

## ML pipeline

- ① Data Ingestion / Read the dataset
- ② EDA / Data cleaning.
- ③ Preparing the data and feature engineering.
- ④ divide the data into  $X$  and  $Y$ 
  - ↓ Independent variable
  - ↓ dependent variable.
- ⑤ Train - test split
- ⑥ Scaling the data (Standardisation)
  - train 70% ←
  - test 30% ←
  - don't scale  $Y$  (target)
  - only scale the independent variable
    - ↓
    - Scaling categorical columns should be avoided.
- ⑦ Model training
- ⑧ Model Evaluation.