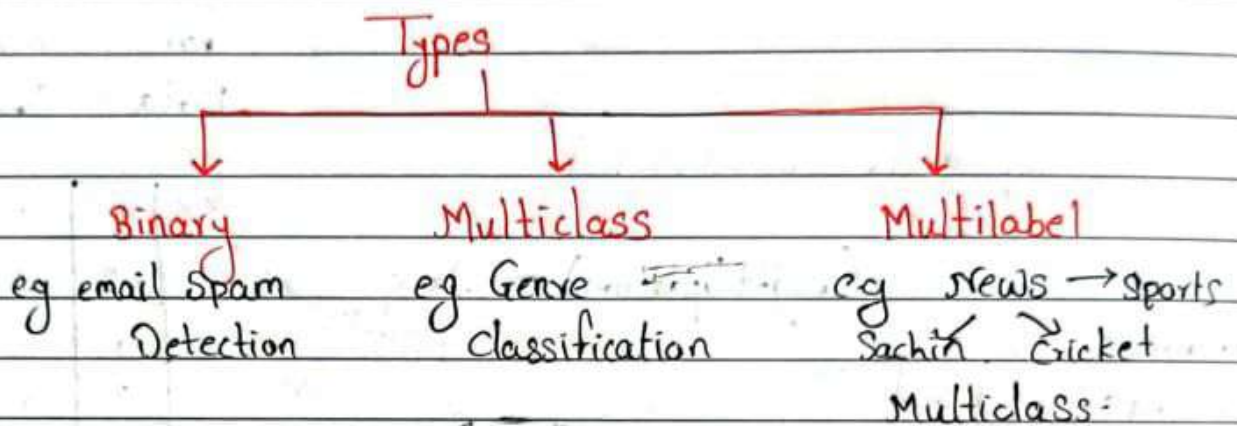


* Text Classification:

→ Classification done of textual data is known as text classification.

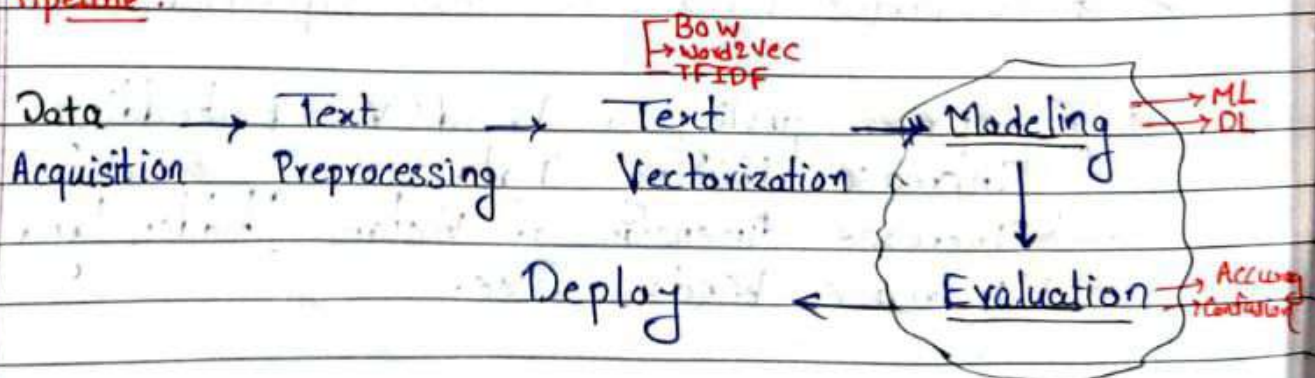
- ↳ 1) Email Spam Detection
- 2) Sentiment Analysis (Positive / Negative)
- 3) message (Sales / Support)



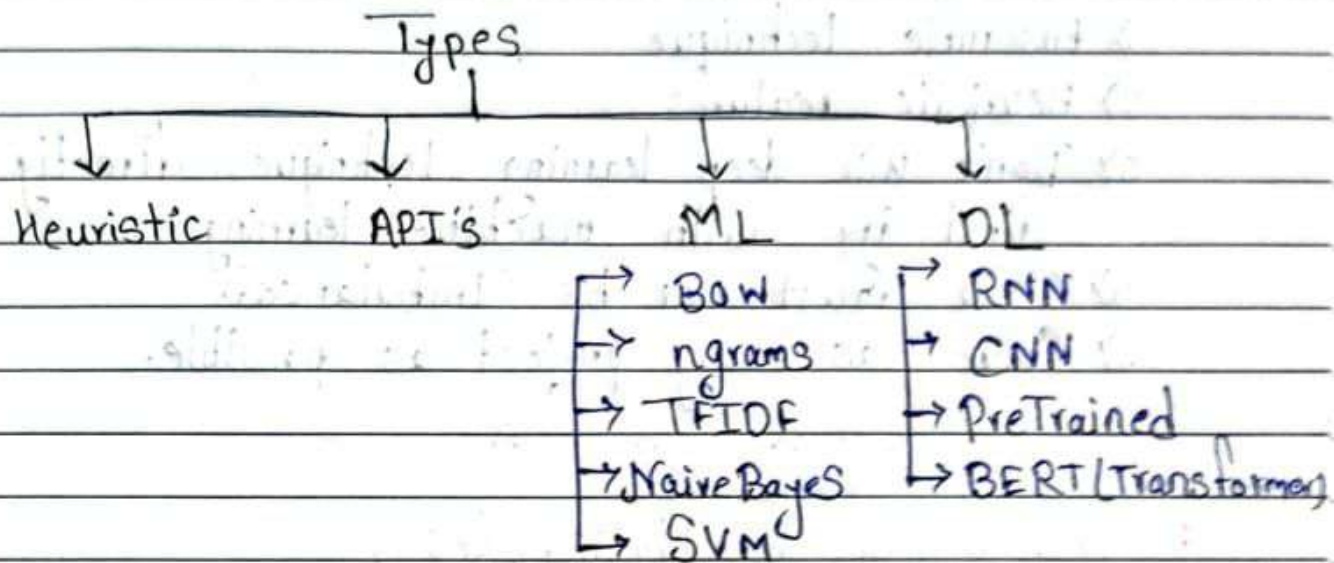
* Applications:

- i) Email Spam Detection
- ii) Customer Support.
- iii) Message Sales / Support Classification
- iv) Sentiment Analysis.
- v) Language Detection
- vi) Fake news Detection

* Pipeline:



* Different Approaches:



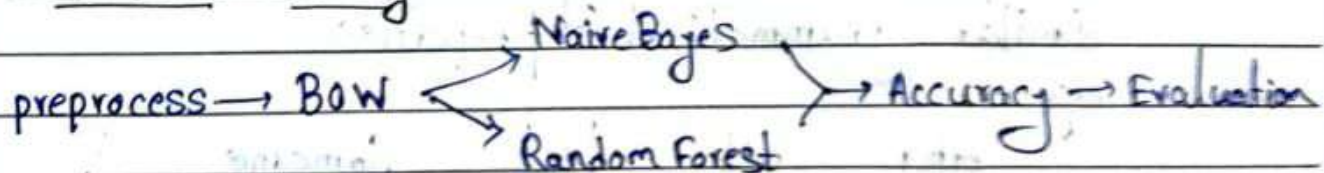
1) Heuristic Approach:

- Creating your own rule and Algorithm.
- Used when there is no data for Training.

2) API's:

- Google Cloud Platform } API
- Amazon Web Services } (AWS Cloud.io)

3) Machine Learning



- you can use n-gram instead of BOW.
- TF-IDF can improve accuracy due to vectorization.

4) Deep Learning

A) Pre-trained Word2Vec model:

- Atleast 80% Similarity between Vocabulary used in pretrained model and our data vocabulary.

B) We need enough data to train our Word2Vec model.

* Practical Advise:

- 1) Ensemble Technique
- 2) Heuristic features
- 3) Don't use deep learning techniques directly instead first try with machine learning
- 4) Data must not be Imbalanced.
- 5) Solve as many project as possible.

— x — x — x —