Corpus Preparation

* Getting the duplicates in the dataset
* Python code

Data\_set.duplicated()

* Output

0 False

1 False

2 False

3 False

4 False

...

4596 False

4597 False

4598 False

4599 False

4600 False

Length: 4601, dtype: bool

* Removing the duplicate values
  + Python code

Data\_set.drop\_duplicates(inplace=True)

Relevance

Compatibility

Implementation

* Code quality (usage of libraries, frameworks, tools etc.)
* Readability (code commenting etc.)
* GIT usage
* Commit history.

Experiments

* Experimental results
* Evaluation criteria
* Evaluation metrics

Limitations & Enhancements

* Limitations
  + Due to the dataset being small, it may affect the accuracy of the models
* Future enhancements
  + Training the models with a larger dataset to improve accuracy.