

Now it's time to start modelling. So, this is where it really depends on your use case and data set how you want to proceed. For example, if you are in an unsupervised context you can choose between an auto-encoder, PCA or clustering. Or if you are in a supervised context you have choice between different state-of-the-art machine learning and deep learning algorithms. But here are some guidelines which are required to follow:

- Choose, justify and apply a model performance indicator (e.g. F1 score, true positive rate, within cluster sum of squared error, ...) to assess your model and justify the choice of an algorithm
- Implement your algorithm in at least one deep learning and at least one non-deep learning algorithm, compare and document model performance
- Apply at least one additional iteration in the process model involving at least the feature creation task and record impact on model performance (e.g. data normalizing, PCA, ...)

Depending on the algorithm class and data set size you might choose specific technologies / frameworks to solve your problem. Please document all your decisions in the ADD (Architectural Decisions Document).

Once you think you have achieved a descent model performance save the notebook according to the process model's naming convention and proceed to the model training task.