Report

Group 5: Jashandeep Singh, Jordan Cao

Assignment3: Task 1

Part A

A steady decrease in training and validation loss shows that the model is learning and improving its performance. When both losses stop decreasing and level off, it means the model has likely reached convergence. For the given hyperparameters, both training and validation loss plateau after 125 epochs, showing no significant changes. At this point, early stopping is applied to prevent overfitting and save resources. Since both testing and validation loss decrease and remain stable after a certain point, it indicates that the model is well-fitted to the data. The good fit is further confirmed by the error of testing, validation and testing. The lower test error could indicate robust generalization.

Fig 1

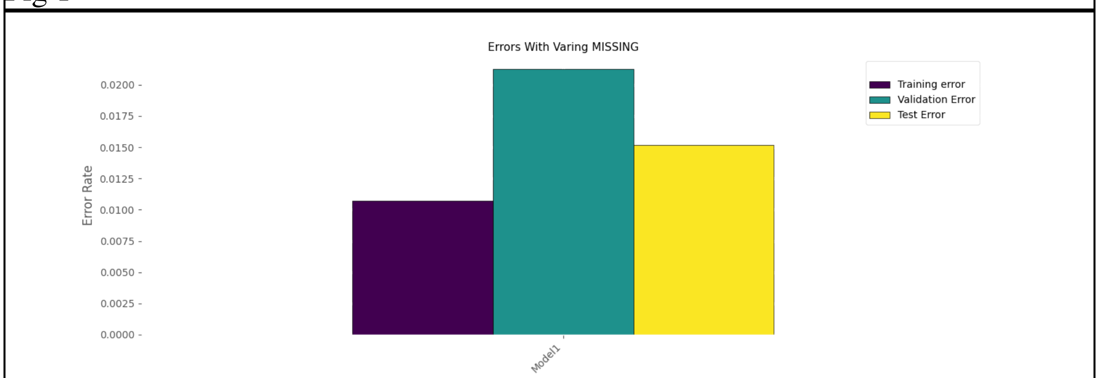

Fig 2

Part B

A neural network can be considered to have converged when the training error (or loss) stops decreasing or has reached a minimum level of acceptable error.

Changes in Learning Rate While Keeping Other Hyperparameters Constant

The learning rate determines how much the model's weights are adjusted during each training step. When training a model with learning rates ranging from 0.0001 to 1.0, we can see that smaller learning rates take more epochs to reach convergence because the weight updates are smaller in each step. On the other hand, larger learning rates require fewer epochs but often lead to a suboptimal solution, as seen in higher training, validation, and testing errors compared to smaller learning rates. The testing, training, and validation errors stay roughly the same as with smaller rates. This means the model trains faster, for .01 learning rate, without sacrificing much accuracy, making this value a good balance between speed and stability.

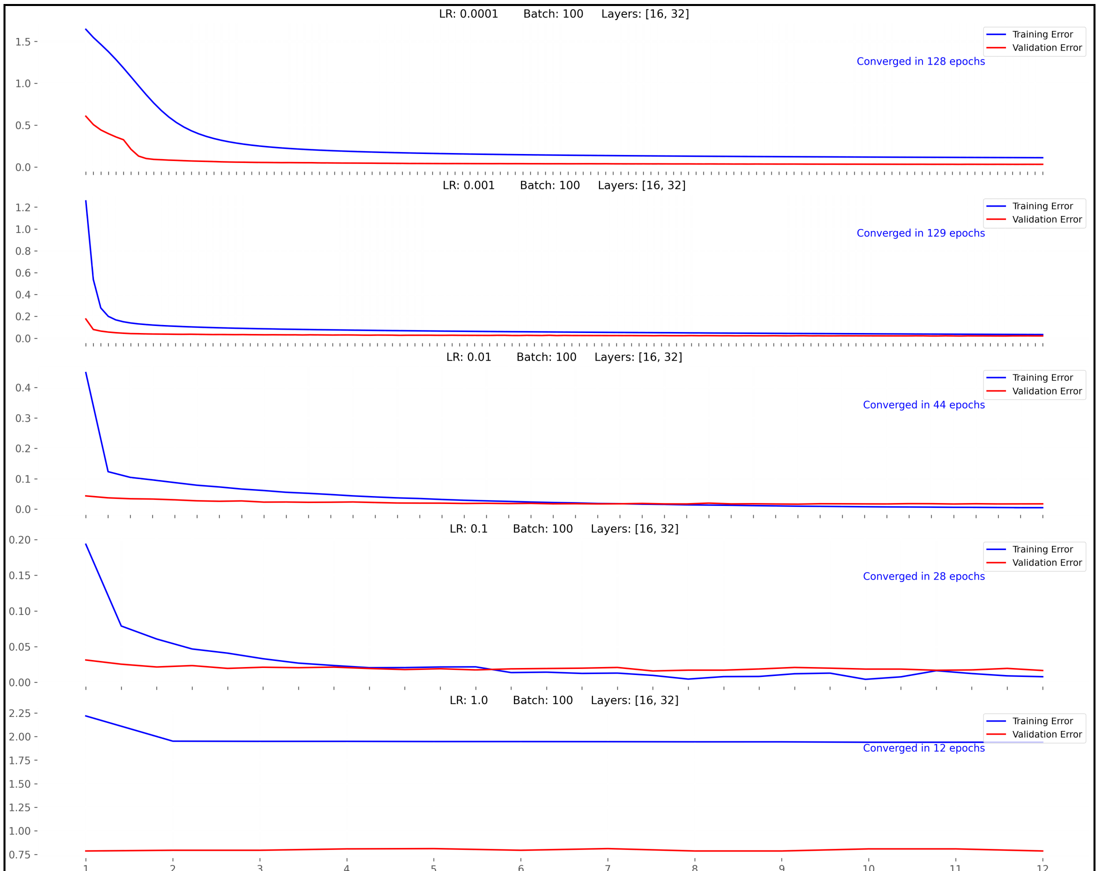

Fig 3

Fig 4

Changes in Batch Sizes While Keeping Other Hyperparameters Constant

From the training, testing, and validation errors, it is clear that larger batch sizes result in poor model generalization. As the batch size increases, the error rates across all three datasets also rise. Based on test results using five different batch sizes, smaller batch sizes not only reach convergence faster but also generalize better.

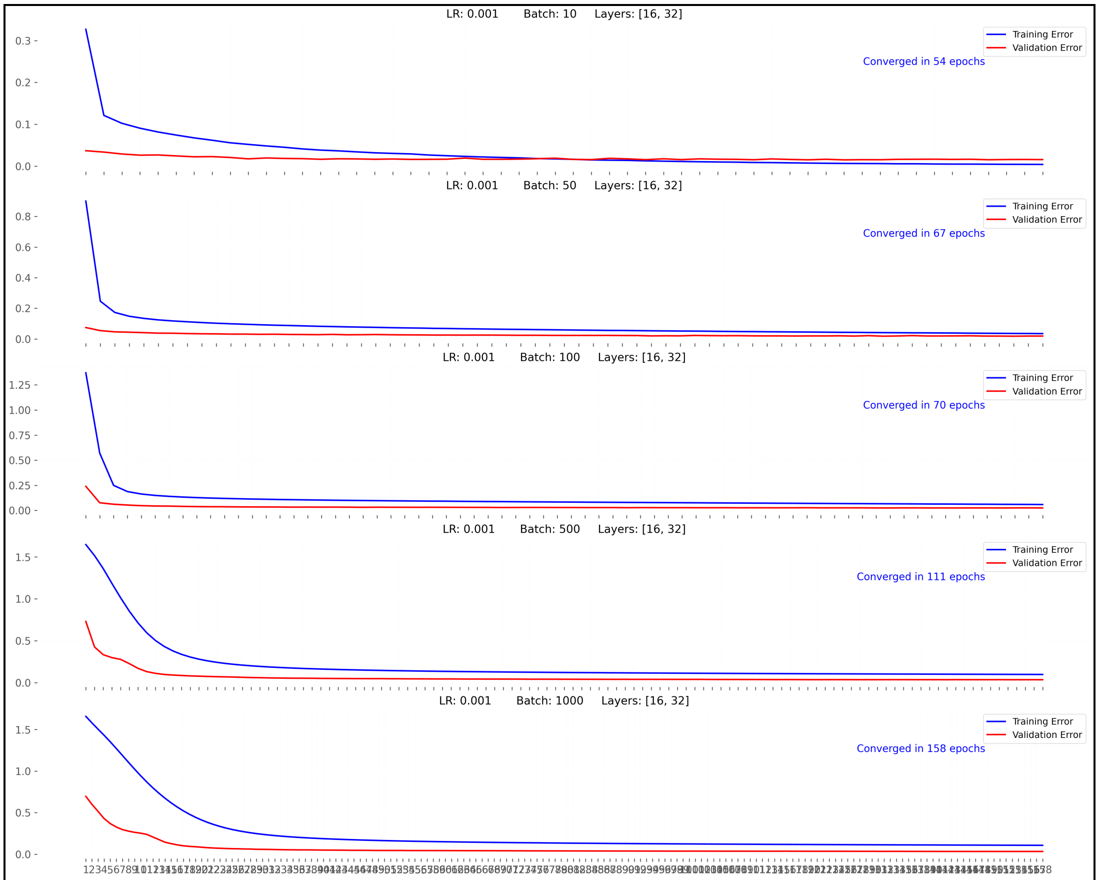

Fig 5

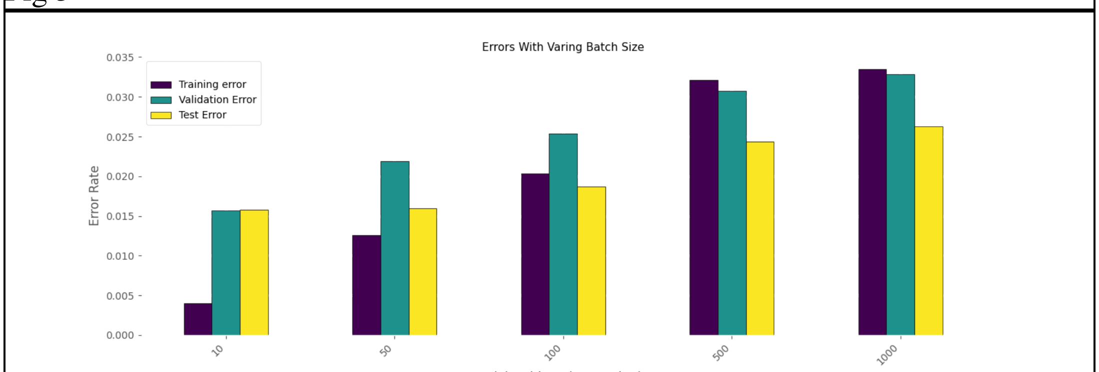

Fig 6

Part C

Changes in Number of Neurons While Keeping Other Hyperparameters Constant

From the experiment conducted understand how changing the number of neurons in the hidden results show that different number of neurons take a similar number of epochs to reach convergence. But it a significant difference is seen in how well the model generalizes as the number of neurons increases. Models with a higher number of neurons in the hidden layer tend to perform better, as they can capture more complex patterns in the data. This suggests that increasing the number of neurons helps the model make more accurate predictions on new, unseen data.

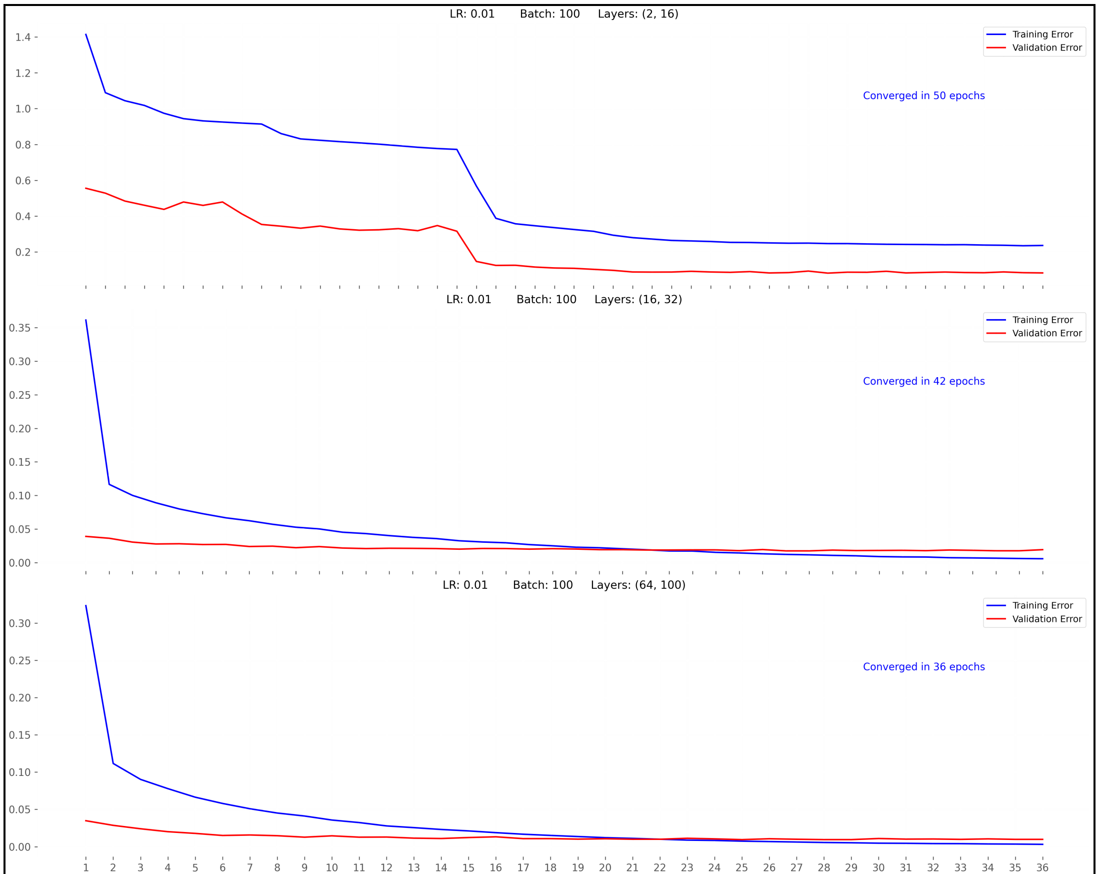

Fig 7

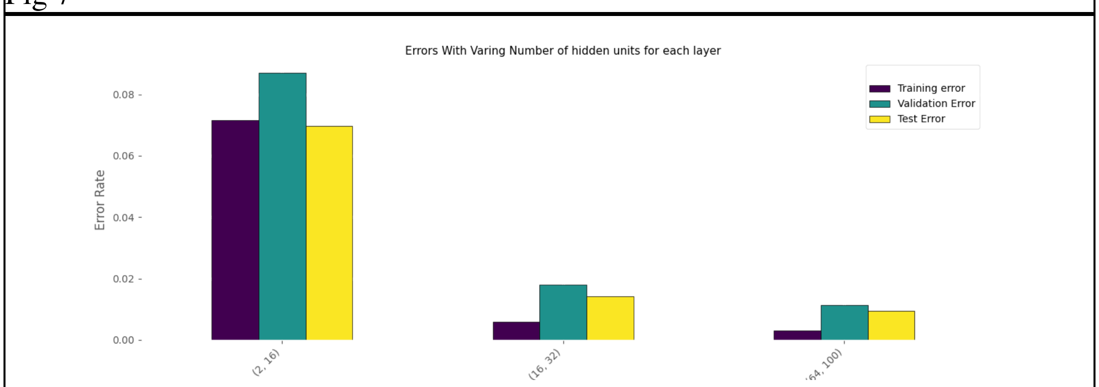

Fig 8

Part D

Based on the test results, the best parameters were chosen: a learning rate of 0.01, a batch size of 100, and 100 units in the hidden layer. Using these optimal hyperparameters, the model was initialized, trained, validated, and tested. Below is the confusion matrix for the best-performing network.

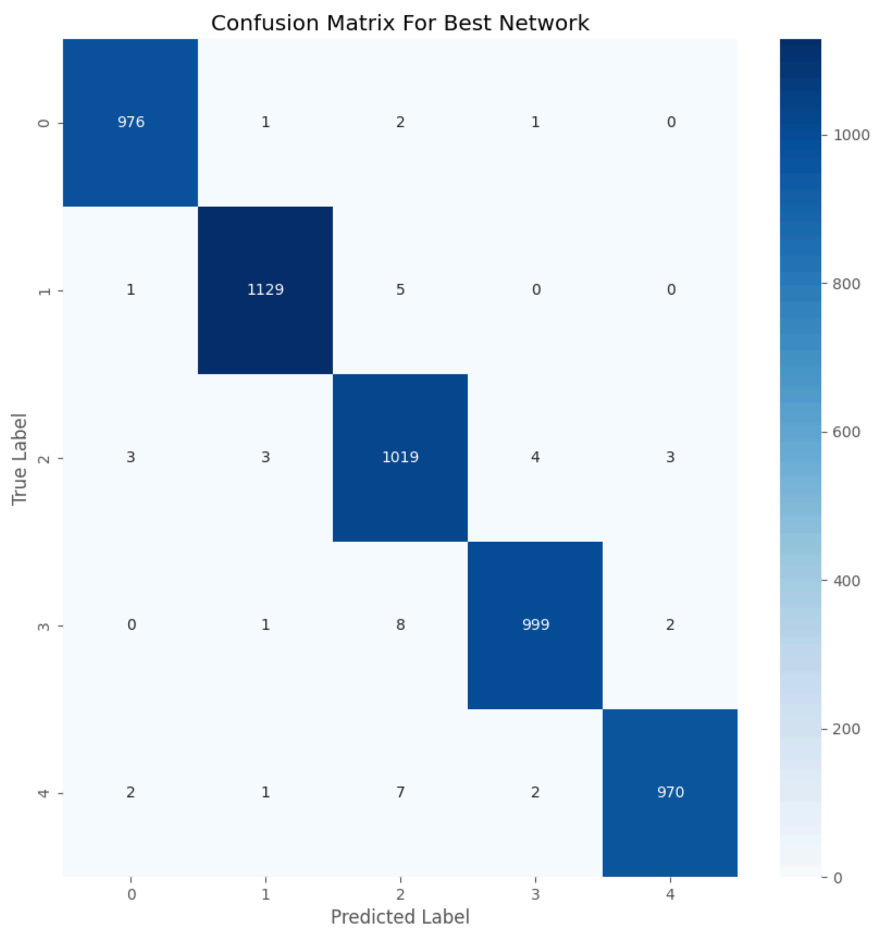

Fig 9

Naïve Bayes News Classifier - Results Analysis

Performance Metrics

- **Accuracy**: The model achieved an accuracy of **78.3%** on the test dataset.
- **Confusion Matrix**:

Actual \ Predicted	Fake (0)	Real (1)
Fake (0)	261	128
Real (1)	85	506

Key Observations

1. **False Positives (128 cases)**: Some fake headlines were misclassified as real, indicating overlap in language usage between fake and real news.
2. **False Negatives (85 cases)**: Some real headlines were classified as fake, there is some room to optimize this algorithm.
3. **Model Bias**: The classifier performs better at detecting real news but has a slightly higher misclassification rate for fake news.