T.C.

MARMARA UNIVERSITY

FACULTY OF ENGINEERING

COMPUTER ENGINEERING DEPARTMENT

CSE4078

Introduction to Natural Language Processing

Group 8

Delivery 3 Report

Group Members

Emirhan Erdoğan

Faruk Akdemir

Eren Başpınar

Alper Özdemir

Şule Koca

1. Introduction

During the third phase of our project, we focused on training a machine learning model and analyzing its outputs. Specifically, we trained the Gamma 2b model using our available data. Throughout the training process, we adjusted parameters such as Rank (r) and Alpha to generate various models. Subsequently, we initiated the inference process using these models alongside our test dataset, encountering challenges along the way

2. Problems Occured

While there were no obstacles during the model creation and loading stages, we faced a significant issue during inference. Despite having a GPU with 40 GB of RAM, we encountered insufficient RAM during the inference process, resulting in incomplete execution. To overcome this limitation, we plan to reduce the dataset size in future phases. By doing so, we anticipate resolving the RAM insufficiency and successfully completing the model's inference processes.

3. Analysis of Training Results and Evaluation Metrics

In this section, we delve into an analysis of the training results obtained for the Gamma 2b model

3.1 Training Results Analysis

Throughout the training phase, we experimented with different parameter configurations, particularly focusing on Rank (r) and Alpha values, aiming to optimize the Gamma 2b model's performance. By varying these parameters, we generated a diverse set of outputs, enabling us to compare and analyze the results obtained from different configurations

3.2 Evaluation Metrics

ROUGE (Recall-Oriented Understudy for Gisting Evaluation): ROUGE1, ROUGE2, and ROUGE-L measure the overlap of n-grams between the generated translations and reference translations. These metrics provide insights into the adequacy and relevance of the generated outputs.

BLEU (Bilingual Evaluation Understudy): BLEU evaluates the quality of translations by comparing the n-grams in the generated translation with those in reference translations. It quantifies the similarity between the generated and reference translations.

BERTScore: BERTScore measures the similarity between the contextual embeddings of the generated and reference translations, providing a more nuanced evaluation of translation quality.

3.6 Conclusion

In conclusion, our analysis of training results and evaluation metrics sheds light on the performance of the Gamma 2b model. Despite encountering RAM insufficiency during inference, hindering metric availability, we remain focused on resolving these challenges in upcoming phases. Our project initially aimed for a large dataset to ensure consistent and robust training. However, resource limitations constrained our ability to complete this within the designated timeframe