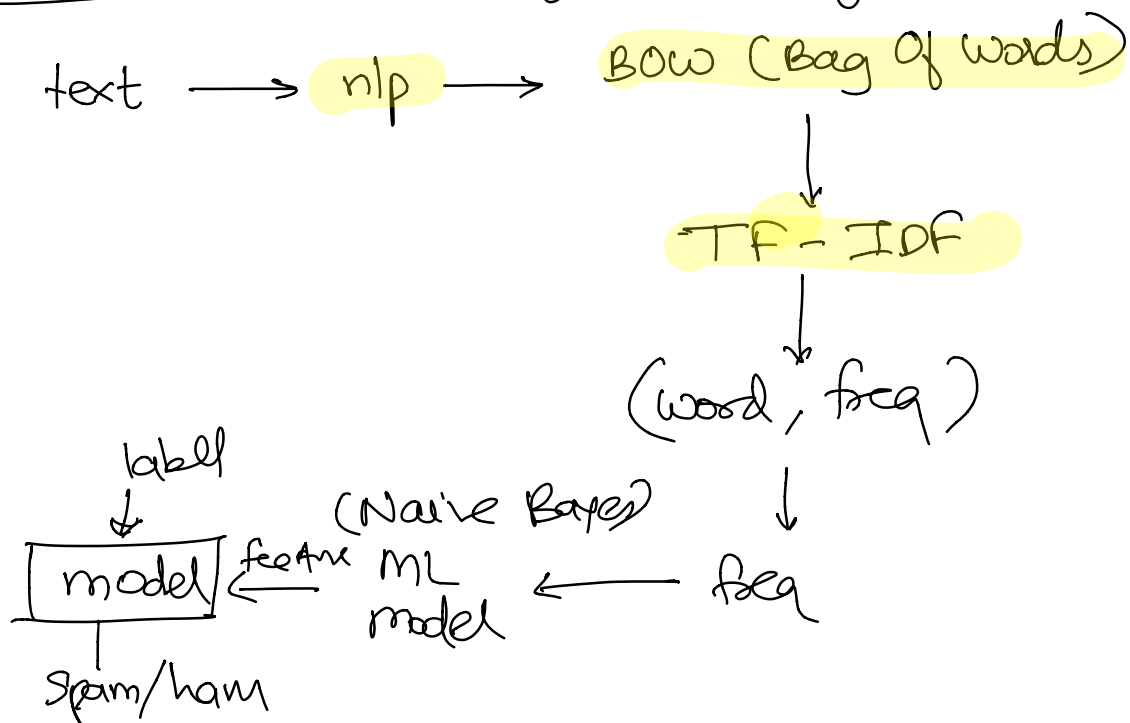


NLP → Natural Language Processing



Two packages:

- ① **nltk** (Natural Language Processing Toolkit)
- ② **sklearn** (nlp + ml model)
- ③ **spacy** (deep learning)

Working with Text

Preprocessing

text preprocessing

- No punctuations
- Remove stopwords
- Convert sentences to words

Bag of words

TF-IDF

features

(Compatible for any ml classification algo)

Logistic Regression, K-NN, Naive Bayes, DTC, RFE, SVC

		predicted	
		ham	spam
actual	ham	4825 ¹	0 ²
	spam	115 ³	632 ⁴

Client is fine
spam → Inbox

ham → Spam

$P_{ham} \rightarrow 0.98$

$R_{spam} \rightarrow 0.85$

Naive Bayes algo

(Base of this algo is Bayes Theorem which $\text{prob}(A|B)$)

