

# mlr3proba: Probabilistic Supervised Learning for “mlr3”

Raphael E.B. Sonabend<sup>1</sup> and Franz J. Kiraly<sup>1, 2</sup>

<sup>1</sup> Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, United Kingdom <sup>2</sup> Shell

DOI: [10.21105/joss.0XXXX](https://doi.org/10.21105/joss.0XXXX)

## Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Editor Name](#) ↗

Submitted: 01 January 1900

Published: 01 January 3030

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

## Summary

mlr3proba is a machine learning toolkit for making probabilistic predictions within the mlr3 (Lang, Binder, et al., 2019) ecosystem. Currently mlr3proba only implements survival analysis through the Surv task, however future updates will introduce probabilistic supervised regression, as well as unconditional distribution estimation.

Key features of mlr3proba are

- A unified fit/predict model interface to any probabilistic predictive model (frequentist, Bayesian, or other)
- Pipeline/model composition
- Task reduction strategies
- Domain-agnostic evaluation workflows using task specific algorithmic performance measures.

mlr3proba is an R (R Core Team, 2017) package that implements survival models in a machine learning framework using the mlr3 (Lang, Binder, et al., 2019) family of packages. mlr3 is the official upgrade to mlr (Bischl et al., 2016), which makes use of the state-of-the-art R6 (Chang, 2018) object-oriented paradigm. mlr3proba builds on mlr3 by introducing survival models and measures with a distributional predict type.

mlr3proba makes use of distr6 (Sonabend & Kiraly, 2019) in order to predict probability distributions that represent survival curves. This has a key advantage over other survival packages as it unifies the return type into a single object.

Related software includes the original mlr (Bischl et al., 2016) package, which includes functionality for some survival learners. Additionally the survival (Therneau, 2015) library implements some classical survival models. Otherwise support for survival analysis is limited to packages that implement particular models or model classes.

## Key Use-Cases

1. **Fitting and predicting survival models** - This is the first and foremost use-case of the package in its current state, though this will later expand to density estimation and probabilistic regression. By making use of the mlr3 train/predict methods, users can treat classical survival models just like machine learning ones.
2. **Inspection of fitted survival models** - Whilst different packages in R have different ways of inspecting fitted models, this is unified by the model field, which is again implemented through mlr3.

3. **Tuning of machine learning models** - Users can make use of `mlr3tuning` (Lang et al., 2019) with any of the implemented survival models, in order to tune and improve the available ML models.
4. **Evaluation of survival models with transparent measures** - Several R packages evaluate survival models however each produces slightly different results and are not transparent. `mlr3proba` implements transparent measures with clear documentation.

## Software Availability

`mlr3proba` is available on [GitHub](#) and [CRAN](#). It can either be installed from GitHub using the `devtools` (Wickham, Hester, & Chang, 2019) library or directly from CRAN with `install.packages`. The package uses the MIT open-source licence. Contributions, issues, feature requests, and general feedback can all be found and provided on the project [GitHub](#). Full tutorials and further details are available on the [project website](#).

## Acknowledgements

`mlr3proba` was originally built from `mlr3survival`, created by Michel Lang. `mlr3survival` included four learners (`coxph`, `glmnet`, `ranger`, `rpart`), two measures (`harrellc`, `unoc`), and original designs for the survival prediction, learner, and task classes. Since merging `mlr3survival` into `mlr3proba`, these have been slightly modified however much of this original code remains. Design discussions of `mlr3proba` were held with Bernd Bischl, Michel Lang, Jakob Richter, and Martin Binder. Additional contributions have been made by Andreas Bender.

## References

- Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., Casalicchio, G., et al. (2016). mlr: Machine Learning in R. *Journal of Machine Learning Research*, 17(170), 1—5. Retrieved from <http://jmlr.org/papers/v17/15-066.html>
- Chang, W. (2018). R6: Classes with Reference Semantics. Retrieved from <https://cran.r-project.org/package=R6>
- Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., Au, Q., et al. (2019). mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software*, 4(44), 1903. doi:10.21105/joss.01903
- Lang, M., Richter, J., Bischl, B., & Schalk, D. (2019). mlr3tuning: Tuning for 'mlr3'. CRAN.
- R Core Team. (2017). R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.
- Sonabend, R., & Kiraly, F. (2019, July). distr6: The Complete R6 Probability Distributions Interface. CRAN. Retrieved from <https://cran.r-project.org/package=distr6>
- Therneau, T. M. (2015). A Package for Survival Analysis in S. Retrieved from <https://cran.r-project.org/package=survival>
- Wickham, H., Hester, J., & Chang, W. (2019). devtools: Tools to Make Developing R Packages Easier. CRAN.