**HW 7- NLP( Natural Language Processing) Report**

Name: Rudraksh Sugandhi ([sugar01@pfw.edu](mailto:sugar01@pfw.edu))

Observation:

1. For the data Set I have used IMDB dataset from Hugging face for which the output is either 0 or 1 where 0- Negative sentiment and 1 means positive sentiment.
2. To pass the dataset to. Model I have converted into csv format and to string format to make prediction.
3. I have used 100 positive sample and 100 negative sample and reshuffle the data to make the prediction more random.
4. While converting the dataset to csv I have removed the column which have none value so that it doesn’t create bias in prediction.
5. I have Used three model from hugging face which are:
   1. distilbert-base-uncased-finetuned-sst-2-english
   2. lvwerra/distilbert-imdb
   3. aychang/roberta-base-imdb
6. As the distilbert-base-uncased-finetuned-sst-2-english model can only work on strings of length 512 tokens or sequence, passing the list of strings reduces the length of the string sequence.
7. One of the issues I faced was trying to implement different type of models I tried using mnoukhov/gpt2-imdb-sentiment-classifier model but it gives value in label\_0 label 1, and label\_2 which is tough to convert into two labels 0 or 1 of just positive and negative.
8. Summary of all the three models :

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Accuracy** | **Precision** | **Recall** |
| **distilbert-base-uncased-finetuned-sst-2-english** | **0.86** | **0.86** | **0.85** |
| **distilbert-imdb** | **0.90** | **0.9090** | **0.9** |
| **roberta-base-imdb** | **0.93** | **0.921** | **0.94** |

1. The first two model give the result or prediction in form of list of dictionary where it give one label i.e. is POSITVE and another is score which is in form of float values.
2. The third Roberta model also give values in the form labels and score but it give output as label: pos or label: neg.
3. Manually converted all the POSTIVE and pos values to 1 and NEGATIVE or neg value to 0 to make get the y\_pred value to compare it to y\_true value to make all the accuracy matrix.

**Limitation**: The models which I used are giving the output in the form of Positive or negative but not any other label but there are different models which give more specific answer in the form of multiclass classification. My models are all the trained on imdb dataset so its parameter are autotune to give good accuracy on the same dataset. If we change the dataset to something else for classification this model will give less accuracy.

**Future Work Impact** : I Can directly use this pretrained model to make prediction and create end-end project by just calling the model in my code and make prediction instead of training model on backend and spending time to train the model on my GPU which anyway not going to give high end accuracy.