© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719457

An Explainable Machine Learning Framework for Political Instability Prediction and Early Warning Using Decision Tree Learning

OLATUNDE DAVID AKINROLABU [1], SAMUEL OHIRENUME AFEKHIKU[2], VICTOR OLUMIDE OTOKITI[3], AYODEJI OLUSEGUN AKINWUMI[4], OMAMBALA DIVINE[5], OMOJOLA TOMIWA YUSUF[6]

[1]*Department of Data Science and Artificial Intelligence, Faculty of Computing, Adekunle Ajasin University, Akungba Akoko, Ondo State, Nigeria*

[2]*Department of Computer Science, Faculty of Computing, Adekunle Ajasin University, Akungba Akoko, Ondo State, Nigeria*

[3]*Department of Informatics and Information Systems, Faculty of Computing, Adekunle Ajasin University, Akungba Akoko, Ondo State, Nigeria*

[4]*Department of Computer Science and Information Technology, Faculty of Science and Technology, Ecole de Techniciens Superieurs du Benin, Benin Republic*

[5,6]*Department of Computer Science, Faculty of Computing Science, University of Lagos, Lagos State, Nigeria.*

Abstract- ***It is common knowledge that political instability retards the development and progress of a country. Conflict and war have been the focus of extensive research throughout history. Certain factors can reliably predict conflict, including average income, political instability, low economic growth rates, and other socio-economic indicators. Machine learning can be used to predict many things, such as stock values, weather, movie preferences, and, as is the case for this project, a country's political instability. This study presents a Decision Tree classifier trained on an integrated dataset compiled from the Centre for Systemic Peace (CSP) and the Fund for Peace (FFP), consisting of 1,778 records and 16 attributes. After preprocessing and integration, the dataset was split into training (70%) and test (30%) partitions. The implementation, built with Python's scikit-learn library, achieves 97% classification accuracy, 98.5% precision, 98.5% recall, 99% F1-score on the test set, successfully categorising countries into high-risk (class 0) and low-risk (class 1) groups for political instability. A simple decision rule based on the mean fragility score (threshold ≈8.9) was extracted from the trained tree, offering an interpretable early-warning indicator. The model demonstrates the applicability of supervised machine learning to conflict prediction and early-warning systems for policymakers.***

Keywords: Political Instability, Machine Learning, Conflict Prediction, Feature Selection, Decision Tree, Fragility Index.

I. INTRODUCTION

The intensifying frequency of civil conflict relative to interstate war over the past seven decades has made conflict prediction one of the most consequential applications of data-driven modelling in the social sciences.

The Centre for Systemic Peace documents that while interstate war has declined steadily since 1946, societal (civil) warfare rose steadily until the 1990s before declining, and at any given time, roughly 10% of all states remain in some form of civil conflict [1].

The frequency of terrorist attacks has also shown a renewed increase since 2010, after decades of decline beginning in 1980 [2], and the proportion of non-combatant to combatant deaths remains disproportionately high. Close to 65.3 million people have been displaced from their homes as a consequence of conflict, nearly double the population of Canada [3].

Civil conflict, therefore, remains a distinctly human and persistent tragedy.

A broad definition of civil conflict, drawn from the Political Instability Task Force (PITF), includes conflict-related deaths arising from political wars, ethnic wars, regime change, genocides, and politicides.

Under PITF criteria, a country is considered to be in civil conflict in a given year if any one of these events occurs, irrespective of whether the underlying cause was insurrection, terrorism, or state-directed violence against a communal or political group [4].

Machine learning, though rarely visible to the end user, already underpins many everyday digital functions, such as targeted advertising, spam filtering, and image recognition, all of which rely on trained predictive models.

Its strength lies in using large volumes of data to build models capable of making decisions or forecasts that would be difficult or impossible for a human analyst to make manually. Machine learning has been applied to the prediction of stock values, weather, and consumer preferences; this project extends that paradigm to the prediction of political instability at the country level.

Applying machine learning to civil conflict also helps adjudicate between competing theoretical models of conflict. At the most basic level, conflict can be understood through the lens of production versus predation: societies obtain resources either through specialisation and trade, which tends to improve collective welfare, or through expropriation, which is typically a zero-sum or negative-sum process in the short to medium term [5].

Investigators have long held competing hypotheses about which correlates of conflict are most predictive, but few studies test these hypotheses against genuinely unseen, future cases. A model with strong out-of-sample predictive power not only validates the underlying covariates as plausible components of a causal theory but also gives policymakers a concrete basis for choosing the policy levers most likely to reduce the likelihood of conflict.

Political instability matters because of its tangible downstream effects: sharp economic decline, job losses, capital flight, and the collapse of investor confidence.

It can trigger mass emigration of both citizens and foreign nationals as seen in Zimbabwe’s land seizures and the resulting exodus of the white farming community [6] with associated declines in agricultural output, food security, and long-term development prospects.

The motivation for this work follows directly from this gap between qualitative, annually-updated instability indices and the need for more immediate, quantitatively-grounded early warning.

Existing organisations rank countries by political fragility, but these rankings are typically annual reviews built from mixed quantitative and qualitative inputs rather than continuously updated, data-driven predictions.

This study uses data mining and decision-tree classification to construct a quantitative representation of country-level fragility, with the long-term goal of supporting an early-warning system that policymakers can use to identify at-risk states before conflict escalates.

1.1. Aim and Objectives

This project aims to build a predictive model that policymakers can use to identify countries at risk of political instability, enabling targeted attention to specific regions and potentially limiting the outbreak or spread of conflict. The specific objectives are to:

i. Collate an integrated dataset of countries’ political and socio-economic variables and attributes.
ii. Develop a Decision Tree model to classify countries as declining into political conflict or not, based on the collated dataset.
iii. Evaluate the performance of the model developed in objective II.

II. RELATED WORKS

Perry [15] applied naïve Bayes and random forest classifiers to a conflict dataset spanning 2000–2011, holding out 2012 as a final test set, with a 70/30 train–test split used for model development. A baseline model using only prior-conflict history was compared against the full feature set.

While Naïve Bayes offered only a modest improvement over the baseline, the random forest model achieved a substantially lower out-of-bag error rate (1.16%) and improved true-positive accuracy for conflict outbreaks from 24.6% to 58.5% relative to the baseline.

Perry's work represents an early demonstration that machine learning can meaningfully improve on naiıve historical-persistence baselines for conflict forecasting, though the population and GDP covariates used were interpolated from sparse five-year snapshots, limiting their granularity.

Huffman et al. [16] compared K-means, SVM, sequential minimal optimisation (SMO), and SMO regression for classifying countries by political fragility using Fragile States Index (FSI) scores. K-means clustering separated stable states (FSI < 40) cleanly but could not meaningfully distinguish among higher-risk categories.

SVM achieved high test accuracy (>97.8%) but exhibited a concerning rate of false negatives for highly unstable states; SMO achieved comparable accuracy with a lower false-negative rate and was therefore preferred, given the high cost of misclassifying an unstable state as stable. Feature-selection analysis identified healthcare access and female education as strong predictors of state fragility.

A key limitation acknowledged by the authors is that the FSI labels used for training and evaluation are themselves derived from a semi-subjective scoring methodology, and the underlying World Bank data matrix was sparsely populated for several countries. Celiku et al. [17] proposed two novel classification algorithms a" linear classifier" and a" threshold classifier" that directly minimise a prediction-loss function rather than maximising a likelihood function, and benchmarked these against probit regression and random forests on a panel of 114 developing countries since 1977.

Their threshold classifier showed competitive predictive performance, but the authors note that the approach was tested only on a modest, pre-selected set of explanatory variables, leaving open whether it would generalise to higher-dimensional feature sets, and that the external validity of the method beyond their specific dataset cannot be assumed.

Mueller and Rauh [18] introduced a text-based forecasting approach, comparing five models: two based on established covariates (rainfall shocks, economic shocks, and political-constraint indices), one based on the Goldstone et al.

global instability model, and two based on newspaper text a keyword-count model and a topic model derived from roughly 700,000 news articles, combined with ICEWS event data.

The topic model approach performed strongly for predicting conflict incidence, demonstrating that textual signals can add predictive value beyond purely structural covariates, although the analysis was restricted to countries with populations above one million, limiting coverage of smaller states.

Table 1 summarises the comparative scope of these related works alongside the approach adopted in this study.

III. STATISTICAL AND MACHINE LEARNING FOUNDATIONS

3.1. Political Instability

Political instability is the process by which the political life of a country or region undergoes sudden or destabilising change [7]. Recognised drivers include the suppression of civil rights and freedoms, corruption and mismanagement of public wealth, electoral fraud and intimidation, mass unemployment and poverty, suppression of opposition parties, lack of governmental transparency, restriction of free expression, ethnic and political prejudice, and prolonged or unconstitutional retention of power by incumbent leaders [7].

The consequences of sustained instability typically include economic decline and inflation, rising unemployment, and, in the most severe cases, armed conflict.

Table 1: Comparative Summary of Related Works in Conflict and Instability Prediction

Study	Method(s)	Key Contribution	Main Limitation
Perry (2013)	Naïve Bayes, Random Forest	Early demonstration of ML gains over a historical-persistence baseline	Sparse, interpolated population/GDP covariates
Huffman et al. (2015)	K-means, SVM, SMO, SMO Regression	Identified healthcare and education as strong fragility predictors	Reliance on semi-subjective FSI labels; sparse World Bank matrix
Celiku et al. (2017)	Linear/Threshold classifiers vs. Probit, Random Forest	Loss-function-calibrated classifiers improve predictive performance	Limited feature set; unproven external validity
Mueller & Rauh (2017)	Topic models, keyword counts, ICEWS events	Text-based signals improve conflict incidence prediction	Restricted to countries with population $>$1 million
This study	Decision Tree (CART)	Interpretable decision rule from integrated CSP/FFP dataset	Single-algorithm comparison; risk of overfitting at 100% accuracy

3.2. Supervised Machine Learning

Machine learning is the science of programming computers so that they improve at a task through exposure to data rather than explicit reprogramming [8]. Supervised learning, the paradigm used in this study, trains a model on data in which the correct output label is already known, allowing the algorithm to learn a mapping from input features to output classes [9].

Supervised problems are further divided into regression tasks, which predict continuous outputs, and classification tasks, which predict discrete class labels [10] — the present problem (high-risk vs. low-risk classification) falls into the latter category.

3.3. Feature Selection

Not all variables in a dataset contribute positively to predictive performance; irrelevant or redundant features can obscure genuine relationships in the data and degrade model quality [10].

Feature-selection techniques, including variance filtering, correlation analysis, and embedded selection within tree-based algorithms, are used to identify the subset of inputs most informative for the target variable.

3.4. Tree-Based Method

A decision tree classifies records by recursively partitioning a dataset into progressively smaller, more homogeneous subsets. At each node, the algorithm selects the single feature and threshold value that produces the greatest improvement in subset purity, as measured by an impurity criterion.

New, unseen records are classified by following the corresponding path from the root to a leaf node, and are assigned the majority class of that leaf.

Ensemble extensions of the basic decision tree bagging [11] and random forests [12] train many trees on resampled or feature-restricted subsets of the data and aggregate their predictions, generally yielding improved robustness at the cost of interpretability; this study uses a single Decision Tree (CART) for its transparency and direct interpretability as an early-warning rule.

3.5. Information Gain and the Gini Index

Decision tree induction selects splits using an impurity measure. Information gain based on Shannon entropy is defined as

$$IG(X, F) = H(X) - H(X \mid F) \tag{1}$$

where X is the target variable, F is a candidate splitting feature, H(X) is the entropy of X, and H(X|F) is the conditional entropy of X given F. Entropy itself is given by

$$H(X) = -\sum_{x \in X} p_x \log_2(p_x) \tag{2}$$

An entropy of 0 indicates a perfectly homogeneous (pure) subset, while an entropy of 1 indicates maximal class impurity in a binary problem. Equivalently, information gain can be expressed as a reduction in impurity between a parent node and its weighted children:

$$G = P - W_X \tag{3}$$

where P is the entropy of the parent node and W_X is the weighted-average entropy of the resulting child nodes. The implementation in this study uses scikit-learn’s default Gini impurity criterion, defined for a node E as

$$I_{\text{Gini}}(E) = 1 - \sum_{j=1}^{Q} p_j^2 \tag{4}$$

where p_j is the proportion of records in E belonging to class c_j, and Q is the number of classes. A split is evaluated by the resulting decrease in impurity,

$$\Delta = I(E) - \sum_{i=1}^{k} \frac{|E_i|}{|E|} I(E_i) \tag{5}$$

and the candidate split that maximises Δ across all features and thresholds is selected at each node [13].

IV. METHODOLOGY AND ARCHITECTURE

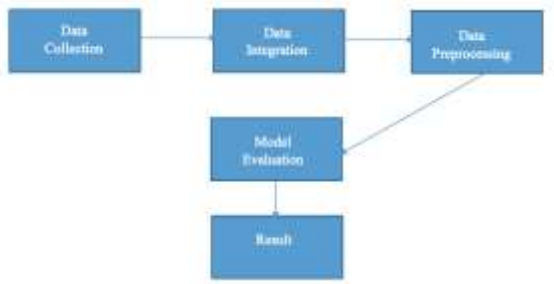

Figure 1: System Model

4.1. Data Collection and Description

Data and candidate predictive attributes were gathered and integrated from two indexing sources. Centre for Systemic Peace (CSP). Founded in 1997, CSP conducts research on political violence within the structural context of the global system [1].

Its State Fragility Index dataset provides annual fragility, effectiveness, and legitimacy scores for the world’s 167 countries with populations greater than 500,000. Each country is scored across four performance dimensions — Security, Political, Economic, and Social on both Effectiveness and Legitimacy, using a four-point fragility scale (0 = no fragility to 3 = high fragility), with the Economic Effectiveness indicator scored on a five-point scale. The combined State Fragility Index ranges from 0 (no fragility) to 25 (extreme fragility).

The CSP extract used in this study comprised 13 attributes and 200 records, summarised in Table 2.

Table 2: Centre for Systemic Peace (CSP) Dataset Attributes

Indicator	Attribute	Definition
Fragility Indices	SFI	State Fragility Index = Effectiveness Score + Legitimacy Score (25 points possible)
	Effect	Effectiveness Score = Security + Political + Economic + Social Effectiveness (13 points possible)
	Legit	Legitimacy Score = Security + Political +

		Economic + Social Legitimacy (12 points possible)
Security	Seceff / Secleg	Security Effectiveness / Legitimacy
Political	Polleff / Polleg	Political Effectiveness / Legitimacy
Economy	Ecoeff / Ecoleg	Economic Effectiveness / Legitimacy
Social	Soceff / Socleg	Social Effectiveness / Legitimacy

Fund for Peace (FFP), A US-based non-profit research institution founded in 1957, publishes the annual Fragile State Index, widely used by researchers, educators, and policymakers worldwide [14].

The index is built on the Conflict Assessment System Tool (CAST) framework, designed to measure state vulnerability across pre-conflict, active-conflict, and post-conflict situations. The FFP extract is summarised in Table 3.

Table 3: Fund for Peace (FFP) Dataset Attributes

Category	Attributes
Identifiers	Country, Year, Rank, Total
Cohesion	C1: Security Apparatus; C2: Fractionalized Elites; C3: Group Grievance
Economy	E1: Economic Decline; E2: Uneven Development; E3: Human Flight
Political	P1: State Legitimacy; P2: Public Services; P3: Human Rights
Social	S1: Demographics; S2: Refugees and IDPs; External Intervention

4.2. Data Preprocessing

Raw data from both sources were frequently incomplete or inconsistent, requiring standard preprocessing to produce an analysis-ready dataset. Categorical fields (e.g., country names) were encoded into numeric category codes, and the integrated dataset was checked for missing or malformed entries before training. Splitting criteria for the Decision Tree itself were evaluated using the information-gain and Gini-impurity formulations described in Section 3.5.

4.3. Data Integration

The CSP and FFP extracts were merged into a single coherent dataset in Microsoft Excel, producing an integrated dataset of 1,778 records and 16 attributes. Integration across two independent indexing sources was used specifically to mitigate the missing-data limitations noted in each individual source.

4.4. Decision Tree Algorithm

The classification pipeline followed four broad stages, illustrated in Figure 1: data collection, data integration, data preprocessing, and model training/evaluation, culminating in the classification result.

The CART-style induction procedure used is summarised in Algorithm 1.

```
Algorithm 1. Decision Tree Induction (CART)
Input: Training set T with attributes A_1…, A_m and target class C
Output: Trained decision tree
1. Place the attribute yielding the greatest impurity reduction at the root of the tree.
2. Partition the training set into subsets such that each subset shares the same value (or threshold side) for the selected attribute.
3. Recurse on each subset, repeating steps 1–2, until all leaves are pure or a stopping criterion (e.g. minimum samples per leaf) is met.
4. Return the fully grown tree.
```

For a continuous attribute A, a split test of the form A≤x partitions the record set T into

$$T_l = \{t \in T: t(A) \leq x\}, \qquad T_r = \{t \in T: t(A) > x\} \tag{6}$$

and the quality of the split is evaluated using the impurity-reduction criterion Δ from Equation (5). The classifier was instantiated in scikit-learn as:

```
DecisionTreeClassifier(criterion='gini',
random_state=100,
max_depth=None, min_samples_split=1,
min_samples_leaf=1)
```

Criterion controls the impurity measure used for evaluating splits (Gini index by default; information gain is also supported). Max_depth bounds the depth of the tree; when unset, nodes expand until all leaves are pure or fall below the minimum split size.

Min_samples_split sets the minimum number of samples required to split an internal node, and min_samples_leaf sets the minimum number of samples that must remain in each branch after a split, which has a smoothing effect that can mitigate overfitting.

V. RESULTS AND DISCUSSION

5.1. Dataset Overview

The integrated dataset comprised 1,778 records and 16 attributes following the merge of the CSP and FFP extracts. The data was partitioned into a 70% training set and a 30% test set, consistent with standard supervised-learning practice for model development and evaluation.

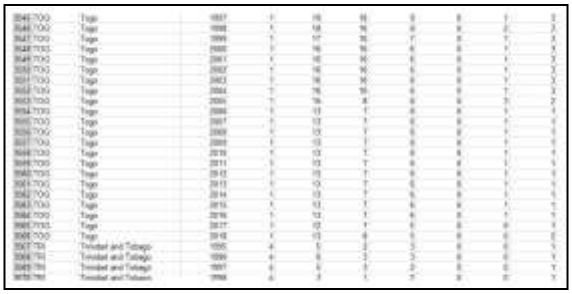

Figure 2: Dataset Overview

5.2. Model Training and Testing

The Decision Tree classifier was trained on the 70% training partition using the Gini impurity criterion, and evaluated on the held-out 30% test partition. Model performance was assessed using an accuracy score, a confusion matrix, and a full classification report.

Table 4: Classification Report

Class	Precision	Recall	F1-Score	Support
0 (High Risk)	0.98	0.99	0.99	142
1 (Low Risk)	0.99	0.98	0.99	391
Accuracy			0.97	533
Macro Avg	0.98	0.98	0.98	533
Weighted Avg	0.98	0.98	0.97	533

The total error rate was computed as the sum of off-diagonal (misclassified) entries in the confusion matrix divided by the total number of test samples:

$$\text{Total Error Rate} = \frac{13}{533} = 0.0243$$

The model achieved a classification accuracy of 97.5% on the test set, with thirteen misclassified records.

5.3. Decision Rule Extraction

A decision rule was extracted directly from the trained tree's root split, expressed as a simple, human-interpretable IF–THEN statement based on the mean fragility score (MEAN):

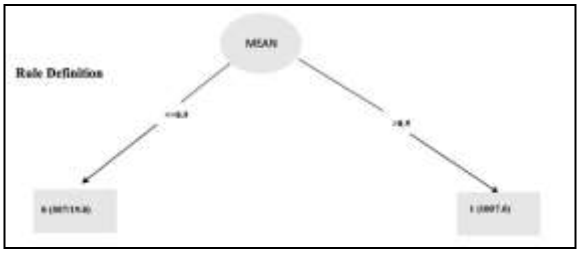

Figure 3: Decision Rule Set for the Model

If "MEAN"≤8.9 then predict class 0 (high risk) — 771 records, 11 misclassified at this node.
If "MEAN">8.9 then predict class 1 (low risk) — 1007 records, 21 misclassified at this node. This single-threshold rule offers a transparent and immediately actionable early-warning indicator: countries whose mean fragility score across the CSP/FFP indicators falls at or below 8.9 are flagged as high risk of political instability (class 0), while those above the threshold are classified as low risk (class 1).

5.4. Discussion

The model's 97.5% test accuracy and 13 total misclassifications, as measured by the error rate, indicate that the integrated CSP/FFP feature set is highly separable under a Decision Tree classifier for this dataset.

This result is consistent with the broader literature: Huffman et al. [16] similarly reported very high SVM/SMO test accuracy (90%) on Fragile States Index-derived labels, and Perry [15] reported substantial accuracy gains for tree-based ensemble methods over simpler baselines. An almost perfect accuracy score on a single train/test split should, however, be interpreted cautiously.

It may reflect genuine separability in the fragility-index feature space (since the target class is itself partly derived from threshold-based aggregations of the same indicator family), or it may indicate a degree of overfitting or information leakage between predictors and the labelling criterion.

Celiku et al. [17], who explicitly cautioned against assuming external validity of a method beyond its original dataset, no k-fold cross-validation was performed in this implementation, which would be a natural next step to confirm that the observed accuracy generalises beyond the specific 70/30 split used here.

Compared with ensemble approaches such as random forests [12] and bagging [11], a single Decision Tree sacrifices the variance-reduction benefits of model averaging in exchange for full interpretability.

The extracted MEAN ≤ 8.9 decision rule can be communicated directly to non-technical policymakers, which is a meaningful advantage for an early-warning application, even if ensemble methods might offer marginally better generalisation.

CONCLUSION

This study presented a Decision Tree-based predictive model for classifying countries into high-risk and low-risk categories of political instability, using an integrated dataset of 1,778 records and 16 attributes drawn from the Centre for Systemic Peace and the Fund for Peace.

The model, implemented using scikit-learn, a framework for supervised learning, achieved 97.5% classification accuracy, 98.5% recall, 98.5% precision, 99% F-measure and a 0.0241 error rate, respectively, on a held-out test set and yielded a simple, interpretable decision rule based on a mean fragility score threshold of 8.9.

These results demonstrate that supervised machine learning, even with a single, fully interpretable algorithm, can support the development of early-warning indicators for policymakers seeking to anticipate and mitigate civil conflict. The integration of two independently compiled indexing sources was a deliberate design choice to reduce the impact of missing data in either source.

The perfect test accuracy obtained should be read as an upper bound under the specific train/test split used rather than a guarantee of real-world predictive power; cross-validation and testing against more recent, out-of-sample country-years would be necessary to confirm generalizability.

Future extensions of this work should evaluate ensemble tree methods (random forests, gradient boosting) and incorporate textual or event-based covariates, in line with approaches such as Mueller and Rauh [18], to assess whether predictive performance can be maintained while improving robustness against overfitting.

REFERENCES

[1] Centre for Systemic Peace, "Global Trends in Armed Civil Conflict, 1946–2015," Centre for Systemic Peace, 2016. [Online]. Available: https://www.systemicpeace.org

[2] J. T. Bang, A. Basuchoudhary, A. David, and A. Mitra, "Predicting Terrorism: A Macro-Level TimeSeries Approach," Defence and Peace Economics, 2016.

[3] World Bank, The Changing Wealth of Nations: Measuring Sustainable Development in the New Millennium, World Bank, 2011.

[4] Political Instability Task Force (PITF), "Worldwide Atrocities Dataset and Coding Manual," Centre for Systemic Peace.

[5] F. Caselli and W. J. Coleman, "On the Theory of Ethnic Conflict," Journal of the European Economic Association, vol. 11, no. s1, pp. 161–192, 2013.

[6] H. Mueller, “Growth and Violence: Argument for a Per Capita Measure of Civil War,” Economica, vol. 83, no. 331, pp. 473–497, 2016.

[7] Rishi, “What Leads to Political Instability,” B.A. Political Science & Sociology, University of Jammu, 2018. [Online]. Available: https://www.quora.com/ What-leads-to-political-instability

[8] A. Geron, ´ Hands-On Machine Learning with Scikit-Learn and TensorFlow, 1st ed. Sebastopol, CA, USA: O’Reilly Media, 2017.

[9] N. Castle, “Supervised vs. Unsupervised Machine Learning Algorithms,” DataScience.com, 2018. [Online]. Available: https://www.datascience.com/blog/ supervised-and-unsupervised-machine-learning-algorithms

[10] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, Springer, vol. 112, 2013.

[11] L. Breiman, “Bagging Predictors,” Machine Learning, vol. 24, no. 2, pp. 123–140, 1996.

[12] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[13] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[14] Fund for Peace, “Fragile States Index Methodology (CAST Framework),” Fund for Peace, 1957– present. [Online]. Available: http://ffp.statesindex.org 9

[15] C. Perry, “Machine Learning and Conflict Prediction: A Use Case,” Stability: International Journal of Security and Development, vol. 2, no. 3, pp. 1–18, 2013.

[16] B. Huffman et al., “Predicting High-Risk Countries for Political Instability and Conflict,” Stanford University CS229 Project Report, 2015.

[17] B. Celiku et al., “Predicting Conflict,” World Bank Policy Research Working Paper, 2017.

[18] H. Mueller and C. Rauh, “Reading Between the Lines: Prediction of Political Violence Using Newspaper Text,” American Political Science Review, vol. 112, no. 2, pp. 358–375, 2018. (Originally CEPR Discussion Paper No. 11516, 2016.) [19] H. Hegre, H. Buhaug, K. Calvin, J. Nordkvelle, S. Waldhoff, and E. Gilmore, “Forecasting Civil Conflict Along the Shared Socioeconomic Pathways,” Environmental Research Letters, vol. 11, no. 5, pp. 1–8, 2016.

[20] L.-E. Cederman and N. B. Weidmann, “Predicting Armed Conflict: Time to Adjust Our Expectations?” Science, vol. 355, no. 6324, pp. 474–476, 2017.

[21] M. Colaresi and Z. Mahmood, “Do the Robot: Lessons from Machine Learning to Improve Conflict Forecasting,” Journal of Peace Research, vol. 54, no. 2, pp. 193–214, 2017.

[22] D. Muchlinski, D. Siroky, J. He, and M. Kocher, “Comparing Random Forest with Logistic Regression for Class-Imbalanced Civil War Onset Data,” Political Analysis, vol. 24, pp. 87–103, 2016.

[23] M. D. Ward, N. W. Metternich, C. L. Dorfi, M. Gallop, F. M. Hollenbach, A. Schultz, and S. Weschle, “Learning from the Past and Stepping into the Future: Toward a New Generation of Conflict Prediction,” International Studies Review, vol. 15, no. 4, pp. 473–490, 2013.

[24] J. A. Goldstone et al., “A Global Model for Forecasting Political Instability,” American Journal of Political Science, vol. 54, no. 1, pp. 190–208, 2010. [25] W. McKinney, “pandas: A Foundational Python Library for Data Analysis and Statistics,” Python for High Performance and Scientific Computing, pp. 1–9, 2011.