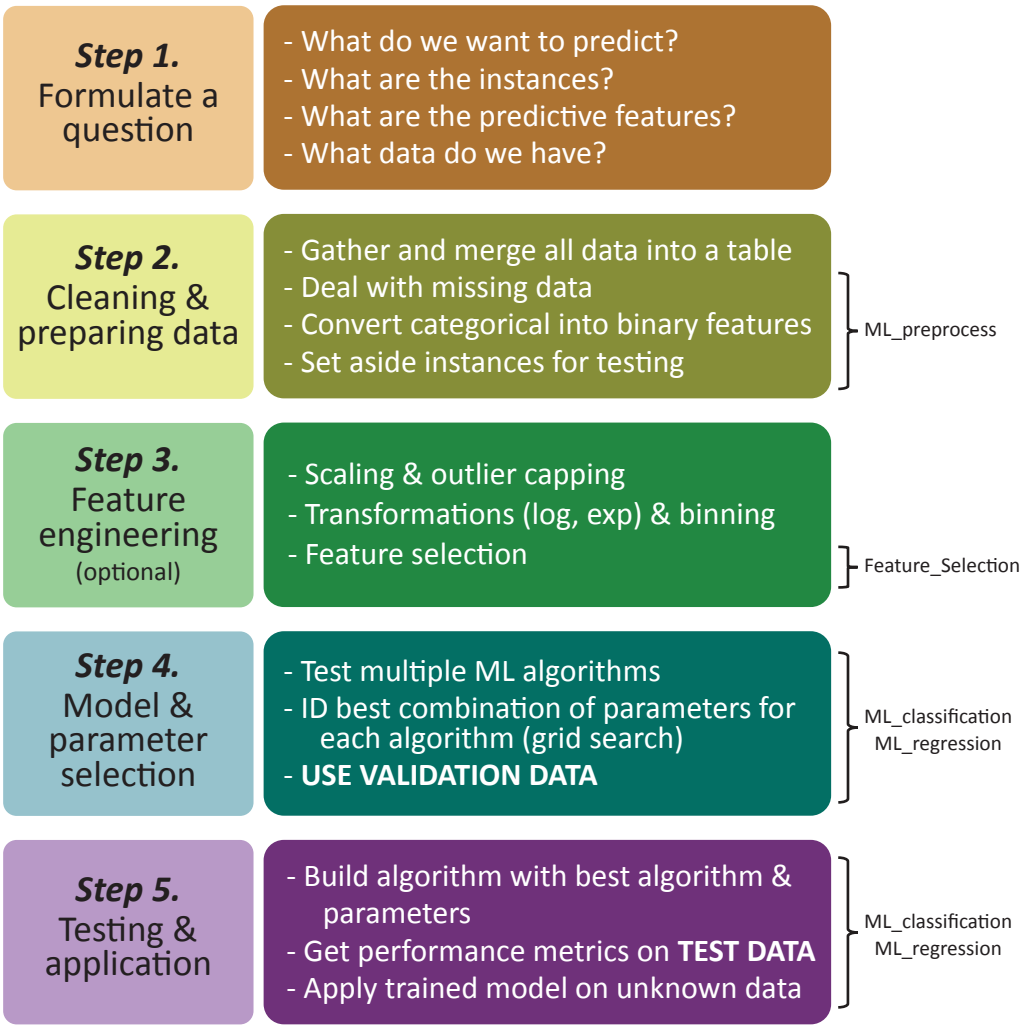
Step 1.
Formulate a question
- What do we want to predict?
- What are the instances?
- What are the predictive features?
- What data do we have?
Step 2.
Cleaning & preparing data
- Gather and merge all data into a table
- Deal with missing data
- Convert categorical into binary features
- Set aside instances for testing
ML_preprocess
Step 3.
Feature engineering
(optional)
- Scaling & outlier capping
- Transformations (log, exp) & binning
- Feature selection
Feature_Selection
Step 4.
Model & parameter selection
- Test multiple ML algorithms
- ID best combination of parameters for each algorithm (grid search)
- USE VALIDATION DATA
ML_classification
ML_regression
Step 5.
Testing & application
- Build algorithm with best algorithm & parameters
- Get performance metrics on TEST DATA
- Apply trained model on unknown data
ML_classification
ML_regression