An Ecosystem of Applications for Modeling Political Violence

Aline Bessa
New York University
aline.bessa@nyu.edu

Sonia Castelo
New York University
s.castelo@nyu.edu

Rémi Rampin
New York University
remi.rampin@nyu.edu

Aécio Santos
New York University
aecio.santos@nyu.edu

Mike Shoemate
Harvard University
shoematem@seas.harvard.edu

Vito D'Orazio
University of Texas at Dallas
dorazio@utdallas.edu

Juliana Freire
New York University
juliana.freire@nyu.edu

ABSTRACT

Conflict researchers face many challenges, including (1) how to model conflicts, (2) how to measure them, (3) how to manage their spatio-temporal character, and (4) how to handle a potential abundance of information and explanation. In this paper, we describe an ecosystem of tools designed for use by subject matter experts that addresses these challenges. Three case studies show workflows that are facilitated by this ecosystem.

CCS CONCEPTS

• **Information systems** → **Information systems applications**;
• **Applied computing** → **Law, social and behavioral sciences**;
• **Human-centered computing** → *Visualization.*

KEYWORDS

Conflict Modeling; Political Violence; Applied Machine Learning

ACM Reference Format:
Aline Bessa, Sonia Castelo, Rémi Rampin, Aécio Santos, Mike Shoemate, Vito D'Orazio, and Juliana Freire. 2021. An Ecosystem of Applications for Modeling Political Violence. In *Proceedings of the 2021 International Conference on Management of Data (SIGMOD '21), June 20–25, 2021, Virtual Event, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3448016.3457235

1 INTRODUCTION

Computational methods are increasingly used to study all types of conflict, including political violence, instability, and social protest. Several projects now existfor these types of events, including the Violence Early Warning System, which is designed to predict sub-national political violence in Africa [22], and CoupCast, which forecasts the risk of coup for each country in the world every month [19]. Another example concerns the United Nations Office for the Coordination of Humanitarian Affair. To assist in the event of food insecurity, outbreaks of disease, and natural disasters, it prepares "anticipatory action frameworks" that are largely built on predictive models [36]. Computational methods are also used in the field of event data, with applications to the study of conflict in both computer and political sciences [12, 35, 40]. Thus, there is a broad set of researchers who can benefit from computational tools that streamline the study of conflicts, including government officials at places such as the Bureau of Conflict and Stabilization Operations at the United States State Department; analysts at think tanks such as RAND [13] and One Earth Future [18]; researchers at the United Nations and other governmental organizations; and researchers affiliated with projects such as the Armed Conflict Location & Event Data project [33].

While studying and analyzing conflicts, researchers face many challenges that our ecosystem has been built to address. Chief among them are (1) how to model conflicts, (2) how to measure them, (3) how to manage their spatio-temporal character, and (4) how to handle a potential abundance of information and explanation.

In an early application of a neural network model [27] to conflict prediction, [4] stated the *modeling challenge* succinctly: "International conflict is a rare event, and the processes that drive it where it is more common are likely to be very different from those elsewhere. As a result, many qualitative researchers expect the relationships to be highly nonlinear, massively interactive, and heavily context dependent or contingent" (p. 22). Although their application is specific to international conflict, it applies to the entire class of political violence, instability, and social protest.

The *measurement challenge* is less specific to conflict–in fact, it is an issue for much of social research. Many of the concepts we study, whether it be conflict and violence or something else, have no objectively true quantification. In practice, we measure them despite well-known conceptual and operational difficulties [1]. As a result, we do not see *a single dataset* on civil war, or on social protest, but rather *many datasets* on these topics. Each dataset has its own conceptual and operational definitions. Consequently, applying multiple datasets to the same problem has several benefits, but most importantly it helps to establish the robustness of an empirical finding by reducing sensitivity in the model [23].

The third challenge concerns the *spatio-temporal nature of conflict datasets*: most conflicts materialize as spatio-temporal events,

SIGMOD '21, June 20–25, 2021, Virtual Event, China

ACM ISBN 978-1-4503-8343-1/21/06…$15.00
https://doi.org/10.1145/3448016.3457235

and they might occur in a variety of spatially and temporally disaggregate levels. Determining what level is ideal for a certain analysis may not be trivial, and if more than one dataset needs to be analyzed in tandem, converting them to a compatible spatio-temporal granularity can be costly.

Last but not least, while the abundance of datasets opens the opportunity for new and more robust analyses, *finding data relevant to a given question is difficult*. Political violence, instability, and social protest are complex issues with no simple explanation. [6] and [20] provide evidence for this statement with respect to civil wars and terrorism. As available data expands, and the number of explanations for these events expands along with it, researchers are left with the task of sorting through large quantities of information to identify what is meaningful for their problem. At its core, this is a problem that arises from an increase of available resources.

In this paper, we describe an ecosystem of tools for spatio-temporal conflict data management and usability, predictive modeling, novelty detection, and visualization. This ecosystem is designed for use by subject matter experts to study all types of conflict, including political violence and social protest. It addresses the modeling, measurement, spatio-temporal, and information exploration challenges detailed above. More specifically, it is built for advanced but simple exploratory visual analysis [3], exploratory model building [24, 31], and engineered to generate insights, assess hypotheses, and construct machine learning models. Since much of what the ecosystem is designed to facilitate is exploratory, it depends on what the subject matter expert considers relevant. Therefore, we do not make assumptions about the relevance of any dataset, nor do we presuppose any notion of a correct answer, or any kind of predefined, linear workflow. Rather, we present an ecosystem to help researchers find and process relevant data, detect and prune useless information, refine the development of hypotheses, and employ advanced models for the study of conflict.

Our main contributions can be summarized as follows: (1) we assembled an ecosystem of tools to empower conflict researchers to carry out complex computational workflows, streamlining data exploration and discovery – this ecosystem supports the spatio-temporal character of datasets, which is essential for conflict analyses and modeling; and (2) we describe case studies that demonstrate the effectiveness of this tool ecosystem.

Related Work. There are various data analytics frameworks, such as Tableau [39] and R [32], that are used to analyze conflict data. However, these tools are general purpose and not domain specific, requiring extensive skills from end users. Moreover, common workflows include different tasks that range from searching for new data to exploring model predictions. While each task can be completed independently with tools such as R and Python, there are extensive synergies in an integrated system. At the other end of the spectrum are systems specifically for conflict data, promoting their collection, curation, and visualization. xSub [41], for example, includes data collections on subnational conflict for over 150 countries, facilitating their aggregation into different spatio-temporal units and allowing for faster dataset comparisons. ACLED [34] collects data on different types of political violence and protest events, and has a dashboard for exploration and visualization. However, systems such as these are basically tools for data accessibility and exploration, and not domain specific research platforms. The set of tools that we present in this paper works as a first step towards filling this gap. The tools fit neatly into the research fields of visual analytics [14], dataset search engines [10, 29], and spatio-temporal relationship mining [2, 5]. They are tailored for conflict research and the kinds of data commonly found in this field.

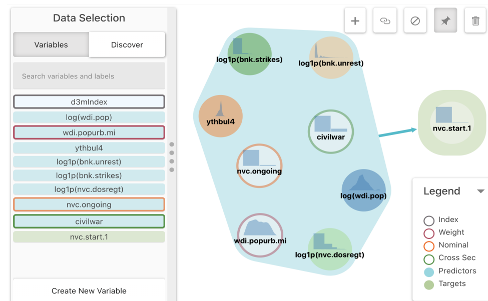

Figure 1: Data exploration and modeling component of TwoRavens. The distribution of each feature is shown, along with a directed graph for model specification.

2 TOOLS FOR CONFLICT EXPLORATION

In this section, we give an overview of the tools we developed to address the challenges described in the introduction.

2.1 TwoRavens

TwoRavens is a Web application for statistical analysis and data exploration [16, 21]. Using a framework called *human guided machine learning*, researchers explore data visualizations, review recommendations about interesting relationships between variables, manipulate the data, and specify models as machine learning tasks. The interface is illustrated in Figure 1, showing a group of predictors and a single target feature. Information about features, including temporal or geospatial tags and summary statistics, are displayed. The machine learning tasks are solved with the backend, AutoML component. Through a common interface, TwoRavens exposes multiple AutoML backends, including AlphaD3M [17], TPOT [25], and an original TwoRavens system. As solutions from AutoML backends stream back, they are cast over interpretation tools that allow users to compare metrics and explore predictions, among other features. The process is iterative and researchers can revise their initial specifications, incorporate new data, and glean new insights.

2.2 TwoRavens for Event Data

TwoRavens for Event Data [16] is designed for structuring raw event data into time-series formats. Researchers can browse event datasets including ICEWS [8] and SPEED [28], can construct queries to select types of events and sets of actors, and view and download resulting time-series. Data may be exported to TwoRavens for further analysis.

Event Data. The unit of observation in event datasets is the event, where typically each row represents an event of interest. Events generally have spatial and temporal attributes, along with source/target actors and classification of action types. Event datasets are often used to model the relationships between actors, which represent

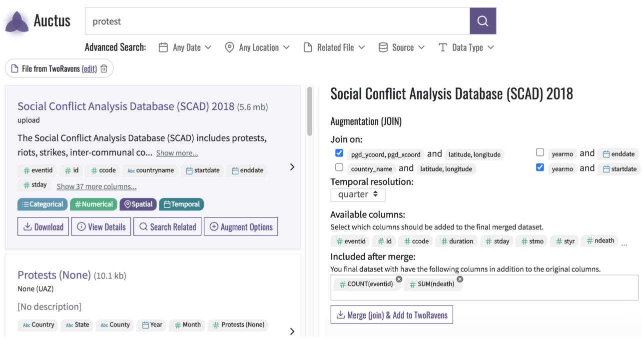

Figure 2: Auctus: The search results that match the keyword 'protest' and the grid dataset are displayed. The SCAD dataset is selected and the augmentation is set up to perform a spatial-temporal join.

entities at various levels of abstraction, including governments, organizations, rebel groups, and prominent individuals. Researchers can construct expressive subset/aggregation constraints via configurable dyads that link groupings of actors together.

Queries. Due to the high frequency of events, event datasets are cumbersome to download and too large to load into memory. Collections such as ICEWS update regularly so data for near-real-time modeling needs to be updated frequently. TwoRavens for Event Data addresses these issues by providing an accessible interface for researchers to build up subset/aggregation queries, and offload query execution to a centrally-hosted MongoDB database. The tool has a single-click transition to move the data into TwoRavens, where researchers can continue their analysis, or into Auctus, where it becomes discoverable for others.

2.3 Auctus Dataset Search Engine

Increasingly, conflict researchers use machine learning to forecast conflict [9], but there is an inherent limitation in this approach: *machine learning models are as good as the training data they use.* Consequently, a question arises: *Given a machine learning task and an initial dataset as input data, how can we find additional, relevant data to build a better model?* Auctus [30], a tool proposed to address this issue, is a dataset search engine tailored for data augmentation. Different from existing dataset search engines, Auctus not only indexes the content of datasets from a variety of sources (including the Web) but also infers consistent metadata that is later used to retrieve datasets that can be joined or appended to a user's data. Beside the Web interface (see Figure 2), Auctus can be accessed through a REST API [37].

2.4 PODS

Since conflicts often materialize as outlying values in datasets (e.g., abnormal peaks in the monthly numbers of violent events), having an automatic mechanism to discriminate data noise from outliers that correspond to events is desirable, especially when the analysis involves a large number of datasets and the manual outlier inspection becomes infeasible. In this context, a useful tool to understand the nature of data outliers is PODS (*Predictable Outliers in DatatrendS*) [5]. Given a collection of temporal datasets, PODS derives explanations for outliers by identifying meaningful relationships between them. It proposes a statistical criterion to determine whether outlier relationships are meaningful, and it does so by verifying if such relationships could have been predicted from non-outliers.

2.5 Novel Integrations

With a high level of compatibility between components' inputs and outputs, our system allows for integration across its tools with little to no programming effort. It is possible and practical, for example, to search for datasets within a region and a time period with Auctus, restrict the search to relationships involving neighboring countries with PODS, and focus on specific spatio-temporal slices for the creation of models with TwoRavens. The integrated architecture of our ecosystem takes the spatio-temporal character of conflict datasets into account, which is generally a challenge for data management. The case studies in Section 3 illustrate how this challenge is addressed. Since users do not need to spend time converting data or implementing scripts to use different tools in a combined fashion, the total of our ecosystem is greater than the sum of its parts.

3 CASE STUDIES

In this section we describe case studies that demonstrate how the ecosystem can be used.

3.1 Group Comparisons

One workflow that researchers may use to study political violence involves comparing different violent extremist organizations (VEOs). For example, [7] compares recruitment strategies of the Islamic State in Iraq and Syria, Al-Qaeda, and the Provisional Irish Republican Army. Comparisons such as this help identify patterns and facilitate the construction of theories of group behavior. We begin with the TwoRavens Event Data system and compare two VEOs in Africa: Al-Shabaab (Somalia & East Africa) and Boko Haram (Nigeria & West Africa). Our research question is: *do these VEOs engage in similar types of armed conflict?* First, we construct time series plots to understand the different types of violence employed by each group. To do so, we perform the following operations: (1) select the ACLED data [34]; (2) group actors, combining all known names for each VEO; (3) subset the data to only include events with these actor groups; (4) aggregate to monthly counts of conflict events; and (5) plot the resulting time series.

The resulting time series for Al-Shabaab is shown in Figure 3. A trend emerges, showing a high number of *battle* events (blue) followed by a sharp drop in 2012, and an increase in 2018. When *battle* events are low, *violence against civilians* (red) tends to increase. This exploration reveals a potentially interesting dependency where one type of conflict drives another. A hypothesis might be that as groups weaken in strength, and thus can no longer engage in direct battle, they increasingly resort to violence against civilians to make their political statements. However, the comparable time series for Boko Haram does not provide additional support for this hypothesis, as *battles* and *violence against civilians* tend to trend together. To gather more insight on how group violence compares, we can broaden the set of groups included in the analysis. Or, since TwoRavens Event Data contains many event datasets, researchers can continue to explore Al-Shabaab and Boko Haram using different conceptualizations and operationalizations of violence. The analyst may also transition to TwoRavens, and use AutoML to forecast trends in conflict events, or join new data from the Auctus system.

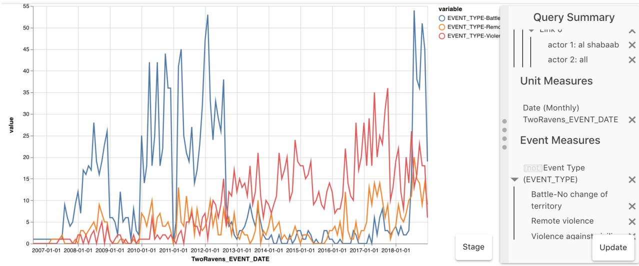

Figure 3: Time series shows three types of conflict events involving Al-Shabaab: battles (blue), remote violence (orange), and violence against civilians (red).

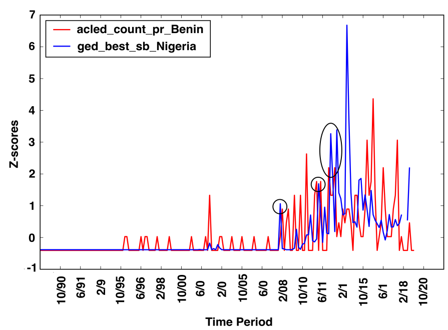

Figure 4: PODS detected a meaningful relationship between "quasi co-occurring" anomalous peaks in state-based violence in Nigeria (*ged_best_sb_Nigeria*) and protest events in Benin (*acled_count_pr_Benin*). Black ellipses enclose a few of the temporally-close peaks in these data features.

3.2 Anomalous Relationships

An analysis of conflict data demonstrates the utility of anomalous relationship detection with PODS. In this example, the data includes all countries in Africa with measures at the monthly level. The target variable is state-based conflict [38], and we want to discover if major changes in levels of state-based conflict are associated with shifts in other data features. For example, it might be that shifts in state-based violence in Nigeria are associated with sudden economic changes in Cameroon or Chad (other states where Boko Haram tends to operate). In our analysis, PODS detects associations between protest in many countries and state-based violence in Nigeria, including Benin (shown in Figure 4). While conflict contagion is a well-known phenomenon [26], it is less common to see contagion across types of conflict and instability, which is what we observe some evidence for here. This suggests additional questions that could be explored and are facilitated with these tools. For example, is this specific to Nigeria? To answer that, PODS could be re-run with a focus on different countries. What are the factors in Nigeria that suggest that nearby protest triggers state-based violence? New data from Auctus may be used to further explore this question.

3.3 Improved Modeling

Conflict researchers use predictive modeling for a number of reasons, including to guide policy-making decisions, and to assess theories of conflict [15]. For each of these purposes, it is crucial to identify new data sources, merge those data, and evaluate the contribution of different features. This is an iterative process loop, which can be facilitated through the integration of TwoRavens and Auctus. In practice, conflict researchers may have to analyze conflict events that appear in datasets with distinct spatial and temporal levels in a combined fashion, and joining these datasets may be challenging. Traditionally, researchers would have to specify functions to identify whether an event falls within a particular grid, if it falls within the grid *in the specified time interval*, and then use an appropriate function to aggregate cases when more than one event occurs. Auctus has features to accomplish this join of grid-level data with arbitrary event data, also allowing for different levels of temporal specification and different aggregation functions.

In this case study, we use grid data for Africa with conflict events aggregated into quarterly counts. Thus, each grid contains four observations per year. The target variable is again state-based conflict [38]. Our hypothesis is that social unrest and destabilizing events contribute to the use of force domestically [11]. The required steps are described as follows: (1) transition to Auctus from TwoRavens (see Figure 2); (2) search keyword *protest* and merge SCAD into the grid data, specifying spatial and temporal levels and selecting aggregation function to *count* the number of destabilizing events and to *sum* the fatalities; (3) transition back to TwoRavens from Auctus; and finally (4) forecast state-based conflict. After exploring the results generated by the AutoML system, we find that the number of destabilizing events at time t is a valuable predictor of state-based violence at time $t+1$. Joining these datasets also enables the researcher to use other features of TwoRavens, including the *discovered problems* component that automatically formulates models to predict features in the data, potentially leading to new research questions. As an example, are there existing variables associated with the merged SCAD data features that are meaningfully related? Upon merging another event dataset, this system feature would apply to combinations of the original data, the SCAD data, and the newly joined data. Thus, this integration between TwoRavens and Auctus provides compound benefits to the researcher.

4 CONCLUSION

In this paper, we describe an ecosystem of tools for domain experts to aid in conflict research. We present three case studies that showcase how these integrated tools facilitate the formulation and refinement of research hypotheses. This ecosystem also provides benefits to data scientists that typically work with domain experts to realize the value of computational methods. Thus, we expect that systems such as ours will have a multiplicative effect on research output. Open data management questions that can be explored in future research include: how to implement mixed-initiative visual interfaces that allow users to explore their needs in more detail; how to aggregate data from the event-level into grid-based or other units, better addressing the spatio-temporal complexity of conflict data; and how to use natural language processing techniques to better make sense of domain experts' information needs.

Acknowledgments. This work was partially supported by the DARPA D3M program award FA8750-17-2-0114, and NSF award OAC-1640864. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF and DARPA.

REFERENCES

[1] Robert Adcock and David Collier. 2001. Measurement validity: A shared standard for qualitative and quantitative research. *American political science review* 95, 3 (2001), 529–546.

[2] Gowtham Atluri, Anuj Karpatne, and Vipin Kumar. 2018. Spatio-Temporal Data Mining: A Survey of Problems and Methods. *Comput. Surveys* 51, 4 (2018), 83:1–83:41.

[3] Leilani Battle and Jeffrey Heer. 2019. Characterizing exploratory visual analysis: A literature review and evaluation of analytic provenance in tableau. *Computer Graphics Forum* 38, 3 (2019), 145–159.

[4] Nathaniel Beck, Gary King, and Langche Zeng. 2000. Improving quantitative studies of international conflict: A conjecture. *American Political science review* 94, 1 (2000), 21–35.

[5] Aline Bessa, Juliana Freire, Tamraparni Dasu, and Divesh Srivastava. 2020. Effective Discovery of Meaningful Outlier Relationships. *ACM/IMS Trans. Data Sci.* 1, 2, Article 12 (June 2020), 33 pages. https://doi.org/10.1145/3385192

[6] Christopher Blattman and Edward Miguel. 2010. Civil war. *Journal of Economic literature* 48, 1 (2010), 3–57.

[7] Mia Bloom. 2017. Constructing expertise: Terrorist recruitment and "talent spotting" in the PIRA, Al Qaeda, and ISIS. *Studies in Conflict & Terrorism* 40, 7 (2017), 603–623.

[8] Elizabeth Boschee, Jennifer Lautenschlager, Sean O'Brien, Steve Shellman, James Starz, and Michael Ward. 2015. ICEWS Coded Event Data. https://doi.org/10.7910/DVN/28075

[9] Thomas Chadefaux. 2017. Conflict forecasting and its limits. *Data Science* 1 (03 2017), 1–11. https://doi.org/10.3233/DS-170002

[10] Adriane Chapman, Elena Simperl, Laura Koesten, George Konstantinidis, Luis Daniel Ibáñez-Gonzalez, Emilia Kacprzak, and Paul T. Groth. 2020. Dataset Search: A Survey. *The VLDB Journal* 29 (2020), 251–272.

[11] Erica Chenoweth and Jay Ulfelder. 2017. Can structural conditions explain the onset of nonviolent uprisings? *Journal of Conflict Resolution* 61, 2 (2017), 298–324.

[12] Daina Chiba and Kristian Skrede Gleditsch. 2017. The shape of things to come? Expanding the inequality and grievance model for civil war forecasts with event data. *Journal of Peace Research* 54, 2 (2017), 275–297.

[13] Rand Corporation. 2021. RAND Corporation Website. https://www.rand.org/.

[14] Guo dao Sun and Ying cai Wu. 2013. A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. *Journal of Computer Science and Technology* 28, 5 (2013), 852–867.

[15] Vito D'Orazio. 2020. Conflict forecasting and prediction.

[16] Vito D'Orazio, Marcus Deng, and Michael Shoemate. 2018. TwoRavens for Event Data. In *2018 IEEE International Conference on Information Reuse and Integration (IRI)*. IEEE, Salt Lake City, USA, 394–401. https://doi.org/10.1109/IRI.2018.00065

[17] Iddo Drori, Yamuna Krishnamurthy, Remi Rampin, Raoni de Paula Lourenco, Jorge Piazentin Ono, Kyunghyun Cho, Claudio Silva, and Juliana Freire. 2018. AlphaD3M: Machine Learning Pipeline Synthesis. In *Proceedings of the ICML Workshop on Automatic Machine Learning*. ACM, Stockholm, Sweden, 8.

[18] One Earth Future Foundation. 2021. *One Earth Future Website*. One Earth Future Foundation.

[19] Matthew Frank, Clayton Besaw, Jay Benson, Jonathan Powell, and Eric Keels. 2019. Annual Risk of Coup Report - 2019. https://oefresearch.org/publications/annual-risk-coup-report-2019.

[20] Khusrav Gaibulloev and Todd Sandler. 2019. What we have learned about terrorism since 9/11. *Journal of Economic Literature* 57, 2 (2019), 275–328.

[21] Yolanda Gil, James Honaker, Shikhar Gupta, Yibo Ma, Vito D'Orazio, Daniel Garijo, Shruti Gadewar, Qifan Yang, and Neda Jahanshad. 2019. Towards Human-guided Machine Learning. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. ACM, New York, NY, USA, 614–624. https://doi.org/10.1145/3301275.3302324

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

[23] Håvard Hegre and Nicholas Sambanis. 2006. Sensitivity analysis of empirical results on civil war onset. *Journal of conflict resolution* 50, 4 (2006), 508–535.

[24] Sungsoo Ray Hong, Jorge Piazentin Ono, Juliana Freire, and Enrico Bertini. 2019. Disseminating Machine Learning to domain experts: Understanding challenges and opportunities in supporting a model building process.

[25] Trang T Le, Weixuan Fu, and Jason H Moore. 2020. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics* 36, 1 (2020), 250–256.

[26] Jessica Maves and Alex Braithwaite. 2013. Autocratic institutions and civil conflict contagion. *The Journal of Politics* 75, 2 (2013), 478–490.

[27] Kevin P. Murphy. 2013. *Machine learning : a probabilistic perspective*. MIT Press, Cambridge, MA, USA.

[28] Peter F Nardulli, Scott L Althaus, and Matthew Hayes. 2015. A progressive supervised-learning approach to generating rich civil strife data. *Sociological methodology* 45, 1 (2015), 148–183.

[29] Natasha Noy, Matthew Burgess, and Dan Brickley. 2019. Google Dataset Search: Building a search engine for datasets in an open Web ecosystem. In *28th Web Conference (WebConf 2019)*. IW3C2, San Francisco, USA, 1365–1375.

[30] NYU. 2020. Auctus. https://docs.auctus.vida-nyu.org/. Library Catalog: gitlab.com.

[31] Jorge Piazentin Ono, Sonia Castelo, Roque Lopez, Enrico Bertini, Juliana Freire, and Claudio Silva. 2020. PipelineProfiler: A Visual Analytics Tool for the Exploration of AutoML Pipelines. arXiv:2005.00160 [cs.HC]

[32] R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

[33] Clionadh Raleigh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. Introducing ACLED: An Armed Conflict Location and Event Dataset. *Journal of Peace Research* 47, 5 (2010), 651–660. https://doi.org/10.1177/0022343310378914

[34] Clionadh Raleigh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. Introducing ACLED: an armed conflict location and event dataset: special data feature. *Journal of peace research* 47, 5 (2010), 651–660.

[35] Naren Ramakrishnan, Patrick Butler, Sathappan Muthiah, Nathan Self, Rupinder Khandpur, Parang Saraf, Wei Wang, Jose Cadena, Anil Vullikanti, Gizem Korkmaz, et al. 2014. 'Beating the news' with EMBERS: forecasting civil unrest using open source indicators. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, New York City, USA, 1799–1808.

[36] Nicolas Rost. 2020. private communication. UN Office for the Coordination of Humanitarian Affairs.

[37] Aécio Santos, Sonia Castelo, Cristian Felix, Jorge Piazentin Ono, Bowen Yu, Sungsoo (Ray) Hong, Cláudio T Silva, Enrico Bertini, and Juliana Freire. 2019. Visus: An Interactive System for Automatic Machine Learning model Building and Curation. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*. ACM, Amsterdam, Netherlands, 6. https://doi.org/10.1145/3328519.3329134

[38] Ralph Sundberg and Erik Melander. 2013. Introducing the UCDP georeferenced event dataset. *Journal of Peace Research* 50, 4 (2013), 523–532.

[39] Tableau. 2021. *Tableau Official Website*. Tableau.

[40] Wei Wang, Ryan Kennedy, David Lazer, and Naren Ramakrishnan. 2016. Growing pains for global monitoring of societal events. *Science* 353, 6307 (2016), 1502–1503.

[41] Yuri Zhukov, Christian Davenport, and Nadiya Kostyuk. 2019. Introducing xSub: A new portal for cross-national data on subnational violence. *Journal of Peace Research* 56, 4 (05 2019), 604–614.