## Random forest: A general purpose classification and regression machine learning method

Matko Bošnjak, Rudjer Boskovic Institute

## Abstract

Random forest is an ensemble of decision tree predictors which is built through a combination of two sources of randomness: bagging (bootstrapping aggregation) and random subspace method. The resulting ensemble benefits on low bias of independent unpruned decision tree while decreasing its variance through decision tree plurality. Through bagging, RF employs internal error estimates proven as unbiased in many tests. The resulting ensemble of predictors obtains very good results in terms of efficiency, quality and reliability while its error rate converges to a limit with the increase of the number of trees. As a machine learning method, RF enables very interesting qualitative interpretation of results through a large number of additional features such as variable importance, multidimensional scaling, prototyping etc. Besides supervised learning, RF can also be used for unsupervised learning.