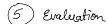
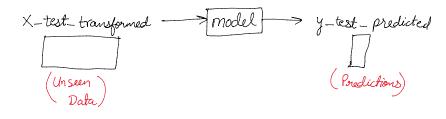
Hands-On Sklearn Module Thursday, May 18, 2023 3:32 PM Train-Test split is performed 1 Data -> Separate the input & output -> (df) nxd So that you can evaluate your model on useen data nxd-1 and get a true idea of how well your model (an handle real world > Split the data into train 4 test data, before you give it to client. percent wise split, use any one. Using Beth is redundant coding (3)+ Data X\_train y-train Data Transformation Cleaning X-test-transformed y-test Data  $\times$  - testData Transformation Cleaning  $\prod$ Lategorical Numerical - One Hot Encading - Flattening Standardization - Lahel Encoding -TF-IDF We apply transformation on numerical data in order Non-numerical to to rescale it to a common rumerical Scale.

Algo

(4) Model Building

X-train-transformed





Iris Pataset Example

Then performing a train test split, the following splits care commonly used: 70-30 for small 75-25 dataset

Text → BOW, TF-IDF

Image -> flattening