Replication of DistilBERT evaluation - Group M

Tai Ho, Nadun Chandrabahu, Muhammad Umer Bashir, Thanushreyas

1 Replication of DistilBERT evaluation

1.1 Introduction

Describe the source paper and talk about our project justification again.

Add a footnote to the original paper: https://arxiv.org/abs/1910.01108 (Sanh et al, 2019)

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter (Version 4). arXiv. https://doi.org/10.48550/ARXIV.1910.01108

1.2 Evaluation Datasets

We'll have a look at the benchmark datasets GLUE, IMdb, SQuAD, we can also talk about how we can use

1.3 Replication of Evaluation Results

Show the coding and use the model on the datasets used in the paper such as GLUE benchmark, IMdb, SQuAD. Calculate the metrics and show how close they are to the ones in the paper

Include link to our github repo. And how we can replicate. Maybe we can include a notebook file running the evaluation on the datasets. We're using python version 3.12.7 to install transformers library so we can get the model. Also need to install pytorch

We can run python code as follows, calculate metrics and report on them. Let's have a discussion about the results on the datasets used in the paper in this section as well.

```python
from transformers import pipeline
print("Hello GroupM!")
```

```
## Hello GroupM!
```

1.4 New data construction

Each member can talk about how they constructed their data.

1.4.1 Tai Ho's dataset

Talk about how you created your dataset, and what your goal /context of the dataset is.

1.4.2 Nadun Chandrahu's dataset

I'll be talking about a question and answer dataset, similar to SQuAD, I can manually anotate whether the answers are correct or not.

1.4.3 Evaluation using new datasets

1.4.4 Tai Ho's dataset

Talk about how your dataset evaluates with the model

1.4.5 Nadun Chandrabahu's dataset

Talk about how your dataset evaluates with the model

1.5 Reflections

Each of us can write a paragraph about our reflections on the project.