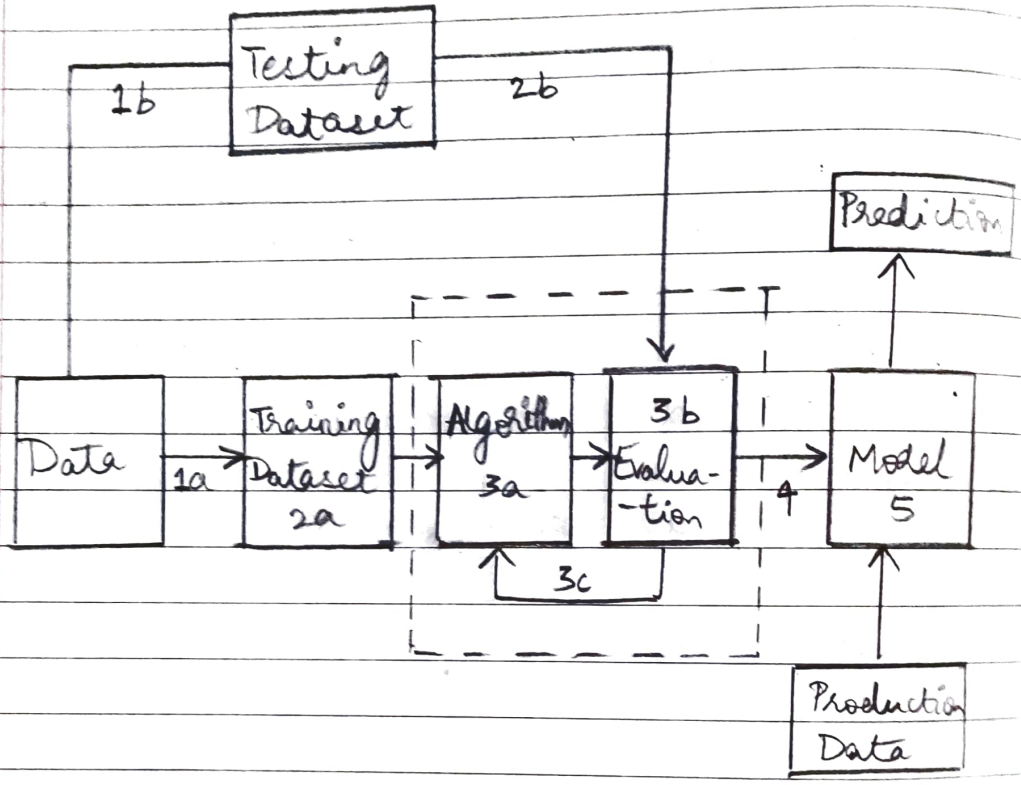
20181CSE0621
Sai Ram

03/06/2021

20181CSE0621
Sai Ram
6-CSE-10

Part-B

Q.3] MACHINE LEARNING WORKFLOW



→ The stages in the ~~workflow~~ workflow are:

1. Gathering Data
2. Data Pre-processing
3. Researching the model ~~for the~~ that will be best
4. Training & testing the model
5. Evaluation.

1] Gathering data: The process of gathering data depends on the type of project we want to do. If it's a real time project we use IoT sensors to constantly monitor & gather data. In the other case we

have some free to use datasets from Kaggle, UCI machine learning repository & so on which can be used.

2] **Data Pre Processing**: It is the process of cleaning raw data. The raw data is not feasible for analysis and hence we clean it to get insights. The various process are filling missing values, outlier handling and also labelling. Operations done are:
- Filling null values
- Outlier handling
- Data wrangling.

3] **Researching the model to be used:**

As we have the clean data we use the data to first decide what task needs to be done. i.e. Supervised, unsupervised or reinforcement learning.

- If the output is continous we use the regression model. like: Linear Regression
  Polynomial Regression
  GLM etc.
- If the output is categorical we use the classification model like: KNN
  Decision tree
  Random Forest etc.
- Unsupervised techniques like K-means, fuzzy c-means are also used.

4] **Training & testing the model:**

- The training of model involves applying the selected model onto training dataset to make the model familiar with features.

2018ICSE0621
Sai Ram.

- The testing is performed in order to check what kind of behaviours the model has with unseen data.

5] Evaluation:

We have this stage to know how well the model is performing. We evaluate the model based on metrics like R^2, adj. R^2. Confusion matrix and so on.