EPPS Knowledge Mining
Comparison between Breiman (2003) and Shmueli (2010)
Prepared for: Dr. Ho
Prepared by: Glen Cooper
Date: January 27, 2022

Breiman and Shmueli both address the nature of modeling. Both authors examine the components of "predictive" models and explore "explanatory" or causal models. While the authors sometimes use slightly different terminology, both authors see predictive models as having the goal of algorithmically forecasting outcomes with the greatest accuracy. The exact causal explanation for why the result is occurring is less, but not necessarily entirely, unimportant. On the other hand dor explanatory models, the theory and causal explanation is the most critical goal.

Breiman's paper assumes an advocation tone. He discusses how statisticians think that they can build good parametric models that can reasonably model complex mechanisms. Breiman claims that this belief might be overly optimistic. To support their faith, statisticians have developed multiple goodness-of-fit standards (e.g., R^2 values, p values, t-tests, residual analysis, etc.). However, these measures have become suspect. Breiman suggests that explanator models have benefits and can be valuable in creating understandable pictures of reality. However, they are limited. Breiman asserts that predictive models, which often abandon understanding the casual relationships between predictor variables and the response variable, can help scientific advancement by providing a new way to get reliable information about complex processes.

Shmueli echoes Breiman's advocacy for predictive modeling but also outlines a step-by-step model development approach to contrast the differences between the two types of modeling. It should be noted that for completeness, Shmueli mentions a third type of modeling known as descriptive modeling. This model type aims to summarize data structures and, unlike explanatory modeling, is not reliant on causal theory. Shmueli outlines the following necessary steps in the development of any statistical model: 1) Define, 2) Design study and collect data, 3) Prepare data, 4) Exploratory data analysis, 5) Choose variables, 6) Choose methods, 7) Evaluate, validate, and model selection, 8) Use model and report. After breaking the steps down in this manner, Shmueli discusses how each step can be analyzed through the lens of either building an explanatory model or a predictive model by highlighting the differences between the models at each stage. In this way, Shmueli provides an excellent model for building either type of model.

In conclusion, both Breiman and Shmueli believe that the scientific community should more fully embrace predictive modeling and integrate its methods into scientific research.

References:

Breiman, L. (2003). Statistical modeling: The two cultures. Quality control and applied statistics, 48(1), 81-82.

Shmueli, G. (2010). To explain or to predict?. Statistical science, 25(3), 289-310.