Summary

In examining the performance between Embedding and Pretrained embedding models, I found that the embedding model achieved high training accuracy (98%), but does not exhibit the same level of performance on the validation accuracy (~52%). This discrepancy suggests a potential overfitting issue, where the model's capacity to generalize is limited due to the small size of the training dataset. On the other hand, the Pretrained embedding model showcases more consistency between its training and validation accuracy, both maintaining around the 58% for training accuracy and an average of 51% for validation accuracy. Cutting reviews after 150 words does not significantly affect the validation accuracy for either model, indicating that the initial words carry sufficient sentiment signals for this task.

I ran these models with a variety of sample sizes to see how the performance of the models would be affected. Both the training and validation loss/accuracy were better on the embedding model than the pretrained embedding model around a sample size of 250. This can be seen in the graphs below. To address overfitting, we could utilize regularization techniques or more sophisticated model architectures. Future experiments could also explore varying the cutoff for review lengths to seek an optimal balance between input size and model performance. The plotted data suggests that neither model significantly outperforms the other under the current configurations and data size. This shows a need for additional experimentation to derive clear conclusions about the most effective model setup for this classification task.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Train Loss** | **Train Accuracy** | **Validation Loss** | **Validation Accuracy** | **Epoch** | **Notes** |
| Embedding | 48.97% | 98.00% | 69.36% | 50.79% | 10 |  |
| Pretrained | 67.99% | 58.00% | 70.15% | 49.49% | 10 |  |
| Embedding 2 | 30.45% | 97.50% | 70.17% | 51.62% | 10 | Cutting off reviews after 150 words |
| Pretrained 2 | 68.34% | 57.50% | 70.10% | 49.49% | 10 | Cutting off reviews after 150 words |

A graph with a red line

Description automatically generated

A graph with a red line

Description automatically generated