# Assignment 9

## Referee Report with Extension Suggestions for:
## ***"Inferring Roll-Call Scores from Campaign Contributions Using Supervised Machine Learning"*** **(Bonica, 2018)**

### Research Question and Quality of Answers

The paper proposed an innovative method that aims to enhance the forecasting power of statistical modeling on "Roll-Call scores" predictions. To rephrase it in the form of a research question: "How will supervised machine learning models, with use of campaign contributions data, improve Roll-Call scores prediction?" (Bonica, 2018, p830) The research question is clear and adequately defined.

Firstly, to better illustrate how supervised machine learning models outperform others, a benchmark is drawn from the widely used unsupervised methods, such as "DW-NOMINATE and other related roll-call scaling models" (Bonica, 2018, p831) The author acknowledges the success these scaling models have achieved, and that they, to some extent, serve the criteria to measure voters' political ideology. But studies of Tausanovitch and Warshaw (2017) cast doubts on their capability to "distinguish between members within the same party". (Bonica, 2018, p831). In contrast, supervised learning models are built upon lesser assumptions and have better accuracy.

Second, the paper outlines the framework for how supervised modeling is constructed to infer DW-NOMINATE scores for candidates. The candidates with DW-NOMINATE provide the pool for the training set. As the target variable is continuous, regression models fit better with the goal to infer DW-NOMINATE scores. Several models are advised, namely Single Vector Regression, Random Forest, for training the regressing machine. The author made deliberate selections on explaining variables and individuals of observation. Only "the donors who have given to at least 15 distinct candidates included in the training set" (Bonica, 2018, p836), so that the data are kept safe from being too sparse. Special caution is laid upon how to think of measurement errors associated with DW-NOMINATE scores, which might cause over-fitting problems. Therefore, the author tried to regularize the estimation by imposing penalty on kernel-regression or ensemble learning.

Third, results are displayed in comparison with other alternative measures of ideology. Good news is that supervised models derived from pre-incumbent data significantly outperform most unsupervised methods, just as well as Nokken-Poole estimates derived from real voting records. The improvement implies that roll-call measures can be "very sensitive" to assumption alteration. The author follows up by analyzing how well they can classify predicted DW-NOMINATE scores against actual votes. Supervised models, especially the Random Forest Model, turn out to "perform only second to DW-NOMINATE scores" (Bonica, 2018, p840) Supervised models are prominently robust to both intraparty and interparty contexts, which is further asserted by the feature analysis.

Generally, I consider the methodology as valid and thoroughly investigated. The author, in the first step, confesses the unsatisfactory facets of unsupervised models and then propels the application of supervised learning that not only generates accurate outcomes but also enables prediction on nonincumbent candidates. However, it is questionable whether the shift from conventional unsupervised models to supervised ones fundamentally improves the interpretability, or resonates more with the underlying political science theories. I will expand on this in the following parts.

## Quality of Answers in a Broader Context

An essential part of the paper "Statement of the Problem" is contributed to dissecting spatial theories that modern political science draws upon, that link the lower-dimensional space of political ideology and higher-dimensional issue preference. The author contrasts this theoretical foundation with statistical models methods in a broader context: the theory construct takes a top-down perspective, which treats voters' issue preference as "holographic interpretation" of their ideological space (Bonica, 2018, p832); while statistical modeling takes a bottom-up strategy, that utilizes issues voted hundreds of times to recover a 'frequency-weighted' estimate of the ideological space. However, the introduction to new statistical techniques, such as supervised learning methods, are unable to resolve this conflicts, or at least remains unsolved in this paper. The author better cites more on discussions on the issue, or how researchers can develop the learning tools to match better with the theory. Besides, There is a minor citation advice. The author brings up an analogy to education testing when describing the methodology of ideal point estimation. However, there is no proper citation to clarify how that is applied in the field of "intelligence or aptitude test". And I would suggest the author considering replace it with other political science illustrations, since there are rich studies on that, like Tahk, A. (2018), Marble and Tyler (2017).

## Grammatical Advice

In the introductory part of the paper, the author briefly summarizes the vantage of supervised learning models, "This is accomplished with a high degree of accuracy and with minimal identifying assumptions. This approach, however, is not limited to roll-call scores." (Bonica, 2018, p831) I feel the 'however' here sounds a bit unfitted. The author implicitly assumes that the readers reckon the model as unique to predicting roll-call scores. But I believe most researchers with computational background will naturally anticipate that this supervised modeling can be applied in wider contexts. I would suggest replacing the original sentence as something like "Also, the model can go beyond predicting roll-call scores".

In the "results" part of the paper, the author compares the supervised model performance to unsupervised ones by saying "Perhaps most telling is that the supervised machine learning models perform on par with the Nokken-Poole first-term estimates in predicting DW-NOMINATE scores." (Bonica, 2018, p837) "Most telling" seems like an improper subject of the sentence, at least to non-native speakers like me. May it can be substituted by "Perhaps what tells us most is that".

### Extension of the Research Method

The paper inspires me by how elegantly the machine learning models have turned financial data into political science attainments. Supervised learning empowers researchers to recover individual ideology status at a lower-dimensional space, and with this knowledge, we can probably extend to areas of finance, marketing, and sociology. Interesting questions can be broached. For example, how can donations indicate investor decisions? Just as what Bonica (2018) did in this research, we can categorize donations into different social domains. By analyzing what donation data reveal about people's behavioral pattern, we can recover what social issues they concern about. Then combining these findings with other commonly used features, like stock returns, P/E ratio, we then may be able to see if investment decisions are predicted better. Benchmark can be set as other supervised learning models without the use of donation data. Let's see if investors' internal ideology status drive them away from solely taking care of pecuniary returns? Could this explain some unwise or irrational investments that are happening out in the market?

## Reference:

**Tausanovitch, Chris, and Christopher Warshaw. 2017.** "Estimating Candidates' Political Orientation in a Polarized Congress." *Political Analysis* 25(2): 167–87.

**Cahoon, Lawrence S., Melvin J.Hinich, and Peter C.Ordeshook. 1976.** "A Multidimensional Statistical Procedure for Spatial Analysis." *Unpublished manuscript*, Carnegie-Mellon University.

**Tahk, A. 2018.** "Nonparametric Ideal-Point Estimation and Inference." *Political Analysis*, 26(2), 131-146.

**Marble, Tyler 2017.** “How Much Should We Trust Ideal Point Estimates from Surveys?” *2017 meeting of the American Political Science Association*, https://web.stanford.edu/~wpmarble/docs/MarbleTyler_SurveyIdealPointsAPSA.pdf