	15043.12	189170.90
gdp_cap_gr	5769	2.00	6.19	-64.99	-0.12	2.09	4.33	140.37
exp_gs	5277	40.51	29.91	0.01	22.14	33.98	50.52	433.22
inf	4885	28.17	395.78	-18.11	1.86	4.11	8.72	23773.13
out_sch_a	2006	13.65	17.48	0	2.12	6.74	16.92	88.82
out_sch_af	1589	17.24	20.29	0.01	3.86	8.82	21.97	92.66
out_sch_am	1582	15.56	16.83	0.03	3.95	9.30	20.81	85.29
out_sch_c	3056	10.62	15.20	0	1.08	4.18	12.31	80.81
out_sch_cf	2191	14.99	17.98	0.00	2.99	7.62	18.48	86.30
out_sch_cm	2191	13.11	14.98	0.00	2.87	7.20	16.67	77.26
hs_grad_f	934	51.75	24.29	0.16	32.68	54.97	72.60	96.43
hs_grad_m	934	55.43	24.65	0.86	34.60	58.19	77.34	96.18
hs_grad	937	53.42	24.26	0.50	33.29	55.97	75.19	96.31
gov_ed	2790	4.45	1.98	0	3.16	4.30	5.45	44.33
lit_r_f	780	78.27	23.65	4.59	65.34	89.69	94.97	100.00
lit_r_m	781	85.79	16.60	18.26	79.87	93.01	97.48	100.00
lit_r	783	81.91	20.05	10.89	72.68	91.18	96.18	100.00
acc_tech	3177	60.96	38.29	0.15	19.97	77.42	97.22	100
acc_ele	4868	79.89	30.30	0.01	66.98	99.22	100	100
acc_ele_r	4843	73.45	36.57	0	45.18	98.69	100	100
acc_ele_u	4868	89.78	19.18	3.5	89.54	99.86	100	100
ag_pct	5645	37.71	22.17	0.45	19.34	38.48	55.23	85.49
wtr_wth	519	111.34	500.68	0.02	3.13	14	33.15	5967.5
arable_pct	5588	13.83	13.41	0.00	3.07	10.17	19.51	73.39
forest_pct	5536	32.19	24.31	0	10.90	31.11	49.97	98.91
low_land_pct	520	5.42	10.81	0	0.53	1.60	3.61	55.88
wtr_str	189	63.42	280.56	0	1.91	7.90	29.86	2603.49
min_rent	5605	1.00	3.34	0	0	0.00	0.27	46.62
nat_gas_rent	5332	0.52	2.63	0	0	0	0.10	67.15
oil_rent	5339	3.63	9.52	0	0	0.00	1.07	78.55
pop_den	6270	391.90	1824.25	0.14	30.26	80.07	197.99	21389.1
pop_den_low	520	7.58	11.08	0	1.27	3.64	8.34	58.51
pop_slum	433	50.86	23.75	3.3	32.7	52.8	68.9	98.9
pop_rural	6314	43.13	24.63	0	23.22	43.56	64.20	94.66
pop_urb	6314	56.87	24.63	5.34	35.80	56.44	76.78	100
death_nonc	724	66.32	23.54	14.9	43.975	73.95	86.35	95.3
death_c	724	24.79	22.55	1.2	6.7	14.5	45.63	78.2

Table 3 Continued

Variable	Preimputation							
	count	mean	std	min	25%	50%	75%	max
sev_wast	578	2.34	2.06	0	0.8	1.8	3	15.9
stunt	751	29.25	16.02	0	16.45	29.2	40.9	73.6
under_n	2944	12.50	12.20	2.5	2.5	7.45	18.225	71.5
under_w	770	15.89	12.78	0	4.6	13.3	23.4	61.5
pop_grow	6339	1.49	1.53	-9.08	0.49	1.36	2.42	17.51
price_diesel	1812	0.80	0.47	0.01	0.41	0.75	1.12	3
price_gas	1830	0.93	0.48	0.00	0.56	0.89	1.25	3.33
phone_fixed	6066	19.36	20.33	0	2.19	12.16	31.38	135.60
pop_internet	4969	24.07	28.44	0	0.90	9.66	42.76	100.00
phone_cell	5965	47.50	52.75	0	0.48	23.39	91.61	345.32
mig_net	1148	6098.20	665061.87	-3266243	-120003	-9004	42357.75	8859954
unemp	5159	8.18	6.22	0.14	3.62	6.66	11.10	37.94
unemp_m	5159	15.85	11.38	0.18	7.22	13.10	21.44	64.82
vul_emp_m	5159	38.09	26.18	0.15	14.64	31.82	58.54	92.12
lf_rt_f	5321	55.49	17.31	6.25	45.71	58.03	67.46	91.95
lf_rt_m	5321	78.44	7.99	40.43	74.76	79.56	83.69	96.20
lf_rt_tot	5321	67.21	10.67	33.26	61.13	68.43	73.86	91.54
pop_tot	6343	29928081.35	120971321.72	8779	570046.5	5267900	18218179	1392730000
gini	1395	39.19	9.44	23.3	31.75	37.2	46.25	65.8
pov_hc	765	29.65	16.84	0.4	17.1	26.2	40.1	83.3
mil_exp	4304	2.33	2.99	0	1.11	1.68	2.71	117.35
idp_conflict	281	224690.79	413256.83	2	6000	75000	246000	3000000
idp_disaster	1069	248001.49	1138859.01	2	1100	8100	52000	18660000
af_pers	4492	1.65	1.97	0	0.51	1.03	1.94	34.85
rugged	6153	1.40	1.38	0.003	0.41	0.967	2.038	7.811
rugged_popw	6153	0.83	0.94	0.003	0.274	0.574	1.041	6.722
rugged_pc	6153	19.86	22.94	0	1.426	12.07	33.976	100
land_area	6123	63574.94	177821.49	1	1100	10025	46993	1638134
soil	6123	38.26	27.22	0	15.85	35.49	56.26	100
desert	6153	3.07	10.57	0	0	0	0	77.28
tropical	6153	43.88	46.33	0	0	14.68	100	100
near_coast	6153	51.30	41.66	0	9.92	42.91	100	100
gemstones	6153	5371.39	29768.60	0	0	0	0	264154
polity2	4767	3.14	6.61	-10	-3	6	9	10
durable	4829	24.77	30.72	0	4	15	33	209
parcomp	4767	3.20	1.44	0	2	3	4	5
corrupt	1069	2.88	0.67	1	2.5	3	3.5	4.5
rule_law	1069	2.89	0.64	1	2.5	3	3.5	4
con1_next	6147	0.25	0.43	0	0	0	1	1
con2_next	5931	0.25	0.43	0	0	0	0	1
con3_next	5715	0.25	0.43	0	0	0	0	1
con5_next	5283	0.24	0.43	0	0	0	0	1
con10_next	4205	0.23	0.42	0	0	0	0	1
con1_1_next	6147	0.18	0.39	0	0	0	0	1
con1_2_next	6147	0.10	0.30	0	0	0	0	1
con1_3_next	6147	0.18	0.38	0	0	0	0	1
con_last3	5982	0.75	1.19	0	0	0	1	3

Table 4: Summary Statistics after Imputation

Variable	Post Imputation							
	count	mean	std	min	25%	50%	75%	max
year	6363	2003.67	8.61	1989	1996	2004	2011	2018
con_ev_1	6363	16.85	108.63	0	0	0	0	3567
con_ev_2	6363	2.23	18.47	0	0	0	0	639
con_ev_3	6363	4.89	23.29	0	0	0	0	524
total_ev	6363	23.98	124.71	0	0	0	1	3722
con_in_1	6363	0.18	0.39	0	0	0	0	1
con_in_2	6363	0.10	0.30	0	0	0	0	1
con_in_3	6363	0.18	0.38	0	0	0	0	1
con_in	6363	0.25	0.43	0	0	0	1	1
net_gs	6363	7.81E+08	3.74E+10	-7.62E+11	-9.25E+08	-6.07E+07	3.00E+08	3.58E+11
imp_gs	6363	52.50	28.75	0	30.57	47.67	74.42	427.58
gdp_cap	6363	12650.42	19132.71	94.56	1147.43	4428.57	17516.16	189170.90
gdp_cap_gr	6363	1.91	5.94	-64.99	-0.21	1.97	4.08	140.37
exp_gs	6363	46.38	30.29	0.01	24.32	39.82	66.94	433.22
inf	6363	22.38	346.93	-18.11	2.01	3.76	6.93	23773.13
out_sch_a	6363	7.39	10.81	0	2.71	4.62	6.48	88.82
out_sch_af	6363	10.63	11.37	0.01	4.54	9.83	12.29	92.66
out_sch_am	6363	9.17	9.26	0.03	5.90	7.30	8.76	85.29
out_sch_c	6363	5.89	11.49	0	0.99	1.91	3.85	80.81
out_sch_cf	6363	6.58	12.19	0.00	1.59	2.51	3.61	86.30
out_sch_cm	6363	6.09	10.24	0.00	1.33	2.58	4.89	77.26
hs_grad_f	6363	34.96	11.62	0.16	32.07	32.07	32.07	96.43
hs_grad_m	6363	35.58	12.52	0.86	32.17	32.17	32.17	96.18
hs_grad	6363	35.25	11.98	0.50	32.12	32.12	32.12	96.31
gov_ed	6363	4.81	1.52	0	4.07	4.63	5.83	44.33
lit_r_f	6363	94.72	10.31	4.59	96.84	96.94	97.02	100.00
lit_r_m	6363	95.95	6.95	18.26	97.19	97.39	97.44	100.00
lit_r	6363	95.30	8.64	10.89	97.06	97.11	97.21	100.00
acc_tech	6363	80.51	33.36	0.15	77.65	100	100	100
acc_ele	6363	84.49	27.78	0.01	83.89	99.66	100	100
acc_ele_r	6363	79.72	33.81	0	71.99	99.75	100	100.12
acc_ele_u	6363	92.18	17.32	3.5	95.75	100.00	100	100.00
ag_pct	6363	34.71	22.51	0.45	11.25	33.10	53.16	85.49
wtr_wth	6363	18.00	145.57	0.02	8.39	9.46	12.98	5967.5
arable_pct	6363	13.50	12.60	0.00	3.73	11.11	17.59	73.39
forest_pct	6363	28.31	24.80	0	3.33	23.73	46.32	98.91
low_land_pct	6363	6.94	3.12	0	7.07	7.07	7.07	55.88
wtr_str	6363	11.28	49.08	0	9.69	9.69	9.69	2603.49
min_rent	6363	0.88	3.15	0	0	0	0.16	46.62
nat_gas_rent	6363	0.44	2.42	0	0	0	0.05	67.15
oil_rent	6363	3.05	8.82	0	0	0	0.54	78.55
pop_den	6363	393.20	1810.93	0.14	30.75	81.79	208.70	21389.1
pop_den_low	6363	7.11	3.17	0	7.07	7.07	7.07	58.51
pop_slum	6363	44.09	10.61	3.3	34.02	45.85	51.20	98.9
pop_rural	6363	43.22	24.55	0	23.33	43.77	64.13	94.66
pop_urb	6363	56.78	24.55	5.34	35.87	56.23	76.67	100
death_nonc	6363	86.18	10.66	14.9	88.49	88.79	88.93	95.3
death_c	6363	7.51	9.81	1.2	5.26	5.27	5.30	78.2

Table 4 Continued

Variable	Post Imputation							
	count	mean	std	min	25%	50%	75%	max
sev_wast	6363	0.25	0.91	-0.01	0.01	0.03	0.09	15.9
stunt	6363	6.00	10.14	0	2.65	3.14	3.29	73.6
under_n	6363	7.13	9.68	2.5	2.5	2.5	6.15	71.5
under_w	6363	2.68	6.62	0	0.79	0.97	1.01	61.5
pop_grow	6363	1.49	1.53	-9.08	0.49	1.36	2.42	17.51
price_diesel	6363	0.65	0.35	0.01	0.34	0.61	0.89	3
price_gas	6363	0.70	0.36	0.00	0.40	0.64	0.92	3.33
phone_fixed	6363	20.12	20.15	0	2.50	13.89	35.28	135.60
pop_internet	6363	21.80	31.06	-16.24	0.13	6.43	40.04	100
phone_cell	6363	49.38	53.86	-12.81	0.57	26.20	95.58	345.32
mig_net	6363	5484.18	282420.28	-3266243	624.51	3965	8335.49	8859954
unemp	6363	8.02	5.62	0.14	4.18	6.91	9.68	37.94
unemp_m	6363	15.67	10.27	0.18	8.37	14.36	19.23	64.82
vul_emp_m	6363	32.43	26.33	0.15	8.37	22.81	52.17	92.12
lf_rt_f	6363	57.79	16.67	6.25	48.80	62.15	69.83	91.95
lf_rt_m	6363	79.02	7.45	40.43	75.79	80.22	83.63	96.20
lf_rt_tot	6363	68.61	10.26	33.26	62.98	70.56	75.85	91.54
pop_tot	6363	29834279.57	120792578.65	8779	551197	5228172	18110640.5	1392730000
gini	6363	37.33	4.80	23.3	34.73	36.70	37.95	65.8
pov_hc	6363	29.82	11.19	0.4	19.19	29.49	39.39	83.3
mil_exp	6363	2.42	2.47	0	1.38	2.27	2.72	117.35
idp_conflict	6363	83662.91	93991.53	2	61054.34	75039.84	87027.42	3000000
idp_disaster	6363	575389.51	512880.75	2	248636.95	715222.20	720962.09	18660000
af_pers	6363	1.39	1.71	0	0.57	0.85	1.47	34.85
rugged	6363	1.37	1.37	0.00	0.42	0.91	1.99	7.81
rugged_popw	6363	0.82	0.92	0.00	0.28	0.56	1.03	6.72
rugged_pc	6363	19.20	22.84	0	0.69	9.74	33.06	100
land_area	6363	61177.69	174855.00	1	887	9150	45286	1638134
soil	6363	37.62	26.89	0	16.79	33.25	56.04	100
desert	6363	2.97	10.41	0	0	0	0	77.28
tropical	6363	45.73	46.65	0	0	27.928	100	100
near_coast	6363	52.91	41.88	0	10.28	48.22	100	100
gemstones	6363	5194.12	29288.90	0	0	0	0	264154
polity2	6363	4.84	6.43	-10	0	8	9.97	10.08
durable	6363	56.16	61.94	0	7	24	140.5	209
parcomp	6363	3.63	1.45	0	3	4	4.98	5.08
corrupt	6363	2.89	0.30	1	2.79	2.86	3	4.5
rule_law	6363	3.00	0.27	1	2.97	3.00	3.07	4
con1_next	6363	0.24	0.43	0	0	0	0	1
con2_next	6363	0.23	0.42	0	0	0	0	1
con3_next	6363	0.22	0.42	0	0	0	0	1
con5_next	6363	0.20	0.40	0	0	0	0	1
con10_next	6363	0.16	0.36	0	0	0	0	1
con1_1_next	6363	0.18	0.38	0	0	0	0	1
con1_2_next	6363	0.10	0.30	0	0	0	0	1
con1_3_next	6363	0.17	0.38	0	0	0	0	1
con_last3	6363	0.70	1.17	0	0	0	1	3

Table 5: Base Specification Results

Algorithm		Base Specification			Spec w/ PCA			Spec w/ FE			Spec w/ FE & PCA		
		Recall	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*
LR	*Score*	0.853	0.825	0.909	0.854	0.828	0.911	0.890	0.826	0.906	0.873	0.827	0.908
	Std Dev	(0.020)	(0.014)	(0.008)	(0.022)	(0.013)	(0.007)	(0.015)	(0.011)	(0.007)	(0.019)	(0.016)	(0.009)
KNN	*Score*	0.889	0.814	0.898	0.887	0.811	0.896	0.936	0.785	0.871	0.944	0.766	0.855
	Std Dev	(0.027)	(0.016)	(0.009)	(0.022)	(0.018)	(0.011)	(0.014)	(0.014)	(0.010)	(0.014)	(0.010)	(0.008)
LDA	*Score*	0.836	0.830	0.914	0.836	0.828	0.913	0.891	0.829	0.907	0.879	0.832	0.911
	Std Dev	(0.019)	(0.015)	(0.008)	(0.023)	(0.017)	(0.009)	(0.018)	(0.016)	(0.009)	(0.018)	(0.015)	(0.009)
SVC	*Score*	0.880	0.829	0.909	0.874	0.829	0.909	0.894	0.830	0.908	0.898	0.834	0.910
	Std Dev	(0.019)	(0.010)	(0.006)	(0.020)	(0.014)	(0.008)	(0.017)	(0.014)	(0.009)	(0.019)	(0.011)	(0.006)
DTREE	*Score*	0.794	0.780	0.887	0.777	0.743	0.865	0.778	0.779	0.889	0.783	0.765	0.879
	Std Dev	(0.035)	(0.025)	(0.012)	(0.036)	(0.030)	(0.016)	(0.040)	(0.020)	(0.007)	(0.040)	(0.025)	(0.012)
RF	*Score*	0.860	0.827	0.909	0.880	0.746	0.850	0.891	0.810	0.895	0.874	0.776	0.873
	Std Dev	(0.021)	(0.012)	(0.007)	(0.023)	(0.012)	(0.008)	(0.021)	(0.015)	(0.009)	(0.022)	(0.025)	(0.016)
XRF	*Score*	0.845	0.837	0.917	0.861	0.820	0.905	0.836	0.839	0.919	0.884	0.833	0.911
	Std Dev	(0.019)	(0.018)	(0.010)	(0.024)	(0.014)	(0.008)	(0.023)	(0.016)	(0.008)	(0.025)	(0.018)	(0.010)

Algorithms Used: LR: Logistic Regression, KNN: K Nearest Neighbors, LDA: Linear Discrimant Analysis, SVC: Support Vector Classifier, DTREE: Decision Tree Classifier, RF: Random Forest Classifier, XRF: XG Boost Random Forest

PCA: Principal Component Analysis accounting for 80% of explained variance

FE: Fixed effects for country, region, and year included

Table 6: State-Based Conflict Specification Results

Algorithm		State Based Conflict											
		Base Specification			Spec w/ PCA			Spec w/ FE			Spec w/ FE & PCA		
		Recall	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*
LR	*Score*	0.890	0.761	0.897	0.883	0.758	0.896	0.897	0.803	0.919	0.901	0.804	0.919
	Std Dev	(0.029)	(0.029)	(0.016)	(0.033)	(0.026)	(0.014)	(0.027)	(0.024)	(0.011)	(0.032)	(0.029)	(0.013)
KNN	*Score*	0.897	0.786	0.910	0.900	0.781	0.908	0.933	0.738	0.878	0.940	0.706	0.856
	Std Dev	(0.024)	(0.028)	(0.015)	(0.033)	(0.021)	(0.010)	(0.016)	(0.032)	(0.020)	(0.022)	(0.030)	(0.021)
LDA	*Score*	0.898	0.758	0.894	0.882	0.754	0.894	0.909	0.797	0.915	0.906	0.797	0.915
	Std Dev	(0.027)	(0.028)	(0.015)	(0.035)	(0.029)	(0.015)	(0.029)	(0.020)	(0.010)	(0.032)	(0.025)	(0.012)
SVC	*Score*	0.887	0.801	0.919	0.902	0.787	0.910	0.892	0.807	0.922	0.899	0.800	0.918
	Std Dev	(0.029)	(0.027)	(0.013)	(0.019)	(0.029)	(0.016)	(0.029)	(0.022)	(0.010)	(0.034)	(0.027)	(0.012)
DTREE	*Score*	0.772	0.755	0.909	0.772	0.711	0.885	0.769	0.765	0.913	0.747	0.716	0.891
	Std Dev	(0.048)	(0.026)	(0.009)	(0.036)	(0.021)	(0.008)	(0.037)	(0.029)	(0.010)	(0.031)	(0.030)	(0.012)
RF	*Score*	0.898	0.760	0.896	0.883	0.701	0.861	0.908	0.744	0.885	0.881	0.716	0.872
	Std Dev	(0.025)	(0.026)	(0.015)	(0.021)	(0.030)	(0.020)	(0.030)	(0.030)	(0.017)	(0.031)	(0.022)	(0.012)
XRF	*Score*	0.855	0.821	0.932	0.888	0.777	0.906	0.850	0.823	0.933	0.896	0.801	0.918
	Std Dev	(0.037)	(0.028)	(0.011)	(0.034)	(0.030)	(0.015)	(0.038)	(0.033)	(0.013)	(0.026)	(0.030)	(0.014)

Algorithms Used: LR: Logistic Regression, KNN: K Nearest Neighbors, LDA: Linear Discrimant Analysis, SVC: Support Vector Classifier, DTREE: Decision Tree Classifier, RF: Random Forest Classifier, XRF: XG Boost Random Forest

PCA: Principal Component Analysis accounting for 80% of explained variance

FE: Fixed effects for country, region, and year included

Table 7: Nonstate Conflict Specification Results

Algorithm		Nonstate Conflict											
		Base Specification			Spec w/ PCA			Spec w/ FE			Spec w/ FE & PCA		
		Recall	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*
LR	*Score*	0.887	0.577	0.868	0.871	0.561	0.862	0.879	0.671	0.913	0.876	0.657	0.908
	Std Dev	(0.042)	(0.031)	(0.013)	(0.046)	(0.035)	(0.015)	(0.035)	(0.026)	(0.008)	(0.050)	(0.023)	(0.008)
KNN	*Score*	0.871	0.671	0.914	0.860	0.667	0.913	0.932	0.638	0.893	0.944	0.601	0.873
	Std Dev	(0.040)	(0.037)	(0.011)	(0.038)	(0.028)	(0.009)	(0.030)	(0.023)	(0.010)	(0.022)	(0.021)	(0.012)
LDA	*Score*	0.910	0.548	0.847	0.886	0.540	0.847	0.907	0.647	0.900	0.911	0.641	0.897
	Std Dev	(0.041)	(0.038)	(0.018)	(0.051)	(0.034)	(0.015)	(0.039)	(0.024)	(0.007)	(0.023)	(0.022)	(0.009)
SVC	*Score*	0.865	0.672	0.914	0.890	0.660	0.907	0.873	0.716	0.930	0.863	0.700	0.925
	Std Dev	(0.040)	(0.031)	(0.010)	(0.040)	(0.035)	(0.012)	(0.048)	(0.023)	(0.005)	(0.044)	(0.022)	(0.005)
DTREE	*Score*	0.715	0.667	0.927	0.676	0.591	0.905	0.675	0.653	0.927	0.646	0.601	0.913
	Std Dev	(0.045)	(0.039)	(0.011)	(0.052)	(0.034)	(0.010)	(0.050)	(0.039)	(0.009)	(0.059)	(0.041)	(0.009)
RF	*Score*	0.903	0.542	0.845	0.894	0.516	0.830	0.900	0.532	0.840	0.860	0.604	0.886
	Std Dev	(0.040)	(0.034)	(0.017)	(0.038)	(0.025)	(0.014)	(0.039)	(0.023)	(0.012)	(0.051)	(0.029)	(0.009)
XRF	*Score*	0.845	0.707	0.929	0.889	0.612	0.886	0.855	0.697	0.925	0.879	0.679	0.916
	Std Dev	(0.050)	(0.035)	(0.010)	(0.039)	(0.030)	(0.012)	(0.038)	(0.034)	(0.011)	(0.041)	(0.020)	(0.006)

Algorithms Used: LR: Logistic Regression, KNN: K Nearest Neighbors, LDA: Linear Discriminant Analysis, SVC: Support Vector Classifier, DTREE: Decision Tree Classifier, RF: Random Forest Classifier, XRF: XG Boost Random Forest

PCA: Principal Component Analysis accounting for 80% of explained variance

FE: Fixed effects for country, region, and year included

Table 8: One-Sided Conflict Specification Results

Algorithm		One-Sided Conflict											
		Base Specification			Spec w/ PCA			Spec w/ FE			Spec w/ FE & PCA		
		Recall	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*
LR	*Score*	0.872	0.743	0.892	0.868	0.726	0.883	0.893	0.773	0.906	0.887	0.762	0.901
	Std Dev	(0.028)	(0.021)	(0.010)	(0.025)	(0.018)	(0.009)	(0.020)	(0.017)	(0.009)	(0.033)	(0.026)	(0.012)
KNN	*Score*	0.878	0.747	0.894	0.869	0.743	0.893	0.933	0.729	0.876	0.951	0.716	0.865
	Std Dev	(0.027)	(0.017)	(0.009)	(0.031)	(0.012)	(0.007)	(0.014)	(0.014)	(0.009)	(0.019)	(0.013)	(0.009)
LDA	*Score*	0.873	0.727	0.883	0.865	0.727	0.884	0.914	0.770	0.902	0.906	0.761	0.898
	Std Dev	(0.019)	(0.019)	(0.009)	(0.019)	(0.022)	(0.011)	(0.014)	(0.017)	(0.010)	(0.022)	(0.014)	(0.007)
SVC	*Score*	0.877	0.771	0.907	0.871	0.754	0.898	0.898	0.792	0.916	0.899	0.787	0.913
	Std Dev	(0.022)	(0.013)	(0.007)	(0.024)	(0.020)	(0.011)	(0.020)	(0.018)	(0.009)	(0.020)	(0.020)	(0.011)
DTREE	*Score*	0.741	0.722	0.898	0.737	0.681	0.877	0.737	0.733	0.904	0.697	0.669	0.877
	Std Dev	(0.031)	(0.019)	(0.008)	(0.035)	(0.025)	(0.010)	(0.036)	(0.023)	(0.008)	(0.042)	(0.026)	(0.009)
RF	*Score*	0.885	0.732	0.884	0.893	0.660	0.835	0.902	0.710	0.868	0.900	0.731	0.881
	Std Dev	(0.023)	(0.020)	(0.010)	(0.023)	(0.018)	(0.013)	(0.027)	(0.017)	(0.009)	(0.023)	(0.020)	(0.011)
XRF	*Score*	0.847	0.788	0.919	0.878	0.748	0.894	0.841	0.783	0.917	0.880	0.773	0.907
	Std Dev	(0.026)	(0.023)	(0.009)	(0.019)	(0.021)	(0.011)	(0.021)	(0.020)	(0.009)	(0.018)	(0.021)	(0.011)

Algorithms Used: LR: Logistic Regression, KNN: K Nearest Neighbors, LDA: Linear Discriminant Analysis, SVC: Support Vector Classifier, DTREE: Decision Tree Classifier, RF: Random Forest Classifier, XRF: XG Boost Random Forest

PCA: Principal Component Analysis accounting for 80% of explained variance

FE: Fixed effects for country, region, and year included

Table 9: Performance Over Time for Base Specification

Algorithm		Base Specification Recall					Base Specification F1					Base Specification Accuracy				
		1	*2*	*3*	*5*	*10*	*1*	*2*	*3*	*5*	*10*	*1*	*2*	*3*	*5*	*10*
LR	*Score*	0.853	0.848	0.837	0.830	0.835	0.825	0.800	0.779	0.760	0.738	0.909	0.894	0.882	0.872	0.861
	Std Dev	(0.020)	(0.028)	(0.030)	(0.052)	(0.045)	(0.014)	(0.008)	(0.017)	(0.037)	(0.033)	(0.008)	(0.004)	(0.011)	(0.020)	(0.018)
KNN	*Score*	0.889	0.904	0.896	0.896	0.883	0.814	0.813	0.797	0.800	0.781	0.898	0.896	0.887	0.891	0.884
	Std Dev	(0.027)	(0.021)	(0.014)	(0.024)	(0.049)	(0.016)	(0.023)	(0.019)	(0.022)	(0.029)	(0.009)	(0.015)	(0.013)	(0.014)	(0.014)
LDA	*Score*	0.836	0.822	0.821	0.808	0.821	0.830	0.806	0.786	0.763	0.737	0.914	0.902	0.889	0.878	0.863
	Std Dev	(0.019)	(0.032)	(0.028)	(0.051)	(0.044)	(0.015)	(0.018)	(0.018)	(0.036)	(0.024)	(0.008)	(0.009)	(0.011)	(0.018)	(0.012)
SVC	*Score*	0.880	0.876	0.882	0.884	0.866	0.829	0.822	0.815	0.809	0.786	0.909	0.906	0.901	0.898	0.889
	Std Dev	(0.019)	(0.023)	(0.033)	(0.029)	(0.038)	(0.010)	(0.015)	(0.014)	(0.028)	(0.025)	(0.006)	(0.009)	(0.009)	(0.016)	(0.013)
DTREE	*Score*	0.794	0.795	0.785	0.775	0.771	0.780	0.775	0.772	0.755	0.759	0.887	0.885	0.885	0.878	0.886
	Std Dev	(0.035)	(0.028)	(0.017)	(0.040)	(0.046)	(0.025)	(0.017)	(0.024)	(0.029)	(0.030)	(0.012)	(0.010)	(0.015)	(0.015)	(0.013)
RF	*Score*	0.860	0.852	0.842	0.832	0.828	0.827	0.806	0.785	0.762	0.740	0.909	0.898	0.886	0.874	0.864
	Std Dev	(0.021)	(0.034)	(0.028)	(0.047)	(0.045)	(0.012)	(0.015)	(0.014)	(0.033)	(0.031)	(0.007)	(0.007)	(0.008)	(0.017)	(0.017)
XRF	*Score*	0.845	0.835	0.829	0.845	0.856	0.837	0.820	0.803	0.804	0.796	0.917	0.909	0.899	0.899	0.897
	Std Dev	(0.019)	(0.043)	(0.032)	(0.042)	(0.042)	(0.018)	(0.022)	(0.020)	(0.031)	(0.021)	(0.010)	(0.010)	(0.011)	(0.016)	(0.010)

Algorithms Used: LR: Logistic Regression, KNN: K Nearest Neighbors, LDA: Linear Discriminant Analysis, SVC: Support Vector Classifier, DTREE: Decision Tree Classifier, RF: Random Forest Classifier, XRF: XG Boost Random Forest

PCA: Principal Component Analysis accounting for 80% of explained variance

FE: Fixed effects for country, region, and year included

Figure 6: Performance Over Time for Base Specification

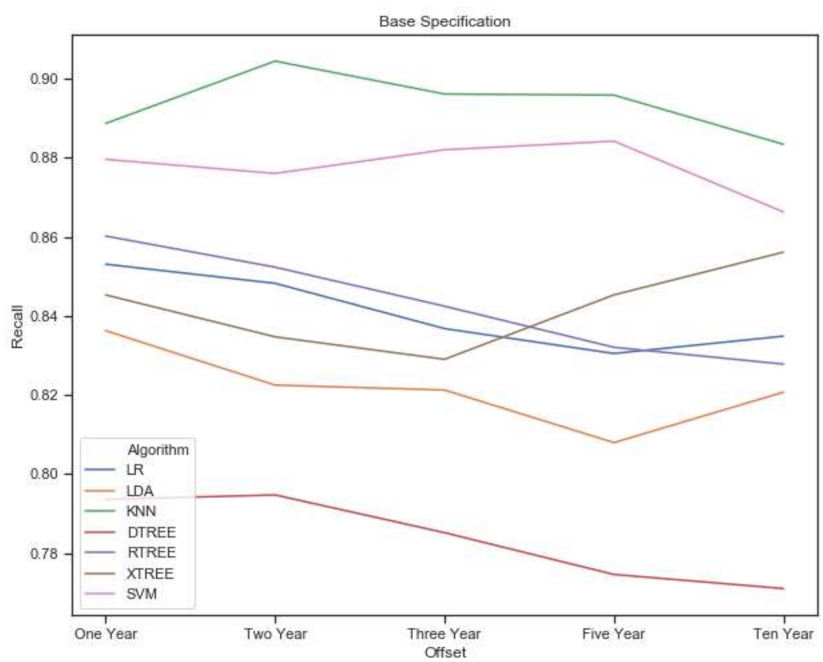
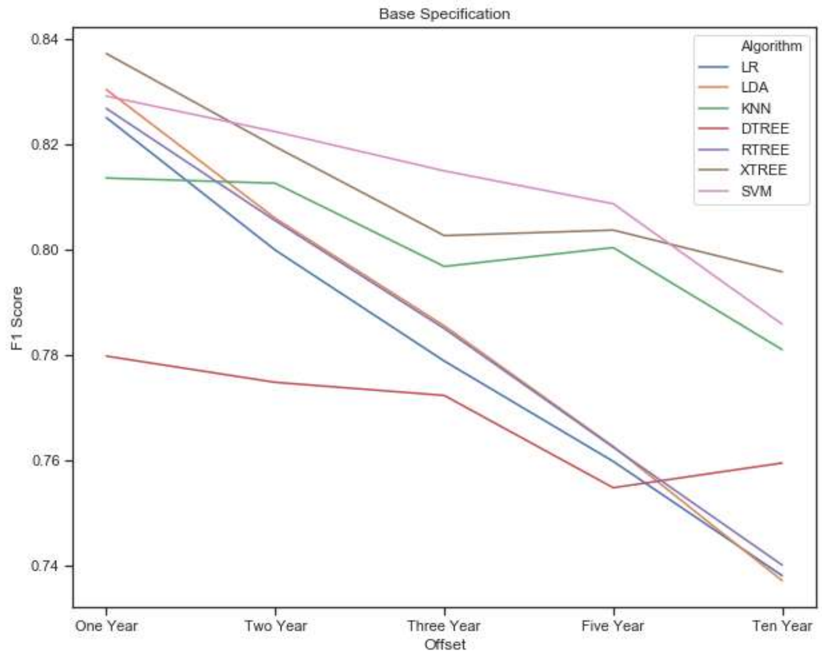


Figure 6 Continued



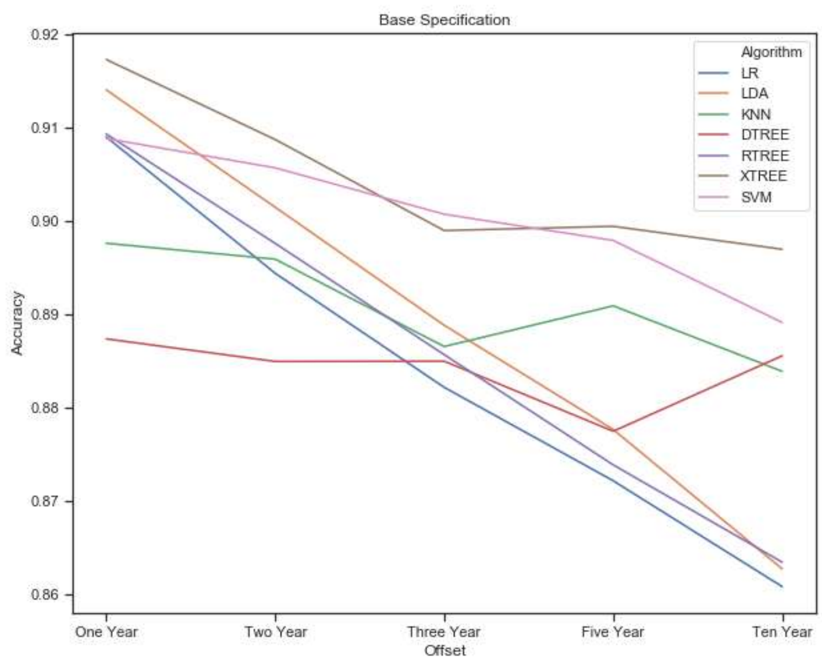


Table 10: Performance Over Time for Specification with PCA

Algorithm		Spec w/ PCA Recall					Spec w/ PCA F1					Spec w/ PCA Accuracy				
		1	*2*	*3*	*5*	*10*	*1*	*2*	*3*	*5*	*10*	*1*	*2*	*3*	*5*	*10*
LR	*Score*	0.854	0.844	0.830	0.813	0.828	0.828	0.805	0.784	0.758	0.733	0.911	0.898	0.887	0.874	0.859
	Std Dev	(0.022)	(0.031)	(0.023)	(0.049)	(0.031)	(0.013)	(0.010)	(0.020)	(0.037)	(0.021)	(0.007)	(0.005)	(0.012)	(0.020)	(0.013)
KNN	*Score*	0.887	0.893	0.900	0.893	0.887	0.811	0.806	0.796	0.791	0.784	0.896	0.893	0.886	0.885	0.885
	Std Dev	(0.022)	(0.018)	(0.022)	(0.031)	(0.036)	(0.018)	(0.018)	(0.019)	(0.023)	(0.028)	(0.011)	(0.011)	(0.013)	(0.015)	(0.015)
LDA	*Score*	0.836	0.819	0.807	0.796	0.810	0.828	0.804	0.784	0.759	0.738	0.913	0.901	0.890	0.877	0.865
	Std Dev	(0.023)	(0.032)	(0.023)	(0.057)	(0.043)	(0.017)	(0.019)	(0.017)	(0.036)	(0.026)	(0.009)	(0.009)	(0.009)	(0.018)	(0.014)
SVC	*Score*	0.874	0.862	0.868	0.878	0.866	0.829	0.809	0.800	0.796	0.771	0.909	0.898	0.893	0.890	0.879
	Std Dev	(0.020)	(0.030)	(0.029)	(0.035)	(0.043)	(0.014)	(0.017)	(0.018)	(0.029)	(0.023)	(0.008)	(0.010)	(0.011)	(0.016)	(0.012)
DTREE	*Score*	0.777	0.768	0.761	0.757	0.728	0.743	0.738	0.733	0.729	0.689	0.865	0.865	0.863	0.863	0.846
	Std Dev	(0.036)	(0.028)	(0.027)	(0.034)	(0.050)	(0.030)	(0.015)	(0.016)	(0.028)	(0.039)	(0.016)	(0.008)	(0.009)	(0.016)	(0.019)
RF	*Score*	0.880	0.867	0.878	0.876	0.879	0.746	0.728	0.708	0.688	0.673	0.850	0.839	0.820	0.807	0.799
	Std Dev	(0.023)	(0.034)	(0.024)	(0.043)	(0.023)	(0.012)	(0.023)	(0.016)	(0.022)	(0.019)	(0.008)	(0.013)	(0.014)	(0.016)	(0.017)
XRF	*Score*	0.861	0.858	0.859	0.841	0.845	0.820	0.805	0.789	0.768	0.757	0.905	0.897	0.886	0.876	0.872
	Std Dev	(0.024)	(0.031)	(0.029)	(0.046)	(0.022)	(0.014)	(0.015)	(0.024)	(0.031)	(0.020)	(0.008)	(0.008)	(0.015)	(0.016)	(0.012)

Algorithms Used: LR: Logistic Regression, KNN: K Nearest Neighbors, LDA: Linear Discrimant Analysis, SVC: Support Vector Classifier, DTREE: Decision Tree Classifier, RF: Random Forest Classifier, XRF: XG Boost Random Forest

PCA: Principal Component Analysis accounting for 80% of explained variance

FE: Fixed effects for country, region, and year included

Figure 7: Performance Over Time for Specification with PCA

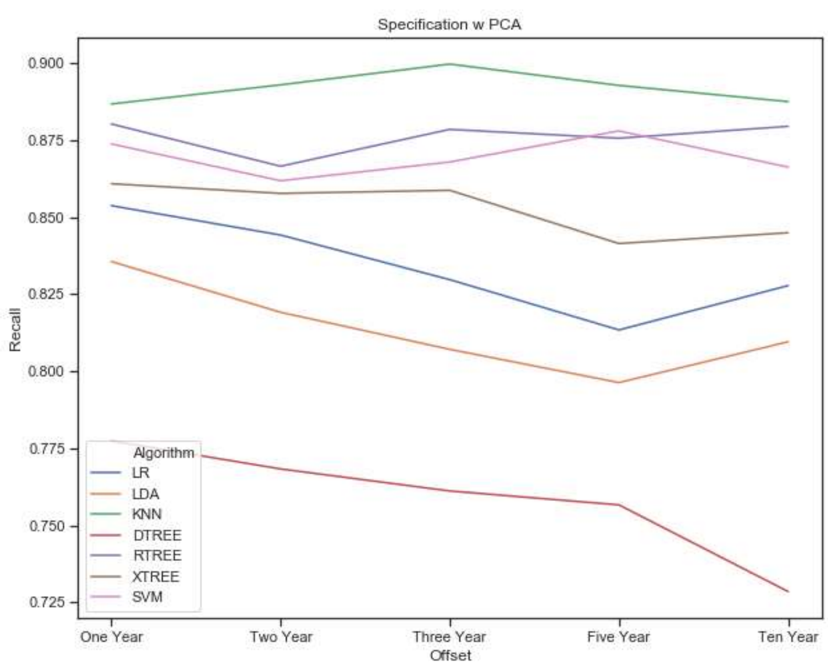
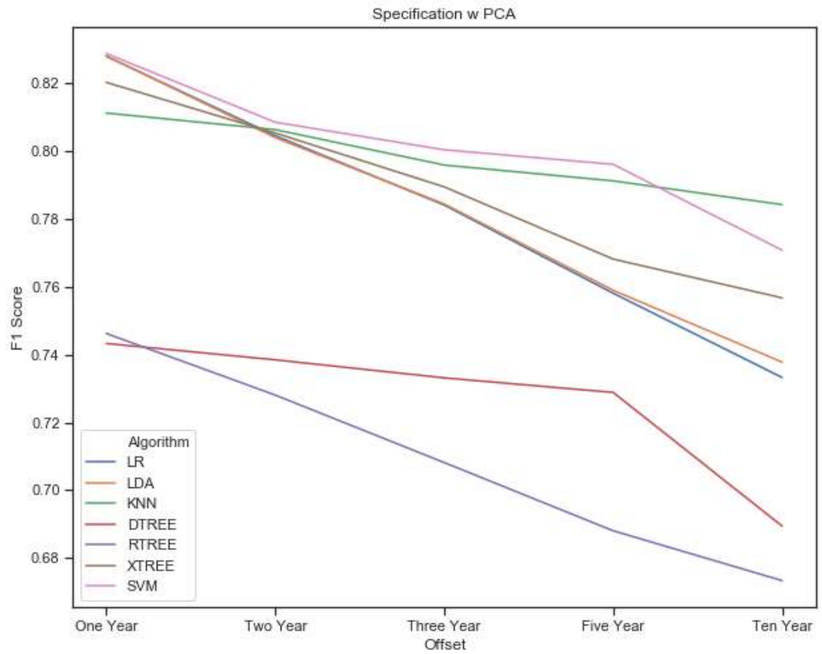


Figure 7 Continued



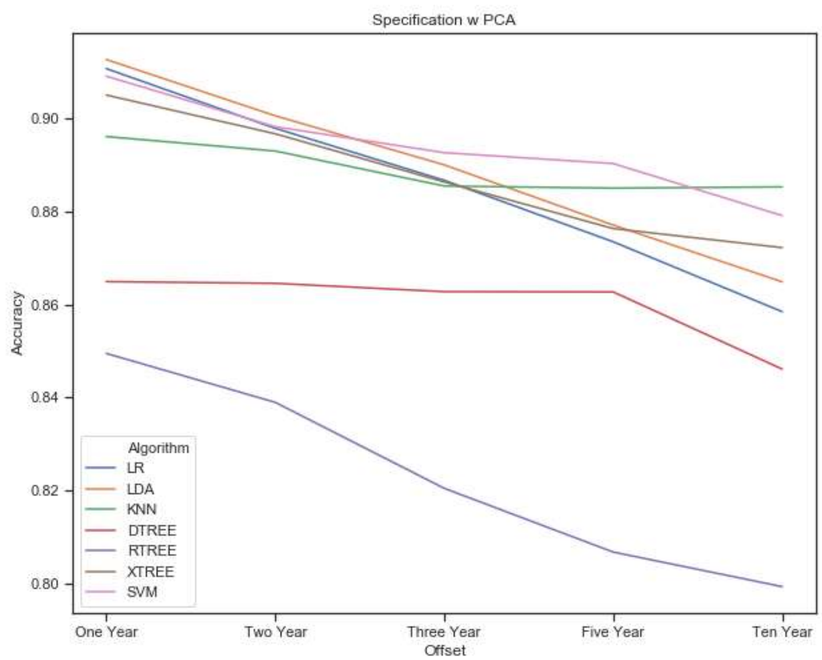


Table 11: Performance Over Time for Specification with Fixed Effects

Algorithm		Spec w/ FE Recall					Spec w/ FE F1					Spec w/ FE Accuracy				
		1	*2*	*3*	*5*	*10*	*1*	*2*	*3*	*5*	*10*	*1*	*2*	*3*	*5*	*10*
LR	*Score*	0.890	0.883	0.887	0.889	0.881	0.826	0.814	0.811	0.801	0.796	0.906	0.900	0.897	0.892	0.894
	Std Dev	(0.015)	(0.025)	(0.019)	(0.034)	(0.038)	(0.011)	(0.020)	(0.026)	(0.025)	(0.017)	(0.007)	(0.012)	(0.016)	(0.015)	(0.008)
KNN	*Score*	0.936	0.930	0.946	0.939	0.913	0.785	0.786	0.791	0.786	0.790	0.871	0.874	0.876	0.875	0.886
	Std Dev	(0.014)	(0.022)	(0.019)	(0.022)	(0.031)	(0.014)	(0.013)	(0.022)	(0.024)	(0.018)	(0.010)	(0.011)	(0.016)	(0.017)	(0.010)
LDA	*Score*	0.891	0.900	0.904	0.907	0.899	0.829	0.816	0.809	0.798	0.797	0.907	0.899	0.894	0.888	0.893
	Std Dev	(0.018)	(0.030)	(0.025)	(0.029)	(0.032)	(0.016)	(0.019)	(0.022)	(0.025)	(0.018)	(0.009)	(0.012)	(0.014)	(0.017)	(0.011)
SVC	*Score*	0.894	0.898	0.900	0.902	0.894	0.830	0.822	0.819	0.812	0.802	0.908	0.903	0.901	0.898	0.896
	Std Dev	(0.017)	(0.031)	(0.024)	(0.031)	(0.040)	(0.014)	(0.015)	(0.024)	(0.022)	(0.020)	(0.009)	(0.009)	(0.015)	(0.014)	(0.010)
DTREE	*Score*	0.778	0.768	0.776	0.763	0.799	0.779	0.758	0.763	0.757	0.779	0.889	0.878	0.880	0.881	0.894
	Std Dev	(0.040)	(0.039)	(0.028)	(0.029)	(0.066)	(0.020)	(0.026)	(0.019)	(0.022)	(0.047)	(0.007)	(0.012)	(0.010)	(0.011)	(0.021)
RF	*Score*	0.891	0.886	0.894	0.872	0.866	0.810	0.787	0.766	0.750	0.727	0.895	0.880	0.864	0.858	0.847
	Std Dev	(0.021)	(0.028)	(0.024)	(0.038)	(0.032)	(0.015)	(0.018)	(0.019)	(0.025)	(0.025)	(0.009)	(0.011)	(0.014)	(0.016)	(0.017)
XRF	*Score*	0.836	0.820	0.819	0.825	0.839	0.839	0.815	0.800	0.793	0.792	0.919	0.907	0.898	0.895	0.897
	Std Dev	(0.023)	(0.035)	(0.029)	(0.050)	(0.045)	(0.016)	(0.022)	(0.018)	(0.033)	(0.030)	(0.008)	(0.011)	(0.010)	(0.016)	(0.014)

Algorithms Used: LR: Logistic Regression, KNN: K Nearest Neighbors, LDA: Linear Discriminant Analysis, SVC: Support Vector Classifier, DTREE: Decision Tree Classifier, RF: Random Forest Classifier, XRF: XG Boost Random Forest

PCA: Principal Component Analysis accounting for 80% of explained variance

FE: Fixed effects for country, region, and year included

Figure 8: Performance Over Time for Specification with Fixed Effects

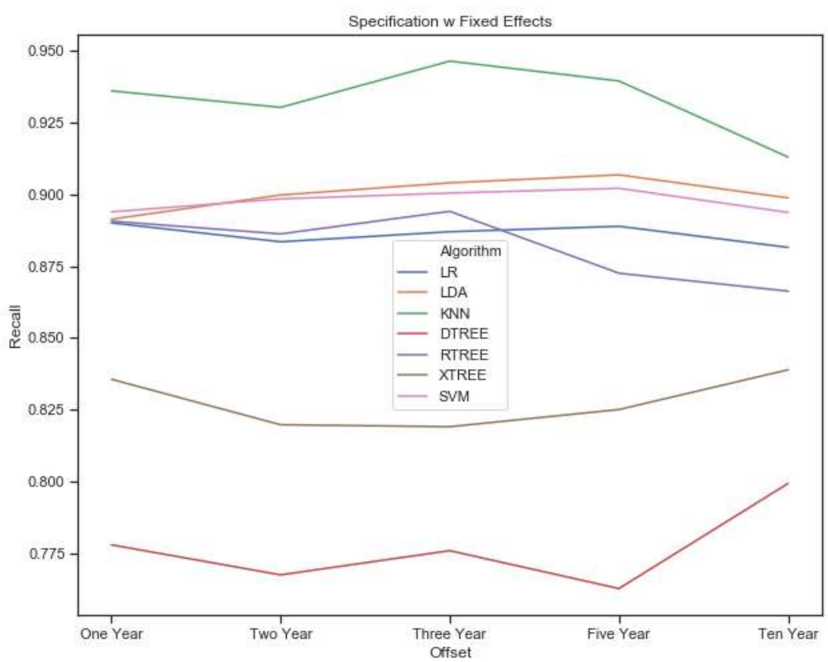
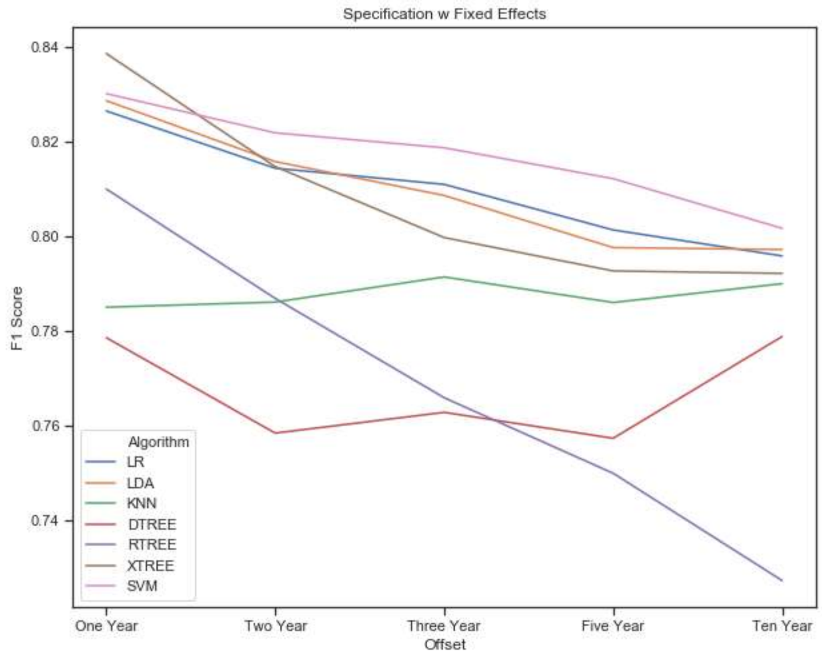


Figure 8 Continued



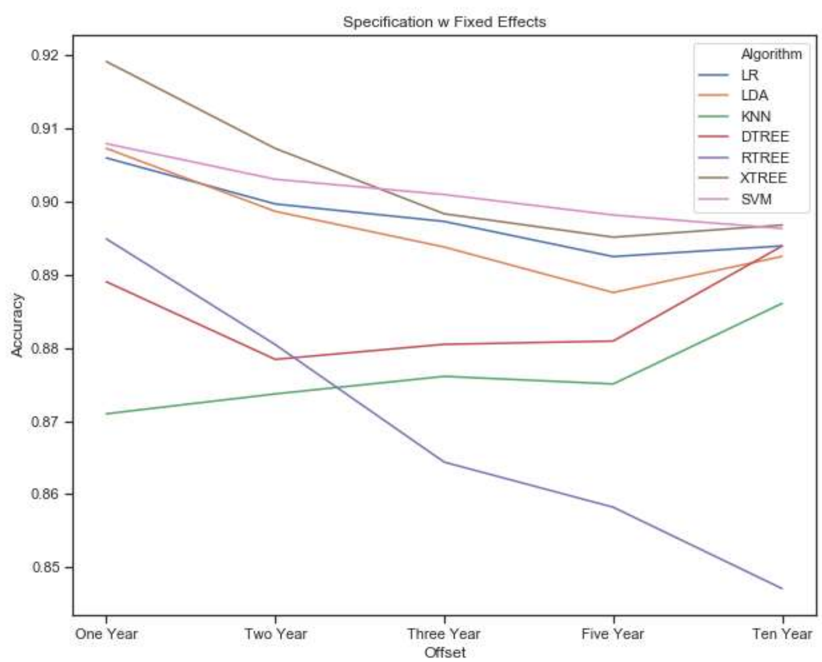


Table 12: Performance Over Time for Specification with Fixed Effects and PCA

Algorithm		Spec w/ FE & PCA Recall					Spec w/ FE & PCA F1					Spec w/ FE & PCA Accuracy				
		1	*2*	*3*	*5*	*10*	*1*	*2*	*3*	*5*	*10*	*1*	*2*	*3*	*5*	*10*
LR	*Score*	0.873	0.862	0.850	0.849	0.843	0.827	0.806	0.789	0.772	0.753	0.908	0.897	0.887	0.878	0.870
	Std Dev	(0.019)	(0.029)	(0.028)	(0.045)	(0.041)	(0.016)	(0.015)	(0.021)	(0.033)	(0.023)	(0.009)	(0.008)	(0.013)	(0.018)	(0.011)
KNN	*Score*	0.944	0.938	0.942	0.948	0.934	0.766	0.770	0.768	0.778	0.775	0.855	0.861	0.858	0.868	0.873
	Std Dev	(0.014)	(0.025)	(0.018)	(0.021)	(0.038)	(0.010)	(0.014)	(0.026)	(0.024)	(0.019)	(0.008)	(0.012)	(0.020)	(0.017)	(0.012)
LDA	*Score*	0.879	0.867	0.864	0.865	0.873	0.832	0.813	0.802	0.788	0.771	0.911	0.901	0.894	0.886	0.878
	Std Dev	(0.018)	(0.030)	(0.027)	(0.042)	(0.039)	(0.015)	(0.015)	(0.019)	(0.036)	(0.026)	(0.009)	(0.008)	(0.011)	(0.021)	(0.014)
SVC	*Score*	0.898	0.899	0.895	0.907	0.891	0.834	0.821	0.814	0.809	0.798	0.910	0.903	0.899	0.896	0.894
	Std Dev	(0.019)	(0.026)	(0.022)	(0.027)	(0.026)	(0.011)	(0.014)	(0.022)	(0.024)	(0.012)	(0.006)	(0.008)	(0.014)	(0.016)	(0.006)
DTREE	*Score*	0.783	0.772	0.766	0.785	0.771	0.765	0.754	0.749	0.757	0.739	0.879	0.875	0.873	0.877	0.872
	Std Dev	(0.040)	(0.036)	(0.030)	(0.054)	(0.041)	(0.025)	(0.028)	(0.026)	(0.024)	(0.019)	(0.012)	(0.014)	(0.015)	(0.011)	(0.008)
RF	*Score*	0.874	0.864	0.875	0.879	0.862	0.776	0.770	0.751	0.750	0.753	0.873	0.871	0.857	0.858	0.867
	Std Dev	(0.022)	(0.030)	(0.027)	(0.031)	(0.027)	(0.025)	(0.025)	(0.014)	(0.011)	(0.034)	(0.016)	(0.016)	(0.011)	(0.007)	(0.022)
XRF	*Score*	0.884	0.878	0.874	0.863	0.866	0.833	0.821	0.812	0.795	0.794	0.911	0.904	0.900	0.892	0.894
	Std Dev	(0.025)	(0.028)	(0.025)	(0.034)	(0.030)	(0.018)	(0.017)	(0.023)	(0.028)	(0.021)	(0.010)	(0.011)	(0.014)	(0.017)	(0.012)

<u>Algorithms Used:</u> LR: Logistic Regression, KNN: K Nearest Neighbors, LDA: Linear Discrimant Analysis, SVC: Support Vector Classifier, DTREE: Decision Tree Classifier, RF: Random Forest Classifier, XRF: XG Boost Random Forest

<u>PCA</u>: Principal Component Analysis accounting for 80% of explained variance

<u>FE</u>: Fixed effects for country, region, and year included

Figure 9: Performance Over Time for Specification with Fixed Effects and PCA

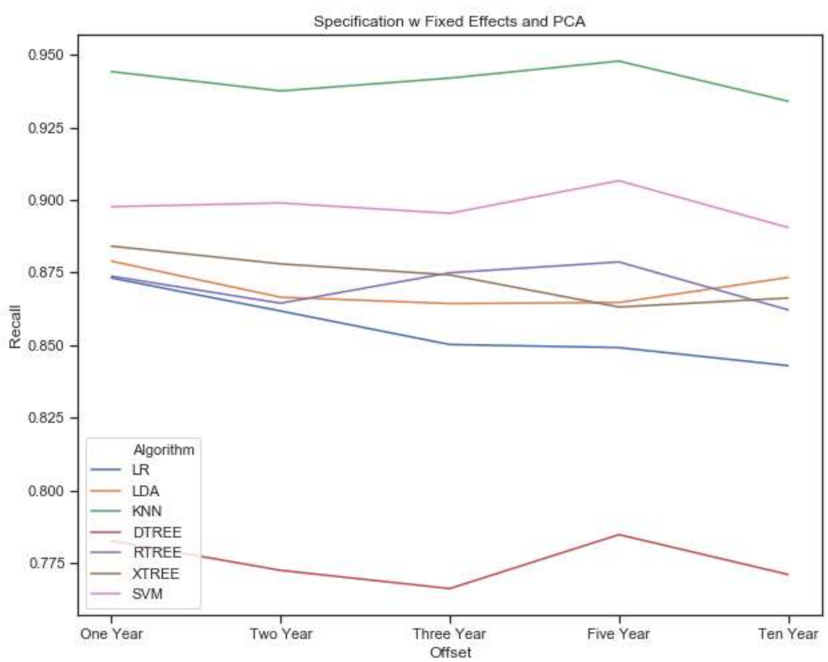
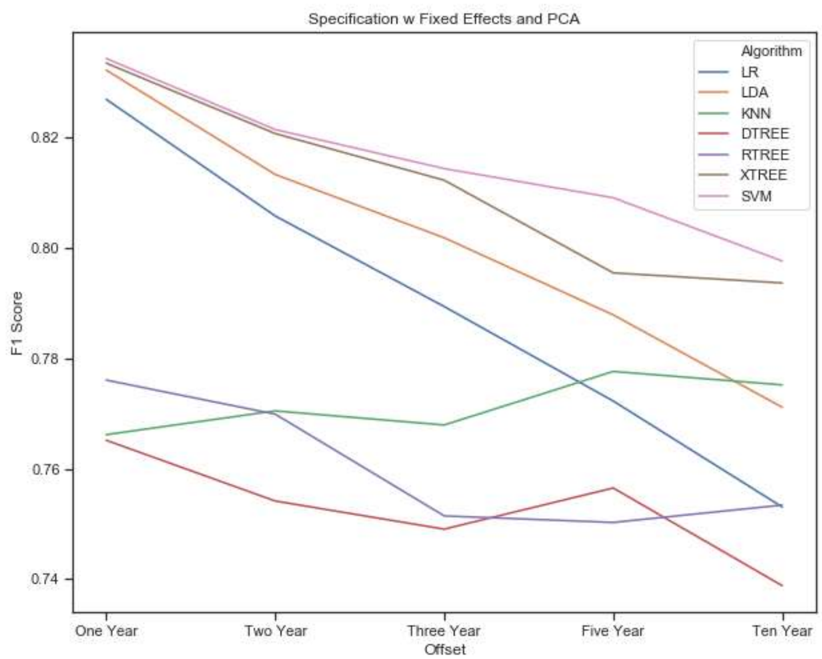


Figure 9 Continued



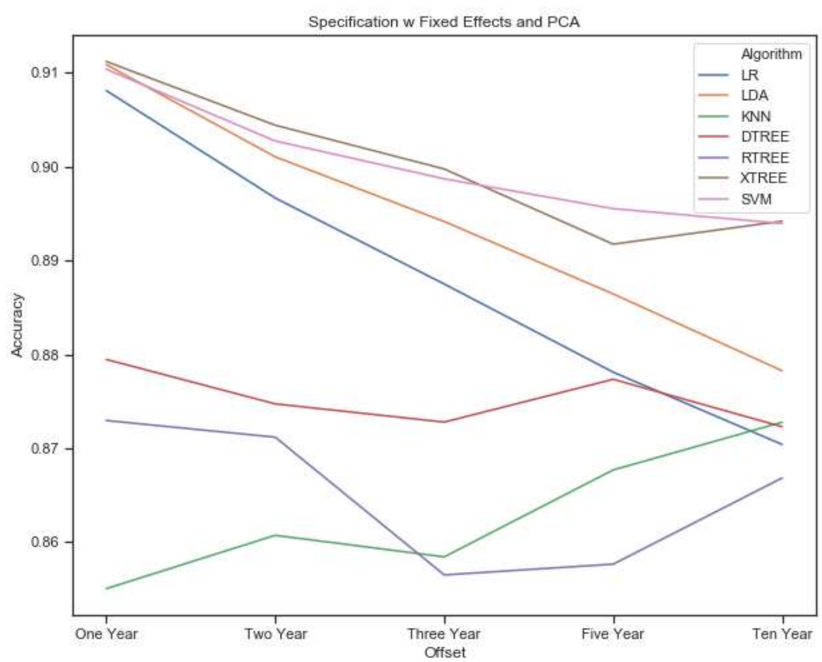


Figure 10: Neural Network Illustration

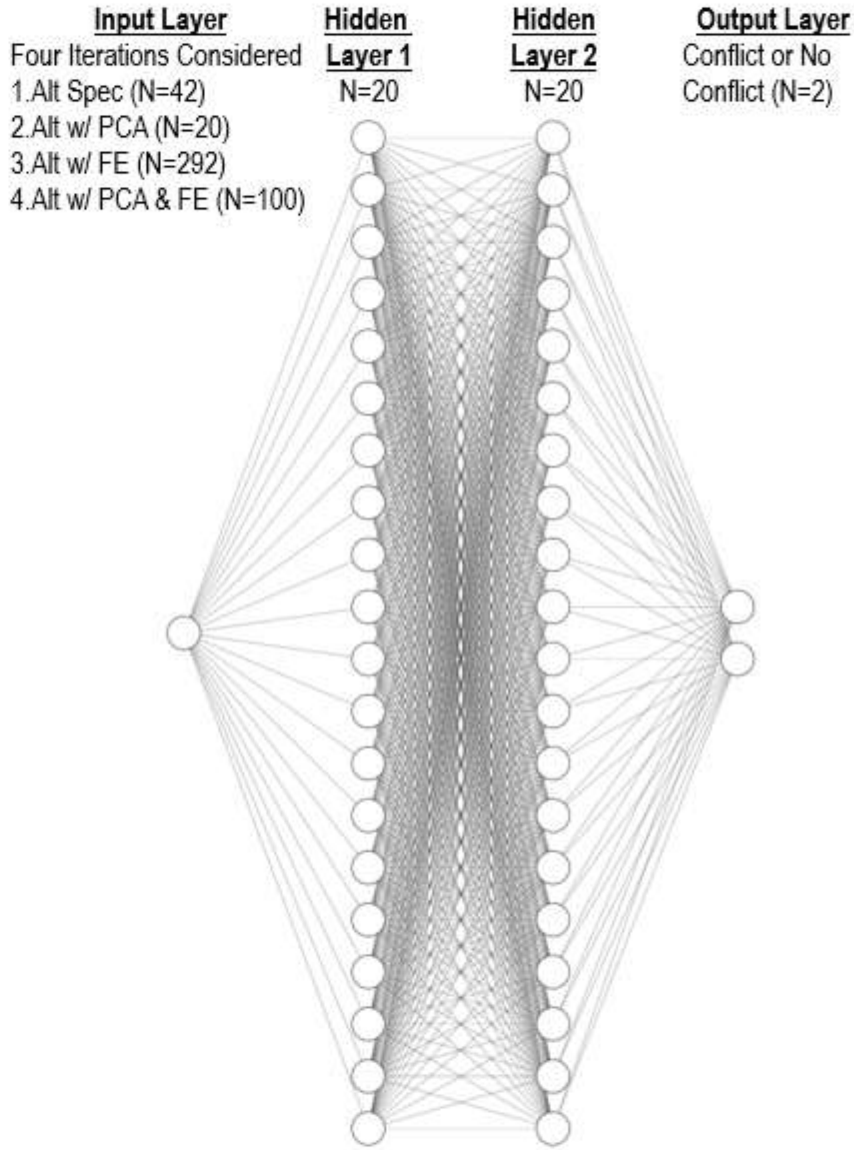


Table 13: Neural Network Results

Algorithm		Base Specification			Spec w/ PCA			Spec w/ FE			Spec w/ FE & PCA		
		Recall	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*	*Recall*	*F1*	*Accuracy*
NN	*Score*	0.809	0.811	0.903	0.816	0.820	0.908	0.794	0.825	0.913	0.840	0.837	0.916
	Std Dev	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)	(0.003)	(0.003)	(0.003)

Algorithms Used: NN: Neural Network with Dense Connections and two hidden layers of 20 nodes

PCA: Principal Component Analysis accounting for 80% of explained variance

FE: Fixed effects for country, region, and year included

Figure 11: 2019 Forecast of Overall Conflict

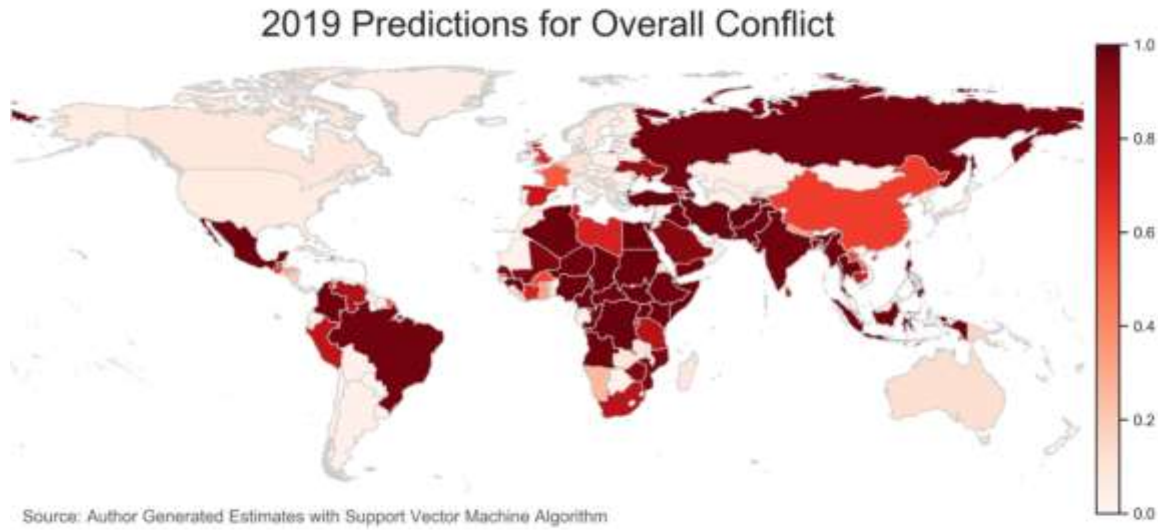


Figure 12: 2020 Forecast of Overall Conflict

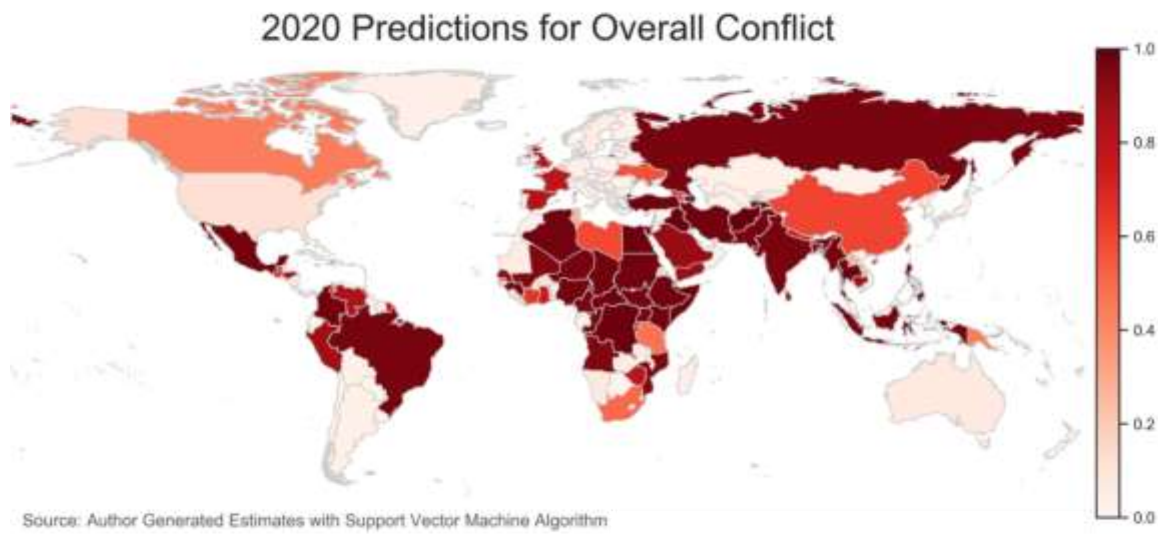


Figure 13: 2021 Forecast of Overall Conflict

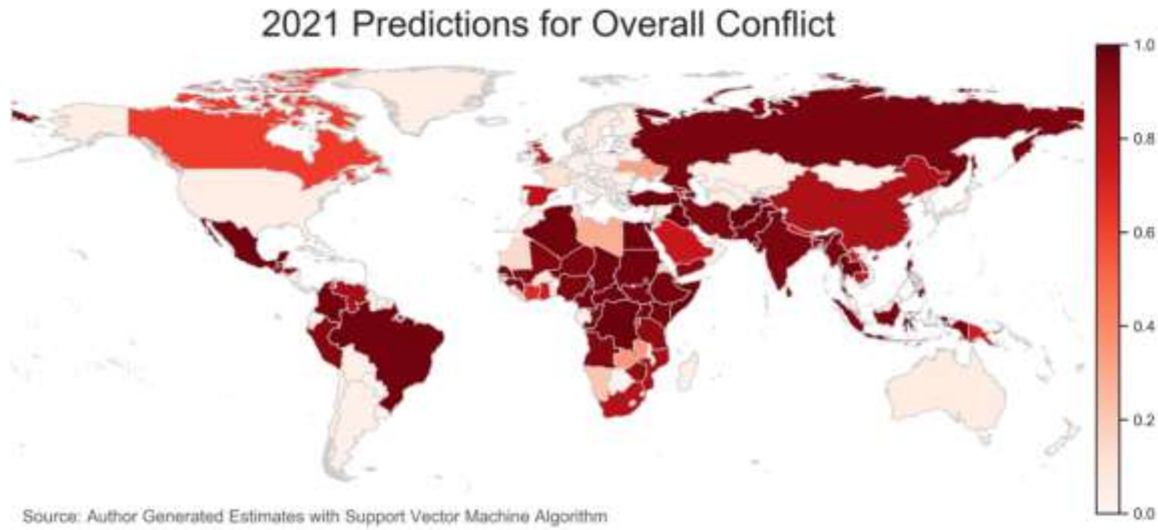


Figure 14: 2023 Forecast of Overall Conflict

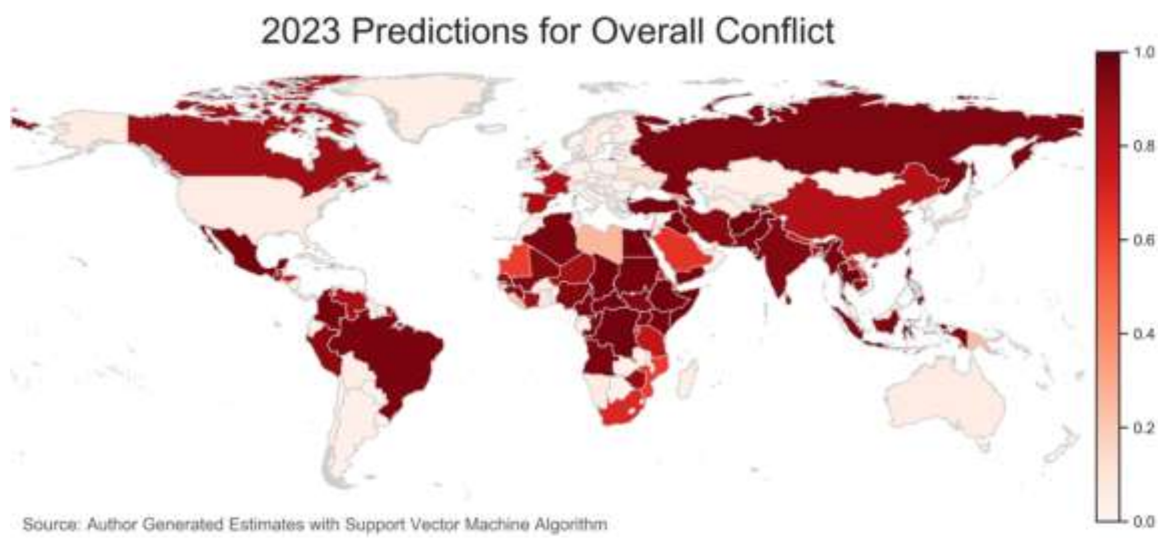
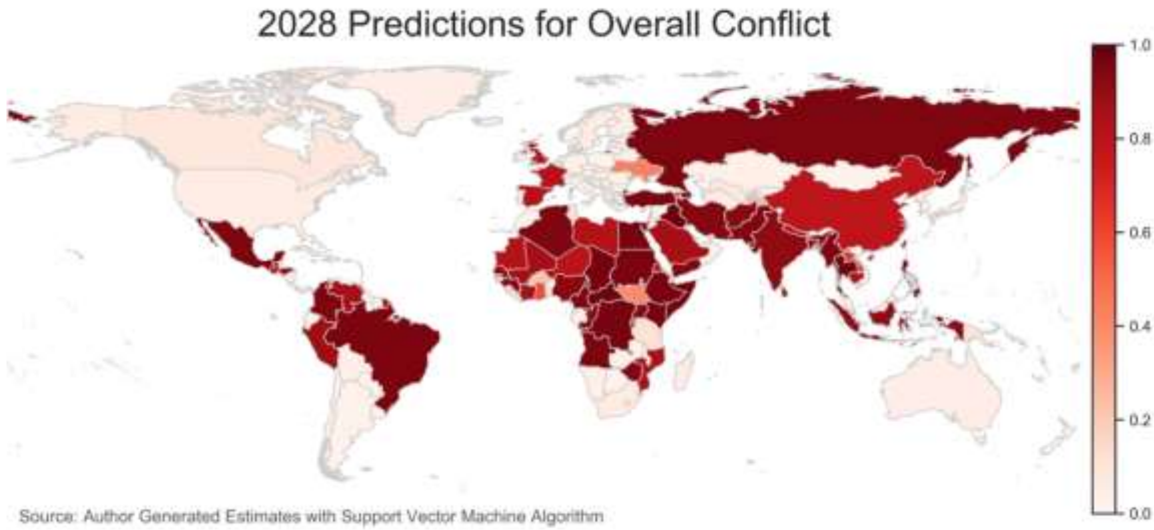


Figure 15: 2028 Forecast of Overall Conflict



References

Abouharb, M. Rodwan, and Anessa L. Kimball. 2007. "A New Dataset on Infant Mortality Rates, 1816-2002." *Journal of Peace Research* 44 (6): 743–54. https://doi.org/10.1177/0022343307082071.

Abraham, Ajith, and Shaza M. Abd Elrahman. 2013. "A Review of Class Imbalance Problem." *Journal of Network and Innovative Computing* 1: 332–40. www.mirlabs.net/jnic/index.html.

Arezki, Rabah Brückner, Markus. 2011. "Food Prices , Conflict , and Democratic Change." 2011–04. Adelaide.

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.

Busby, Joshua. 2018. "Taking Stock: The Field of Climate and Security." *Current Climate Change Reports*. https://doi.org/10.1007/s40641-018-0116-z.

Busby, Joshua W, Clionadh Raleigh, Todd G Smith, Roudabeh Kishi, Nisha Krishnan, Charles Wight, and Management Systems International. 2018. "The Intersection of Global Fragility and Climate Risks," no. September: 70. https://pdf.usaid.gov/pdf_docs/PA00TBFH.pdf.

Cederman, Lars Erik, and Nils B. Weidmann. 2017. "Predicting Armed Conflict: Time to Adjust Our Expectations?" *Science*. American Association for the Advancement of Science. https://doi.org/10.1126/science.aal4483.

Chadefaux, Thomas. 2017. "Conflict Forecasting and Its Limits." *Data Science* 1 (1–2): 7–17. https://doi.org/10.3233/ds-170002.

Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. "SMOTE: Synthetic Minority Over-Sampling Technique." *Journal of Artificial Intelligence Research* 16: 321–57. https://doi.org/10.1613/jair.953.

Chen, Tianqi. 2020. "Random Forests in XGBoost." XGBoost 1.1.0-Snapshot Documentation. 2020. https://xgboost.readthedocs.io/en/latest/tutorials/rf.html.

Chollet, François. 2015. "Home - Keras Documentation." 2015. https://keras.io/.

Claesen, Marc, and Bart De Moor. 2015. "Hyperparameter Search in Machine Learning." In *The XI Metaheuristics International Conference*. Agadir, Morocco.

https://www.codalab.org/competitions/2321.

Collier, Paul. 2000. “Economic Causes of Civil Conflict and Their Implications for Policy.”

Collier, Paul, and Anke Hoeffler. 2000. “Greed and Grievance in Civil War.” 2355. *Conflict, Political Accountability and Aid*. https://doi.org/10.4324/9780203842256.

Collier, Paul, Anke Hoeffler, and Måns Söderbom. 2004. “On the Duration of Civil War.” *Journal of Peace Research* 41 (3): 253–73.

Fearon, James D, and David D Laitin. 2003. “Ethnicity, Insurgency, and Civil War.” *The American Political Science Review*. Vol. 97.

Hayes, Jarrod. 2011. “The Democratic Peace and the New Evolution of an Old Idea.” *European Journal of International Relations* 18 (4): 767–91. https://doi.org/10.1177/1354066111405859.

Hegre, Håvard, Joakim Karlsen, Håvard Mokleiv, Håvard Strand, and Henrik Urdal. 2011. “Predicting Armed Conflict, 2010-2050.”

Helle, Valeria, Andra-stefania Negus, and Jakob Nyberg. 2018. “Improving Armed Conflict Prediction Using Machine Learning ViEWS +.” Uppsala.

Lemaitre, Guillaume, Fernando Nogueira, and Christos K. Aridas. 2017. “Imbalanced-Learn 0.5.0 Documentation.” Imbalanced-Learn 0.5.0 Documentation. 2017. https://imbalanced-learn.readthedocs.io/en/stable/over_sampling.html.

Marshall, Monty G, Ted Robert Gurr, and Keith Jaggers. 2019. “POLITY™ IV PROJECT Political Regime Characteristics and Transitions, 1800-2018 Dataset Users’ Manual.” www.systemicpeace.org.

O’Brochta, William. 2019. “A Meta-Analysis of Natural Resources and Conflict.” *Research and Politics* 6 (1). https://doi.org/10.1177/2053168018818232.

Pedregosa, Fabian, Gael Varoquaux, Alexandre Gramfort, and Vincent Michel. 2020. “Scikit-Learn 0.22.2 Documentation.” Scikit-Learn 0.22.2 Documentation. 2020. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html?highlight=random forest#sklearn.ensemble.RandomForestClassifier.

Percival, Val, and Thomas Homer-Dixon. 1996. “Environmental Scarcity and Violent Conflict : The Case of Rwanda.” *The Journal of Environment & Development* 5 (3):

270–91.

Post, Riley, Darren Hudson, Donna Mitchell, Patrick Bell, Arie Perliger, and Ryan Williams. 2016. "Rethinking the Water-Food-Climate Nexus and Conflict: An Opportunity Cost Approach." *Applied Economic Perspectives and Policy* 38 (4): 563–77. https://doi.org/10.1093/aepp/ppw027.

Probst, Philipp, Anne-Laure Boulesteix, and Bernd Bischl. 2019. "Tunability: Importance of Hyperparameters of Machine Learning Algorithms." *Journal of Machine Learning Research* 20. http://www.jmlr.org/papers/volume20/18-444/18-444.pdf.

Puga, Diego, and Nathan Nunn. 2012. "Data on Terrain Ruggedness and Other Geographical Characteristics of Countries." "Ruggedness: The Blessing of Bad Geography in Africa." 2012. https://diegopuga.org/data/rugged/#country.

Raleigh, Clionadh, and Håvard Hegre. 2009. "Population Size, Concentration, and Civil War. A Geographically Disaggregated Analysis." *Political Geography* 28 (4): 224–38. https://doi.org/10.1016/j.polgeo.2009.05.007.

Ross, Michael L. 2015. "What Have We Learned about the Resource Curse?" *Annual Review of Political Science* 18 (1): 239–59. https://doi.org/10.1146/annurev-polisci-052213-040359.

Stina, Högbladh. 2019. "UCDP GED Codebook Version 19.1." Uppsala University: Department of Peace and Conflict Research.

The World Bank. 2019. "The World Development Indicators." 2019. https://data.worldbank.org/indicator.

Urdal, Henrik. 2005. "People vs. Malthus: Population Pressure, Environmental Degradation, and Armed Conflict Revisited." *Journal of Peace Research* 42 (4): 417–34. https://doi.org/10.1177/0022343305054089.