

Writing papers

Tom Rochette <tom.rochette@coreteks.org>

December 21, 2025 — 77e1b28a

0.1 Context

Machine learning is a very young field of science, yet it is one in which we have the benefit of using computers. With that benefit should also come the benefit of high reproducibility of experiments, which is currently not the case. In this article we attempt to list best practices in order to improve the quality of the papers written and published so that other scientists may reproduce the experiments you have tried and that they may possibly extend on your approach or test alternative models using the same dataset.

0.2 Learned in this study

0.3 Things to explore

1 Overview

- Always list the datasets that were used as train/validation/test split
- Always link to where you retrieved those datasets and indicate when you acquired them (similar to how wikipedia does bibliographies)
- Always indicate how to reproduce your train/validation/test split
 - In the best scenario, provide code that replicates how you've split your data
- In the best case, provide all the code you wrote to prepare your paper

2 Notes

3 See also

4 References