

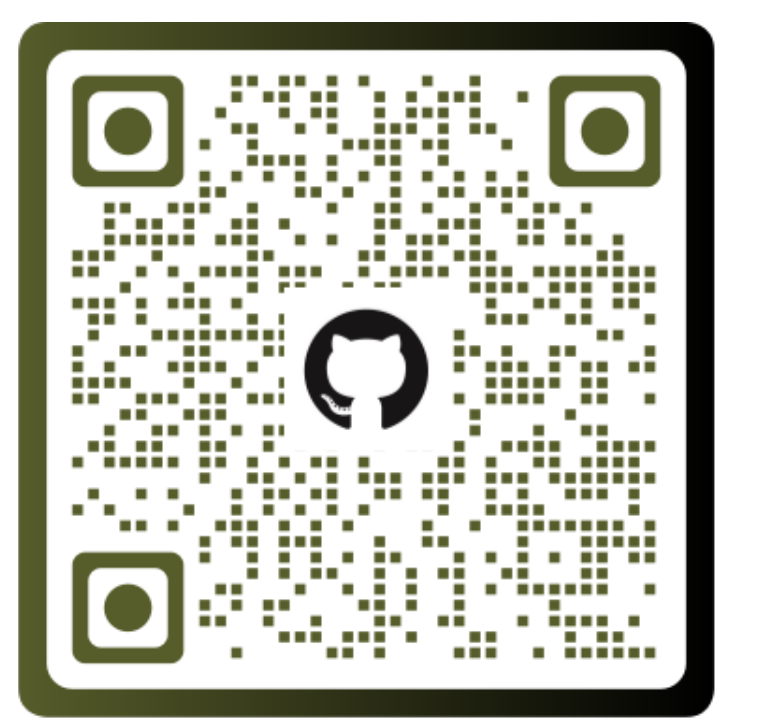


forester: growing transparent tree-based models for everyone

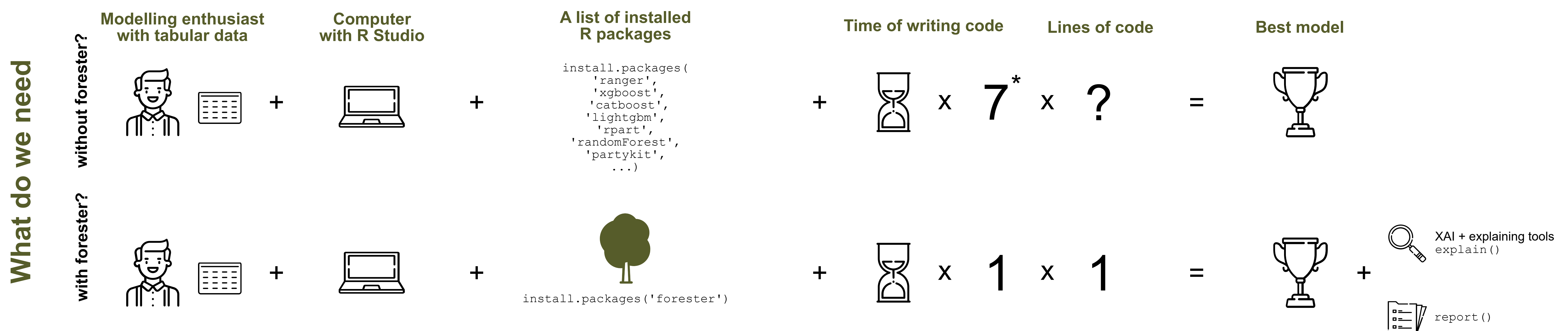
Anna Kozak¹, Adrianna Grudzień¹, Hubert Ruczyński¹,

Patryk Słowakiewicz¹, Przemysław Biecek^{1,2}

¹MI2.AI, Warsaw University of Technology ²MI2.AI, University of Warsaw



Let's talk about AutoML, tree-based models, explainable AI (XAI), exploratory data analysis (EDA)!



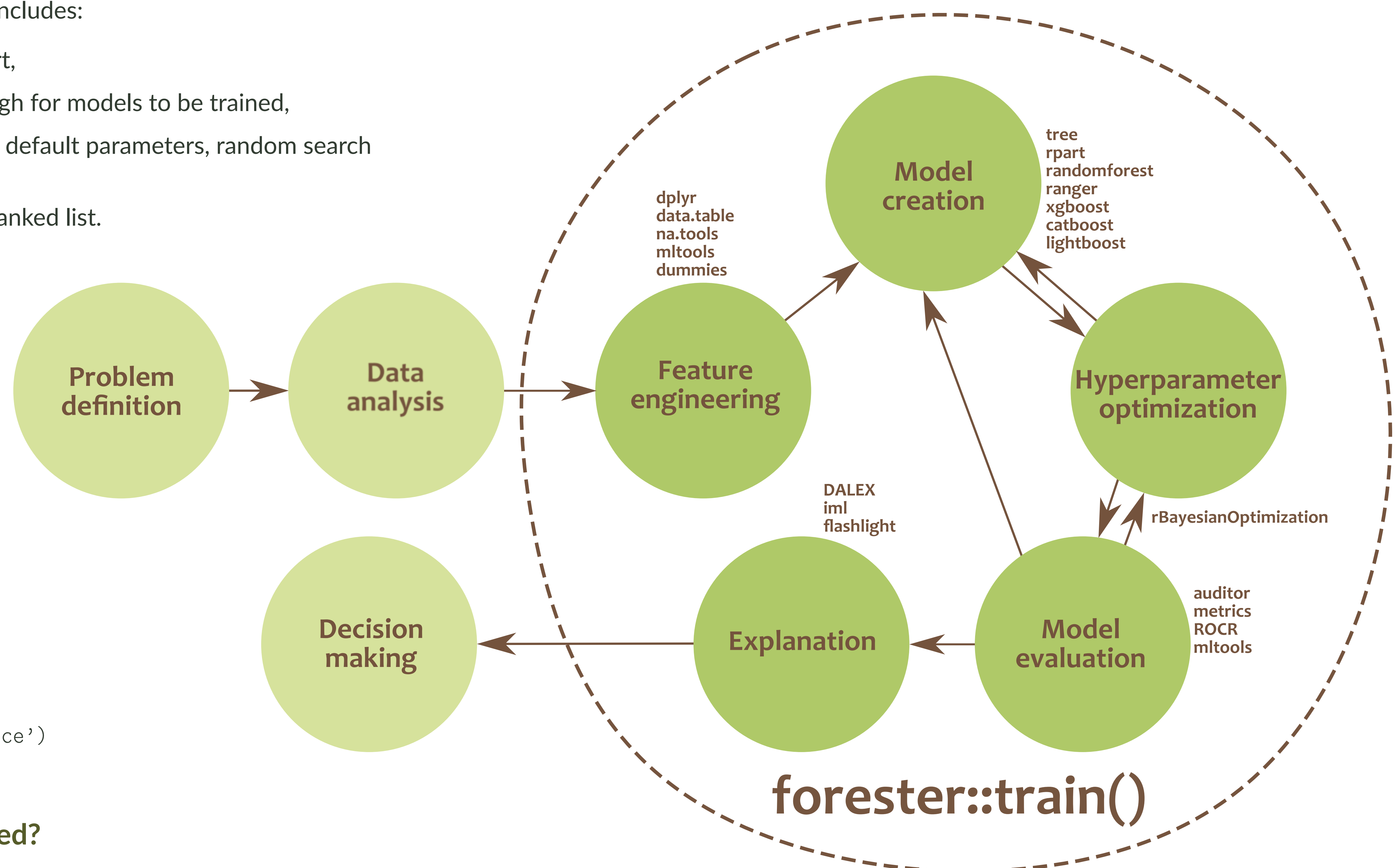
How to build tree-based models in R?

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.

Contact info

✉ anna.kozak@pw.edu.pl
<https://github.com/ModelOriented/forester>

References

P. Biecek. DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research*, 19(84):1-5, 2018. URL <https://jmlr.org/papers/v19/18-416.html>.

A. Kozak, H. Ruczyński, P. Słowakiewicz, A. Grudzień, and P. Biecek. *forester: Quick and Simple Tools for Training and Testing of Tree-based Models*, 2022. URL <https://github.com/ModelOriented/forester>. R package version 1.0.0.

Prepare meaningful report less than in 60 seconds!

As data scientists, we are fully aware that there are some time expensive processes in our work. One of them is creating a report with meaningful results. That's why one of the most powerful *forester* feature, which makes it an efficient tool for both experienced users and 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*.

See the report yourself!

