Chapter -3- Machine Learning Pipelines

Level-0

One of the most important thing to approach Machine Learning problems in ML pipeline is to ask the right question that indicates the output of the model.

Suppose, you want to lend some money to your friend but you are also concerned whether he or she will be able to pay you back within time. So, which of the following primary question (question can change when you move on to build your model) would be suitable for building a ML model you think. You can also add your own question if you think it is more suitable.

1. Should I provide loan to him/her?
2. What aspects should I consider before providing loan to him/her?
3. What available features need to be considered to make a decision to provide the loan to him/her?

Level-1

Data can be in various format. Use the following link to access the dataset "loan_prediction"

Dataset_00

1. Using general insight do you think all the features need to be considered for predicting whether a specific person would be provided loan or not?
2. What is the most useless (disallowed) feature of this dataset for loan prediction, you think? And why? If you could identify any such unnecessary column, please drop the column before moving forward. Otherwise, you can use as it is.

Level-2

If you think the dataset needs to be prepared before deploying to a Machine Learning model (like removing rows with missing values), then apply necessary python instructions to amend the dataset.

Level-3

1. Split the dataset into 80% training data and 20% testing data for evaluating and selecting a Machine Learning model.
2. Use two popular and basic ML model such as Logistic Regression and Decision tree classifier for predicting loan status.
3. Finally, select the ML model which has a better performance rate, for loan status prediction.