

# HOMER: Human Oriented autoMated machinE leaRning

Machine learning models are used everywhere. Predictive modelling fundamentally changed data-driven disciplines like health-care, biology, finance, legal, military, security, transportation, and many more. Increasing availability of large annotated data sources combined with recent developments in machine learning models leads to the next industrial revolution.

**BUT:** Predictive models are handcrafted by data scientists in a tedious and laborious process. Most of time spend on data exploration and training is a set of try-and-error experiments. Models become more and more complex. Lack of understanding of complex models and poor automation results in problems with replicability and quality of models. This leads to harmful situations.

- Models are not working properly and are hard to debug. E.g. *Watson for Oncology* was criticized by oncologists for delivering *unsafe and inaccurate* recommendations (Ross and Swetliz, 2018).
- Results are biased in a systematic ways. E.g. AI giant, Amazon failed with system for CV screening, as it was biased against woman (Dastin, 2018), or COMPAS recidivism algorithm model discriminates against race (Larson et al., 2016). These are serious violations of fairness and ethical principles.
- Data drift leads to the deterioration in models performance. E.g. very popular model *Google Flu* after two years gave worse predictions than the baseline (Salzberg, 2014).
- Model prediction is wrong but no-one can explain which inputs drive this particular prediction. Many examples for such problems ssmay be found in the book (O’Neil, 2016) with expressive subtitle „*How Big Data Increases Inequality and Threatens Democracy*”.

Most of these problems would be spotted sssswith better human oriented methods for automated debugging, exploration and explanation of machine learning models.

**The main goal of this project is to develop new methods for human oriented model exploration, interpretable model audits and automated model assembly. The newly appointed research team will create a grammar for human oriented model development.**

It is a general long term goal for the research team. We also set up seven specific research questions and corresponding seven tasks for this project.

- Q1 Can we use a complex machine learning model as surrogate to extract interpretable features and interpretable model?
- Q2 Would interpretable scoring systems like Elo ranking improve the comparison of machine learning models.
- Q3 Would Elo-based embeddings increase the efficiency of Meta learning in optimisation of hyperparameters?
- Q4 Would interpretable Elo scores help in detection of concept drift? The decrease in predictive power measures with Elo rating may be an efficient statistic for identification of drift in data?
- Q5 Would Reinforcement Learning help in automated of data exploration?
- Q6 Would Reinforcement Learning improve feature selection process?
- Q7 How process oriented methodologies developed for Software Engineering fits Data Science projects?

These questions are based on our recent results published in Biecek (2018), Gosiewska and Biecek (2019), Staniak and Biecek (2019) and Biecek (2019).

## References

- Biecek, P. (2018). DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research*, 19:1–5.
- Biecek, P. (2019). Model Development Process. <https://arxiv.org/abs/1907.04461>.
- Dastin, J. (2018). Amazon scraps secret ai recruiting tool that showed bias against women. *Reuters*.
- Gosiewska, A. and Biecek, P. (2019). ibreakdown: Uncertainty of model explanations for non-additive predictive models. *CoRR*, abs/1903.11420.
- Larson, J., Mattu, S., Kirchner, L., and Angwin, J. (2016). How We Analyzed the COMPAS Recidivism Algorithm. *ProPublica*.
- O’Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group, New York, NY, USA.
- Ross, C. and Swetliz, I. (2018). Ibm’s watson super-computer recommended ‘unsafe and incorrect’ cancer treatments, internal documents show. *Statnews*.
- Salzberg, S. (2014). Why google flu is a failure. *Forbes*.
- Staniak, M. and Biecek, P. (2019). The Landscape of R Packages for Automated Exploratory Data Analysis. <https://arxiv.org/abs/1904.02101>.

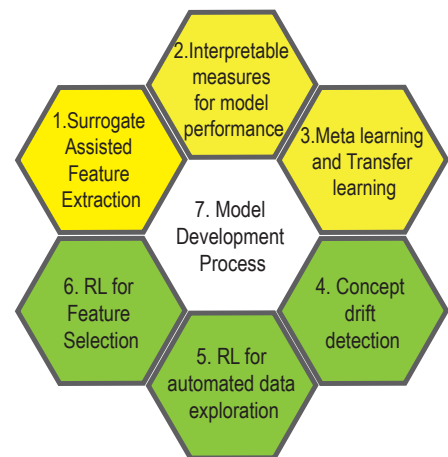


Figure 1: HOMER tasks. Yellow for human-model interaction, green for automation of model assembly.