

Pandas → data cleaning
Numpy → maths operations
Sklearn → machine Learning

How Count-Vectorizer Work.

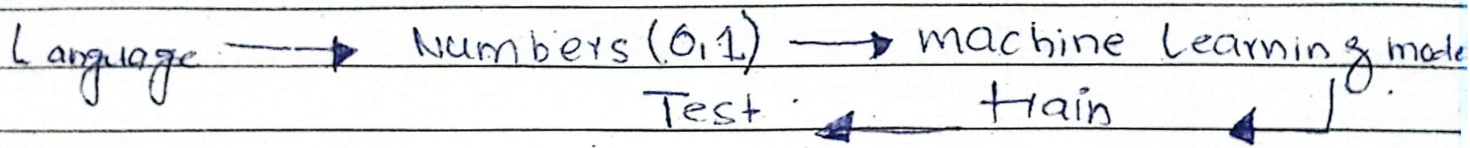
First count-vectorizer extract the unique words in the sentence and make array of these words. This technique known as bag of words.

I Love my country → [I, Love, my, Country
I love my home home]

And then change every unique word in the Number.

Bag of Words: only work on the words, not depend on the grammar, ordering.

Flow of Project

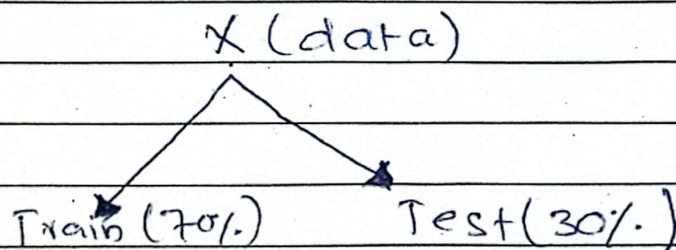


Work of train test split

Use to train, test the model.

- * We use Naive Bayes Algorithm use to Language detection problem
 \rightarrow also use in the spam email, where it detect the words and put our email in the Spam.

ML Model



- 70% data we use for training
- 30% data we use for testing which is unique for the machine