

# Balancing Accuracy and Interpretability with Prediction Rule Ensembles

Marjolein Fokkema

## **Abstract:**

Most statistical prediction methods provide a trade-off between accuracy and interpretability. For example, single classification trees are easy to interpret, but provide lower predictive accuracy than methods like tree ensembles. Tree ensembles, on the other hand, are more difficult to interpret, sometimes even termed black boxes. Prediction rule ensembles (PREs) aim to strike a balance between accuracy and interpretability. PREs consist of a small set of prediction rules, which can be depicted as very simple decision trees, and as such are easy to interpret and apply. Friedman and Popescu (2008) proposed the RuleFit algorithm for deriving PREs, which has been found to provide accuracy competitive with random forests and boosted tree ensembles. The R package "pre" implements the RuleFit algorithm in R, with several adjustments. In this talk, I will illustrate the methodology and package with real data examples, and discuss current developments, like PREs for multilevel data.