

# Individual Final Report

Natural Language Processing

Final Project

December 9, 2021

Abdulaziz Gebril

In this project, we investigated utilizing transformers-based models to train on downstream NLP tasks in the GLUE benchmark. Our focus was to understand BERT model architecture and make a model that uses pretrained models and add BiLSTM and an output layer. We followed lecture code and Hugging Face GitHub repo to build pipelines required to train our models. We also compared BERT model performance with RoBERTa model.

In this project, I contributed to coding, training and finetuning pretrained models. In addition, I contributed to writing the report and creating slides. I customized RoBERTa model and created the “CustomRoberta.py”. The file can train all GLUE tasks by changing the “task” variable. I also wrote the description of model section in the report and slides, where I provided an overview and differences between BERT and RoBERTa and how they differentiate from previous Architectures such as OpenAI GPT and ELMO. I trained all tasks with the customized RoBERTa model

The results obtained from the RoBERTa model was better for all tasks. The results obtained are shown in the following table.

Task	Batch size	Epoch	Learning rate	RoBERTa
CoLA	64	30	5e <sup>-5</sup>	Matthew's Corr: 0.6131
SST-2	64	5		Accuracy: 0.938
MRPC	16	5		F1: 0.927 Accuracy: 0.8995
STS-B	8	5		Pearson: 0.90488 Spearman Corr: 0.902
QQP	32	5		F1: 0.885 Accuracy: 0.9144
MNLI	32	5		Accuracy: 0.865
QNLI	32	5		Accuracy: 0.9218

In conclusion, We Implemented different Transformers-based models (BERT & RoBERTa). I contributed in finetuning and training a customized RoBERTa model. For future work, I would like to explore other based transformer models. In addition, due to GPU limitations, we were not able to conduct hyperparameter tuning. I would like to explore that for future work. Finally, it would be interesting to create our own transformer-based model.

#### References:

1. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova (2019, May 24). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Retrieved December 7, 2021, from <https://arxiv.org/pdf/1810.04805.pdf>
2. Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov (2019, July 26). RoBERTa: A Robustly Optimized BERT Pretraining Approach. Retrieved December 7, 2021, from <https://arxiv.org/pdf/1907.11692.pdf>
3. Ashwin Geet D'Sa's answer on Stackoverflow, <https://stackoverflow.com/questions/65205582/how-can-i-add-a-bi-lstm-layer-on-top-of-bert-model?rq=1>
4. NLP lecture codes, [https://github.com/amir-jafari/NLP/blob/master/Lecture\\_09/Lecture%20Code/12-training.py](https://github.com/amir-jafari/NLP/blob/master/Lecture_09/Lecture%20Code/12-training.py)
5. Hugging Face transformers notebooks, [https://colab.research.google.com/github/huggingface/notebooks/blob/master/examples/text\\_classification.ipynb#scrollTo=7k8ge1L1rJk](https://colab.research.google.com/github/huggingface/notebooks/blob/master/examples/text_classification.ipynb#scrollTo=7k8ge1L1rJk)