

ML pipeline

- ① Data ingestion | Read the dataset
- ② EDA
- ③ Data preparation | Data cleaning | Feature Engineering
- ④ Divide the data into X & Y
 \downarrow Independent variable \uparrow dependent variable.
- ⑤ Train-test split.

train 70
test 30 []
- ⑥ Scaling the data
(Most of ML — Standardization)

\rightarrow don't scale Y (target)
 \rightarrow only scale independent variable
- ⑦ Model training
- ⑧ Model Evaluation

Scaling Categorical Column should be avoided.