REPORT
FOR

Transformers Architecture Task

PREPARED FOR : generative ai consultant internship test

Table of Content

Objective:
Develop a simple transformer-based model to solve a specific problem, such as text classification, sentiment analysis, or language translation.

Model Architecture:
The transformer model is designed with a 2-layer transformer block. The self-attention mechanism and the feed-forward network are utilized in this model. Furthermore, the neural network architecture contains an embedding matrix, a global mean pooling layer, a fully connected hidden layer, and a concluding prediction layer for producing outputs.

Dataset:
The binary sentiment analysis dataset is used for this task. This dataset contains reviews from Amazon, and each review is labeled as positive or negative.

Challenges:
One challenge faced during the implementation was to adjust the hyperparameters, such as the learning rate and the number of units in the dense layer, to achieve optimal model performance. Another challenge was to manage the sequence length and padding to efficiently process the input sequences.

Model Evaluation Results:

Training Accuracy: 98.25%

Testing Accuracy: 85.55%

Classification Report: precision recall f1-score support negative 0.844309 0.951567 0.894625 2029 positive 0.870434 0.842989 0.856299 1845

The model achieved an overall accuracy of 0.855519 with a macro average of 0.857372, weighted average of 0.897277, and weighted class average of 0.872643 when evaluating 3874 samples, mirroring its original accuracy, weighted, and sample counts.