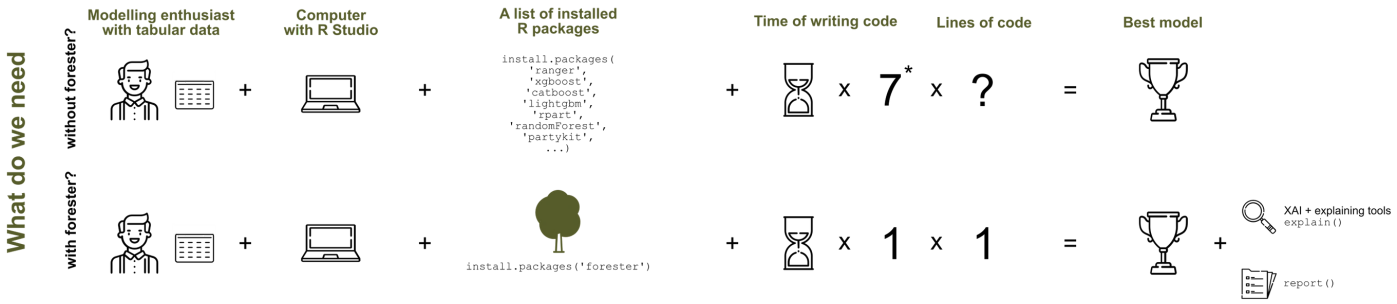
forester: an R package for automated building of tree-based machine learning models

Adrianna Grudzień, Hubert Ruczyński, Patryk Słowakiewicz

thesis supervisor: mgr Anna Kozak

consultation: dr hab. inż. Przemysław Biecek, prof. uczelni



How to build tree-based models in R?

What is *forester*?

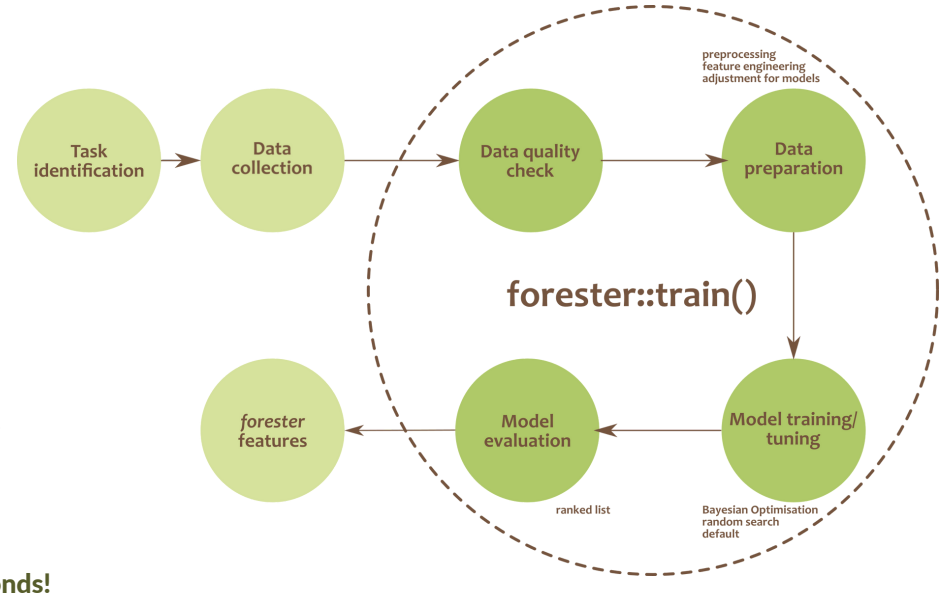
- full automation of the process of training tree-based models
- no demand for ML expertise
- powerful tool for making high-quality baseline models for experienced users

The *forester* package is **an AutoML tool in R** that wraps up all machine learning processes into a single `train()` function, which includes:

1. rendering a brief **data check** report,
2. **preprocessing** initial dataset enough for models to be trained,
3. **training** 5 tree-based models with default parameters, random search and Bayesian optimisation,
4. **evaluating** them and providing a ranked list.

How to use it?

```
library(forester)
data('lisbon')
train_output <- train(lisbon, 'Price')
```



For whom is this package created?

The *forester* package is designed for beginners in data science, but also for more experienced users. They get an easy-to-use tool that can be used to prepare high-quality baseline models for comparison with more advanced methods or a set of output parameters for more thorough optimisations.

Prepare meaningful report less than in 60 seconds!

As data scientists, we are fully aware that there are some time expensive processes in out work. One of them is creating a report with meaningful results. That's why one of the most powerful *forester* feature, which makes it a efficient tool for both experienced users and the newcomers, is a report() function. This single-line command is designed to **provide a holistic view on the outcomes of the ML process** happening inside of the forester.

Faculty of Mathematics and Information Science
WARSAW UNIVERSITY OF TECHNOLOGY