Task 2

Training vs. Validation loss-epoch diagram

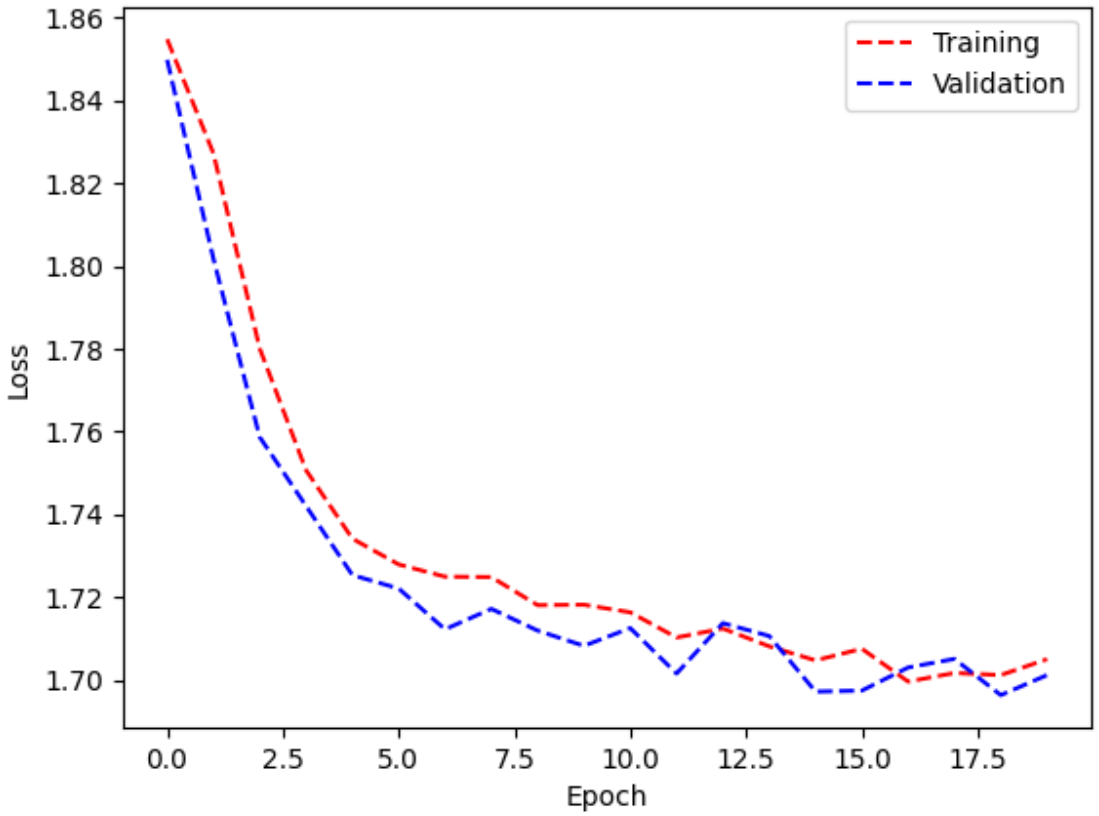

Figure 1: Training and validation loss across epochs.

Results

My model achieved a total F1 score of **0.332**. The F1 scores for each genre can be seen in Table 1.

Genre	F1 Score
Blues	0.0
Country	0.0
Metal	0.571
Pop	0.571
Rap	0.625
Rock	0.222

Table 1: F1 Scores across different genres.

Analysis of the result

The model failed to classify the genres correctly. For instance, the model did not classify any songs of the Blues or Country genre correctly as we can observe from Table 1. For other genres, the model performed better, achieving a total F1 score of 0.332. The performance of this model reflects the difficulty of the task. I have decided to extract the following list of features from the lyrics: Number of Words, Vocabulary Size, Average Line Length, Average Word Length, Stop Word Ratio, Rhyme density, Line Length Variation, Sentiment Score, and Part-of-Speech Distribution. I have also tried extracting and using Named Entity Recognition (NER), but I found that NER did not have a strong correlation with the song genres, so I decided to exclude it from the final model.

Genre classification based on solely lyrics features is difficult as there are not many clear relationships between the features and the genres. To help visualize these relationships I provide a correlation heatmap in Figure 2.

I believe that extracting more features or having more training data might help with the model's performance. I did not expect great results from this experiment as genre classification solely from the lyrics can pose difficulties to many people. Song lyrics lack audio features like tempo, energy, and instrumentalness which carry much more valuable information about the song genre than just text. Choosing a different approach (like in Task 3) is probably the best way to achieve the best results.

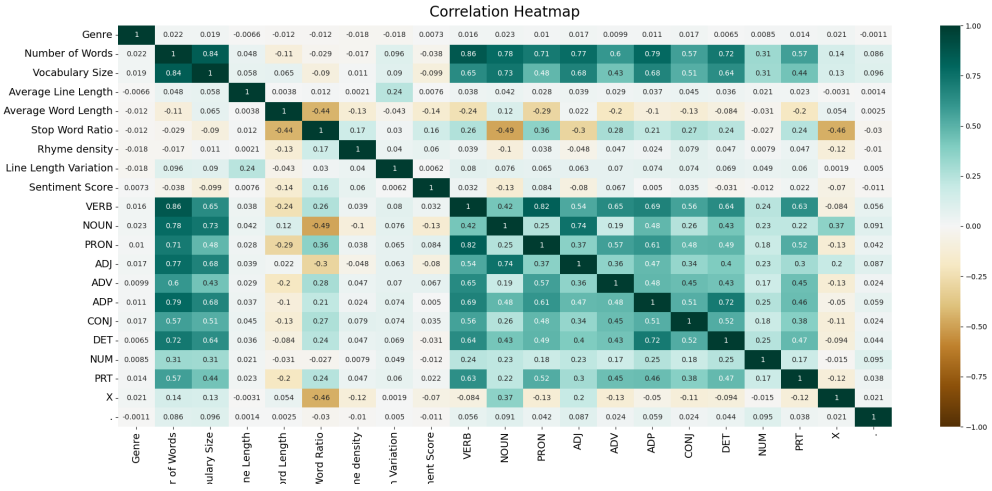

Figure 2: Correlation heatmap between the lyrics features and song genre.