GLUE Benchmark Proposal

NLP Project Proposal

Authors: Jongchan Kim, Nigel Martis, Rohan Paul, Bradley Reardon

• What problem did you select and why did you select it?

We will be working on the GLUE Benchmark project competing with fellow classmates and scholars around the world in an attempt to obtain the highest performance metrics possible while testing various pre-trained transformer models. Our goal is to learn to apply and understand the architecture of said transformer models while applying any changes we see fit.

• What database/dataset will you use?

The datasets we will use are the benchmark datasets provided on the GLUE website. These datasets will be used to complete their corresponding task(s).

• What NLP methods will you pick from the concept list? Will it be a classical model or will you have to customize it?

We plan to test various transformer models found in the Transformer package to see how they fare against one another on the GLUE Benchmark tasks. Some customization to the transformers might occur if we feel we can make improvements

• What packages are you planning to use? Why?

The packages we will use are NumPy, Pandas, Matplotlib, PyTorch, and Transformers. NumPy and Pandas will be used for data cleaning and preprocessing, Matplotlib will be used for visualizing our findings, PyTorch used to implement, train, and test our models, and transformers will be used to import and pipeline the transformer base models like Bert, GPT, etc..

• What NLP tasks will you work on?

The reference materials we will use to obtain sufficient background on applying the chosen network to our vehicle image classification are the NLP course materials, the official documentation websites for the packages we will be using, the GLUE Benchmark site, and various websites such as medium.com.

• How will you judge the performance of the network? What metrics will you use?

We will judge the performance of our network by checking the accuracy score of our test dataset. The metrics we will calculate are the ones mentioned in the GLUE benchmark like accuracy, F1 score and Pearson's scores. We may use accuracy as a metric, depending on how balanced the datasets are. This will be examined during exploratory data analysis.

• Provide a rough schedule for completing the project

The table below outlines the progress and deliverables that we plan to achieve at each date.

Step	Target Date
Decide on project (GLUE)	2021-11-03
Familiarize ourselves with the GLUE project & datasets	2021-11-09
Choose which transformers to use; test & compare results	2021-11-16
Optimize models and determine which is best	2021-11-23
Draft of final report	2021-11-27
Finalize report	2021-12-04
Create and finalize presentation	2021-12-05