Language Classification Report

Report on the implementation of Language Classification project

1. Libraries that I have Used

* Scikit Learn
* Pandas
* Numpy

1. Model Architecture

* Input data from language classification spreadsheet.
* Split the data into train dataset and test dataset.
* Used Gaussians Naïve Bayes machine learning algorithm.
* Train the model with training datasets.
* Make predictions with the testing datasets.
* Evaluate the model in terms of how accurate is the model.

1. Splitting data

* I split the data into 60% training dataset and 40% testing dataset.
* I used the training dataset to induce the model.
* Make predictions with the testing dataset.
* I split the dataset to avoid overfitting, a model evaluated on a training dataset is more likely to have higher accuracy than a model evaluated on a testing dataset.

1. Training Process

* Used the Gaussian Naïve Bayes Algorithm
* Used the training dataset to train the model