

## Individual Report

The project that my group worked on is Sentiment Analysis for Yelp Reviews using neural networks. We decided to use different kinds of models to train and classify the reviews. The project was divided into traditional LSTM model and pre-trained models and their results were compared.

I worked completely on the pre-trained models. I wrote the entire two python files namely, Transformers\_train.py and Transformers\_test.py. I also created the shell file to download the data for Mac users. For the train and test files, I used the basic code structure from the Hugging Face website. I started modifying and adding lines to accommodate our data. I tried to train and test 9 different pre-trained models including one ensemble model. Unfortunately, due to AWS and VPN limitations, 3 of these models could not be trained. But all the other models performed well.

For the training of the models, I used only the first 140,000 datapoints of the training set, and used only 3 epochs due to the restriction of AWS and VPN. The training set was split into training and validation sets. Both the set were tokenized using the tokenizer corresponding to each of the model used for the training. Then, I used the data loaders, for training and validation set, mentioned in the Hugging Face website to load the data in batches for training and validation. Next, I set the optimizer as AdamW and criterion as Cross Entropy Loss. I used the criterion only for ensemble model. Following this was the actual training of the models using the training set and calculating the performance of the models on the validation set. The models were then saved for the test file to load and classify the test set.

The results of the remaining models were very good. Flaubert model had the weakest performance among the models with accuracy around 94% whereas the rest of the models had accuracy around 96%. The table below shows the models used and their accuracies.

| MODEL                               | ACCURACY |
|-------------------------------------|----------|
| DistilBert                          | 0.9633   |
| Electra                             | 0.9655   |
| DistilRoberta                       | 0.9672   |
| MobileBert                          | 0.9645   |
| Flaubert                            | 0.9454   |
| Ensemble of Distilbert and Electra* | 0.9463   |

\* The ensemble model was trained for 1 epoch due to difficulties with AWS and VPN.

From the table above it can be seen that Distil Roberta has the maximum accuracy. However, the ensemble model seems likely to be best model if trained for three epochs like the others.

Excluding the import lines of code, I have used approximately 35%-40% of the code from the two reference websites.

While doing this project, I was able to learn how to handle huge amounts of data, understood the importance of clean and efficient code, most importantly, I learnt how text information gets converted into numerical data and how the pre-trained models were able to classify the text. I also learned how to use the pre-trained models and modify their structure to create new architectures. My future work for this project would be to use the pre-trained models and modify their architecture in order to try and perform better than the existing models. Also, I would like to make the code cleaner and more efficient.

## **References**

- Hugging Face website.
  - <https://huggingface.co/docs/transformers/training>
- Ensemble base code
  - <https://discuss.pytorch.org/t/custom-ensemble-approach/52024/4>