# Determining a model usefulness

Tom Rochette <tom.rochette@coreteks.org>

November 2, 2024 — 36c8eb68

## 1  Question

How do you determine whether you have a useful model?

## 2  Answer

In machine learning, one way you can determine that you have a useful model is to compare it against baseline models. In a field such as time series, one can create models that are based on previous values, such as lag 0, which predicts that the next value will be equal to the current value, or a moving average, which takes the X last values and averages them and returns this average as the next predicted value. In this field, we expect that a model that can predict more accurately than those baseline models may prove to be useful.

In other cases, we may already have an existing model from which we can generate predictions. This model may also serve as a baseline which other models will have to beat in order to replace it.

There is however a case where the answer isn't clear: what happens when out of various models, one of the baseline models is the best? Then it becomes a question of whether the prediction interval produced by the model satisfies your needs for your problem.

Note that even when you beat the baseline models, if the best model still does not satisfy the error requirements that have been defined on the metrics you care about, the model may still not be useful. For example, an OCR model that has 95% accuracy at the character level will still produce 5 errors every 100 characters, which may be too high for the requirements of the system to be produced.

Another factor to keep in mind when comparing models is whether the improvements on the metric you are to optimize for are significant. The mean absolute error (MAE) may be lower for a model compared to another, but if its confidence interval is larger compared to the other model and their intervals overlap, then you cannot claim that one model is really better than the other. There may even be cases where you will prefer a model with higher MAE simply because its confidence interval is smaller than the other model with lower MAE but larger confidence interval.

There might be other attributes of the model that you may need to take into account. A model that takes days to train may be useless when you need it to be up to date every hour. A faster to train but less accurate model may be more useful in this case.

### 2.1  Summary

- Can the model under evaluation beat baseline models?
- Does the model under evaluation satisfy error requirements?
- Is the model under evaluation significantly better than existing models?
- Does the model under evaluation satisfy requirements such as time to train (e.g., less than 6 hours) or necessary resources (cpu, gpu, ram)?