Assignment 4

For the initial model setup, 100 training samples were selected, each containing reviews of up to 150 words. A total vocabulary of 10,000 words was utilized as input for the model. Additionally, the model underwent validation using 10,000 samples, encompassing both positive and negative reviews. The chosen loss function for this classification model was binary cross-entropy, while the optimizer used was RMSProp. ReLU was employed as the activation function throughout the model.

100 Training Samples:
There are four models trained, validated, and tested using the initial setup with performance metric as accuracy.

- **One hot encoded sequence:** Accuracy of .654 and loss of .6386.
- **Embedded model without masking:** Accuracy of .558 and loss of .6827.
- **Embedded model with masking:** Accuracy of .597 and loss of .683.
- **A pre-trained model Global Vectors for word representation (GloVe):** Accuracy of .536 and loss of .6998.

Training Across Several Sample Sizes:

sample size	one hot encoded sequence		Embedded		Embedded masked		Pretrained	
	Test Loss	Test Accuracy	Test Loss	Test Accuracy	Test Loss	Test Accuracy	Test Loss	Test Accuracy
100	0.6386	0.654	0.6827	0.558	0.6830	0.597	0.6998	0.536
800	0.579	0.708	0.6203	0.658	0.6610	0.574	0.6186	0.659
1600	0.5138	0.741	0.5812	0.697	0.5541	0.724	0.913	0.681
2400	0.5762	0.7639	0.7168	0.759	0.6617	0.65	0.5745	0.7795

Conclusion:
The analysis results showed that in the task of sentiment analysis, RNNs with embedded layers performed noticeably better than other word embedding techniques such one-hot encoded sequences. Test accuracy and test loss were consistently higher with the embedded layer-based models than with other methods. Furthermore, it was found that higher sample sizes enhanced the performance of RNN-based models. The test loss of the RNN-based models dropped from roughly 64% to less than 58% as the sample size rose from 100 to 2400 samples, demonstrating the positive effect of larger training data sets on model performance.

Additionally, it was observed that standard embedded layer-based models showed somewhat better test accuracy than masked embedded layer-based models when comparing various types of embedded layers, such as standard embedded and masked embedded layers. Although the masking strategy can improve performance by allowing the model to concentrate only on meaningful word representations, its effect was not noticeable in this implementation using the IMDb dataset.

When compared to training embedded layers from scratch, the IMDb movie review dataset analysis demonstrated the superiority of employing pre-trained word embeddings—more specifically, GloVe embeddings—in creating more effective and efficient models. After training on 2400 samples, the pre-trained model outperformed both masked and standard embedded layers in terms of test accuracy, achieving a test loss of 57.4%.

The experiment's findings indicate, in summary, that higher sample numbers are often associated with better model performance, with 1600 and 2400 sample sizes consistently producing positive outcomes for a variety of embedding strategies. Furthermore, the pre-trained GloVe embeddings consistently performed better than other embedding methods, exhibiting reduced loss and higher accuracy, especially in situations where training data was scarce. Because pre-trained GloVe embeddings provide standardised representations and make use of a wealth of pre-existing knowledge from big corpora, they can thus greatly improve sentiment analysis applications.