Phuong Pham

Assignment 1

Table of Contents:

1. Introduction: Objectives and overview of the work.

The project is about using MLP for classification problem on IMDB reviews dataset. The project involves 3 main parts: comparing character-level and word-level tokenization, build the MLP model that can adjust hyperparameters, and visualizing the performance. The project mainly focuses on the influence of changing hyperparamters on the performance of the model.

2. Methodology: Detailed explanation of tokenization changes and hyper-parameter optimization strategy.

For tokenization change, the project added word level tokenization using built in Tokenizer library from Tensorflow. First of all, the function word_level_tokenizer fits the text data into numerical tokens. Then it transform the input texts to a sequence. After that, it ensures that all inputs have the same length by using pad_sequences.

The project conducted a grid search manually to find the most optimized hypermarameter that starts with random seed 42 and character-level-tokenization. Character level tokenization are kept during the first 10 tests, and other parameters are changed one by one for each search. After that, the project is conducted with 10 other tests on word level tokenization. The training

loss and val loss for word level tokenization could be very high, >1000 when the activation is ReLU or LeakyReLU, so the test focuses more on Tanh activation function for word level tokenization. All results are recorded to find the best parameter combination. If there are any combination that work mostly better than other tests, the experiment would keep that combination of parameters and change other parameters.

3. Experiments and Results:

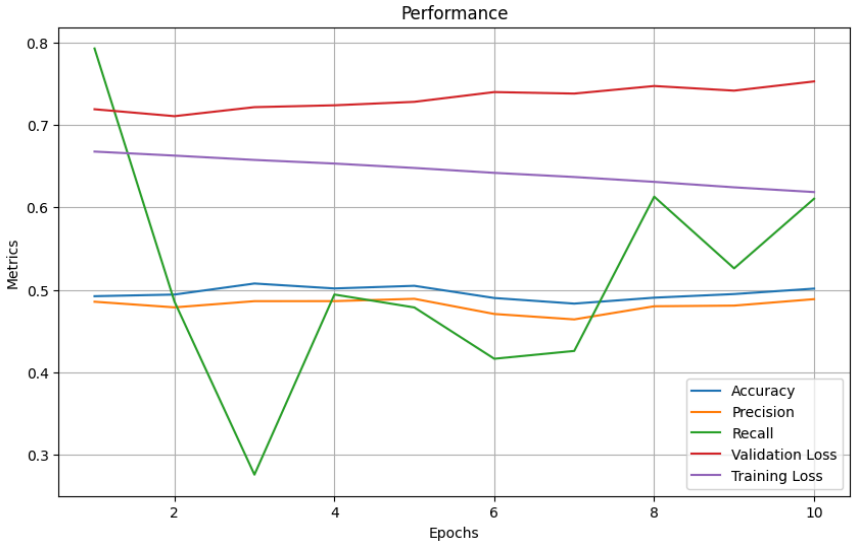
a. Comparison between character-level and word-level tokenization.

The vocab size of word-level is longer than character-level, but the sequence length of word-level is shorter than character-level. Since the character level tokenizer breaks the word into individual character, it can work with unseen data, while word level tokenizer cannot handle unseen data. However, since the character level tokenizer has long sequence length, it can be more computational expensive.

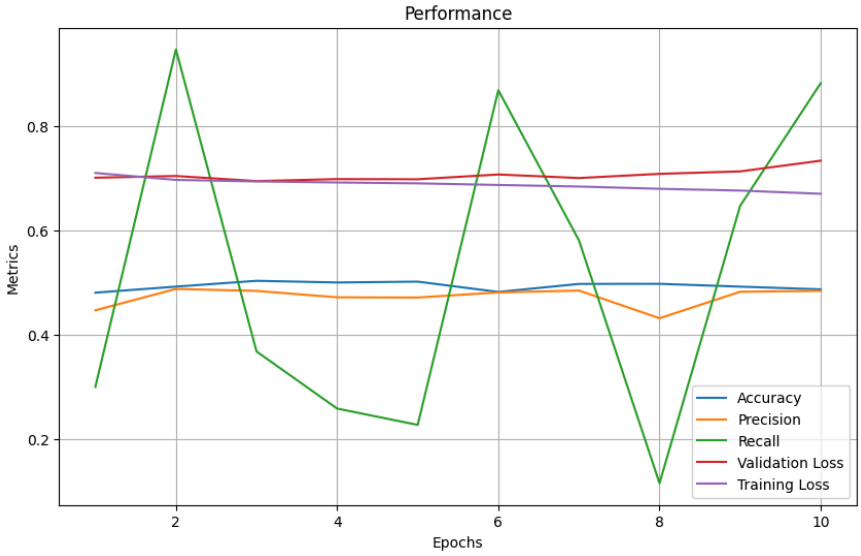
In this project, the performance of word level tokenizer and character level tokenizer are almost the same.

b. Tables/graphs for hyper-parameter experiments.

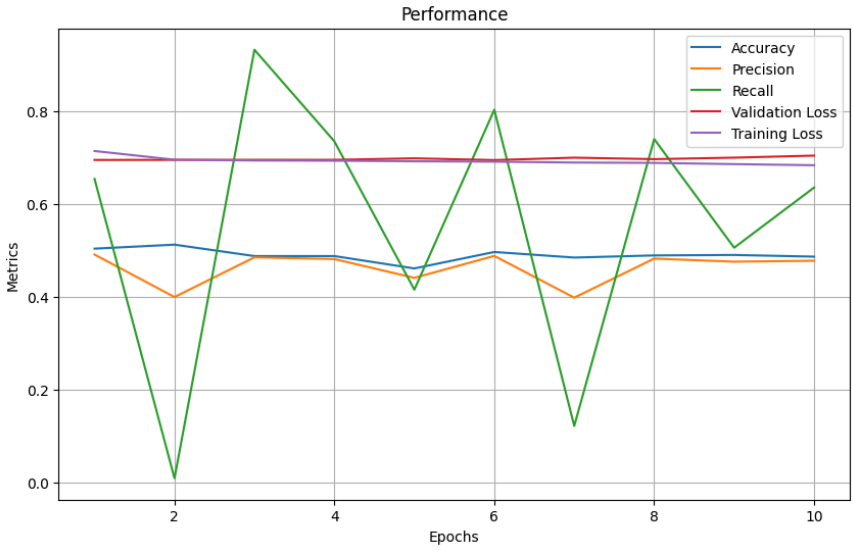
1. Seed 42, Character Level Tokenization, Learning Rate 0.0005, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 32, Optimizer Adam, Activation Relu



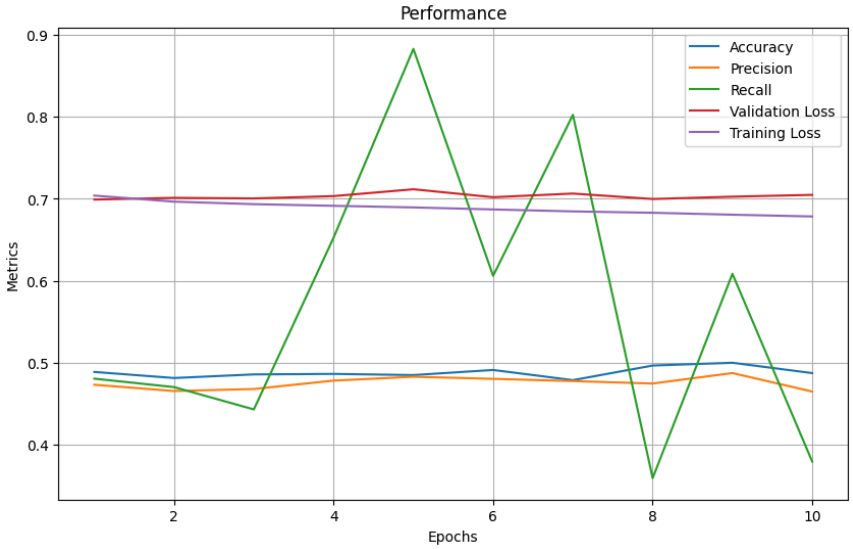
2. Seed 42, Character Level Tokenization, Learning Rate 0.0005, 2 Hidden Layers, Hidden Layer 1st 256, Hidden Layer 2nd 512, Batch Size 32, Optimizer Adam, Activation Relu



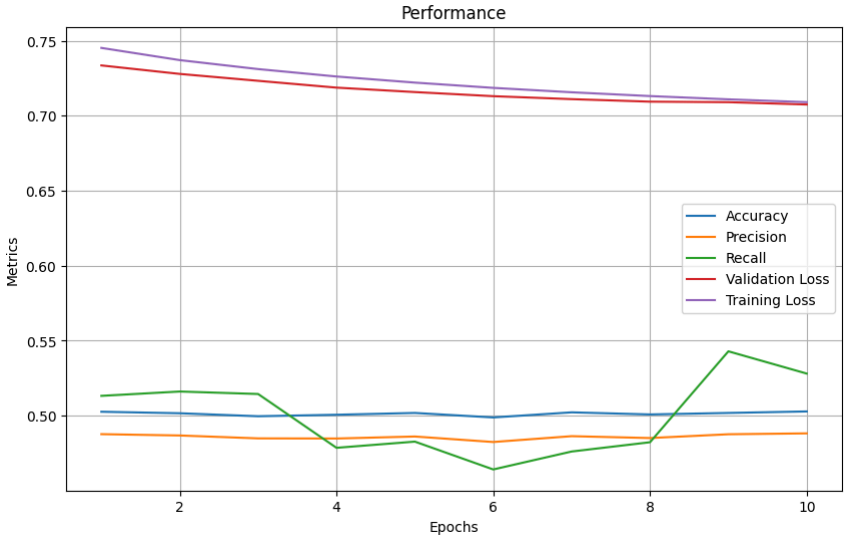
3. Seed 42, Character Level Tokenization, Learning Rate 0.001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 32, Optimizer Adam, Activation Relu



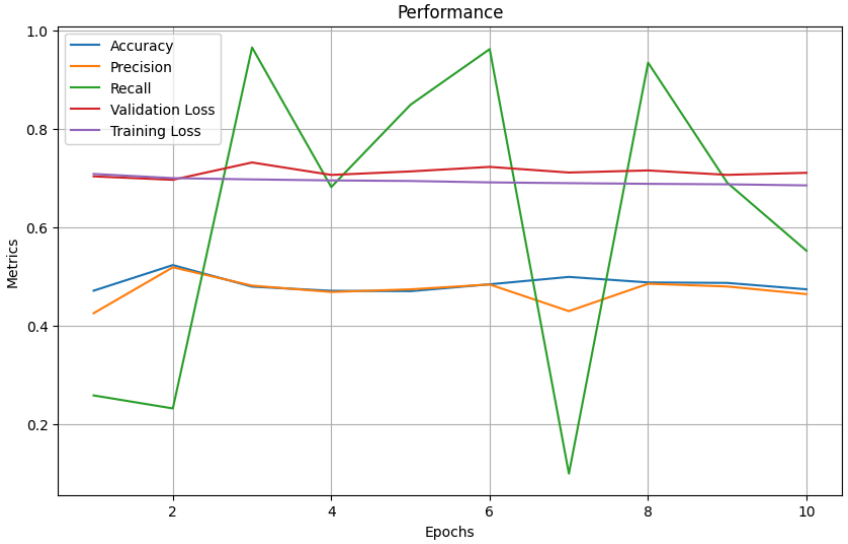
4. Seed 42, Character Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 32, Optimizer Adam, Activation Relu



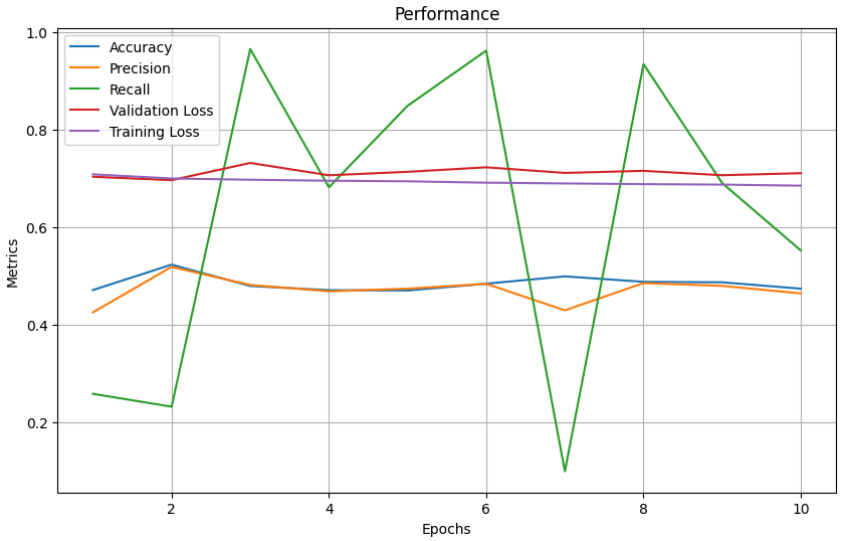
5. Seed 42, Character Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 32, Optimizer SDG, Activation Relu



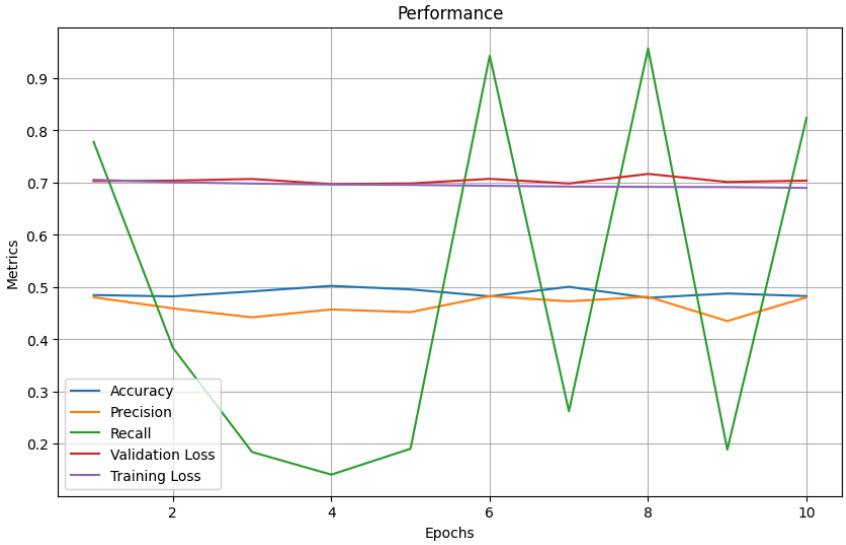
6. Seed 42, Character Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 32, Optimizer RMSprop, Activation Relu



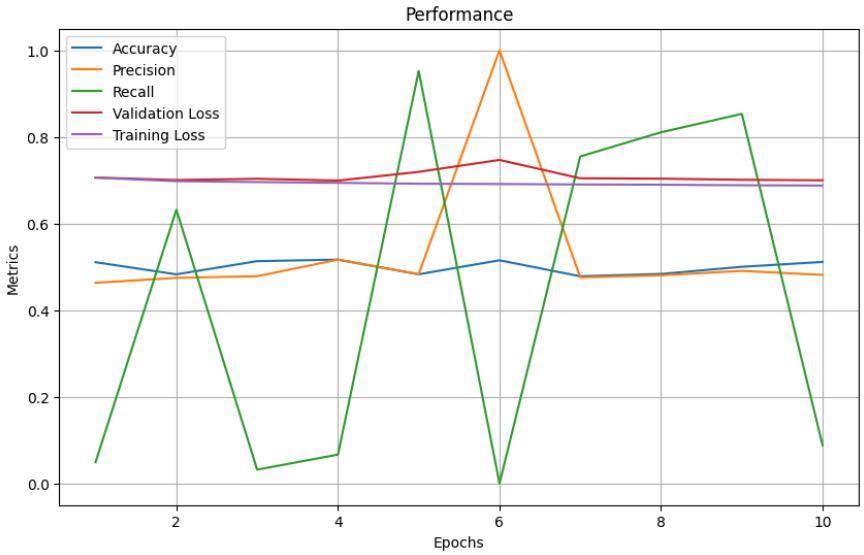
7. Seed 42, Character Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 32, Optimizer RMSprop, Activation Tanh



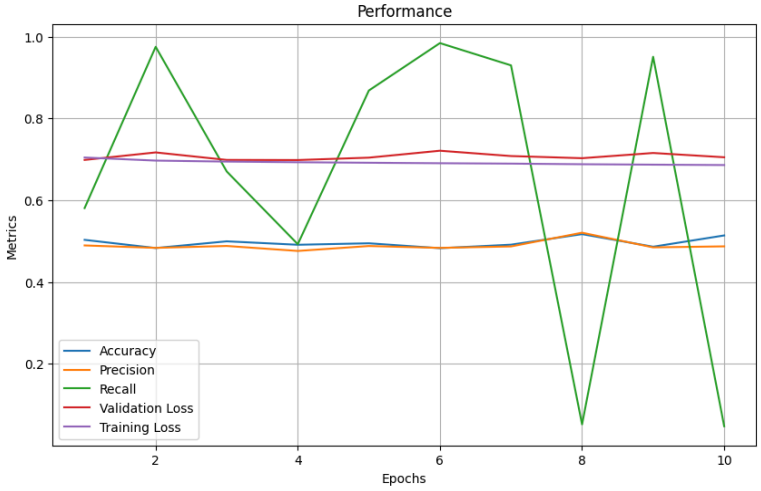
8. Seed 42, Character Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 32, Optimizer RMSprop, Activation LeakyReLU



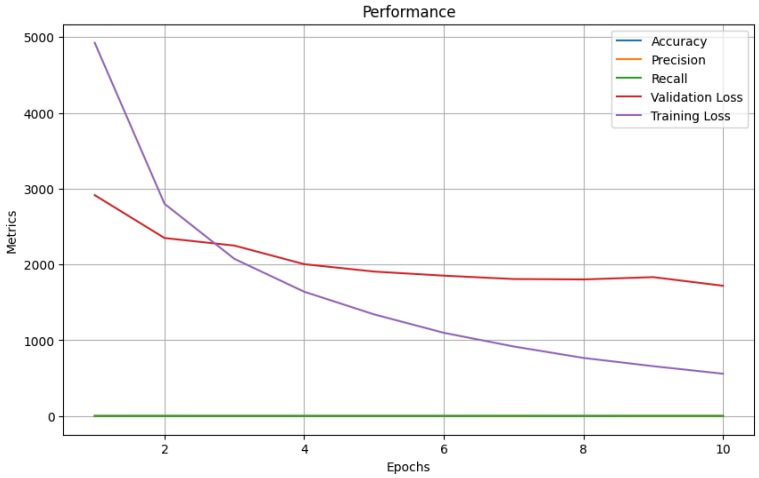
9. Seed 42, Character Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 64, Optimizer RMSprop, Activation ReLU



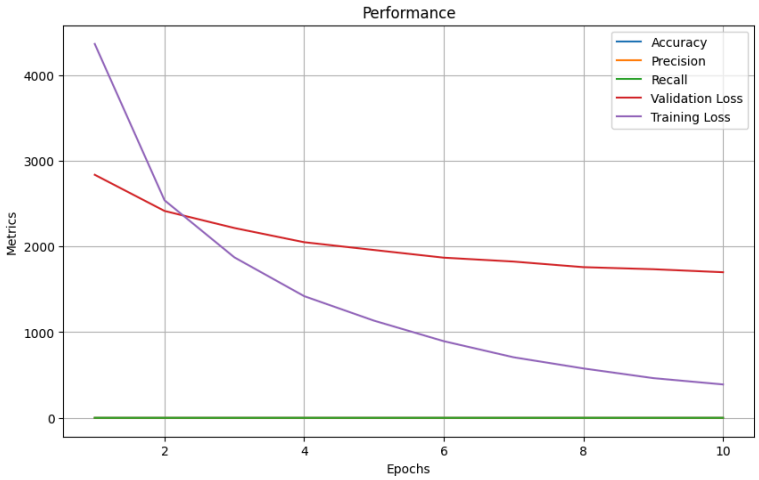
10. Seed 42, Character Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 128, Optimizer RMSprop, Activation ReLU



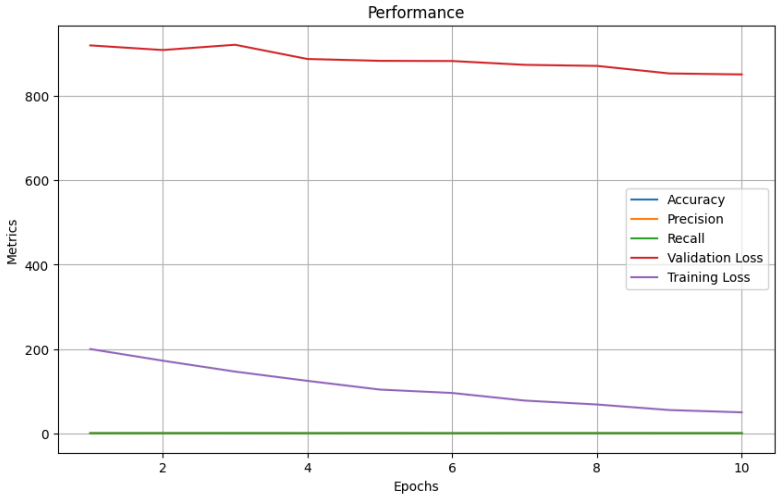
11. Seed 42, Word Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 128, Optimizer RMSprop, Activation ReLU



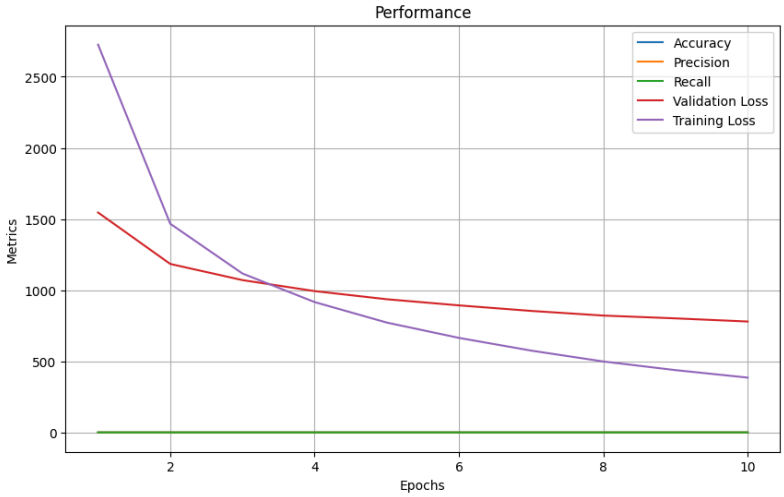
12. Seed 42, Word Level Tokenization, Learning Rate 0.0001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 512, Hidden Layer 3rd 512, Batch Size 128, Optimizer RMSprop, Activation ReLU



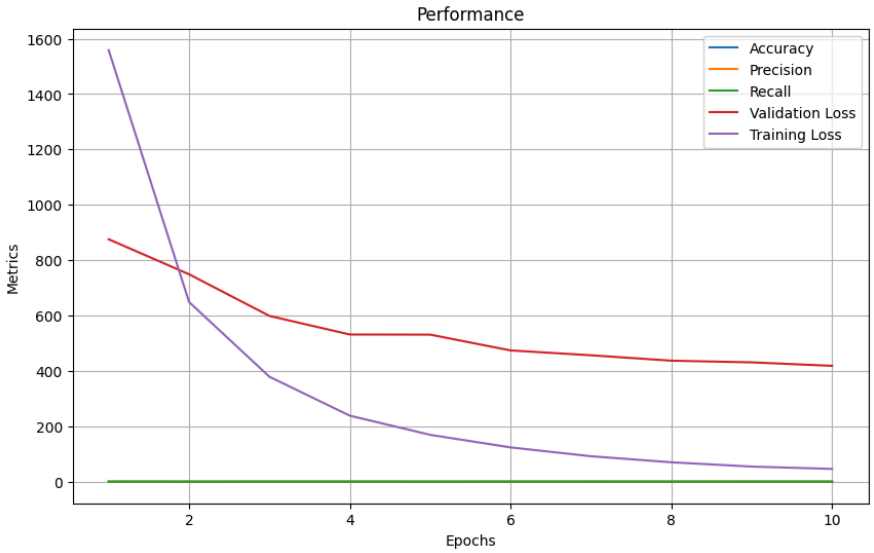
13. Seed 42, Word Level Tokenization, Learning Rate 0.0001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 128, Optimizer Adam, Activation ReLU



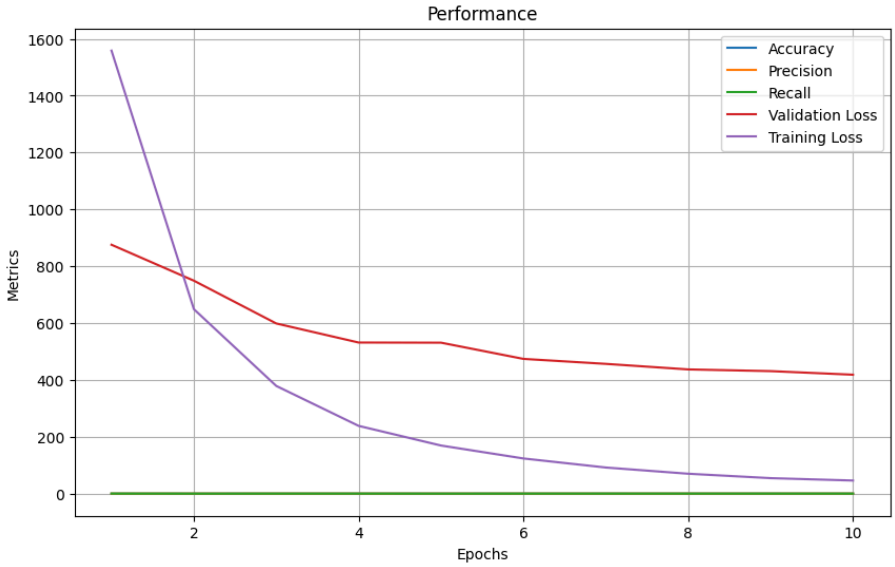
14. Seed 42, Word Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 128, Optimizer Adam, Activation ReLU



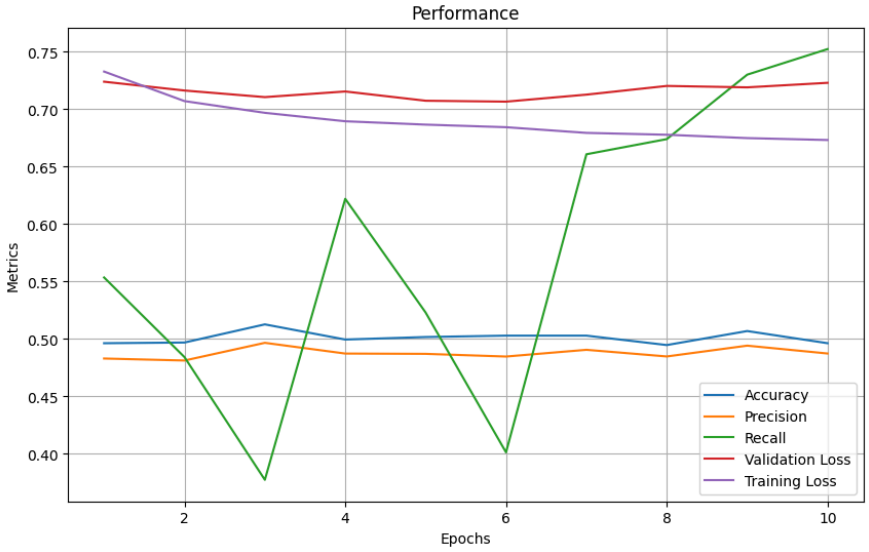
15. Seed 42, Word Level Tokenization, Learning Rate 0.0005, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 128, Optimizer Adam, Activation ReLU



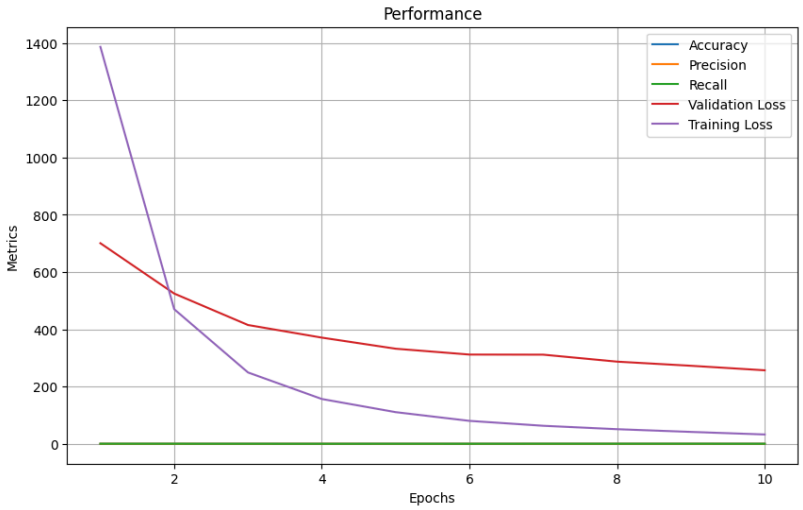
16. Seed 42, Word Level Tokenization, Learning Rate 0.001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 128, Optimizer Adam, Activation ReLU



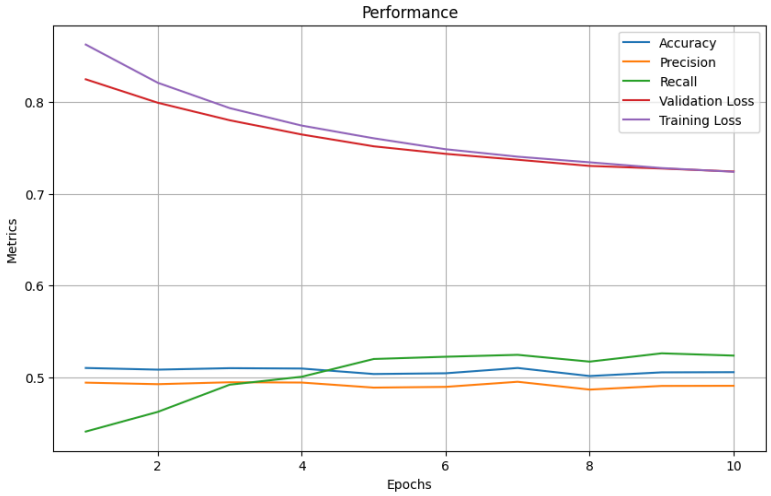
17. Seed 42, Word Level Tokenization, Learning Rate 0.001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 128, Optimizer Adam, Activation Tanh



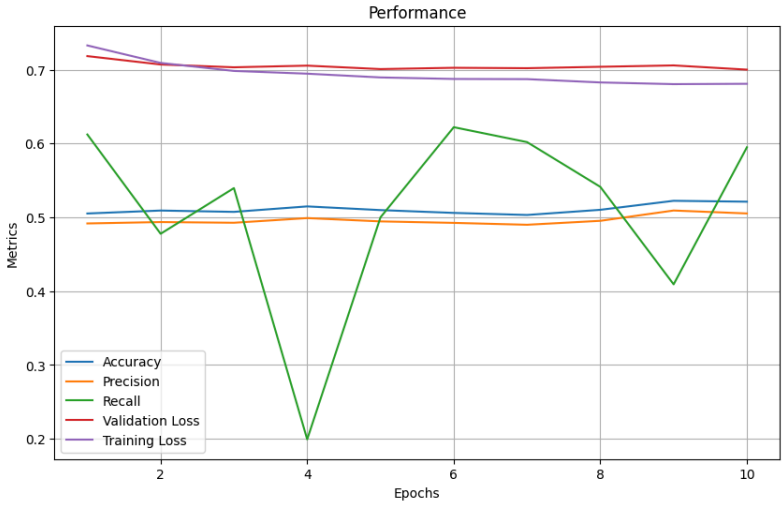
18. Seed 42, Word Level Tokenization, Learning Rate 0.001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 128, Optimizer Adam, Activation LeakyReLU



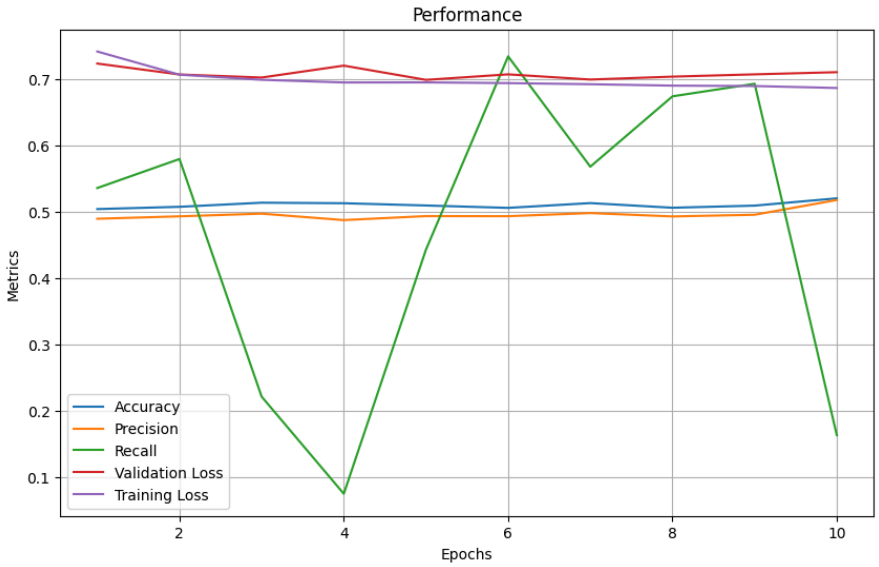
19. Seed 42, Word Level Tokenization, Learning Rate 0.001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 128, Optimizer SGD, Activation Tanh



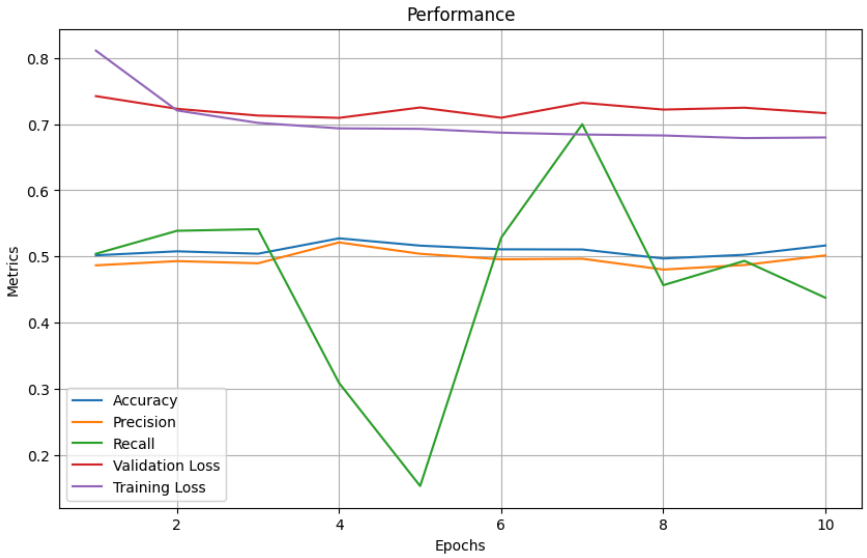
20. Seed 42, Word Level Tokenization, Learning Rate 0.001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 64, Optimizer Adam, Activation Tanh



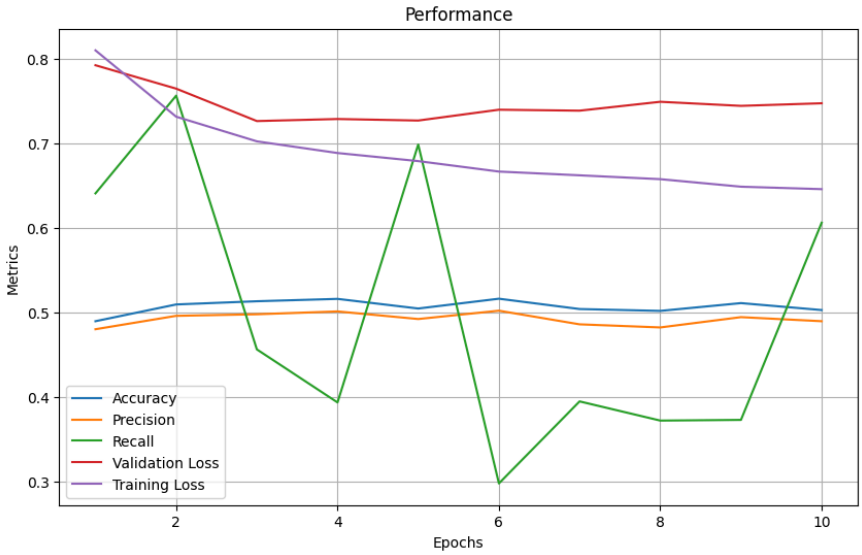
21. Seed 42, Word Level Tokenization, Learning Rate 0.001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 32, Optimizer Adam, Activation Tanh



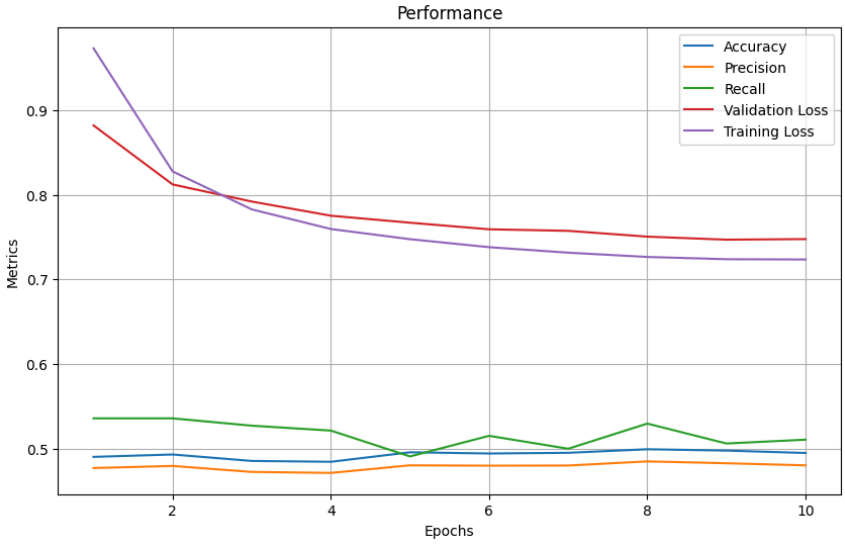
22. Seed 42, Word Level Tokenization, Learning Rate 0.001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 32, Optimizer Adam, Activation Tanh



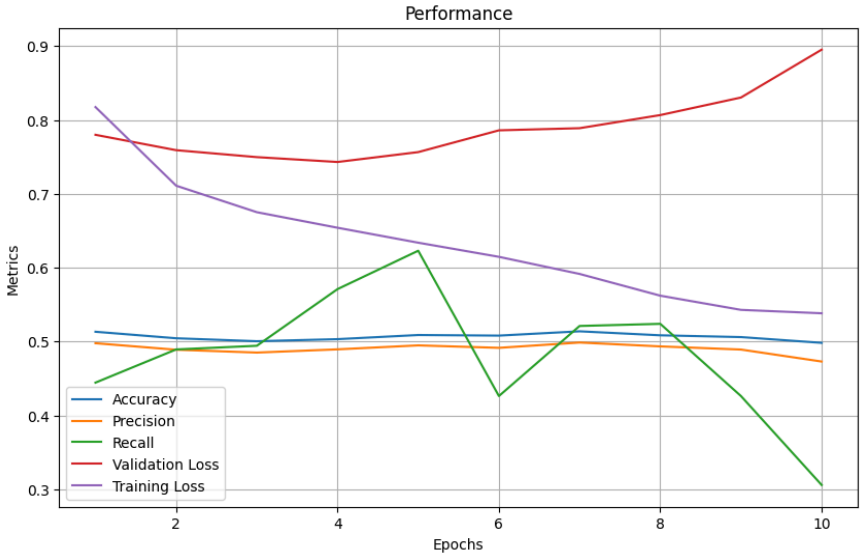
23. Seed 42, Word Level Tokenization, Learning Rate 0.001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 64, Optimizer Adam, Activation Tanh



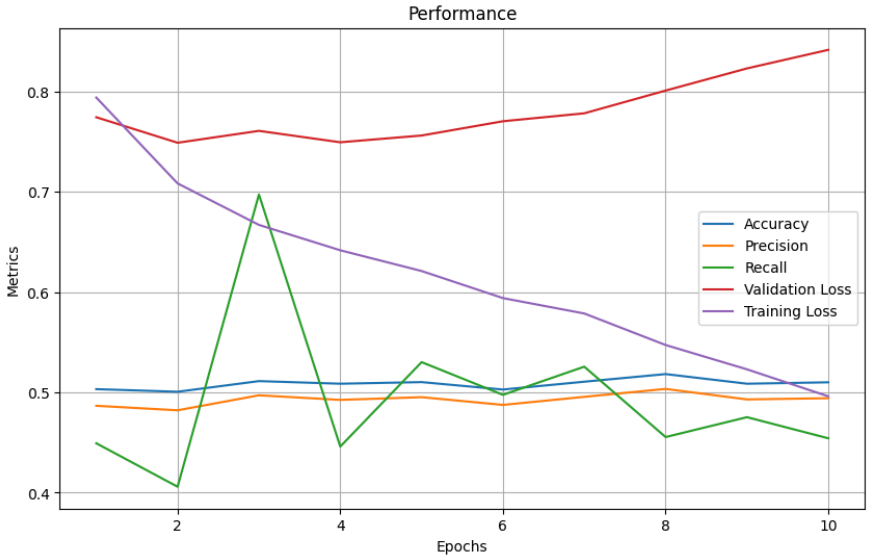
24. Seed 42, Word Level Tokenization, Learning Rate 0.001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 64, Optimizer SGD, Activation Tanh



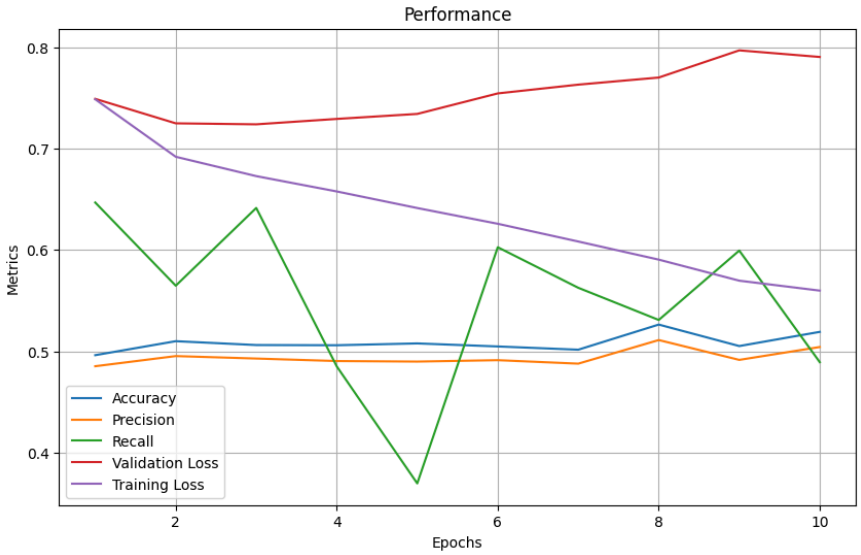
25. Seed 42, Word Level Tokenization, Learning Rate 0.001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 64, Optimizer RMSprop, Activation Tanh



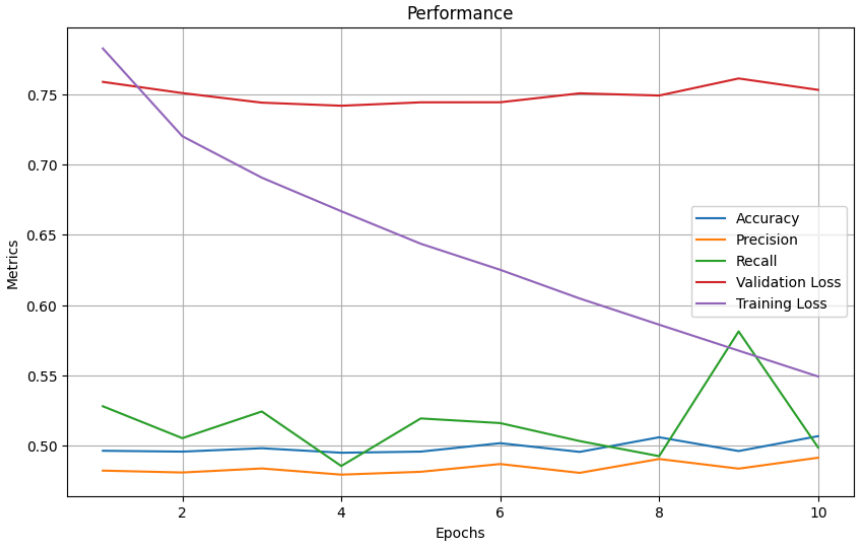
26. Seed 42, Word Level Tokenization, Learning Rate 0.001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 256, Batch Size 132, Optimizer RMSprop, Activation Tanh



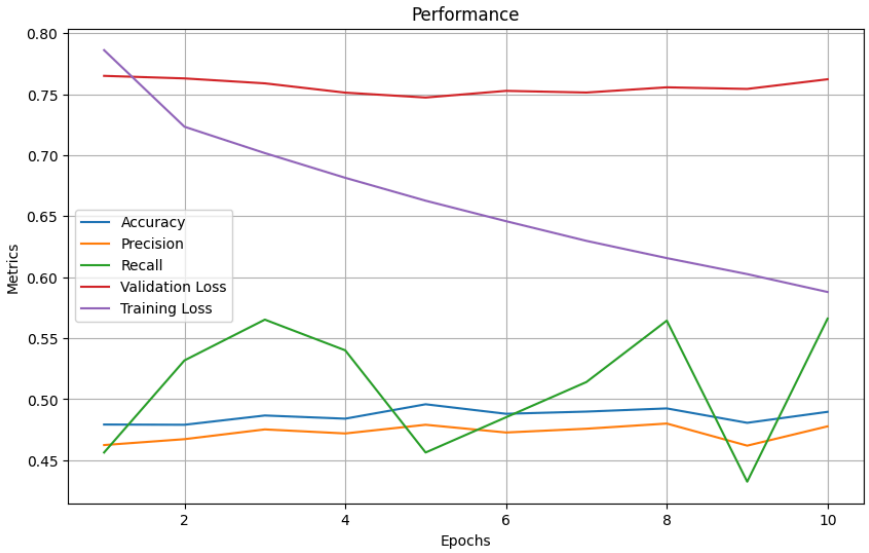
27. Seed 42, Word Level Tokenization, Learning Rate 0.001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 256, Batch Size 64, Optimizer RMSprop, Activation Tanh



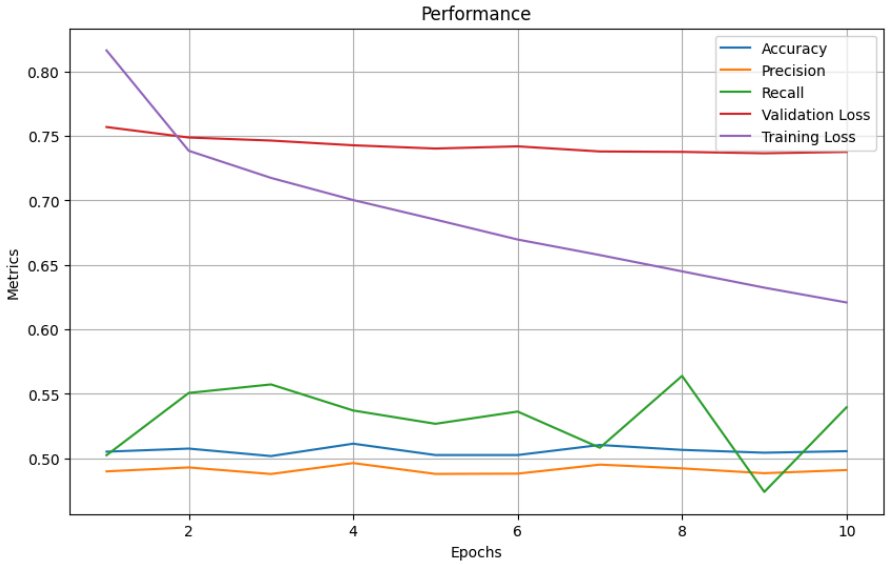
28. Seed 42, Word Level Tokenization, Learning Rate 0.0001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 32, Optimizer RMSprop, Activation Tanh



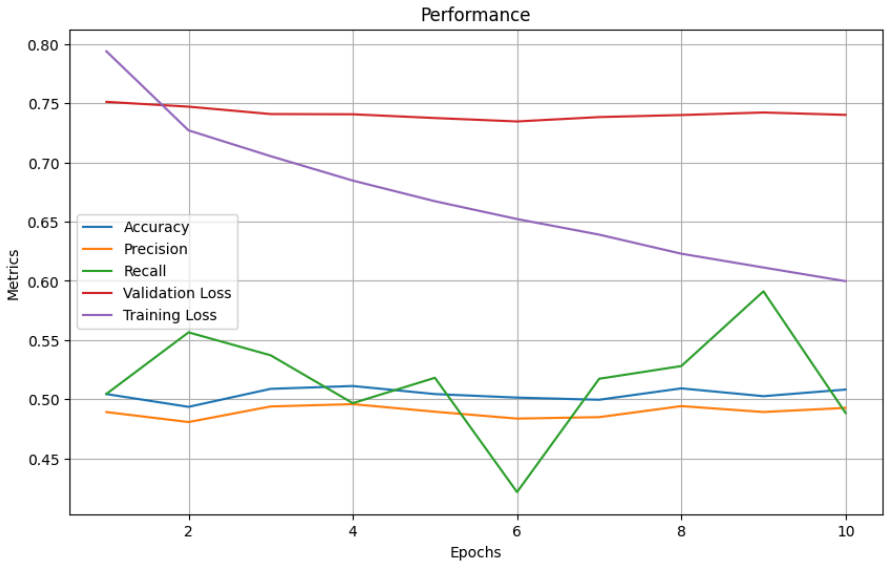
29. Seed 42, Word Level Tokenization, Learning Rate 0.0001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 64, Optimizer RMSprop, Activation Tanh



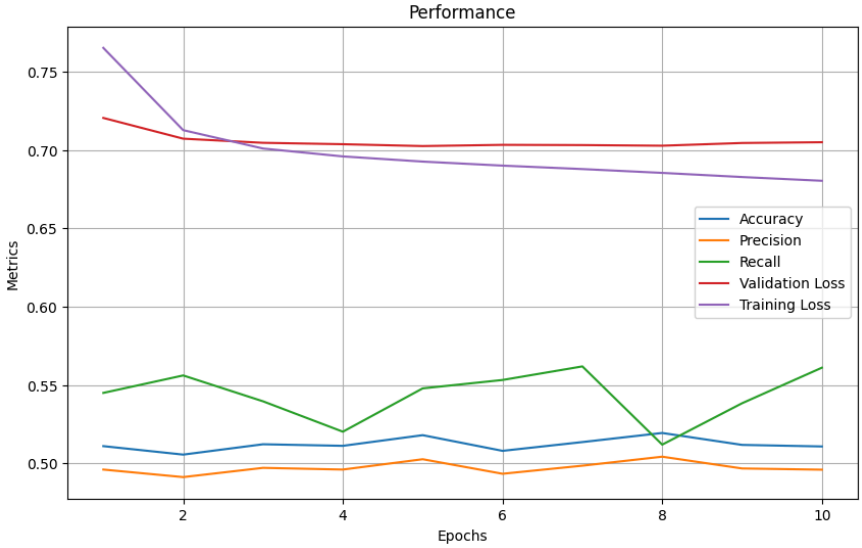
30. Seed 42, Word Level Tokenization, Learning Rate 0.0001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 132, Optimizer RMSprop, Activation Tanh



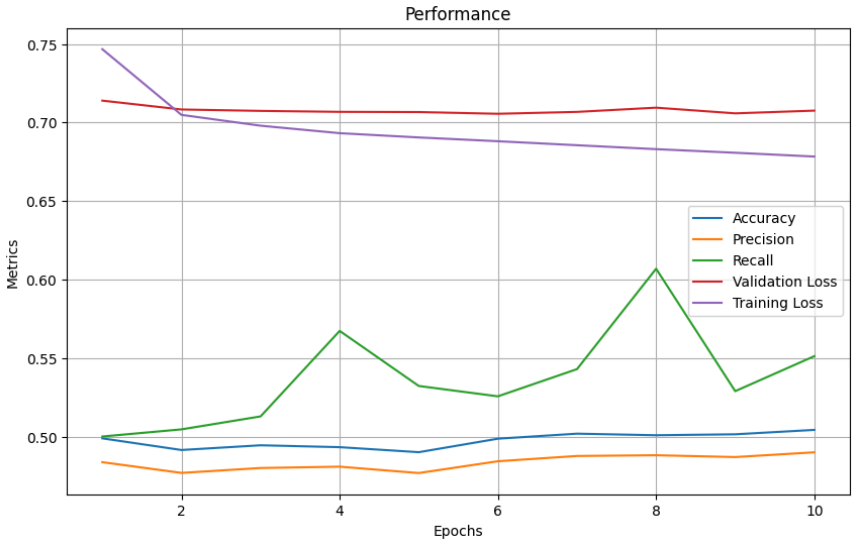
31. Seed 42, Word Level Tokenization, Learning Rate 0.0001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 64, Optimizer Adam, Activation Tanh



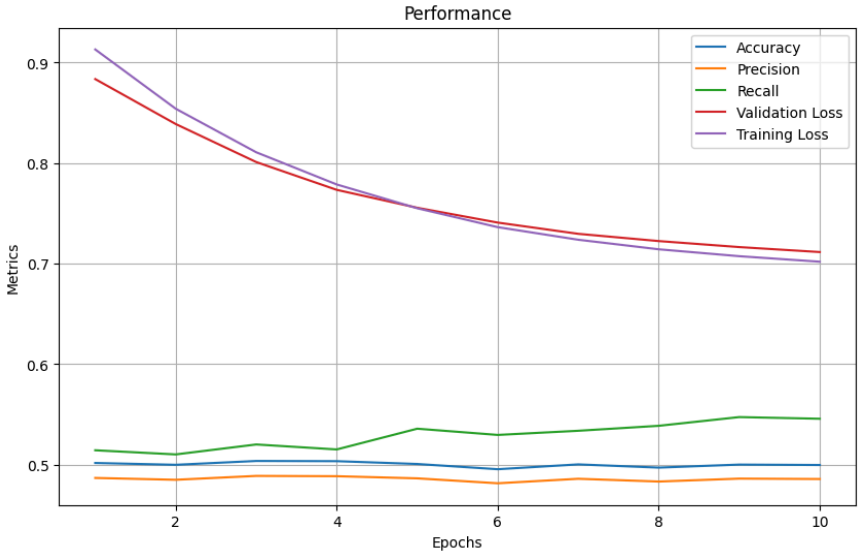
32. Seed 42, Word Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 128, Batch Size 64, Optimizer Adam, Activation Tanh



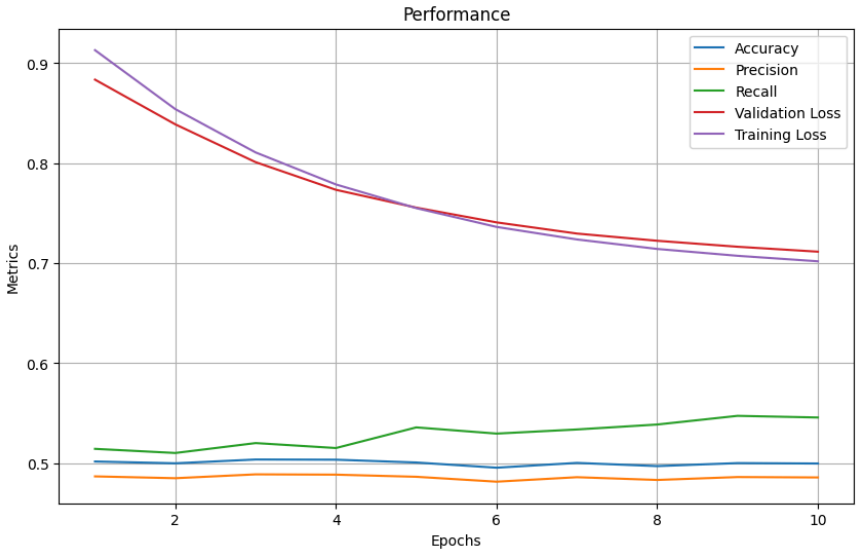
33. Seed 42, Word Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 128, Batch Size 32, Optimizer Adam, Activation Tanh



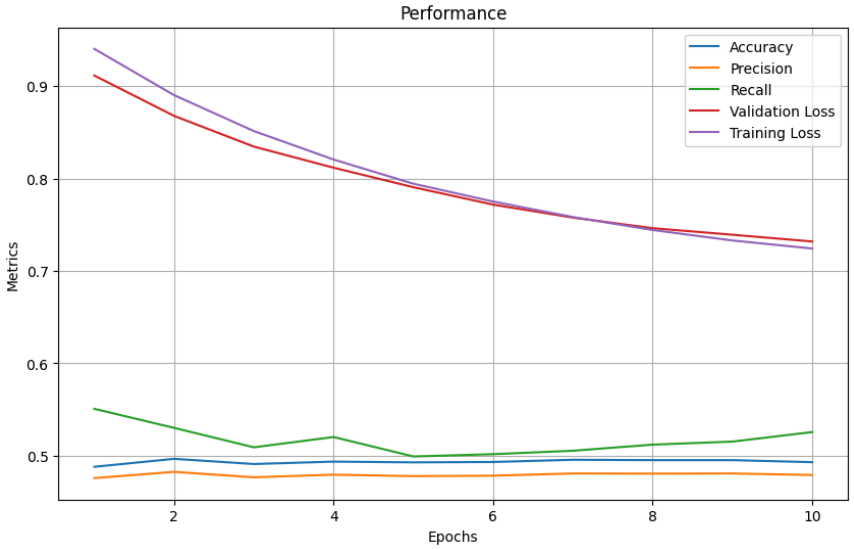
34. Seed 42, Word Level Tokenization, Learning Rate 0.0001, 1 Hidden Layer, Hidden Layer 1st 128, Batch Size 64, Optimizer Adam, Activation Tanh



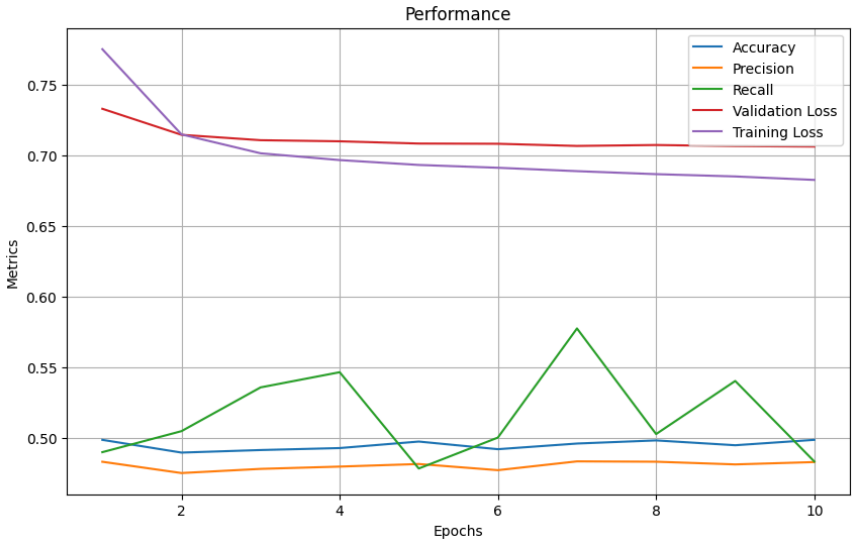
35. Seed 42, Word Level Tokenization, Learning Rate 0.0001, 1 Hidden Layer, Hidden Layer 1st 128, Batch Size 32, Optimizer Adam, Activation Tanh



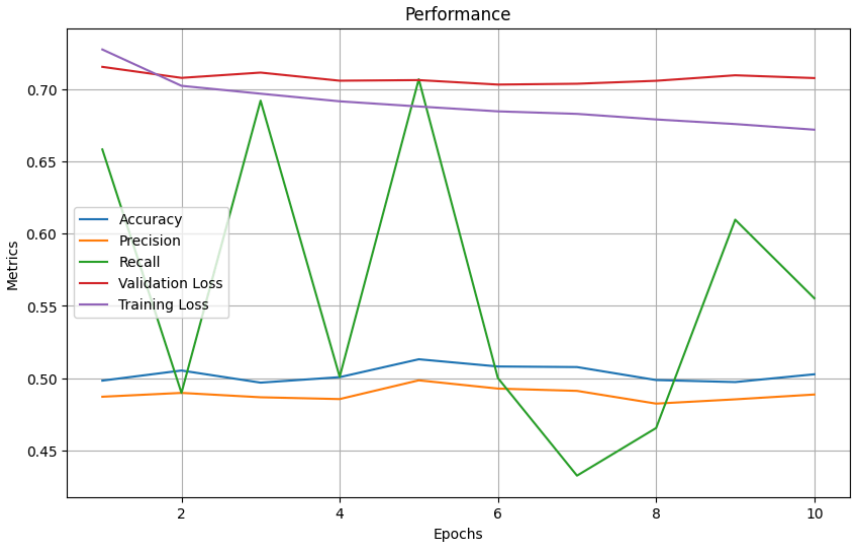
36. Seed 42, Word Level Tokenization, Learning Rate 0.0001, 1 Hidden Layer, Hidden Layer 1st 128, Batch Size 128, Optimizer Adam, Activation Tanh



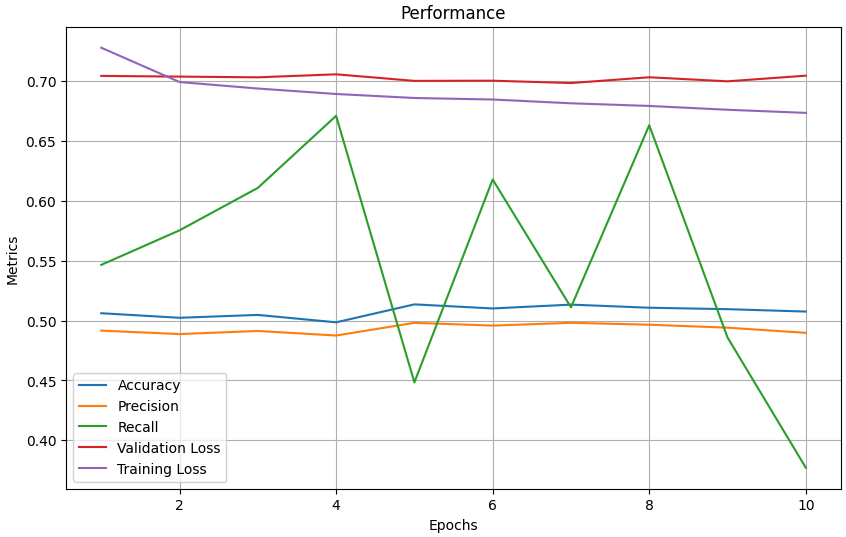
37. Seed 42, Word Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 256, Hidden Layer 2nd 512, Batch Size 64, Optimizer Adam, Activation Tanh



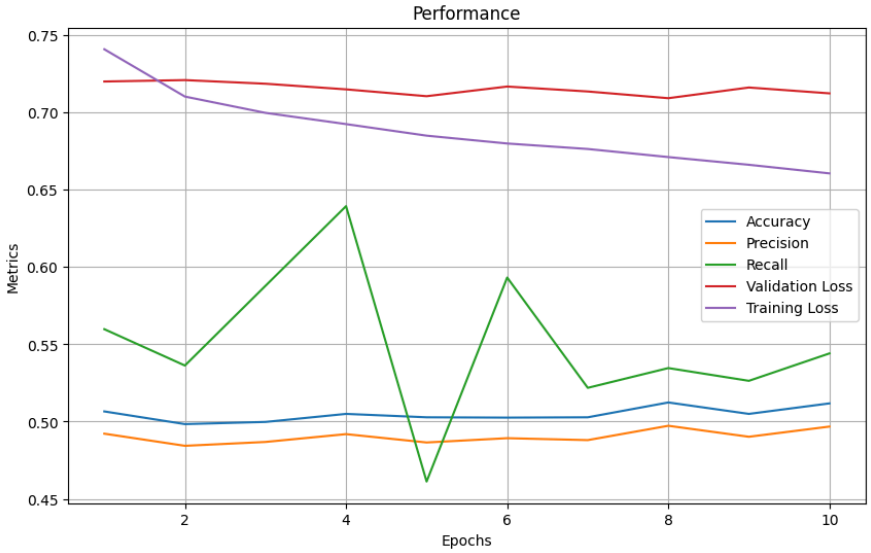
38. Seed 42, Word Level Tokenization, Learning Rate 0.0005, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 128, Batch Size 64, Optimizer Adam, Activation Tanh



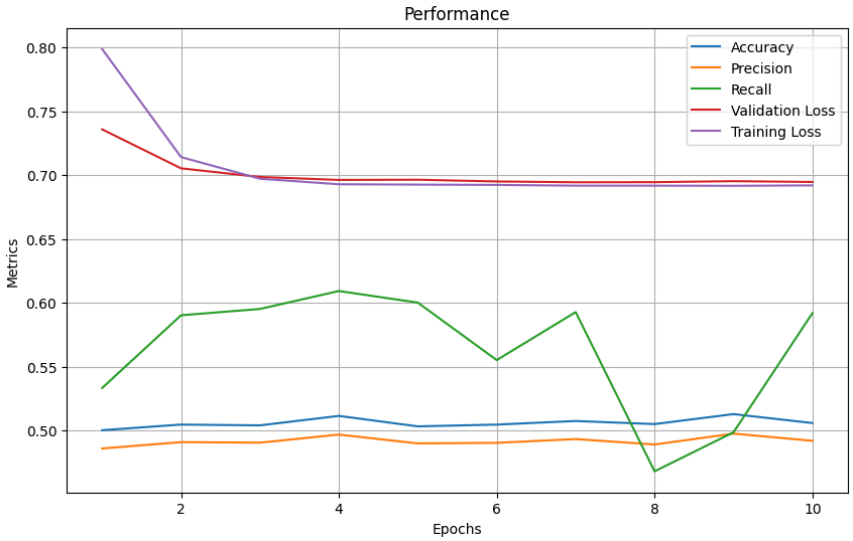
39. Seed 42, Word Level Tokenization, Learning Rate 0.0005, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 128, Batch Size 132, Optimizer Adam, Activation Tanh



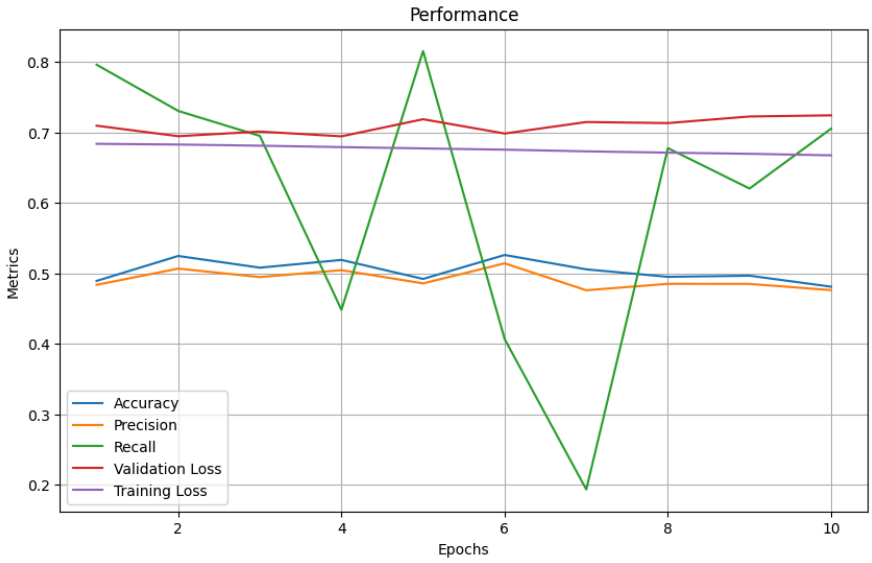
40. Seed 42, Word Level Tokenization, Learning Rate 0.0005, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 64, Optimizer Adam, Activation Tanh



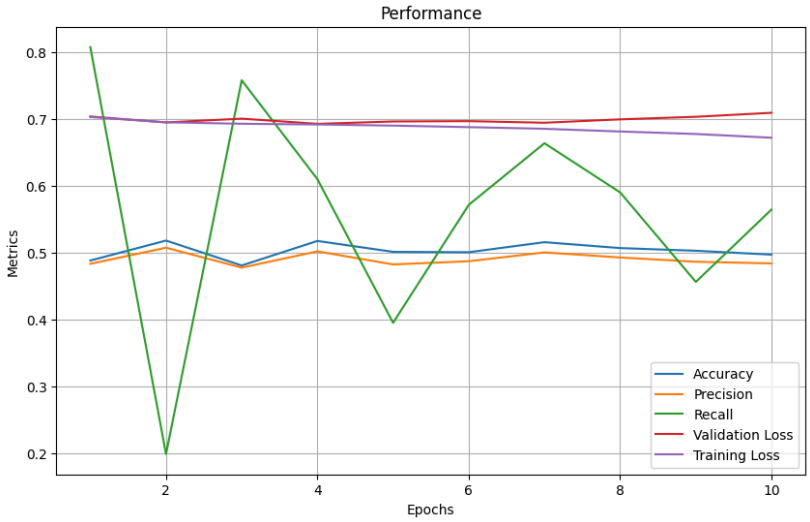
41. Seed 42, Word Level Tokenization, Learning Rate 0.0005, 1 Hidden Layer, Hidden Layer 1st 128, Batch Size 64, Optimizer Adam, Activation Tanh



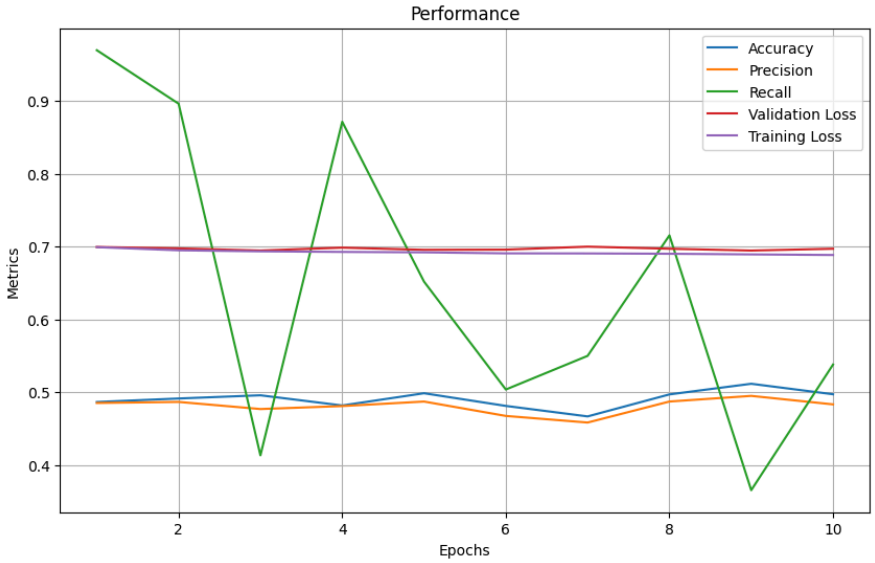
42. Seed 42, Character Level Tokenization, Learning Rate 0.001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 128, Optimizer RMSprop, Activation LeakyReLU



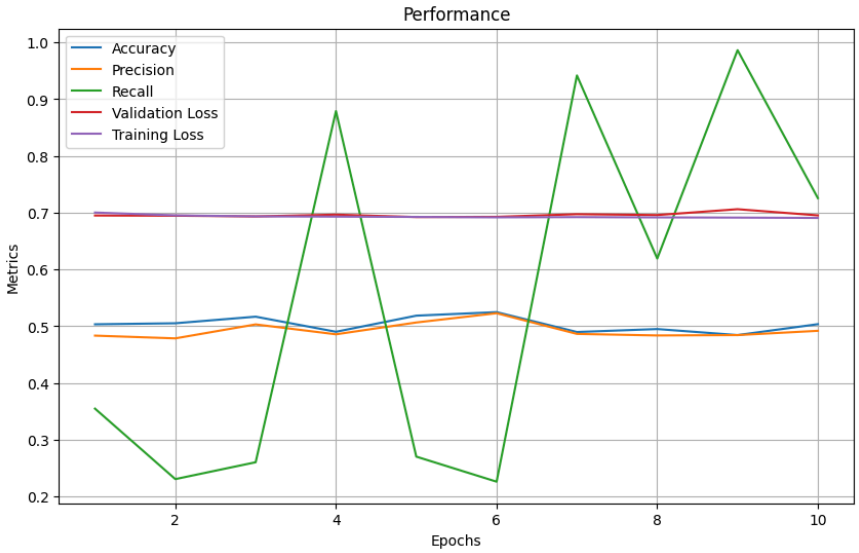
43. Seed 42, Character Level Tokenization, Learning Rate 0.0005, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 32, Optimizer Adam, Activation Relu



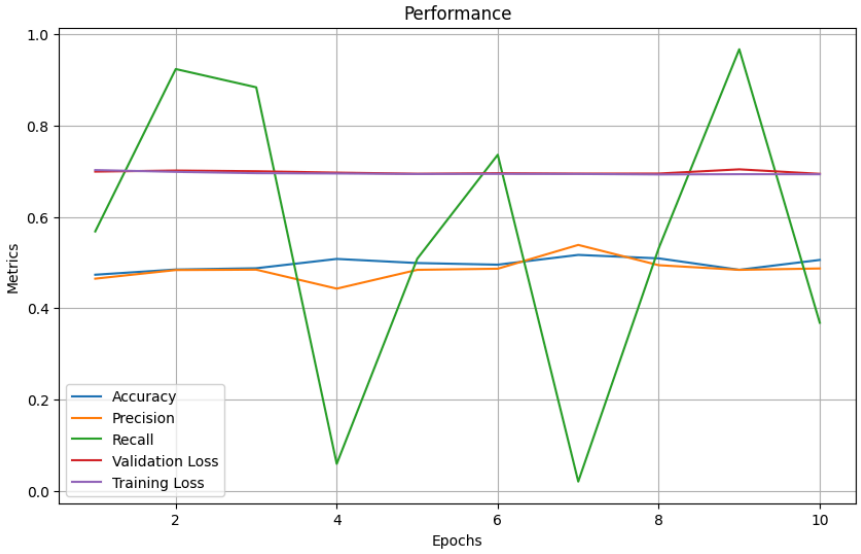
44. Seed 42, Character Level Tokenization, Learning Rate 0.001, 1 Hidden Layer, Hidden Layer 1st 128, Batch Size 32, Optimizer Adam, Activation Relu



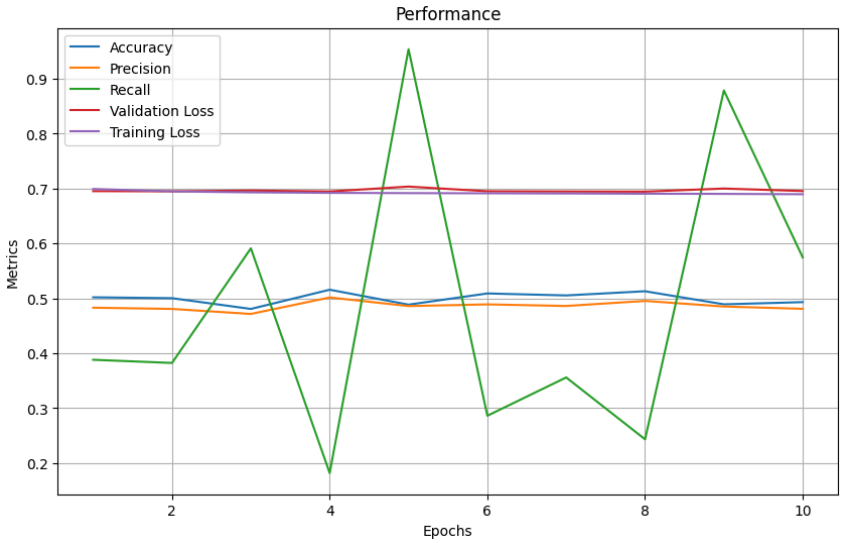
45. Seed 42, Character Level Tokenization, Learning Rate 0.001, 1 Hidden Layer, Hidden Layer 1st 128, Batch Size 32, Optimizer Adam, Activation LeakyReLU.



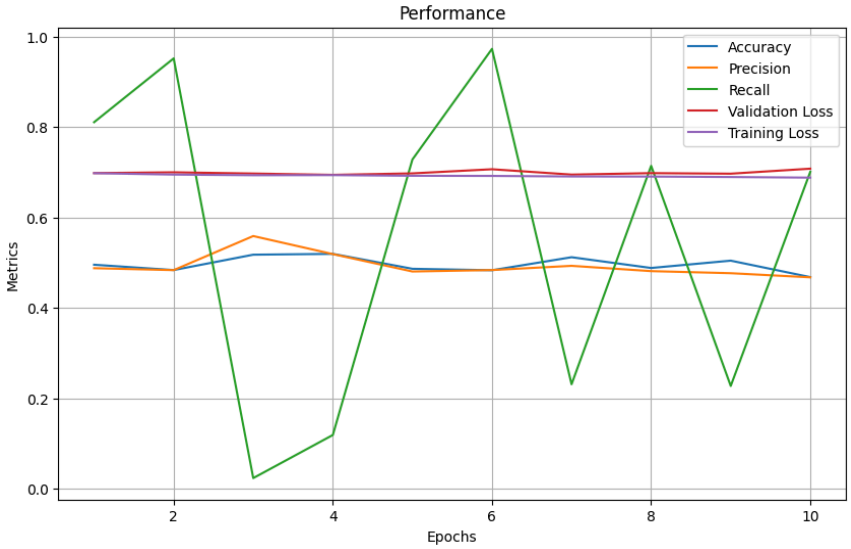
46. Seed 42, Character Level Tokenization, Learning Rate 0.001, 1 Hidden Layer, Hidden Layer 1st 128, Batch Size 32, Optimizer Adam, Activation Tanh



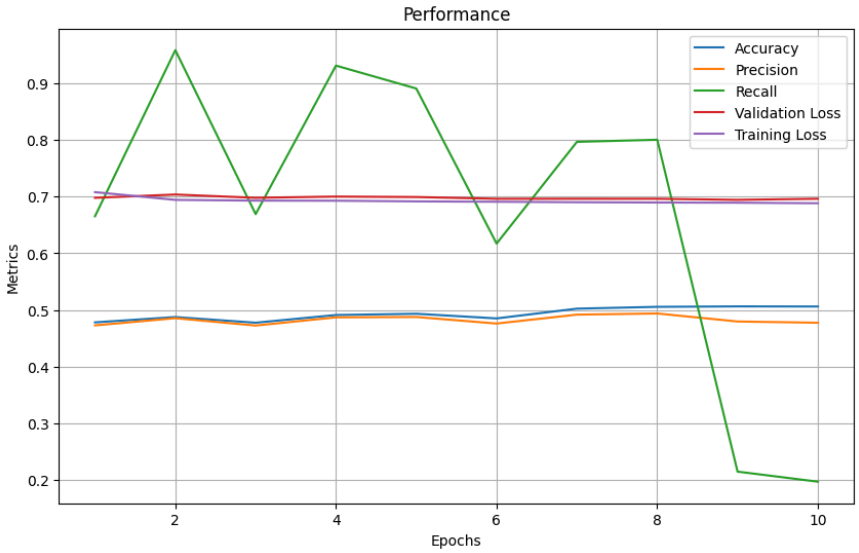
47. Seed 42, Character Level Tokenization, Learning Rate 0.0001, 1 Hidden Layer, Hidden Layer 1st 128, Batch Size 32, Optimizer Adam, Activation Relu



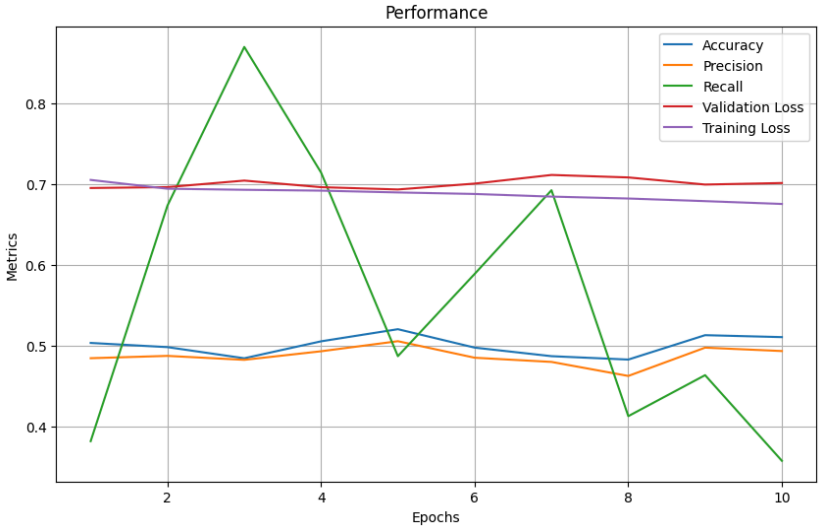
48. Seed 42, Character Level Tokenization, Learning Rate 0.0005, 1 Hidden Layer, Hidden Layer 1st 128, Batch Size 32, Optimizer Adam, Activation Relu



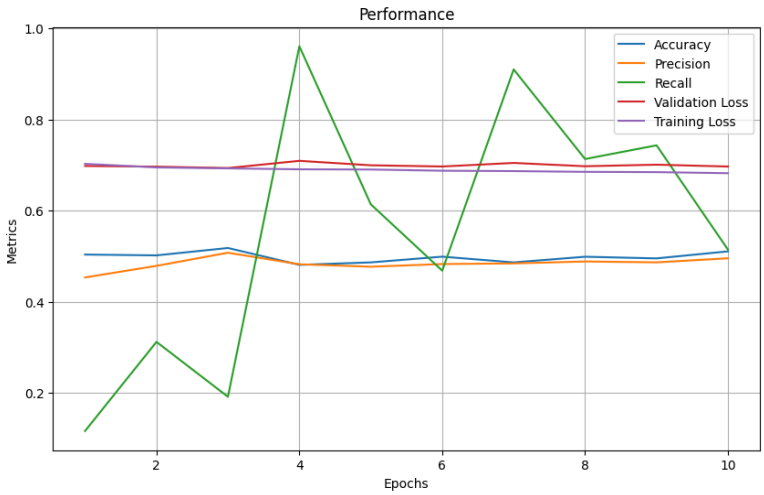
49. Seed 42, Character Level Tokenization, Learning Rate 0.0005, 1 Hidden Layer, Hidden Layer 1st 128, Batch Size 128, Optimizer Adam, Activation Relu



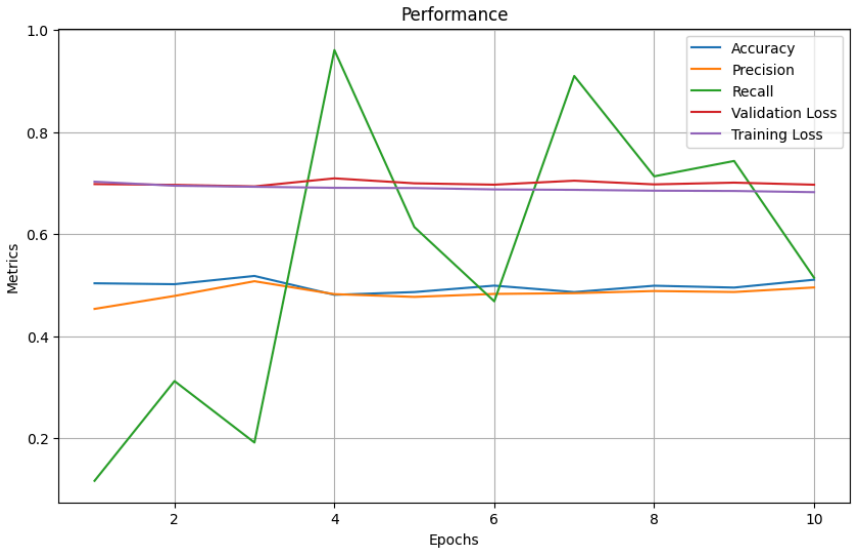
50. Seed 42, Character Level Tokenization, Learning Rate 0.0005, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 512, Batch Size 32, Optimizer Adam, Activation Relu



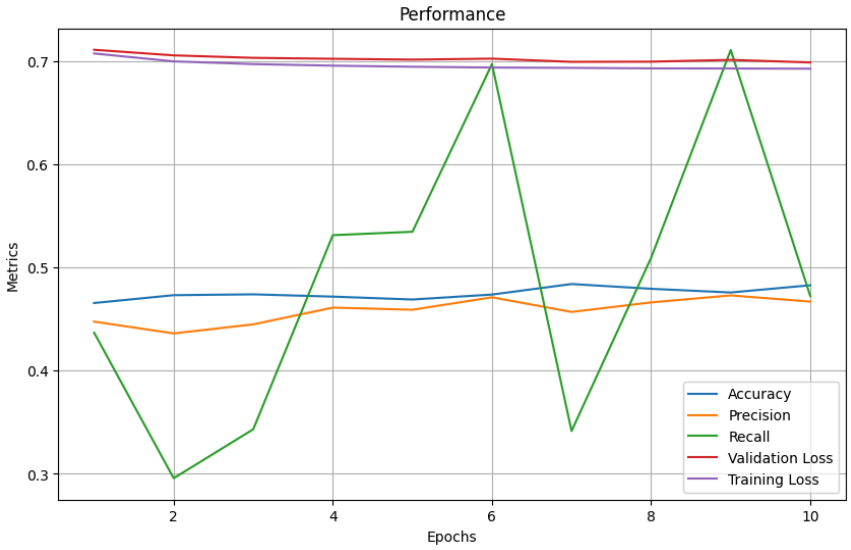
51. Seed 42, Character Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 512, Batch Size 32, Optimizer Adam, Activation Relu



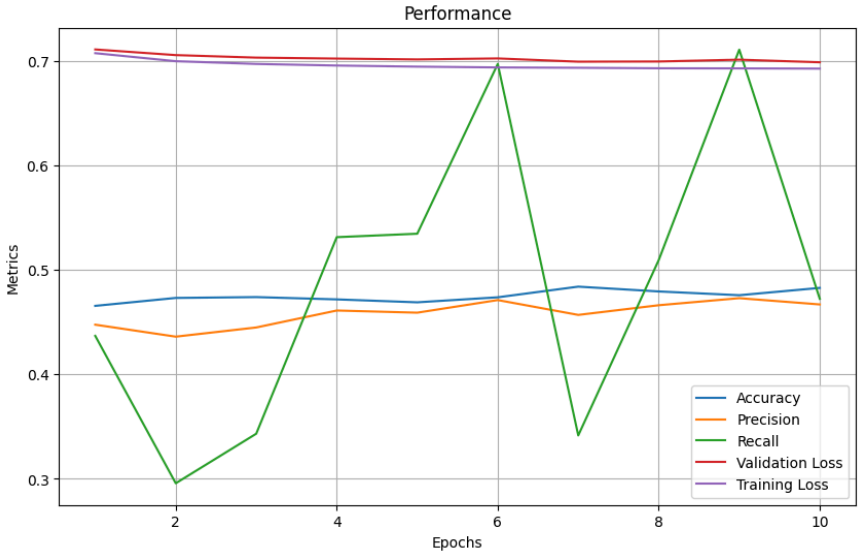
52. Seed 42, Character Level Tokenization, Learning Rate 0.001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 512, Batch Size 32, Optimizer Adam, Activation Relu



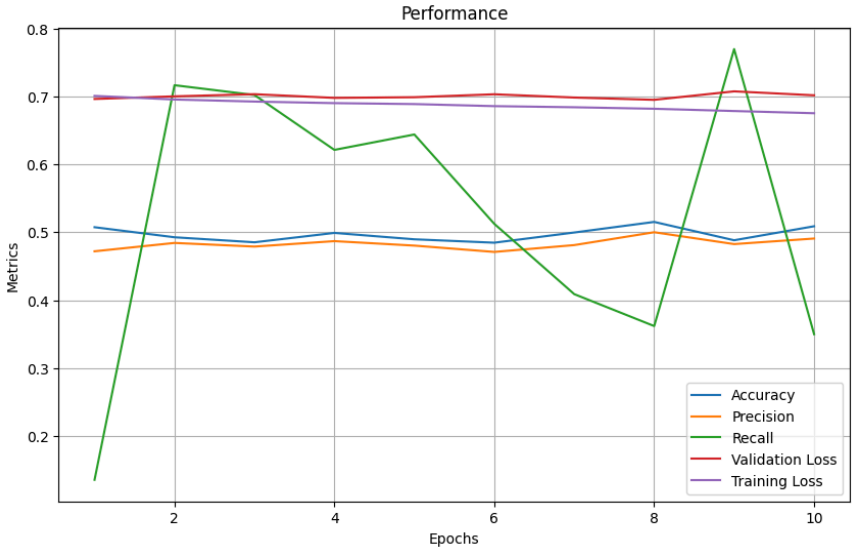
53. Seed 42, Character Level Tokenization, Learning Rate 0.001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 512, Batch Size 32, Optimizer SGD, Activation Relu



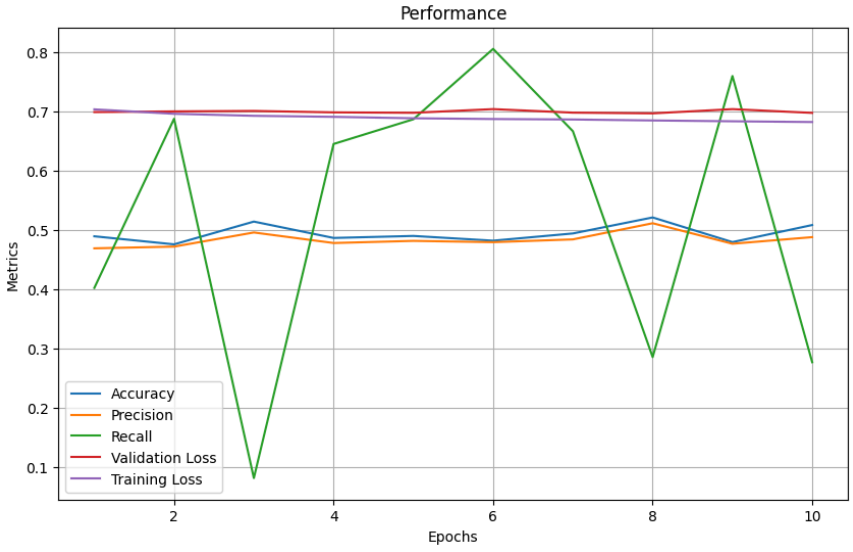
54. Seed 42, Character Level Tokenization, Learning Rate 0.0001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 32, Optimizer Adam, Activation Relu



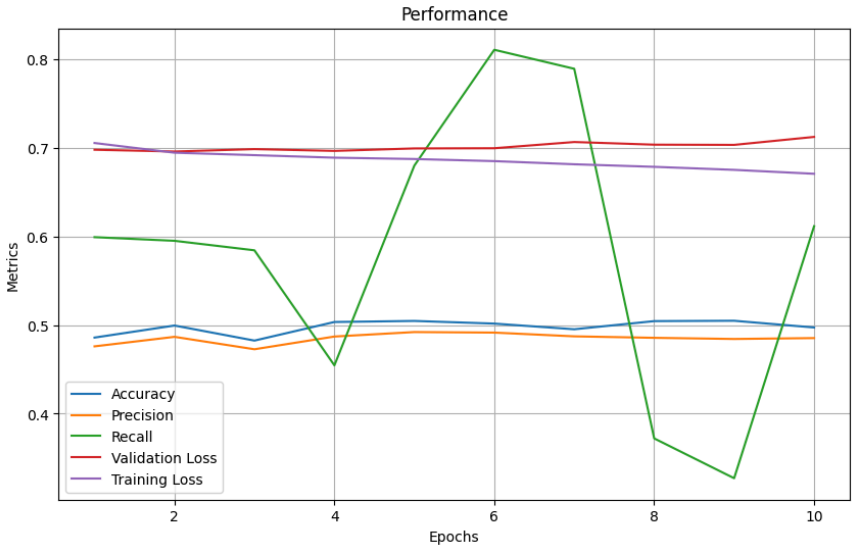
55. Seed 42, Character Level Tokenization, Learning Rate 0.0001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 64, Optimizer Adam, Activation Relu



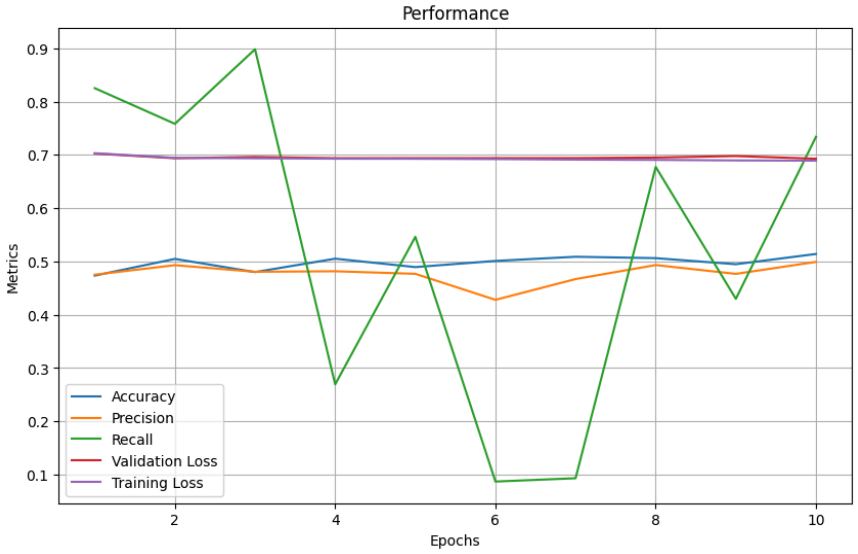
56. Seed 42, Character Level Tokenization, Learning Rate 0.0001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 128, Optimizer Adam, Activation Relu



57. Seed 42, Character Level Tokenization, Learning Rate 0.0005, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 128, Optimizer Adam, Activation Relu

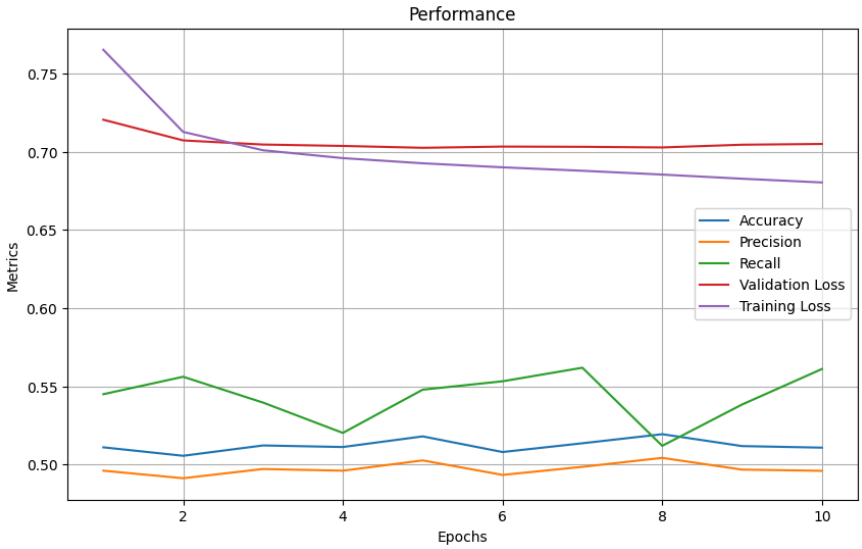


58. Seed 42, Character Level Tokenization, Learning Rate 0.001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 32, Optimizer Adam, Activation Relu



c. Final model performance with mean accuracy and standard error.

The best model: seed 42, Word Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 128, Batch Size 64, Optimizer Adam, Activation Tanh

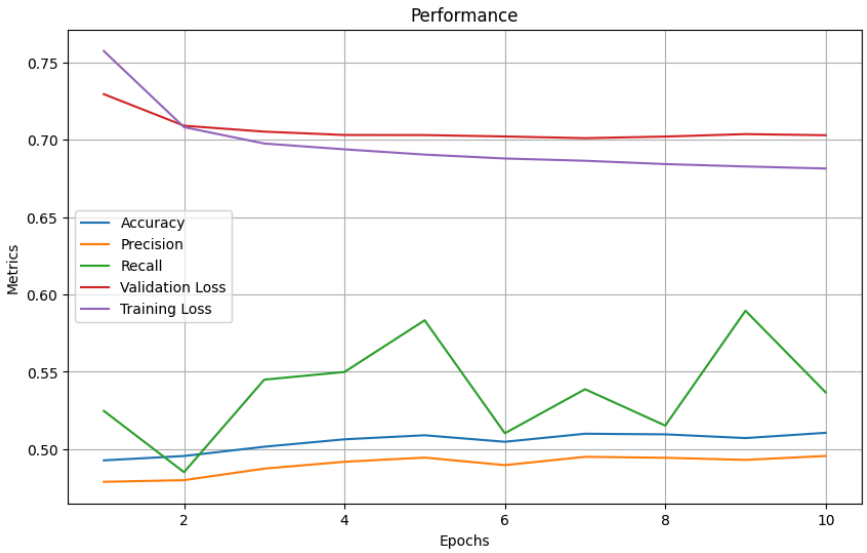


Mean accuracy: 0.5067

Standard error of accuracy: 0.0031

Compare the Seed 42 with other seeds:

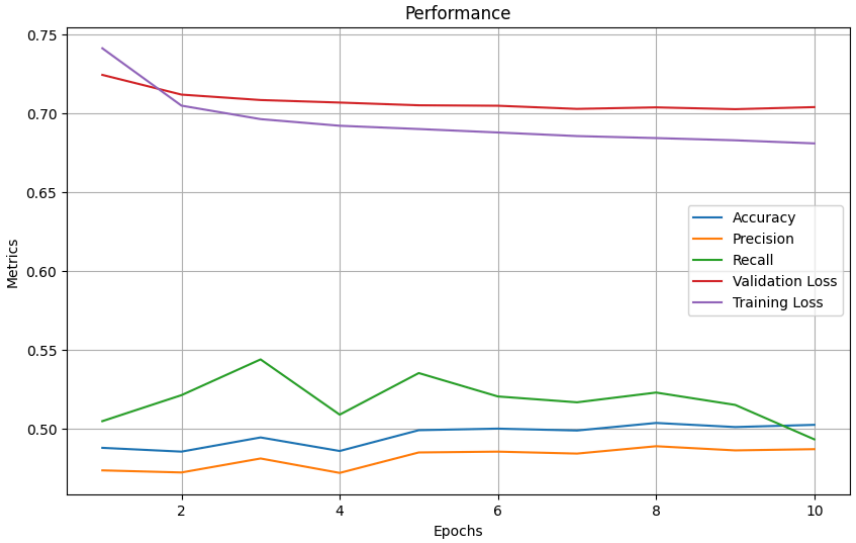
Seed 1:



Mean accuracy: 0.5048

Standard error of accuracy: 0.0059

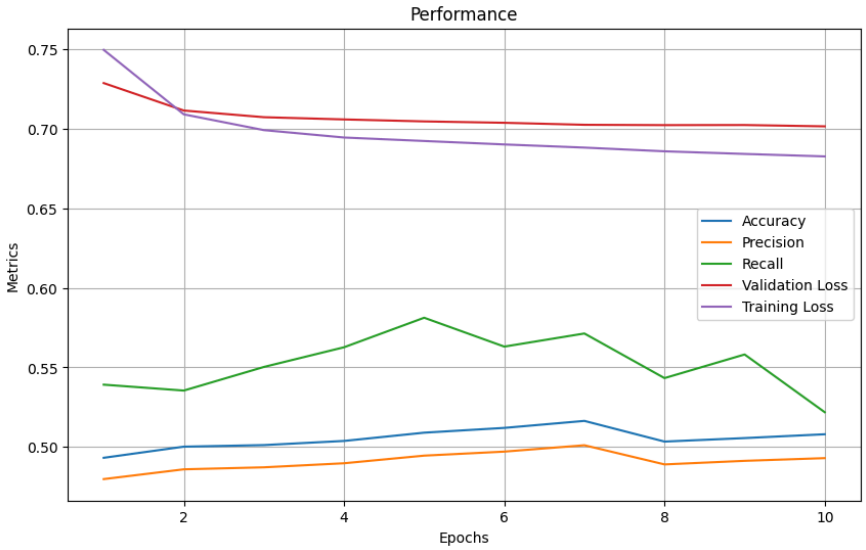
Seed 1234:



Mean accuracy: 0.4956

Standard error of accuracy: 0.0067

Seed 0:



Mean accuracy: 0.5053

Standard error of accuracy: 0.0062

=> In this case, Seed 42 slightly outperforms other seeds.

4. Discussion: Analysis of results, challenges encountered, and insights.

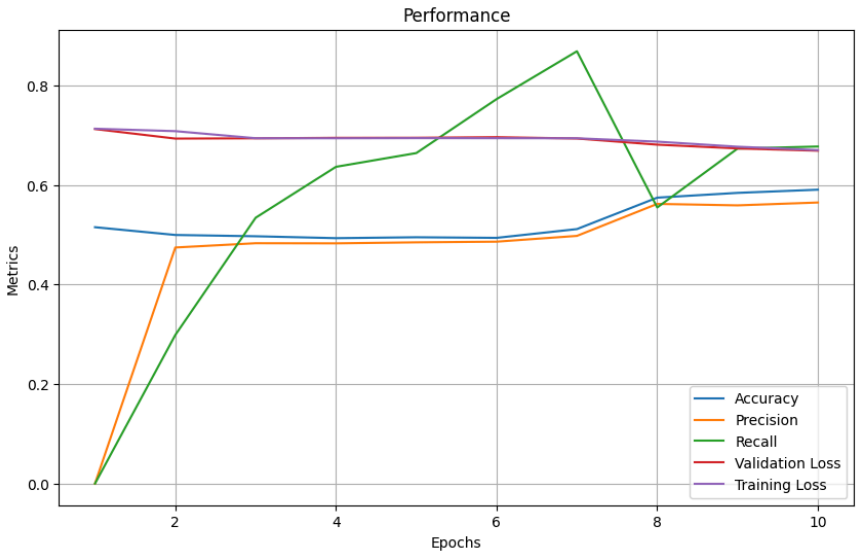
Analysis of results:
Adam optimizer usually works better in this case compared to other optimizers.
The accuracies of the model don't change much during each epoch training indicates that the model don't learn meaningful patterns.
Activation ReLU and LeakyReLU with word level tokenization in this model result in very high training and val loss, the first epoch could be >1000.
The mean accuracy to be 0.5067 indicates that the performance of the model is poor. The standard error of accuracy of 0.0031 indicates that the accuracy through out each epoch is stable.

Challenges:
One of the biggest challenges I encountered during training is the accuracy was very low, and it's always lower than 62%. I thought there was something wrong in my code, so I spent a lot of time to debug it. However, the accuracy was still not improving.
Some of the accuracies of some model are lower than 50%, which indicates that it's lower than random guess in binary classification probability. It also means that the model don't learn meaningful patterns from the data. It was challenging to find parameter values to improve the accuracy, but it didn't improve much.

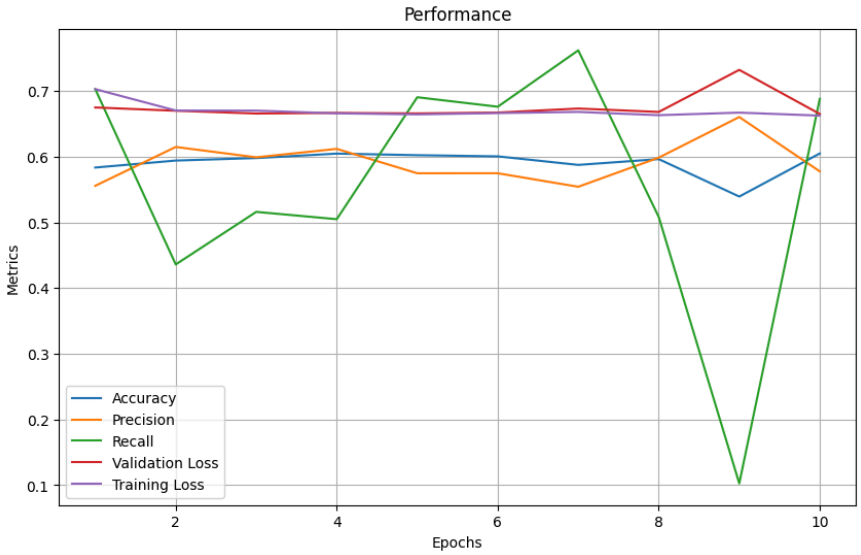
Insights:
One of the biggest insights I learned from this is MLP model is very limit on NLP problems. It doesn't take contexts into account and doesn't support sequence to sequence due to its infrastructure limitation. It performed very poor during in my training model.
Larger batch size results in faster training speed.
Character level tokenizer in this problem works better with ReLU activation function. While word level tokenizer in this problem works better with Tanh activation function.

5. Experiment with MLP_FA: 10 charts

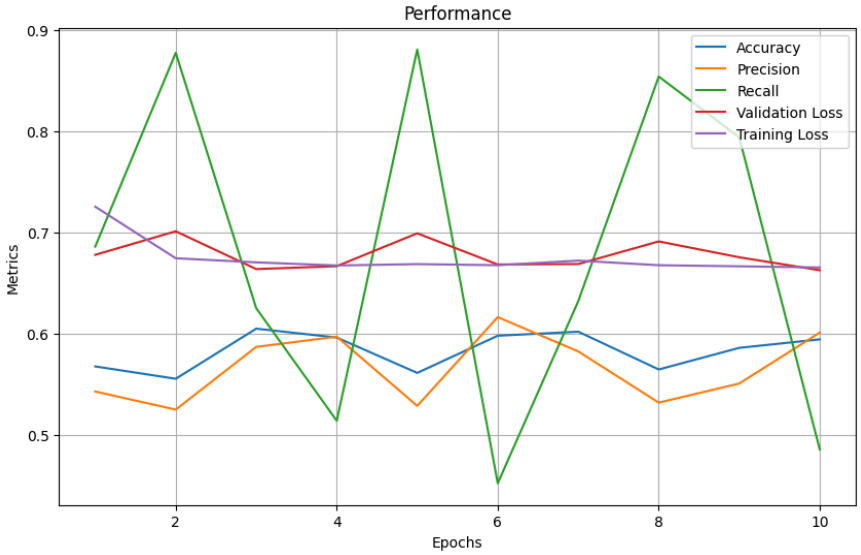
1. Seed 1234, Character Level Tokenization, Learning Rate 0.0001, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 32, Optimizer Adam, Activation Relu



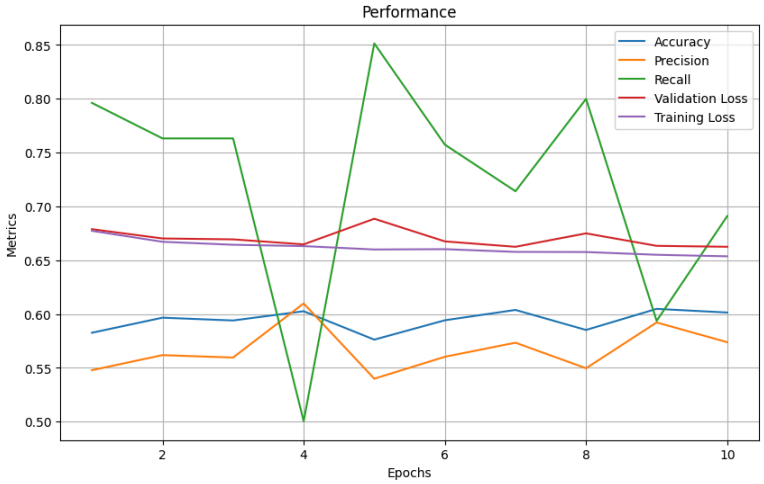
2. Seed 1234, Character Level Tokenization, Learning Rate 0.0005, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 64, Optimizer Adam, Activation Relu



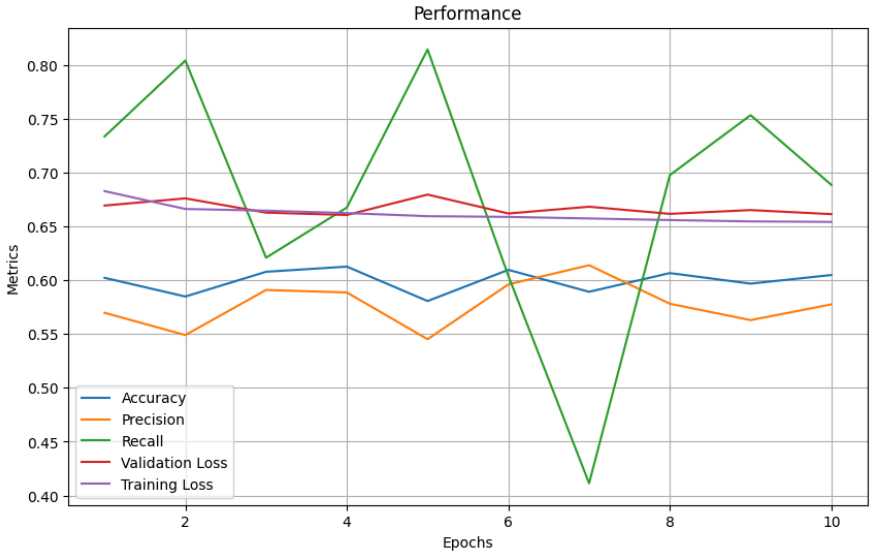
3. Seed 1234, Character Level Tokenization, Learning Rate 0.0005, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 64, Optimizer Adam, Activation Tanh



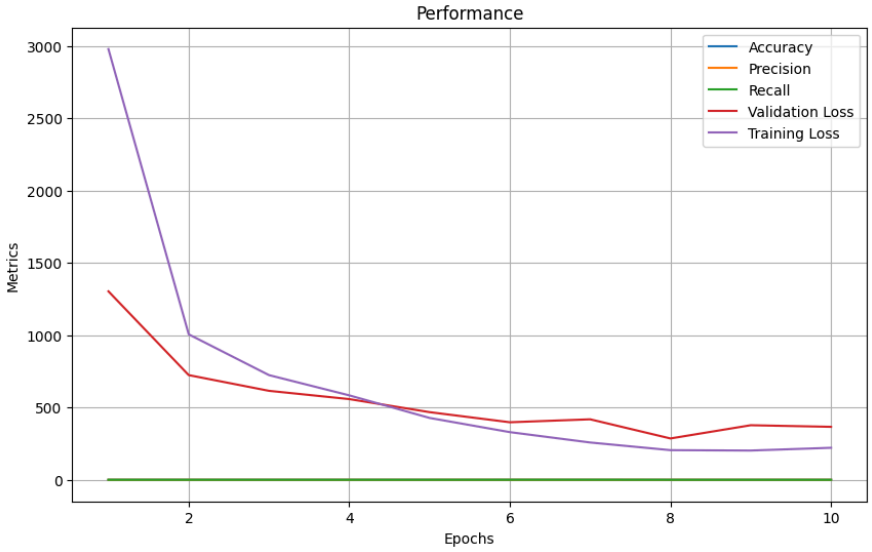
4. Seed 1234, Character Level Tokenization, Learning Rate 0.0005, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 64, Optimizer Adam, Activation ReLU



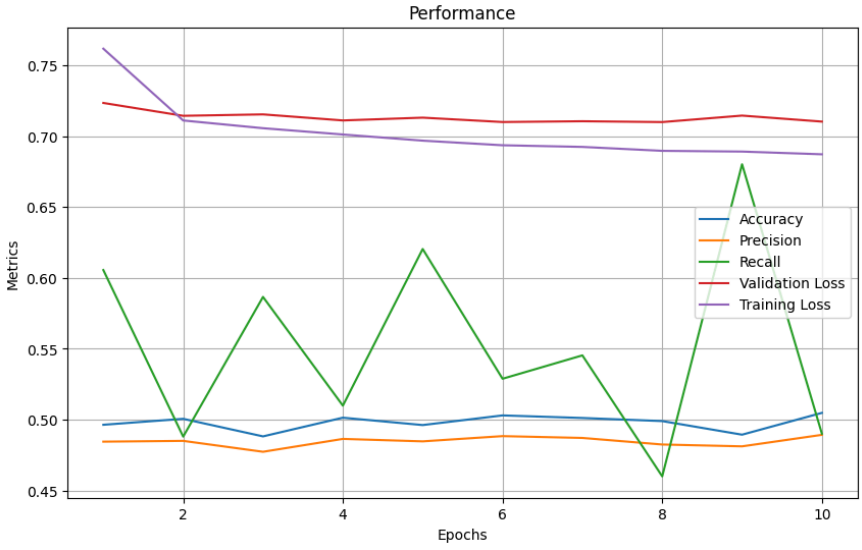
5. Seed 42, Character Level Tokenization, Learning Rate 0.0005, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 64, Optimizer Adam, Activation ReLU



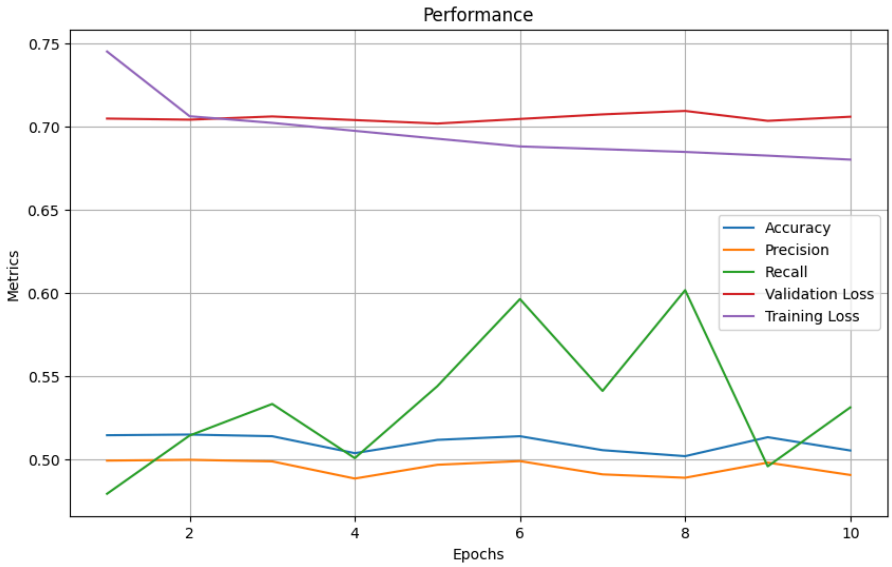
6. Seed 42, Word Level Tokenization, Learning Rate 0.0005, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 64, Optimizer Adam, Activation ReLU



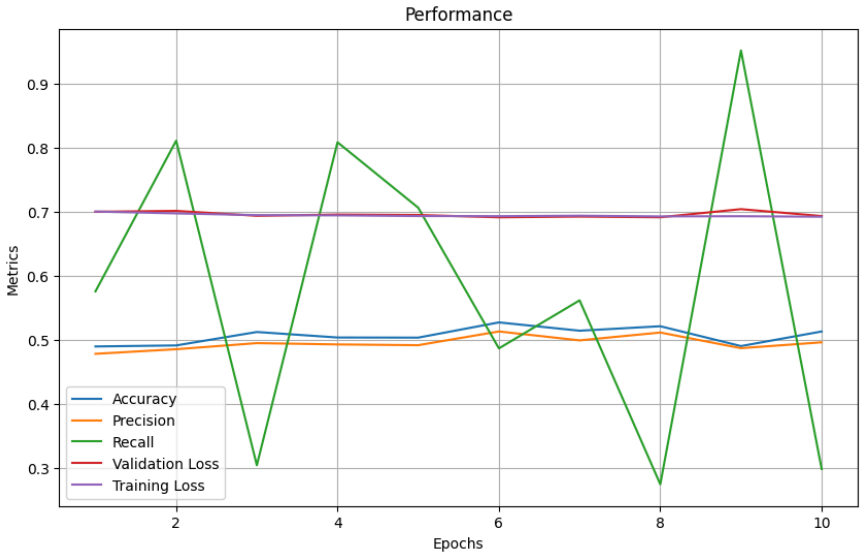
7. Seed 42, Word Level Tokenization, Learning Rate 0.0005, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 64, Optimizer Adam, Activation Tanh



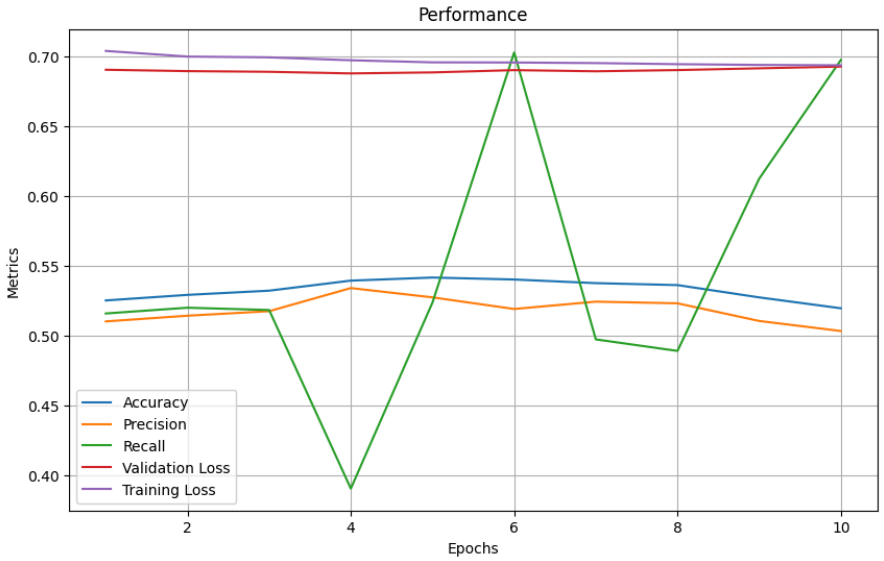
8. Seed 1234, Word Level Tokenization, Learning Rate 0.0005, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 64, Optimizer Adam, Activation Tanh



9. Seed 1234, Character Level Tokenization, Learning Rate 0.0005, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 132, Optimizer Adam, Activation ReLU



10. Seed 1234, Character Level Tokenization, Learning Rate 0.0001, 2 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Batch Size 132, Optimizer Adam, Activation ReLU



=> The Best model after using MLP_FA:

Seed 1234, Character Level Tokenization, Learning Rate 0.0005, 3 Hidden Layers, Hidden Layer 1st 128, Hidden Layer 2nd 256, Hidden Layer 3rd 512, Batch Size 64, Optimizer Adam, Activation Relu

Mean accuracy: 0.5977
Standard error of accuracy: 0.0148

6. Conclusion: Summarize the key findings.

- Vocab size of word level tokenization is higher, but its sequence length is lower than character level tokenization.
- Using activation ReLU and LeakyReLU with word level tokenizer can cause the training loss and validation loss to be extremely high.
- MLP is not a good option for this type of problem of NLP classficationsince it doesn't consider sequential data for NLP, especially in sentiment analysis where the context is important.
- The lowest model performance accuracy and the highest model performance accuracy in this project can be as much as 10%.
- MLP_FA worked better than Random_MLP in some cases.