Faculty of Sciences

Doctor of Philosophy in
Data Science
XXXIII Cycle

Measuring well-being through novel digital data

Candidate
Vasiliki Voukelatou

Supervisors:
Fosca Giannotti, Scuola Normale Superiore, Pisa, Italy
Luca Pappalardo, CNR, Pisa, Italy

March, 2022

To the rebellion of the soul against the intellect...

“Education is the most powerful weapon
which you can use to change the world.”

Nelson Mandela

Abstract

Well-being is an important value for people's lives, and it is crucial for societal progress. 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. The current Ph.D. thesis focuses on the safety dimension, and in particular on peace, which is an emerging challenge nowadays. Peace is the way out of inequity and violence, and its measurement is crucial, considering that the world is constantly under socio-economic, political, and military instability. Novel digital data streams and AI tools foster peace studies during the last years. Following this direction, we exploit information extracted from a new digital database called Global Data on Events, Location, and Tone (GDELT) to capture the Global Peace Index (GPI), a well-known official peace index. Applying predictive machine learning models, we demonstrate that news media attention from GDELT can be used as a proxy for measuring GPI at a higher frequency than the official yearly index cost- and time-efficiently. Additionally, we conduct variable importance analysis, and we use explainable AI techniques to understand better the models' behaviour, peace, and its determinants. This in-depth analysis highlights each country's profile and explains the predictions, prediction errors, and events that drive these errors. We believe that novel digital data exploited by researchers, policymakers, and non-governmental organisations, with data science tools as powerful as machine learning, could maximize the societal benefits and minimize the risks to peace and well-being as a whole.

Keywords: *AI for Social Good*, *Machine learning*, *Novel digital data*, *GDELT*, *news*, *Explainable AI*, *SHAP*, *Well-being*, *Peace*, *Global Peace Index*

List of publications

1. Voukelatou, V., Pappalardo, L., and Giannotti, F. (2019). Understanding societal well-being through the eyes of the news world media. HUSO 2019 : The Fifth International Conference on Human and Social Analytics

2. Voukelatou, V., Gabrielli, L., Miliou, I., Cresci, S., Sharma, R., Tesconi, M., and Pappalardo, L. (2020). Measuring objective and subjective well-being: dimensions and data sources. International Journal of Data Science and Analytics, 11(4), 279-309.

3. Voukelatou, V., Pappalardo, L., Miliou, I., Gabrielli, L., and Giannotti, F. (2020, October). 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). IEEE.

4. Voukelatou, V., Miliou, I., and Pappalardo, L. (2021). Stima dell'indice di pace attraverso notizie digitali. Lettura ragionata dell'Enciclica Papale "Fratelli tutti", alla luce dell'Obiettivo 16 dell'Agenda Onu 2030. Quaderni dell' ASviS (pp. 57-60)

5. Voukelatou, V., Miliou, I., Giannotti, F., and Pappalardo, L. (2022). Understanding peace through the world news. EPJ Data Sci. 11, 2. `https://doi.org/10.1140/epjds/s13688-022-00315-z`

Acknowledgements

With immense pleasure and deep sense of gratitude, I wish to express my sincere thanks to my supervisors Fosca Giannotti and Luca Pappalardo, who supported me by all means and challenged me during the Ph.D. period. Without their advice, teaching, inspiration, motivation, continuous encouragement and enthusiasm, this research would not have been successfully completed.

I wish to extend my profound sense of gratitude to Dino Pedreschi not only for making possible such a prestigious Ph.D. course but also for believing in me and my research objectives since the very first moment.

With pleasure, I am also grateful to Ioanna Miliou who closely collaborated with me. She patiently introduced me to the Computer Scientist world and generously shared with me all the programming and machine-learning “secrets”.

Special thanks go to Stefano Maria Iacus, Stan Matwin, Monica Pratesi, Chiara Roda for their feedback and inspiration and to Francesca Chiaromonte, Anna Monreale, Mirco Nanni, Lorenzo Gabrielli, Michella Natilli, Vittorio Romano, Francesca Pratesi, Roberto Trasardi, and Silvia Zappulla, who have rendered support to my research and my research-related activities. Additionally, I deeply thank Daniele Fadda for his invaluable support on data visualization. I also thank my Ph.D. fellows Cecilia Panigutti, Tommaso Radicioni, Giorgio Tripodi, Gevorg Yeghikyan, Jisu Kim, Luca Insolia and Elisa Ferrari for making together the initial steps, my fresh new team members Matteo Böhm, Giuliano Cornacchia, Giovanni Mauro and Massimiliano Luca for their feedback, and Armando Montanari and Barbara Staniscia for supporting me during the Ph.D. applications period.

Last but not least, I would like to emphatically thank my family and friends for supporting me personally and encouraging me whenever required, directly or indirectly. I warmly thank Donato for constantly encouraging me and believing in me and for the endless time spent giving feedback on my work. I deeply thank Vangelis, Sevi, and Natalia for supporting my dream and my choice of doing a Ph.D. From the depth of my heart, I also thank my sister Konstantina, my mother and father, for supporting my dream and encouraging me daily. Last, I thank John, Tasos, Paola, Theodosia, Luisa, Virginia, Anaïs, Antonio, Sonali, Yvonni, my Italian folk group friends...

Again, the words seem not to be enough to express how thankful to all these people I am.

TABLE OF CONTENTS

List of figures

List of tables

Chapter 1

Introduction

The global challenges regarding people's well-being that today's society faces are manifold. In a major attempt to face them, researchers of various disciplines, from psychologists to computer scientists, governments, non-governmental organizations, and policy-making centers, are actively working on identifying the most critical societal challenges related to well-being. This way, they can provide better decision support and solutions that can drive to higher levels of people's well-being, which is fundamental since a country's level of well-being reflects its societal progress [1, 2].

Since 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]. Therefore, the scientific community is focused on carefully specifying its measurable dimensions [5, 6]. For example, the United Nations Development Programme (UNDP) introduced the Sustainable Development Goals (SDGs) during the United Nations Conference on Sustainable Development in Rio de Janeiro in 2012 [7]. In total, the SDGs are 17 universal objectives [8, 9, 10], such as "Good Health and Well-Being", "No Poverty", "No hunger", "Gender Equality", "Reduced inequalities", and "Sustainable cities and communities". By realizing these objectives, the global community can ensure higher levels of societal well-being.

The data revolution and Artificial Intelligence (AI) play an important role in exploring well-being and realising the goals set. It is not surprising that the UN underlines that unprecedented availability of large-scale human behavioral data harnessed with AI tools can crucially contribute to the investigation of patterns of phenomena related to people's health and well-being [11, 12]. Data and AI are profoundly changing the world we live in and are the lifeblood of decision-making. Without data, we cannot know how many people are born and at what age they die. We cannot know how many men, women, and children live in poverty and how many children need education. We also cannot know whether greenhouse gas emissions are increasing or the fish stocks in the ocean are dangerously low, to name but a few.

It is therefore evident that AI for well-being is an emerging field. This thesis conducts complete research on the topic, from exploring the major measurable well-being dimensions to conducting an applied analysis with novel digital data and AI tools. Following this direction, we aim to demonstrate that novel digital data and AI tools can capture well-being dimensions and contribute considerably to research with traditional data.

We first introduce well-being and its two core concepts, objective well-being and subjective well-being. We then perform an in-depth review of the SDGs set by the UNDP, the Better Life Index (BLI) created by the OECD, and the project named "Benessere Equo e Sostenibile" (BES) that stands for "Fair and Sustainable Well-being" owned by the Italian National Institute of Statistics (ISTAT) [13]. We also review existing surveys and studies (such as the survey by Diener et al. [14]). We identify the most important and measurable dimensions of objective well-being, such as health and safety, and the dimensions of subjective well-being, such as the social environment. This theoretical analysis reveals that research in safety, particularly in peace that is interrelated with safety, is still at the very beginning. Indeed, to date, the use of big data and AI to foster research in the safety and peace field is lacking [15, 16], leaving ample space for contribution. The Expert Panel also confirms this on Technology and Innovation in UN Peacekeeping, which recognizes the importance of harnessing the data revolution for the benefit of the international community, safety, and peace [17].

The thesis's core content and most analytical part focus on AI for peace. Armed violence is constantly on the rise, and it is challenging to prevent it [18]. Since 2011, at least 100,000 people have been killed in deadly conflicts, with the majority of them in Afghanistan, Iraq, and Syria. Although the rate of major wars declined over the past decades, the number of civil conflicts and terrorist attacks increased in the last few years, even in developed countries [16]. In addition, the war expenses for the war-torn countries weaken their economies. For example, since 1996, the Democratic Republic of Congo has spent on war almost one-third of its gross domestic product [19]. For this reason, recently, the UNDP included on their goals the SDG 16, i.e., "Peace, Justice, and Strong Institutions".

Official indexes created to measure peace usually fail to give an early warning of socio-economic, political, or military events. Governments and the international community often have little warning of abrupt changes in peace and safety. For example, the Global Peace Index, the world's leading measurement of national peace, produced by the Institute for Economics and Peace [20], is an annual index and, thus, it neglects short-term fluctuations of peace. The reason behind the yearly GPI fluctuation is, among others, the measurement of GPI by institutional surveys and governmental data, which are usually expensive and time-consuming [4]. Thus the final estimates are produced only after a series of yearly revisions.

Consequently, the main objective of this thesis is to demonstrate that an official peace index, such as the GPI, can be estimated with the use of AI and novel digital data at a higher time-frequency than the official index score. Therefore, the main research question of the thesis is:

- *RQ1: Can we measure GPI through novel digital data and AI tools at a higher time frequency as compared to the official index?*

To tackle this task, we exploit machine learning and information extracted from a digital database called Global Data on Events, Location, and Tone (GDELT) [21]. 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. News media records generally describe a variety of subject domains (e.g., economic events, political events) and represent a wide range of targets (e.g., opposing politicians) [22]. Considering that GDELT is a free access database updated daily, it can contribute to the monthly estimation of GPI as compared to the real annual GPI. Besides, GPI through GDELT is produced at a low cost and time-efficient way, compared to the traditional methodology.

We perform our analysis for all countries around the world. In particular, our models exploit the information from GDELT to provide GPI predictions from 1-month-ahead up to 6-months-ahead. Our results demonstrate that GDELT variables are a good proxy for measuring GPI at a monthly level. There are country models that show high performance, such as the United Kingdom and Yemen, countries that show medium performance, such as Chile and Libya, and others that show low performance, such as Estonia and Cyprus. The reasons for the low model's performance could be various, such as the under-representation or over-representation of some countries through the GDELT news [23].

Furthermore, we deepen our analysis by setting an additional research question:

- *RQ2: Can we explain peace and its determinants for each country?*

To tackle this task, we conduct variable importance analysis and use explainable AI (XAI) techniques [24, 25, 26]. We identify the relationships between the GDELT variables and peace and explain the models' behavior. This analysis allows us to unveil 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.

Frequent estimation updates of the GPI score through the GDELT database could flag conflict or war spots months in advance and reveal considerable month-to-month peace fluctuations and significant events that would be otherwise neglected. For example, even though Yemen is a war-torn country, and it is currently at war, it might be the case that a month is less warn-torn. This fluctuation can be easily observed by monthly GPI measurements and can be difficult to capture by yearly GPI measurements.

Overall, we believe that the research conducted for the purposes of thesis offers great value to the scientific community and especially to researchers interested in the so called "Data Science for Social Good" (DS4SG) field or similarly "Artificial Intelligence for Social Good" (AI4SG) [27]. Particularly, it can be beneficial to non-governmental organizations, such as the United Nations and its agencies, to organize early peace interventions. Similarly, this study could be extended to any well-being dimension, to foster well-being research and decision-making. This way, detrimental societal effects could be prevented, and the world could be closer to lasting societal progress.

Structure of the thesis

The present Ph.D. thesis is a cumulative thesis. The content of the chapters are essentially modified versions of papers published in the course of the Ph.D. program [4, 28, 29, 30]. In what follows, we present the thesis structure and a brief synoptic outline of each chapter.

- In **Chapter 2 - State of the art**, we introduce the reader to core concepts related to the topic of the Ph.D. work, and we provide an extensive literature review. We firstly start from the general concept of interest, i.e., well-being. We introduce the theory of well-being, present its dimensions, and focus on the safety dimension and its interrelation with peace. Furthermore, we present the novel digital data used to measure peace and peace- or safety-related indicators, as well as their advantages and disadvantages. Last, we present studies that use novel digital data and AI technologies to study peace.

- In **Chapter 3 - Setting the stage**, we set the stage for the research. We first define the objectives of our work and formulate the main research question (RQ1) we attempt to answer. Secondly, we describe the data used for the conduction of the research, i.e., the official GPI (ground truth data), the GDELT data and the variables created from the GDELT data (exogenous data), and we discuss which of these variables could cover the GPI indicators. In addition, we present the prediction models we use for the predictions of the GPI. Last, we introduce the SHAP methodology, which we use to answer the additional research question (RQ2), to interpret better our models and explain the results or errors in the predictions.

- In **Chapter 4 - Measuring peace through the world news** we mainly focus on the RQ1. Particularly, we present the methodological approach used to conduct the research and the estimation framework we apply to measure and predict the GPI. Then, we present the models' results and validate the GPI predictions from 1-month-ahead to 6-months-ahead. We also compare the results between the machine learning models, and we illustrate that countries demonstrate stable behavior between the different algorithms. Last, we present and interpret the results of country models which show high, medium, or low performance.

- In **Chapter 5 - Understanding peace**, we mainly focus on the RQ2. In particular, we conduct variable importance analysis, and we apply explainable AI methodologies for the high performance models. We show that the most important variables that the models use for the predictions reveal the profile of the country, e.g., Yemen has a war-torn country profile. In addition, the explainable AI techniques contribute to the interpretation of the models' results, explain peace and its determinants throughout the months, and help to understand better the models' behavior, such as large errors produced in the predictions.

- In **Chapter 6 - Conclusion**, we summarise the main research objectives of the thesis, the methodologies we use to tackle them, and the main findings. We also discuss the advantages and we point out drawbacks and biases of the presented approach in particular, and for well-being studies in the digital era in general. Last, we extensively discuss future research lines and conclude the thesis.

Chapter 2

State of the art

2.1 Well-being and the relevant dimensions

The concept of well-being is crucial for the societal progress, despite being vague and not clearly defined up to date. Economists and policy-makers have traditionally considered Gross Domestic Product (GDP) as a good indicator of well-being in society, mainly because it is linked with the standard of living indicators [31]. However, GDP has been criticized as a weak indicator of well-being and, therefore, a misleading tool for public policies [2]. Although GDP reflects current economic activity, it ignores the destruction of the natural environment, safety, health, and other factors associated to well-being. Consequently, in 2009, the Stiglitz Commission [32] observed that other tools should be used, complementary to GDP, for the measurement of well-being. In line with the aforementioned criticisms and suggestions, non-governmental organisations and researchers with various backgrounds, from economists to psychologists, created alternative ways to measure well-being. For example, the United Nations Development Programme (UNDP) created the Human Development Index (HDI) [33], which evaluates the extent people have a long and healthy life, are knowledgeable and have a decent standard of living. Similarly, the OECD organisation created the Better Life Index (BLI) [34] which evaluates other well-being dimensions, such as safety, life-satisfaction, and work-life balance.

Well-being dimensions might be either of objective or subjective nature. Indeed, researchers working on the field have suggested two type of concepts for measuring the well-being, i.e., the objective well-being, which is represented by the objective dimensions, and the subjective well-being, which is represented by the subjective well-being dimensions. For example, it might be different how healthy an elderly is, and how healthy an elderly feels [35].

Since defining objective well-being is a challenging task, researchers have focused on exploring its dimensions rather than its definition [5, 6]. For example, the Organisation for Economic Co-operation and Development (OECD) has identified 11 essential topics labeled as OECD well-being framework [3]. Similarly, the United Nations Development Programme (UNDP) has identified 17 sustainable development goals (SDGs) [8]. From the initiatives mentioned above, it is evident that different institutions propose different well-being dimensions, which are sometimes vague and hard to capture. Therefore, as already anticipated in Section 1, we identify six concrete, objective and measurable dimensions of well-being: health, job opportunities, socio-economic development, environment, safety, and politics.

On the contrary, subjective well-being, also called happiness, is commonly defined as the degree to which an individual assesses the overall quality of her life-as-a-whole favorably [36]. This might as well be different as compared to GDP, which cannot be representative of societal happiness. Indeed, GDP explains only a small proportion of its variations on humans [37] and might be different from people's perceptions of their well-being [38]. Subjective well-being is traditionally captured through studies based on data collected by self-reports. These studies highlight five main dimensions of subjective well-being: the role of human genes, which is fairly heritable [39, 40, 41, 42, 43, 44, 45, 46, 47]; universal needs, meaning basic and psychological needs [48, 49, 50]; social environment, such as education and health [51, 52, 53, 54, 55]; economic environment, including a lot of research on income [56, 57, 58, 59, 60]; and political environment, such as democracy and political freedom [61, 62].

Taking into consideration the multi-dimensionality of both objective and subjective well-being, studies are usually focused on exploring one of the dimensions (e.g., [63, 64, 65, 66, 67]). This approach is also followed for the purposes of the current thesis. Specifically, this thesis aims to explore safety, an objective well-being dimension.

2.2 Safety and its interrelation with peace

Safety is one of the most essential objective well-being dimensions for a democratic society [4]. Particularly, nowadays that the world is under socio-political instability and under constant conflicts, safety is a core element for people's well-being. According to BLI, the safety dimension covers the risk of people being physically assaulted or falling victim to other types of crime. Crime may lead to loss of life and property, as well as physical pain [68]. For example, based on the latest OECD data, the average homicide rate in the OECD countries is 3.6 murders per 100,000 inhabitants [68]. Besides, the Italian BES project [13] suggests that safety is characterized by two determinants, i.e., criminality and violence. Particularly, the concept of safety covers substantially the targets "Reduce violence everywhere", "Combat organized crime and illicit financial and arms flows", and "Strengthen national institutions to prevent violence and combat terrorism and crime", which are created to promote the SDG 16 [69], i.e., peace justice and strong institutions. Thus, considering that peace is associated with a world free of violence and war [70], where individuals live in a safe environment, the concepts of safety and peace can be addressed with interrelation.

In particular, according to Galtung [70] a common conception of peace is Negative Peace, or actual peace. Negative peace is the absence of violence and the absence of war. Achieving negative peace is often the first goal for maintaining a peaceful society, as outright violence is an obvious indicator that a society is not peaceful. In addition, it is negative because something undesirable stopped happening (e.g. the violence stopped, the oppression ended). Researchers, policymakers and peacekeepers use this definition as flagship for developing and measuring peace indicators [71]. For example, the Global Peace Index (GPI), the world's leading measurement of national peace produced by the Institute for Economics and Peace (IEP) measures negative peace [20] (we describe GPI in detail in Section 3.2.1).

Similarly to most official well-being related indexes, such as HDI and BLI, the GPI is captured by official data, like surveys and governmental data, economic data, police data, etc. Likewise, researchers study peace or peace indicators related to violence with the use of official data. For example, Brückner and Ciccone [72] use official conflict, economic, and environmental data to examine whether civil wars outbreaks are more likely due to certain economic conditions in Sub-Saharan African countries. Hegre et al. [73] use demographic, mortality, and other official data to predict changes in global and regional incidences of armed conflict for the 2010–2050 period. Furthermore, Chadefaux [74] uses financial data of government bond yields to show, among others, that wars involving democracies lead to greater market shocks.

However, traditional data might bring biases and limitations. For example, although data collected through surveys have been proven to be valid, they are costly, time-consuming [75, 4], and might include errors brought from social desirability biases due to participants' inaccurate answers [76, 4]. In addition, socio-economic data and other governmental data are hard to collect, not yearly updated and could have a lag of up to two or three years. Thus, they might not be accurately representing the corresponding year of the peace measurement.

As conflicts and violence become increasingly complex, policymakers and peacekeepers search for novel approaches to tackle the growing challenge. The revolution of digital data and AI may help overcome the aforementioned difficulties providing cost-efficient, time-efficient, and almost real-time estimates of peace. In other words, novel digital data harnessed with AI techniques are powerful tools to measure peace or peace-related indicators, produce early warnings of peace changes, and complement the estimations produced from official data. The crucial role of the novel digital data and AI 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 [77]. However, the application of new technologies and novel data in the peace area is, in comparison to the rest of SDGs, still in an early phase.

2.3 Peace in the digital age

In line with the UN strategy, independent researchers explore data-driven and technology–based solutions for the pursue of peace. Similarly to other well-being dimensions, safety, peace or peace-related indicators are captured by social media data, mobile phone records, GPS data, web search queries, crowdsourced data, and news data.

Social media data

Social media platforms such as Twitter, Facebook, and Instagram can be considered a digital database of information about online users, hence rendering individuals' online activities accessible for analysis. Given this enormous potential, researchers, governments, and corporations are turning their interest on social media to understand human behavior and interactions better [78]. Among all social media, Twitter is the most popular, since it provides public access to data through APIs with the least restrictive policy. The Twitter APIs return information about locations, date of the event, interactions with other users, or tags inserted in the tweet. Twitter also returns some information about the user profile. However, despite their indubitable usefulness, social media data may also encounter some concerns [79]. First of all, they may reflect social desirability biases, since individuals manage their online profiles [80]. Besides, social media users may not be as representative of the general population as traditional anonymized self-reports conducted through a chosen representative sample [65].

Social media data are primarily used to assess indicators related to peace since they render individuals' online activities accessible for analysis. For example, Curiel et al. [81] collect Twitter data in 18 Spanish-speaking countries in Latin America and classify the tweets as crime-related or not. By comparing the number of tweets related to crime against the number of murders from official data they demonstrate that tweets reflect the fear of crime. Chen and Neill [82] use Twitter data and propose a methodology that can forecast with high accuracy and lead time civil unrest events. Similarly, Spangler and Smith [83] analyze Twitter data to explore public dissent and civil unrest. Particularly, they demonstrate that the estimates of public dissent in Canada and Kenya can predict civil unrest events days before they occur in both countries. Additionally, Twitter data are used to study early detection of the global terrorist activity [84]. For example, Azizan and Aziz [85] collect Twitter data to study terrorism. Specifically, applying machine learning techniques on users' tweets they sense their act leading to terrorism. Moreover, Zeitzoff [86] uses Twitter data and other social media sources to analyze the short-term dynamics of the Gaza Conflict (2008–2009). He measures changes in Israel's and Hamas's military response dynamics. Particularly, he demonstrates that both sides responses to provocations increase shortly after the ground invasion. Additionally, after the UN Security Council vote, Israel's response decreases, whereas Hamas's slightly increases. Siapera et al. [87] analyze tweets posted in the period of the Operation Protective Edge (July 2014). Specifically, they use data mining and sentiment analysis techniques to identify and understand how the Gaza attack over the summer of 2014 was mediated. Also, Tucker et al. [88] analyse Twitter data to study public violence and private conflict in Boston. Zeitzoff et al. [89] use Twitter data to demonstrate that some foreign policy networks, such as English and Farsi Twitter networks can accurately reflect policy positions and salient cleavages. Zagheni et al. [90] use Twitter data for users in OECD countries from May 2011 to April 2013. They present an approach which can be used to predict turning points in migration trends and understand the relationships between internal and international migration. In addition, Kadar et al. [91] use Foursquare data to describe urban crime. Furthermore, Zagheni et al. [92] query data from the Facebook's advertising platform to show the feasibility of nowcasting stocks of migrants within and across countries and discuss the limitations of the data. Finally, Mazoyer et al. [93] have created a French corpus of 38 million tweets, from July to August 2018, annotated for event detection tasks, such as conflict, war and peace, crime, and justice.

Researchers do not exclusively use social media data to study peace, but also a combination of official and Twitter data. For example, Chen et al. [94] use Twitter data combined with official weather data, to predict the time and location in which a specific type of crime will occur in Chicago in the United States. In addition, Alexander et al. [95] combine Facebook and surveys data and propose a statistical framework to produce timely nowcasts of migrant stocks from Mexico, India and Germany, by state in the United States.

Mobile phone records

Mobile phone records collect geographical, temporal, and interaction information on mobile phone use [96, 97, 63, 64, 98, 99, 100], hence providing a comprehensive picture of human behavior at a societal scale. Each time an individual makes a call, the mobile phone operator registers the connection between the caller and the callee, the duration of the call, and the coordinates of the phone tower communicating with the served phone. Researchers use mobile phone records since they offer an additional advance; that is the calling and texting activity of users, because they guarantee the repeatability of experiments in different countries and on different scales given the worldwide diffusion of mobile phones [101]. Note that mobile phone records suffer from different types of bias [102, 103]. For example, the position of a user is known at the granularity level of phone towers, and only when they make a phone call. Moreover, phone calls are sparse in time, i.e., the time between consecutive calls follows a heavy tail distribution [104, 105]. In other words, since users are inactive most of their time, mobile phone records allow reconstructing only a subset of a user's behavior.

Particularly, many researchers study peace by analysing mobile phone records, usually in combination with traditional data as well as novel digital data. For example, Bogomolov et al. [106] use mob records for 3 weeks from the 9th to the 15th of December 2012, and from the 23rd December 2012 to the 5th of January 2013, in combination with demographic data from December 2012 to January 2013, to predict crime in the city of London. Experimental results show 70% of accuracy in predicting whether an area could be a crime hotspot or not. Wu et al. [107] combine mobile phone data with official data, such as poverty and unemployment statistics, to study mobility-based crime predictions. Specifically, they present a novel model that utilizes domain knowledge about biases in reported crime data to characterize and enhance fairness and accuracy in mobility-based crime predictors. In addition, Ferrara et al. [108] use mobile phone records to deeply understand hierarchies within criminal organizations, discover members who play central role and provide connection among sub-groups. Their work concludes illustrating the adoption of our computational framework for a real-word criminal investigation. Moreover, researchers combine social media data with mobile phone records to infer migration events and population movements [109]. For example, Chi et al. [110] use four years of mobile phone data from Rwanda's near monopoly mobile phone operator, and three years of Twitter data to study migration. In particular, the researchers propose a new segment-based approach to measuring migration.

GPS data

GPS data can track the movements of the individuals [111, 112, 113, 114] providing time and location coordinates information, which can be used to link locations with environments and to calculate the speed of movements [66]. For insurance reasons, some vehicles have on-board devices that record the vehicle's position at regular intervals and sends it to a GPRS server [115, 97]. GPS data can also cover rural areas, as opposed to other data, mostly collected among citizens of urban areas [116]. Comparing to the traditional data, usually extracted by self-reports assessed with questionnaires, GPS does not bring any biases and misclassification, [117, 116], as it eliminates the social desirability usually brought by self-report participants [118, 119]. Another advantage of GPS data is that they provide real-time monitoring.

However, while there are studies based on GPS data covering hundreds of thousands of individuals [115], most of the GPS studies are conducted with a few participants [120, 116], usually due to privacy issues. Apart from this drawback, when a GPS is used indoors, the spatial accuracy of the measurements is fairly detected [121], which might create problems in research.

GPS data are used for a variety of peace-related studies, such as crime, safety, civil unrest, protests, to name a few. For example, Robinson et al. [122] collect GPS, accelerometer and personal data to study the relationship between the spatial distribution of crime incidences and moderate-to-vigorous physical activity (MVPA) among adolescents in Massachusetts between 2011 and 2012. Applying correlation and regression analysis they demonstrate that a strong positive association between crime and adolescent MVPA. Daviera etl al. [123] collect GPS data for eight days in the city of Chicago, as well as subjective and objective data. Conducting analysis of geo-narratives they show that perceptions of safety and danger are related to environmental, social, and temporal cues. Moreover, Robinson et al. [122] collect GPS, accelerometer and personal data to study the relationship between the spatial distribution of crime incidences and moderate-to-vigorous physical activity (MVPA) among adolescents in Massachusetts between 2011 and 2012. They demonstrate that a strong positive association between crime and adolescent MVPA. Daviera etl al. [123] collect GPS data for eight days in the city of Chicago, as well as subjective and objective data. They show that perceptions of safety and danger are related to environmental, social, and temporal cues. Ariel et al. [124] use GPS data to replicate findings published from US official research on the effect of hot spots policing for the prevention of crime in England and Wales and demonstrate that victim-generated crimes increase in both the near vicinity and in catchment areas.

Web search queries data

Web search queries data report the frequency of specific terms over time, entered into a web search engine from users to satisfy their information needs. Data are represented as time series of the frequency. Comparing to other data sources that require customized and often complicated collection strategies, search data can be collected for many domains simultaneously. They can also be easily analyzed across several countries or regions in real-time. Search data are often helpful in making forecasts. However, their utility for predicting real-world events is based on convenience, speed, and flexibility and has less to do with their superiority over other data sources. Goel et al. [125] provide a useful survey in this area and describe some of the limitations of this data source. First, for different domains, the size of the relevant population varies considerably, along with difficulty in identifying relevant queries. Additionally, in specific domains, searching may be more closely tied to the measured outcomes than in others.

In particular, researchers use Google Trends to study peace-related indicators. For example, Qi et al. [126] show that a simple low-level indicator of civil unrest can be obtained from online data at an aggregate level through Google Trends or similar tools. The study covers countries across Latin America from 2011 to 2014 in which diverse civil unrest events took place. In each case, they find that the combination of the volume and momentum of searches from Google Trends surrounding pairs of simple keywords, tailored for the specific cultural setting, provide useful indicators of periods of civil unrest. Similarly, Qi et al. [127] study online search activity from Google Trends surrounding the topics of social unrest over several countries in Latin America during 2011–2014. They find that the volume and momentum of searches surrounding mass protest language, can detect – and may even pre-empt – the macroscopic on-street activity. They also find that the most crucial search keywords differ subtlety from country to country, even though the language may be the same. They explain this by the fact that civil unrest is a time-varying coordinated interaction between individuals, groups, or populations within a given cultural and socio-economic setting. Muchow and Amuedo-Dorantes [128] use data on calls for service dispatched to LAPD patrols from 2014 through 2017 to assess if heightened awareness of immigration enforcement, as captured by a novel Google Trends index on related searches, is associated with reduced calls to report domestic violence in predominately Latino non citizen neighborhoods. They show that domestic violence calls per capita dropped in LAPD reporting districts with a higher concentration of Latino non citizens as awareness about immigration enforcement increased.

Crowdsourced data

Crowdsourced data is another promising data source. In 2008, Kleemann and Rieder [129] have defined crowdsourcing as the "the intentional mobilization for commercial exploitation of creative ideas and other forms of work performed by consumers". In other words, crowdsourcing involves obtaining work, information, or opinions from a large group of people who submit their data via the Internet, smartphone apps, etc. Naturally, crowdsourcing brings several advantages. Crowdsourcing can provide researchers with a huge amount of data, which can be accessed quickly and at a relatively low cost. Besides, comparing to traditional research (such as studies using traditional surveys), the use of crowdsourcing can provide researchers with data from samples that are more diverse [130]. However, crowdsourcing yields various challenges. Firstly, it may bring relatively low-quality results, e.g., a participant of a crowdsourced study may intentionally give wrong answers. Secondly, mobile platforms pose new challenges for crowdsourced data management.

Researchers use crowdsourced data to study various concepts related to peace. For example, Goodney Lea et al. [131] collect data through the Safecity.in crowdsourced platform to map violence against women in 2012 in response to a brutal attack on a woman in India who was out with her boyfriend at 8.30 p.m. They identify patterns, categorize the assaults by location and type and propose a crowdmapping tool which it can allow contribute to danger awareness and women empowerment. Palakodety et al. [132] use YouTube data for analyzing the international crisis between India and Pakistan for the dispute over Kashmir. In particular, they argue, among other, the importance of hope-speech detection, which automatically detects web content that could play a positive role in diffusing hostility on social media triggered by heightened political tensions during a conflict. Ozkan et al. [133] use crowdsourced police-involved killings data from FatalEncounters.org, as well as media data, to control whether police killings is counted and reported correctly in the aforementioned unofficial data, as compared to official data in the city of Dallas. Results mostly show consistency between all data sources. Additionally, Rumi et al. [134] use crowdsourced check-in and real-time emergency and propose a model, which is efficient in preventing crime events and robust to emergency situations.

News data

Last, news data contain information extracted from newspapers around the world. They generally describe a variety of subject domains (e.g., economic events, political events), represent a wide range of targets (e.g., opposing politicians) [22] and are continuously updated, containing even archived historical news of the last decades. Nevertheless, news data contain three main biases [135]. The gatekeeping bias, i.e., the editors or the journalists decide on which event to publish; the coverage bias, related to the coverage of an event (e.g., western countries are over-covered, whereas African countries are under-covered); the statement bias, when the content written by the journalist, even if tried to be objective, is favorable or unfavorable towards certain events.

Researchers combine news data, such as ACLED [136] and GDELT [21], with other official data to capture peace indicators [137, 138]. For example, Ide [139] uses ACLED data to assess of the impact of COVID-19 on armed conflict based on data from the first six months of 2020. He shows, between others, that the armed conflict in Colombia between the government and the Ejército de Liberación Nacional de-escalated considerably after the pandemic struck the country, whereas in India it seems that the rebels use the lack of state presence and economic deprivation caused by a heavy lockdown to recruit for future offensives (Bhardwaj, 2020, Kujur, 2020). Hossain et al. [140] use news data to present, among others, an approach to converting predictions of the proposed models to real-world warnings. Particularly, they extract features from the Arabia Inform news articles [141], a corpus of news documents originating from North Africa (MENA) countries to predict violent events in the Middle East and MENA region over a year from 1 August 2016 to 30 September 2017. For evaluation they use manually curated violent MANSA events, called the gold standard report (GSR), which is provided by the Center for Analytics at New Haven. [140].

The Global Data on Events, Location, and Tone database (GDELT) is a major news data source that describes the worldwide socio-economic and political situation through the eyes of the news media, making it ideal for measuring well-being and peacefulness [21]. GDELT is mainly used to explore social unrest, protests, civil wars and coups, crime, migration, and refugee patterns. For example, Qiao et al. [142] use GDELT to build a framework that predicts indicators associated with country instability. The framework utilizes the temporal burst patterns in GDELT event streams to uncover the underlying event development mechanics and formulates the social unrest event prediction in five countries in Southeast Asia. Galla and Burke [143] use themes and events from the GDELT database associated with social unrest. They apply machine learning techniques to identify regions at state and county level in the United States where social unrest might occur in near future. Alsaqabi et al. [144] use GDELT data to predict the crime distribution over Saudi Arabia and to provide the indicators of specific areas which may become a criminal hotspot. Their main goal is to demonstrate that their suggested feature selection is more accurate than others, using the example of the Saudi Arabia. Joshi et al. [145]

created Social Unrest Reconnaissance Gazetteer (or SURGE), a Webbased application that provides an open system to visualize and integrate spatio-temporal data about social unrest events in South Asia. The system displays eight categories of unrest for India, Pakistan and Bangladesh, based mainly on the GDELT database and the Global Terrorism Database (GTD). This is an important system since it contributes, between other, to deeper analysing the data, finding patterns, trends and outliers, but also visualising and communicating important information. Qiao et al. [146] use GDELT data to predict social unrest events in Thailand. Particularly, they suggest a new more effective framework which extracts historical events captured from GDELT to characterize the transitional process of the social unrest events' evolutionary stages, uncovering the underlying event development mechanics and formulates the social unrest event prediction.

As discussed above, GDELT is also used for the exploration of severe internal and external conflicts. For example, Keertipati et al. [147] use GDELT data to study the Sri Lankan civil war, and the 2006 Fijian coup. Specifically, they demonstrate that the data extracted from news items can capture the global events accurately. Therefore, they create a framework which can effectively identifies significant conflicting events. Yuan et al. [148] use cooperative and conflictual scales of event data from GDELT to investigate the interactions between the USA, Russia and China after the end of the Cold War and particularly in two periods: from January 1991 to September 2001 and from October 2001 to December, 2016. Their results provide insights into the direct interactions between the three dyads and helps the scientific community to understand their interactions better in the post-Cold War period. As an example, the United States was always an essential factor in affecting the interactions between Russia and China in both periods, but China's behavior only played a limited role in influencing the interactions between the United States and Russia dyad. GDELT is also used from researchers who study violence related issues. For example, Yonamine [149] uses GDELT data to forecast future levels of violence in Afghanistan. He demonstrates that the forecast accuracy decreases as the degree of geo-spatial aggregation increases, i.e., forecasts at the district-month, province-month and country-month level. He also observes that a major spike in violence during a specific period of time in a specific sub-state location is followed by a rebound-effect. Merari and Germán [150] are motivated to use the GDELT data to discover violence-related issues in Iraq, due to the lack of open governmental data. Analysing the data they discover violence-related social issues in terms of refugees, humanitarian aid, violent protests, fights with artillery and tanks, and mass killings. They also created a software which software classifies the zones of Iraq with available or unavailable data by using the latitude and longitude values of the area they focus their study. Qiao et al. [126] use GDELT data to detect protest events. Particularly, they demonstrate that their suggestion of a novel graph-based framework can more effectively detect the "Occupy Wall Street" in New York and the "Occupy Central" in Hong Kong as compared to baseline models. Keneshloo et al. [151] use GDELT data to forecast domestic political crisis in Brazil, Colombia, Mexico, Argentina, and Venezuela. The data are collected from January 2003 to December 2013. The researchers demonstrate the use of frequent subgraph mining to identify signatures preceding domestic political crisis, and the predictive utility of these signatures through both qualitative and quantitative results. Lastly, news data from GDELT are combined with other data sources, such as socio-economic indicators to study various peace related indicators. For example, Ahmed et al. [152] use GDELT data combined with socio-economic indicators to study migration. GDELT data help capturing information about policy changes, such as the close of the Hungary border and other external events provided from the news data. The authors use the European refugee crisis as a case study to present a system for scenario analysis and forecasting of mass migration. the mobility Beine et al. [153] combine refugee data with GDELT data to study the mobility of refugees across provinces in Turkey. Applying standard econometric techniques they demonstrate that non-refugees move further and more frequently compared to refugees, and that the standard determinants of mobility for immigrants also apply for refugees, i.e., income of the origin province, distance between provinces and network effects. Bertolini et al. [154] combine housing market data with GDELT data to conduct exploratory analyses to illustrate the possible research avenues. Particularly, they show there is a positive correlation between events and refugee call volume while the housing data reveal that real estate prices did not increase as much as expected with the increase, even in the Southeast, the region with the largest relative number of refugee inflows.

Carammia et al. [155] combine GDELT with Google Trends, and official migration data to study the asylum applications lodged in countries of the European Union by nationals of all countries of origin worldwide. Their approach monitors potential drivers of migration in countries of origin to detect changes early onset, models individual country-to-country migration flows, estimates the effects of individual drivers, provides forecasts of asylum applications up to four weeks ahead, and assesses how patterns of drivers shift over time.

However, despite the aforementioned studies, the peace and security pillar is still in the early stages of exploring data-driven and new technology–based solutions. Up do data most studies using novel digital data focus on exploring peace-related indicators, and do not study peace as a composite concept. Furthermore, the existing studies focus their analysis at a country level, and not at a global level. Therefore, this thesis is different from previous work in two important aspects. First, GDELT news data are harnessed with AI techniques to estimate a composite peace index as GPI, which covers domestic and international conflicts, safety and security, migration phenomena, etc. The wide variety of GDELT event categories can cover most GPI indicators. Second, we perform our analysis at a global scale to study peace over all countries in the world.

We believe that the news data contribute to the advances made through this research since they hold two considerable advantages for peace-related studies as compared to the rest of the data (Table 2.1). Firstly, news data can cover all countries in the world. For example, GDELT covers all countries since 1979. This allows researchers to conduct peace and safety analysis at a global scale. Secondly, available news databases usually cover a pre-defined wide variety of socio-political events. For example GDELT covers more than 200 event categories. Therefore, they can be powerful in covering many peace and safety GPI indicators. Last, GDELT database is preferred from ACLED database due to its time and geographical coverage. GDELT covers all countries since 1979, whereas ACLED is limited. It covers Asia since 2010, the Middle East since 2016, and Europe since 2018.

Table 2.1 provides a summary of the novel digital data sources used for peace studies, highlighting the pros and cons of each one.

Data Source	**Pros**	**Cons**
mobile phone records	temporal and social dimensions, world wide diffusion, repeatability	not publicly available, sparsity, geographically imprecise
GPS	coverage of rural areas, unbiased and classified, real-time monitoring	privacy issues, indoor spatial inaccuracy
Social Media	measuring social dynamics, publicly available	privacy issues, overrepresentation, social desirability bias
Web Search	publicly available, speed, convenience, flexibility, ease of analysis	population size varies across domains, hard identifying relevant queries
Crowdsourcing	large number of data, speed, relative low cost	risk of low-quality results, trade-off between quality and cost
News	variety of subject domains, range of targets, all countries coverage, archived historical news	gatekeeping bias, coverage bias, statement bias

Table 2.1 Pros and cons for each data source used for the measurement of well-being dimensions, including safety and peace.

Chapter 3

Setting the stage

3.1 Problem formulation

Let $Y^j = Y_1^j, \dots, Y_T^j$ denote the time-series of T yearly real-valued GPI observations for a country j, with each $Y_i \in \mathbb{R}$. In addition, let $X^j = \{\mathcal{X}_1^{j,n}, \dots, \mathcal{X}_t^{j,n}, \dots, \mathcal{X}_{t+k}^{j,n}\}$ denote a set of monthly exogenous GDELT variables of event counts for a country j. We aim to define a function $f^j(.)$, which measures the GPI at a monthly frequency and produces monthly GPI predictions $y^j = y_{t+1}^j, \dots, y_{t+k}^j$ for k-months-ahead, $t+1, \dots, t+k$, given the corresponding GDELT variables $\{\mathcal{X}_{t+1}^{j,n}, \dots, \mathcal{X}_{t+k}^{j,n}\}$ as input. The function is constructed taking into account both the historical observations of GPI, as well as the exogenous GDELT variables up to time t.

Below follows Section 3.2 which presents an exploratory analysis and more details on the data-preprocessing for the objectives of this study. In particular, Section 3.2.1 focuses the GPI data, Section 3.2.2 focuses on the GDELT data, and Section 3.2.3 presents the matching of the GDELT data with the GPI indicators.

3.2 Datasets

3.2.1 Global Peace Index (GPI)

The Global Peace Index (GPI) [156], created by the Institute for Economics and Peace (IEP), measures the relative position of countries' peacefulness. The index ranks 163 independent states and territories according to their level of peacefulness. GPI data are available from 2008 until 2020 at a yearly level (see, e.g., GPI report 2020 [20]).

The GPI is constructed from 23 indicators related to Ongoing Domestic and International Conflict, Societal Safety and Security, and Militarisation domains [20]. In particular:

- Ongoing Domestic and International Conflict includes: “Number and duration of internal conflicts”, “Number of deaths from external organized conflict”, “Number of deaths from internal organized conflict”, “Number, duration and role in external conflicts”, “Intensity of organized internal conflict”, and “Relations with neighbouring countries”.

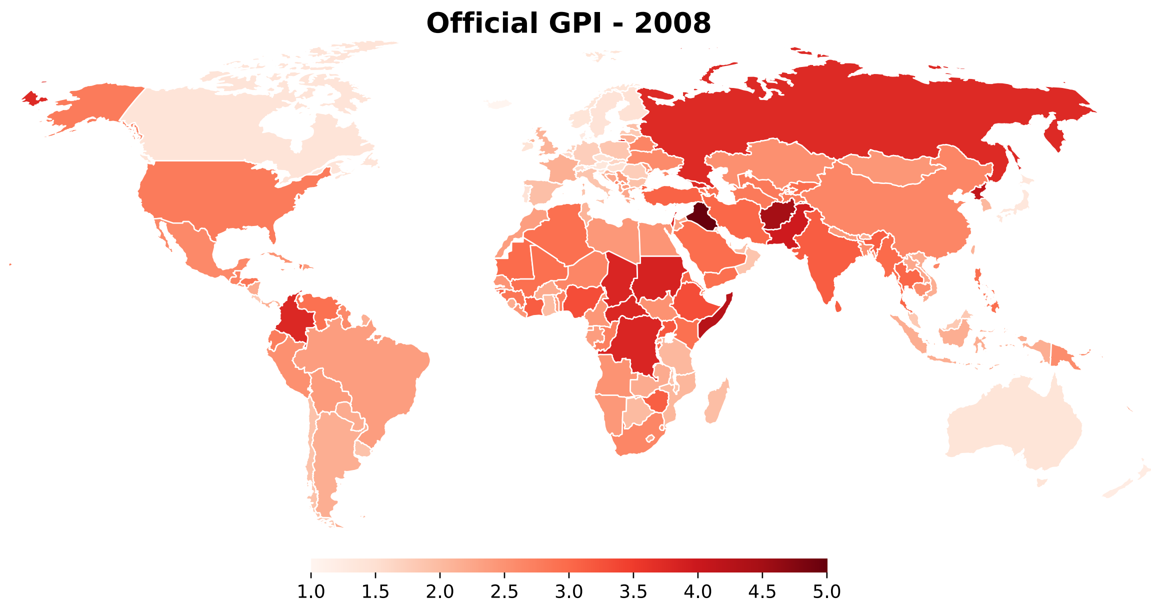
- Societal Safety and Security encompasses: “Level of perceived criminality in society”, “Number of refugees and internally displaced people as a percentage of the population” , “Political instability”, “Political Terror Scale”, “Impact of terrorism”, “Number of homicides per 100,000 people”, “Level 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”.

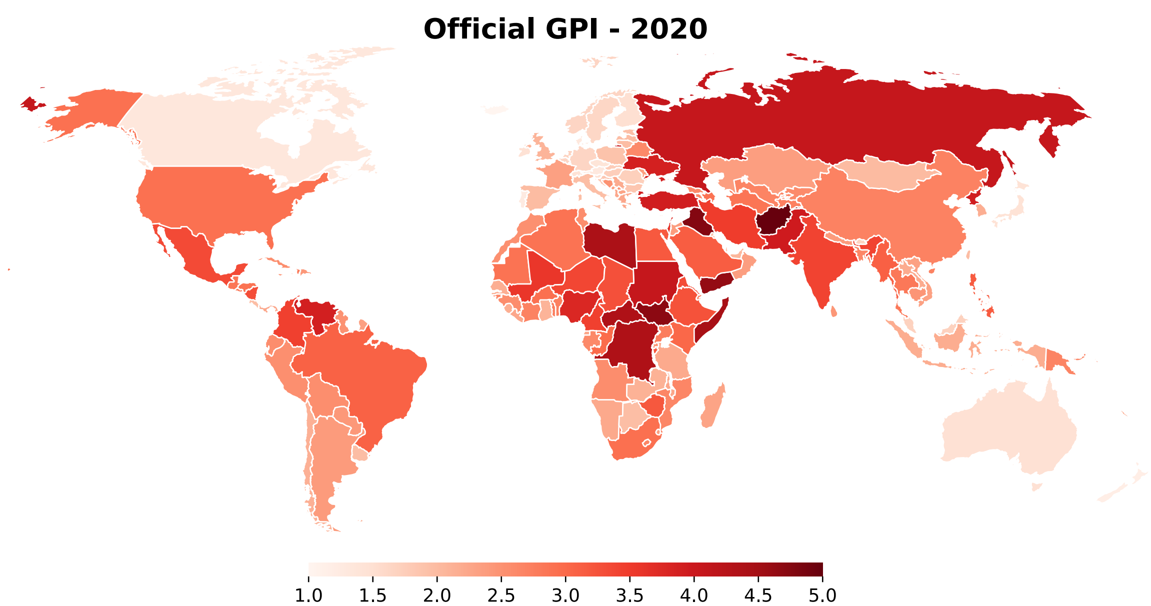
For the construction of each GPI indicator presented above data are derived from official sources, such as governmental data, institutional surveys, and military data. The indicators are then weighted and combined into one overall score which is the composite GPI. Although the actual values of each indicator are not available and we therefore cannot reproduce them, the weights for the GPI indicators can be retrieved from the GPI reports [20]. The score for each country’s composite GPI is continuous, normalized on a scale of 1 to 5, where the higher the score, the less peaceful a country is.

Figure 3.1 presents the official GPI around the world for 2008 (Figure 3.1a) and for 2020 (Figure 3.1b). Least peaceful countries are colored with darker red, whereas more peaceful countries are colored with lighter yellow. For example, from 2008 to 2020 Russia remains one of the least peaceful countries, whereas Canada remains one of the most peaceful. In addition, we observe that throughout the years peace around the world deteriorates, particularly in African and Middle East countries. For example, peace in Libya considerably deteriorates with the years.

For the purposes of this study, we increase the frequency of GPI from yearly to monthly data using linear interpolation. Every yearly GPI value is assigned to March of the



(a) Official GPI in 2008 for all countries around the world. The least peaceful countries are colored with dark red and the most peaceful countries are colored with light red.



(b) Official GPI in 2020 for all countries around the world. The least peaceful countries are colored with dark red and the most peaceful countries are colored with light red.

Fig. 3.1 Official GPI around the world. Official GPI in 2008 (a) and in 2020 (b) for all countries around the world. The least peaceful countries are colored with dark red and the most peaceful countries are colored with light red. We observe that throughout the years the peace deteriorates around the world.

corresponding year, since most of the annual GPI indicators are measured until this month. The linear upsampling is 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 months in total (March 2008 - March 2020).

The reason for increasing the frequency from yearly to monthly is that a month might contain some important events that are distorted from the yearly index. In other words, the yearly GPI data might not indicate abrupt peacefulness changes that happen at a higher frequency because these changes are usually smoothed out on the yearly GPI value. Therefore, monthly GPI estimations could reveal events neglected from the yearly GPI. At the same time, we do not increase the frequency at a weekly or daily level to keep a trade-off between the noisy GDELT information and the official GPI. Daily or weekly estimates could indicate fluctuations that would not significantly change the stability of the country for weeks or even months after taking place.

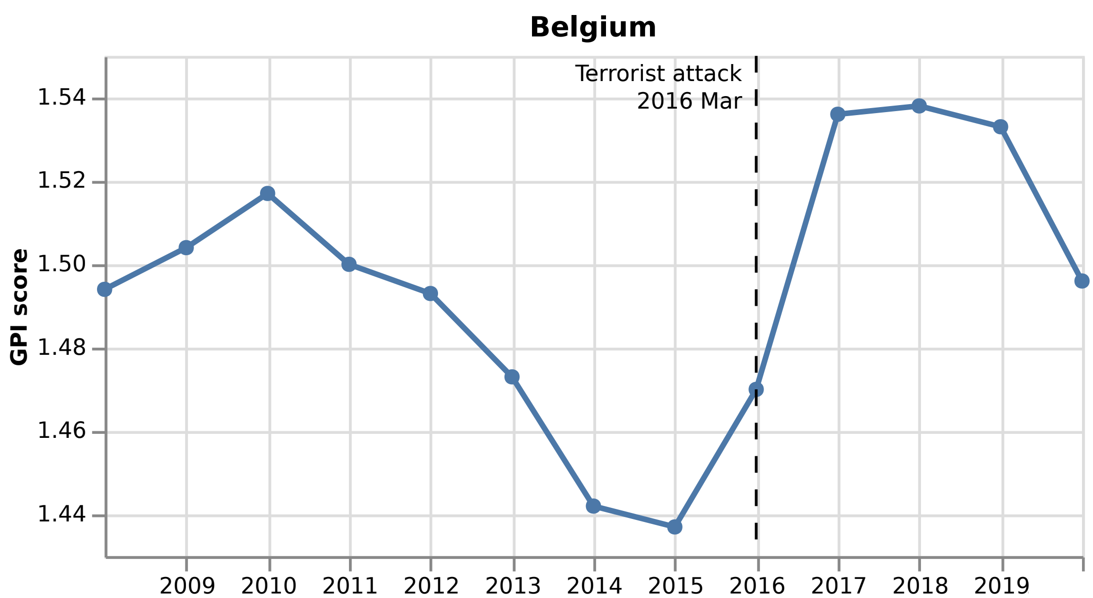


Fig. 3.2 Monthly Global Peace Index for Belgium from 2008 to 2020. In March 2016, the terrorist attack took place in Belgium, and as a result the GPI increases

To further explain the reason of our choice to increase the GPI frequency at a monthly level, we present Figure 3.2 and Figure 3.3. These plots show the monthly GPI for Belgium and Yemen, respectively, from 2008 to 2020.

In Figure 3.2, we annotate the terrorist attack that took place in Belgium in March 2016, which brought a deterioration in the peacefulness level of the country, increasing GPI from 1.47 to 1.536. However, this is depicted in the real yearly GPI only a year later, in 2017. On the contrary, when we introduce the monthly GPI score, we expect our model to depict the increase more timely, just one month after the attack.

In Figure 3.3, we annotate the start of the Civil War in Yemen in September 2014, which brings a deterioration in the peacefulness level of the country, increasing GPI from 2.735 to 2.84. Since the real GPI is only published once a year, it seems that the increase starts from March 2014, i.e., six months before the actual event. With the use of the monthly GPI score, we expect our model to capture this change in the GPI on time, just one month after the start of the Civil War.

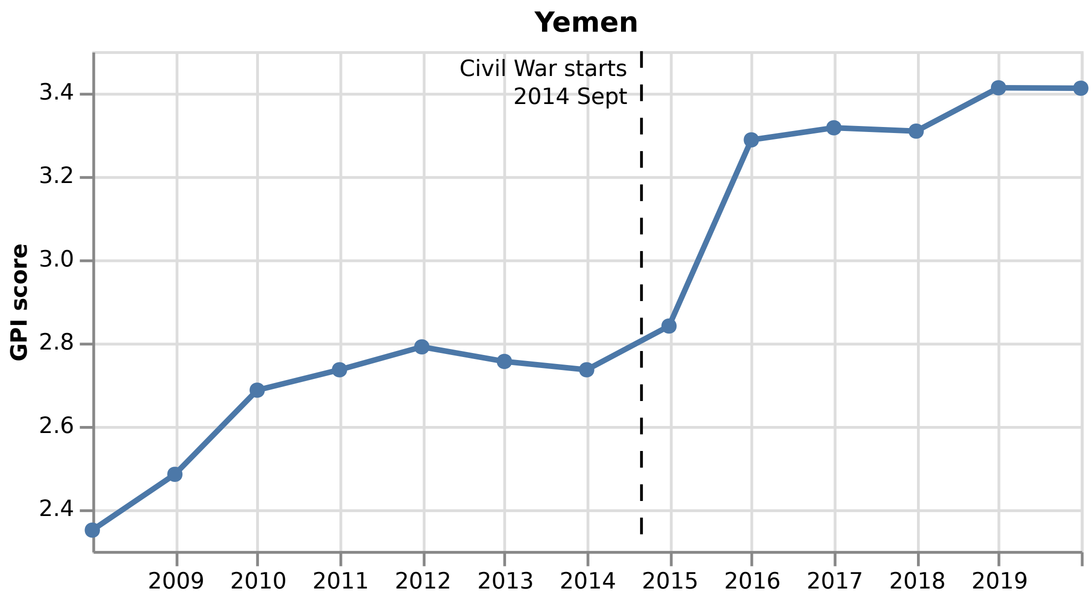


Fig. 3.3 Monthly Global Peace Index for Yemen from 2008 to 2020. In September 2014, the Civil War started in Yemen, and as a result the GPI increases.

In other words, these beforehand or delayed changes of the GPI are caused by the fact that peace fluctuations are depicted on the annual GPI, updated the following year. As a consequence, a monthly system that adequately corresponds to the peacefulness fluctuations has the potential to quickly inform the placement of peacekeepers and the deployment of non-governmental organization (NGO) resources, making it potentially easier to save lives and prevent devastation [157].

3.2.2 GDELT data

GDELT (Global Database of Events, Language, and Tone) [21] is a Google-supported and publicly available digital news database related to socio-political events. It is a collection of geopolitical event data extracted from international English-language news sources, such as Associated Press, The New York Times, Agence France Presse, Associated Press Online and Google News. Particularly, GDELT data records provide information on events from digital news articles identified with the Tabari system (Textual Analysis by Augmented Replacement Instructions system) [158]. The Tabari system is applied to digital news articles to extract all events contained in each article. The system uses pattern recognition to identify different events 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 [159]. Specifically, the system extracts, among others, the verbs that identify the action performed that determines the event code [160].

The subsequent are pieces of articles provided as examples to show how Tabari system identifies and codes events from article phrases:

- "...Senior Hungarian and Romanian officials agreed on Wednesday that their countries should cooperate to encourage Romanian refugees in Hungary to return home...". In this phrase the event identified is coded as "Express intent to cooperate" (030),
- "...Palestinians of the Israeli-occupied West Bank shunned work on Monday to protest at settlement of Soviet Jewish immigrants on Arab land...". In this phrase the event identified is coded as "Conduct strike or boycott" (143),
- "...One Serb policeman was murdered in an attack on a police patrol by Kosovo Albanians near the border with Kosovo, state agency Tanjug reported Sunday...". In this phrase the event identified is coded as "Use conventional military force" (190).
- "...Switzerland said today it had expelled two Soviet diplomats based in Geneva for spying, adding to a long series of espionage scares...". In this phrase the event identified is coded as "Reduce or break diplomatic relations (161)".

In total, 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 [159]. 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 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”.

The Tabari geocoding post-processing system is also enabled to georeference each event back to the specific country (geographic landmark) it is associated with [160]. In addition, the database offers various information for each coded event, such as the date and the URL of the news article the event is found in. In this thesis, we use GDELT 1.0, which is updated on a daily basis. Therefore data are available at a daily frequency. Historical data are also available since 1979 [160].

For example, in Figure 3.4, 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, from which the events are extracted. 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. Therefore, it is demonstrated that GDELT news can depict the worldwide sociopolitical and conflictual reality with a small lag, i.e., a day.

For the purposes of the current thesis, several variables are derived from GDELT database. These variables correspond to the total number of events (No. events) of each available GDELT category at country and monthly level. On average, the number of variables per country is 87, varying from 25 to 141. This indicates that some event

Fig. 3.4 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).

categories may not be present in the news of a country. To extract the GDELT variables the BigQuery [161] data manipulation language in the Google Cloud Platform is used. Listing 1 (3.1) presents the query used for the extraction of the variables.

Listing 3.1: 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<202010)
AND(ActionGeo_CountryCode<>'null')
GROUP BY ActionGeo_CountryCode,MonthYear,EventBaseCode
ORDER BY ActionGeo_CountryCode,MonthYear,EventBaseCode
```

Specifically, the used attributes in the listing represent:

- **ActionGeo_CountryCode** (string): the location of the event. It specifies the geographic resolution and returns the country,
- **MonthYear** (integer): the date of the event. It specifies the date and returns the month and the year,
- **EventBaseCode** (string): the CAMEO event code. It specifies the event and it returns the code in a three-level taxonomy. For an event related to "Appeal to yield" it returns the code "025".

Table 3.1 presents an indicative example of the GDELT data records for the United States in February and March 2018. For example, in February 2018, the No. events for category "Investigate crime" is 680, and in March 2018 it is 799. In February 2018, the No. events for category "Conduct non-military bombing" is 523, and in March 2018 it 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.

Table 3.1 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 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 non-military bombing	1099	2018/03
⋮	⋮	⋮	⋮

In addition, it is interesting to explore the share of each event category over all news. For example, Table 3.2 presents the 10 GDELT variables with the largest share of No. events for the United States from March 2008 to March 2020. We notice that, the GDELT variable "Make statement" has the largest share and is followed by "Make a visit" and "Host a visit" variables. Additional GDELT data description can be found in Appendix A.

Table 3.2 The ten GDELT variables with the largest share of the number of news for the United States over the whole dataset, i.e., from March 2008 to March 2020.

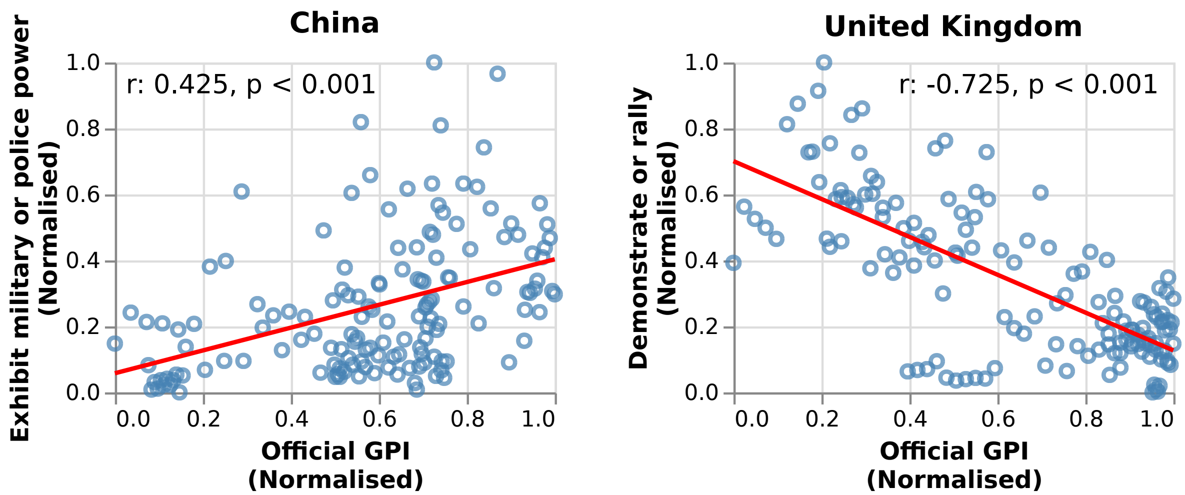
Event Code	Event Category	Share over all news
010	Make statement	7.73 %
042	Make a visit	7.52 %
043	Host a visit	6.97 %
020	Make an appeal or request	6.61 %
051	Praise or endorse	5.80 %
040	Consult	5.59 %
036	Express intent to meet or negotiate	4.50 %
173	Arrest, detain, or charge with legal action	4.08 %
190	Use conventional military force	3.72 %
046	Engage in negotiation	2.85 %

3.2.3 Matching GPI indicators with GDELT variables

Considering that GDELT is a database with socio-political events, we believe that the wide variety of its categories can cover most GPI indicators. Table 3.3 presents ten examples of GPI indicators matched with GDELT event categories. For example, the GPI indicator "Number of Internal Security Officers and Police per 100,000 People" could be covered by the GDELT variable "Exhibit military or police power". In addition, the GPI indicators "Ease of Access to Small Arms and Light Weapons" and "Volume of Transfers of Major Conventional Weapons, as recipient (imports) per 100,000 people" could be covered by "Fight with small arms and light weapons" and "Use conventional military force" or "Conduct non-military bombing" GDELT variables, respectively. Similarly, "Nuclear and Heavy Weapons Capabilities" GPI indicator could be covered by the "Employ aerial weapons" GDELT variable. Also, the GPI indicator "Likelihood of violent demonstrations" could be covered by "Engage in political dissent", "Protest violently, riot" or "Demonstrate or rally" GDELT variables. Last, "Financial Contribution to UN Peacekeeping Missions" GPI indicator could potentially be covered by the GDELT variables "Appeal for aid" or "Provide humanitarian aid".

We conduct correlation analysis to investigate whether there is a relationship between some of the aforementioned variables and the official GPI data. Figure 3.5 presents two scatterplots with the Pearson's correlation between official GPI data and GDELT variables. In particular, Figure 3.5a demonstrates that there is a moderate positive correlation (0.425) between the China GPI and GDELT variable "Exhibit military or police power". Figure 3.5b shows that there is a high negative correlation (-0.725) between the United

Kingdom GPI and the GDELT variable "Demonstrate or rally". GPI and the GDELT variables are normalised on a scale from 0 to 1 for visualisation.



(a) Moderate positive correlation (0.425) between the China GPI and GDELT variable "Exhibit military or police power".

(b) High negative correlation (-0.725) between the United Kingdom GPI and the GDELT variable "Demonstrate or rally".

Fig. 3.5 Correlation between official GPI and GDELT variables. Two examples of the correlation between official GPI data and GDELT variables for China and United Kingdom. GPI and the GDELT variables are normalised on a scale from 0 to 1 for visualisation.

The examples presented on Table 3.3 are intuitive matches. It is obvious that the model might return different but interpretable variables. Machine learning models are black boxes. Therefore, the relationship learned between the GDELT variables and the GPI might not be easily guessed in advance.

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 contribute to the GPI measurement.

Table 3.3 Ten examples of GPI indicators matched with GDELT event categories.

GPI indicators	**GDELT event category**
Number of Internal Security Officers and Police per 100,000 People	Exhibit military or police power
Ease of Access to Small Arms and Light Weapons	Fight with small arms and light weapons
Volume of Transfers of Major Conventional Weapons, as recipient (imports) per 100,000 people	Use conventional military force or Conduct non-military bombing
Nuclear and Heavy Weapons Capabilities	Employ aerial weapons
Likelihood of violent demonstrations	Engage in political dissent or Protest violently, riot or Demonstrate or rally
Financial Contribution to UN Peacekeeping Missions	Appeal for aid or Provide humanitarian aid
Relations with neighbouring countries	Reduce relations or Express intent to cooperate
Number of refugees and displaced persons as percentage of population	Grant asylum or Use as human shield
Level of violent crime	Investigate crime, corruption or Investigate war crimes

We use Linear Regression, Elastic Net, Decision Tree, Support Vector Regression, Random Forest, and Extreme Gradient Boosting, to investigate the relationship between the GPI score and the GDELT variables at a country level. Specifically, we aim to develop GPI estimates 1-month-ahead to 6-months-ahead of the latest ground-truth GPI value. Firstly, we introduce simple models, i.e., Linear Regression, Elastic Net and Decision Tree, which are easy to implement and interpret. Next, we apply SVR, Random Forest, and XGBoost models, which are superior in terms of accuracy but harder to interpret, and they need additional methodologies for the interpretation of the results (e.g., SHAP [24, 25]). Our main goal is to find the model with the highest performance.

Linear regression

Linear regression, one of the simplest and most widely used regression techniques, calculates the estimators of the regression coefficients (the predicted weights) by minimising the sum of squared residuals [162]. One of its main advantages is the ease of interpreting results.

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 [163]. 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 [163]. 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.

Support Vector Regression (SVR)

SVR [164] 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.

Random Forest

Random Forest limits the risk of a Decision Tree to overfit the training data [163]. 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 [165] 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.

For each of the models introduced, we estimate the best hyperparameters in each training phase through 10-fold cross-validation. The following paragraph presents the hyperparameters we tune for each model, except for Linear regression, which has no hyperparameters.

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.

3.4 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:

$$r = \frac{\sum_{t=1}^{n}(y_t - \bar{y})(x_t - \bar{x})}{\sqrt{\sum_{t=1}^{n}(y_t - \bar{y})^2}\sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2}} \ . \tag{3.1}$$

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:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - y_t)^2} \ . \tag{3.2}$$

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

$$MAPE = \left(\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - x_t}{y_t}\right|\right) \times 100 \ . \tag{3.3}$$

3.5 Explainable AI tools

Understanding a model's prediction is important for trust, actionability, accountability, debugging, and many other reasons. Indeed, one of the biggest challenges in adopting machine learning models, in particular advanced machine learning models, such as Random Forest or XGBoost, is their difficulty for interpretability. To understand predictions from tree-based machine learning models, such as the models mentioned above, 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.

However, it is crucial for researchers to be able to understand and explain the models' behaviour. Therefore, Lundberg et al. [24, 25] propose SHAP (SHapley Additive exPlanation). SHAP is based on game theory [166] and local explanations [167], 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. This provides local explanations with theoretical guarantees of local accuracy and consistency. Consistency in terms that if a model is changed so that it relies more on a particular variable, then the method must not attribute less importance to that variable. Accuracy in terms that the total contribution of each variable must sum up to the total contribution in the whole model. Lundberg and Lee also highlight that some commonly used variable importance approaches (including the Gain method that we use for the variable importance in Section 5.1) do not satisfy these properties. On the contrary, they propose SHAP as the only additive variable attribution method that satisfies these two properties based on results from game theory.

Additionally, 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. Combining many local explanations, a global structure can be represented while retaining local faithfulness [168] to the original model, which generates detailed and accurate representations of model behavior. In particular, SHAP can provide a breakdown of the key drivers for one particular record in the data (i.e. a local explanation). The SHAP values for each variable represent their contribution towards a higher or lower final prediction.

Last but not least, SHAP methodology can be applied for the interpretation of the results of the machine learning models, since it identifies the relationship between the independent variables, either internal or external, and the dependent variable. The relationship between the independent variables and the dependent variable does not need to be causal, as SHAP can fail to accurately answer causal questions [26].

In this study, SHAP serves as a tool to identify which external GDELT variables drive the GPI estimations. This can be useful for explaining peace and its determinants, explaining the models' behavior overall and for every prediction, and for diagnosing errors in the predictions. In Section 5.2 we compute the variable importance through the SHAP and we conduct in-depth analysis for four country models.

Chapter 4

Measuring peace through the world news

In this chapter, we mainly focus on the RQ1, i.e., the measurement of the GPI with the use of GDELT data at a monthly frequency compared to the official yearly index score. To tackle this task, we design the methodology described in Figure 4.1. As explained in Chapter 3 we extract news media attention variables from GDELT at a monthly and country level. We also upsample the official GPI data from a yearly level to a monthly level. Therefore, we construct every country model, using the corresponding country data as input, i.e., the upsampled GPI, as ground-truth data (dependent variable), and the GDELT data, as exogenous (independent) variables. Then, we set the estimation framework for the training of our models. Last, we measure and predict the monthly GPI values per country and then we validate the models.

4.1 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 [169, 170].

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, widely used in business and finance [171, 172]. 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.

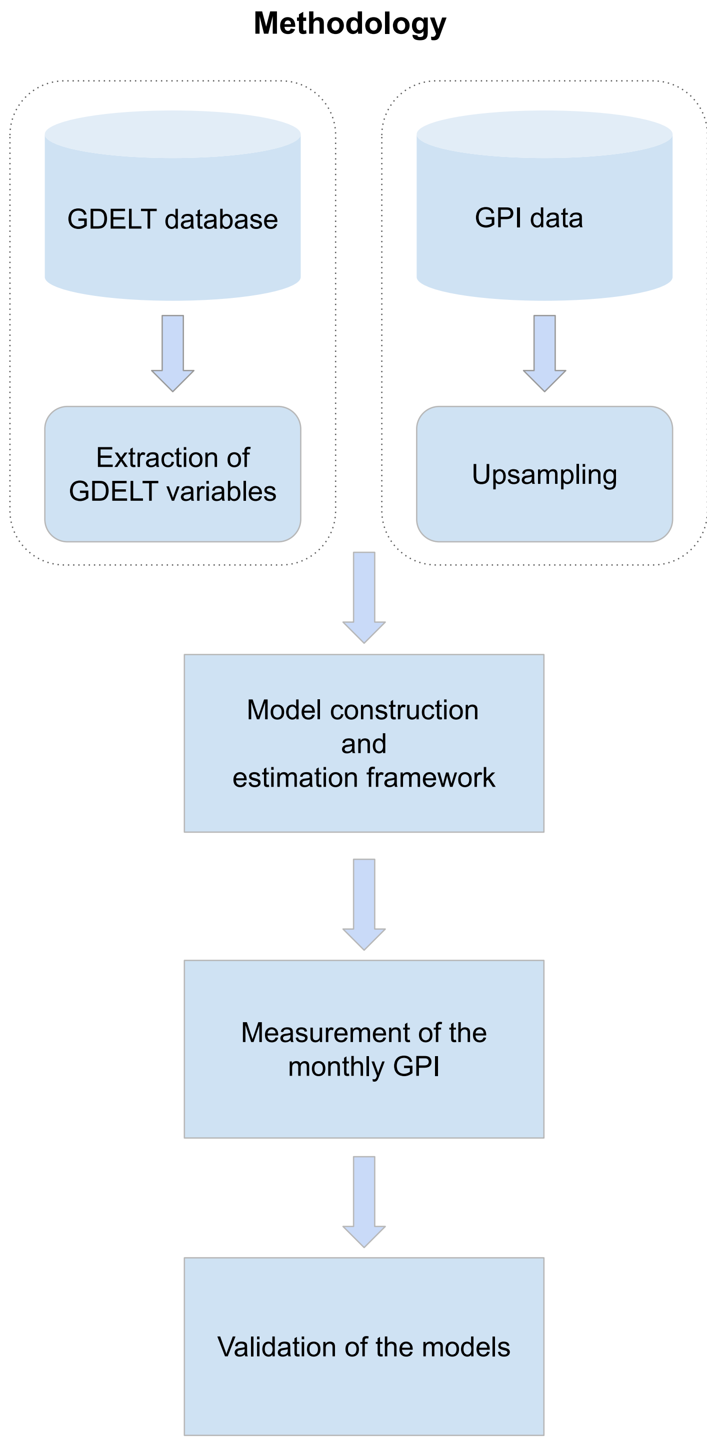


Fig. 4.1 Main methodological approach. For the construction of every country model, the upsampled GPI is used as ground-truth data (dependent variable), and the GDELT news media attention data are used as exogenous (independent) variables. Then, the models are trained, the monthly GPI values are estimated, and the models are validated.

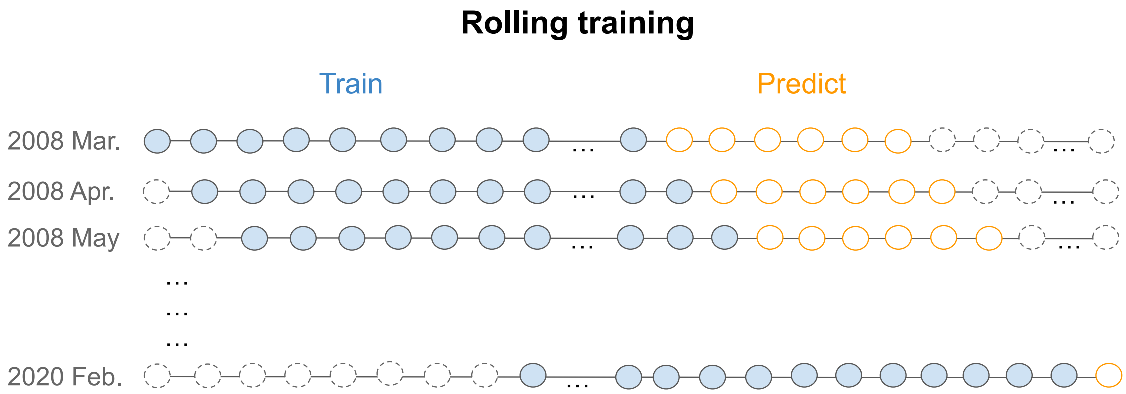


Fig. 4.2 Rolling training: Data from March 2008 to February 2014 are used to train the model and predict 6-months-ahead GPI values, data from April 2008 to March 2014 are used to train the model and predict the 6-months-ahead GPI values of April 2014 up to September 2014, and so on, till the last training, which includes data from March 2014 to February 2020 to make only 1-month-ahead GPI prediction.

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. Next, we drop one month from the beginning of the training set and add another month to the end of the training set. We then 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.

Figure 4.2 presents a visualisation of the rolling training. 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 $k = 1, 2, \dots, 6$ 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. As mentioned above, our estimation

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

framework obtains from 1 up 6-months-ahead GPI values. Long-term predictions, such as 6-months-ahead peacefulness estimations, are an important tool for policy-makers, since it is a "policy-relevant lead time" consistent with other forecasting work; that is, a period of time sufficiently long that there could be a policy response [173].

For each of the models mentioned in Section 3.3, we estimate the best hyperparameters in each training phase. Section 3.3 includes all the details for the hyperparameters we tune for each model.

4.2 Prediction results and validation

The five prediction models, Elastic Net, Decision Tree, SVR, Random Forest, and XG-Boost, (see Section 3.3), are constructed for every country to produce the GPI 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) [162, 174, 175, 100] (see more details in Section 3.4).

The analysis is conducted for all 163 countries that have a GPI score. As discussed in Section 4.1, our estimation framework is not limited to the 1-month-ahead predictions, but it generates GPI estimates up to 6-months-ahead. Figure 4.3 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. Similarly, Figure 4.4 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 4.3a-b, with Figures 4.3c-d, and with Figures 4.3e-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

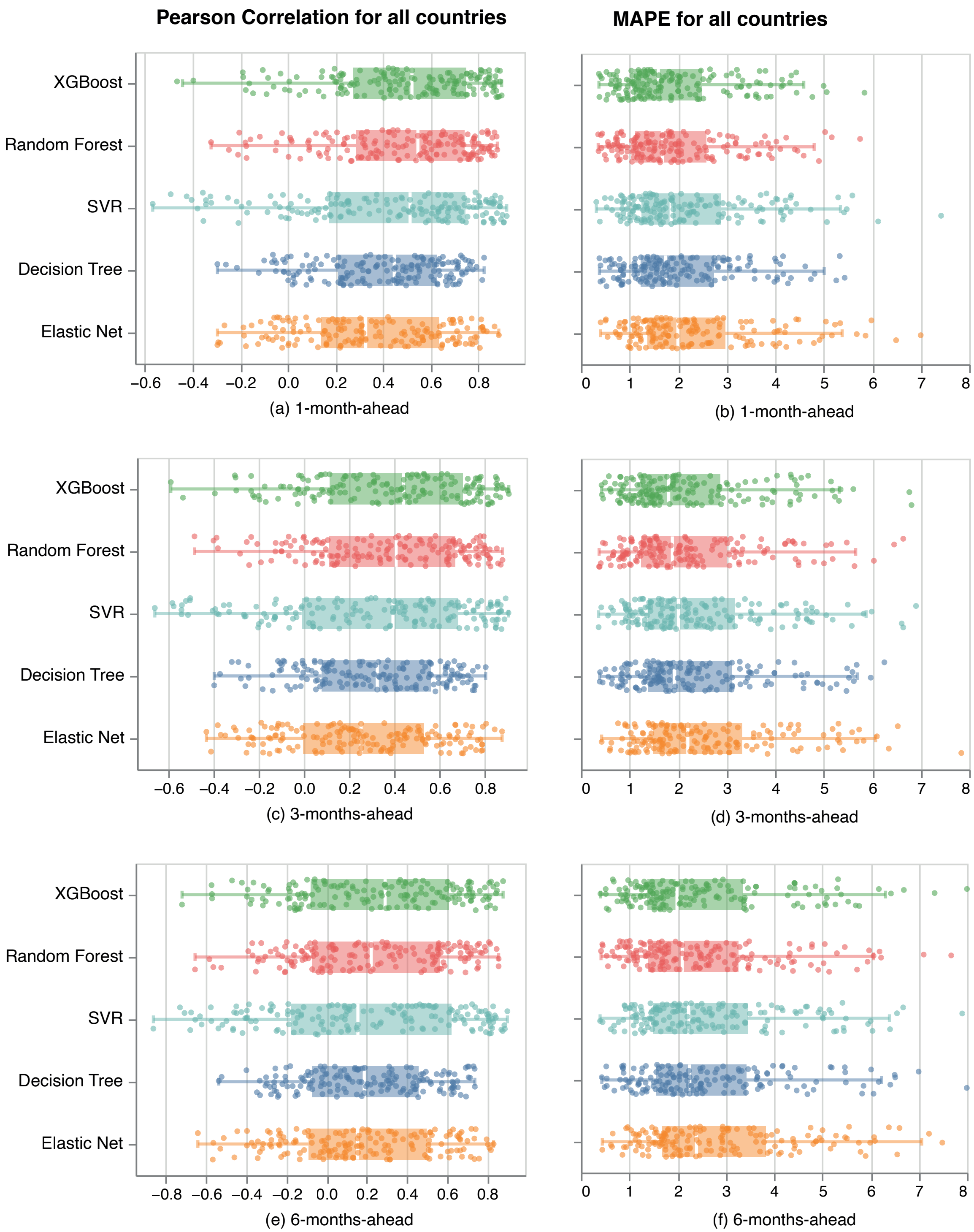


Fig. 4.3 Pearson Correlation and MAPE for all country models. Pearson Correlation and MAPE between the real and the predicted 1-, 3-, and 6-months-ahead GPI values at a country level, for all prediction models. The boxplots represent the distribution of the aforementioned performance indicators for all country models. The plots' data points correspond to each country model. Overall, XGBoost models outperform the rest of the four models.

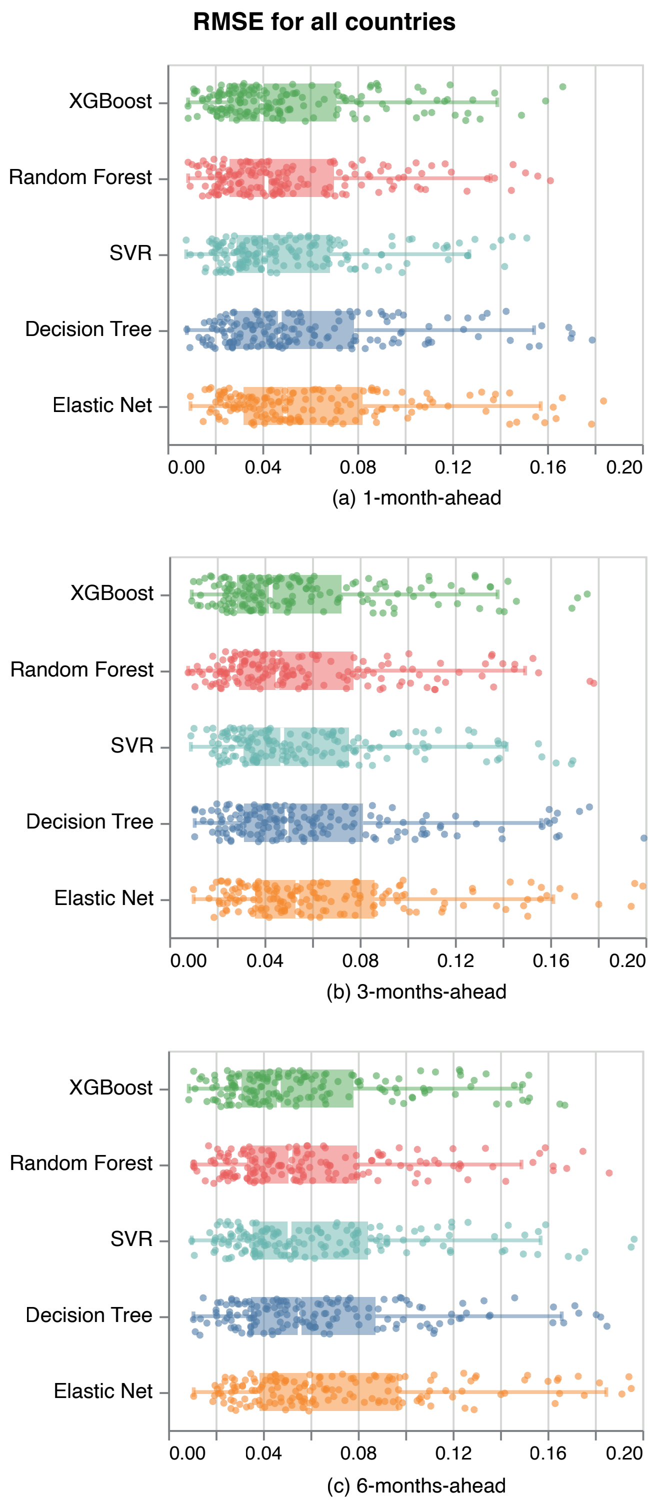


Fig. 4.4 RMSE for all country models. RMSE between the real and the predicted 1-, 3-, and 6-months-ahead GPI values at a country level, for all prediction models. The boxplots represent the distribution of the aforementioned performance indicators for all country models. The plots' data points correspond to each country model. Overall, XGBoost models outperform the rest of the four models.

to the 1-month-ahead predictions.

It is noticeable from Figure 4.3 that some models show high performance and others show low performance. For example, the XGBoost 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 XGBoost 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.

As illustrated above, our results show that XGBoost achieves the highest performance compared to the other models, in particular when compared to Elastic net and Decision tree. At this point, it is important to control the stability of the behaviour of the countries for every algorithm. Figure 4.5 presents five scatter plots with the countries' performance for every algorithm applied, i.e., Elastic net (Figure 4.5a), Decision tree (Figure 4.5b), SVR (Figure 4.5c), Random Forest (Figure 4.5d), and XGBoost (Figure 4.5e). The countries with the highest and the lowest performance are the same for all algorithms. For example, the United States (USA) and the United Kingdom (GBR) have high performance for all algorithms. On the contrary, Moldavia (MDA), Cyprus (CYP), Uruguay (URY), Estonia (EST), and Gabia (GAB) have low performance for all algorithms. In addition, the plots illustrate that most countries, such as Pakistan (PAK), Turkey (TUR), Uganda (UGA), and Sweden (SWE), have high performance for the most effective algorithms, i.e., for SVR, Random Forest, and XGBoost, and improved performance as compared to Elastic net and Decision tree.

Therefore, considering that XGBoost provides the best results on average across all countries, and after controlling the stability of the countries, we focus our analysis on XGBoost models results. Particularly, we use XGBoost for the deeper analysis that follows in the next chapters.

4.3 Country models' performance

To further study and understand the models' behaviour, we split the countries into three categories based on their performance for the 1-month-ahead predictions. We consider high performance models those with Pearson Correlation >= 0.7 and MAPE < 5 (48 countries in total), low performance models those with Pearson Correlation <= 0.2 (26 countries), and we consider the rest of the models as medium performance models (89 countries) [176, 177, 177]. Figure 4.6 presents the countries with high, medium, and

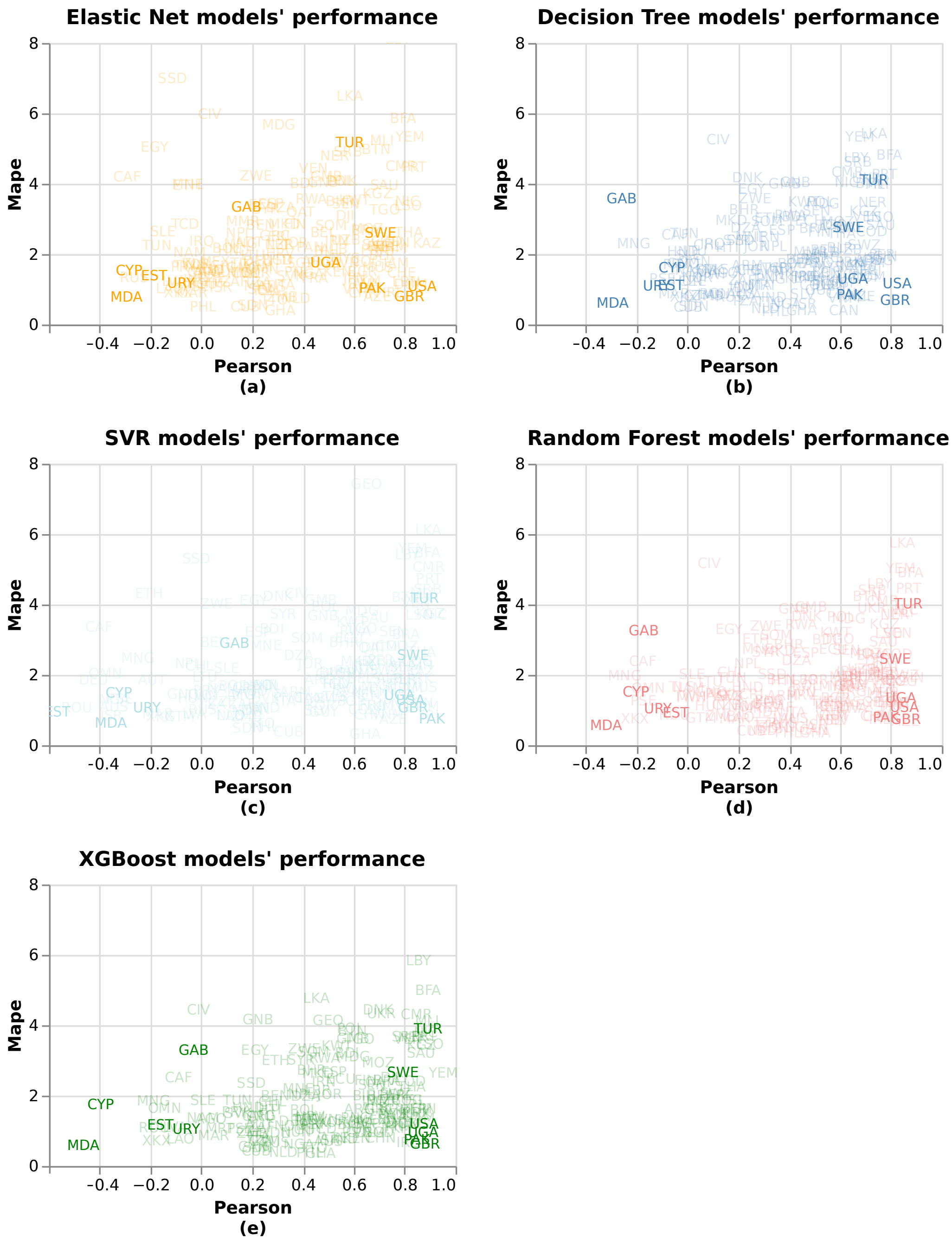


Fig. 4.5 Stability of countries behaviour. The countries that demonstrate the highest and the lowest performance, in this example for the 1-month-ahead predictions, are the same for all algorithms. For example, the United States (USA) demonstrates high performance for all algorithms, whereas Moldavia (MDA) demonstrates low performance for all algorithms. Most countries countries, such as Pakistan (PAK) show improved performance for the most effective algorithms, i.e., for SVR, Random Forest, and XGBoost, compared to Elastic net and Decision tree.

low performance for the 1-month-ahead predictions. For example, Uganda (UGA), Pakistan (PAK), Turkey (TUR), the United Kingdom (GBR), and Sweden (SWE) are high-performance models, with a Pearson Correlation > 0.8 (very high). We also observe medium performance countries, such as Libya (LYB) with high Pearson Correlation and high MAPE, and India (IND) with low Pearson Correlation and low MAPE. Besides, there are country models, such as Cyprus (CYP), Estonia (EST), Moldova (MDA), Mongolia (MNG), and Romania (ROU), with a negative Pearson Correlation. The reasons for the low model's performance could be various, such as the GDELT news coverage or the under- or over-representation of some countries through the GDELT news. We further explore the reasons for the countries' low performance in Section 4.3.2.

4.3.1 High performance models

In this section we present high performance countries' results. We choose countries with different military, socio-economic, and political history and current situation, to cover a variety of scenarios. In particular, 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., [178]). Additionally, we use various sources, such as the official GPI ranking [20], to choose three of the most peaceful countries (Portugal, Iceland, and New Zealand) and three of the most war-torn countries (DR Congo, Pakistan, and Yemen).

Table 4.1 reports the models' performance for the 1-month-ahead up to 6-months-ahead GPI estimates for the nine selected countries. Overall, 1-month-ahead GPI estimates are more accurate than the other estimates, especially with respect to the 6-months-ahead time horizon estimates. There are countries, such as Portugal, for which the performance remains stable over all 6 months predictions, and countries like Yemen for which the performance falls for each additional in future prediction. An explanation to these different behaviors could be, for example in the case of Portugal, that the military, socio-economic, and political situation remains stable over time, and therefore the most important variables contribute to a more accurate prediction even further in the future. On the contrary, in war-torn countries like Yemen, the country's situation changes constantly and the variables are not much relevant anymore. For this reason, for Yemen we also conduct a training with the 36 most recent monthly values (Yemen * in Table 4.1), 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

Fig. 4.6 High, medium, and low performance country models. High, medium, and low performance country models for the 1-month-ahead predictions. There are country models that show high performance, such as the United Kingdom (GBR), models that show medium performance, such as Libya (LBY), and models that show low performance, such as Mongolia (MNG).

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 DR Congo.

Table 4.1 Performance indicators with respect to GPI ground-truth of nine high performance country models. Overall, 1-month-ahead GPI estimates are significantly more accurate compared to the rest future estimates, especially to the 6-months-ahead time horizon. For the training of Yemen *, 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.

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
Pakistan	Pearson	0.848	0.772	0.720	0.668	0.672	0.640	0.720
	MAPE(%)	0.749	0.858	0.922	1.006	1.052	1.036	0.937
	RMSE	0.029	0.033	0.036	0.040	0.040	0.040	0.036
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

Additionally, Figure 4.7 presents four scatter plots of the real and the estimated GPI values for the 1-month-ahead predictions. In particular, Figure 4.7a presents the scatter plot of the real and predicted GPI values of all the countries regardless of their performance, highlighting three high performance models, i.e., Iceland (green data points), Saudi Arabia (blue data points), and Pakistan (purple data points). The highlighted countries illustrate their superior performance as compared to other countries. Figure

4.7b-d present the corresponding scatter plots of the real and predicted GPI values of Iceland, Saudi Arabia, and Pakistan. The examples of the countries indicate that the models show high performance for either low, medium, or high GPI values.

4.3.2 Medium and low performance models

There are countries whose models have medium performance (Figure 4.6), such as Colombia and Chile (Pearson Correlation = 0.63 and MAPE = 0.96, and Pearson Correlation = 0.28 and MAPE = 1.83, respectively, for the 1-month-ahead predictions). To get insights about the reasons behind the medium performance, we further study these country models.

Colombia ranks in the 11th place out of 163 countries on the list presenting the economic cost of violence ranked by percentage of GDP. Particularly, its economic cost of violence is 169,517.8 (in million 2019 PPP U.S. dollars) [20]. Thus, in line with the study's purposes, it would be important to at least understand and explain the reasons that drive the model to have a medium performance. Figure 4.8 presents Colombia model predictions, with respect to the real GPI score. Colombia has been pursuing peace since 1964. We have therefore selected a sample of important events to show how well our model is capturing peacefulness fluctuations and why predictions may vary compared to the real GPI score.

In January 2015, President Santos said the government was ready for a bilateral ceasefire with Farc, after welcoming Farc's December unilateral ceasefire. The estimated GPI captures the decrease of GPI, as opposed to the real GPI that continues increasing. In March 2016, the government and Farc delayed the signing of a final agreement. In this case, the estimated GPI adequately captures the GPI increase compared to GPI that decreases. Similarly, in September 2016, the government and Farc signed a historic peace accord. The estimated GPI is correctly decreased this month, as compared to the real GPI that continues increasing. Last, in August 2019, the Farc rebel group commander defied 2016 peace agreement and calls on supporters to take up arms again. The GPI score should increase and Colombia model adequately on time captures this peace fluctuation as compared to the real GPI that continues decreasing. The reason that the real GPI score does not depict these peacefulness changes is because it is a monthly index upsampled from a yearly index. Therefore, some small changes are smoothed out on the real index or if important ones are depicted later on the next year (see Section 3.2.1 for further details on the upsampled GPI).

In addition to Colombia, we choose to further analyse Chile to get a better under-

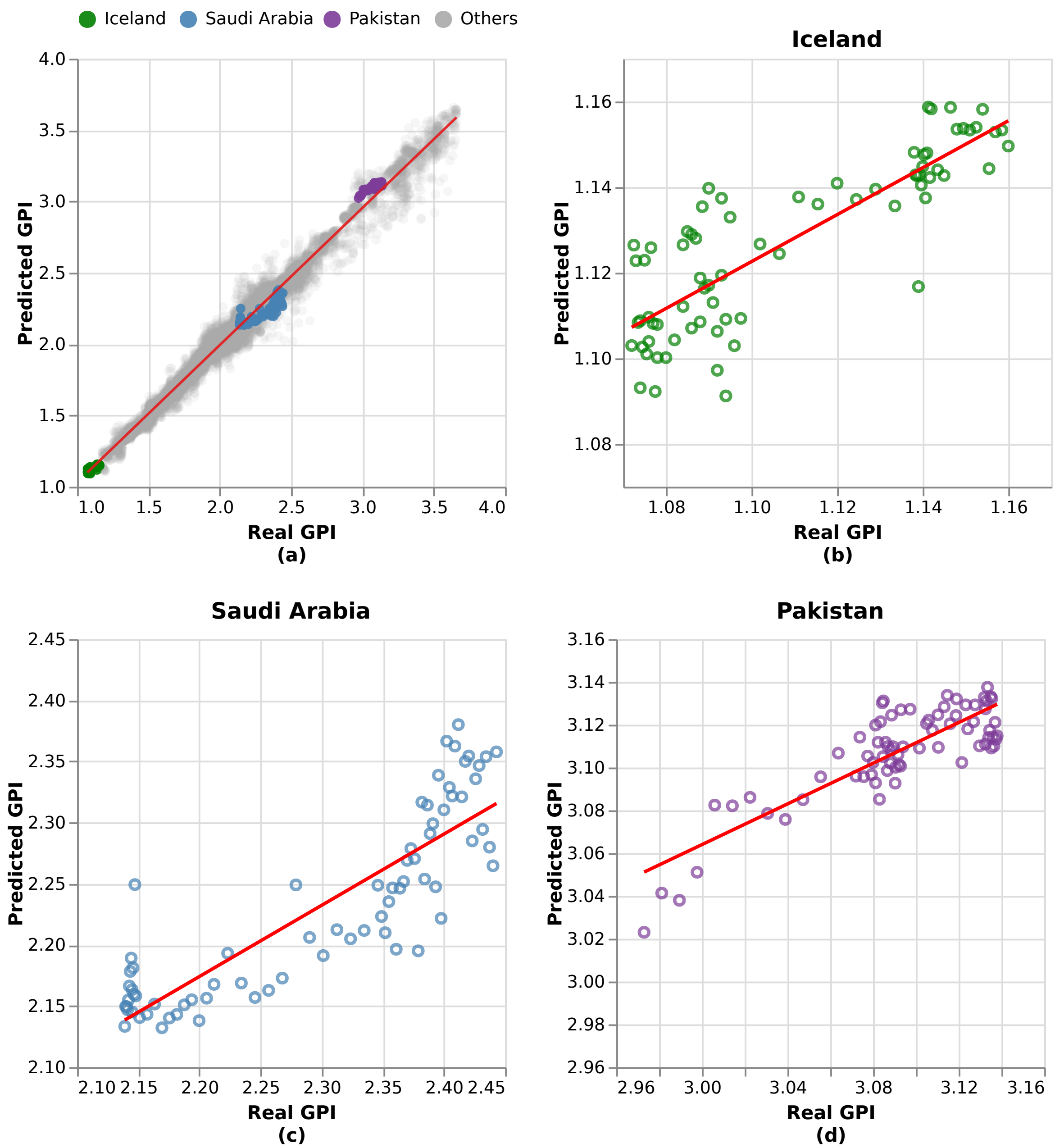


Fig. 4.7 Scatter plots of the real and estimated GPI values. (a) Scatter plots of the real and estimated GPI values for all country models, (b) Real versus estimated GPI values for Iceland (b), Saudi Arabia (c), and Pakistan (d).

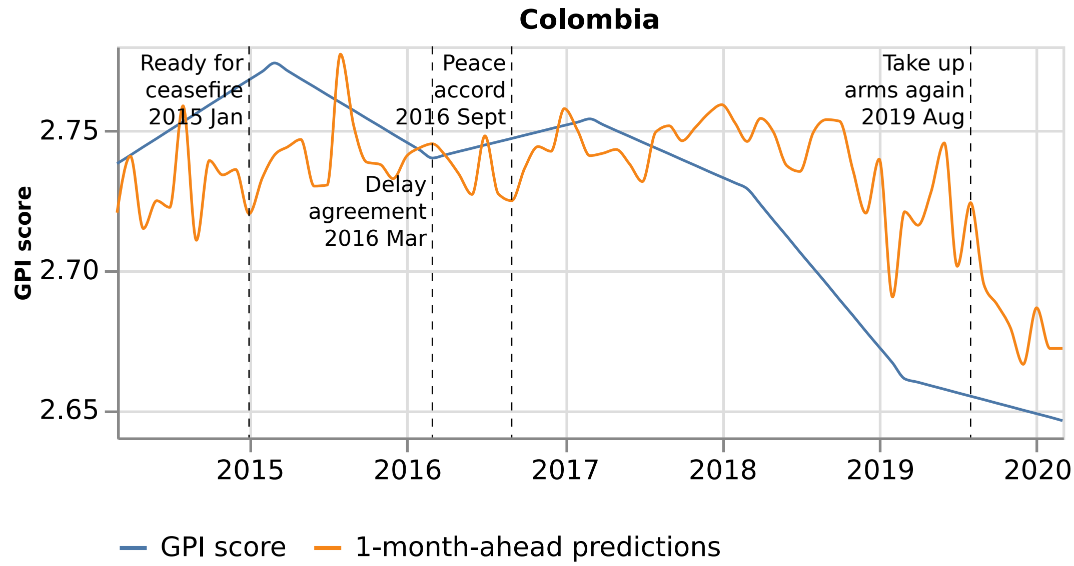


Fig. 4.8 Colombia predictions, with respect to the real GPI score. Colombia 1-month-ahead predictions (blue curve), with respect to the real GPI score (orange curve). The estimated GPI score adequately captures the changes in peace in January 2015, March 2016, September 2016, and August 2019, as compared to the real GPI score.

standing of its medium performance. Based on the 2020 GPI report [20], Chile has its lowest levels of peacefulness since the inception of the GPI. Figure 4.9 compares Chile's model predictions with the real GPI score, showing that the predictions curve follows the real GPI curve till March 2019. In March 2019, we observe the real GPI score increasing abruptly till March 2020, and the predictions curve does not follow the real GPI score till October 2019. In October 2019, Chile was rocked by mass protests at economic inequality, prompted by a subsequently-reversed rise in Santiago metro fares. The estimated GPI score, in contrast with the real one, captures this score increase in adequate time. The reason that the real GPI score anticipates this increase may be the fact that the GPI score is yearly and upsampled to a monthly index. Therefore it depicts the abrupt peacefulness turbulence already from March 2019.

We also deepen the analysis to find out the reasons that drive some country models to show low performance. To control to what extent these countries are covered from the GDELT news, we investigate if there is any correlation between each country's average yearly news and model's performance. Figure 4.10a demonstrates that the Pearson correlation between the two is very low. In addition, we investigate if there is any correlation between each country's number of the monthly news and model's monthly performance. Figures 4.10b presents Portugal, a high performance model, and 4.10c and 4.10d present two low performance models Zambia, and Moldova, respectively. All figures demonstrate

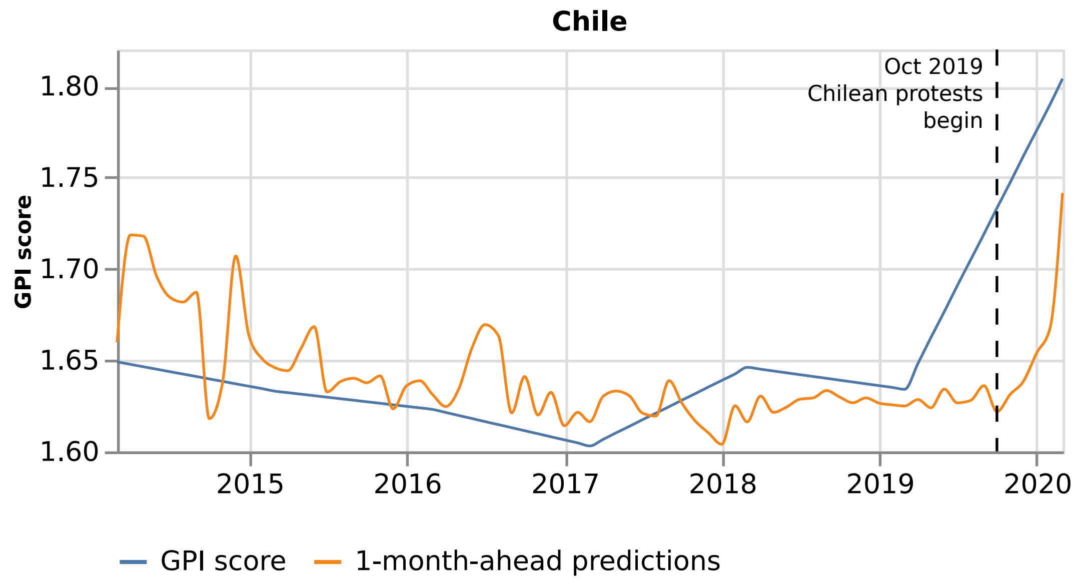
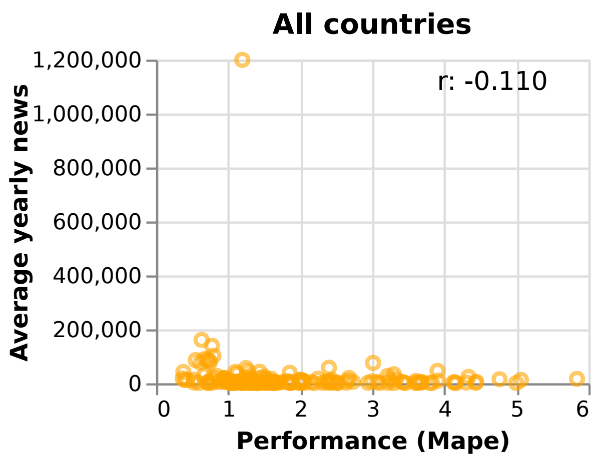
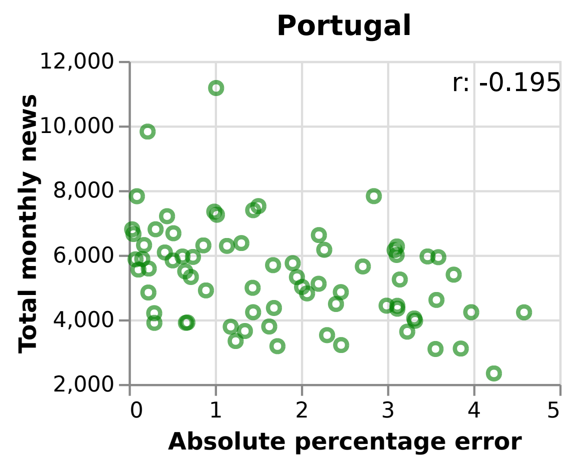


Fig. 4.9 Chile predictions, with respect to the real GPI score. Chile 1-month-ahead predictions (blue curve), with respect to the real GPI score (orange curve). The estimated GPI score adequately captures the disturbance in peace in October 2019, that the Chilean protests begun, as compared to the real GPI score.

that the Pearson correlation between each country's monthly number of news and model's monthly performance is very low.

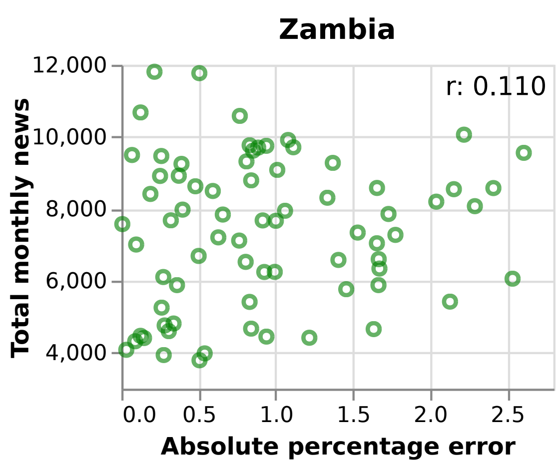
Another possible explanation for some countries low performance, which could be further explored, is that some countries might be under-represented through the GDELT news or even over-represented [23]. For example, a lot of United States news media, which is the strongest player in the media industry, are tracked by GDELT. The United States news in the English language might not cover well events happening in foreign countries or non-English speaking countries. Additionally, news media could introduce further biases in the study. First, they sometimes misrepresent reality. For example, they give a distorted version of the crimes within a city with a significant bias towards violence [179]. Second, news media datasets contain the gatekeeping bias, i.e., the journalists decide on which event to publish, the coverage bias, related to the over-coverage or under-coverage of an event, and the statement bias, i.e., when the content of an article might be favorable or unfavorable towards certain events [135].

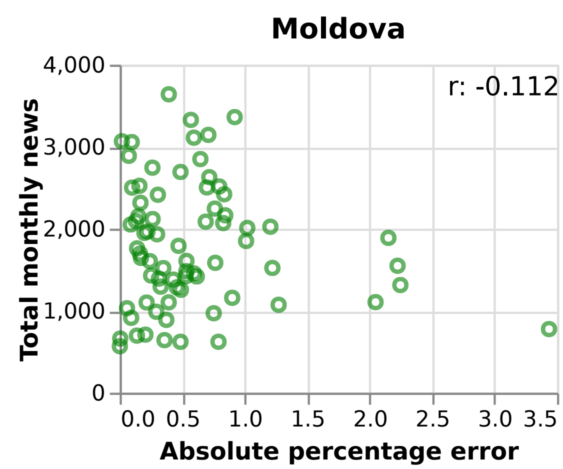




(a) Low Pearson correlation (-0.110) between each country's average yearly news and the models' overall performance.

(b) Low Pearson correlation (-0.195) between Portugal's monthly news and the monthly absolute percentage error.





(c) Low Pearson correlation (0.110) between Zambia's monthly news and the monthly absolute percentage error.

(d) Low Pearson correlation (-0.112) between the Moldova's monthly news and the monthly absolute percentage error.

Fig. 4.10 GDELT news coverage and countries' performance. The average yearly news and all the model's performance are not correlated. Similarly, the total monthly news and the monthly absolute percentage error are not correlated for neither high performance models, such as Portugal, nor for low performance models, such as Zambia and Moldova.

Chapter 5

Understanding peace

In this chapter, we mainly focus on the RQ2, i.e., to explain peace and its determinants. To tackle this task, we conduct variable importance analysis, and we apply explainable AI methodologies for high performance models.

5.1 Variable importance via Gain

The most important variables that the model uses for the predictions can help interpret the results of the models. There are multiple ways of analysing the variable importance. XG-Boost performs variable selection and provides variable importance through the method called Gain, i.e., the average improvement in model fit each time a variable is used in the trees. A variable's higher value of this metric when compared to another variable's value implies it is more important for generating a prediction. At every training phase, a Gain value is assigned to each variable, capturing its contribution to the accuracy of the prediction.

Considering that we apply a rolling methodology for the training of the models (Section 4.1), we obtain 72 different Gain values for each variable in each of the 72 training phases. Therefore, to find out the average importance for each variable over all training processes, we calculate the arithmetic mean of each variable's Gain values. We define this arithmetic mean as *Importance* and identify the most important variables per country, on average, for all GPI predictions. We apply the same process to estimate the *Importance* of each variable for the rest of the models as well.

In Figures 5.1, 5.2, 5.3, we present the top 10 variables of the United States (powerful country), Portugal (peaceful country), and Pakistan (war-torn country), respectively. For the United States, the most important variables are related to military engagements, weapons, ethnic cleansing, assassinations, as well as cooperations, forgiveness, relations, and agreements (Figure 5.1). These variables illustrate a profile of a strong player in the military economic and political foreground.

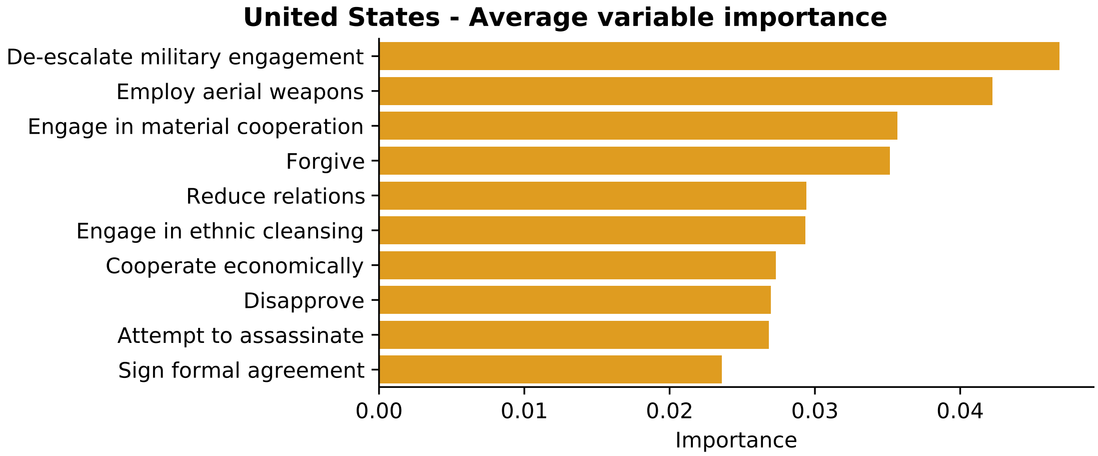


Fig. 5.1 Average variable importance for the United States. Average variable importance for the United States, a powerful country, calculated through the Gain method. The variables are related to military engagements, weapons, ethnic cleansing, assassinations, as well as cooperations, forgiveness, relations, and agreements.

For Portugal, the most important variables are related to visits, administrative sanctions, symbolic, cooperations, and aids, as well as criticisms or denouncements, arrests, detains or charges with legal action, and fights with small arms or light weapons (Figure 5.2). Apart from the latter, the rest of the variables reveal a peaceful country profile for Portugal. On the other hand, Figure 5.3 presents the most important variables for Pakistan, illustrating a war-torn country profile. Specifically, Pakistan's variables are related to the use of conventional military force, fights with artillery and tanks, reductions of aids, lawsuits, rejection of requests or demands for political reforms, denial of responsibility, strike or boycott, as well as consults, and diplomatic recognition.

Overall, we observe that variable importance analysis confirms the categorization of the countries presented in Table 4.1 into powerful, peaceful, and war-torn since it reveals the profile of each country.

In addition, we select the three most important variables (top3) of each high-performance country and we control their frequency in the top3 of all high performance countries. Figure 5.4 presents the frequency of each variable in the top3 most important variables overall countries. We observe that the most frequent most important variable is "Appeal" , which is selected from 11 country models. "Criticize or denounce" is the second most frequent and "Make a visit" is the third most frequent most important variable, which are selected for 10 and 9 country models, respectively.

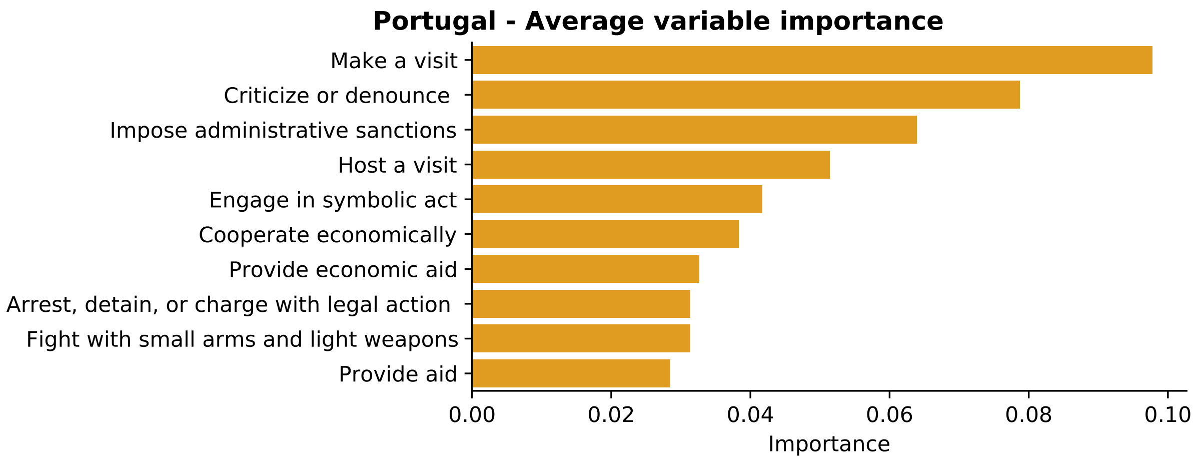


Fig. 5.2 Average variable importance for Portugal. Average variable importance for Portugal, a peaceful country, calculated through the Gain method. The variables are related to visits, administrative sanctions, symbolic, cooperations, and aids, as well as criticisms or denouncements, arrests, detains or charges with legal action, and fights with small arms or light weapons.

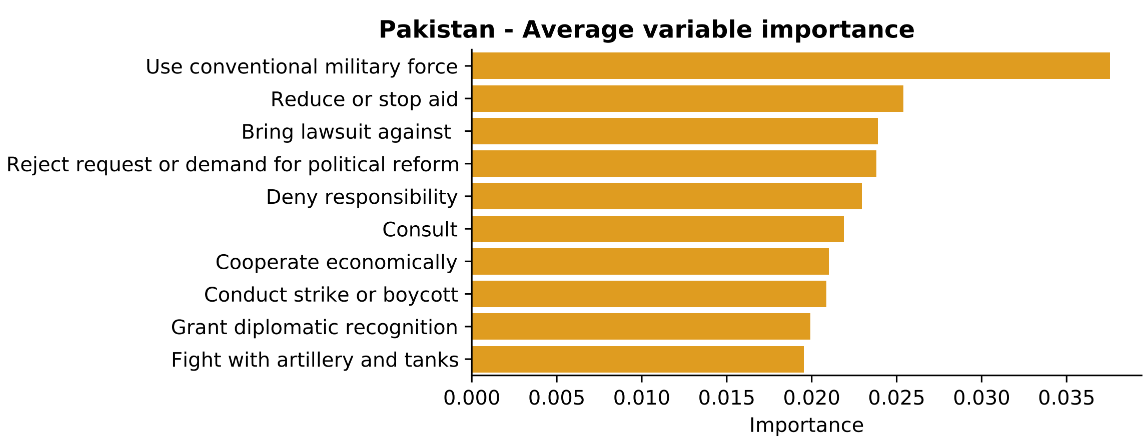


Fig. 5.3 Average variable importance for Pakistan. Average variable importance for Pakistan, a war-torn country, calculated through the Gain method. The variables are related to the use of conventional military force, fights with artillery and tanks, reductions of aids, lawsuits, rejection of requests or demands for political reforms, denial of responsibility, strike or boycott, as well as consults, and diplomatic recognition.

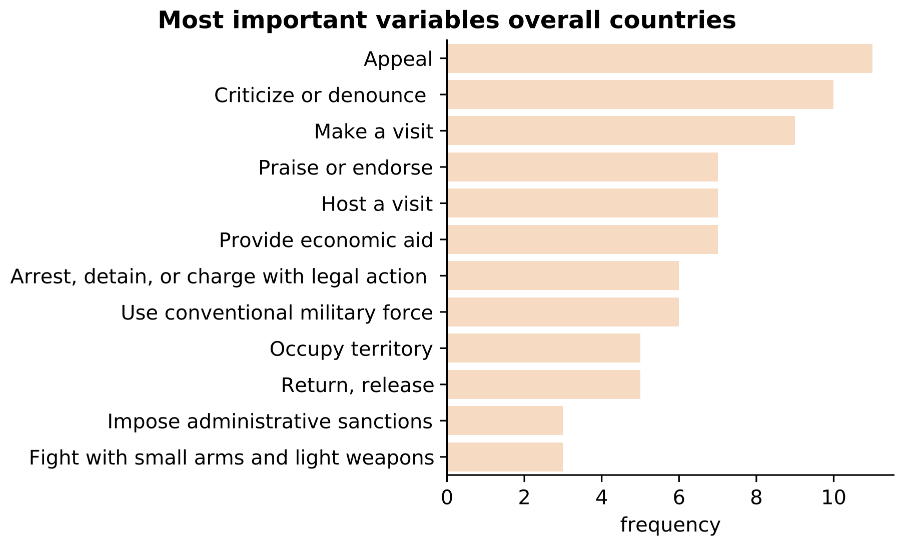


Fig. 5.4 Comparison of countries' variable importance. The frequency of the variables in the three most important variables overall countries. We observe that the most frequent most important variable is "Appeal" , which appears in the top3 for 11 countries.

Furthermore, Figure 5.5 presents the most important variables overall countries, as shown in Figure 5.4, and examples of countries which include or do not include each variable in their top3 most important variables. We notice that Iceland (ISL) and Portugal (PRT) have the same top3 most important variables, i.e., "Appeal", "Criticise or denounce", and "Make a visit". This could be explained by the fact that, on average, Iceland and Portugal are between the most peaceful countries in the world, and they both have a low GPI score (they are represented by the blue square points). Additionally, we observe that New Zealand (NZL) and Peru (PER) have two common variables in the top3 variable list, i.e., "Appeal" and "Criticise or denounce". New Zealand is more peaceful than Peru since the average GPI value of the former is lower than the latter (they are represented by blue and orange square points , respectively). This demonstrates that the same variable for different country profiles might be related to different in nature events. For example, the variable "Appeal" for the New Zealand might be related to appeal for diplomatic cooperations or for intelligence cooperations. The same variable for Peru could be related to appeals for military cooperation or aid since very often conflictual events take place in the country. Last, it is interesting to notice that there are variables, such as "Appeal" that are mostly found in the top3 of medium GPI countries (orange squared points), whereas there are variables such as "Criticise or denounce" and "Impose administrative sanctions" that are mostly found in the top3 of low GPI countries (blues squared point) or variables, such as “Use unconventional military force” that are found in the top3 of medium and high GPI countries (orange and red squared points, respectively). Besides, there are variables, such as “Provide economic aid” which are found in the top3 of either low, medium and high GPI countries (blue, orange and red squared points, respectively).

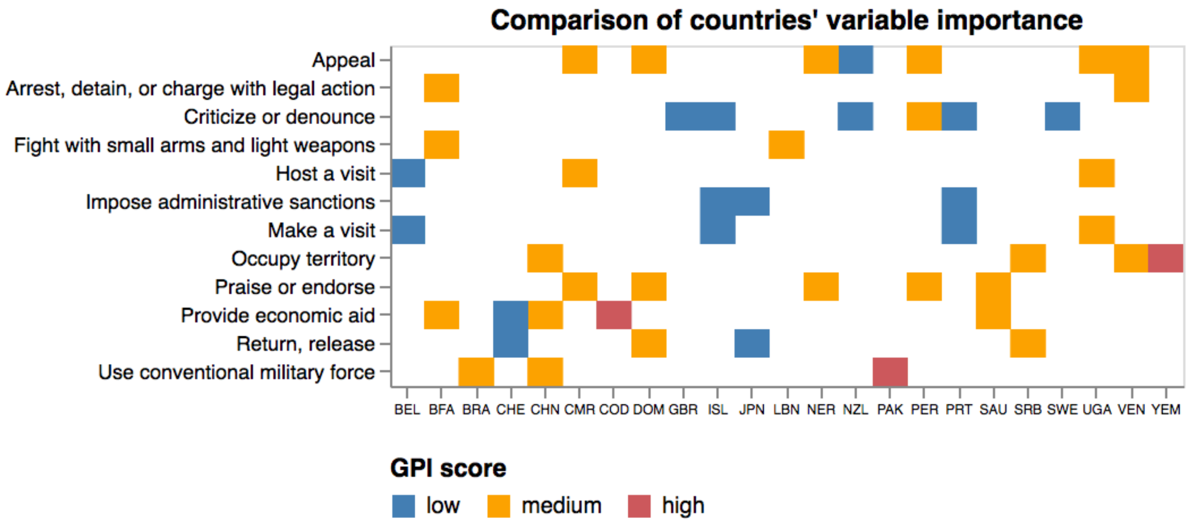


Fig. 5.5 Most important variables overall countries. The most important variables overall countries, as well as the countries which include or do not include the variables in their top3 variable importance list. The color of the squared points represents the level of the GPI score.

5.2 Variable importance via SHAP

As explained above there are various methods for variable importance analysis. In Section 5.1 we present the Gain approach, which calculates the variable importance through the Gain method. In this section we present the variable importance calculated through the SHAP method (see Section 3.5). In particular, we demonstrate how SHAP method can significantly contribute not only to the better interpretability of the model results, but also to the explanation of the models’ behaviour and of the large estimation errors produced. Both methods calculate global variable importance and can therefore reveal the countries’ profile. However, SHAP provides local interpretability which stands between its advantages and it cannot be provided by the Gain method (we provide a comparison of the two methods in Appendix C).

To conduct the in-depth analysis we select to study four countries: (i) Saudi Arabia (Section 5.2.1), (ii) Yemen (Section 5.2.2), (iii) United Kingdom (Section 5.2.4), and (iv) the United States (Section 5.2.3).

5.2.1 Saudi Arabia

Based on the G20 list of countries [178], Saudi Arabia is one of the most powerful countries in the world in terms of military alliances, international alliances, political influence, economic influence, and leadership. Consequently, for our research purposes, it is of great interest.

Figure 5.6 presents the percentage error of Saudi Arabia for the 6-months-ahead GPI estimations. The performance is high, even for the 6-months-ahead GPI predictions, and the percentage error varies, in absolute values, from 4.05% to 11.38%. A positive percentage error indicates that the estimated GPI is higher than the real GPI, and therefore the model overestimates the monthly value. On the contrary, a negative percentage error illustrates that the estimated GPI is lower than the real GPI, and thus the model underestimates the monthly value. We obtain the largest negative percentage error for the GPI estimation for October 2018.

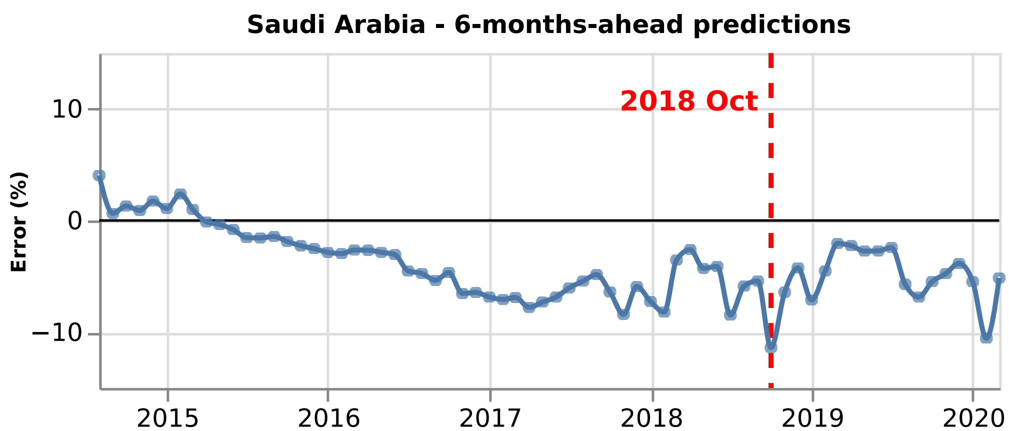


Fig. 5.6 Percentage error for the Saudi Arabia model. Percentage error of Saudi Arabia for the 6-months-ahead GPI estimations (blue curve). The performance is very high and the percentage error varies, in absolute values, from 4.05% to 11.38%. We obtain the largest negative percentage error for the GPI estimation for October 2018 (vertical dashed red line).

The analysis of the feature importance through SHAP reveals the country's profile and help understand the largest model errors. Figure 5.7 shows the global feature 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. The feature importance shows the profile of a powerful country in military, socio-economic and political terms: the important variables are related to embargo, boycott, or sanctions, diplomatic relations, mediations, economic cooperations, and appeals for aid, fights with military arms, military engagement, assaults, and endorsements. In Figure 5.7, we observe that "Fight with artillery and tanks" and "Appeal for aid" are among the most important variables for Saudi Arabia. This is reasonable, as they these GDELT variables could correspond to the "Volume of Transfers of Major Conventional Weapons, as recipient (imports) per 100,000 people" and the "Financial Contribution to UN Peacekeeping Missions" GPI indicators, respectively.

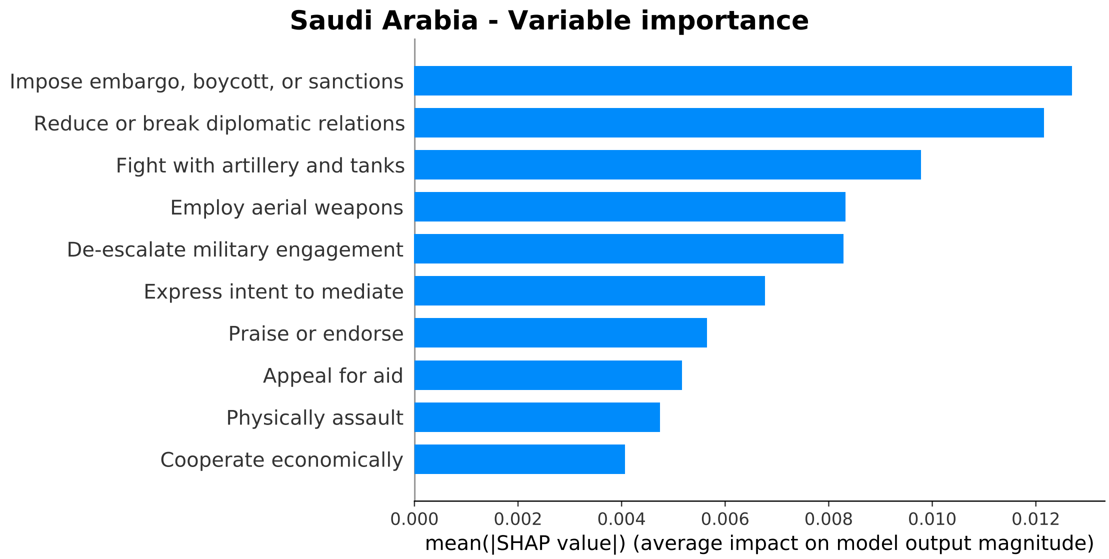


Fig. 5.7 Global variable importance plot for Saudi Arabia. Global feature importance plot for Saudi Arabia. The barplot orders the variables based on their importance in the estimation of the GPI score. Overall, we show that the variables demonstrate a profile of a powerful country in military, socio-economic, and political terms.

To better explain why the model has the worst performance in October 2018, we show the individual SHAP plot for Saudi Arabia, which indicates the most important variables that the model uses for the GPI estimation of October 2018 (Figure 5.8). 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 underestimates the GPI value (see Figure 5.6), we focus on the variables that push the GPI estimation lower. The most important variables to October 2018's 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 5.7). In October 2018, the journalist Jamal Khashoggi was assassinated at the Saudi consulate in Istanbul, Turkey. This event provoked a

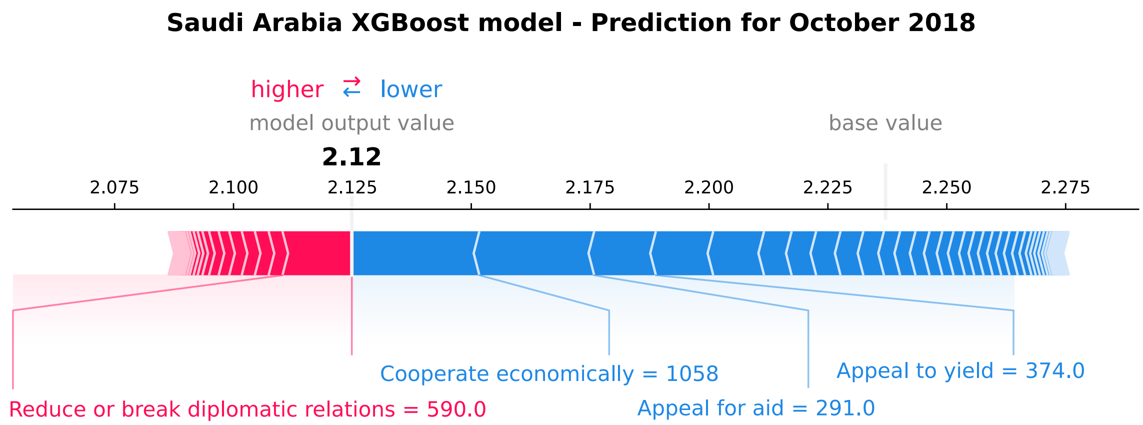


Fig. 5.8 Individual SHAP Value plot for Saudi Arabia. Individual SHAP Value plot for Saudi Arabia. It presents the model output value, i.e., the estimation of the GPI for October 2018, and the base value, which is the value that would be predicted if the variables for the current output were unavailable. 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.

series of news on the topics mentioned above in Saudi Arabia. Figure 5.9 presents Saudi Arabia model predictions with respect to the real GPI score and the variable "Cooperate economically". This variable shows an abrupt increase in October 2018 and pushes GPI prediction lower, showing a more peaceful month. Similarly, Figure 5.10 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 5.9, and 5.10 respectively) the model underestimates the monthly value. 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.

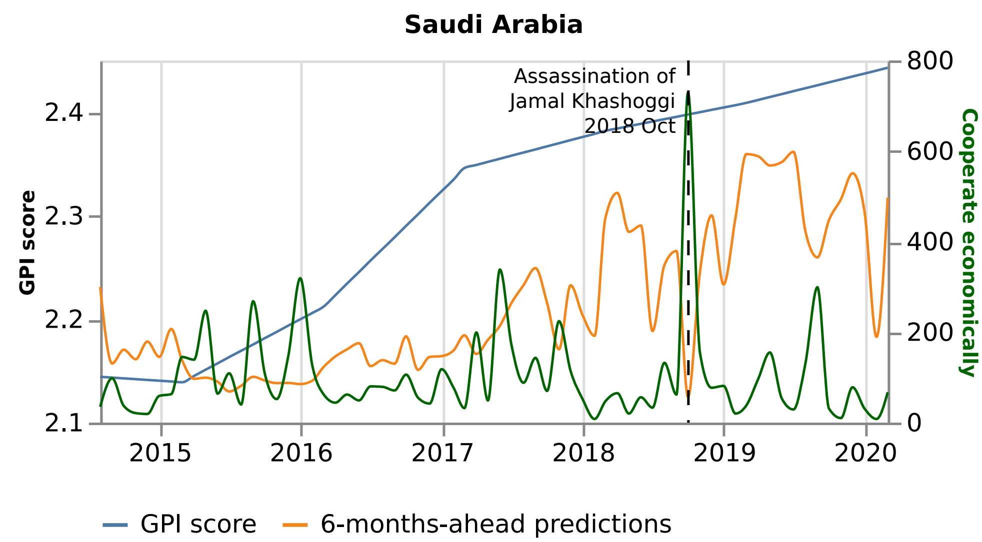


Fig. 5.9 Saudi Arabia predictions, with respect to the real GPI score, and the variable "Cooperate economically". Saudi Arabia 6-months-ahead predictions (orange curve), with respect to the real GPI score (blue curve), and the variable "Cooperate economically" (green curve). This variable pushes the model to underestimate the monthly value in October 2018 (vertical dashed black line). The reason for this error is the assassination of Jamal Khashoggi in this specific month.

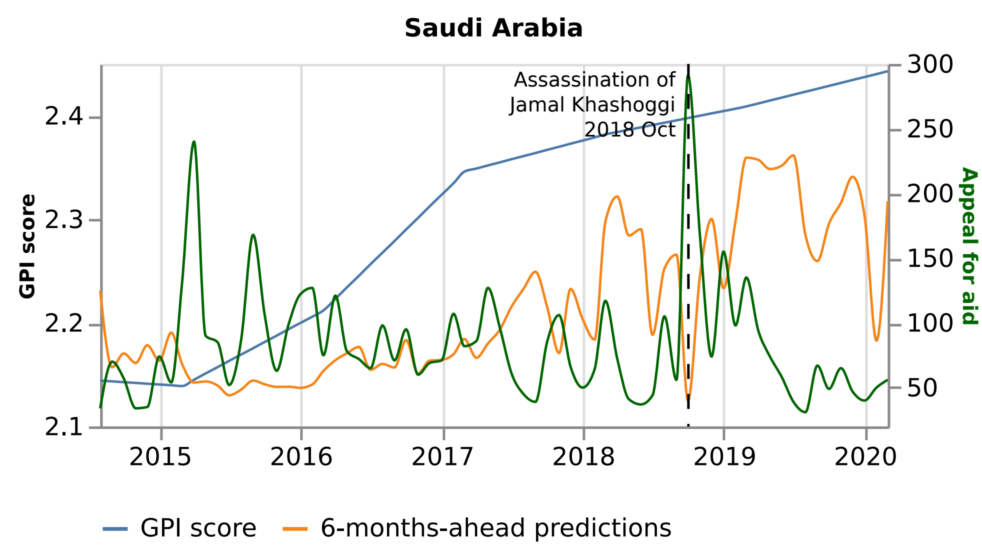


Fig. 5.10 Saudi Arabia predictions, with respect to the real GPI score, and the variable "Appeal for aid". Saudi Arabia predictions (orange curve), with respect to the real GPI score (blue curve), and the variable "Appeal for aid" (green curve). This variable pushes the model to underestimate the monthly value in October 2018 (vertical dashed black line). The reason for this error is the assassination of Jamal Khashoggi in this specific month.

5.2.2 Yemen

Based on the official GPI ranking [156], Yemen is one of the most war-torn countries in the world. Thus, for the current research purposes, it would be interesting to understand in-depth the model's behavior for such a country's profile.

For all country models, the training dataset has 72 values (six years). 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 real GPI value, which abruptly increases in 2015 (see [156]). Therefore, as explained in Section 4, 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. Therefore, six years of training data related to the pre-war period would not be useful for the model to predict peace after the start of the war, since the No. events related to the military, economic, and political situation of the country changes. By decreasing the training set to three years and using the rolling methodology, the model throws the pre-war historical data more quickly and learns from the most recent and relevant data related to the post-war period. Therefore, for Yemen we use data from March 2015 to March 2020 to understand the model's behavior during the Civil War period. Additionally, we study the 1-month-ahead predictions for the Yemen XGBoost model.

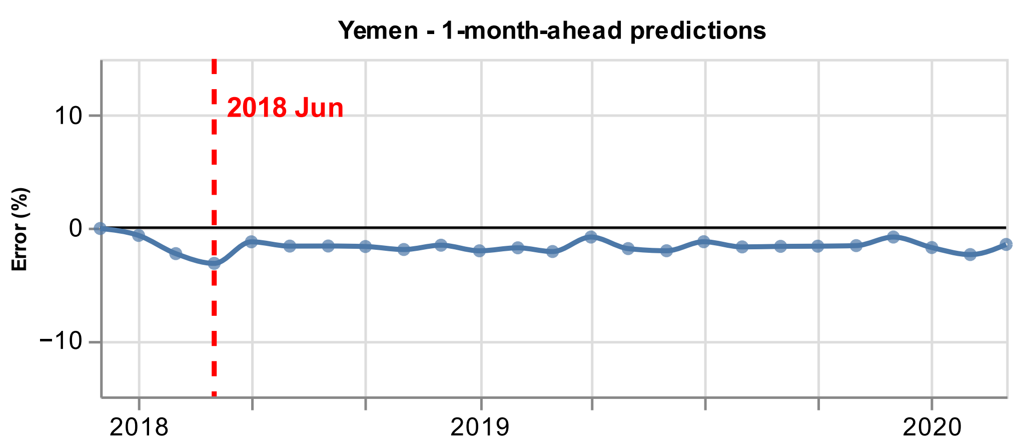


Fig. 5.11 Percentage error for Yemen. Percentage error for Yemen for the 1-month-ahead GPI estimations (blue curve). The percentage error varies, in absolute values, from 0.07% to 3.18%. We obtain the largest negative percentage error for the GPI estimation in June 2018 (vertical dashed red line).

Figure 5.11 presents the percentage error 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%. We obtain the largest negative percentage error (underestimation of GPI) for June 2018.

Figure 5.12 presents the global feature importance plot, which orders the variables based on their importance in the estimation of the GPI score. Each variable importance is calculated through SHAP, with a training period from June 2015 to May 2018. 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, the most important variables reveal a war-torn country profile: they are related to military aid, territory occupation, bombing, as well as negotiations, discussions, yields, visits, international involvements, and consults. In Figure 5.12, "Conduct non-military bombing" is among the most important variables. This is reasonable since this GDELT variable could correspond to the "Volume of Transfers of Major Conventional Weapons" GPI indicator.

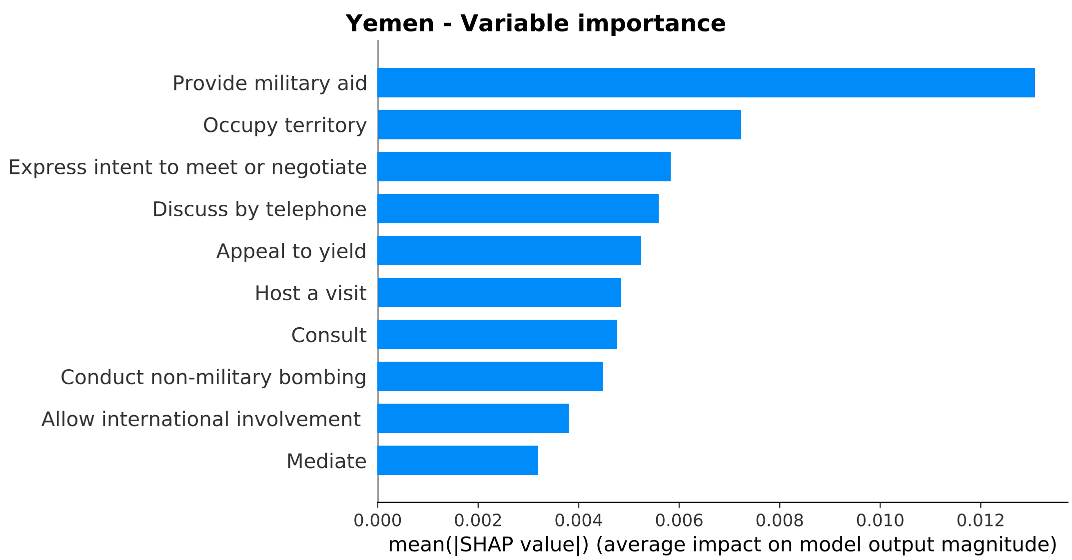


Fig. 5.12 Global feature importance plot for Yemen. Global feature importance plot for Yemen. The barplot orders the variables based on their importance in the estimation of the GPI score. Overall, we show that the variables mostly demonstrate a country with a war-torn profile.

Similarly to Saudi Arabia, we analyze at a local level to deeply understand why the model produces the highest percentage error in June 2018. Figure 5.13 presents the individual SHAP value plot, revealing 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. The base value is the GPI value that would be predicted if the variables for the current output were unavailable. The red arrows represent the variables that push the GPI estimation higher, i.e., “Conduct non-military bombing”. The blue arrows represent the variables that push the GPI estimation lower, i.e., “Discuss by telephone” and “Provide military aid”. Considering that in June 2018 the model underestimates the monthly value (see Figure 5.11), we focus our analysis on the latter variables.

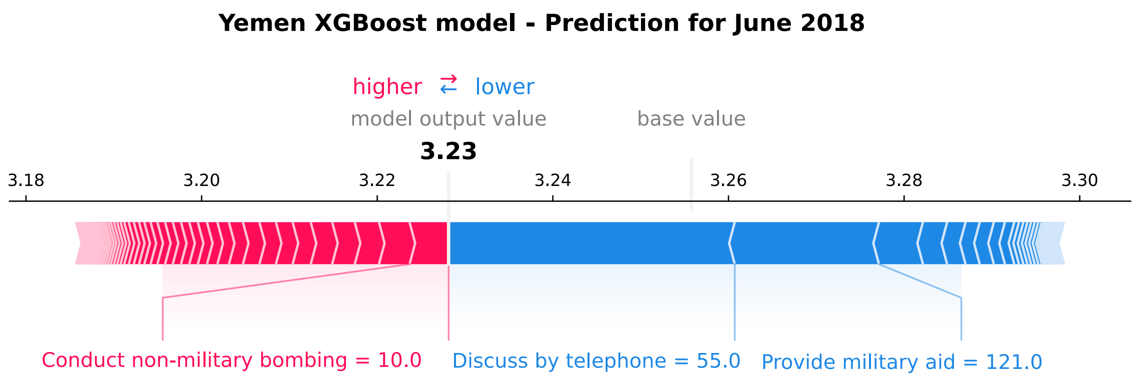


Fig. 5.13 Individual SHAP Value plot for Yemen. Individual SHAP Value plot for Yemen. It presents the model output value, i.e., the estimation of the GPI for June 2018, and the base value, which is the value that would be predicted if the variables for the current output were unavailable. The plot also 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.

In June 2018, the number of events on “Discuss by telephone” is 55, higher than the median value (14) of the previous three years. Similarly, the number of events on “Provide military aid” is 121, higher than the median value (72) of the previous three years. In June 2018, 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”. 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 “Provide military aid”.

Figure 5.14 and Figure 5.15 show that the variables’ higher monthly value and their mostly negative relationship with the GPI drive the model to underestimate the monthly value in June 2018. In other words, the model’s behavior reveals that this month the GPI value should be lower, and consequently, June 2018 results more peaceful that it really was. On the one hand, the model makes a wrong prediction, resulting in the largest percentage error. On the other hand, the model might give an interesting signal: although Yemen is involved in constant conflicts, June 2018 results more peaceful since the UN-brokered ceasefire agreement managed the withdrawal of the warring parties from Al Hudaydah in Yemen. Although we notice additional abrupt increases of the two variables' values, e.g., in November 2020 (Figure 5.14 and Figure 5.15), the model does not reproduce an abrupt decrease of the GPI. Consequently, the model shows its power to learn from its mistakes.

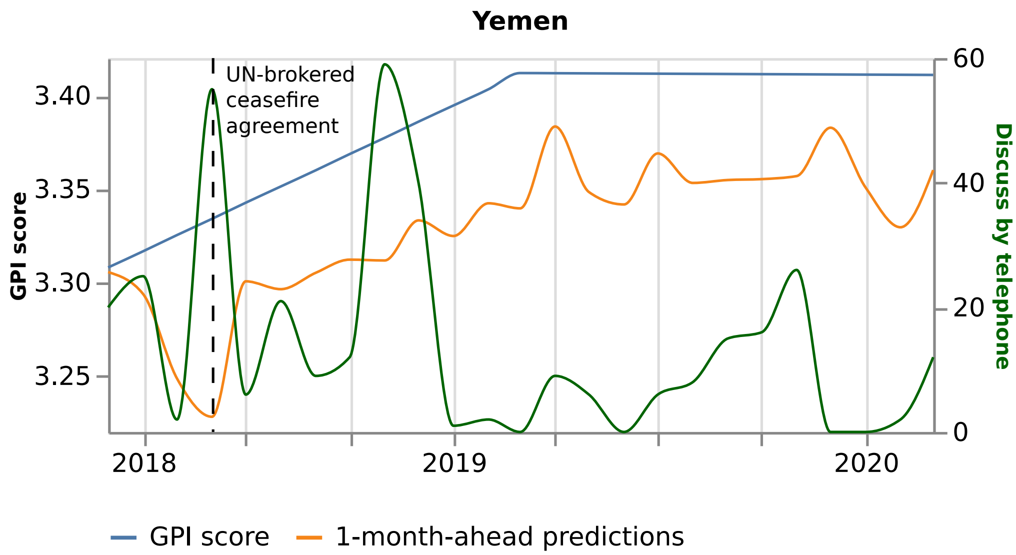


Fig. 5.14 Yemen predictions, with respect to the real GPI score and the variable "Discuss by telephone". Yemen 1-month-ahead predictions (orange curve), with respect to the real GPI score (blue curve) and the variable "Discuss by telephone" (green curve). This variable pushes the model to underestimate the monthly value of June 2018. The reason for this error is the increase of the news on the topic in this specific month.

5.2.3 United States

The United States is considered the most powerful country in the world [178]. The United States model has a high performance (see Table 4.1) and can provide policy-makers and peacekeepers with useful insights into the country's peacefulness before the real GPI score becomes available.

Figure 5.16 shows the most important variables for the training period between April 2014 and March 2020. Overall, these variables indicate a country profile of a strong player in the military, socio-economic, and political foreground. 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. In Figure 5.16, we observe that "Employ aerial weapons", "Fight with small arms and light weapons", and "Protest violently, riot" are among the most important variables for the United States.

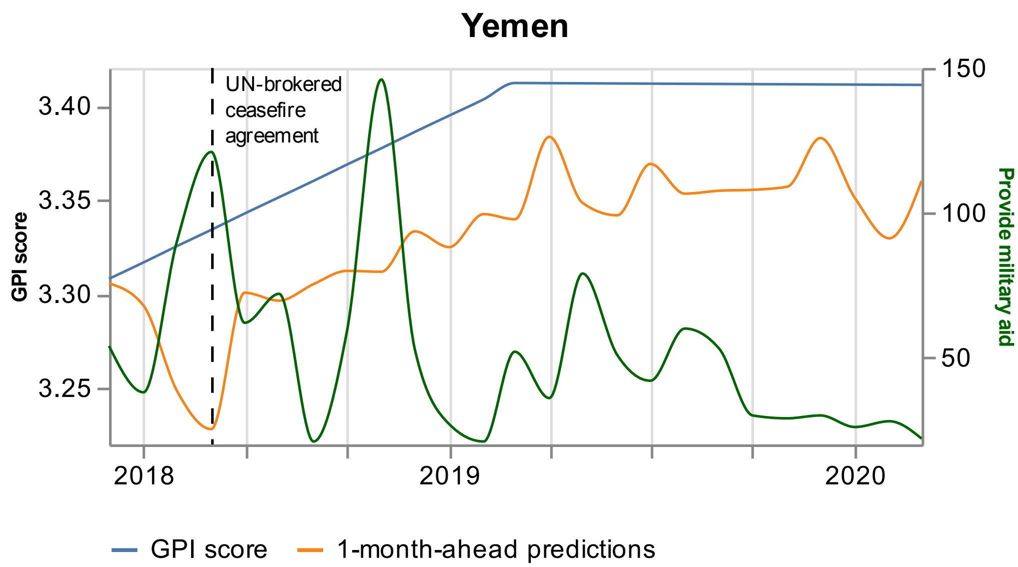


Fig. 5.15 Yemen predictions, with respect to the real GPI score and the variable "Provide military aid". Yemen 1-month-ahead model predictions (orange curve), with respect to the real GPI score (blue curve) and the variable "Provide military aid" (green curve). This variable pushes the model to underestimate the monthly value 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.

This is reasonable as these GDELT variables could correspond to GPI indicators "Nuclear and Heavy Weapons Capabilities", "Ease of Access to Small Arms and Light Weapons", and "Likelihood of violent demonstrations", respectively. Last, we compare the variables in Figure 5.16 with the ten variables that have the largest share over all news (see Table 3.2 in Section 3.2.2). None of the variables that have the largest share over all news is among the most important variables for the United States. This confirms that the model is not biased to learning only from the variables with the largest share, but it selects the variables that adequately serve for making the peacefulness prediction.

We now focus on the murder of George Floyd, which took place on May 25, 2020. Several protests followed this event at the end of May and for the whole of June 2020, provoking an amount of news concentrated on the topic. Figure 5.17 shows the local SHAP explanation for the prediction of June 2020. The estimated GPI (3-months-ahead prediction) is 2.30, indicating that the GPI value will remain stable 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). Particularly, "Protest violently, riot" is the variable that pushes the GPI estimation the lowest. 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

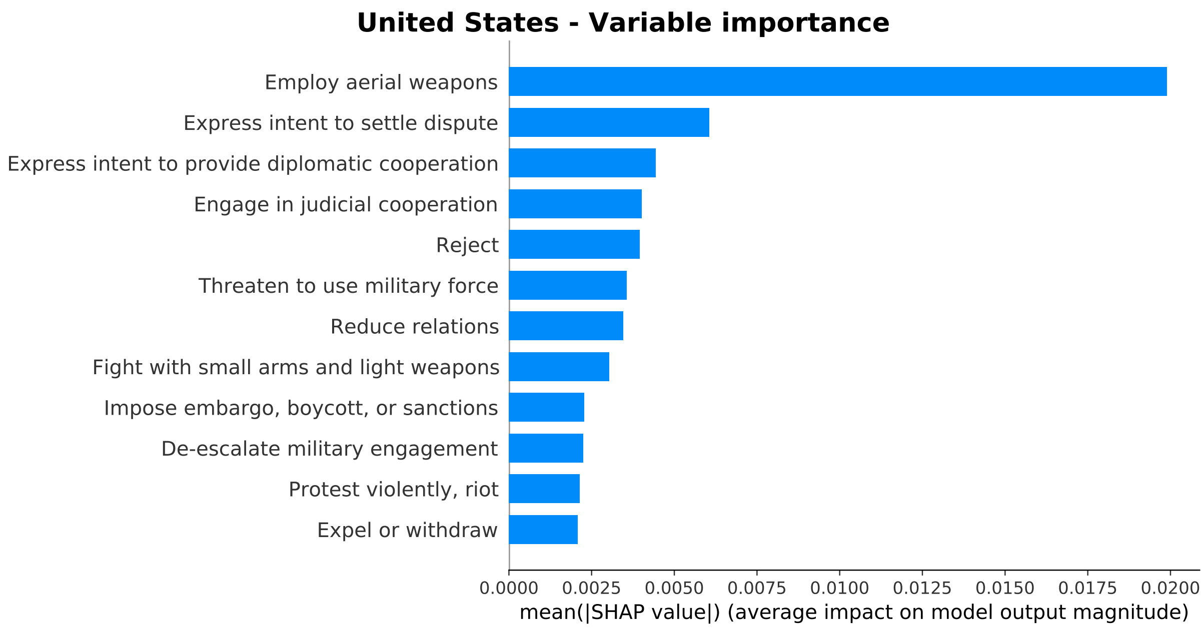


Fig. 5.16 Global variable importance plot for the United States. Global variable importance plot for the United States. The barplot orders the variables based on their importance in the estimation of the GPI score. Overall, we show that the variables indicate a country profile of a strong player in the military, socio-economic, and political foreground.

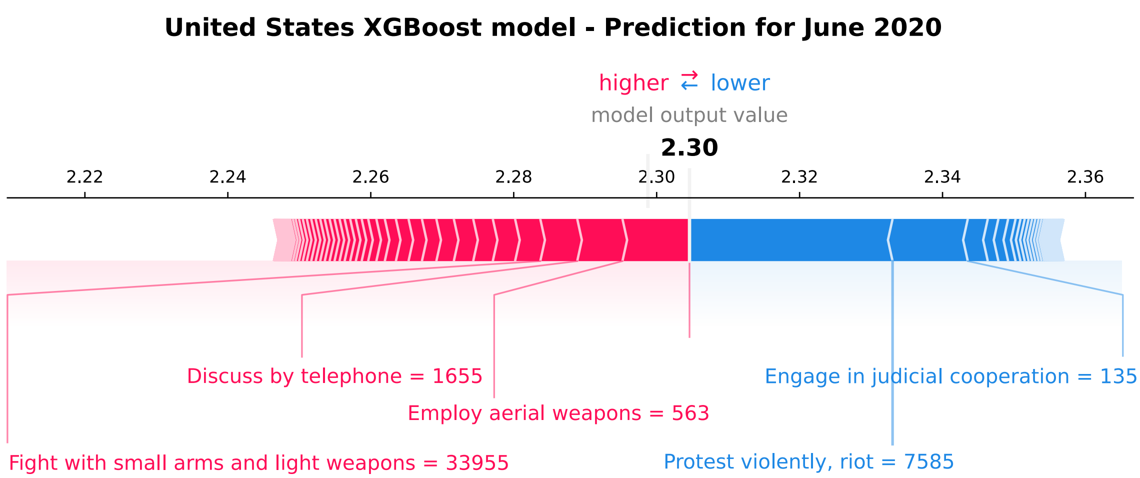


Fig. 5.17 Individual SHAP Value plot for the United States. Individual SHAP Value plot for the United States. It presents the model output value, i.e., the estimation of the GPI for June 2020, and the base value, which is the value that would be predicted if the variables for the current output were 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.

with the GPI. It seems that protesting in the United States contributes to the improvement of various socio-political situations, and as a consequence to peace-building.

The rest of the variables displayed in Figure 5.17 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 point out that, in this particular prediction, the most important variable for the overall training period, i.e., "Employ aerial weapons" (Figure 5.16) 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.

5.2.4 United Kingdom

Similar to the United States and Saudi Arabia, based on the list of G20 [178], the United Kingdom is considered one of the most powerful countries in the world. It is hence interesting for the European social policy-making to anticipate the level of peacefulness after the last ground-truth data, i.e., after March 2020.

Here we focus on the GPI prediction for July 2020, where various restrictions related to Covid-19 and the civilians' protection were announced. Figure 5.18 presents the global feature importance plot for a training period from April 2014 to March 2020. The figure highlights a country where various socio-political events occur: the important variables are mostly related to strikes or boycotts, appeals, negotiations, yields, relationships, and sanctions. "Engage in political dissent" is among the most important variables for the United Kingdom (Figure 5.18). This is reasonable as this variable could correspond to the GPI indicator "Likelihood of violent demonstrations".

To study peacefulness in July 2020, we need to deep into the analysis at a local level. Figure 5.19 presents the individual SHAP value plot for the United Kingdom: the GPI value is 1.8, and it is the model output value for the 4-months-ahead prediction. 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 (12,026), and the latter variable's value is 120, which is slightly lower than the median value of the previous six years (126). These results show that lower values of the aforementioned event categories decrease internal peace in the United Kingdom. The decrease in the events of these

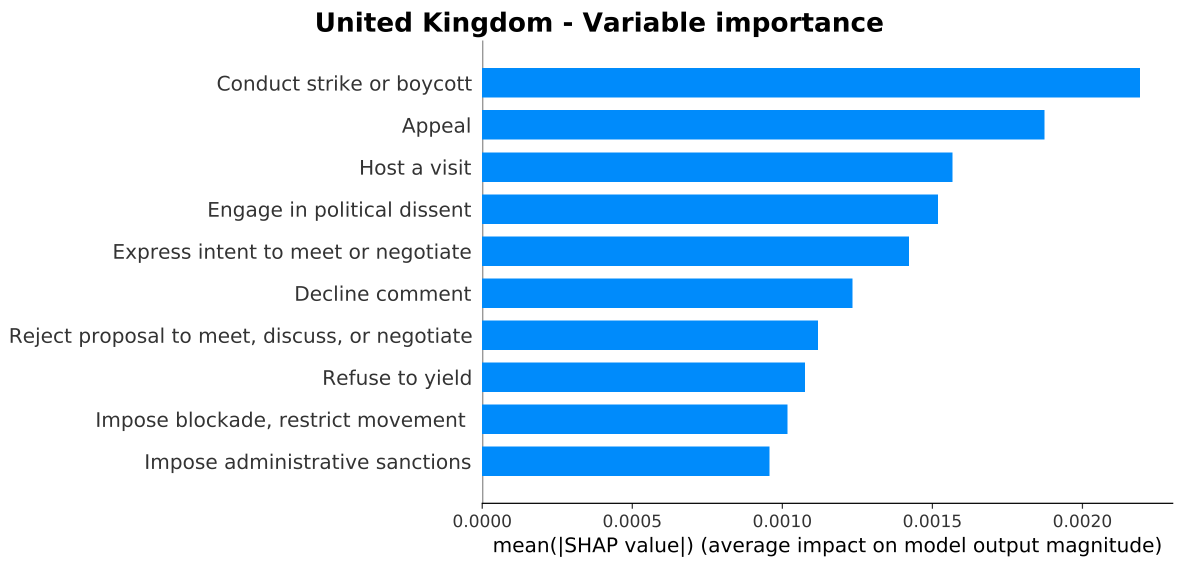


Fig. 5.18 Global variable importance plot for the United Kingdom. Global variable importance plot for the United Kingdom. The barplot orders the variables based on their importance in the estimation of the GPI score. Overall, we show that the variables mostly demonstrate a country where various socio-political events occur.

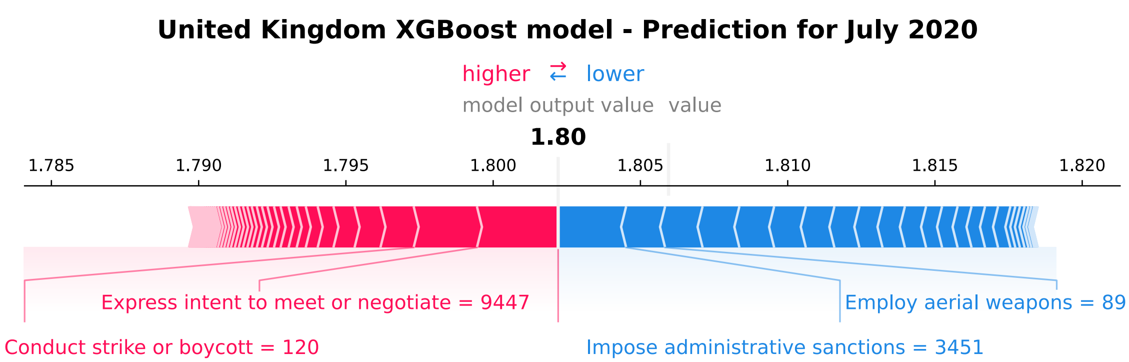


Fig. 5.19 Individual SHAP Value plot for the United Kingdom. Individual SHAP Value plot for the United Kingdom. It presents the model output value, i.e., the estimation of the GPI for July 2020, and the base value, which is the value that would be predicted if the variables for the current output were 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.

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). The news related to "Impose administrative sanctions" regard 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.

5.3 A tool for exploring countries' peace and its determinants through time

Our research is preeminently addressed to policy-makers. Therefore, it is crucial to communicate effectively our results in policy-making systems, and to understand how policy-makers process evidence and the environment in which they operate. Combining psychology and policy-studies, underline that the first step, for an adequate communication with policy-makers, is to "Understand your audience and tailor your response" [180]. Policy-makers should not be bombarded with evidence: they have too much information to process, and they use heuristics to filter information to make decisions quickly. It is therefore important to decode and synthesise the results produced from a research to help you tailor it to the ways in which policymakers demand and understand information.

Following the same path, for the current research, we created a dashboard that synthesises the most crucial information for the policy-makers, avoiding complicated academical data visualisation, and thus minimising the cognitive load. Telling the story of our results produced through this dashboard aims to maintain the messages clear and coherent, providing time for processing and reflection, with the use of specific examples. We provide the link of the dashboard in `http://experiments.sobigdata.eu/gpi_prediction/` [1].

In our dashboard, the user can choose the date and the country of interest to explore its peace. Figure 5.20 presents the example of Brazil for the October 2019 (the square bullet point is colored in dark grey). On the left side of the dashboard (Figure 5.20),

[1]The code for the creation of the dashboard can be find in `https://github.com/danielefadda/gpi_dashboard`

the user observes the map. The country selected, Brazil, is colored in dark grey. In October 2019, its predicted GPI is 2.228 and the real GPI is 2.342. On the right side of the dashboard (Figure 5.20), the user observes the plot depicting the real versus the predicted GPI for all dates, and can extract information for the percentage error. On the yellow area of the plot, which is zoomed in below as well, the user is informed for the predicted GPI for the dates the real GPI is not available yet.

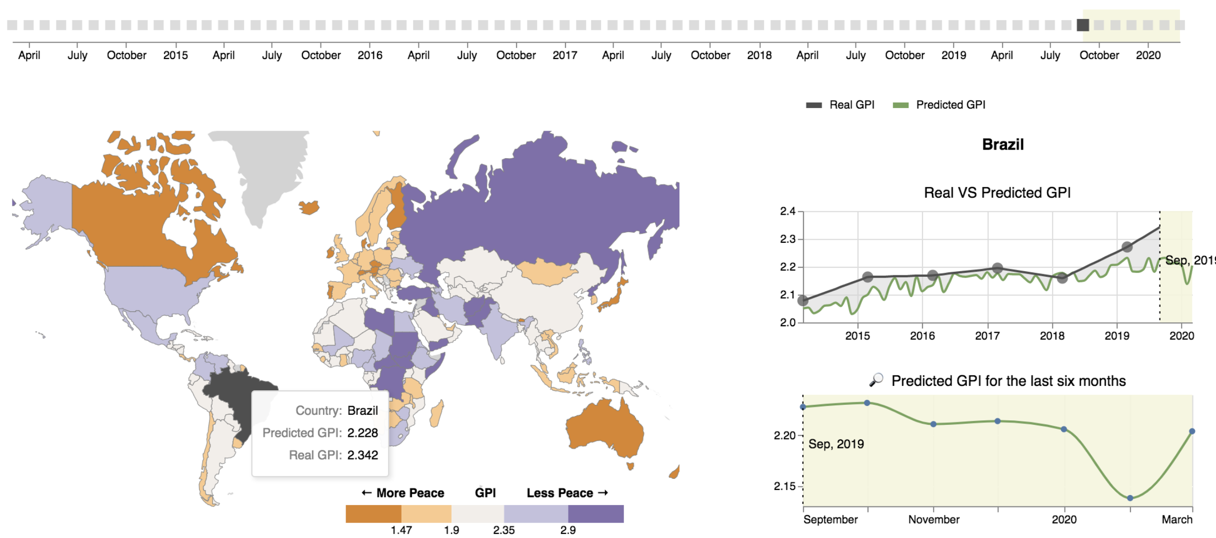


Fig. 5.20 The example of Brazil on the dashboard. The example of Brazil for October 2019. On the left side of the Figure the user observes the map. On the right side the user observes the plot depicting the real versus the predicted GPI for all dates, and can extract information for the percentage error. On the yellow area of the plot, which is zoomed in below as well, the user is informed for the predicted GPI for the dates the real GPI is not available yet.

Figure 5.21 presents additional capabilities of the dashboard. On the left side, we observe the world divided in regions and the clickable country buttons. This is an alternative to the map, particularly for countries that are not easy to look for on the map. On the right side, a user can find the prediction error for all GPI predictions (Figure 5.21). The variable importance plot illustrates to the user the most important factors that influence the GPI prediction for the selected date.

Last, the dashboard provides the user with all the information which might be useful. For example, the github repository for code reproducibility, the most important references of the study, as well as with answers for a list of frequently asked questions, such as where the scientific papers can be found.

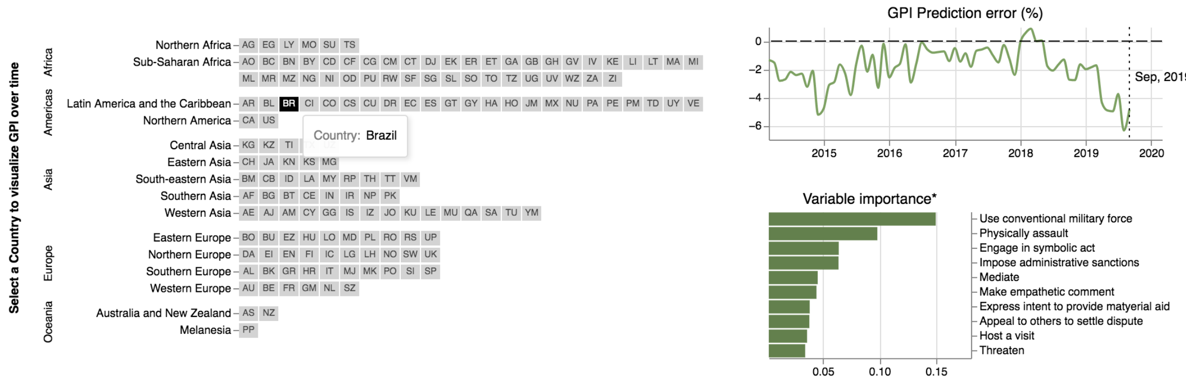


Fig. 5.21 The example of Brazil on the dashboard. Additional capabilities of the dash board. On the left side of the plot we observe the world divided in regions and the clickable country buttons. This is an alternative to the map, particularly for countries that are not easy to look for on the map. On the right side a user can find the prediction error for all GPI predictions. In addition, the variable importance plot illustrates to the user the most important factors that influence the GPI prediction for the selected date.

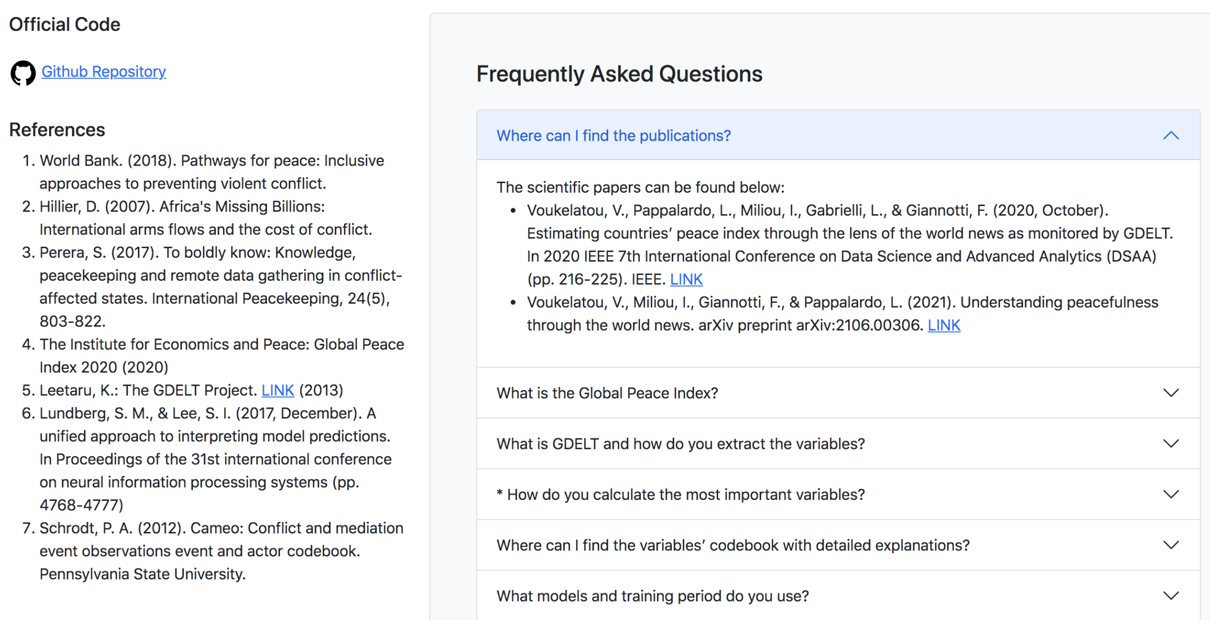


Fig. 5.22 Additional information on the dashboard. The user can find the link of the official code, the references, and frequently asked questions on the dash board.

Chapter 6

Conclusion

New technologies are becoming crucial in well-being research, offering many opportunities to advance it further. In particular, new digital data streams harnessed with AI techniques allow for predictive analytics to enhance early warning about emerging global challenges related to well-being, such as conflicts and operational risks, cost- and time-effectively.

This thesis starts with the conduction of an extensive literature review on well-being. We introduce the theoretical background of well-being and present the corresponding dimensions. This in-depth literature review shows that one of the most emerging well-being dimensions is safety, closely interrelated with peace. We then present the novel digital data sources used to capture peace, peace- and safety-related determinants and the corresponding studies conducted on the topic. This analysis reveals that although new technologies have been increasingly acknowledged as critical tools to foster peace [181, 182], research on AI for peace is still at the very beginning.

Given those mentioned above, the main aim of the thesis is to capture peace though the GPI using novel digital and AI tools. To tackle this task, we exploit GDELT, a database containing digital news related to socio-political events, and we use AI techniques to measure the monthly GPI values. Measuring the GPI at a monthly level indicates 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. In addition, 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. We observe that there are countries for which the model performance is high, while for others, the model performance is medium or low. In particular, we aim to demonstrate that GDELT is a good proxy for estimating the GPI values. For this reason, we conduct in-depth analysis on the models that strongly confirm this hypothesis, i.e., country models with high performance. Among others, we select Saudi Arabia, Yemen, the United States, and the United Kingdom, and we apply explainability techniques, i.e., the SHAP methodology. This allows us to explain the countries' peace and the determinants of peace. We also 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 model are related to military aid, territory occupation, bombing, negotiations, discussions, yields, visits, international involvements, and consults, revealing a war-torn country profile. Additionally, we use explainable AI techniques to provide local explanations of the models' behavior to understand how the contribution of the most important variables changes 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 some aspects of our study that we should take into consideration. Considering that GPI is a yearly index, we upsampled its yearly values linearly to monthly values. The linear upsampling is an assumption since the monthly data generated do not correspond to the real monthly value, which might add bias to the models' results. In addition, from a statistical point of view, upsampling the GPI data with linear estimates, which do not correspond to the true values, increases the observations compared to the available observations. This might wrongly increase the confidence of the models' results, for example, by biasing confidence intervals [183]. Alternatively, another assumption could be to increase the frequency of GPI through stochastic differential equation (SDE) methods [184], a more complex methodology than simple linear interpolation. Considering that both solutions are assumptions and that our main goal is to demonstrate that monthly peace can be captured through the news data, we choose the simplest one. Future studies could deepen the analysis by trying different upsampling methodologies. An alternative solution could be replacing GPI with a monthly index, which would not require upsampling.

Additionally, news media might introduce biases. Consequently, these biases might influence the models' results. Any bias could drive the models to show low performance in predicting the GPI value. Therefore, future research could study, for example, in-depth the representativeness of GDELT news, as some countries might be under-represented or over-represented, which could help to explain why some models fail to demonstrate high or at least medium performance.

Another line of future research lies in analyzing the results per country. One approach to improve the models' performance is to change the training data length based on the history of the country, usually depicted on the GPI. For example, as we show for Yemen, the performance improves by changing the training data from the most recent 72 months to the most recent 36 months.

Furthermore, additional independent variables could be added to the models to improve the performance of our models. For example, we conduct a preliminary analysis by adding a 12-months-lagged GPI to the XGBoost country models. This variable improves the models' performance since it incorporates feedback effect to the models over time (see Appendix D for more details on the results). We also conduct another preliminary analysis by adding a salience variable to the XGBoost models. This variable could illustrate whether the global news is concentrated on a specific country for a certain month. Results demonstrate that the countries' performance slightly improves (see Appendix E for more details on the results). Future studies could dig deeper into these aspects.

Moreover, as explained in Section 3.2 some GDELT event categories might not be present in the news of a country. Other data sources, such as ACLED news data [136] or social-media data, such as Twitter data, could be combined with the GDELT data to help overcome this limitation.

Also, future work could engage peace experts, such as peace-makers, to provide domain expert interpretations of the models' results. For example, although in Section 5.2 we interpret the most important variables indicated by the countries' models, interpretations could be improved by experts in the field. Similarly, the tool built (Section 5.3) to inform the policy-makers could be evaluated for its utility. For example, a case study could be conducted to collect feedback from policy-makers on the usefulness of the tool.

In addition, we highlight that machine learning models are a powerful tool for solving prediction problems. Still, they are not inherently causal, and interpreting them with explainable AI techniques fails to answer causal questions accurately [26]. Therefore, we indicate two additional points that can improve early-warning conflict systems: more information about the causes of conflicts and war and theoretical models representing the complexity of social interactions and human decision-making. In particular, future AI-based conflict models should offer explanations for conflicts and war and plans for preventing them. This is a difficult task because conflict and war dynamics are multi-dimensional, and the data collected today are too narrow, sparse, and disparate [16].

We have learned a lot about safety and peace from an objective point of view. However, as discussed in Section 2.1 the subjective perspective is very important for well-being studies. For example, it is different how peaceful a place is and how peaceful people perceive a place. Future studies could include the subjective approach in the analysis for more thorough peace measurements. For example, researchers could extract subjective measurements of peace from the news articles by applying Natural Language Processing (NLP) techniques.

Safety is a crucial well-being dimension and might influence other well-being dimensions. For example, war and conflict can cause food insecurity and hunger, just as hunger and food insecurity can cause latent conflicts to flare up and trigger the use of violence [185]. Despite this, future research could replicate this study to capture other well-being dimensions. For instance, the World Food Programme (WFP) is currently interested in using novel digital data and predictive machine learning techniques to create a nutrition deficiency index (e.g., [186]). This research would cover health, another objective well-being dimension. Similarly, World Health Organization is interested in using novel digital data and NLP techniques for creating an index measuring how age-friendly older people perceive a city. This research would cover the health well-being dimension as well, but from a subjective point of view [187].

We want to highlight that we do not mean that novel digital data should substitute official data. Innovative data, if harnessed correctly through AI tools, can be combined with official data to inform and empower the social good decision-making [188]. Indeed, researchers highlight both the positive opportunities that are created through data-driven studies and the potential negative consequences that practitioners should be aware of and address in to take advantage of the potentialities of this emergent field [189]. For example, since, usually, the data used are personal, if not sensitive, and are analyzed to shape policy and to make decisions [190, 191], ethical concerns may arise, such as privacy and respect to human rights. In the European Union, additional attention to the topic has been brought after the implementation of the General Data Protection Regulation (GDPR). Researchers need to take into consideration the ethical challenges and not overlook them but address them successfully. Only by facing ethical problems researchers can maximize the contributing value of data science studies for society.

Last but not least, we would like to underline that AI technologies can bring forward a society, or the same technologies, if in the wrong hands, can be used as weapons [192]. Any technology can be hacked, weaponized, and used in a way researchers did not intend. Therefore, we need to ensure we do not create unintended consequences even with good intentions. We need a framework for AI's ethical and safe deployment in well-being studies, such as the peace study we conducted in this thesis. We need to define the ethical use of AI and embed those ethical standards in innovative global governance systems based on international law. Furthermore, people knowledgeable in AI must ask the right questions at the right time. This way, we can proactively detect unintended uses and consequences of AI and have a timely action to design safeguards against them.

All in all, we have a lot to learn, improve and safeguard from. However, there is a lot of hope. We believe our results show great promise for peace and all well-being studies and can benefit humanity. This research could be valuable to policy-makers, non-governmental organizations, UN agencies, and the scientific community, especially researchers interested in Data Science and AI for Social Good. We aim to witness positive changes in well-being and in society overall, brought by AI and novel digital data once we turn back in 2030 to observe how the previous decade looked like and what advances have been made.

References

[1] Nenad Tomašev, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle C M Belgrave, Daphne Ezer, Fanny Cachat van der Haert, Frank Mugisha, Gerald Abila, Hiromi Arai, Hisham Almiraat, Julia Proskurnia, Kyle Snyder, Mihoko Otake-Matsuura, Mustafa Othman, Tobias Glasmachers, Wilfried de Wever, Yee Whye Teh, Mohammad Emtiyaz Khan, Ruben De Winne, Tom Schaul, and Claudia Clopath. AI for social good: unlocking the opportunity for positive impact. *Nature Communications*, 11(1):2468, 2020.

[2] Marc Fleurbaey. 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] Vasiliki Voukelatou, Lorenzo Gabrielli, Ioanna Miliou, Stefano Cresci, Rajesh Sharma, Maurizio Tesconi, and Luca Pappalardo. Measuring objective and subjective well-being: dimensions and data sources. *International Journal of Data Science and Analytics*, pages 1–31, 2020.

[5] Rachel Dodge, Annette P Daly, cJan Huyton, and Lalage D Sanders. The challenge of defining wellbeing. *International journal of wellbeing*, 2(3), 2012.

[6] Sabina Alkire. Dimensions of human development. *World development*, 30(2):181–205, 2002.

[7] UN. United Nations Conference on Sustainable Development, Rio+20 . `UnitedNationsConferenceonSustainableDevelopment,Rio+20`, 2012.

[8] UNDP. Sustainable Development Goals. `https://sustainabledevelopment.un.org/sdgs`, 2015.

[9] Christian Kroll, Anne Warchold, and Prajal Pradhan. Sustainable development goals (sdgs): Are we successful in turning trade-offs into synergies? *Palgrave Communications*, 5(1):1–11, 2019.

[10] David Le Blanc. Towards integration at last? the sustainable development goals as a network of targets. *Sustainable Development*, 23(3):176–187, 2015.

[11] United Nations. A world that counts: mobilizing the data revolution for sustainable development. Technical report, 2014.

[12] Sustainable Development Solutions Network. Indicators and a Monitoring Framework for the Sustainable Development Goals, Launching a Data Revolution for the SDGs, 2015. United Nations New York.

[13] Rapporto, BES. Il benessere equo e sostenibile in Italia, 2015. ISTAT.

[14] Ed Diener, Shigehiro Oishi, and Louis Tay. Advances in subjective well-being research. *Nature Human Behaviour*, page 1, 2018.

[15] Martin Wählisch. Big data, new technologies, and sustainable peace: Challenges and opportunities for the un. *Journal of Peacebuilding and Development*, 15(1):122–126, 2020.

[16] Weisi Guo, Kristian Gleditsch, and Alan Wilson. Retool ai to forecast and limit wars, 2018.

[17] Suda Perera. To boldly know: Knowledge, peacekeeping and remote data gathering in conflict-affected states. *International Peacekeeping*, 24(5):803–822, 2017.

[18] World Bank. Pathways for peace: Inclusive approaches to preventing violent conflict, 2018.

[19] Debbie Hillier. Africa's missing billions: International arms flows and the cost of conflict. 2007.

[20] The Institute for Economics and Peace. Global Peace Index 2020, 2020.

[21] Kalev Leetaru. The GDELT Project. `https://www.gdeltproject.org/`, 2013.

[22] Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik Van Der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. Sentiment analysis in the news. *arXiv preprint arXiv:1309.6202*, 2013.

[23] Haewoon Kwak and Jisun An. A first look at global news coverage of disasters by using the gdelt dataset. In *International Conference on Social Informatics*, pages 300–308. Springer, 2014.

[24] Scott M Lundberg, Gabriel G Erion, and Su-In Lee. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*, 2018.

[25] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*, 2017.

[26] Eleanor Dillon, Jacob LaRiviere, Scott Lundberg, Jonathan Roth, and Vasilis Syrgkanis. Be careful when interpreting predictive models in search of causal insights. `https://shap.readthedocs.io/en/latest/example_notebooks/overviews/Be%20careful%20when%20interpreting%20predictive%20models%20in%20search%20of%20causal%C2%A0insights.html#Be-careful-when-interpreting-predictive-models-in-search-of-causal%C2%A0insights`, 2018.

[27] Shi, Zheyuan Ryan and Wang, Claire and Fang, Fei. Artificial Intelligence for Social Good: A Survey. *arXiv preprint arXiv:2001.01818*, 2020.

[28] Vasiliki Voukelatou, Luca Pappalardo, Ioanna Miliou, Lorenzo Gabrielli, and Fosca Giannotti. 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)*, pages 216–225. IEEE, 2020.

[29] Vasiliki Voukelatou, Ioanna Miliou, and Luca Pappalardo. Stima dell'indice di pace attraverso notizie digitali. *Lettura ragionata dell'Enciclica Papale "Fratelli tutti", alla luce dell'Obiettivo 16 dell'Agenda Onu 2030*, pages 57–60, 2021.

[30] Vasiliki Voukelatou, Ioanna Miliou, Fosca Giannotti, and Luca Pappalardo. Understanding peacefulness through the world news. *EPJ Data Science*, 2022.

[31] Carmen M Reinhart and Vincent R Reinhart. After the fall. Technical report, National Bureau of Economic Research, 2010.

[32] Joseph Eugene Stiglitz, Amartya Sen, and Jean-Paul Fitoussi. Report by the Commission on the Measurement of Economic Performance and Social Progress, 2009. The Commission Paris.

[33] Sudhir Anand and Amartya Sen. Human Development Index: Methodology and Measurement., 1994.

[34] OECD. OECD Better Life Index. `https://www.oecdbetterlifeindex.org/#/11111111111`, 2011. (Online; accessed October 2019).

[35] WHO. world report on ageing and health. `https://apps.who.int/iris/bitstream/handle/10665/186463/9789240694811_eng.pdf?sequence=1&isAllowed=y`, 2015.

[36] R Veenhoven. Conditions of happiness, Reidel (now Springer), Dordrecht, The Netherlands, 1984.

[37] Bruno S Frey and Alois Stutzer. What can economists learn from happiness research? *Journal of Economic literature*, 40(2):402–435, 2002.

[38] Joseph E Stiglitz, Amartya Sen, and Jean-Paul Fitoussi. Measurement of economic performance and social progress. *Online document http://bit. ly/JTwmG. Accessed June*, 26:2012, 2009.

[39] Meike Bartels and Dorret I Boomsma. Born to be happy? The etiology of subjective well-being. *Behavior genetics*, 39(6):605, 2009.

[40] Meike Bartels, Viatcheslav Saviouk, Marleen HM De Moor, Gonneke Willemsen, Toos CEM van Beijsterveldt, Jouke-Jan Hottenga, Eco JC De Geus, and Dorret I Boomsma. Heritability and genome-wide linkage scan of subjective happiness. *Twin Research and Human Genetics*, 13(2):135–142, 2010.

[41] Ragnhild Bang Nes and Espen Røysamb. The heritability of subjective well-being: Review and meta-analysis. *The genetics of psychological well-being: The role of heritability and genetics in positive psychology*, pages 75–96, 2015.

[42] Ragnhild Bang Nes, N Czajkowski, and K Tambs. Family matters: happiness in nuclear families and twins. *Behavior genetics*, 40(5):577–590, 2010.

[43] RB Nes, E Røysamb, K Tambs, JR Harris, and T Reichborn-Kjennerud. Subjective well-being: genetic and environmental contributions to stability and change. *Psychological medicine*, 36(7):1033–1042, 2006.

[44] Espen Røysamb, Jennifer R Harris, Per Magnus, Joar Vittersø, and Kristian Tambs. Subjective well-being. Sex-specific effects of genetic and environmental factors. *Personality and individual differences*, 32(2):211–223, 2002.

[45] Espen Røysamb, Kristian Tambs, Ted Reichborn-Kjennerud, Michael C Neale, and Jennifer R Harris. Happiness and health: environmental and genetic contributions to the relationship between subjective well-being, perceived health, and somatic illness. *Journal of personality and social psychology*, 85(6):1136, 2003.

[46] Jason Schnittker. Happiness and success: Genes, families, and the psychological effects of socioeconomic position and social support. *American Journal of Sociology*, 114(S1):S233–S259, 2008.

[47] Pleeging, Emma and Burger, Martijn and van Exel, Job. The relations between hope and subjective well-being: A literature overview and empirical analysis. *Applied Research in Quality of Life*, pages 1–23, 2020.

[48] Douglas T Kenrick, Vladas Griskevicius, Steven L Neuberg, and Mark Schaller. Renovating the pyramid of needs: Contemporary extensions built upon ancient foundations. *Perspectives on psychological science*, 5(3):292–314, 2010.

[49] Richard M Ryan and Edward L Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1):68, 2000.

[50] Louis Tay and Ed Diener. Needs and subjective well-being around the world. *Journal of personality and social psychology*, 101(2):354, 2011.

[51] Andrew E Clark and Andrew J Oswald. Satisfaction and comparison income. *Journal of public economics*, 61(3):359–381, 1996.

[52] Michael A Shields, Stephen Wheatley Price, and Mark Wooden. Life satisfaction and the economic and social characteristics of neighbourhoods. *Journal of Population Economics*, 22(2):421–443, 2009.

[53] Nattavudh Powdthavee. How much does money really matter? Estimating the causal effects of income on happiness. *Empirical economics*, 39(1):77–92, 2010.

[54] Boris Nikolaev. Living with mom and dad and loving it... or are you? *Journal of Economic Psychology*, 51:199–209, 2015.

[55] Paul Dolan, Tessa Peasgood, and Mathew White. Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of economic psychology*, 29(1):94–122, 2008.

[56] Richard A Easterlin. Does economic growth improve the human lot? Some empirical evidence. In *Nations and households in economic growth*, pages 89–125. Elsevier, 1974.

[57] Ruut Veenhoven. Is happiness relative? *Social indicators research*, 24(1):1–34, 1991.

[58] Ed Diener, Louis Tay, and Shigehiro Oishi. Rising income and the subjective well-being of nations. *Journal of personality and social psychology*, 104(2):267, 2013.

[59] Ruut Veenhoven and Floris Vergunst. The Easterlin illusion: economic growth does go with greater happiness. *International Journal of Happiness and Development*, 1(4):311–343, 2014.

[60] Daniel W Sacks, Betsey Stevenson, and Justin Wolfers. The new stylized facts about income and subjective well-being. *Emotion*, 12(6):1181, 2012.

[61] Benjamin Radcliff and Gregory Shufeldt. Direct democracy and subjective well-being: the initiative and life satisfaction in the American states. *Social Indicators Research*, 128(3):1405–1423, 2016.

[62] Ruut Veenhoven. Social conditions for human happiness: A review of research. *International Journal of Psychology*, 50(5):379–391, 2015.

[63] Luca Pappalardo, Salvatore Rinzivillo, and Filippo Simini. Human mobility modelling: Exploration and preferential return meet the gravity model. *Procedia Computer Science*, 83:934 – 939, 2016. The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016) / The 6th International Conference on Sustainable Energy Information Technology (SEIT-2016) / Affiliated Workshops.

[64] Roberto Pellungrini, Luca Pappalardo, Francesca Pratesi, and Anna Monreale. A data mining approach to assess privacy risk in human mobility data. *ACM Trans. Intell. Syst. Technol.*, 9(3):31:1–31:27, December 2017.

[65] Johannes C Eichstaedt, Hansen Andrew Schwartz, Margaret L Kern, Gregory Park, Darwin R Labarthe, Raina M Merchant, Sneha Jha, Megha Agrawal, Lukasz A Dziurzynski, Maarten Sap, et al. Psychological language on twitter predicts county-level heart disease mortality. *Psychological science*, 26(2):159–169, 2015.

[66] Marta M Jankowska, Jasper Schipperijn, and Jacqueline Kerr. A framework for using GPS data in physical activity and sedentary behavior studies. *Exercise and sport sciences reviews*, 43(1):48, 2015.

[67] Yury Kryvasheyeu, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian. Rapid assessment of disaster damage using social media activity. *Science advances*, 2(3):e1500779, 2016.

[68] OECD. OECD Better Life Index: Safety. `http://www.oecdbetterlifeindex.org/topics/safety/`, 2011. (Online; accessed October 2019).

[69] UNDP. peace, justice and strong institutions. `https://sdgs.un.org/goals/goal16`, 2015.

[70] Johan Galtung. *Peace by peaceful means: Peace and conflict, development and civilization*, volume 14. Sage, 1996.

[71] Royce Anderson. A definition of peace. *Peace and Conflict: Journal of Peace Psychology*, 10(2):101, 2004.

[72] Markus Brückner and Antonio Ciccone. International commodity prices, growth and the outbreak of civil war in Sub-Saharan Africa. *The Economic Journal*, 120(544):519–534, 2010.

[73] Håvard Hegre, Joakim Karlsen, Håvard Mokleiv Nygård, Håvard Strand, and Henrik Urdal. Predicting armed conflict, 2010–2050. *International Studies Quarterly*, 57(2):250–270, 2013.

[74] Thomas Chadefaux. Market anticipations of conflict onsets. *Journal of Peace Research*, 54(2):313–327, 2017.

[75] David J Solomon. Conducting web-based surveys. *Practical assessment research and evaluation*, 7(19), 2001.

[76] Philip S Brenner and John DeLamater. Lies, damned lies, and survey self-reports? identity as a cause of measurement bias. *Social psychology quarterly*, 79(4):333–354, 2016.

[77] C Melamed, L Morales, Y Hsu, J Poole, B Rae, I Rutherford, and A Jahic. A World That Counts-Mobilising the Data Revolution for Sustainable Development. *United Nations. Accessed March*, 12:2018, 2014.

[78] L. Spinsanti, M. Berlingerio, and L. Pappalardo. *Mobility and Geo-Social Networks*, page 315–333. Cambridge University Press, 2013.

[79] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kiciman. Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data*, 2:13, 2019.

[80] Mattias Rost, Louise Barkhuus, Henriette Cramer, and Barry Brown. Representation and communication: challenges in interpreting large social media datasets. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 357–362. ACM, 2013.

[81] Rafael Prieto Curiel, Stefano Cresci, Cristina Ioana Muntean, and Steven Richard Bishop. Crime and its fear in social media. *Palgrave Communications*, 6(1):1–12, 2020.

[82] Feng Chen and Daniel B Neill. 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*, pages 1166–1175. ACM, 2014.

[83] Ethan Spangler and Ben Smith. Let them tweet cake: Estimating public dissent using twitter. *Defence and Peace Economics*, pages 1–20, 2021.

[84] Esraa Najjar and Salam Al-augby. Sentiment analysis combination in terrorist detection on twitter: A brief survey of approaches and techniques. *Research in Intelligent and Computing in Engineering*, pages 231–240, 2021.

[85] Sofea Azrina Azizan and Izzatdin Abdul Aziz. Terrorism detection based on sentiment analysis using machine learning. *Journal of Engineering and Applied Sciences*, 12(3):691–698, 2017.

[86] Thomas Zeitzoff. Using social media to measure conflict dynamics: An application to the 2008-2009 Gaza conflict. *Journal of Conflict Resolution*, 55(6):938–969, 2011.

[87] Eugenia Siapera, Graham Hunt, and Theo Lynn. # GazaUnderAttack: Twitter, Palestine and diffused war. *Information, Communication and Society*, 18(11):1297–1319, 2015.

[88] Riley Tucker, Daniel T O'Brien, Alexandra Ciomek, Edgar Castro, Qi Wang, and Nolan Edward Phillips. 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*, pages 1–27, 2021.

[89] Thomas Zeitzoff, John Kelly, and Gilad Lotan. 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.

[90] Emilio Zagheni, Venkata Rama Kiran Garimella, Ingmar Weber, and Bogdan State. Inferring international and internal migration patterns from twitter data. In *Proc. of the 23rd International Conference on World Wide Web*, pages 439–444, 2014.

[91] Cristina Kadar, Raquel Rosés Brüngger, and Irena Pletikosa. Measuring ambient population from location-based social networks to describe urban crime. In *International conference on social informatics*, pages 521–535. Springer, 2017.

[92] Emilio Zagheni, Ingmar Weber, and Krishna Gummadi. Leveraging facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, pages 721–734, 2017.

[93] Béatrice Mazoyer, Julia Cagé, Nicolas Hervé, and Céline Hudelot. A french corpus for event detection on twitter. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 6220–6227, 2020.

[94] Xinyu Chen, Youngwoon Cho, and Suk Young Jang. Crime prediction using Twitter sentiment and weather. In *2015 Systems and Information Engineering Design Symposium*, pages 63–68. IEEE, 2015.

[95] Monica Alexander, Kivan Polimis, and Emilio Zagheni. Combining social media and survey data to nowcast migrant stocks in the united states. *Population Research and Policy Review*, pages 1–28, 2020.

[96] Nathan Eagle and Alex Sandy Pentland. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7):1057–1066, 2009.

[97] Luca Pappalardo, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. Returners and explorers dichotomy in human mobility. *Nature communications*, 6:8166, 2015.

[98] Luca Pappalardo and Filippo Simini. Data-driven generation of spatio-temporal routines in human mobility. *Data Mining and Knowledge Discovery*, 32(3):787–829, 2018.

[99] F. Giannotti, L. Pappalardo, D. Pedreschi, and D. Wang. *A Complexity Science Perspective on Human Mobility*, page 297–314. Cambridge University Press, 2013.

[100] Massimiliano Luca, Gianni Barlacchi, Bruno Lepri, and Luca Pappalardo. A survey on deep learning for human mobility, 2021.

[101] Vincent D Blondel, Adeline Decuyper, and Gautier Krings. A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(1):10, 2015.

[102] Gyan Ranjan, Hui Zang, Zhi-Li Zhang, and Jean Bolot. Are call detail records biased for sampling human mobility? *ACM SIGMOBILE Mobile Computing and Communications Review*, 16(3):33–44, 2012.

[103] Corina Iovan, Ana-Maria Olteanu-Raimond, Thomas Couronné, and Zbigniew Smoreda. Moving and calling: Mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies. In *Geographic information science at the heart of Europe*, pages 247–265. Springer, 2013.

[104] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *nature*, 453(7196):779, 2008.

[105] Albert-Laszlo Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207, 2005.

[106] Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi, and Alex Pentland. Once upon a crime: towards crime prediction from demographics and mobile data. In *Proc. of the 16th international conference on multimodal interaction*, pages 427–434. ACM, 2014.

[107] Jiahui Wu, Enrique Frias-Martinez, and Vanessa Frias-Martinez. Addressing underreporting to enhance fairness and accuracy in mobility-based crime prediction. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, pages 325–336, 2020.

[108] Emilio Ferrara, Pasquale De Meo, Salvatore Catanese, and Giacomo Fiumara. Detecting criminal organizations in mobile phone networks. *Expert Systems with Applications*, 41(13):5733–5750, 2014.

[109] Alina Sîrbu, Gennady Andrienko, Natalia Andrienko, Chiara Boldrini, Marco Conti, Fosca Giannotti, Riccardo Guidotti, Simone Bertoli, Jisu Kim, Cristina Ioana Muntean, et al. Human migration: the big data perspective. *International Journal of Data Science and Analytics*, pages 1–20, 2020.

[110] Guanghua Chi, Fengyang Lin, Guangqing Chi, and Joshua Blumenstock. A general approach to detecting migration events in digital trace data. *PloS one*, 15(10):e0239408, 2020.

[111] Kai Elgethun, Richard A Fenske, Michael G Yost, and Gary J Palcisko. Time-location analysis for exposure assessment studies of children using a novel global positioning system instrument. *Environmental Health Perspectives*, 111(1):115–122, 2003.

[112] Daniela Dias and Oxana Tchepel. Modelling of human exposure to air pollution in the urban environment: a GPS-based approach. *Environmental Science and Pollution Research*, 21(5):3558–3571, 2014.

[113] Johan Beekhuizen, Hans Kromhout, Anke Huss, and Roel Vermeulen. Performance of gps-devices for environmental exposure assessment. *Journal of Exposure Science and Environmental Epidemiology*, 23(5):498, 2013.

[114] Luca Pappalardo, Filippo Simini, Gianni Barlacchi, and Roberto Pellungrini. scikit-mobility: a python library for the analysis, generation and risk assessment of mobility data, 2019.

[115] Luca Pappalardo, Salvatore Rinzivillo, Zehui Qu, Dino Pedreschi, and Fosca Giannotti. Understanding the patterns of car travel. *The European Physical Journal Special Topics*, 215(1):61–73, Jan 2013.

[116] Paul Kelly, Patricia Krenn, Sylvia Titze, Peter Stopher, and Charlie Foster. Quantifying the difference between self-reported and global positioning systems-measured journey durations: a systematic review. *Transport Reviews*, 33(4):443–459, 2013.

[117] Henk Meurs and Rinus Haaijer. Spatial structure and mobility. *Transportation Research Part D: Transport and Environment*, 6(6):429–446, 2001.

[118] Melody Oliver, Hannah Badland, Suzanne Mavoa, Mitch J Duncan, and Scott Duncan. Combining GPS, GIS, and accelerometry: methodological issues in the assessment of location and intensity of travel behaviors. *Journal of Physical Activity and Health*, 7(1):102–108, 2010.

[119] Swann Arp Adams, Charles E Matthews, Cara B Ebbeling, Charity G Moore, Joan E Cunningham, Jeanette Fulton, and James R Hebert. The effect of social desirability and social approval on self-reports of physical activity. *American journal of epidemiology*, 161(4):389–398, 2005.

[120] Basile Chaix, Yan Kestens, Dustin T Duncan, Ruben Brondeel, Julie Méline, Tarik El Aarbaoui, Bruno Pannier, and Juan Merlo. A GPS-based methodology to analyze environment-health associations at the trip level: case-crossover analyses of built environments and walking. *American journal of epidemiology*, 184(8):579–589, 2016.

[121] Jacqueline Kerr, Scott Duncan, and Jasper Schipperjin. Using global positioning systems in health research: a practical approach to data collection and processing. *American journal of preventive medicine*, 41(5):532–540, 2011.

[122] Alyssa I Robinson, Fei Carnes, and Nicolas M Oreskovic. Spatial analysis of crime incidence and adolescent physical activity. *Preventive medicine*, 85:74–77, 2016.

[123] Andrea L DaViera, Amanda L Roy, Marbella Uriostegui, and Denise Fiesta. 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.

[124] Barak Ariel and Henry Partridge. Predictable policing: Measuring the crime control benefits of hotspots policing at bus stops. *Journal of Quantitative Criminology*, 33(4):809–833, 2017.

[125] Sharad Goel, Jake M Hofman, Sébastien Lahaie, David M Pennock, and Duncan J Watts. Predicting consumer behavior with web search. *Proceedings of the National academy of sciences*, 107(41):17486–17490, 2010.

[126] Hong Qi, Pedro Manrique, Daniela Johnson, Elvira Restrepo, and Neil F Johnson. Open source data reveals connection between online and on-street protest activity. *EPJ data science*, 5(1):18, 2016.

[127] Hong Qi, Pedro Manrique, Daniela Johnson, Elvira Restrepo, and Neil F Johnson. Association between volume and momentum of online searches and real-world collective unrest. *Results in physics*, 6:414–419, 2016.

[128] Ashley N Muchow and Catalina Amuedo-Dorantes. Immigration enforcement awareness and community engagement with police: Evidence from domestic violence calls in los angeles. *Journal of urban economics*, 117:103253, 2020.

[129] Frank Kleemann, G Günter Voß, and Kerstin Rieder. Un (der) paid innovators: The commercial utilization of consumer work through crowdsourcing. *Science, technology and innovation studies*, 4(1):5–26, 2008.

[130] Tara S Behrend, David J Sharek, Adam W Meade, and Eric N Wiebe. The viability of crowdsourcing for survey research. *Behavior research methods*, 43(3):800, 2011.

[131] Suzanne Goodney Lea, Elsa D'Silva, and Abhijith Asok. 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.

[132] Shriphani Palakodety, Ashiqur R. KhudaBukhsh, and Jaime G. Carbonell. Hope Speech Detection: A Computational Analysis of the Voice of Peace, 2019.

[133] Turgut Ozkan, John L Worrall, and Haley Zettler. Validating media-driven and crowdsourced police shooting data: a research note. *Journal of crime and justice*, 41(3):334–345, 2018.

[134] Shakila Khan Rumi, Wei Shao, and Flora D Salim. Realtime predictive patrolling and routing with mobility and emergency calls data. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 964–968, 2020.

[135] Ali Dehghan, Lloyd Montgomery, Maryi Arciniegas-Mendez, and Maria Ferman-Guerra. Predicting News Bias, 2016.

[136] Raleigh Clionadh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. Introducing ACLED-Armed Conflict Location and Event Data. *Journal of Peace Research*, 47(5):651–660, 2010.

[137] Jurgen Brauer and Charles H Anderton. Conflict and Peace Economics: Retrospective and Prospective Reflections on Concepts, Theories, and Data. *Defence and Peace Economics*, pages 1–23, 2020.

[138] Pamina Firchow and Roger Mac Ginty. Measuring peace: Comparability, commensurability, and complementarity using bottom-up indicators. *International Studies Review*, 19(1):6–27, 2017.

[139] Tobias Ide. Covid-19 and armed conflict. *World development*, 140:105355, 2021.

[140] KSM Tozammel Hossain, Shuyang Gao, Brendan Kennedy, Aram Galstyan, and Prem Natarajan. 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.

[141] Arabia Inform. Arabia Inform. `http://arabiainform.com/`, 1998.

[142] Fengcai Qiao, Pei Li, Xin Zhang, Zhaoyun Ding, Jiajun Cheng, and Hui Wang. Predicting social unrest events with hidden Markov models using GDELT. *Discrete Dynamics in Nature and Society*, 2017, 2017.

[143] Divyanshi Galla and James Burke. Predicting social unrest using GDELT. In *International Conference on Machine Learning and Data Mining in Pattern Recognition*, pages 103–116. Springer, 2018.

[144] Anadil Alsaqabi, Fatimah Aldhubayi, and Saleh Albahli. Using Machine Learning for Prediction of Factors Affecting Crimes in Saudi Arabia. In *Proc. of the 2019 International Conference on Big Data Engineering*, pages 57–62, 2019.

[145] Deepti Joshi, Sudeep Basnet, Hariharan Arunachalam, Leen-Kiat Soh, Ashok Samal, Shawn Ratcliff, and Regina Werum. SURGE: Social Unrest Reconnaissance GazEteer. In *Proc. of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–4, 2017.

[146] Qiao Fengcai, Deng Jinsheng, and Wei Li. 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)*, pages 176–182. IEEE, 2020.

[147] Swetha Keertipati, Bastin Tony Roy Savarimuthu, Maryam Purvis, and Martin Purvis. 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*, pages 33–40, 2014.

[148] Lihua Yuan, Changqing Song, Changxiu Cheng, Shi Shen, Xiaoqiang Chen, and Yuanhui Wang. 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.

[149] James E Yonamine. Predicting future levels of violence in afghanistan districts using gdelt. *Unpublished manuscript*, 2013.

[150] Merari González and Germán H Alférez. Application of data science to discover violence-related issues in iraq. *arXiv preprint arXiv:2006.07980*, 2020.

[151] Yaser Keneshloo, Jose Cadena, Gizem Korkmaz, and Naren Ramakrishnan. Detecting and forecasting domestic political crises: A graph-based approach. In *Proc. of the 2014 ACM conference on Web science*, pages 192–196, 2014.

[152] Mohammed N Ahmed, Gianni Barlacchi, Stefano Braghin, Francesco Calabrese, Michele Ferretti, Vincent Lonij, Rahul Nair, Rana Novack, Jurij Paraszczak, and Andeep S Toor. A Multi-Scale Approach to Data-Driven Mass Migration Analysis. In *SoGood@ ECML-PKDD*, 2016.

[153] Michel Beine, Luisito Bertinelli, Rana Cömertpay, Anastasia Litina, Jean-François Maystadt, and Benteng Zou. Refugee Mobility: Evidence from Phone Data in Turkey. In *Guide to Mobile Data Analytics in Refugee Scenarios*, pages 433–449. Springer, 2019.

[154] Simone Bertoli, Paolo Cintia, Fosca Giannotti, Etienne Madinier, Caglar Ozden, Michael Packard, Dino Pedreschi, Hillel Rapoport, Alina Sîrbu, and Biagio Speciale. Integration of Syrian refugees: insights from D4R, media events and housing market data. In *Guide to Mobile Data Analytics in Refugee Scenarios*, pages 179–199. Springer, 2019.

[155] Marcello Carammia, Stefano Maria Iacus, and Taddy Wilkin. Forecasting asylum applications in the european union with machine learning and data at scale. *arXiv preprint arXiv:2011.04348*, 2020.

[156] The Institute for Economics and Peace. VISION of HUMANITY. `http://visionofhumanity.org//`, 2017.

[157] Håvard Hegre, Marie Allansson, Matthias Basedau, Michael Colaresi, Mihai Croicu, Hanne Fjelde, Frederick Hoyles, Lisa Hultman, Stina Högbladh, Remco Jansen, et al. Views: a political violence early-warning system. *Journal of peace research*, 56(2):155–174, 2019.

[158] Rebecca H Best, Christine Carpino, and Mark JC Crescenzi. An analysis of the TABARI coding system. *Conflict Management and Peace Science*, 30(4):335–348, 2013.

[159] Philip A Schrodt. Cameo: Conflict and mediation event observations event and actor codebook. *Pennsylvania State University*, 2012.

[160] Kalev Leetaru and Philip A Schrodt. Gdelt: Global data on events, location, and tone, 1979-2012. In *ISA annual convention*, volume 2, pages 1–49. Citeseer, 2013.

[161] Sérgio Fernandes and Jorge Bernardino. What is bigquery? In *Proceedings of the 19th International Database Engineering and Applications Symposium*, IDEAS '15, page 202–203, New York, NY, USA, 2015. Association for Computing Machinery.

[162] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

[163] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction.* Springer Science and Business Media, 2009.

[164] Mariette Awad and Rahul Khanna. Support vector regression. In *Efficient learning machines*, pages 67–80. Springer, 2015.

[165] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[166] Erik Štrumbelj and Igor Kononenko. Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems*, 41(3):647–665, 2014.

[167] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " 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*, pages 1135–1144, 2016.

[168] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Anchors: High-precision model-agnostic explanations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[169] Ramalingam Shanmugam. Introduction to time series and forecasting, 1997.

[170] Rob J Hyndman and Anne B Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.

[171] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice.* OTexts, 2018.

[172] Thomas L Zeller, Lawrence M Metzger, et al. Good bye traditional budgeting, hello rolling forecast: has the time come? *American Journal of Business Education (AJBE)*, 6(3):299–310, 2013.

[173] Philip A Schrodt. Forecasting political conflict in asia using latent dirichlet allocation models. In *Annual Meeting of the European Political Science Association, Dublin*, 2011.

[174] Alboukadel Kassambara. *Machine Learning Essentials: Practical Guide in R*. sthda, 2018.

[175] Arnaud De Myttenaere, Boris Golden, Bénédicte Le Grand, and Fabrice Rossi. Mean absolute percentage error for regression models. *Neurocomputing*, 192:38–48, 2016.

[176] David A Swanson. On the relationship among values of the same summary measure of error when used across multiple characteristics at the same point in time: An examination of malpe and mape. *Review of Economics and Finance*, 5(1), 2015.

[177] Haldun Akoglu. User's guide to correlation coefficients. *Turkish journal of emergency medicine*, 18(3):91–93, 2018.

[178] Andrew F Cooper and Ramesh Thakur. *The group of twenty (G20)*. Routledge, 2013.

[179] Meghan E Hollis, Sharece Downey, Alex del Carmen, and Rhonda R Dobbs. The relationship between media portrayals and crime: perceptions of fear of crime among citizens. *Crime prevention and community safety*, 19(1):46–60, 2017.

[180] Paul Cairney and Richard Kwiatkowski. How to communicate effectively with policymakers: combine insights from psychology and policy studies. *Palgrave Communications*, 3(1):1–8, 2017.

[181] Michael Colaresi and Zuhaib Mahmood. Do the robot: Lessons from machine learning to improve conflict forecasting. *Journal of Peace Research*, 54(2):193–214, 2017.

[182] Sanjana Hattotuwa. Big data and peacebuilding. *Stability: International Journal of Security and Development*, 2(3), 2013.

[183] James Honaker and Gary King. What to do about missing values in time-series cross-section data. *American journal of political science*, 54(2):561–581, 2010.

[184] Stefano M Iacus and Nakahiro Yoshida. Simulation and inference for stochastic processes with yuima. *A comprehensive R framework for SDEs and other stochastic processes. Use R*, 2018.

[185] NobelPrize.org. Nobel Prize Outreach AB 2022. The Nobel Peace Prize for 2020. https://www.nobelprize.org/prizes/peace/2020/press-release/, 2022.

[186] World Food Programme. Food fortification: An effective and safe way to fight micronutrient malnutrition and its consequences. https://docs.wfp.org/api/documents/WFP-0000139908/download/?_ga=2.234963844.44014634.1657183207-982293757.1656686435, 2022.

[187] World Health Organisation. Measuring the age-friendliness of cities. https://apps.who.int/iris/bitstream/handle/10665/203830/9789241509695_eng.pdf, 2015.

[188] Coulton, Claudia J and Goerge, Robert and Putnam-Hornstein, Emily and de Haan, Benjamin. Harnessing big data for social good: A grand challenge for social work. *Cleveland: American Academy of Social Work and Social Welfare*, pages 1–20, 2015.

[189] Lepri, Bruno and Staiano, Jacopo and Sangokoya, David and Letouzé, Emmanuel and Oliver, Nuria. The tyranny of data? the bright and dark sides of data-driven decision-making for social good. In *Transparent data mining for big and small data*, pages 3–24. Springer, 2017.

[190] Luciano Floridi and Mariarosaria Taddeo. What is data ethics?, 2016. The Royal Society.

[191] David J Hand. Aspects of data ethics in a changing world: where are we now? *Big data*, 6(3):176–190, 2018.

[192] Branka Panic. Peace in the Age of Artificial Intelligence - The Promise and Perils. https://tinyurl.com/3hbupea9, 2020.

[193] Haewoon Kwak and Jisun An. Two tales of the world: Comparison of widely used world news datasets gdelt and eventregistry. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 10, pages 619–622, 2016.

[194] Tsan-Kuo Chang. All countries not created equal to be news: World system and international communication. *Communication research*, 25(5):528–563, 1998.

Appendices

Appendix A

GDELT data description

To better understand the data over the countries, we plot the distribution of the monthly No. events for six countries. Particularly, Figure A.1 presents the distribution of the total monthly No. events for the United States (a), Yemen (b), the United Kingdom (c), Saudi Arabia (d), China (e), and Chile (f). We notice that the United States has a much higher monthly No. events compared to the other countries. The second country with the highest No. events is the United Kingdom. The third and fourth countries with the highest No. events are Saudi Arabia and China. Last, Yemen and Chile have the lowest monthly No. events. Previous research has studied the reasons behind this country difference in the No. events. Particularly, Kwak and An [193] have demonstrated that the countries with the highest No. events are the English-speaking countries, explaining the higher monthly No. events for the United States and the United Kingdom in our data. Kwak and An [193] discuss that this is additional evidence that journalists have a concept of "world hierarchy" [194] in determining which is more important foreign news.

In addition, it is interesting to explore the share of each event category over all news. For example, Table A.1 presents the five GDELT variables with the largest share of No. events for the United States (a), Yemen (b), the United Kingdom (c), Saudi Arabia (d), China (e), and Chile (f) over the whole dataset, i.e., from March 2008 to March 2020. We notice that for all six countries among the GDELT variables with the highest share of No. events are "Make statement", "Make a visit" and "Host a visit" variables. We also observe that although these variables are between the variables with the highest share of No. events for Yemen, contrary to the other countries the variable with the highest share of No. events is "Use conventional military force".

Based on the observations mentioned above, we choose the variables 'Make statement' and 'Use conventional military force' to explore how they are distributed over the months for the six countries presented above. Figure A.2 shows the distributions of the monthly

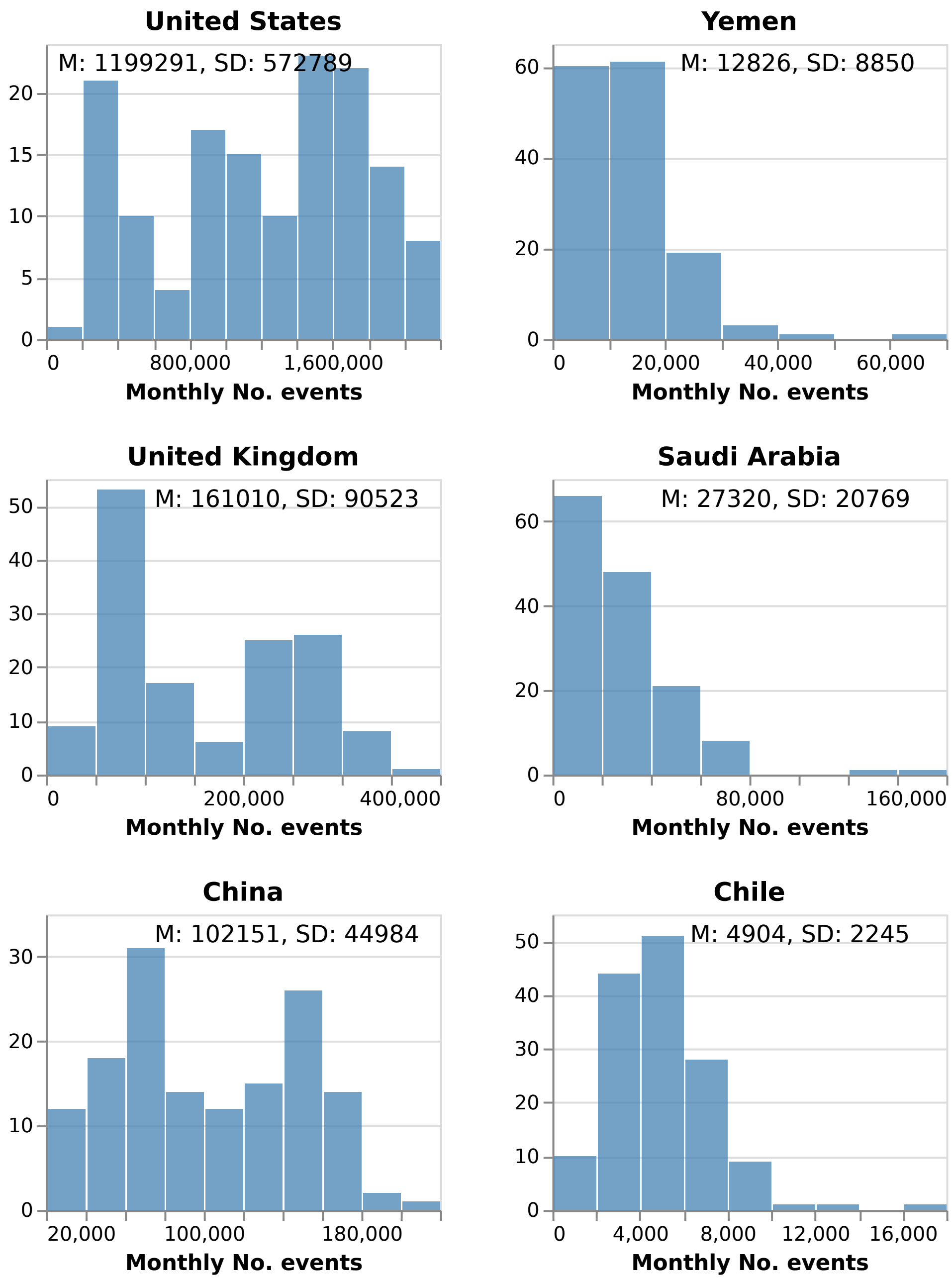


Fig. A.1 The distribution of the total monthly No. events (news) for 6 countries. The distribution of the total monthly No. events (news) for the United States (a), Yemen (b), the United Kingdom (c), Saudi Arabia (d), China (e), and Chile (f).

Table A.1 The five GDELT variables with the largest share of the number of news for the United States (a), Yemen (b), the United Kingdom (c), Saudi Arabia (d), China (e), and Chile (f) over the whole dataset, i.e., from March 2008 to March 2020.

Event code	Event Category	Share of news
010	Make statement	7.73 %
042	Make a visit	7.52 %
043	Host a visit	6.97 %
020	Make an appeal or request	6.61%
051	Praise or endorse	5.80 %

(a) United States

Event code	Event Category	Share of news
190	Use conventional military force	13%
010	Make statement	7.17%
042	Make a visit	5.83%
040	Consult	5.25 %
043	Host a visit	5.24 %

(b) Yemen

Event code	Event Category	Share of news
042	Make a visit	8.16 %
010	Make statement	7.86 %
043	Host a visit	7.57 %
020	Appeal	6.47 %
051	Praise or endorse	5.80 %

(c) United Kingdom

Event code	Event Category	Share of news
042	Make a visit	7.68%
043	Host a visit	7.12 %
040	Consult	6.76%
010	Make statement	6.63%
051	Praise or endorse	5.39%

(d) Saudi Arabia

Event code	Event Category	Share of news
042	Make a visit	8.68%
043	Host a visit	7.87%
010	Make statement	7.64%
040	Consult	6.51%
020	Appeal	5.40%

(e) China

Event code	Event Category	Share of news
010	Make statement	9.91%
042	Make a visit	9.18%
043	Host a visit	8.39%
040	Consult	6.05%
020	Appeal	5.58%

(f) Chile

No. events related to 'Make statement' for the United States (a), Yemen (b), the United Kingdom (c), Saudi Arabia (d), China (e), and Chile (f). We observe that the mean No. events on 'Make statement' for the United States is 92687, and the standard deviation is 44047. On the contrary, Chile's mean No. events on 'Make statement' is 486, and the standard deviation is 215. Furthermore, Figure A.3 presents the distributions of the monthly No. events related to 'Use conventional military force' for the United States (a), Yemen (b), the United Kingdom (c), Saudi Arabia (d), China (e), and Chile (f). We observe that the mean No. events on 'Use conventional military force' for the United States is 44597, and the standard deviation is 20709. On the contrary, Chile's mean No. events on 'Use conventional military force' is 149, and the standard deviation is 83.

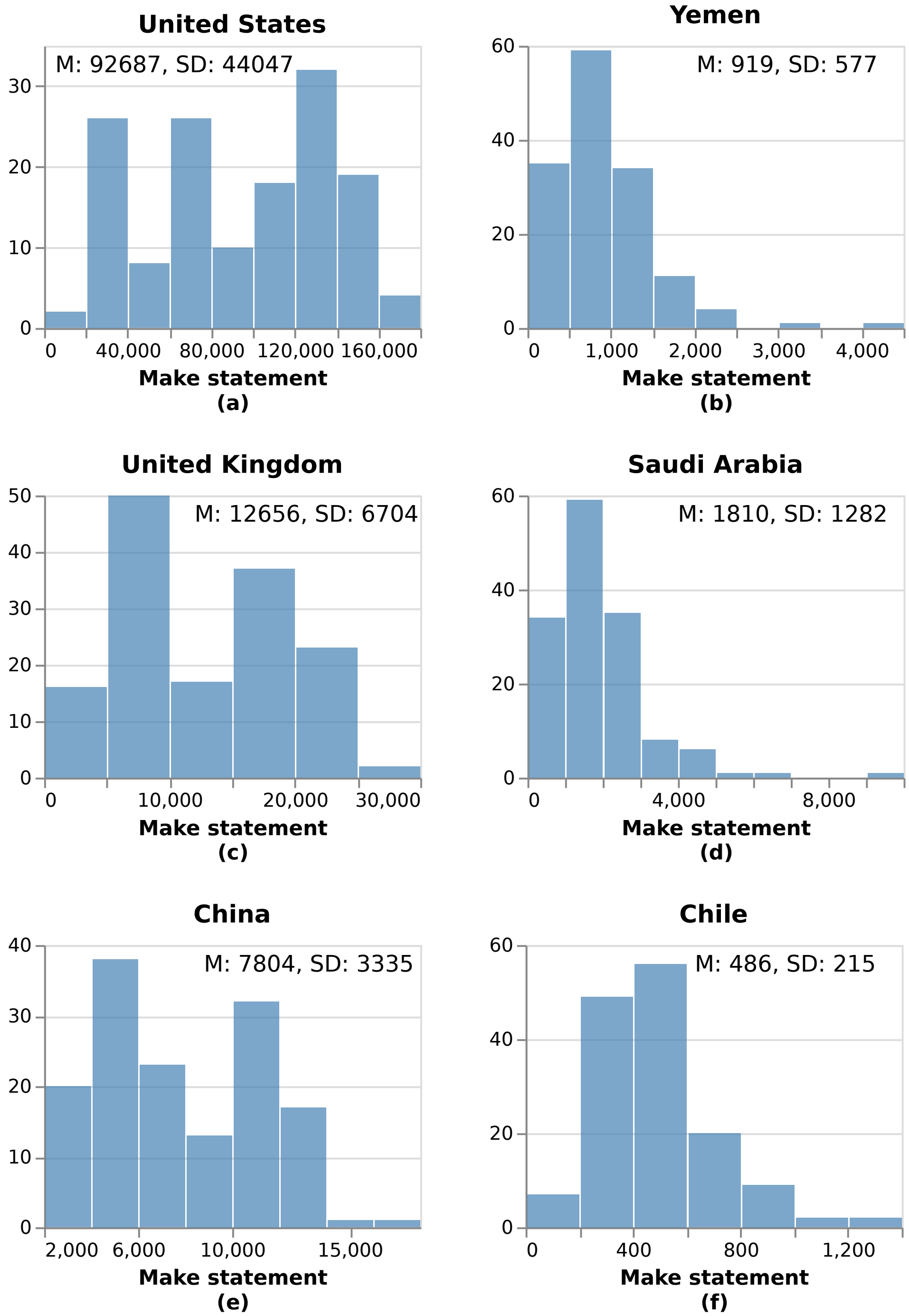


Fig. A.2 The distribution of the monthly No. events (news) related to 'Make statement' for 6 countries. The distribution of the monthly No. events (news) related to 'Make statement' for the United States (a), Yemen (b), the United Kingdom (c), Saudi Arabia (d), China (e), and Chile (f).

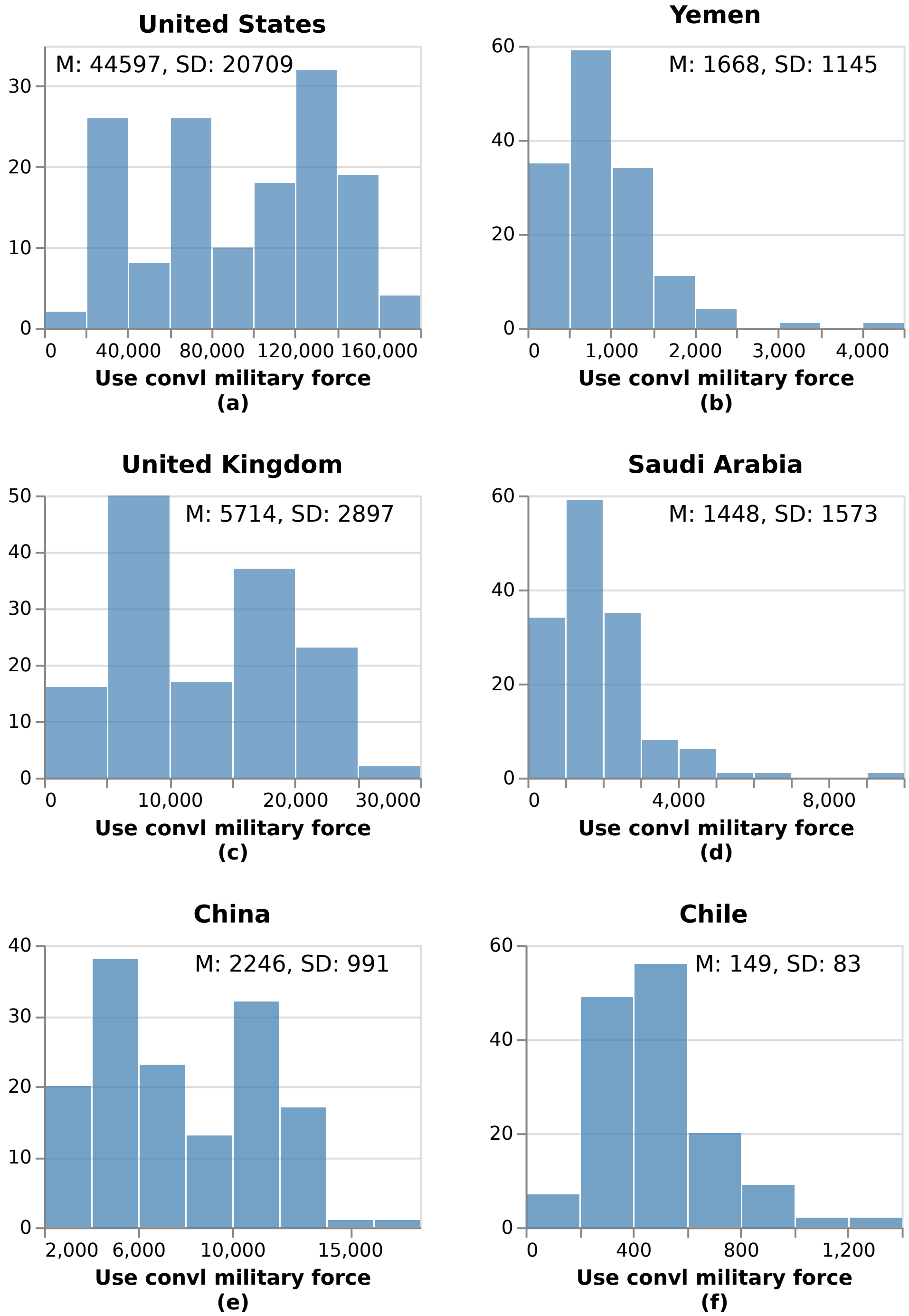


Fig. A.3 The distribution of the monthly No. events (news) related to 'Use conventional military force' for 6 countries. The distribution of the monthly No. events (news) related to 'Use conventional military force' for the United States (a), Yemen (b), the United Kingdom (c), Saudi Arabia (d), China (e), and Chile (f).

Appendix B

Linear models

The median Pearson Correlation for the Linear models for the 1-month-ahead predictions is 0.069, and the median MAPE is 39.273. These results demonstrate that Linear models show lower performance not only from the XGBoost models (0.521, and 1.593, respectively), but also from the Elastic Net models (0.327, and 1.997, respectively), already from the 1-month-ahead predictions.

Appendix C

Comparing the variable importances

In Section 5.2 we demonstrate how SHAP method can significantly contribute to the interpretability of the model results and the explanation of their behaviour as well as of the large estimation errors produced. Local interpretability stands between its advantages and it cannot be provided by the Gain method. However, the global variable importance can be provided from both methods. Consequently, it would be interesting to make a comparison between the variable importance provided by Gain method (Section 5.1) and the variable importance provided by SHAP method (Section 5.2) for the same country models and training.

For example, Figure C.1 presents the most important variables for the United States as calculated between the training period from April 2014 to March 2020. Comparing Figure C.1 with Figure 5.17 we observe we observe some similarities and some changes between the plots. Particularly, we observe that the majority of the most important (top 12) variables are the same, and the profile of the country remains the same, i.e., a country profile of a strong player in the military, socio-economic, and political foreground. In addition, we observe that the first most important variable as calculated by both methods is "Employ aerial weapons". This variable is considerably more important compared to the rest of the variables. However, the order of rest of the most important variables changes (the red numbers correspond to the variable importance order given from SHAP method). The Gain method places the variable "Reject" as the second-most important variable, although it might make a lower contribution to prediction. Indeed SHAP method places it further down the list, as seventh-most important variable. Similarly, the Gain method places the variable "Use as human shield" although SHAP method does not include it in the top 12 most important variables. This behaviour is due to the Gain method averaging the contribution of variables across all instances they appear in the trees. This dilutes the calculated importance of certain variables, which are used as a splitter many times although not always improving the model by a large amount each time they are used. On the contrary, SHAP method measures the influence of a variable by comparing model predictions with and without the variable, which can make the model and potential large errors easier to interpret.

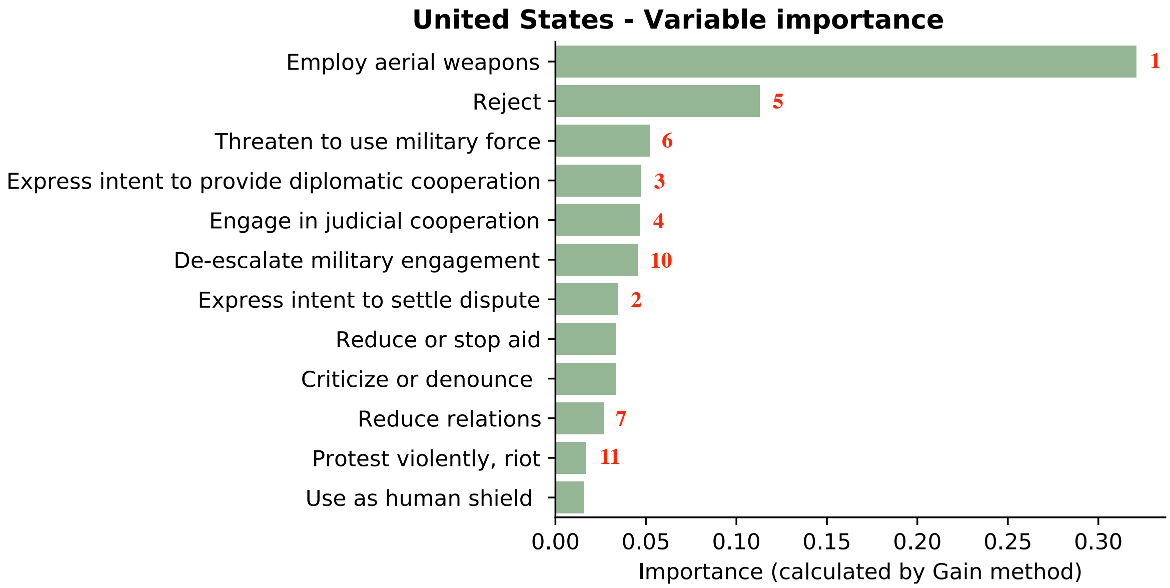


Fig. C.1 Variable importance for the United States. Variable importance for the United States as calculated by the Gain method with a training period from April 2014 to March 2020. Comparing with SHAP method, we observe some similarities and some changes. For example, the first most important variable remains the same, whereas the rest of the variables are ordered differently, and others are not included in the top12 at all (the red numbers correspond to the variable importance order given from SHAP method).

Appendix D

Adding lagged GPI to XGBoost

As explained in Section 3.3 we have not used autoregressive models due to the nature of the linearly interpolated GPI. However, an alternative to our models, which use as input only GDELT variables, i.e., external variables, could be to add as input lagged GPI time-series. The objective would be to incorporate feedback effect to the model over time. We select the algorithm with the highest performance, i.e., XGBoost algorithm, to conduct a preliminary analysis adding to the model predictors a 12-months lagged GPI. For example, in order to predict the monthly GPI values of 2018, we use the monthly GPI values of 2017. As expected, results show the XGBoost models without the lagged GPI demonstrate a lower performance comparing to XGBoost models with the lagged GPI. Figure D.1 presents the MAPE for all XGBoost models without or with the lagged GPI. The boxplots represent the distribution of the aforementioned performance indicator for all country models. The plots' data points correspond to each country model. As expected, the plots demonstrate that, overall, XGBoost models without (w/o) the lagged GPI as independent variable have a lower performance than the XGBoost models with (w/) the lagged GPI as independent variable. For example, median MAPE for the 1-month-ahead predictions for the XGBoost models w/ is 1.33, whereas for the XGBoost models w/o is 1.59, i.e., 19.5% higher.

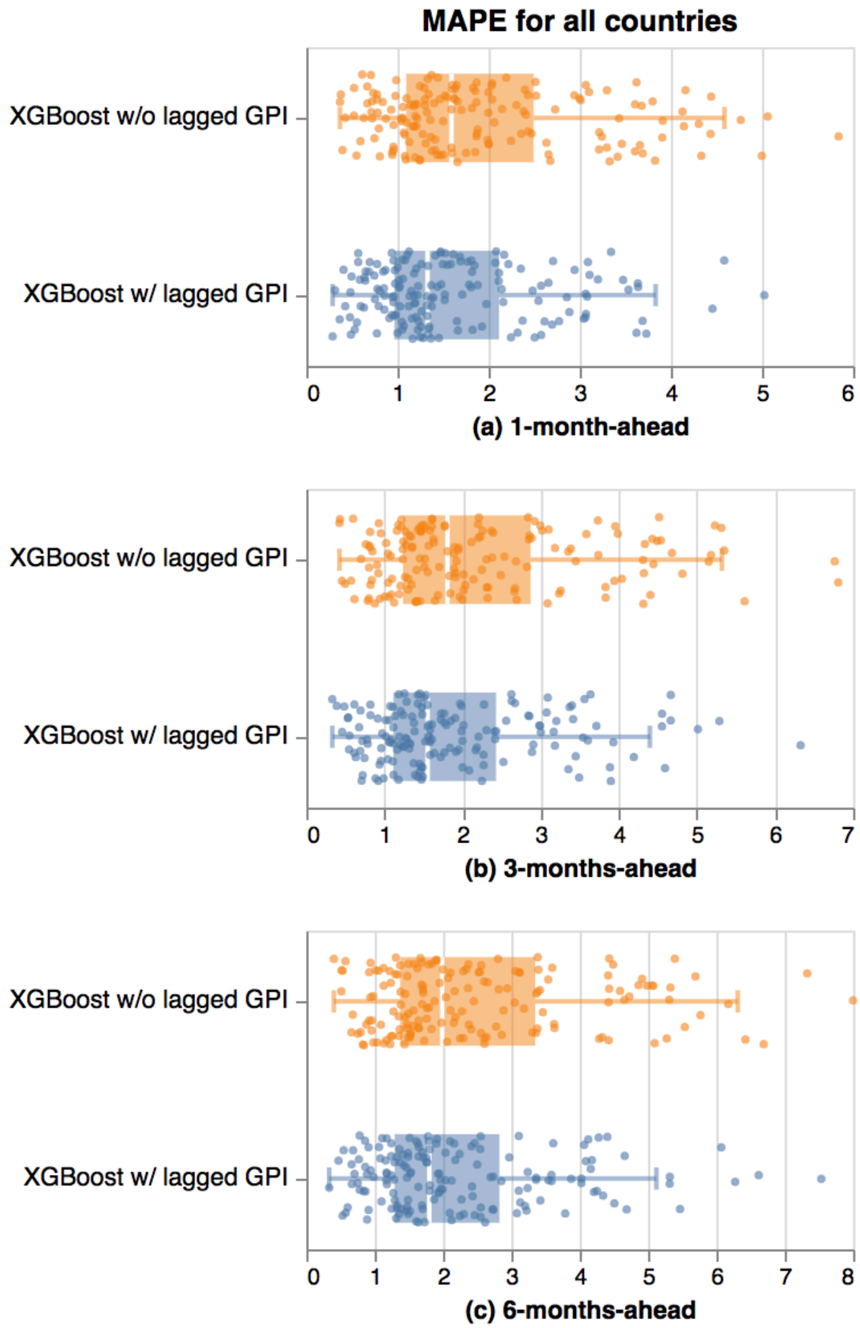


Fig. D.1 MAPE for all XGBoost models without or with the lagged GPI. MAPE between the real and the predicted 1-, 3-, and 6-months-ahead GPI values at a country level, for XGBoost models without (w/o) and with (w/) the lagged GPI as independent variable. The boxplots represent the distribution of the aforementioned performance indicator for all country models. The plots' data points correspond to each country model. Overall, XGBoost without the lagged GPI has a lower performance comparing to XGBoost with lagged GPI.

Appendix E

Adding salience to XGBoost

The monthly count of all news for a country to the total count of all monthly news for all countries could be added to the models as an independent input variable. This variable could contribute to the models' performance by giving a sign of whether the global news is concentrated to a specific country for a certain month. We call this variable salience, and we train all country models, including salience as an independent variable. We select the algorithm with the highest performance, i.e., XGBoost algorithm, to conduct , to conduct a preliminary analysis. Results show the XGBoost models without (w/o) the salience demonstrate almost the same performance with the XGBoost models with (w/) the salience. Figure E.1 presents the MAPE for all XGBoost models w/o or w/ the salience. The boxplots represent the distribution of the aforementioned performance indicator for all country models. The plots' data points correspond to each country model. The plots demonstrate that, overall, XGBoost models without (w/o) the salience as independent variable perform almost the same with the XGBoost models with (w/) the salience. For example, median MAPE for the 1-month-ahead predictions for the XGBoost models w/ is 1.33, whereas for the XGBoost models w/o is 1.59.

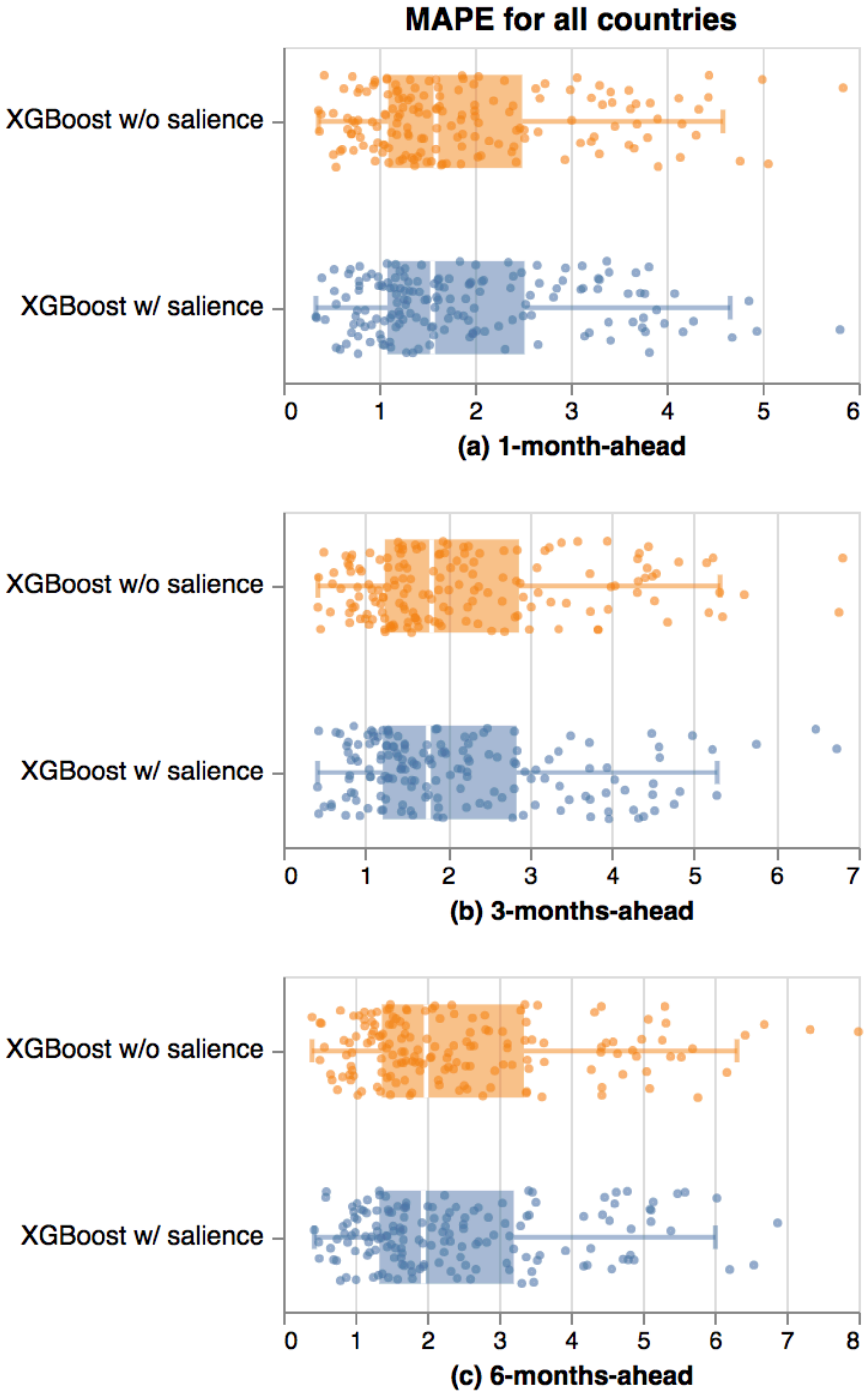


Fig. E.1 MAPE for all XGBoost models without or with the salience. MAPE between the real and the predicted 1-, 3-, and 6-months-ahead GPI values at a country level, for XGBoost models without (w/o) and with (w/) the salience independent variable. The boxplots represent the distribution of the aforementioned performance indicator for all country models. The plots' data points correspond to each country model. Overall, XGBoost without or with the salience have a similar performance.