Lecture 5
Pipeline

Dr. Jochen Cremer
Dr. Pedro P. Vergara
Dr. Simon Tindemans

TU Delft

Pipeline and Machine Learning Workflows

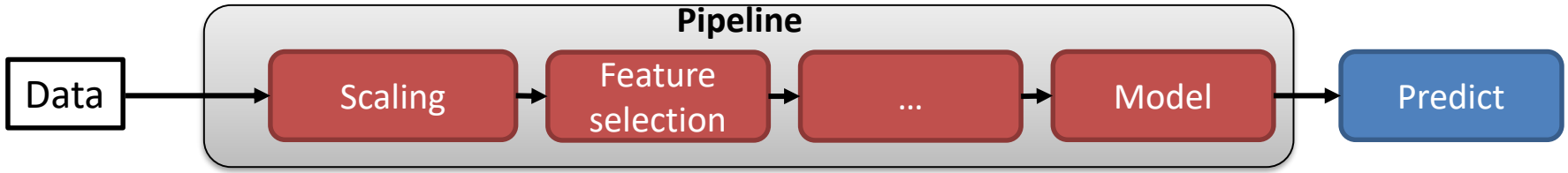
- There are standard **workflows** in a machine learning project that can be **automated**.
- In scikit-learn, **Pipeline** helps to define and automate these workflows.
- Useful when we have a fixed sequence of steps.

Photo from here.

TU Delft https://scikit-learn.org/stable/modules/compose.html#pipeline

Pipeline and Machine Learning Workflows

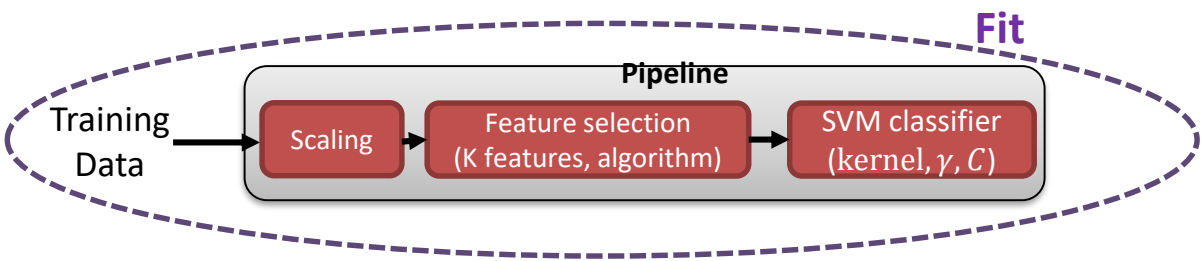
- Pipeline can be used to chain a fixed sequence of steps together.
- Example:

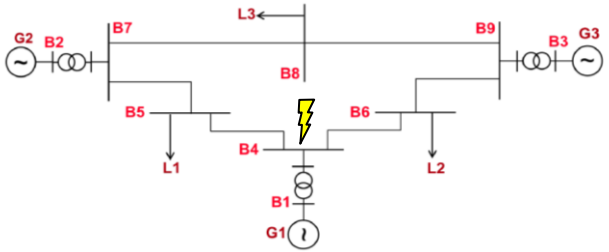


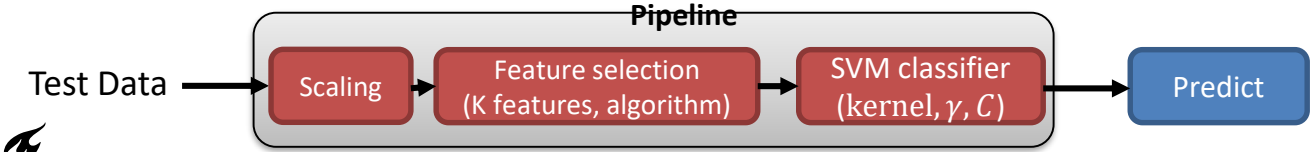
- Advantages:
  1. Convenience: You only have to call fit and predict once on your data to fit a whole sequence of estimators.
  2. Joint parameter selection: You can grid search over parameters of all estimators in the pipeline at once.
  3. Safety: avoid leaking data from your test data into the trained model

Example: Security Assessment

- Suppose we wish to predict security in the system using a binary classifier.
- **Desired output: 0**: secure, **1**: unsecure.
- **Input:** Operational variables of system: (i.e., $V_n, \theta_n, P_G, Q_G, P_D, Q_D, P_{nm}, Q_{nm}$)
- We have **1000 samples** and **100 features**.

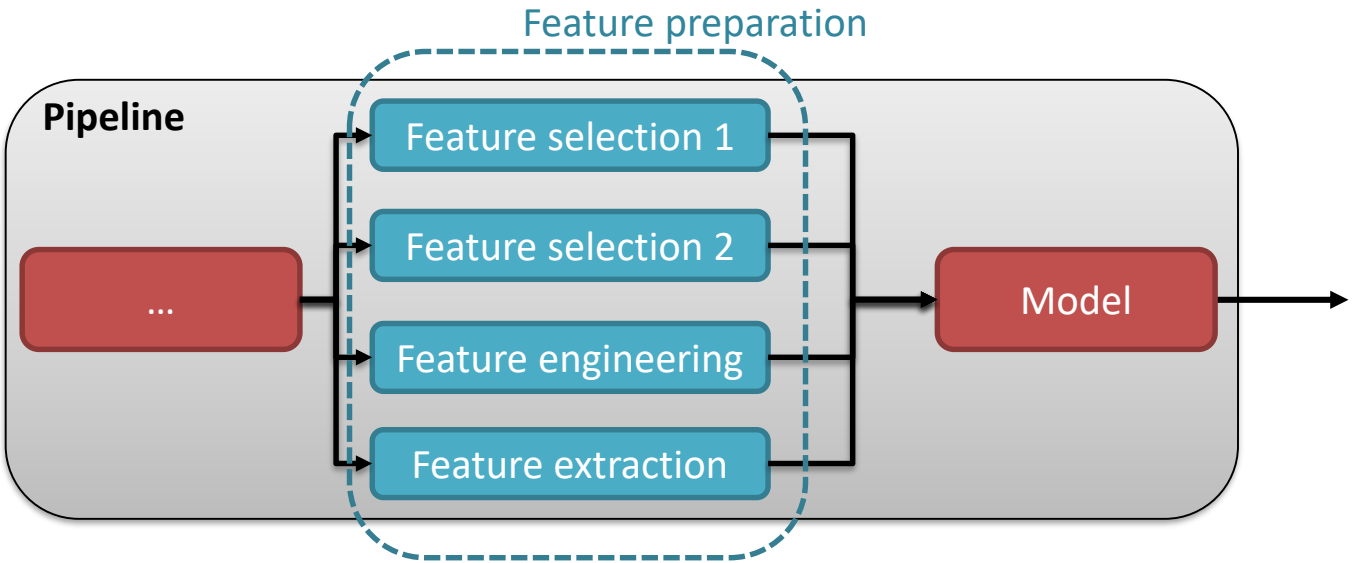






TU Delft

Feature Union



https://scikit-learn.org/stable/modules/compose.html#feature-union
https://scikit-learn.org/stable/modules/compose.html#pipeline

TU Delft

Code

- Data (**1000 samples** and **100 features**)
- Train/test split