

# STEPS IN MAKING ML MODEL

16

Data Preprocessing

DATA

Remove NULL VALUE + Duplicate Rows

CHECK BALANCED DATA OR NOT

EDA

- Scatter Plot
- Boxplot (outlier)
- Pair Plot (Linear or Non Linear)
- Histogram
- Correlation (Heatmap)

Data Preprocessing

REMOVE OUTLIER (IF NECESSARY)

→ you also need to check whether or not problem outlier are necessary

Data Preprocessing (FEATURE ENGINEERING)

FEATURE SCALING (Standardization or Normalization) (If Necessary)

Have To Do on Features not on target

Data Preprocessing (FEATURE ENGINEERING)

FEATURE ENCODING (If categorical variables)



# FEATURE ENGINEERING

FEATURE  
SELECTION

→ Remove By Threshold  
which don't contribute  
-ute to Pred.

2 → Remove Correlated  
Feature  
i) Apply PCA  
ii) Remove Col.

Split Data into  
Input & O/p

TRAIN TEST  
SPLIT

MODEL SELECTION

HYPERPARAMETER  
TUNE

TRAIN

TEST

PERFORMANCE METRICS

MODEL DEPLOYMENT