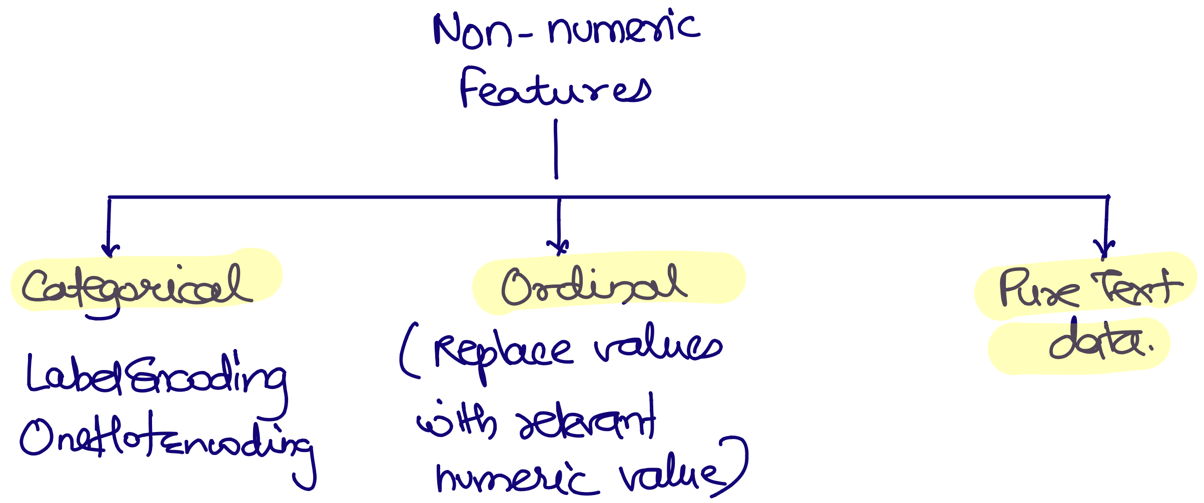
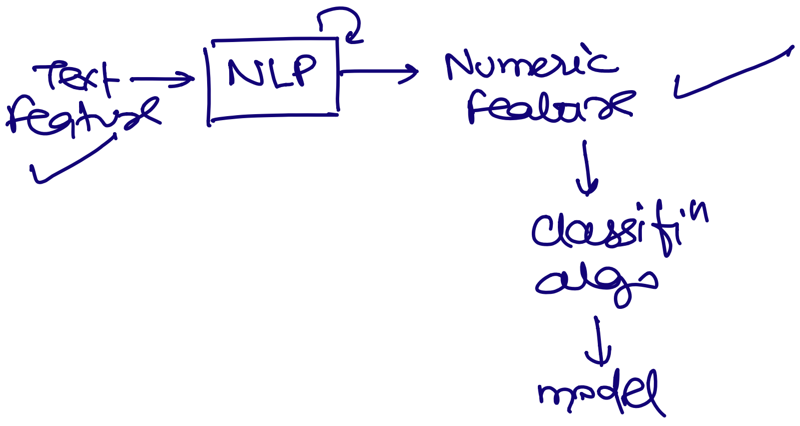
Agenda.

NLP- Natural Language Processing



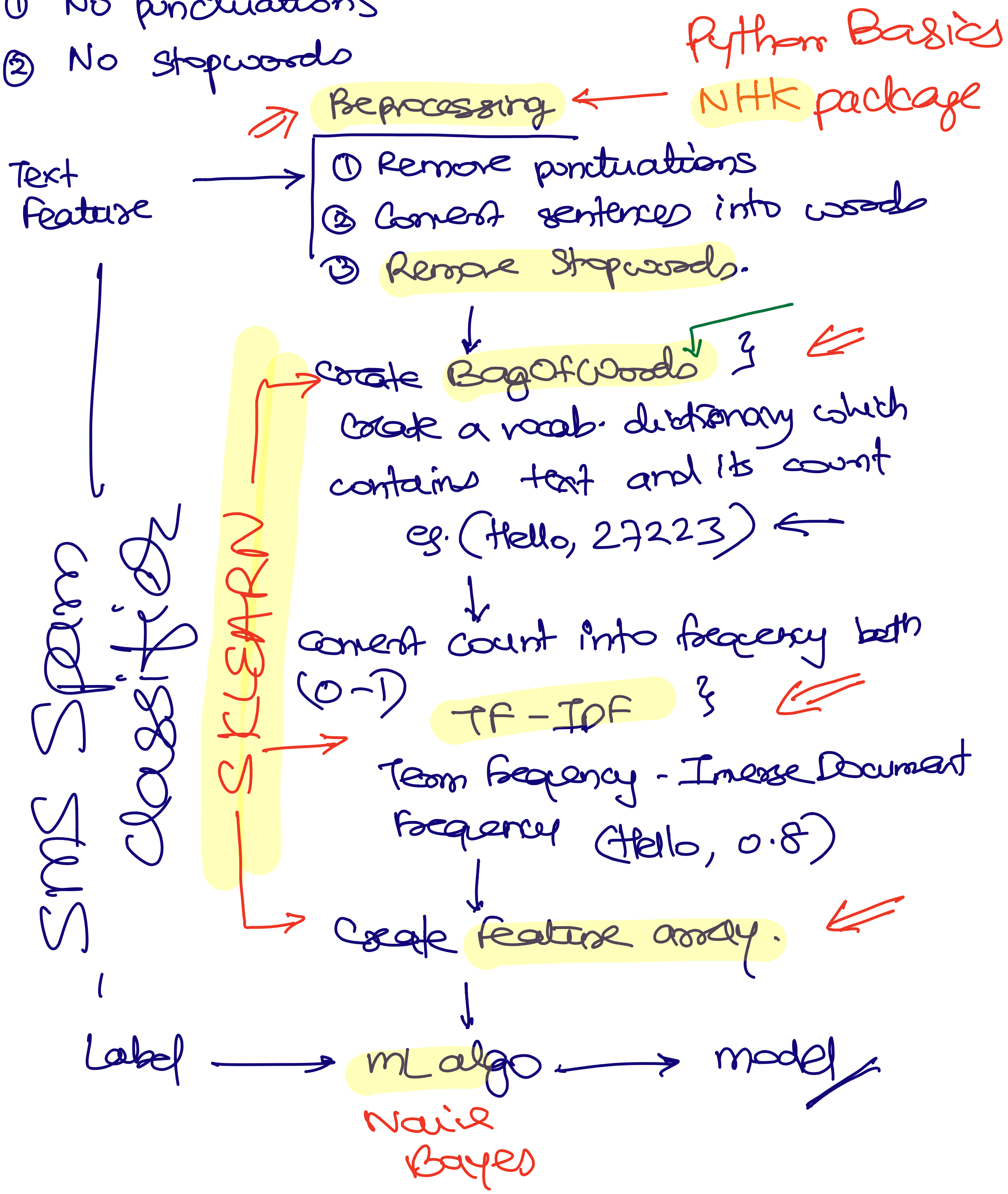
The scope of NLP is to handle pure txt data.



NLP program rules

① No punctuations

② No stopwords



"win
Lottery
guaranteed" → model → Spam

eg:

dataset = ["welcome to Simplilearn",
"welcome welcome welcome",
"simplilearn, Simplilearn welcome"]

↓

① Preprocess.

(welcome, simplilearn)

(welcome, welcome, welcome)

(simplilearn, Simplilearn, Simplilearn)

↓ BOW → Vocab. Dict

→ (Welcome, 1, 4) (simplilearn, 1, 4)

→ (welcome, 3, 4)

→ (Simplilearn, 3, 4)

welcome, 4
simplilearn, 4

sparse matrix

Convert BOW into freq using TFIDF

↓

TFIDF Transformer

→ fit(bow) →

{ (welcome 1/4) (Simplilearn, 1/4)
(welcome, 3/4)
(simplilearn, 3/4) sparse matrix

←

feature

0.25, 0.25
0.75
0.75

sparse matrix