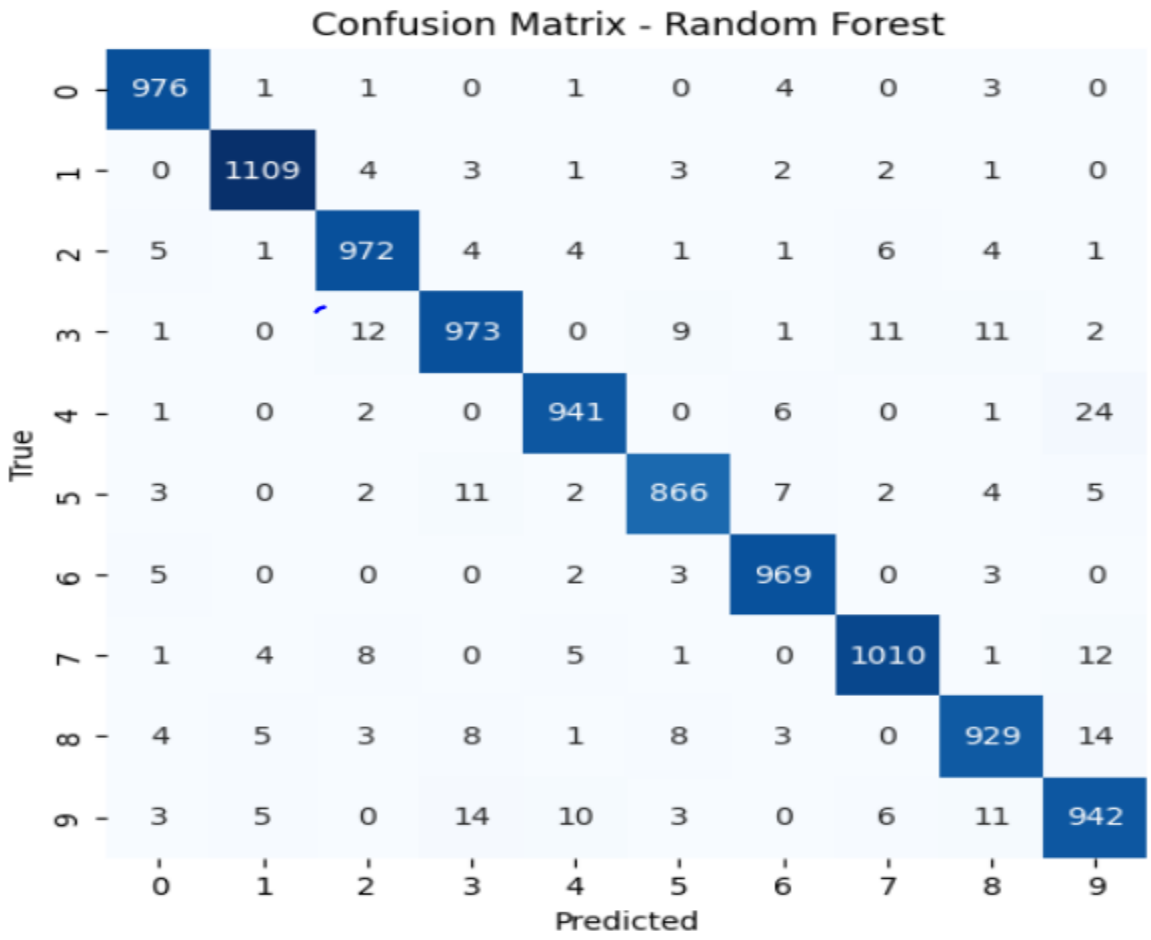
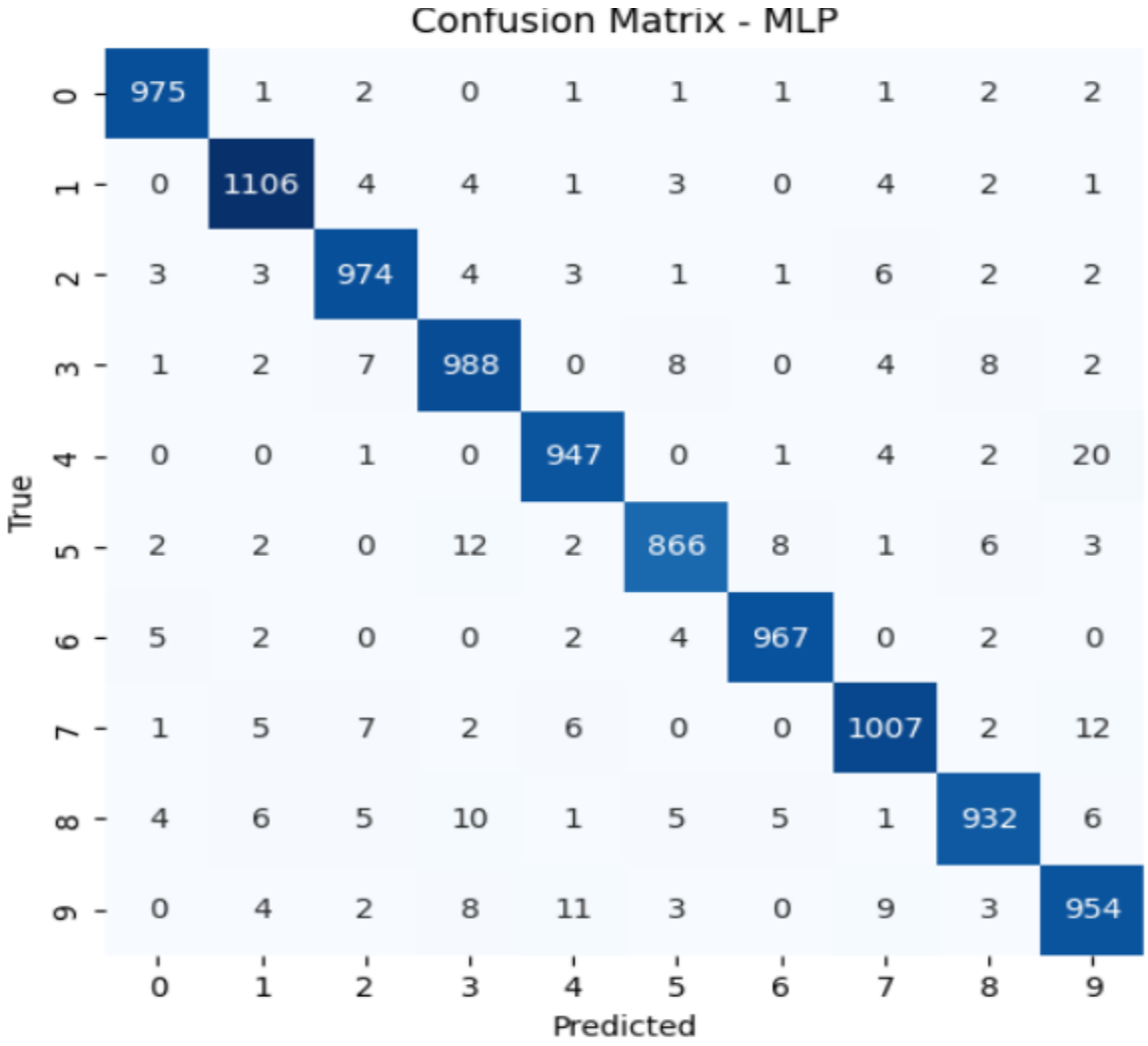
Model: Random Forest
F1-score: 0.9686810271583829

Confusion Matrix:



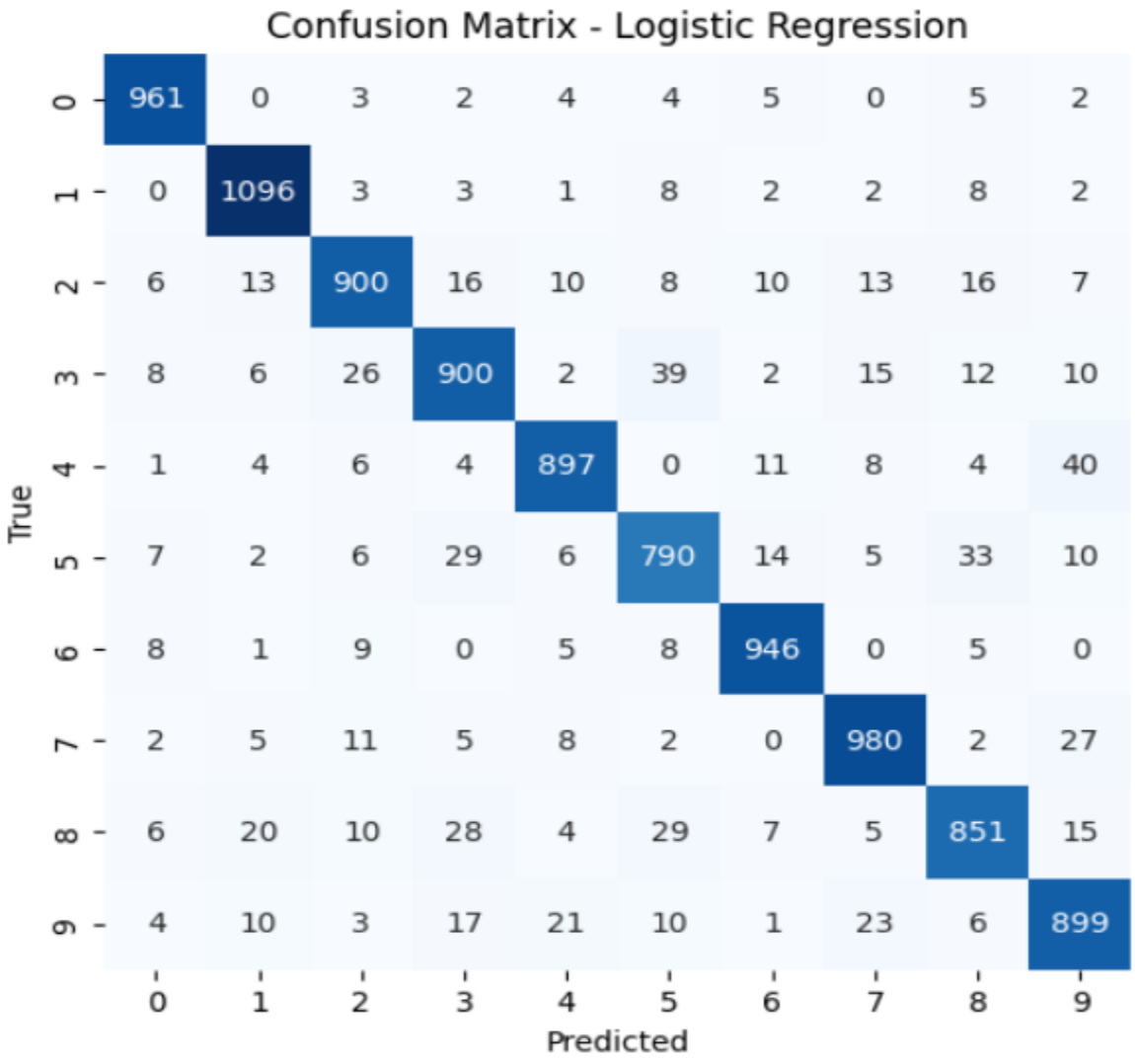
Model: MLP
F1-score: 0.9715917529638729

Confusion matrix:



Model: Logistic Regression

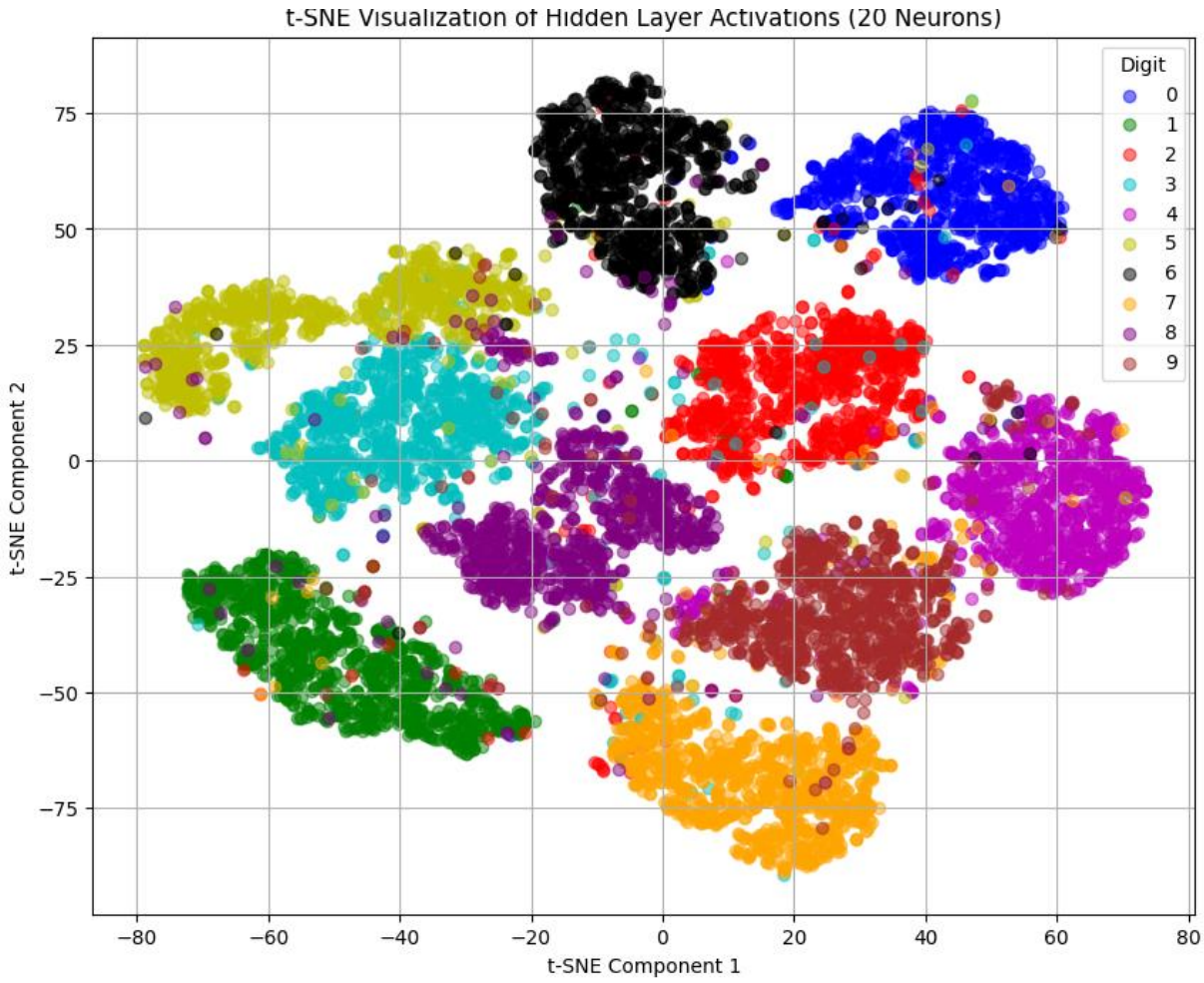
F1-score: 0.9217920708506417



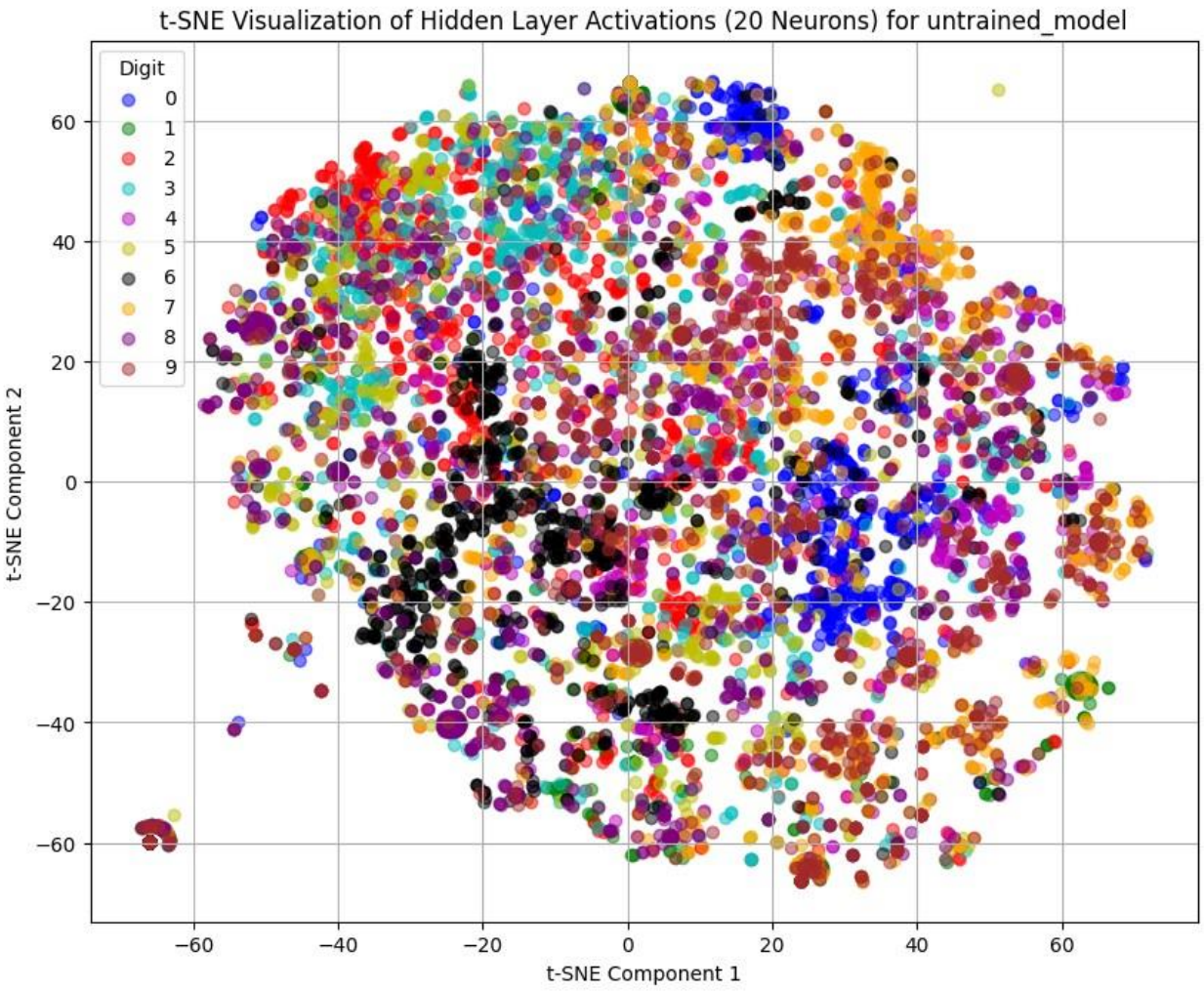
We observe that the MLP classifier has the highest F1-score among the three classifiers used. The digits that are commonly confused are:-

- 4 and 9
- 3 and 5
- 3 and 8

t-SNE plot to visualize the output of the layer containing 20 neurons



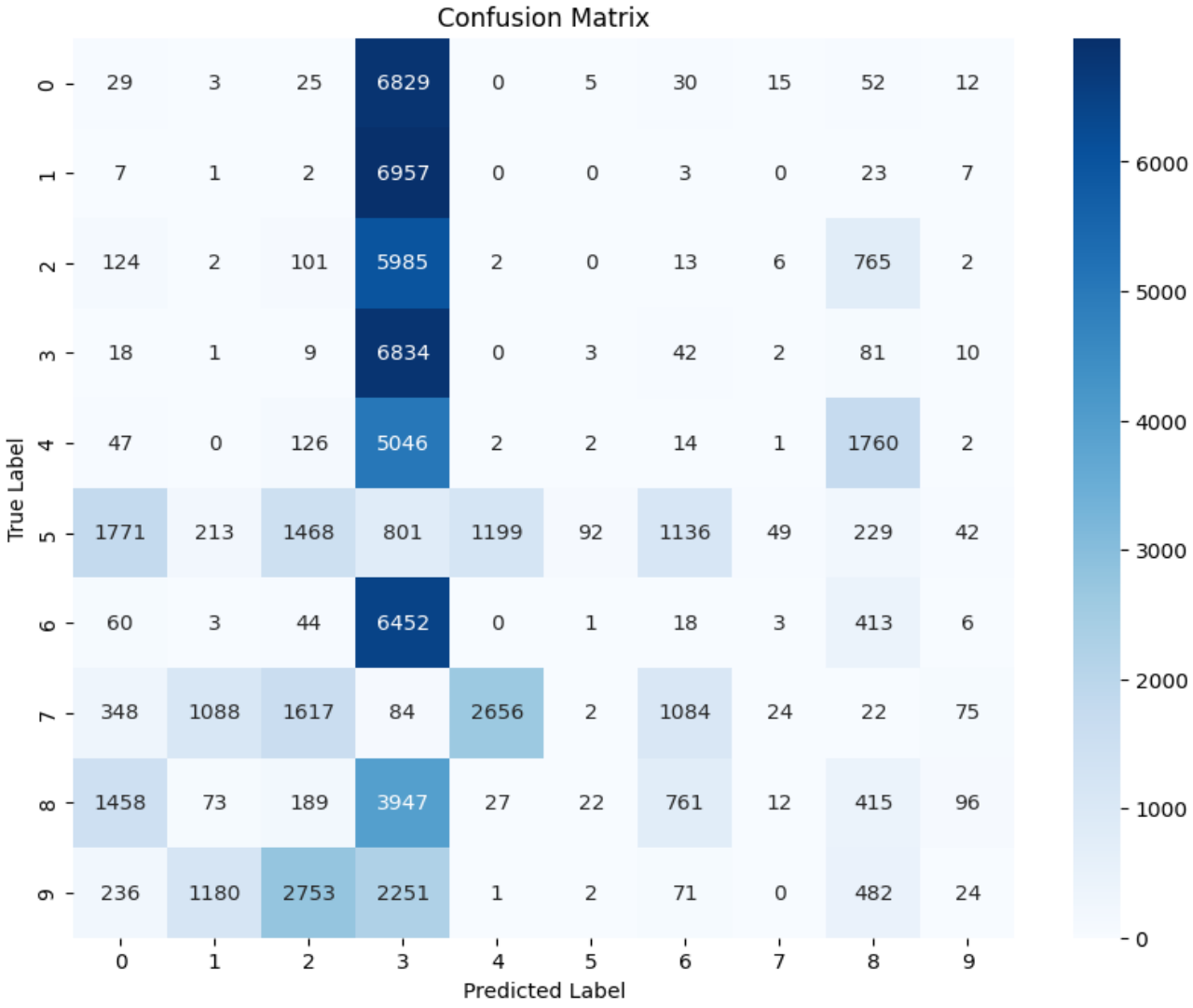
For untrained



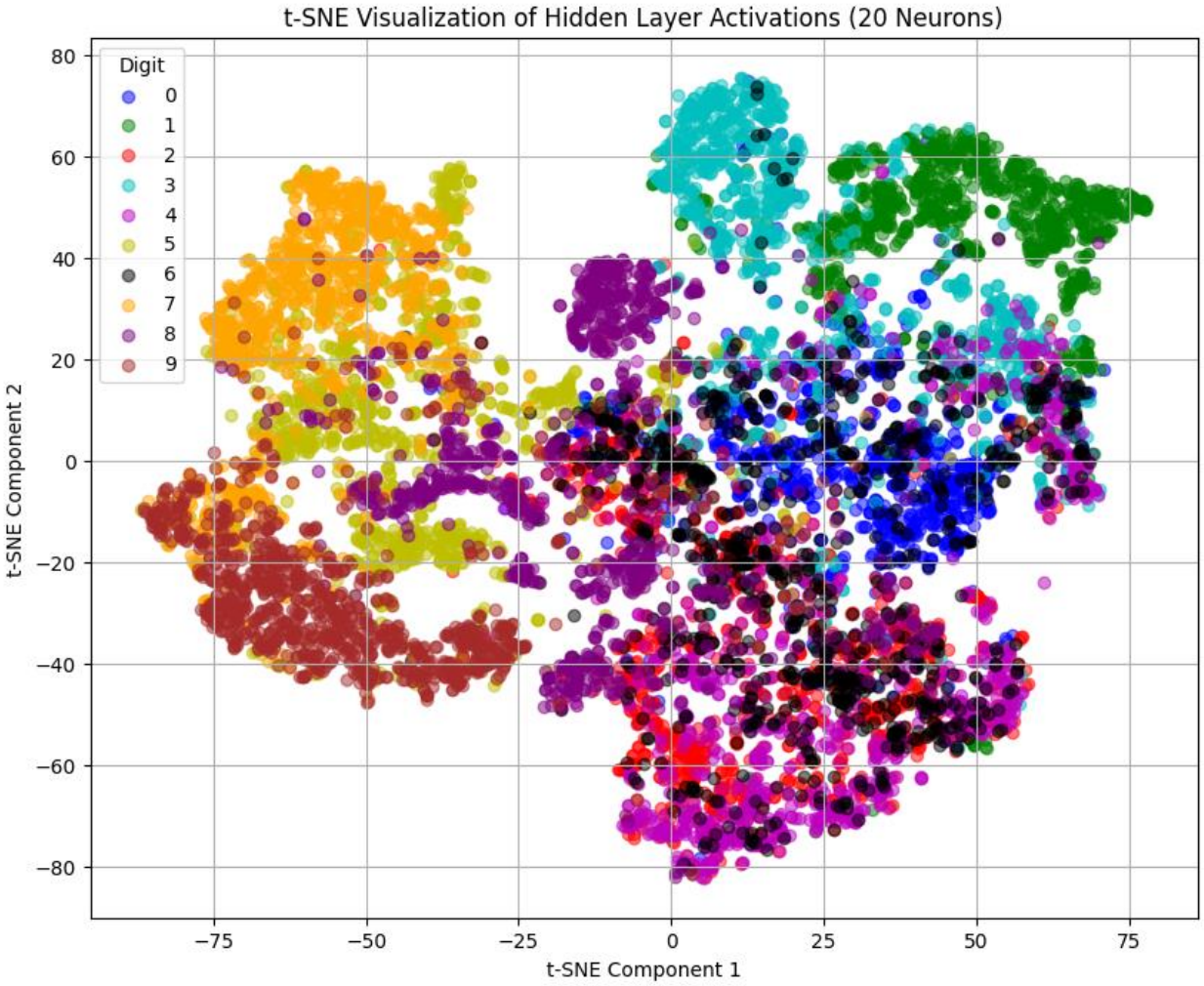
We can see that the untrained plot has no clustering since it is untrained and in the trained one clusterings have been formed.

Confusion matrix and f1 score for the Fashion-MNIST dataset

F1-score: 0.0399



t-SNE plot to visualize the output for the layer containing 20 neurons for the Fashion - MNIST dataset



Since the MLP trained on the MNIST dataset achieves a F1 score of 0.97, it indicates that the model has learned certain representations of the data that are optimized for the MNIST dataset. When we apply the same trained model to the Fashion-MNIST dataset, which contains different types of images, it's likely that the model's learned representations (embeddings) may not be as effective for this new dataset.

The different types of images in the Fashion-MNIST dataset may require different features to be accurately classified compared to the digits in the MNIST dataset. As a result, we observe that the F1 score for fashion- MNIST dataset is 0.0399 and rightly so the t-SNE plot for the same shows a few mischaracterization as the representations were trained for the MNIST dataset.