**Assignment 1: Neural Network Model Evaluation**
**IMDB Sentiment Analysis**

**Tara Jameel**

# 1. Objective

In this project, I built and evaluated several neural network models to predict whether a movie review is positive or negative. By testing different model structures and settings, I aimed to identify the configuration that provided the best balance between accuracy and generalization.

## 2. Data Review

The IMDB dataset was used for this project. Each movie review was converted into a numerical format by limiting the vocabulary to the 10,000 most frequent words.

The dataset was split into training, validation, and test sets. The training set was used to fit the models, the validation set was used to compare different configurations, and the test set was used to evaluate the final model's performance on unseen data.

# 3. Modeling Approach

To evaluate performance, I tested several neural network models with different configurations. I changed the number of hidden layers, the number of units in each layer, the activation functions, and the loss functions. I also experimented with adding dropouts to see if it helped reduce overfitting. The goal was to compare these variations and determine which setup performed best on the validation set.

# 4. Key Results

The table below compares the performance of the different models that were tested.

| Model Configuration | Validation Accuracy | Test Accuracy |
|---|---|---|
| 2 layers, 32 units, ReLU, BCE | 0.8696 | 0.8588 |
| 2 layers, 32 units, ReLU, MSE | 0.8694 | 0.8590 |
| 3 layers, 32 units, ReLU, BCE | 0.8667 | 0.8566 |

| | | |
|---|---|---|
| 1 layer, 32 units, ReLU, BCE | 0.8678 | 0.8581 |
| 2 layers, 64 units, ReLU, BCE | 0.8696 | 0.8588 |
| 2 layers, 32 units, tanh, BCE | 0.8647 | 0.8514 |

# 5. Interpretation of Results

The two-layer model with 32 units and ReLU activation achieved the highest validation accuracy while also maintaining similar performance on the test set. Adding a third layer did not improve results and slightly lowered validation accuracy in some cases.

Increasing the number of units from 32 to 64 also did not lead to noticeable improvement. In general, models using binary crossentropy performed slightly better than those using mean squared error. ReLU activation also showed slightly stronger results compared to tanh.

Overall, the results suggest that a moderately sized network was sufficient for this task, and increasing model complexity did not provide additional benefits.

# 6. Recommendation

The two-layer model with 32 units per layer, ReLU activation, and binary crossentropy loss is recommended as the final model. It achieved strong predictive performance and demonstrated stable generalization. Increasing model complexity did not provide additional gains, making the simpler configuration both efficient and effective.

# 7. Conclusion

The results suggest that increasing model complexity does not automatically improve accuracy. In this case, a simpler two-layer network performed just as well as more complex architectures.

Future work could explore different types of models, such as embedding based methods or recurrent networks, to see if they provide additional improvements. For now, the selected model offers a strong and practical solution for sentiment classification.