RESEARCH

Understanding peacefulness through the world news

Vasiliki Voukelatou[1*], Ioanna Miliou[2], Fosca Giannotti[3] and Luca Pappalardo[3]

[*]Correspondence: vasiliki.voukelatou@sns.it
[1]Scuola Normale Superiore, Pisa, Italy
Full list of author information is available at the end of the article

Abstract

Peacefulness is a principal dimension of well-being for all humankind and is the way out of inequity and every single form of violence. Thus, its measurement has lately drawn the attention of researchers and policy-makers. During the last years, novel digital data streams have drastically changed the research in this field. In the current study, we exploit information extracted from a new digital database called Global Data on Events, Location, and Tone (GDELT), to capture peacefulness through the Global Peace Index (GPI). Applying predictive machine learning models, we demonstrate that news media attention from GDELT can be used as a proxy for measuring GPI at a monthly level. Additionally, we use the SHAP methodology to obtain the most important variables that drive the predictions. This analysis highlights each country's profile and provides explanations for the predictions overall, and particularly for the errors and the events that drive these errors. We believe that digital data exploited by Social Good researchers, policy-makers, and peace-builders, with data science tools as powerful as machine learning, could contribute to maximize the societal benefits and minimize the risks to peacefulness.

Keywords: well-being; peacefulness; Global Peace Index; GDELT; news; Data Science for Social Good; SHAP

1 Introduction

Measuring well-being for adequate policy making is a crucial objectives of every country, as high or low levels of well-being reflect a country's societal progress [1, 2]. Considering that well-being is a vague and multi-dimensional concept, it cannot be captured as a whole, but through a set of health, socio-economic, safety, environmental, and political dimensions [3, 4].

The United Nations Development Programme (UNDP) embodies the aforementioned dimensions into 17 Sustainable Development Goals (SDGs) [5, 6, 7], such as "No Poverty", "Reduced inequalities", "Peace, Justice, and Strong Institutions". Apart from the SDGs, the UNDP has created the Human Development Index (HDI) [8] to capture well-being and human development. Similarly, the Organisation for Economic Co-operation and Development (OECD) created the Better Life Index (BLI) [3] to capture various well-being dimensions. Another prominent well-being tool is the Global Peace Index (GPI) [9], which specifically captures peacefulness in all countries around the world.

Well-being indexes, such as the GPI, are traditionally captured by institutional surveys and governmental data, which are usually expensive and time-consuming. Additionally, these indexes are generally determined with a lag, and final estimates are produced only after a series of revisions, usually a few months later. The revolution of digital data and artificial intelligence may help overcome the aforementioned difficulties providing cost-efficient, time-efficient, and almost real-time estimates of well-being. This was also highlighted by the United Nations, in 2014, that recognized the importance of harnessing the data revolution to put the best available tools and methods to work in achieving the SDGs [10].

Therefore, the objective of this study is to measure well-being in terms of peacefulness through the estimation of GPI [9] at a higher time frequency, as compared to the official annual GPI score. To tackle this task, we exploit machine learning and information extracted from a digital data source called Global Data on Events, Location, and Tone (GDELT) [11]. In particular, we use news media attention from GDELT as a proxy for estimating GPI, to complement the knowledge obtained from the traditional data sources and overcome their limitations. Considering that GDELT is free access database updated daily, it can contribute to the estimation of GPI at a higher frequency as compared to the official GPI, which is updated at a yearly level. Besides, GPI though GDELT is produced at a low cost and in a time-efficient way, as compared to the traditional methodology. In this study, we expand the approach introduced in [12].

Our results demonstrate that GDELT variables are a good proxy for measuring GPI at a monthly level and that news media attention from GDELT can capture GPI from 1-month-ahead up to 6-months-ahead. There are countries for which the model's performance is high, such as the United Kingdom, and others for which the model's performance is low, such as Estonia. The reasons for the low model's performance could be the limited GDELT news coverage for some countries. In addition, we use explainable AI techniques [13] to identify the variables that most affect peacefulness. Through these variables analysis we highlight each country's profile. For example, the most important variables for the United States indicate a powerful country in military, socio-economic, and political terms. In contrast, the most important variables for Iceland denote a peaceful country. We also identify the events that drive the errors in the predictions, which is crucial since GDELT events might reveal signals that the official GPI values could neglect.

Frequent estimation updates of the GPI score through the GDELT database could be beneficial for researchers, policy-makers, and peacekeeping organizations, such as the United Nations and its agencies UNESCO and UNICEF. In particular, almost real-time GPI estimations can reveal considerable month-to-month peacefulness fluctuations, and significant events that would be otherwise neglected. As a consequence, peace-makers could be additionally empowered to timely react on applying adequate policies, preventing detrimental societal effects, and contributing effectively to social well-being and progress.

The remainder of the paper is structured as follows. Section 2 presents an overview of the literature in peacefulness indicators and research conducted with the use of the GDELT data. Section 3 describes the datasets used for the study, the prediction models, the estimation framework, and the SHAP methodology. Section 4 presents and analyzes the results. Finally, Section 5 discusses the conclusions derived from the research, limitations, and future work.

2 Related Works

Peacefulness is traditionally captured with official data, such as surveys and socioeconomic data [14, 15, 16]. In assessing peacefulness, the GPI explores the ongoing domestic and international conflicts, and militarisation, at a country and yearly level. It also seeks to determine a nation's level of harmony or disagreement, through indicators that evaluate safety and security in society (see **A.1 Indicators of GPI** or [17] for a detailed list of the GPI indicators). For example, low level of violent crime, low number of homicides, low presence of police forces, and harmonious relations with neighboring countries can be suggestive of peacefulness [18].

With the growth of technology, researchers are inclined to use new data sources to measure the aforementioned GPI indicators, as an alternative or complement to traditional data. To begin with, social media, such as Twitter, have been primarily used to assess public safety, external conflicts, foreign policy, and migration phenomena, as they render individuals' online activities accessible for analysis. Given this enormous potential, researchers use social media data to predict crime rates or detect the fear of crime [19, 20, 21, 22] and to track civil unrest and violent crimes [23, 24, 25, 26, 27]. Similarly, Twitter data are used to study early detection of the global terrorist activity [28], military conflicts in Gaza Strip [29, 30], and foreign policy discussions between Israel and Iran [31]. In addition, social media data are useful in estimating turning points in migration trends [32] and stocks of migrants [33, 34]. Finally, researchers have created a French corpus of tweets annotated for event detection, such as conflict, war and peace, crime, and justice [35].

Besides social media data, many researchers use mobility data, such as mobile phone records and GPS traces [36, 37, 38, 39, 40], in combination with traditional data, to predict and prevent crime [41, 42, 43, 44, 45], compare how the different factors correlate with crime in various cities [46], and estimate deprivation and well-being [47, 48, 49, 50]. In addition, researchers combine social media data with phone records to infer migration events and population [51, 52, 53, 54, 40] and use GPS data, combined with subjective and objective data, to study perceived safety [55].

Additionally, the volume and momentum of web search queries, such as Google Trends, provide useful indicators of periods of civil unrest over several countries [56, 57], and contribute in capturing a decline in domestic violence calls per capita when immigration enforcement awareness increases [58].

Moreover, crowdsourced data are used to map violence against women [59], for police-involved killings [60], and for analyzing the international crisis between India and Pakistan for the dispute over Kashmir [61]. Crowdsourced data are also used for creating an efficient model in preventing crime events and emergency situations [62], as well as to capture the fear of crime [63].

Recently, researchers have started exploring remote sensing data, such as satellite images, to map refugee settlements [64, 65], to study conflicts, in particular in zones where field observations are sparse or non-existent [65], ethnic violence [66], and humanitarian crises [67].

Finally, researchers combine conflict-related news databases such as ACLED [68] with other traditional data to capture peace indicators and measure conflict risks [69, 70], or demonstrate the relatively short-term decline in conflict events during the COVID-19 pandemic [71], or such as Arabia Inform [72] to extract variables for generating military event forecasts [73].

GDELT is another major news data source, yet barely explored. It describes the worldwide socioeconomic and political situation through the eyes of the news media, making it an ideal data source for measuring well-being indexes and indicators related to peacefulness.

GDELT is mostly used to explore social unrest, protests, civil wars and coups, crime, migration, and refugee patterns. Many researchers try to explain and predict social unrest events in several geographic areas around the world, such as in Egypt [74], in Southeast Asia [75], in the United States [76], in Saudi Arabia [77], or recognize social unrest patterns in the countries of India, Pakistan, and Bangladesh [78], or reveal the causes or evolution of future social unrest events in Thailand [79]. GDELT is a valuable source of data for the detection of protest events [80] and violence related social issues [81], as well as for detecting and forecasting domestic political crises [82]. GDELT is also used for the exploration of severe internal and external conflicts, such as the Sri Lankan civil war, the 2006 Fijian coup [83], the Afganistan violence events [84], and helps understand the direct cooperative and conflictual interactions between the dyads of China-Russia, Russia-USA and USA-China since the end of the Cold War [85]. Last, news data from GDELT are combined with other data sources, such as socioeconomic indicators [86], refugee data [87] and housing market data [88], Google Trends and official migration data [89], to analyze and produce short and medium-term forecasts of migration patterns.

The main contribution of our study is the use of GDELT to capture the monthly peacefulness as a whole, through the estimation of the GPI. The wide variety of GDELT event categories can cover most GPI indicators, and the daily updates of its data allow the GPI estimation at a higher frequency.

3 Methodology

In this Section, we describe the data used in our study, highlighting their characteristics and explaining how they are used in our analysis. Additionally, we describe the five models we use to produce the GPI estimates: Elastic Net, Decision Tree, Random Forest, Extreme Gradient Boosting, Support Vector Regression (SVR), and the process of their rolling training. We provide the data and the code of our study for reproducibility in `https://github.com/VickyVouk/GDELT_GPI_SHAP_project`.

3.1 GPI data

GPI [9] measures the relative position of nations' and regions' peacefulness. The index ranks 163 independent states and territories according to their level of peacefulness, and it is created by the Institute for Economics & Peace (IEP). GPI score data are available from 2008 until 2020 at a yearly level (see, e.g., GPI report 2020 [17]). The score for each country is continuous, normalized on a scale of 1 to 5, where the higher the score, the less peaceful a country is. For example, in 2019, Iceland has been the most peaceful country with GPI 1.072, whereas Somalia has been the least peaceful country with GPI 3.574. The index is constructed from 23 indicators related to Ongoing Domestic and International Conflict, Societal Safety and Security, and Militarisation domains (see **A.1 Indicators of GPI** or [17] for a detailed list of the indicators). These indicators are weighted and combined into one overall score. For GPI construction, data are derived from official sources, such as governmental data, institutional surveys, and military data.

For the purposes of this study, the frequency of the GPI increases from yearly to monthly data. In particular, the GPI data are upsampled linearly. Every yearly GPI value is assigned to March of the corresponding year. The upsampling is definitively an assumption, since the monthly data generated do not correspond to the real monthly GPI. However, considering that monthly data are not available, linear upsampling is the simplest assumption. After upsampling, from 13 yearly values (2008 - 2020), we obtain 145 monthly values in total (March 2008 - March 2020).

3.2 GDELT data

GDELT [11] is a publicly available digital news database related to socio-political events, and it is supported by Google. In particular, it is a collection of international English-language news sources, such as Associated Press, The New York Times, etc. GDELT data are based on news reports coded with the Tabari system [90], which extracts the events from the media and assigns the corresponding code to each event. Events are coded based on an expanded version of the dyadic CAMEO format, a conflict, and mediation event taxonomy [91]. GDELT compiles a list of 200 categories of events, from riots and protests to peace appeals and diplomatic exchanges, from public statements and consulting to fights and mass violence. Examples of identified events are "Express intent to cooperate", "Conduct strike or boycott", "Use conventional military force", and "Reduce or break diplomatic relations" (see **A.2 Topics of GDELT** or [91] for a detailed list of the topics covered in GDELT).

The database offers various information for each event, such as the date, location, and the URL of the news article the event is found in. We use GDELT 1.0 database, which is updated on a daily basis. Therefore data are available at a daily, monthly, and yearly frequency. Historical data are also available since 1979 [92].

In Figure 1 we present an example of the number of events related to engagement in political dissent, such as civilian demonstrations, derived from the GDELT news on the United States, from the middle of December 2020 to the middle of January 2021. We also present three examples of news articles published on the 6th and 7th of January. The plot depicts a noticeable rise in these events on the 6th of January 2021, the day of the "Storming of the United States Capitol", and a peak of news related to the topic on the 7th of January 2021, demonstrating that GDELT news can depict the worldwide sociopolitical and conflictual reality.

For the prediction of GPI, we derive several variables from GDELT. These variables correspond to the total number of events (No. events) of each GDELT category at country and monthly level. On average, the total number of variables per country is 87, varying from 25 to 141. This indicates that some event categories may not be present in the news of a country. We use the BigQuery [93] data manipulation language (DML) in the Google Cloud Platform to extract the GDELT variables (see Listing 1).

Listing 1 Query for the extraction of GDELT variables.

```
SELECT ActionGeo_CountryCode,MonthYear,EventBaseCode,
COUNT(EventBaseCode) AS No_events,
FROM 'gdelt-bq.full.events'
WHERE(MonthYear>200802)AND(MonthYear<202004)
AND(ActionGeo_CountryCode<>'null')
GROUP BY ActionGeo_CountryCode,MonthYear,EventBaseCode
ORDER BY ActionGeo_CountryCode,MonthYear,EventBaseCode
```

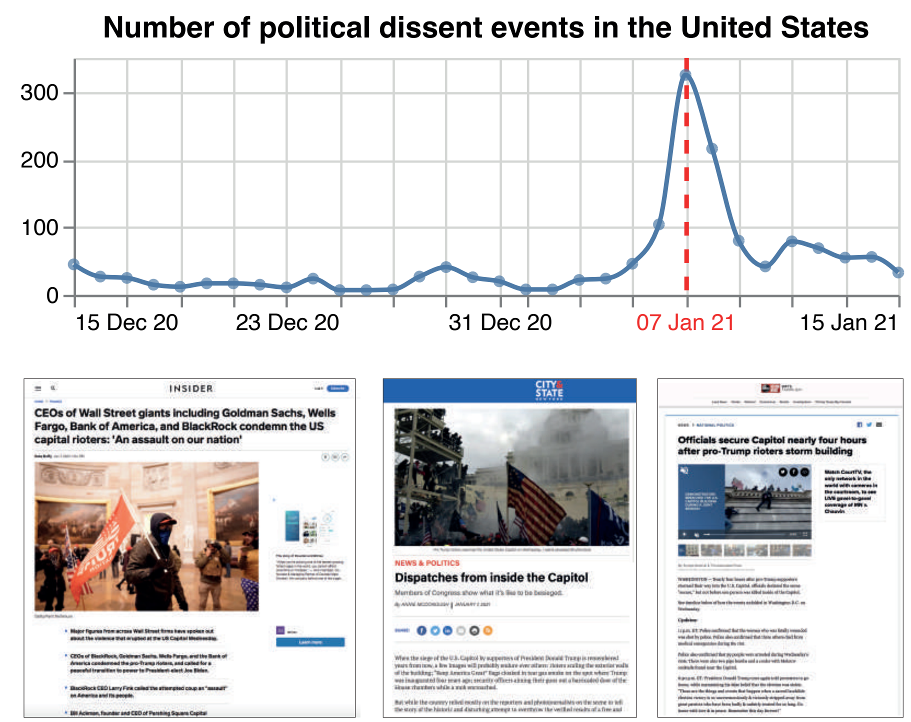

Figure 1 Number of political dissent events in the United States. Daily number of political dissent events (blue curve) derived from the GDELT news in the United States, from the middle of December 2020 to the middle of January 2021, and three examples of news articles published on the 6th and 7th of January. GDELT depicts a noticeable rise of the events related to political dissent on the 6th of January 2021, the day of the "Storming of the United States Capitol", and a peak of news related to the topic on the 7th of January 2021 (vertical dashed red line).

Table 1 shows some random examples of the United States variables in February and March 2018. For example, in February 2018, the No. events for the event category "Investigate crime" is 680, and in March 2018 is 799. In February 2018, the No. events for the event category "Conduct suicide, car, or other non-military bombing" is 523, and in March 2018 is 1099. The latter variable's value has increased a lot from February to March 2018. This is explained by the "Austin serial bombings" which occurred between March 2 and March 22, 2018, mostly in Austin, Texas, where in total, five package bombs exploded.

3.3 Prediction models

Models handling time series are used to predict future values of indices by extracting relevant information from historical data. Traditional time series models are based on various mathematical approaches, such as autoregression. Autoregressive models specify that the output variable depends linearly on its previous values and a stochastic term. Considering that our data are upsampled linearly, it is not feasible to apply autoregressive models, because of the linear relationship between the dependent variable (GPI) and its past values. Besides, our objective is not only to measure GPI, but also to understand and explain how different peacefulness topics captured by GDELT, such as militarisation, contribute to the GPI measurement.

Table 1 Examples of the United States variables in February and March 2018. The event code and category that describe the event are reported. The No. events that occurred are also presented.

Event Code	Event Category	No.events	Date
⋮	⋮	⋮	⋮
022	Appeal for diplomatic cooperation	2168	2018/02
091	Investigate crime	680	2018/02
122	Reject, request or demand for material aid	501	2018/02
183	Conduct suicide, car, or other non-military bombing	523	2018/02
⋮	⋮	⋮	⋮
022	Appeal for diplomatic cooperation	2561	2018/03
091	Investigate crime	799	2018/03
122	Reject, request or demand for material aid	534	2018/03
183	Conduct suicide, car, or other non-military bombing	1099	2018/03
⋮	⋮	⋮	⋮

We use Elastic Net, Decision Tree, Random Forest, Extreme Gradient Boosting, and Support Vector Regression models, to describe the relationship between the GPI score and the GDELT variables at a country level. Specifically, the aim is to develop GPI estimates at least one month in advance of the latest ground-truth GPI value.

Elastic Net

Elastic Net is a regularized and variable selection regression method. One of the essential advantages of Elastic Net is that it combines penalization techniques from the Lasso and Ridge regression methods into a single algorithm [94]. Lasso regression penalizes the sum of absolute values of the coefficients (L1 penalty), Ridge regression penalizes the sum of squared coefficients (L2 penalty), while Elastic Net imposes both L1 and L2 penalties. This means that Elastic Net can completely remove weak variables, as Lasso does, or reduce them by bringing them closer to zero, as Ridge does. Therefore, it does not lose valuable information, but still imposes penalties to lessen the impact of certain variables.

Decision Tree

Decision trees are used to visually and explicitly represent decisions, in the form of a tree structure. A Decision Tree is called regression tree when the dependent variable takes continuous values [94]. The goal of using a Regression Decision Tree is to create a training model that can predict the value of the dependent variable by learning simple decision rules inferred from the training data. In particular, Decision Tree divides the dataset into smaller data groups, while simultaneously, an associated decision tree is incrementally developed. The final tree consists of decision nodes and leaf nodes. A decision node has two or more branches, each representing values for the variable tested. A leaf node represents a decision on the value of the dependent variable. The topmost decision node, called the root node, corresponds to the most important variable.

Random Forest

Random Forest limits the risk of a Decision Tree to overfit the training data [94]. As the names “Tree” and “Forest” imply, a Random Forest Regression is essentially a collection of individual Regression Decision Trees that operate as a whole. A Decision Tree is built on the entire dataset, using all the variables of interest. On the contrary, Random Forest builds multiple Decision Trees from randomly selecting observations and specific variables and then combines the predictions into a single model. Individually, predictions made by Decision Trees may not be accurate, but combined, are, on average, closer to the true value.

Extreme Gradient Boosting (XGBoost)

XGBoost [95] is a scalable machine learning regression system for tree boosting. It uses a gradient descent algorithm and incorporates a regularized model to prevent overfitting. Comparing to Random Forest that builds each tree independently and combines results at the end of the process, XGBoost builds one tree at a time and combines results along the way. In particular, XGBoost corrects the previous mistakes made by the model, learns from it and its next step enhances the performance until there is no scope of further improvements. Its main advantage is that it is fast to execute and gives high accuracy.

Support Vector Regression (SVR)

SVR [96] is a regression learning approach which, comparing to other regression algorithms that try to minimize the error between the real and predicted value, uses a symmetrical loss function that equally penalizes high and low misestimates. In particular, it forms a tube symmetrically around the estimated function (hyperplane), such that the absolute values of errors less than a certain threshold are penalised both above and below the estimate, but those within the threshold do not receive any penalty. The most commonly used kernels, for finding the hyperplane, is the Radial Basis Function (RBF) kernel, that we also use for our analysis. One of the main advantages of SVR is that its computational complexity does not depend on the dimensionality of the input space. Moreover, it has excellent generalization capability, and provides high prediction accuracy.

3.4 Estimation framework

Traditionally, before modeling, researchers start by dividing the data into training data and test data. Training data are used to estimate and generate the models' parameters, and the test data are used to calculate the accuracy of the models. Because the test data are not taken into account to fit the model, they should be a reliable indicator of the models' predictive power on new data [97, 98].

Considering that the socio-economic and political situation around the world is not stationary and more recent events are relevant for the prediction, we train our models using the rolling methodology [99], widely used in business and finance [100]. The rolling methodology updates the training set by an add/drop process, while keeping it stable, and retrains the model before each k-months-ahead predictions.

The rolling training's set period of time for all models is half of our data, i.e., 72 months. First, we train the model to predict 6-months-ahead GPI values. After the first training, one month is dropped from the beginning of the training set and another month is added to the end of the training set. Then, we perform the training again to predict the next 6-months-ahead GPI values. We continue this rolling training's first in/first out process for all subsequent months, until we predict the last monthly value. This process ensures that the training set always covers the same amount of time and it is always updated with the most recent information.

In particular, we use the data from March 2008 to February 2014 (72 values) to train the model and predict the GPI values of March 2014 up to August 2014, the data from April 2008 to March 2014 (72 values) to train the model and predict the GPI values of April 2014 up to September 2014, and so on. We repeat this procedure until the last training, which includes data from March 2014 to February 2020 (72 values), to make only 1-month-ahead prediction of the GPI, corresponding to March 2020, the last value of the time series.

At every step, we obtain up to 6-months-ahead predicted GPI values. Specifically, by the end of each rolling training described above, we have k-months-ahead GPI predictions, where $k = 1, 2, \ldots, 6$ months. By the end of all the trainings, we have 72 1-month-ahead GPI predictions[1], 71 2-months-ahead GPI predictions, and so on. We evaluate the accuracy of the predictions for each k-months ahead time horizon with respect to the corresponding test set, that contains the real GPI values.

For each of the models mentioned in Section 3.3, we estimate the best hyperparameters in each training phase through 10-fold cross-validation. **A.3 Hyperparameters** includes all the details for the hyperparameters we tune for each model.

3.5 Model interpretation

Understanding a model's prediction is important for trust, actionability, accountability, debugging, and many other reasons. To understand predictions from tree-based machine learning models, such as Random Forests or XGBoosts, importance values are typically attributed to each variable. Yet traditional variable attribution for trees is inconsistent, meaning it can lower a variable's assigned importance when the true impact of that variable actually increases.

Therefore, for the interpretation of the importance of the model variables and for understanding the drivers of every single GPI estimation we compute the SHAP (SHapley Additive exPlanation) values, proposed by Lundberg et al. [101, 102]. SHAP is based on game theory [103] and local explanations [104], and it offers a means to estimate the contribution of each variable. By focusing specifically on tree-based models, the authors developed an algorithm that computes local explanations based on exact Shapley values in polynomial time. On the one hand, this provides local explanations with theoretical guarantees of local accuracy and consistency. On the other hand, the ability to efficiently compute local explanations using Shapley values over a dataset enables the development of a range of tools to interpret and understand the global behavior of a model. Specifically, by combining many local explanations, a global structure can be represented while retaining local faithfulness [105] to the original model, which generates detailed and accurate representations of model behavior.

4 Results

4.1 Model comparison

The five prediction models, Elastic Net, Decision Tree, Random Forest, XGBoost, and SVR (see Section 3.3), are constructed for every country to produce the GPI

[1] according to the initial test set's length

estimates. In these models, each country's GPI values are the ground-truth data (dependent variable), and the GDELT variables are the exogenous (independent) variables. We consider standard performance indicators to evaluate the performance of the prediction models: the Pearson Correlation coefficient, the Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) [106, 107, 108, 40] (see more details in **A.4 Performance Indicators**).

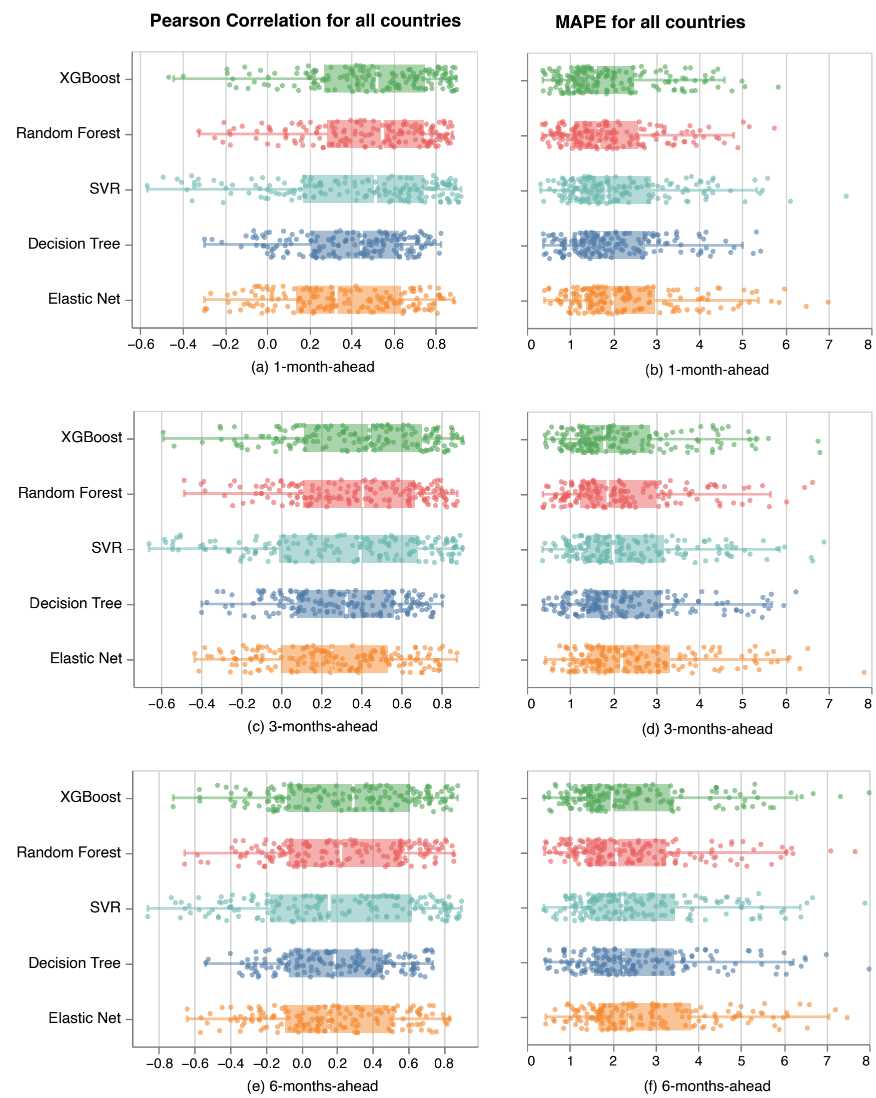

Figure 2 Pearson Correlation and MAPE for all country models. Pearson Correlation and MAPE between the real and the predicted 1-, 3-, and 6-months-ahead GPI values at a country level, for all prediction models. The boxplots represent the distribution of the aforementioned performance indicators for all country models. The plots' data points correspond to each country model. Overall, XGBoost models outperform the rest of the four models.

The analysis is conducted for all 163 countries that have a GPI score. As discussed in Section 3.4, our estimation framework is not limited to the 1-month-ahead predictions, but it generates GPI estimates up to 6-months-ahead. Figure 2 presents Pearson correlation and MAPE performance indicators between the real and the 1-, 3-, and 6-months-ahead predicted GPI values at a country level for all prediction models. The boxplots represent the distribution of the aforementioned performance indicators for all country models. The plots' data points correspond to each country model. Figure 17 in **A.5 RMSE for all country models** presents RMSE performance indicator as well. We observe that XGBoost, Random Forest, and SVR models show similar performance and outperform Decision Tree and Elastic Net models. Overall, XGBoost shows the highest performance. This is more evident for the 6-months-ahead predictions.

For the estimation of the GPI, the models use the historical data of the military, social, and political situation of the corresponding country. For each additional future estimation, we move further away from the last training data, while the country's reality change, and we therefore expect a lower model performance. Indeed, comparing Figures 2 a-b, with Figures 2 c-d, and with Figures 2 e-f, we demonstrate that the performance of the models decreases for every additional month-ahead prediction. For example, we observe a 13,43% increase of the median MAPE for the 3-months-ahead predictions, and a 25.61% increase of the median MAPE for the 6-months-ahead predictions, as compared to the 1-month-ahead predictions.

Concentrating our analysis on the XGBoost boxplot, it is noticeable that some models show high performance and others show low performance. For example, the models for Cameroon, Mali, Turkey, the United Kingdom, and Portugal indicate a very strong correlation, higher than 0.8, and maintain this behavior even for the 6-months-ahead predictions. However, there are models, such as the models for the Central African Republic, Estonia, Moldova, Mongolia, and Romania, that indicate a negative correlation, even for the 1-month-ahead predictions. Notwithstanding that the reasons for the low model's performance are rather complicated, we deduce that GDELT news coverage is not sufficient for some countries.

4.2 Predicting GPI with the XGBoost model

Since our analysis is worldwide, and each country has a different military, socioeconomic, and political history and current situation, it would be interesting to present the performance indicators of various countries' models, so that we cover a variety of country profiles.

We present three of the most powerful countries (United States, United Kingdom, and Saudi Arabia) since they shape global economic patterns and influence decision- and policy-making (see, e.g., [109]). Additionally, we use various sources, such as the official GPI ranking [17], to choose three of the most peaceful countries (Portugal, Iceland, and New Zealand) and three of the most war-torn countries (DR Congo, Libya, and Yemen).

Considering that XGBoost provides the best results on average across all countries, we choose it for the results presentation, as well as for the analysis that follows. Table 2 reports the performance indicators for the XGBoost models for the 1-month-ahead up to 6-months-ahead GPI estimates for nine countries. Overall, 1-month-ahead GPI estimates are significantly more accurate compared to the rest future estimates, especially to the 6-months-ahead time horizon. We observe that there are countries, such as Portugal, for which the model performance remains stable over all 6 months predictions, and countries, such as Yemen, for which the

Table 2 Performance indicators with respect to GPI ground-truth of the prediction models, for nine countries. Overall, 1-month-ahead GPI estimates are significantly more accurate compared to the rest future estimates, especially to the 6-months-ahead time horizon.

Countries	Performance indicators	Prediction framework						Mean
		1-month-ahead	2-months-ahead	3-months-ahead	4-months-ahead	5-months-ahead	6-months-ahead	
United States	Pearson	0.876	0.838	0.813	0.782	0.750	0.710	0.795
	MAPE(%)	1.197	1.367	1.465	1.592	1.700	1.899	1.537
	RMSE	0.037	0.040	0.042	0.045	0.048	0.053	0.044
United Kingdom	Pearson	0.880	0.849	0.848	0.845	0.853	0.850	0.854
	MAPE(%)	0.632	0.742	0.787	0.821	0.826	0.981	0.798
	RMSE	0.015	0.017	0.017	0.018	0.018	0.020	0.017
Saudi Arabia	Pearson	0.864	0.848	0.849	0.814	0.772	0.781	0.822
	MAPE(%)	3.213	3.406	3.733	4.126	4.396	4.590	3.911
	RMSE	0.089	0.094	0.101	0.111	0.119	0.123	0.106
Portugal	Pearson	0.876	0.868	0.868	0.838	0.835	0.820	0.851
	MAPE(%)	3.691	4.241	4.539	5.221	5.067	5.538	4.716
	RMSE	0.057	0.065	0.067	0.077	0.075	0.080	0.070
Iceland	Pearson	0.840	0.833	0.827	0.810	0.770	0.731	0.802
	MAPE(%)	1.867	2.014	2.114	2.256	2.283	2.367	2.150
	RMSE	0.025	0.027	0.028	0.030	0.030	0.031	0.028
New Zealand	Pearson	0.780	0.748	0.725	0.692	0.689	0.650	0.714
	MAPE(%)	1.444	1.538	1.633	1.651	1.741	1.793	1.633
	RMSE	0.023	0.024	0.025	0.026	0.026	0.027	0.025
DR Congo	Pearson	0.820	0.815	0.790	0.762	0.740	0.728	0.776
	MAPE(%)	2.409	2.792	2.856	2.899	2.957	3.120	2.839
	RMSE	0.088	0.099	0.103	0.105	0.107	0.113	0.103
Libya	Pearson	0.854	0.846	0.777	0.835	0.752	0.709	0.796
	MAPE(%)	5.841	6.099	6.765	7.324	7.948	8.603	7.096
	RMSE	0.210	0.225	0.258	0.259	0.289	0.314	0.259
Yemen	Pearson	0.832	0.771	0.746	0.722	0.687	0.662	0.737
	MAPE(%)	5.063	6.033	6.810	7.287	7.801	7.999	6.832
	RMSE	0.207	0.243	0.267	0.283	0.300	0.309	0.268
Yemen *	Pearson	0.953	0.945	0.934	0.922	0.908	0.898	0.892
	MAPE(%)	2.645	2.990	3.440	3.652	3.914	4.171	4.287
	RMSE	0.116	0.129	0.144	0.154	0.166	0.176	0.180

* For the training of this model, the most recent 36 monthly values are used, as compared with the rest of the countries' models that are trained with the most recent 72 monthly values.

model performance falls for each additional in future prediction. An explanation to these different model behaviors could be that, for example for Portugal the military, socio-economic, and political situation remains stable over time, and therefore the most important variables can contribute to a more accurate prediction even further in the future. On the contrary, in war-torn countries, such as Yemen, the country situation changes constantly, and as a consequence, the variables the model uses for the future predictions are not that much relevant anymore. For this reason, for the Yemen model we also conduct a training with the 36 most recent monthly values (Yemen annotated in the Table 2 with an asterisk), as opposed to the 72 values used for the rest of the countries' models. The model's performance improves considerably: the mean Pearson Correlation increases from 0.737 to 0.892, the mean MAPE drops from 6.832 to 4.287, and the mean RMSE decreases from 0.268 to 0.180. However, we do not observe the same improvement in the models' performance when decreasing the training set for the other war-torn countries, such as Libya and DR Congo.

4.3 Case studies

We make a selection of four countries to study their level of peacefulness further. We aim to capture various scenarios on the models' accuracy and the models' explanation of the global and local predictive behavior. In particular, we choose Saudi Arabia and Yemen to understand better and interpret the results and errors of the predictive models based on historical data. Additionally, we choose the United Kingdom and the United States to estimate their future GPI values to gain some initial insights into the country's peace before the official GPI score becomes available.

Saudi Arabia

Based on the World Population Review [109], Saudi Arabia is considered one of the most powerful countries in the world in terms of military alliances, international alliances, political influence, economic influence, and leadership. Consequently, for the current research purposes, it is of great interest.

In particular, we focus on the 6-months-ahead predictions performance for the Saudi Arabia XGBoost model. Figure 3 presents the percentage error of the Saudi Arabia XGBoost model for the 6-months-ahead GPI estimations. We observe that the model performance is high, even for the 6-months-ahead GPI predictions. The percentage error varies, in absolute values, from 4.05% to 11.38%. A positive percentage error demonstrates that the estimated GPI is higher than the real GPI, and therefore the model overshoots. On the contrary, a negative percentage error illustrates that the estimated GPI is lower than the real GPI, and thus the model undershoots. We obtain the largest negative percentage error for the GPI estimation for October 2018.

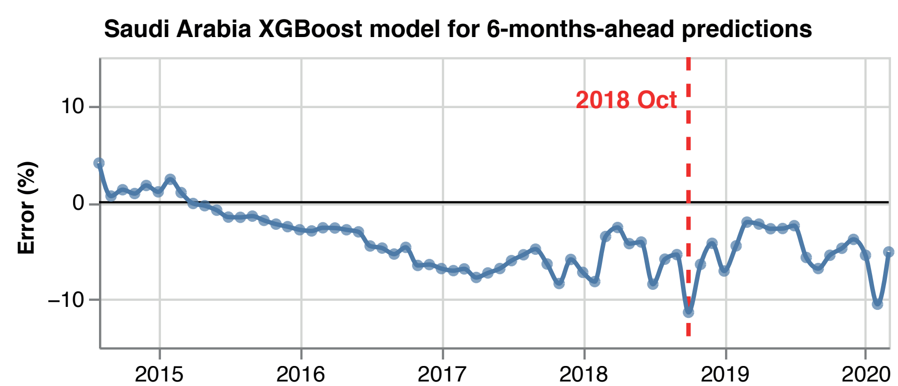

Figure 3 Percentage error for the Saudi Arabia XGBoost model. Percentage error for the Saudi Arabia XGBoost model for the 6-months-ahead GPI estimations (blue curve). The model performance is very high, and the percentage error varies, in absolute values, from 4.05% to 11.38%. We obtain the largest negative percentage error for the GPI estimation for October 2018 (vertical dashed red line).

The analysis of the variables importance through the SHAP methodology reveals the country's profile, but most importantly, it provides us with deeper insights to better understand the larger errors of the model. Figure 4 shows the Global variable importance plot that orders the variables based on their importance in the estimation of the GPI score. Each importance is calculated by combining many local explanations, and the model is trained between May 2012 to April 2018. Overall, we show that the variables demonstrate a profile of a powerful country in military, socio-economic and political terms. This is evident since the variables are related to embargo, boycott, or sanctions, diplomatic relations, mediations, economic co-operations, and appeals for aid, fights with military arms, military engagement, assaults, and endorsements.

To better explain why the model has the worst performance in October 2018, we need to deep into the analysis at a local level. Figure 5 displays the individual SHAP plot for Saudi Arabia's model. The local interpretability can help us explain

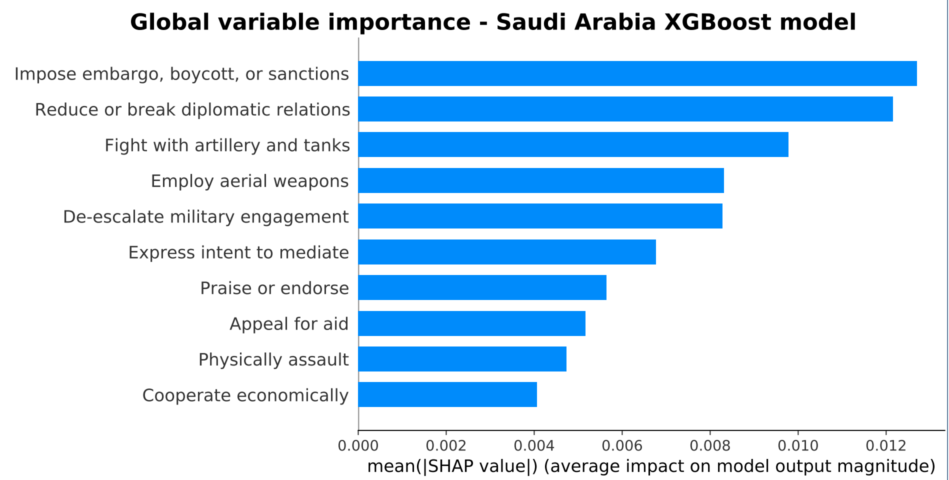

Figure 4 Global variable importance plot for the Saudi Arabia XGBoost model. Global variable importance plot for the Saudi Arabia XGBoost model. The barplot orders the variables based on their importance in the estimation of the GPI score. Overall, we show that the variables demonstrate a profile of a powerful country in military, socio-economic, and political terms.

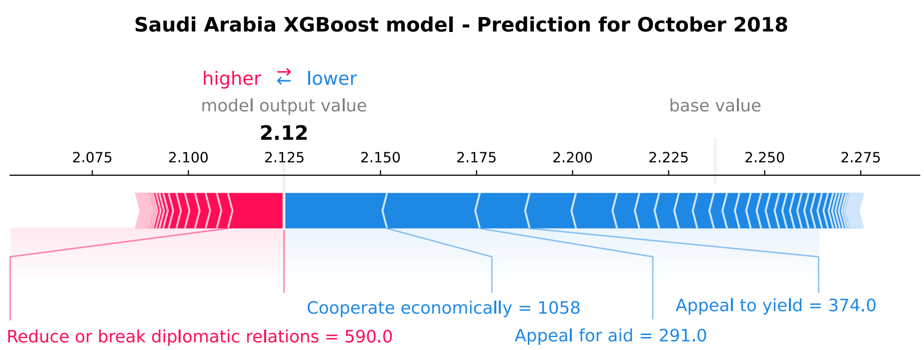

Figure 5 Individual SHAP Value plot for the Saudi Arabia XGBoost model. Individual SHAP Value plot for the Saudi Arabia XGBoost model. It presents the model output value, i.e., the estimation of the GPI for October 2018, and the base value, which is the value that would be predicted if the variables for the current output were unavailable. The plot also displays the most important variables that the model uses for the estimation, such as 'Cooperate economically' and 'Appeal for aid'. The red arrows are the variables that push the GPI estimation higher, and the blue ones push the estimation lower.

the estimation of the GPI for October 2018 through the most important variables that the model uses for this estimation. The model output value is 2.12, and it corresponds to the 6-months-ahead prediction. The base value is smaller than the estimated GPI, and it is the value that would be predicted if the variables for the current output were unavailable. The red arrows are the variables that push the GPI estimation higher (to the right), and those blue push the estimation lower (to the left). Considering that this month the model undershoots (see Figure 3), we focus on the variables that push the GPI estimation lower.

The most important variables to this particular prediction are 'Cooperate economically' and 'Appeal for aid', although they are 10th and 8th respectively in the model's overall ranking of importance (see Figure 4). In October 2018, the journalist Jamal Khashoggi was assassinated at the Saudi consulate in Istanbul, Turkey. This event provoked a series of news on the topics mentioned above in Saudi Arabia. Figure 6 presents Saudi Arabia's model predictions with respect to the real GPI score and the variable 'Cooperate economically'. We notice that this variable shows an abrupt increase this month and pushes GPI prediction lower, showing a more peaceful month. Similarly, Figure 7 shows an abrupt increase of the variable 'Appeal for aid' in October 2018 and drives the prediction lower, showing a more peaceful month. Considering that the assassination of the journalist is a negative event, one would expect a less peaceful month. However, looking at the news, the articles discuss possible spills into oil markets and economic cooperation between Saudi Arabia and other countries, such as the United States, in an attempt to overcome a dispute over Khashoggi. In addition, the news is also concentrated on the investigation of the Khashoggi case, such as Amnesty International asking for a United Nations inquiry. Therefore, considering that the variables 'Cooperate economically', and 'Appeal for aid' have a negative relationship with GPI (see Figure 6, and 7 respectively) the model undershoots. Therefore, we observe that through the eyes of the world news, the presentation of peace is not always at the level we would intuitively expect.

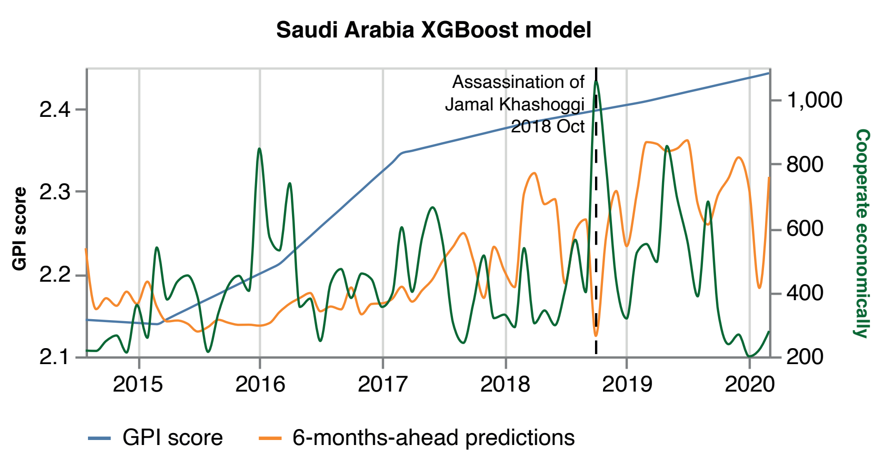

Figure 6 Saudi Arabia XGBoost model predictions, with respect to the real GPI score, and the variable 'Cooperate economically'. Saudi Arabia XGBoost model 6-months-ahead predictions (orange curve), with respect to the real GPI score (blue curve), and the variable 'Cooperate economically' (green curve). This variable pushes the model to undershoot in October 2018 (vertical dashed black line). The reason for this error is the assassination of Jamal Khashoggi in this specific month.

Yemen

In the current world, there is no absence of violent conflict and war. Therefore, peace-builders need to understand the conflict dynamics in war-torn countries to develop entry points for engagement. Based on the official GPI ranking [9], and the World Population Review [110], Yemen is one of the most war-torn countries in the world. Thus it would be interesting to understand in-depth such a country's profile model behavior.

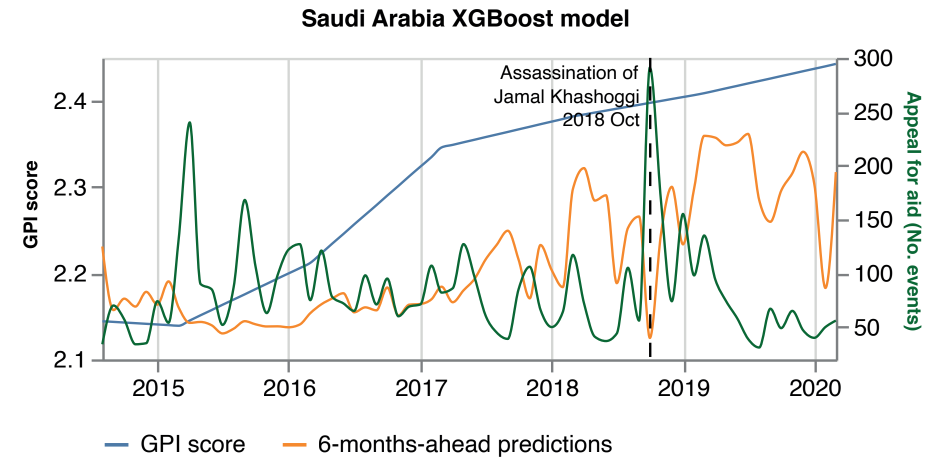

Figure 7 Saudi Arabia XGBoost model predictions, with respect to the real GPI score, and the variable 'Appeal for aid'. Saudi Arabia XGBoost model predictions (orange curve), with respect to the real GPI score (blue curve), and the variable 'Appeal for aid' (green curve). This variable pushes the model to undershoot in October 2018 (vertical dashed black line). The reason for this error is the assassination of Jamal Khashoggi in this specific month.

For all country models, the training dataset has 72 values (six years). However, the situation in Yemen constantly changes due to the Civilian War that broke out in September 2014. The change of peacefulness in the country is depicted in the official GPI value, which abruptly increases in 2015 (see [9]). Therefore, as explained in Section 4.2, it makes sense to shorten the training data from the most recent six years to three years to use more representative data for the prediction. In this case study, we focus our analysis using data from March 2015 to March 2020 to understand the model's behaviour during the Civil War period. Additionally, we study the 1-month-ahead predictions for the Yemen XGBoost model.

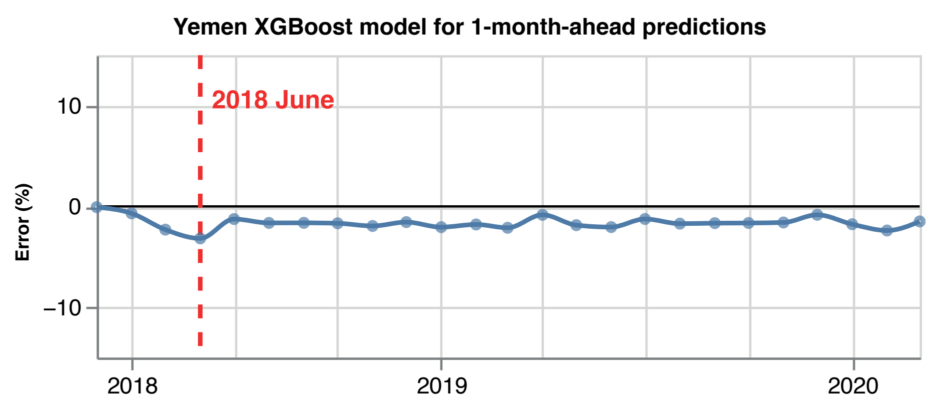

Figure 8 Percentage error for the Yemen XGBoost model. Percentage error for the Yemen XGBoost model for the 1-month-ahead GPI estimations (blue curve). The percentage error varies, in absolute values, from 0.07% to 3.18%. We obtain the largest negative percentage error for the GPI estimation in June 2018 (vertical dashed red line).

Figure 8 presents the percentage error of Yemen's model for 1-month-ahead GPI estimations from March 2018 to March 2020 with a training period of 36 months. The model has a high performance, with a low percentage error that varies from 0.07% to 3.18% with a median value of 1.66%. As discussed before, a positive percentage error illustrates that the model overshoots. On the contrary, a negative percentage error demonstrates that the model undershoots. We obtain the largest negative percentage error for the GPI estimation in June 2018.

Figure 9 presents the global variable importance plot, which orders the variables based on their importance in the estimation of the GPI score. Each variable importance is calculated through the SHAP methodology, with a training period from June 2015 to May 2018. As discussed previously, since each variable importance is calculated with the combination of many local explanations, the plot can give us an overview of the situation in Yemen relevant to the GPI estimation and a general understanding of the model's behavior. Overall, we see that the most important variables reveal a war-torn country profile. Particularly, the variables are related to military aid, territory occupation, bombing, as well as negotiations, discussions, yields, visits, international involvements, and consults.

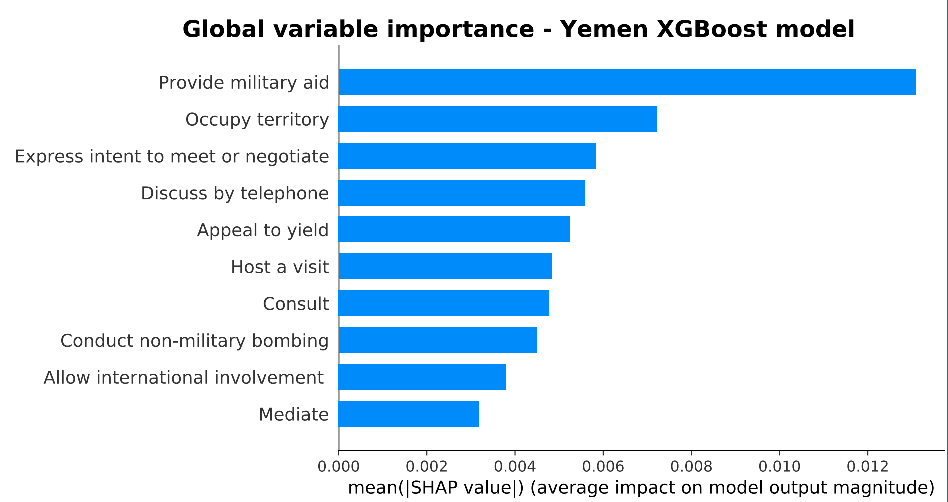

Figure 9 Global variable importance plot for the Yemen XGBoost model. Global variable importance plot for the Yemen XGBoost model. The barplot orders the variables based on their importance in the estimation of the GPI score. Overall, we show that the variables mostly demonstrate a country with a war-torn profile.

Similarly to the Saudi Arabia case study, we need to analyze at a local level to deeply understand why the model produces the highest percentage error for GPI estimation in June 2018. Figure 10 presents the Individual SHAP value plot for Yemen's model, revealing that the GPI prediction and the variables that drive the prediction in June 2018. The model output value is 3.23, and it corresponds to the 1-month-ahead prediction. As explained previously, the base value is the GPI value that would be predicted if the variables for the current output were unavailable. The red arrow represents the variable that pushes the GPI estimation higher, i.e., 'Conduct non-military bombing'. The blue arrow represents the variables that push the GPI estimation lower, i.e., 'Discuss by telephone' and 'Provide military aid'.

Considering that in June 2018, the model undershoots (see Figure 8), we focus our analysis on the latter variables.

This month the number of events on ‘Discuss by telephone’ is 55 and is higher than the median value (14) of the previous three years’ training data. Similarly, the number of events on ‘Provide military aid’ is 121, and it is higher than the median value (72) of the previous three years’ training data. This specific month the United Arab Emirates Armed Forces (UAE) announced a pause to the military operations on 23 June 2018, because of UN-brokered talks. This is depicted in the increase of the news on ‘Discuss by telephone’ topic. In addition, the United States turned down UAE request for aid in the offensive against rebel-held Yemeni port, thanks to the UN efforts. This denial has been discussed a lot on the news, which explains the increase of the news on the ‘Provide military aid’ topic.

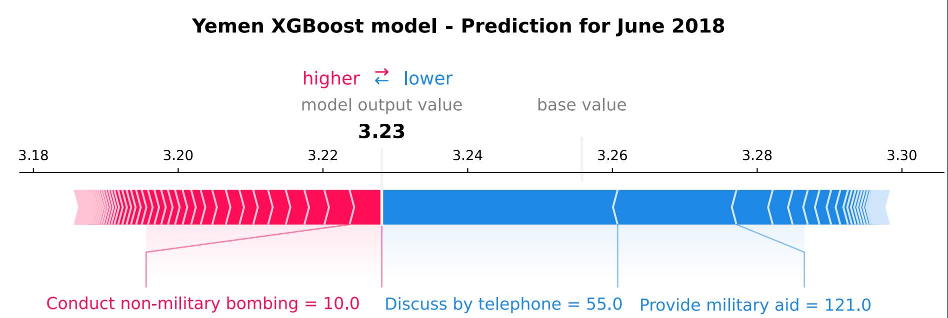

Figure 10 Individual SHAP Value plot for the Yemen XGBoost model. Individual SHAP Value plot for the Yemen XGBoost model. It presents the model output value, i.e., the estimation of the GPI for June 2018, and the base value, which is the value that would be predicted if the variables for the current output were unavailable. The plot also displays the most important variables that the model uses for the estimation, such as ‘Discuss by telephone’ and ‘Provide military aid’. The red arrows are the variables that push the GPI estimation higher, and the blue ones push the estimation lower.

Figure 11 and Figure 12 show that the variables’ higher monthly value and their mostly negative relationship with the GPI drive the model to undershoot in June 2018. In other words, the model’s behavior reveals that this month the GPI value should be lower, and consequently, the month results more peaceful. On the one hand, the model fails to make the correct prediction, since in June 2018, the percentage error is the largest. On the other hand, the model might give an interesting signal; although Yemen is involved in constant conflicts, this month results more peaceful since the UN-brokered ceasefire agreement managed the withdrawal of the warring parties from Al Hudaydah in Yemen. We would like to point out that although we notice additional abrupt increases of the two variables’ values, e.g., in November 2020 (see Figure 11 and Figure 12), the model does not reproduce an abrupt decrease of the GPI. Consequently, the model demonstrates its power to learn from its mistakes.

United States

The United States is considered the most powerful country in the world [109]. On that account, it could be very interesting to study this country beyond our

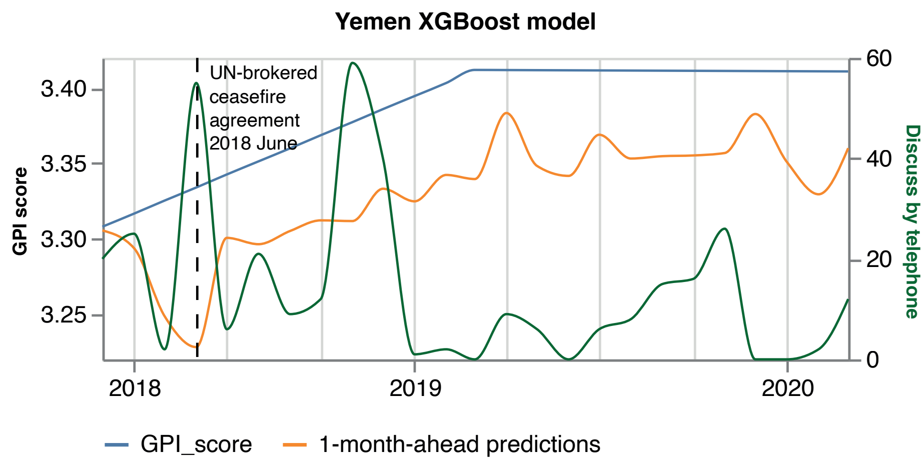

Figure 11 Yemen XGBoost model predictions, with respect to the real GPI score and the variable 'Discuss by telephone'. Yemen XGBoost model 1-month-ahead predictions (orange curve), with respect to the real GPI score (blue curve) and the variable 'Discuss by telephone' (green curve). This variable pushes the model to undershoot in June 2018. The reason for this error is the increase of the news on the topic in this specific month.

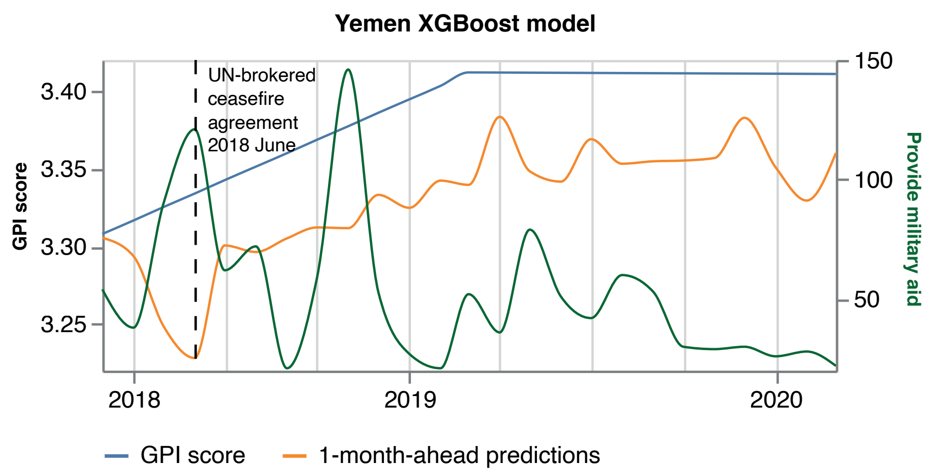

Figure 12 Yemen XGBoost model predictions, with respect to the real GPI score and the variable 'Provide military aid'. Yemen XGBoost 1-month-ahead model predictions (orange curve), with respect to the real GPI score (blue curve) and the variable 'Provide military aid' (green curve). This variable pushes the model to undershoot in June 2018 (vertical dashed black line). The reason for this error is the increase of the news on the topic in this specific month.

ground-truth data, i.e., study its peacefulness after March 2020. The United States's model demonstrates a high performance (see Table 2), and therefore it could provide policy-makers and peace-builders with good initial insights into the country's peacefulness before the official GPI score becomes available. In particular, in this case study, we focus our analysis on the murder of George Floyd, which took place on May 25, 2020, in the United States. Other events followed this extreme event at the end of May and for the whole of June 2020, such as protests, and it provoked an amount of news concentrated on the topic. Thus, it would be interesting to study the United States' level of peacefulness for June and the variables that drive the predicted GPI value.

To start with, Figure 13 displays the global variable importance plot, which presents the most important variables for the training period between April 2014 and March 2020. Overall, we obtain variables that indicate a country profile of a strong player in the military, socio-economic, and political foreground. In particular, we show that the most important variable is related to aerial weapons, and it mainly concerns events that take place overseas. Additionally, the rest of the variables are mostly related to fights with small arms, military de-escalations, embargoes, threats, protests, cooperations, and relations.

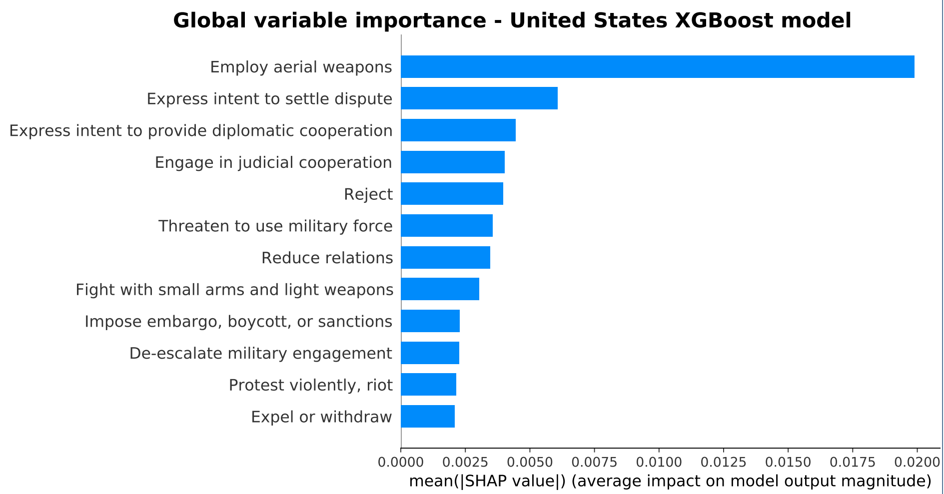

Figure 13 Global variable importance plot for the United States XGBoost model. Global variable importance plot for the United States XGBoost model. The barplot orders the variables based on their importance in the estimation of the GPI score. Overall, we show that the variables indicate a country profile of a strong player in the military, socio-economic, and political foreground.

Since we are interested in predicting the GPI value in June 2020, as well as understanding the most important variables that the model uses for this prediction, we study the local results presented on the individual SHAP plot displayed in Figure 14. The local interpretability depicted in the figure illustrates that the estimated GPI is 2.30 that corresponds to the model output value for 3-months-ahead prediction. This value indicates that the GPI value will remain stably high in June 2020 compared with the last ground-truth value on March 2020 (2.31) and the median GPI value of the previous three years (2.34). The base value is the same as the model output value, and it is the value that would be predicted if the variables for the current output were unavailable. The red arrows are the variables that push the GPI estimation higher (to the right), and those blue push the estimation lower (to the left). Particularly, the variable 'Protest violently, riot' is the variable that pushes the GPI estimation lower. Indeed, in June 2020, the news was concentrated on a series of protests, followed by the murder of George Floyd against police brutality and racism. This variable pushes for a more peaceful month since it has a negative relationship with the GPI. It seems that protesting in the United States contributes

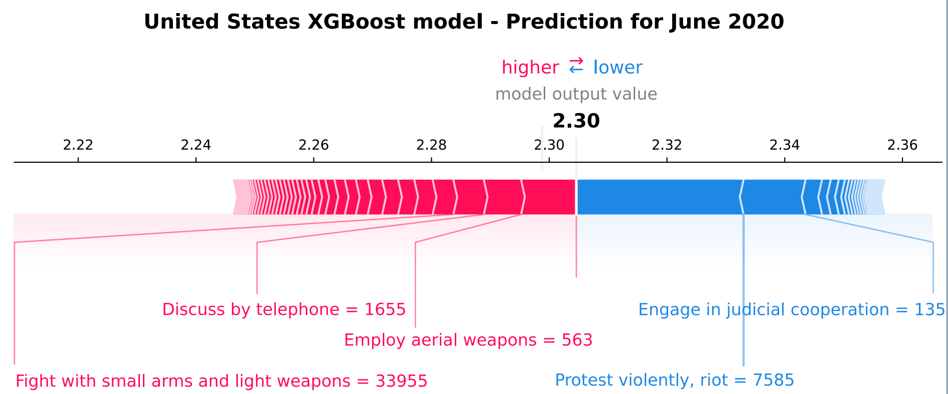

Figure 14 Individual SHAP Value plot for the United States XGBoost model. Individual SHAP Value plot for the United States XGBoost model. It presents the model output value, i.e., the estimation of the GPI for June 2020, and the base value, which is the value that would be predicted if the variables for the current output were unavailable. The plot also displays the most important variables that the model uses for the estimation, such as 'Protest violently, riot'. The red arrows are the variables that push the GPI estimation higher, and the blue ones push the estimation lower.

to the improvement of various socio-political situations, and as a consequence to peace-building.

Furthermore, the rest of the variables displayed in Figure 14 that drive the prediction have lower values than their corresponding median values of the training period, confirming that the news of the month was concentrated on the United States racial unrest and the Black Lives Matter movement. We want to point out that in this particular prediction, the most important variable for the overall training period, i.e., 'Employ aerial weapons' (see Figure 13) has a less important contribution to the model output as compared with the variable 'Protest violently, riot'. This proves the power of SHAP in identifying the role of each variable for every single prediction.

United Kingdom

Similar to the United States and Saudi Arabia, based on the World Population Review [109], the United Kingdom is considered one of the most powerful countries in the world. In addition, the United Kingdom is located in Europe, and it would be interesting for the European social policy-making to anticipate the level of peacefulness after the last ground-truth data, i.e., after March 2020.

In this case study, our objective is to predict the United Kingdom GPI value of July 2020. This month various restrictions related to Covid-19 and the civilians' protection were announced. Figure 15 presents the global variable importance plot that orders the variables based on their importance in the estimation of the GPI score, as estimated through the SHAP methodology with a training period from April 2014 to March 2020. Considering that each importance is calculated by combining many local explanations, the plot can provide us with some general insights on the situation of the United Kingdom in the previous six years and contribute to a better understanding of the model's behavior. Overall, we notice that the variables mostly demonstrate a country where various socio-political events occur. This

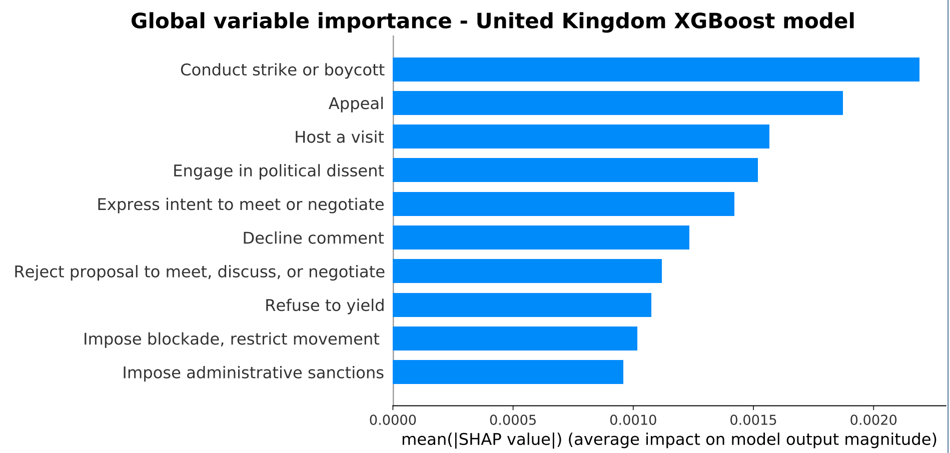

Figure 15 Global variable importance plot for the United Kingdom XGBoost model. Global variable importance plot for the United Kingdom XGBoost model. The barplot orders the variables based on their importance in the estimation of the GPI score. Overall, we show that the variables mostly demonstrate a country where various socio-political events occur.

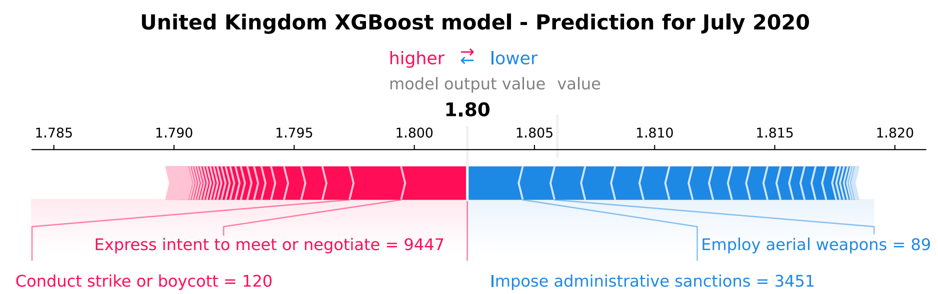

Figure 16 Individual SHAP Value plot for the United Kingdom XGBoost model. Individual SHAP Value plot for the United Kingdom XGBoost model. It presents the model output value, i.e., the estimation of the GPI for July 2020, and the base value, which is the value that would be predicted if the variables for the current output were unavailable. The plot also displays the most important variables that the model uses for the estimation, such as 'Express intent to meet or negotiate' and 'Conduct strike or boycott'. The red arrows are the variables that push the GPI estimation higher, and the blue ones push the estimation lower.

is evident since the variables are mostly related to strikes or boycotts, appeals, negotiations, yields, relationships, and sanctions.

To study the United Kingdom peacefulness in July 2020, we need to deep into the analysis at a local level. The analysis of the variables importance through the SHAP methodology can provide us with an estimation of the GPI value. Additionally, it can offer us deeper insights for understanding which are the variables that drive the prediction for this specific month. For example, Figure 16 presents the Individual SHAP Value plot for the United Kingdom's model. The GPI value is 1.8, and it is the model output value for the 4-months-ahead prediction. As explained before, the base value is the value that would be predicted if the variables for the current output were unavailable. The GPI value in July 2020 is slightly higher than the last ground-truth value (1.77), and it is stable compared to the median GPI value of the previous three years (1.8).

The most important variables that push the GPI value higher are 'Express intent to meet or negotiate' and 'Conduct strike or boycott'. The former variable's value is 9447, which is lower than the median value of the previous six years (12026), and the latter variable's value is 120, which is slightly lower than the median value of the previous six years (126). These results demonstrate that lower values of the aforementioned event categories decrease internal peace in the United Kingdom. The decrease in the events of these categories could be due to COVID-19 restrictions or due to the news concentrated on the COVID-19 pandemic. Besides, the blue arrows represent the variables that push the GPI estimation higher. In particular, 'Impose administrative sanctions' and 'Employ aerial weapons' are the variables that drive the GPI prediction lower. The former's value in July 2020 is 3451, and it is higher than the variable's median value of the previous six years (2590). Looking at the 'Impose administrative sanctions' news articles collected from GDELT, we observe that they are related to discussions on restrictions due to the pandemic, despite the easing of the lockdown. Additionally, many articles discuss the ban to Huawei from the 5G network due to security risks and the ban on junk food advertising and promotion in-store. Consequently, the model has learned that although 'Impose administrative sanctions' events restrict people, the deeper aim of the restrictions is to protect them and promote their well-being. Last, the variable 'Employ aerial weapons' value is much lower than the median value of the previous six years (167) and therefore pushes the GPI value lower. This variable is referred to overseas events that the United Kingdom is involved. The decrease in its value might demonstrate that the news does not discuss it due to previous de-escalations or due to the fact that the news is concentrated on other topics.

5 Conclusion

The analysis of well-being is taking off with the digital era and machine learning revolution. Standard governmental or survey-based measures of well-being are now captured with data science models relying on new digital data streams. Hence, compared to traditional well-being research, data science research allows cost-effective and finer analysis across time.

This study exploits GDELT, a database containing digital news related to socio-political events, to estimate the monthly peacefulness values through GPI. Measuring the GPI score at a monthly level can indicate trends at a much finer scale than it is possible with the yearly official measurements, capturing month-to-month fluctuations and significant events that would be otherwise neglected. Using machine learning, we estimate the GPI values from 1-month-ahead up to 6-months-ahead for 163 countries worldwide, with different socio-economic, political, and military profiles. The application of the SHAP technique allows us to obtain a general explanation of each model's most important GDELT variables, indicating the profile of each country. For example, the most important variables for the Yemen's model are related to military aid, territory occupation, bombing, as well as negotiations, discussions, yields, visits, international involvements, and consults, revealing a war-torn country profile. Additionally, we use SHAP values to provide local explanations of the models' behavior so to understand how the contribution of the most important variables change for each specific prediction. This analysis allows us to explain the errors in the predictions and identify the events that drive the errors.

There are two aspects of our study that we should take into consideration. Firstly, traditional media sometimes misrepresent reality. For example, they give a distorted version of the crimes within a city with a significant bias towards violence [111]. Consequently, the prediction of GPI through the news might be influenced by media biases. Secondly, since the GPI is a yearly index, we upsampled its yearly values linearly to monthly values. The linear upsampling is definitively an assumption since the monthly data generated do not correspond to the real monthly GPI. However, considering that monthly data are not available, linear upsampling is the simplest assumption. Future studies could deepen more the analysis by trying different upsampling methodologies. An alternative solution to this bias could be replacing GPI with a monthly index, which would not require upsampling.

Another line of future research lies in the analysis of the results per country. As discussed in Section 4, for certain countries, the models show low performance in predicting the GPI value. One approach to improve the model's performance is to change the training data length based on the history of the country, usually depicted on the GPI. For example, as shown in Section 4.2 for Yemen, the model's performance improves by changing the training data from the most recent 72 months to the most recent 36 months. In addition, studying in depth the representativeness of GDELT news, as not all countries are equally covered, could explain why some countries' model fails in any case.

The analysis of our results shows great promise for the estimation of GPI through GDELT, yet an unexplored data source. We believe that this study is valuable to policy-makers and the scientific community, especially to researchers interested in "Data Science for Social Good". In other words, GDELT could be used not only for peacefulness but for any other well-being dimension and socio-economic index related to societal progress.

Appendix

A.1 Indicators of GPI

The GPI is a composite index of these 23 indicators weighted and combined into one overall score. The GPI comprises 23 indicators of the absence of violence or fear of violence aggregated into three major categories: Ongoing Domestic & International Conflict, Societal Safety & Security, and Militarization:

- Ongoing Domestic & International Conflict includes: "Number and duration of internal conflicts", "Number of deaths from external organised conflict", "Number of deaths from internal organised conflict", "Number, duration and role in external conflicts", "Intensity of organised internal conflict", and "Relations with neighbouring countries".
- Societal Safety & Security encompasses: "Level of perceived criminality in society", "Number of refugees and internally displaced people as a percentage of the population" , "Political instability", "Political Terror Scale", "Impact of terrorism", "Number of homicides per 100,000 people", "Level of violent crime", "Likelihood of violent demonstrations", "Number of jailed population per 100,000 people", "Number of internal security officers, and police per 100,000 people".

- Militarization contains: “Military expenditure as a percentage of GDP”, “Number of armed services personnel per 100,000 people”, “Volume of transfers of major conventional weapons as recipient (imports) per 100,000 people”, “Volume of transfers of major conventional weapons as supplier (exports) per 100,000 people”, “Financial contribution to UN peacekeeping missions”, “Nuclear and heavy weapons capabilities”, and “Ease of access to small arms and light weapons”.

A.2 Topics of GDELT

The GDELT event categories we use are related to 20 topics, as described below. For each topic, we provide a short description and a few examples of event categories:

Make Public Statement refers to public statements expressed verbally or in action, such as “Make statement”, “Make pessimistic comment”, and “Decline comment”. Appeal refers to requests, proposals, suggestions and appeals, such as “Appeal for material cooperation”, “Appeal for economic cooperation”, and “Appeal to others to settle dispute”. Express Intent To Cooperate refers to offer, promise, agree to, or otherwise indicate willingness or commitment to cooperate, such as “Express intent to engage in material cooperation” and “Express intent to provide material aid”. Consult refers to consultations and meetings, such as “Discuss by telephone” and “Host a visit”. Engage In Diplomatic Cooperation refers to initiate, resume, improve, or expand diplomatic, non-material cooperation or exchange, such as “Sign formal agreement” and “Praise or endorse”. Engage In Material Cooperation refers to initiate, resume, improve, or expand material cooperation or exchange, such as “Cooperate economically” and “Share intelligence or information”. Provide Aid refers to provisions and extension of material aid, such as “Provide economic aid” and “Provide humanitarian aid”. Yield refers to yieldings and concessions, such as “Accede to requests or demands for political reform”, “De-escalate military engagement”, and “Return, release”. Investigate refers to non-covert investigations, such as “Investigate crime, corruption” and “Investigate human rights abuses”. Demand refers to demands and orders, such as “Demand political reform” and “Demand settling of dispute”. Disapprove refers to the expression of disapprovals, objections, and complaints, such as “Criticize or denounce” and “Complain officially”. Reject refers to rejections and refusals, such as “Reject request or demand for material aid” and “Reject mediation”. Threaten refers to threats, coercive or forceful warnings with serious potential repercussions, such as “Threaten with military force” and “Threaten with administrative sanctions”. Protest refers to civilian demonstrations and other collective actions carried out as protests such as “Demonstrate or rally” and “Conduct strike or boycott”. Exhibit Force Posture refers to military or police moves that fall short of the actual use of force, such as “Exhibit military or police power” and “Increase military alert status”. Reduce Relations refers to reductions in normal, routine, or cooperative relations, such as “Reduce or break diplomatic relations” and “Halt negotiations”. Coerce refers to repression, violence against civilians, or their rights or properties, such as “Arrest, detain” and “Seize or damage property”. Assault refers to the use of different forms of violence, such as “Conduct suicide, car, or other non-military bombing” and “Abduct, hijack, take hostage”. Fight refers to uses of conventional force and acts of war, such as "Use conventional military force" and "Fight with small arms and light weapons". ENGAGE IN UNCONVENTIONAL MASS VIOLENCE refers to uses of unconventional force that are meant to cause mass destruction, casualties, and suffering, such as "Engage in ethnic cleansing" and "Detonate nuclear weapons".

A.3 Hyperparameters

The hyperparameters we tune for Elastic Net are α, which is the relative importance of the L1 (LASSO) and L2 (Ridge) penalties, and λ, which is the amount of regularization used in the model. For Decision Tree, we tune the complexity parameters $maxdepth$, which is the maximum depth of the tree), $minsamplessplit$, which is the minimum number of samples required to split an internal node, and $minsamplesleaf$, which is the minimum number of samples required to be at a leaf node. For Random Forest, similarly to Decision Tree, we tune the $maxdepth$, the $minsamplessplit$, and the $minsamplesleaf$. We also tune the $nestimators$, which accounts for the number of number of trees in the model, and the $maxfeatures$, which corresponds to the number of variables to consider when looking for the best split. For XGBoost, we tune the $nestimators$, similarly to Random Forest, and the $maxdepth$, similarly to Decision Tree. We also tune the $learningrate$, a value that in each boosting step, shrinks the weight of new variables, preventing overfitting or a local minimum, and $colsample_bytree$, which represents the fraction of columns to be subsampled, it is related to the speed of the algorithm and it prevents overfitting. Last, for SVR RBF model we tune the regularization parameter C, which imposes a penalty to the model for making an error, and $gamma$ parameter, which defines how far the influence of a single training example reaches.

A.4 Performance Indicators

We consider the following indicators to assess the performance of the prediction models with respect to the ground-truth GPI values. Our notation is as follows: y_t denotes the observed value of the GPI at time t, x_t denotes the predicted value by the model at time t, $\bar{y}$ denotes the mean or average of the values y_t and similarly $\bar{x}$ denotes the mean or average of the values x_t.

Pearson Correlation, a measure of the linear dependence between two variables during a time period $[t_1, t_n]$, is defined as:

$$r = \frac{\sum_{t=1}^{n}(y_t - \bar{y})(x_t - \bar{x})}{\sqrt{\sum_{t=1}^{n}(y_t - \bar{y})^2}\sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2}} \ . \tag{1}$$

Root Mean Square Error (RMSE), a measure of prediction accuracy that represents the square root of the second sample moment of the differences between predicted values and actual values, is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - y_t)^2} \ . \tag{2}$$

Mean Absolute Percentage Error (MAPE), a measure of prediction accuracy between predicted and true values, is defined as:

$$MAPE = \left(\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - x_t}{y_t}\right|\right) \times 100 \ . \tag{3}$$

A.5 RMSE for all country models

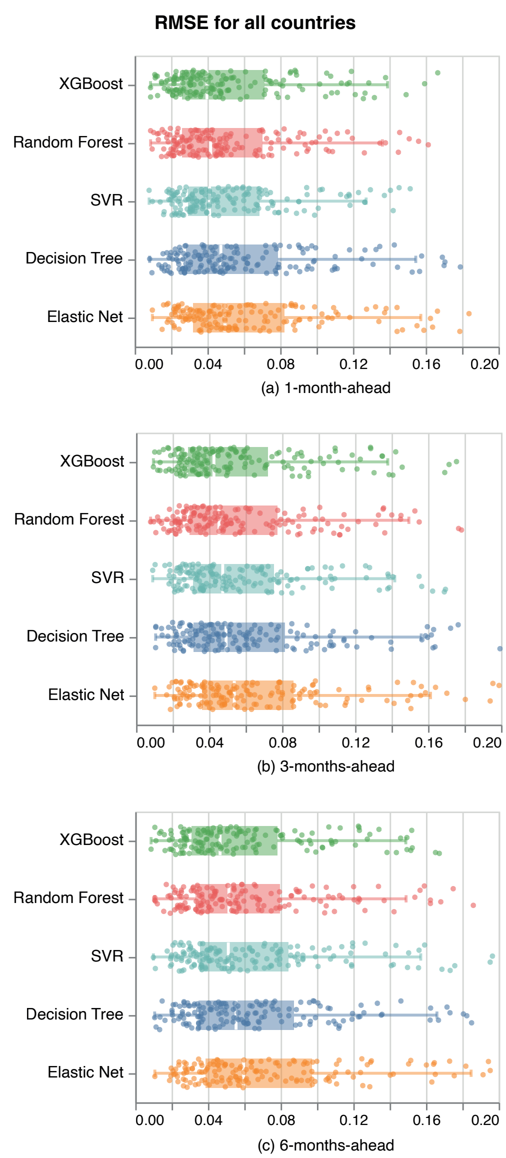

Figure 17 RMSE for all country models. RMSE between the real and the predicted 1-, 3-, and 6-months-ahead GPI values at a country level, for all prediction models. The boxplots represent the distribution of the aforementioned performance indicators for all country models. The plots' data points correspond to each country model. Overall, XGBoost models outperform the rest of the four models.

Availability of data and materials
The code to reproduce the study is available at https://github.com/VickyVouk/GDELT_GPI_SHAP_project.

Competing interests
The authors declare that they have no competing interests.

Funding
This work has been partially funded by EU project H2020 SoBigData++ #871042.

Author's contributions
VV : study conceptualization, data preprocessing and analysis, experiment running, code implementation, interpretation of results, writing, plots, IM: study conceptualization, data preprocessing and analysis, experiment running, code implementation, interpretation of results, writing, FG: interpretation of results and study direction, LP: study conceptualization, experiment design, interpretation of results, writing, study direction.

Acknowledgements
This work is supported by H2020 SoBigData++ #871042. We thank Stefano-Maria Iacus, Stan Matwin, Francesca Chiaromonte, and Donato Farina for their valuable feedback and inspiration. We also thank Daniele Fadda for support on data visualization.

Author details
[1]Scuola Normale Superiore, Pisa, Italy. [2]Department of Computer & Systems Sciences, Stockholm University, Stockholm, Sweden. [3]Institute of Information Science and Technologies (ISTI), National Research Council of Italy (CNR), Pisa, Italy.

References
1. Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., Belgrave, D.C.M., Ezer, D., van der Haert, F.C., Mugisha, F., Abila, G., Arai, H., Almiraat, H., Proskurnia, J., Snyder, K., Otake-Matsuura, M., Othman, M., Glasmachers, T., de Wever, W., Teh, Y.W., Khan, M.E., Winne, R.D., Schaul, T., Clopath, C.: AI for social good: unlocking the opportunity for positive impact. Nature Communications **11**(1), 2468 (2020). doi:10.1038/s41467-020-15871-z
2. Fleurbaey, M.: Beyond GDP: The quest for a measure of social welfare. Journal of Economic literature **47**(4), 1029–75 (2009)
3. Organisation for Economic Co-operation and Development: How's Life?: Measuring Well-being. OECD, Paris (2011)
4. Voukelatou, V., Gabrielli, L., Miliou, I., Cresci, S., Sharma, R., Tesconi, M., Pappalardo, L.: Measuring objective and subjective well-being: dimensions and data sources. International Journal of Data Science and Analytics, 1–31 (2020)
5. UNDP: Sustainable Development Goals. https://sustainabledevelopment.un.org/sdgs (2015)
6. Kroll, C., Warchold, A., Pradhan, P.: Sustainable development goals (sdgs): Are we successful in turning trade-offs into synergies? Palgrave Communications **5**(1), 1–11 (2019)
7. Le Blanc, D.: Towards integration at last? the sustainable development goals as a network of targets. Sustainable Development **23**(3), 176–187 (2015)
8. Anand, S., Sen, A.: Human Development Index: Methodology and Measurement. (1994)
9. The Institute for Economics and Peace: VISION of HUMANITY. http://visionofhumanity.org// (2017)
10. Melamed, C., Morales, L., Hsu, Y., Poole, J., Rae, B., Rutherford, I., Jahic, A.: A World That Counts-Mobilising the Data Revolution for Sustainable Development. United Nations. Accessed March **12**, 2018 (2014)
11. Leetaru, K.: The GDELT Project. https://www.gdeltproject.org/ (2013)
12. Voukelatou, V., Pappalardo, L., Miliou, I., Gabrielli, L., Giannotti, F.: Estimating countries' peace index through the lens of the world news as monitored by gdelt. In: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), pp. 216–225 (2020). IEEE
13. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D.: A survey of methods for explaining black box models. ACM Comput. Surv. **51**(5) (2018). doi:10.1145/3236009
14. Brückner, M., Ciccone, A.: International commodity prices, growth and the outbreak of civil war in Sub-Saharan Africa. The Economic Journal **120**(544), 519–534 (2010)
15. Gries, P., Fox, A., Jing, Y., Mader, M., Scotto, T.J., Reifler, J.: A new measure of the 'democratic peace': what country feeling thermometer data can teach us about the drivers of American and Western European foreign policy. Political Research Exchange **2**(1), 1716630 (2020)
16. The Institute for Economics & Peace: Structures of peace: identifying what leads to peaceful societies (2011)
17. The Institute for Economics and Peace: Global Peace Index 2020 (2020)
18. The Institute for Economics and Peace: Global Peace Index 2015 (2015)
19. Chen, X., Cho, Y., Jang, S.Y.: Crime prediction using Twitter sentiment and weather. In: 2015 Systems and Information Engineering Design Symposium, pp. 63–68 (2015). IEEE
20. Al Boni, M., Gerber, M.S.: Predicting crime with routine activity patterns inferred from social media. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 001233–001238 (2016). IEEE
21. Kadar, C., Brüngger, R.R., Pletikosa, I.: Measuring ambient population from location-based social networks to describe urban crime. In: International Conference on Social Informatics, pp. 521–535 (2017). Springer
22. Curiel, R.P., Cresci, S., Muntean, C.I., Bishop, S.R.: Crime and its fear in social media. Palgrave Communications **6**(1), 1–12 (2020)
23. Chen, F., Neill, D.B.: Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs. In: Proc. of the 20th ACM SIGKDD Intenational Conference on Knowledge Discovery and Data Mining, pp. 1166–1175 (2014). ACM

24. Nobles, M., Neill, D.B., Flaxman, S.: Predicting and Preventing Emerging Outbreaks of Crime (2014)
25. Neill, D.B., Gorr, W.L.: Detecting and preventing emerging epidemics of crime. Advances in Disease Surveillance **4**(13) (2007)
26. Tucker, R., O'Brien, D.T., Ciomek, A., Castro, E., Wang, Q., Phillips, N.E.: Who 'tweets' where and when, and how does it help understand crime rates at places? measuring the presence of tourists and commuters in ambient populations. Journal of Quantitative Criminology, 1–27 (2021)
27. Spangler, E., Smith, B.: Let them tweet cake: Estimating public dissent using twitter. Defence and Peace Economics, 1–20 (2021)
28. Najjar, E., Al-augby, S.: Sentiment analysis combination in terrorist detection on twitter: A brief survey of approaches and techniques. Research in Intelligent and Computing in Engineering, 231–240 (2021)
29. Zeitzoff, T.: Using social media to measure conflict dynamics: An application to the 2008-2009 Gaza conflict. Journal of Conflict Resolution **55**(6), 938–969 (2011)
30. Siapera, E., Hunt, G., Lynn, T.: # GazaUnderAttack: Twitter, Palestine and diffused war. Information, Communication & Society **18**(11), 1297–1319 (2015)
31. Zeitzoff, T., Kelly, J., Lotan, G.: Using social media to measure foreign policy dynamics: An empirical analysis of the Iranian-Israeli confrontation (2012-13). Journal of Peace Research **52**(3), 368–383 (2015)
32. Zagheni, E., Garimella, V.R.K., Weber, I., State, B.: Inferring international and internal migration patterns from twitter data. In: Proc. of the 23rd International Conference on World Wide Web, pp. 439–444 (2014)
33. Zagheni, E., Weber, I., Gummadi, K.: Leveraging facebook's advertising platform to monitor stocks of migrants. Population and Development Review, 721–734 (2017)
34. Alexander, M., Polimis, K., Zagheni, E.: Combining social media and survey data to nowcast migrant stocks in the united states. Population Research and Policy Review, 1–28 (2020)
35. Mazoyer, B., Cagé, J., Hervé, N., Hudelot, C.: A french corpus for event detection on twitter. In: Proceedings of The 12th Language Resources and Evaluation Conference, pp. 6220–6227 (2020)
36. Toch, E., Lerner, B., Ben-Zion, E., Ben-Gal, I.: Analyzing large-scale human mobility data: a survey of machine learning methods and applications. Knowledge and Information Systems **58**(3), 501–523 (2019)
37. Pappalardo, L., Simini, F., Barlacchi, G., Pellungrini, R.: Scikit-mobility: a Python library for the analysis, generation and risk assessment of mobility data (2021). 1907.07062
38. Blondel, V.D., Decuyper, A., Krings, G.: A survey of results on mobile phone datasets analysis. EPJ Data Science **4**(1), 10 (2015). doi:10.1140/epjds/s13688-015-0046-0
39. Andrienko, G., Andrienko, N., Boldrini, C., Caldarelli, G., Cintia, P., Cresci, S., Facchini, A., Giannotti, F., Gionis, A., Guidotti, R., *et al.*: (so) big data and the transformation of the city. International Journal of Data Science and Analytics **11**(4), 311–340 (2021)
40. Luca, M., Barlacchi, G., Lepri, B., Pappalardo, L.: A Survey on Deep Learning for Human Mobility (2021). 2012.02825
41. Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., Pentland, A.: Once upon a crime: towards crime prediction from demographics and mobile data. In: Proc. of the 16th International Conference on Multimodal Interaction, pp. 427–434 (2014). ACM
42. Ariel, B., Partridge, H.: Predictable policing: Measuring the crime control benefits of hotspots policing at bus stops. Journal of Quantitative Criminology **33**(4), 809–833 (2017)
43. Ferrara, E., De Meo, P., Catanese, S., Fiumara, G.: Detecting criminal organizations in mobile phone networks. Expert Systems with Applications **41**(13), 5733–5750 (2014)
44. Robinson, A.I., Carnes, F., Oreskovic, N.M.: Spatial analysis of crime incidence and adolescent physical activity. Preventive medicine **85**, 74–77 (2016)
45. Wu, J., Frias-Martinez, E., Frias-Martinez, V.: Addressing under-reporting to enhance fairness and accuracy in mobility-based crime prediction. In: Proceedings of the 28th International Conference on Advances in Geographic Information Systems, pp. 325–336 (2020)
46. De Nadai, M., Xu, Y., Letouzé, E., González, M.C., Lepri, B.: Socio-economic, built environment, and mobility conditions associated with crime: a study of multiple cities. Scientific reports **10**(1), 1–12 (2020)
47. Pappalardo, L., Vanhoof, M., Gabrielli, L., Smoreda, Z., Pedreschi, D., Giannotti, F.: An analytical framework to nowcast well-being using mobile phone data. International Journal of Data Science and Analytics **2**(1), 75–92 (2016)
48. Pappalardo, L., Pedreschi, D., Smoreda, Z., Giannotti, F.: Using big data to study the link between human mobility and socio-economic development. In: 2015 IEEE International Conference on Big Data (Big Data), pp. 871–878 (2015). IEEE
49. Stefano, M., Caterina, G., Monica, P., Nicola, S., Fosca, G., Dino, P., Salvatore, R., Luca, P., Lorenzo, G.: Small Area Model-Based Estimators Using Big Data Sources. Journal of Official Statistics **31**(2), 263–281 (2015). doi:10.1515/jos-2015-0017
50. Eagle, N., Macy, M., Claxton, R.: Network diversity and economic development. Science **328**(5981), 1029–1031 (2010)
51. Chi, G., Lin, F., Chi, G., Blumenstock, J.: A general approach to detecting migration events in digital trace data. PloS one **15**(10), 0239408 (2020)
52. Sîrbu, A., Andrienko, G., Andrienko, N., Boldrini, C., Conti, M., Giannotti, F., Guidotti, R., Bertoli, S., Kim, J., Muntean, C.I., et al.: Human migration: the big data perspective. International Journal of Data Science and Analytics, 1–20 (2020)
53. Pappalardo, L., Ferres, L., Sacasa, M., Cattuto, C., Bravo, L.: Evaluation of home detection algorithms on mobile phone data using individual-level ground truth. EPJ Data Science **10**(1), 29 (2021). doi:10.1140/epjds/s13688-021-00284-9
54. Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., Blondel, V.D., Tatem, A.J.: Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences **111**(45), 15888–15893 (2014). doi:10.1073/pnas.1408439111
55. DaViera, A.L., Roy, A.L., Uriostegui, M., Fiesta, D.: Safe spaces embedded in dangerous contexts: How

chicago youth navigate daily life and demonstrate resilience in high-crime neighborhoods. American journal of community psychology **66**(1-2), 65–80 (2020)
56. Qi, H., Manrique, P., Johnson, D., Restrepo, E., Johnson, N.F.: Open source data reveals connection between online and on-street protest activity. EPJ data science **5**(1), 18 (2016)
57. Qi, H., Manrique, P., Johnson, D., Restrepo, E., Johnson, N.F.: Association between volume and momentum of online searches and real-world collective unrest. Results in physics **6**, 414–419 (2016)
58. Muchow, A.N., Amuedo-Dorantes, C.: Immigration enforcement awareness and community engagement with police: Evidence from domestic violence calls in los angeles. Journal of urban economics **117**, 103253 (2020)
59. Lea, S.G., D'Silva, E., Asok, A.: Women's strategies addressing sexual harassment and assault on public buses: an analysis of crowdsourced data. Crime prevention and community safety **19**(3-4), 227–239 (2017)
60. Ozkan, T., Worrall, J.L., Zettler, H.: Validating media-driven and crowdsourced police shooting data: a research note. Journal of crime and justice **41**(3), 334–345 (2018)
61. Palakodety, S., KhudaBukhsh, A.R., Carbonell, J.G.: Hope Speech Detection: A Computational Analysis of the Voice of Peace (2019). 1909.12940
62. Rumi, S.K., Shao, W., Salim, F.D.: Realtime predictive patrolling and routing with mobility and emergency calls data. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 14, pp. 964–968 (2020)
63. Solymosi, R., Buil-Gil, D., Vozmediano, L., Guedes, I.S.: Towards a place-based measure of fear of crime: A systematic review of app-based and crowdsourcing approaches. Environment and Behavior, 0013916520947114 (2020)
64. Quinn, J.A., Nyhan, M.M., Navarro, C., Coluccia, D., Bromley, L., Luengo-Oroz, M.: Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **376**(2128), 20170363 (2018)
65. Witmer, F.D.: Remote sensing of violent conflict: eyes from above. International Journal of Remote Sensing **36**(9), 2326–2352 (2015)
66. Marx, A., Loboda, T.: Landsat-based early warning system to detect the destruction of villages in darfur, sudan. Remote sensing of environment **136**, 126–134 (2013)
67. Li, X., Li, D.: Can night-time light images play a role in evaluating the syrian crisis? International Journal of Remote Sensing **35**(18), 6648–6661 (2014)
68. Clionadh, R., Linke, A., Hegre, H., Karlsen, J.: Introducing ACLED-Armed Conflict Location and Event Data. Journal of Peace Research **47**(5), 651–660 (2010)
69. Brauer, J., Anderton, C.H.: Conflict and Peace Economics: Retrospective and Prospective Reflections on Concepts, Theories, and Data. Defence and Peace Economics, 1–23 (2020)
70. Firchow, P., Ginty, R.M.: Measuring peace: Comparability, commensurability, and complementarity using bottom-up indicators. International Studies Review **19**(1), 6–27 (2017)
71. Ide, T.: Covid-19 and armed conflict. World development **140**, 105355 (2021)
72. Inform, A.: Arabia Inform. `http://arabiainform.com/` (1998)
73. Hossain, K.T., Gao, S., Kennedy, B., Galstyan, A., Natarajan, P.: Forecasting violent events in the middle east and north africa using the hidden markov model and regularized autoregressive models. The Journal of Defense Modeling and Simulation **17**(3), 269–283 (2020)
74. Wu, C., Gerber, M.S.: Forecasting civil unrest using social media and protest participation theory. IEEE Transactions on Computational Social Systems **5**(1), 82–94 (2017)
75. Qiao, F., Li, P., Zhang, X., Ding, Z., Cheng, J., Wang, H.: Predicting social unrest events with hidden Markov models using GDELT. Discrete Dynamics in Nature and Society **2017** (2017)
76. Galla, D., Burke, J.: Predicting social unrest using GDELT. In: International Conference on Machine Learning and Data Mining in Pattern Recognition, pp. 103–116 (2018). Springer
77. Alsaqabi, A., Aldhubayi, F., Albahli, S.: Using Machine Learning for Prediction of Factors Affecting Crimes in Saudi Arabia. In: Proc. of the 2019 International Conference on Big Data Engineering, pp. 57–62 (2019)
78. Joshi, D., Basnet, S., Arunachalam, H., Soh, L.-K., Samal, A., Ratcliff, S., Werum, R.: SURGE: Social Unrest Reconnaissance GazEteer. In: Proc. of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 1–4 (2017)
79. Fengcai, Q., Jinsheng, D., Li, W.: An online framework for temporal social unrest event prediction using news stream. In: 2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), pp. 176–182 (2020). IEEE
80. Qiao, F., Li, P., Deng, J., Ding, Z., Wang, H.: Graph-based method for detecting occupy protest events using gdelt dataset. In: 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, pp. 164–168 (2015). IEEE
81. González, M., Alférez, G.H.: Application of data science to discover violence-related issues in iraq. arXiv preprint arXiv:2006.07980 (2020)
82. Keneshloo, Y., Cadena, J., Korkmaz, G., Ramakrishnan, N.: Detecting and forecasting domestic political crises: A graph-based approach. In: Proc. of the 2014 ACM Conference on Web Science, pp. 192–196 (2014)
83. Keertipati, S., Savarimuthu, B.T.R., Purvis, M., Purvis, M.: Multi-level analysis of peace and conflict data in GDELT. In: Proc. of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis, pp. 33–40 (2014)
84. Yonamine, J.E.: Predicting future levels of violence in afghanistan districts using gdelt. Unpublished manuscript (2013)
85. Yuan, L., Song, C., Cheng, C., Shen, S., Chen, X., Wang, Y.: The cooperative and conflictual interactions between the united states, russia, and china: A quantitative analysis of event data. Journal of Geographical Sciences **30**(10), 1702–1720 (2020)
86. Ahmed, M.N., Barlacchi, G., Braghin, S., Calabrese, F., Ferretti, M., Lonij, V., Nair, R., Novack, R., Paraszczak, J., Toor, A.S.: A Multi-Scale Approach to Data-Driven Mass Migration Analysis. In: SoGood@

ECML-PKDD (2016)
87. Beine, M., Bertinelli, L., Cömertpay, R., Litina, A., Maystadt, J.-F., Zou, B.: Refugee Mobility: Evidence from Phone Data in Turkey. In: Guide to Mobile Data Analytics in Refugee Scenarios, pp. 433–449. Springer, (2019)
88. Bertoli, S., Cintia, P., Giannotti, F., Madinier, E., Ozden, C., Packard, M., Pedreschi, D., Rapoport, H., Sîrbu, A., Speciale, B.: Integration of Syrian refugees: insights from D4R, media events and housing market data. In: Guide to Mobile Data Analytics in Refugee Scenarios, pp. 179–199. Springer, (2019)
89. Carammia, M., Iacus, S.M., Wilkin, T.: Forecasting asylum applications in the european union with machine learning and data at scale. arXiv preprint arXiv:2011.04348 (2020)
90. Best, R.H., Carpino, C., Crescenzi, M.J.: An analysis of the TABARI coding system. Conflict Management and Peace Science **30**(4), 335–348 (2013)
91. Schrodt, P.A.: Cameo: Conflict and mediation event observations event and actor codebook. Pennsylvania State University (2012)
92. Leetaru, K., Schrodt, P.A.: Gdelt: Global data on events, location, and tone, 1979-2012. In: ISA Annual Convention, vol. 2, pp. 1–49 (2013). Citeseer
93. Fernandes, S., Bernardino, J.: What is bigquery? In: Proceedings of the 19th International Database Engineering & Applications Symposium. IDEAS '15, pp. 202–203. Association for Computing Machinery, New York, NY, USA (2015). doi:10.1145/2790755.2790797. https://doi.org/10.1145/2790755.2790797
94. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, (2009)
95. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)
96. Awad, M., Khanna, R.: Support vector regression. In: Efficient Learning Machines, pp. 67–80. Springer, (2015)
97. Shanmugam, R.: Introduction to time series and forecasting. Taylor & Francis Group (1997)
98. Hyndman, R.J., Koehler, A.B.: Another look at measures of forecast accuracy. International journal of forecasting **22**(4), 679–688 (2006)
99. Hyndman, R.J., Athanasopoulos, G.: Forecasting: Principles and Practice. OTexts, (2018)
100. Zeller, T.L., Metzger, L.M., *et al.*: Good bye traditional budgeting, hello rolling forecast: has the time come? American Journal of Business Education (AJBE) **6**(3), 299–310 (2013)
101. Lundberg, S.M., Erion, G.G., Lee, S.-I.: Consistent individualized feature attribution for tree ensembles. arXiv preprint arXiv:1802.03888 (2018)
102. Lundberg, S., Lee, S.-I.: A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874 (2017)
103. Štrumbelj, E., Kononenko, I.: Explaining prediction models and individual predictions with feature contributions. Knowledge and information systems **41**(3), 647–665 (2014)
104. Ribeiro, M.T., Singh, S., Guestrin, C.: " why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144 (2016)
105. Ribeiro, M.T., Singh, S., Guestrin, C.: Anchors: High-precision model-agnostic explanations. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
106. James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning vol. 112. Springer, (2013)
107. Kassambara, A.: Machine Learning Essentials: Practical Guide in R. sthda, (2018)
108. De Myttenaere, A., Golden, B., Le Grand, B., Rossi, F.: Mean absolute percentage error for regression models. Neurocomputing **192**, 38–48 (2016)
109. World Population Review organization: Most Powerful Countries Population. `https://worldpopulationreview.com/country-rankings/most-powerful-countries` (2020)
110. World Population Review organization: Most War-torn Countries Population. `https://worldpopulationreview.com/country-rankings/war-torn-countries` (2021)
111. Hollis, M.E., Downey, S., del Carmen, A., Dobbs, R.R.: The relationship between media portrayals and crime: perceptions of fear of crime among citizens. Crime prevention and community safety **19**(1), 46–60 (2017)