

# Assignment 1: Algorithm Selection

## COSC 4010/5010: Practical Machine Learning

Lars Kotthoff  
larsko@uwyo.edu

For this assignment, we will have a look at the simplest version of Automated Machine Learning and choose the best type of machine learning model for a given task. The end result (i.e. the predictive performance) is not important; how you get there is. Your deliverable will be a report, written in a style that it would be suitable for inclusion in an academic paper as the “Experimental Setup” section or similar. If unsure, check an academic paper of your choice, for example [https://www.eecs.uwyo.edu/~larsko/papers/pulatov\\_opening\\_2022-1.pdf](https://www.eecs.uwyo.edu/~larsko/papers/pulatov_opening_2022-1.pdf). The level of detail should be higher than in a typical academic paper though. Your report should be at most five pages, including references and figures but excluding appendices. It should have the following structure:

**Introduction** What problem are you solving, how are you going to solve it.

**Dataset Description** Describe the data you’re using, e.g. how many features and observations, what are you predicting, any missing values, etc.

**Experimental Setup** What specifically are you doing to solve the problem, i.e. what programming languages and libraries, how are you processing the data, what machine learning algorithms are you considering, how are you evaluating them, etc.

**Results** Description of what you observed, including plots.

**Appendices** For example the code you’ve used. Optional.

There is no required format for the report. You could, for example, use an iPython notebook. Please submit a PDF file.

## Data

We will have a look at Wine Quality dataset<sup>1</sup> for this assignment. Choose the one that corresponds to your preference in wine. You may also use a dataset of your choice, for example one that’s relevant to your research.

---

<sup>1</sup><https://archive-beta.ics.uci.edu/dataset/186/wine+quality>

## ML Algorithm Selection

Choose a small number of different machine learning algorithms. For example, you could use a random forest, support vector machine, linear/logistic regression, a decision/regression tree learner, and gradient boosting. You will also have to choose their hyperparameters, e.g. the default values. Determine the best machine learning algorithm for your dataset, where the “best” algorithm could be a set of algorithms. Make sure that the way you evaluate this avoids bias and overfitting. You could, but do not have to, use statistical tests to make this determination.

The report needs to convince me that your conclusions are correct and that you have obtained them in a reasonable way.