An interpretable machine learning workflow with an application to economic forecasting

Marcus Buckmann, Andreas Joseph
Bank of England, United Kingdom.

Abstract

We propose a generic workflow for the use of machine learning models to inform decision making and to communicate modelling results with stakeholders. It involves three steps: (1) a comparative model evaluation, (2) a feature importance analysis and (3) statistical inference based on Shapley value decompositions. We discuss the different steps of the workflow in detail and demonstrate each by forecasting changes in US unemployment one year ahead using the well-established FRED-MD dataset. We find that universal function approximators from the machine learning literature, including gradient boosting and artificial neural networks, outperform more conventional linear models. This better performance is associated with greater flexibility, allowing the machine learning models to account for time-varying and nonlinear relationships in the data generating process. The Shapley value decomposition identifies economically meaningful nonlinearities learned by the models. Shapley regressions for statistical inference on machine learning models enable us to assess and communicate variable importance akin to conventional econometric approaches. While we also explore high-dimensional models, our findings suggest that the best trade-off between interpretability and performance of the models is achieved when a small set of variables is selected by domain experts.

Keywords: *machine learning, model interpretability, forecasting, unemployment, Shapley values.*