› PROJECT 1: NAIVE BAYES CLASSIFIER

Praktikum maschinelles Lernen

Friedrich Carrle
Prof. Alexander Windberger
Wintersemester 2023/24

PROJECT 1: NAIVE BAYES CLASSIFIER

1.   (5 Points) Implement a Python class `NaiveBayes`.

   >   The class contains an __init__, fit, predict_probability, and evaluate_on_data method
   >   The implementation should be able to deal with, both, discrete and continuous features. A template file with correspondant signatures and documentation will be handed out.
   >   Feel free to implement as many other methodes as you need.
   >   **The only package you're allowed to import for this task is Pandas**

PROJECT 1: NAIVE BAYES CLASSIFIER

2. (2 Points) Evaluate your Naive Bayes implementation using the „inflammation_diagnosis.csv" data set.

   > Combine the columns „inflammation" and „nephritis" to a new column „deasease" which is „False" if „inflammation" and „nephritis" is „False" and „True" otherwise. Use this column as target for the classifer

   > Split the data into a train and a test set using 20% of the data for the test set. Show that no train data samples are contained in the test set.

   > Fit the Naive Bayes classifier using the train set and evaluate it in the test set.

   > Calculate and plot the confusion matrix.

   > **You're only allowed to import your NaiveBayes class, Pandas, and Matplotlib (or Seaborn) for this task**

PROJECT 1: NAIVE BAYES CLASSIFIER

- The submission consists of two files:

1. A file „naiver_bayes_klassifizierer.py" containing the NaiveBayes class implementation (task 1)
2. A Jupyter Notebook containing the preprocessing, the training, and the evaluation of the Naive Bayes classifier (task 2)

- Due Date is **October 27th**

- **Remember: There are 3 Points for presentation, style, and creativity**
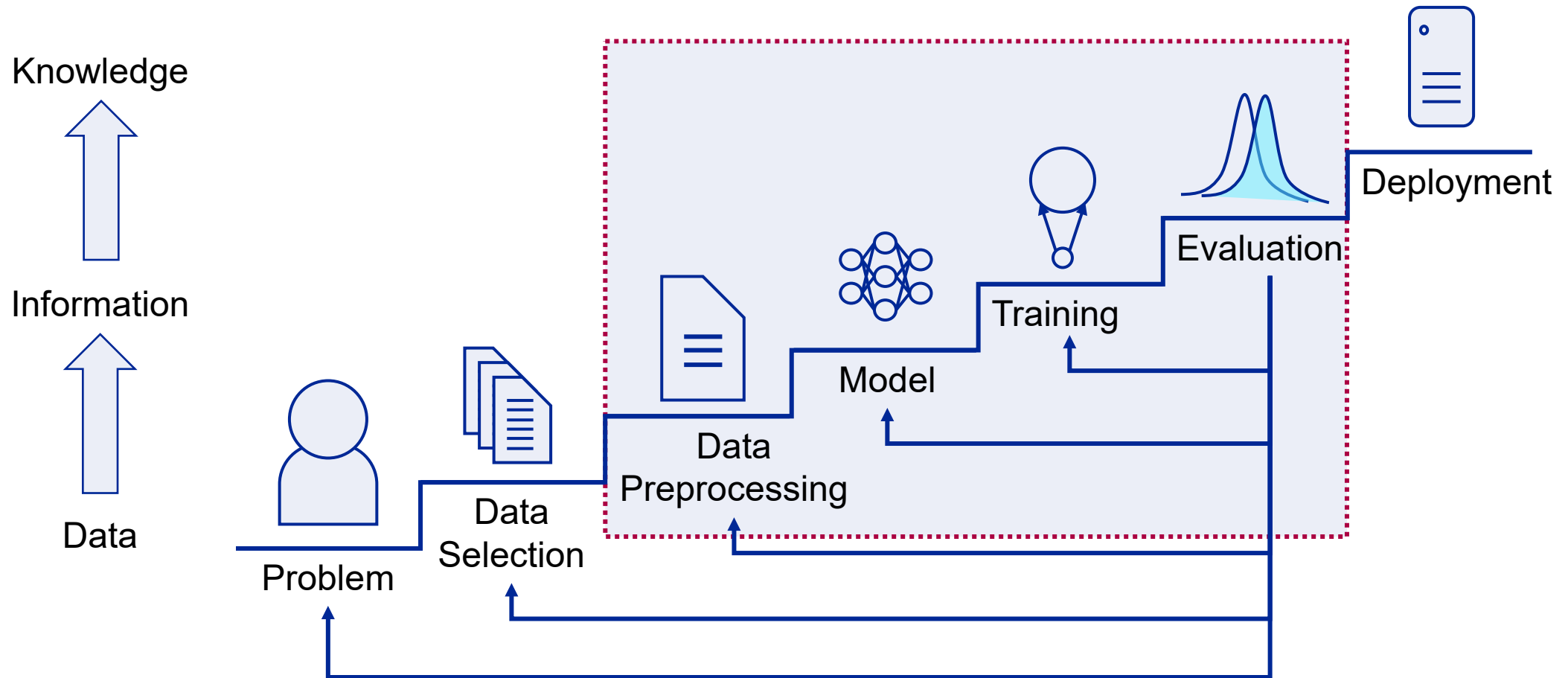
THE TOOLS
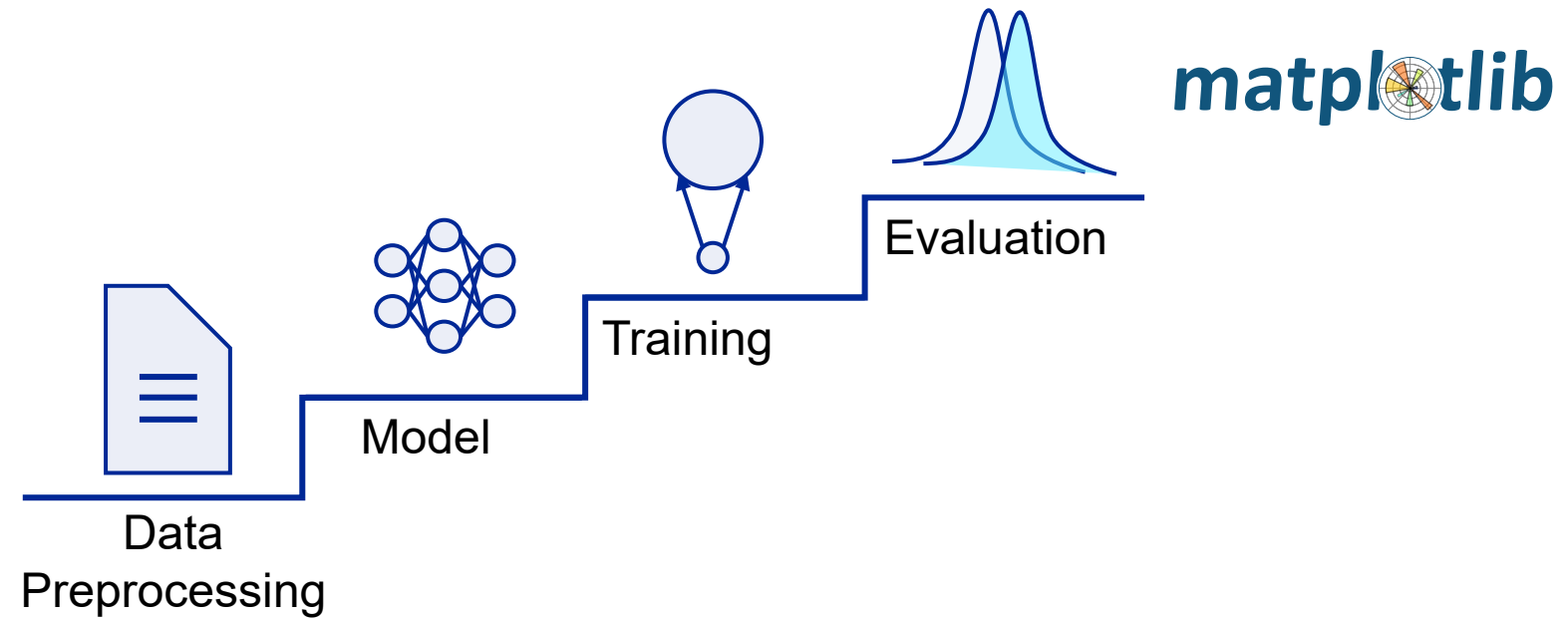
WHAT'S TO BE LEARNED

- **Learn how to extract information from data**

  - Deal with different data types (structured and unstructured)

  - ML algorithms

  - Explore

  - Statistics

  - Linear algebra (why live in 3 dimensions, when you can master 1000s?)

  - Use analysis to find the sweet spots

WHAT'S TO BE LEARNED

- **Learn how to interpret your findings**

  - What are the methods and metrics?
  - What correlations can be found?
  - What's significant and what's not?

WHAT'S TO BE LEARNED

- **Communicate and present your results**

  - Make pretty graphs

  - Your data contains a story. Tell it!

  - Make mistaces, harness criticism, defend your theses

  - Convince your peers, boss, customer, client, referee …