TP N° 03

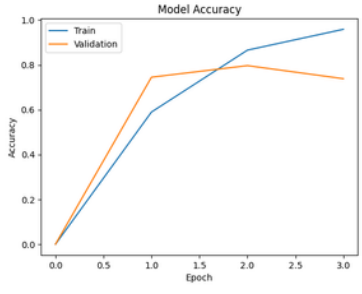
TEXT CLASSIFICATION USING BERT MODEL

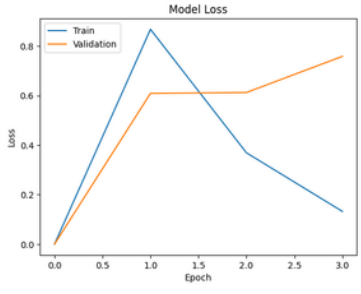
I have developed a text classification model using BERT's uncased pre-trained version. I'll share the results and compare them to my previous model, which utilized Skip-Gram with MLP.

BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained language model that captures contextual word meanings. In my text classification model, I fine-tuned BERT on my dataset to extract relevant features and make predictions.

Let's start with the performance of my **BERT** model (**Note** that this impressive resulats in just **3 itreations** and **1%** from our Orginal **dataset**):

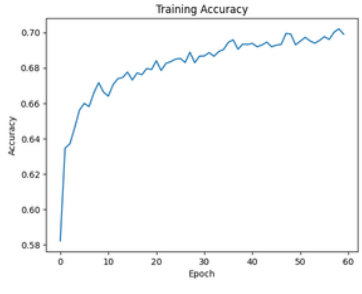
- During the first epoch, the model achieved a loss of **0.8676** and an accuracy of **0.5896**. The validation set showed a loss of **0.6088** and an accuracy of **0.7450**.
- Moving on to the second epoch, the loss significantly decreased to **0.3690**, and the accuracy improved to **0.8658**. The validation loss remained similar at **0.6124**, while the validation accuracy increased to **0.7967**.
- In the third and final epoch, the loss dropped even further to **0.1323**, with an impressive accuracy of **0.9588**. **However**, the validation loss slightly increased to **0.7583**, resulting in a validation accuracy of **0.7383**.

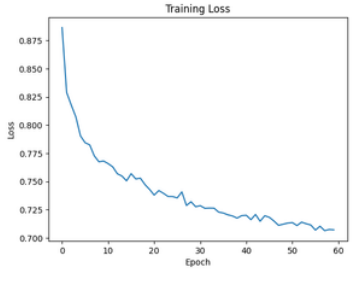




After training, I evaluated my BERT model on the test set, where it achieved a test accuracy of **0.7383** with a corresponding test loss of **0.7583**.

Comparing these results to my previous **Skip-Gram** with **MLP** model, which reached an accuracy of **0.6990** after **60** epochs, it's clear that BERT **outperforms** it in terms of accuracy.





Interpreting these results, we can see that the pre-training of BERT on a large corpus **allows** it to capture more semantic relationships between words and contexts, resulting in **superior text classification performance**. However, it's worth noting that the model seemed to exhibit some **overfitting**, as the validation accuracy plateaued after the second epoch, and the test accuracy was slightly **lower** than the validation accuracy.

To further **improve** my BERT model, I plan to conduct a thorough **error analysis** and explore **regularization** techniques to mitigate overfitting. **Additionally**, I will consider **hyperparameter** tuning to optimize the model's performance.

Overall, I'm pleased with the results achieved by my BERT-based text classification model.

// ABDECHAKOUR MECHRI