template: clone this doc for each reading group paper

Model Info Sheet for Detecting and Preventing Leakage in ML-based Science (abridged)

**About model info sheets**

Completing this model info sheet requires the researcher to provide precise arguments to justify that predictive models used for making scientific claims do not suffer from leakage. It is inspired by the model cards introduced by Mitchell et al.[[1]](#footnote-0)

Model info sheets are intended to accompany the paper or report that introduces the model: for instance, as an appendix or supplemental material.

This is a beta version of our model info sheet template. We are soliciting feedback and will continue to update the template. For more information about model info sheets and to obtain the latest version of the template, see [reproducible.cs.princeton.edu](https://reproducible.cs.princeton.edu/). For feedback or questions, contact: [sayashk@princeton.edu](mailto:sayashk@princeton.edu)

The model info sheet starts on the next page. After filling it out, save it starting from that page. To cite the paper that introduces the model info sheets, use the bibliography file available at [reproducible.cs.princeton.edu/citation.bib](https://reproducible.cs.princeton.edu/citation.bib)

Model Info Sheet

**Section 1: Information about paper or report**

1) Author(s): Names of the authors of the paper or report

2) Title of the paper or report which introduces the model

3) DOI or permanent link to the paper or report (for example, link to arxiv.org webpage)

4) Email address of the corresponding author

**Section 2: Scientific claim(s) of interest**

5) Does your paper make a generalizable claim based on the ML model? If yes, what is the scientific claim?

7) Is the scientific claim made about a distribution or population from which you can sample? If yes: (a) what is the population or distribution about which the scientific claim is being made? (b) What is the sample used for the study?

8) Does the scientific claim only apply to certain subsets of the distribution mentioned in Q6?

**Section 3: Train-test split is maintained across all steps in creating the model**

9) Train-test split type: How was the dataset split into train and test sets? (For example, cross-validation; separate train and test sets).

10) Are there duplicates in the dataset? If yes, explain how duplicates are handled to ensure the train-test split.

11) In case the dataset has dependencies (e.g., multiple rows of data from the same patient), describe how the dependencies were addressed (for example, using block-cross validation).

12) List all the pre-processing steps used in creating your model. For example, imputing missing data, normalizing feature values, selecting a subset of rows from the dataset for building the model.

13) How was the train-test split observed during each pre-processing step? If applicable, use a separate line for each step mentioned in Q12.

14) List all the modeling steps used in creating your model. For example, feature selection, parameter tuning, model selection.

15) How was the train-test split observed during each modeling step? If applicable, use a separate line for each step mentioned in Q14.

16) List all the evaluation steps used in evaluating model performance. For example, cross-validation, out-of-sample testing.

17) How was the train-test split observed during each evaluation step? If applicable, use a separate line for each step mentioned in Q16.

**Section 4: Test set is drawn from the distribution of scientific interest.**

18) Why is your test set representative of the population or distribution about which you are making your scientific claims?

19) Explain the process for selecting the test set and why this does not introduce selection bias in the learning process.

20) In case your model is used to predict a future outcome of interest using past data, detail how data in the training set is always from a date earlier than the data in the test set.

**Section 5:** **Each feature used in the model is legitimate for the task**

21) List the features used in the model, alongside an argument for their legitimacy. A legitimate feature is one that would be available when the model is used in the real world and is not a proxy of the outcome being predicted. You can also include this list in an appendix and reference the relevant section of your Appendix here.

1. Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. "Model cards for model reporting." In *Proceedings of the conference on fairness, accountability, and transparency*, 2019. [↑](#footnote-ref-0)