**“Machine Learning and Conflict Prediction: A Use Case”**

**Review 1**

**Overall Feedback.** This paper constitutes a useful contribution to the literature on conflict early warning / early response. Its value-added appears to lie in its use of sub-national indicators for conflict; this is an important step beyond existing, well-known applications of machine learning to EWER (though, as the authors are aware, others have taken this step as well). The language needs to be revised to appeal to readers from the policy / practice community.

**Suggested Changes**. Most of this article will not be accessible to a policy / practice audience. Following are several suggestions for improving accessibility:

* The definitions of (1) classification, (2) machine learning, (3) supervised vs. unsupervised learning, (4) naïve Bayes and (5) random forests will be difficult for many readers to understand. (I found them difficult to understand, and I am very familiar with these methods.) Can the definitions be reworked to include less jargon, and more analogies to real-world applications that readers will recognize and understand—Google, Netflix, etc.? (The cat example is too atypical of real-world applications to be helpful.)
* I found the description of the datasets and predictors unnecessarily dense. Can some of this material be moved to footnotes, or to a technical appendix?
* It would be helpful to include more easily digestible results alongside the confusion matrices and the performance plots. Just some descriptive statistics would help. What are the accuracy, sensitivity and specificity rates for the various models? What is the ratio of true positives to false positives, and true negatives to false negatives? The reader can calculate these statistics by hand using the results in the confusion matrices, but they should be included in the body of the paper. The second paragraph on p. 10 begins to do this, but should go further.
* It would be helpful to include a more intuitive baseline against which to judge the ML-based models. For example, in this application, what would happen if the authors simply predicted violence in time t everywhere that violence occurred in time t-1? Does ML outperform this intuitive decision rule, and by how much? The punchline of the paper, and the corresponding lesson for policymakers, depends very much on the answer to this question.

**Justifications and limitations**

* I found the justification for the use of random forests and (especially) naïve Bayes a bit lacking. Why did you choose these? And what should a policymaker learn from the authors’ choices? Is there a methodological / analytical principle at work here?
* It would be helpful to include a sentence or two on the limitations of ACLED. This dataset has been widely criticized for its idiosyncratic coverage and quality (see [Kristine Eck’s work](http://www.uu.se/digitalAssets/83/83553_CoCo__Eck__final_.111204.release_vers.pdf), for example), and readers should be aware of the biases that may arise as a result.
* It would also be helpful to include some discussion of the more general limitations of ML for conflict early warning / early response. For example, governments, NGOs and peacekeepers in these countries typically decide where to intervene on the basis of information gathered at relatively high frequency and high granularity. What are the advantages and disadvantages of ML relative to this approach?

**Miscellaneous**

* The transition from applications of machine learning in international development to conflict early warning on p. 2 is awkward. Why not cite some of the existing research that uses machine learning to predict onsets of violence? The work of the Political Stability Task Force comes immediately to mind, as does work by Jay Ulfelder and others on predicting mass atrocities. See also Tikuisis, Carment and Samy’s 2012 piece in the JCR (entitled “Prediction of Intrastate Conflict Using State Structural Factors and Events Data”), and Arva et al.’s 2013 paper presented at EPSA (entitled “Improving Forecasts of International Events of Interest”).

**Review 2**

This article presents a non-technical introduction to machine learning algorithms as applied to the case of conflict prediction. The articles make three key contributions: i) situating the emergence of machine learning and big data, ii) applying machine learning to the question of conflict prediction, and iii) doing so with a newly constructed dataset at the subnational level.

The introduction opens with the well-known public call to early warning systems and the commensurate follow in data and new empirical approaches. The advances noted – in data management, predictive analytics, and parallelization - are non-unique to big data; many of the algorithims used in machine learning were analytically derived prior to advances in computation. Moreover, much prediction has been done on ‘small data’. It is worth discussing this in depth, particularly as it connects to conflict prediction (see for example, Schrodt, Ulfelder, Goldsmith, Hegre, Gledtisch, etc.).

In your introduction to machine learning, I think it is worth discussing the divergence between hypothesis testing and prediction. A large body of research and policy has historically focused on understanding the causes of conflict; this hypothesis-oriented process is quite different from prediction and there are very different implications for what we learn. For example, we may have a predictive model that massacres are highly predictive of conflict, but we wouldn’t learn much from that; in you’re model the greatest predictor of your binary predictor of battles is a past battle: what should policy makers make of this? For an audience unfamiliar with the motivation of prediction, differentiating it from explanation would be fruitful. (Additionally, it seems as though many of the variables you use in your model are theoretically motivated; how did you choose though? It would be useful to explain how variable selection in machine learning works, particularly in an era of big data where analysts may have thousands of variables with which to work). In discussing these issues, I would also encourage you to focus on key methodological and conceptual issues rather than processing and management which are second order for individuals unfamiliar with predictive analytics.

The data collection process you engaged in is quite useful and a contribution in itself. It’s worth noting the dearth of systematic subnational data in countries affected by conflict. A few points of clarity would improve this section. First, it is a bit vague what the final unit of analysis is – you begin with district-years, but given that you introduce many variables and drop data, what are you left with? Assuming you do have a panel, which variables are time-invariant (ethinc composion, diamond deposits, petroleum, and a few others may be)? It would be helpful to provide summary statistics as well. Much of the focus on the national level conflict prediction has been fueled by a lack of data, but also by the desire to predict civil conflict. What percentage of the countries that have violent events in your dataset are not officially engaged in a civil war – in other words, to what extent are you predicting episodes of violence across the globe in any country or predicting episodes of violence conditional on civil war (ACLED is quite focused on the latter)? These are two substantively different exercises.

With respect to your conclusion, I think it would be useful to motivate the focus on subnational variation more – why does this inherently matter more? Is this because subnational data is a priori better or does it capture different dynamics of conflict? The vast majority of conflict data doesn’t require parallel processing – this comment needs to be connected directly to the call for data that does require this (i.e. connect points 2 and 4; if point 2 is not met, point 4 is less salient).

Lastly, your conclusion should also wrap up by answering a few questions for non-technical audiences: what should policy makers and scholars take away from your model? From conflict prediction more generally? Do you think that the lack of intervention or response is about a lack of information about conflict? If it isn’t, then what is the role of predictive analytics? What can we learn about the nature of conflict from your predictions?