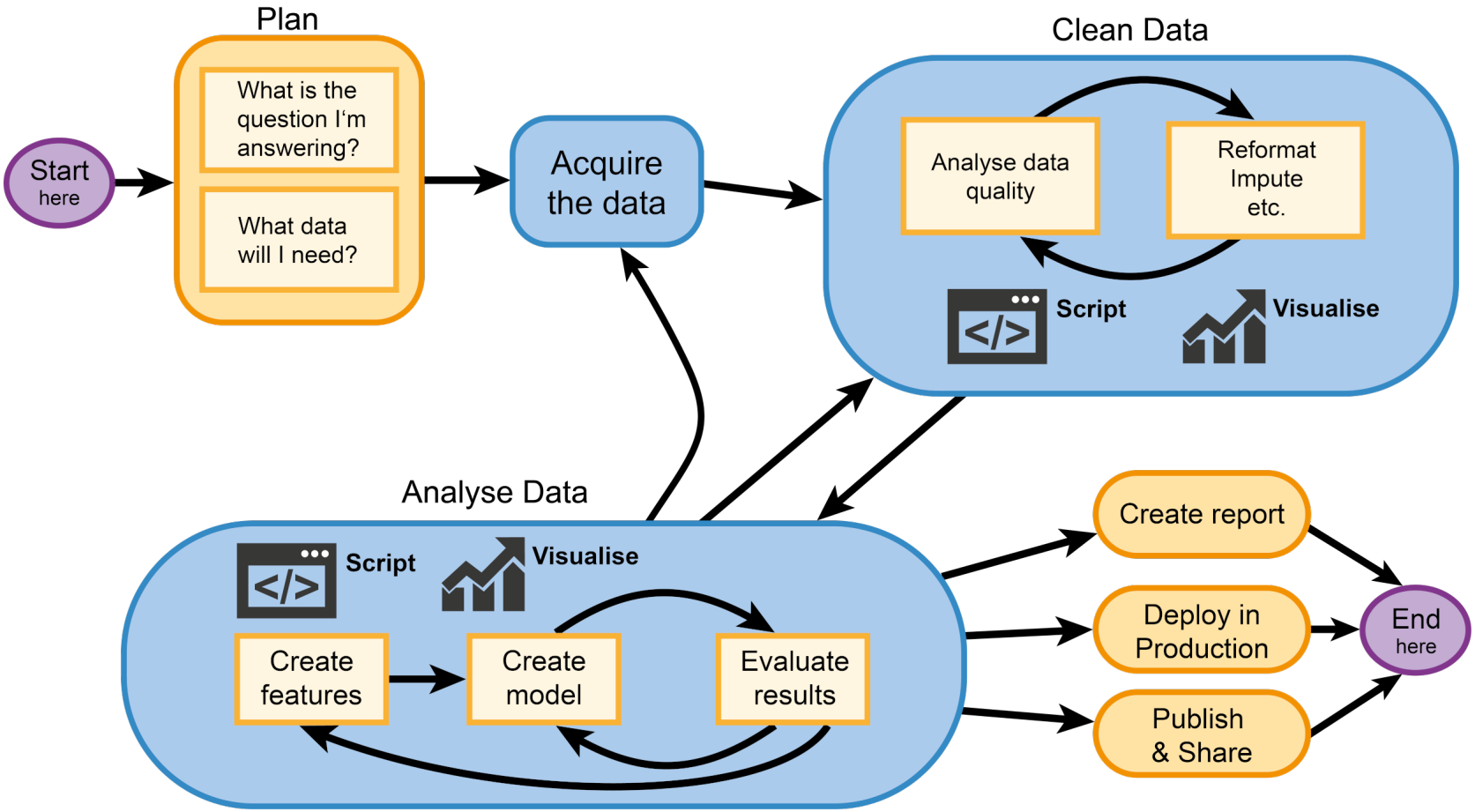
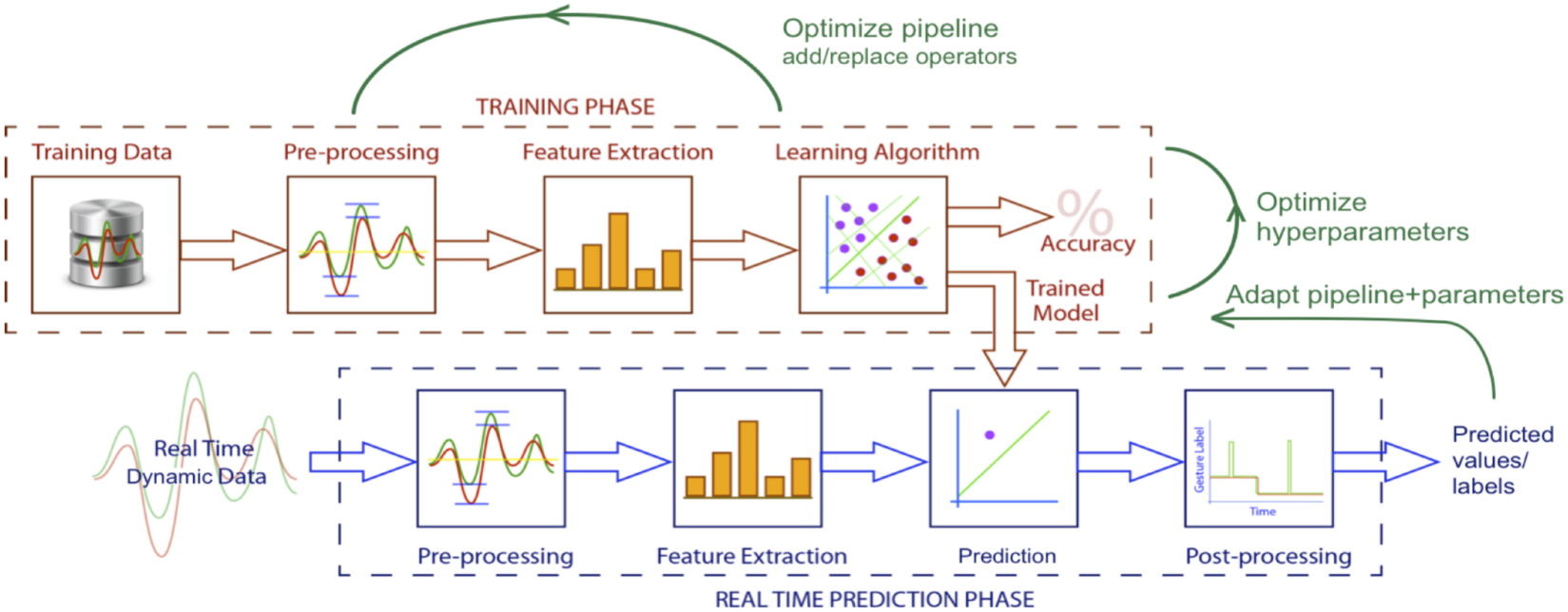
MACHINE LEARNING WORKFLOW



MACHINE LEARNING PIPELINES



Choose pipeline structure and optimize pipeline parameters w.r.t. the estimated prediction error, on an independent test set, or measured by cross-validation.

IMPORTANT TYPES OF FEATURE ENGINEERING

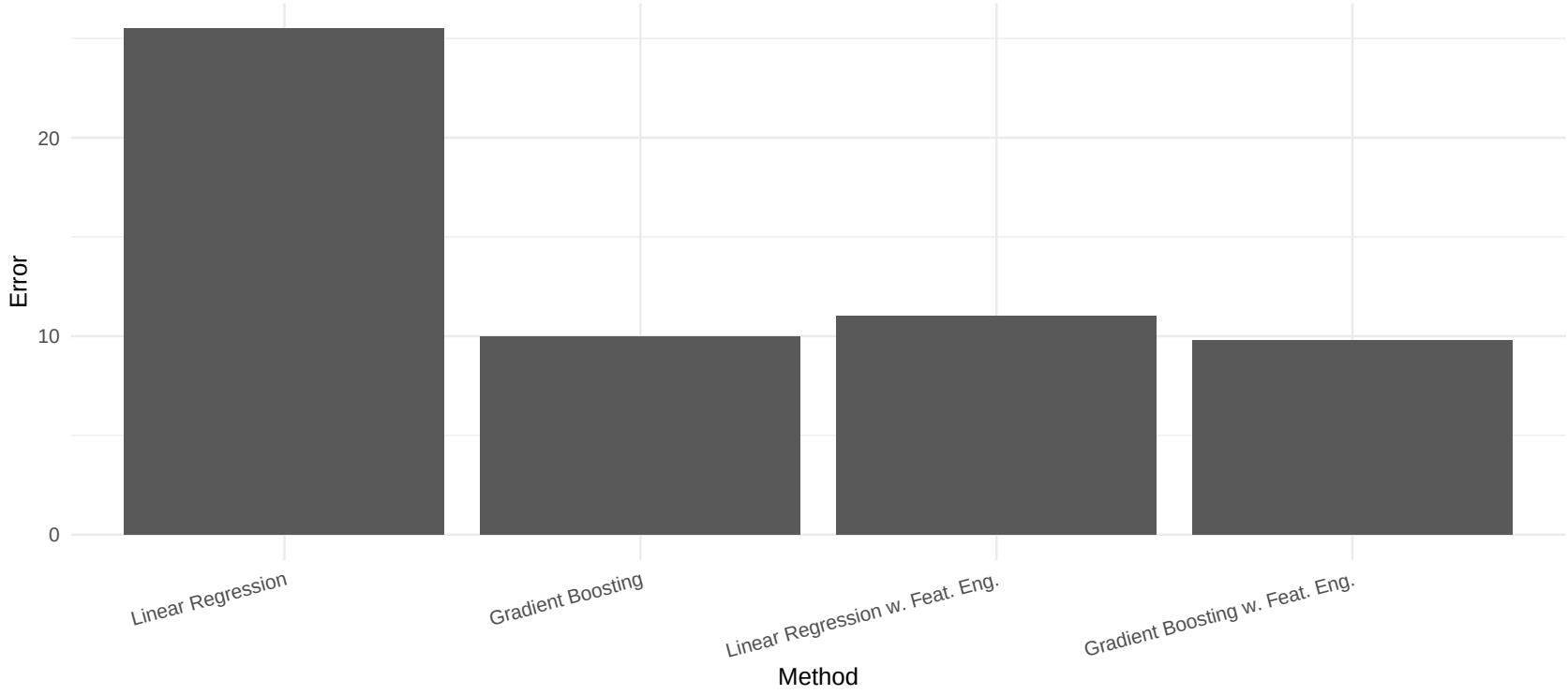
Feature engineering is on the intersection of **data cleaning**, **feature creation** and **feature selection**.

The goal is to solve common difficulties in data science projects, like

- skewed/*weird* feature distributions,
- (high cardinality) categorical features,
- functional (temporal) features,
- missing observations,
- high dimensional data,
- ...

and **improve model performance**.

WHY FEATURE ENGINEERING IS IMPORTANT



Choice between a simple **interpretable** model with feature engineering or a complex model without.

FEATURE ENGINEERING AND DEEP LEARNING

One argument for deep learning is often the idea of *“automatic feature engineering”*, i.e., that no further preprocessing steps are necessary.

This is mainly true for special types of data like

- Images
- Texts
- Curves/Sequences

Many feature engineering problems for regular **tabular** data are not solved by deep learning.

Furthermore, choosing the architecture and learning hyperparameters poses its own new challenges.