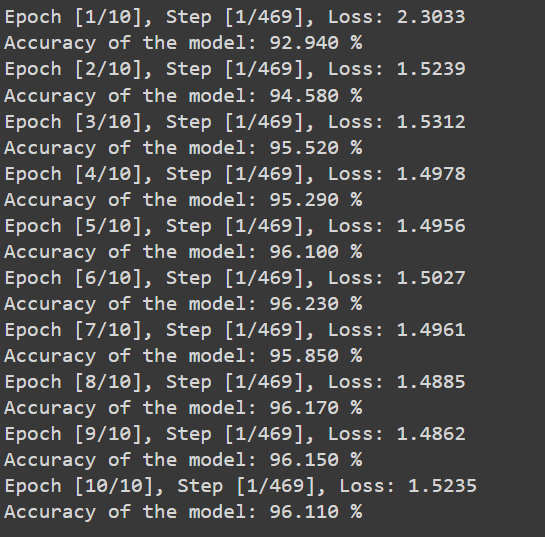
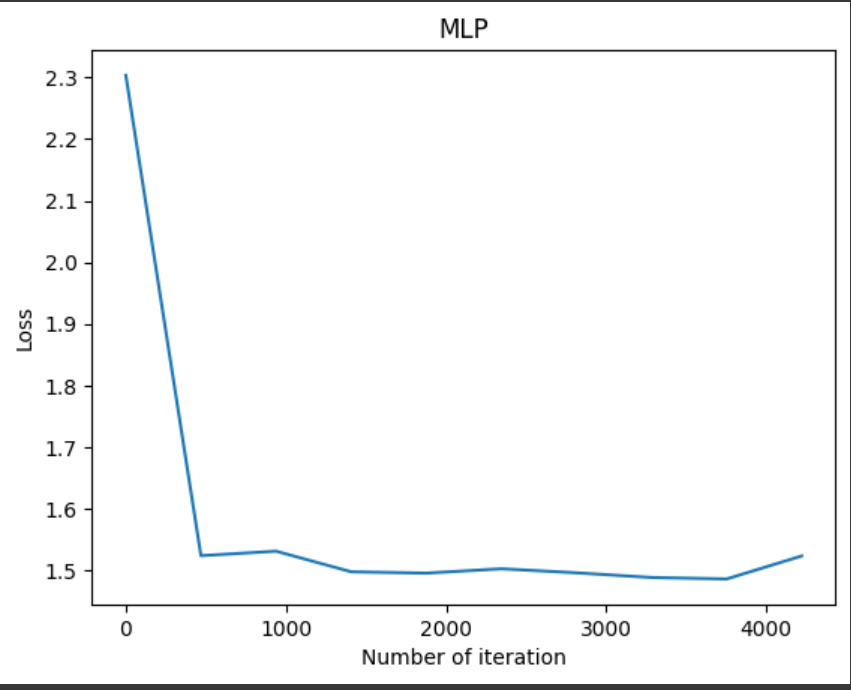
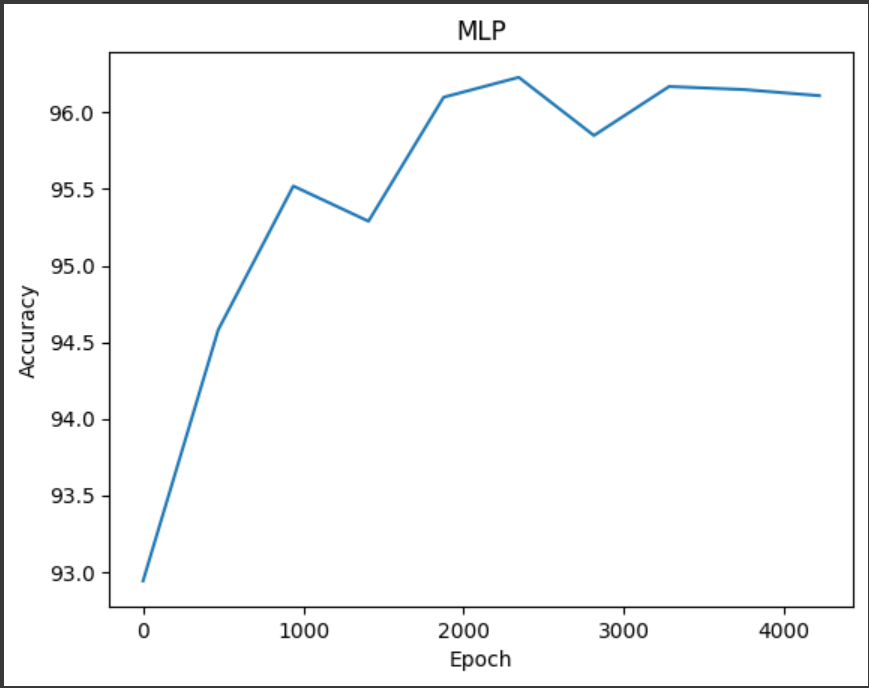
**MLP models on MNIST**

Imp observation- Use of CPU led to the running time of 10 min for 10 epochs but on the other hand using the GPU from the google colab helped to reduce the running time by 90% and decreased the time to just 1 min. Hence the GPU is used in the complete codes.

The results with hyper parameters as:

* Lr = 0.002
* Dropout rate for Regularization- 0.3
* Hidden neurons- 256, 64
* Batch size = 128



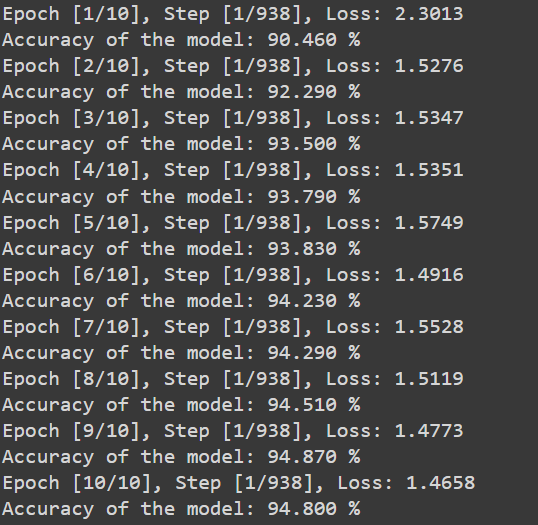
 

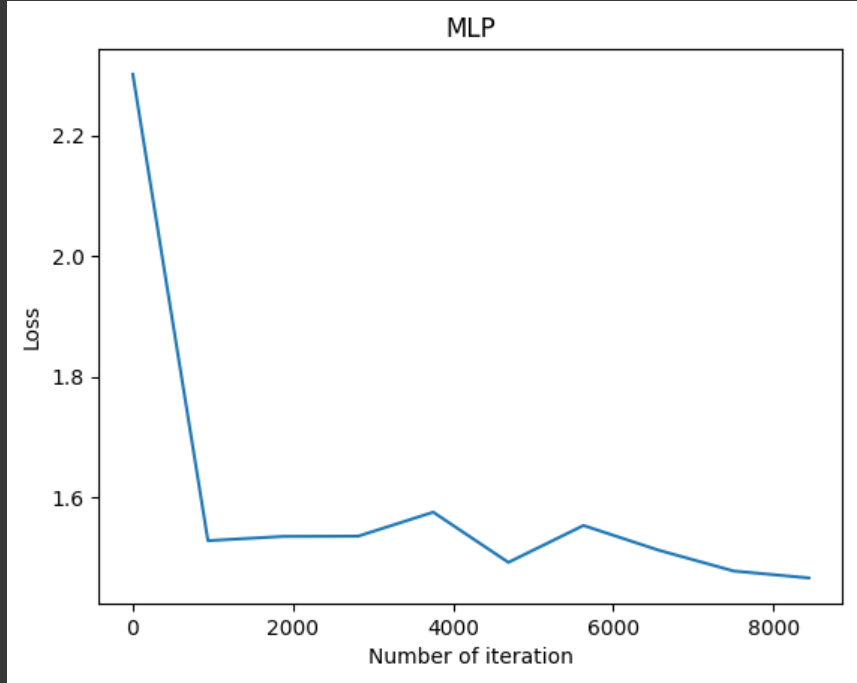
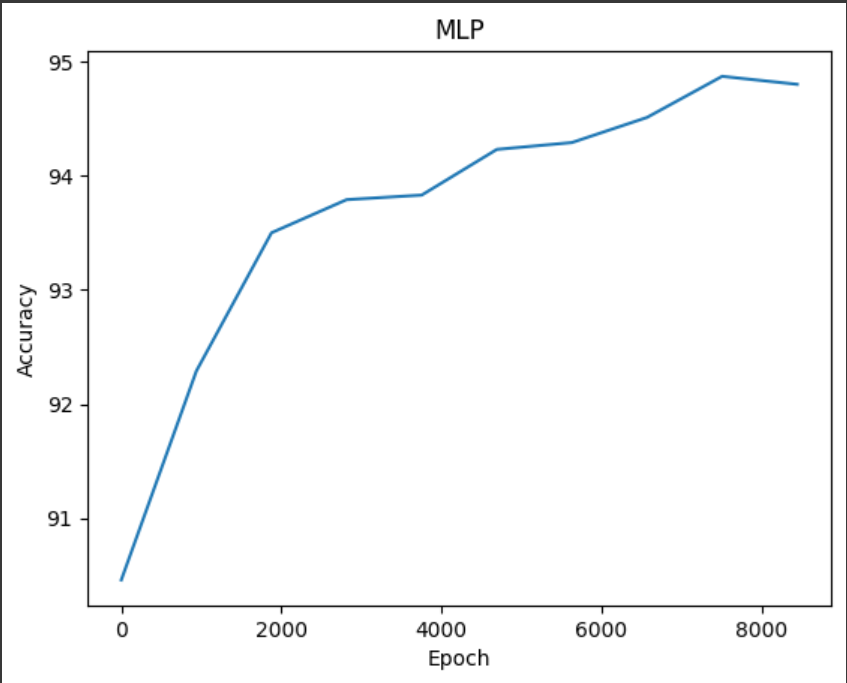
Conclusion- The output is quite good but some overfitting maybe is happening as the test accuracy is non monotonous.

**MLP models on MNIST**

The results with hyper parameters as:

* Lr = 0.001
* Dropout rate for Regularization- 0.5
* Hidden neurons- 128, 64
* Batch size = 64



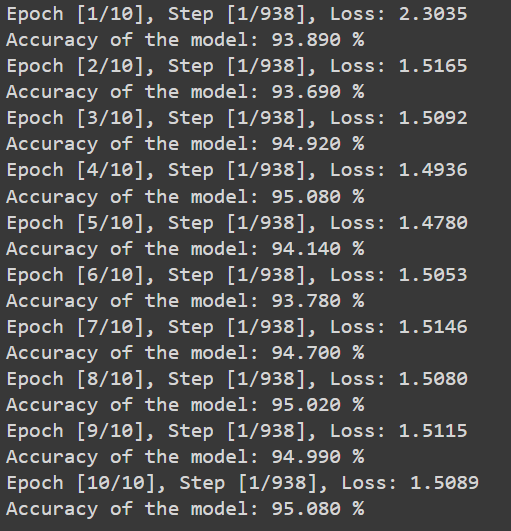
 

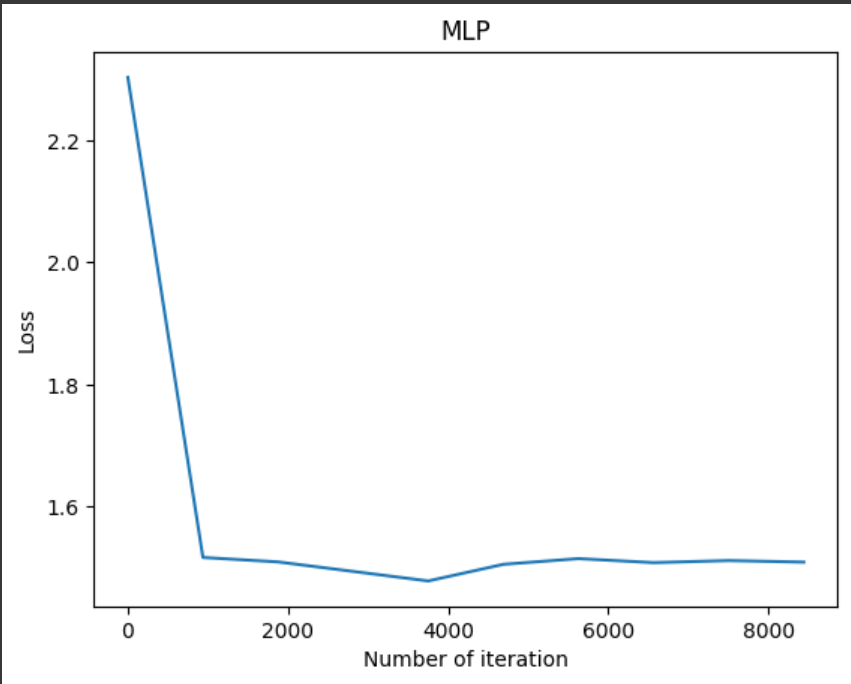
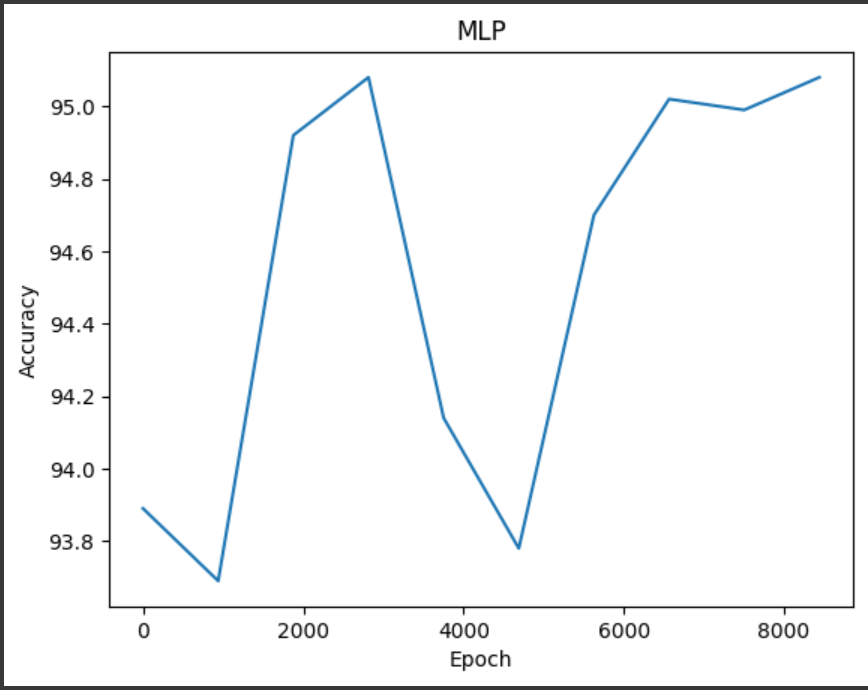
Conclusion- The model is performing well but the desired accuracy can’t be obtained here. So, we need to change the hyperparameters further for better accuracy/

**MLP models on MNIST**

The results with hyper parameters as:

* Lr = 0.002
* Dropout rate for Regularization- 0.3
* Hidden neurons- 512, 128
* Batch size = 64



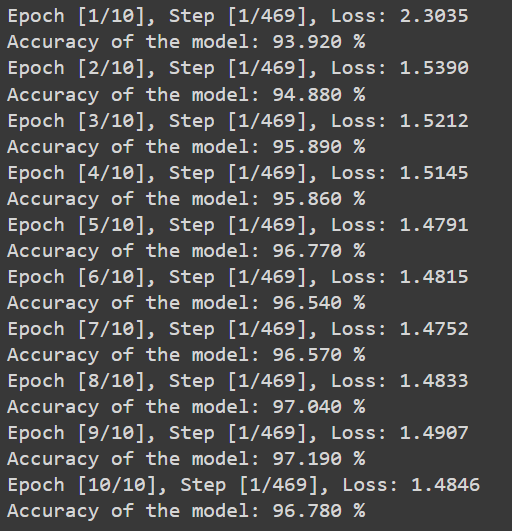
 

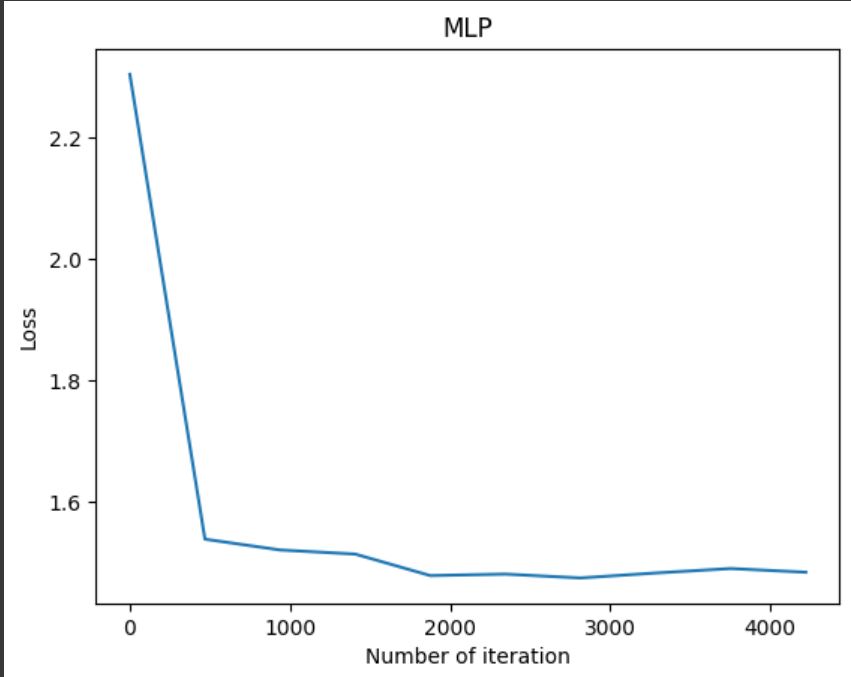
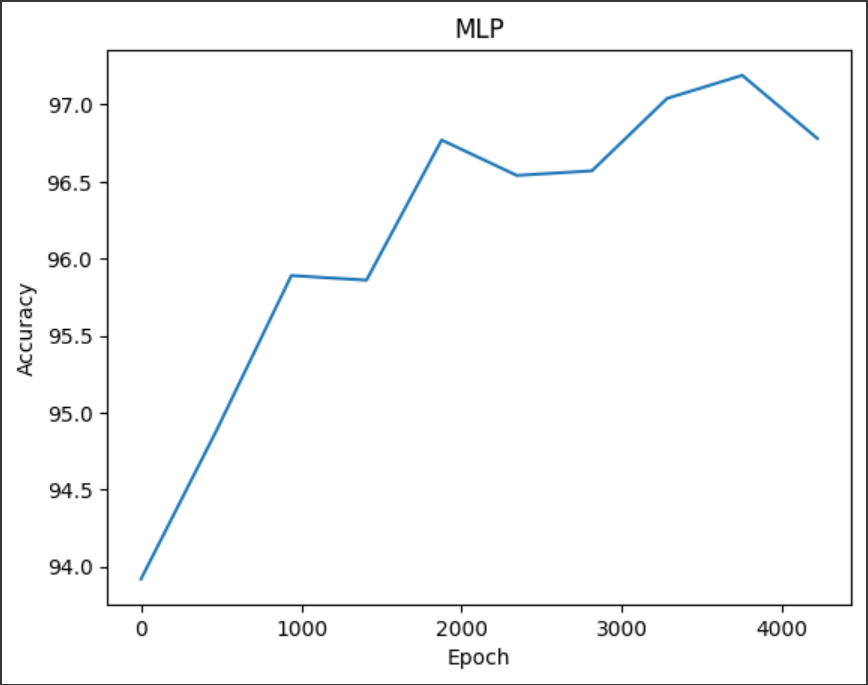
Conclusion- As the accuracy of the model is unstable and changing on every epoch. Hence the model seems to be overfitting.

**MLP models on MNIST**

The results with hyper parameters as:

* **Lr = 0.001**
* **Dropout rate for Regularization- 0.3**
* **Hidden neurons- 512, 128**
* **Batch size = 128**



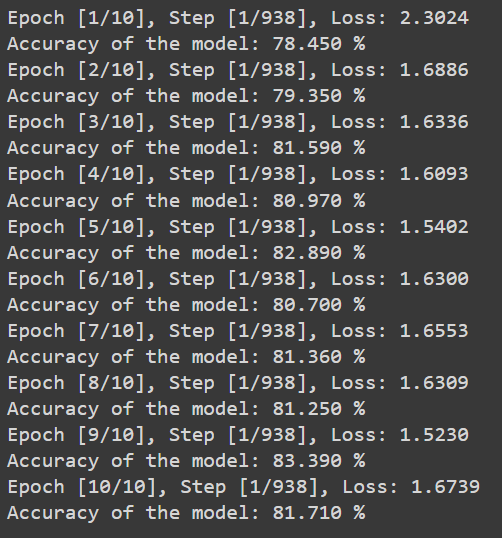
Conclusion- The output is quite good but some overfitting maybe is happening as the test accuracy is non monotonous. The achieved accuracy here is the best one among tested models.

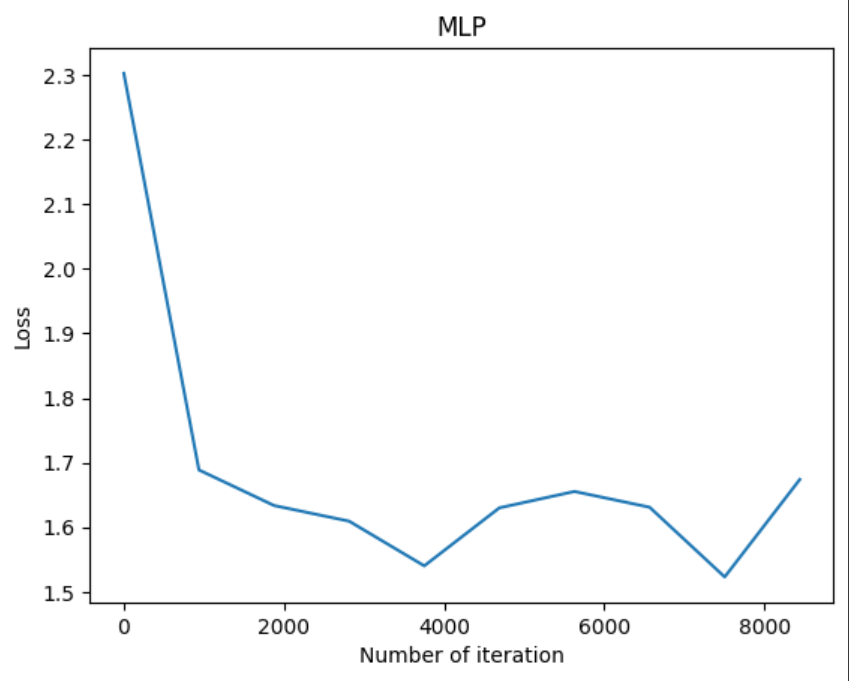
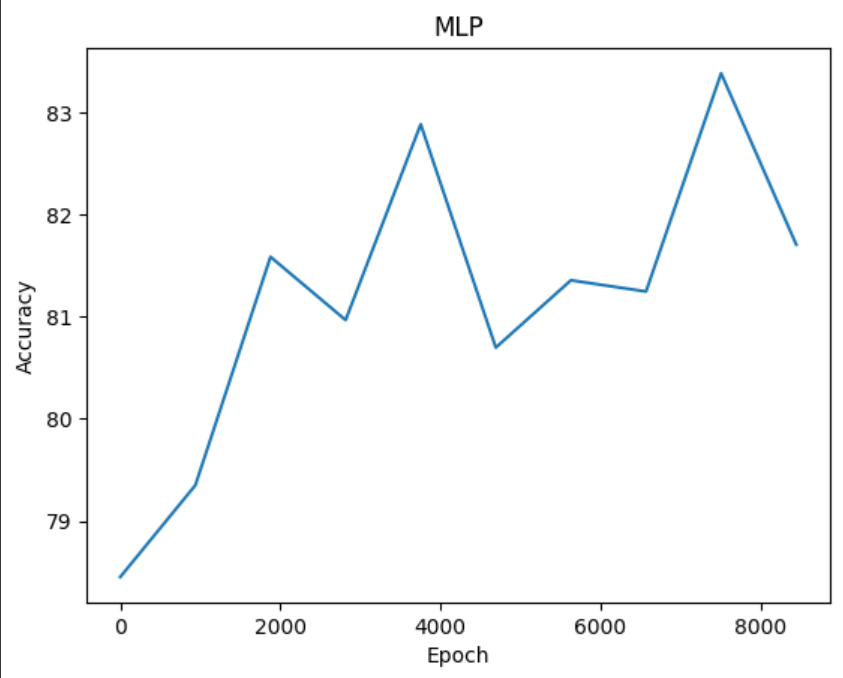
**The best Test Accuracy achieved= 97.04%**

**MLP models on Fashion\_MNIST**

The results with hyper parameters as:

* Lr = 0.001
* Dropout rate for Regularization- 0.3
* Hidden neurons- 512, 256
* Batch size = 64



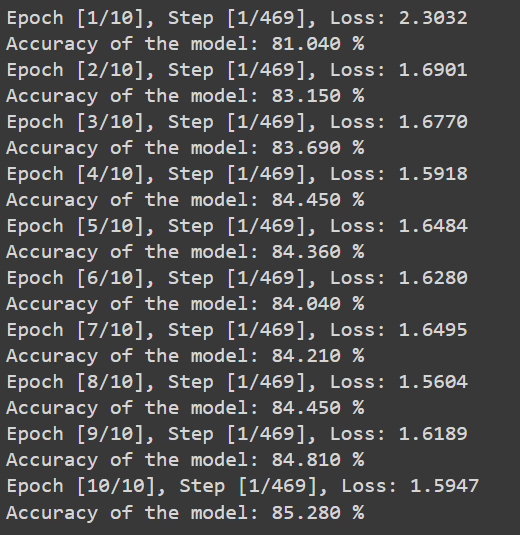
 

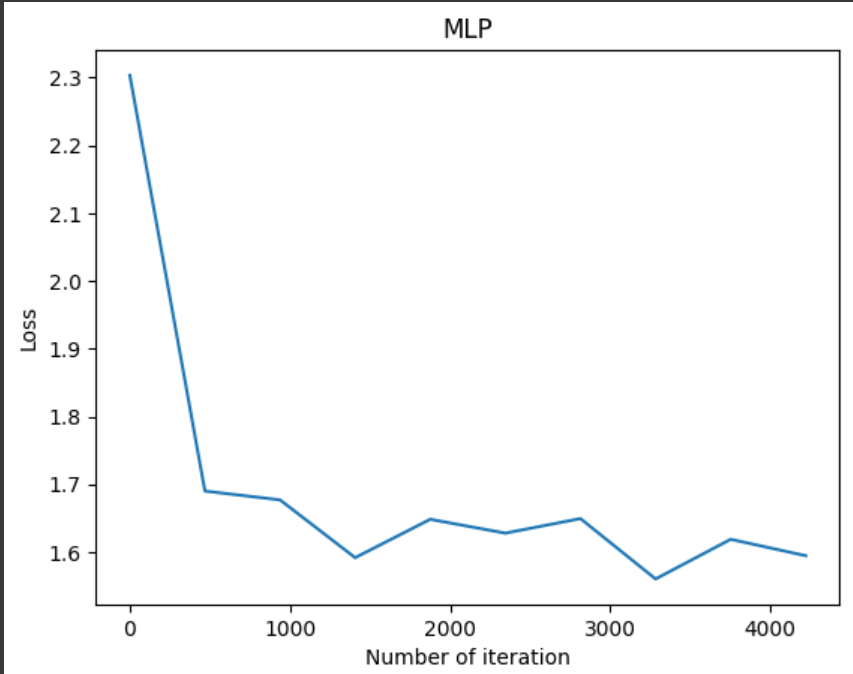
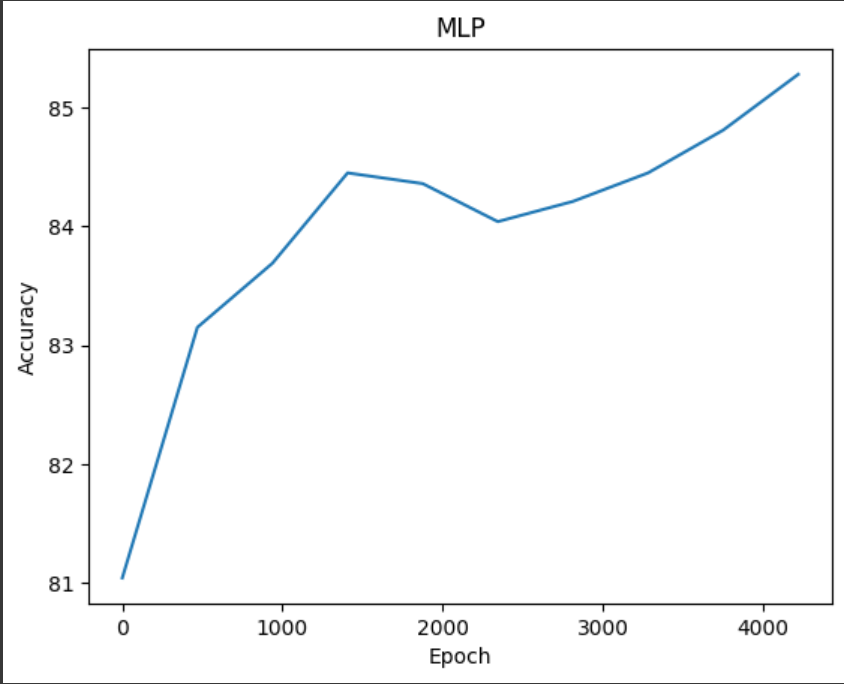
Conclusion- As the accuracy of the model is unstable and changing on every epoch. Hence the model seems to be overfitting. The accuracy is also quite less.

**MLP models on Fashion\_MNIST**

The results with hyper parameters as:

* **Lr = 0.001**
* **Dropout rate for Regularization- 0.3**
* **Hidden neurons- 512, 128**
* **Batch size = 128**



Conclusion- As the accuracy of the model is quite good and monotonous. So, it’s chosen as the best model.

These are the final results that I obtained on using the above hyper parameters.

**The best Test Accuracy achieved= 85.28%**

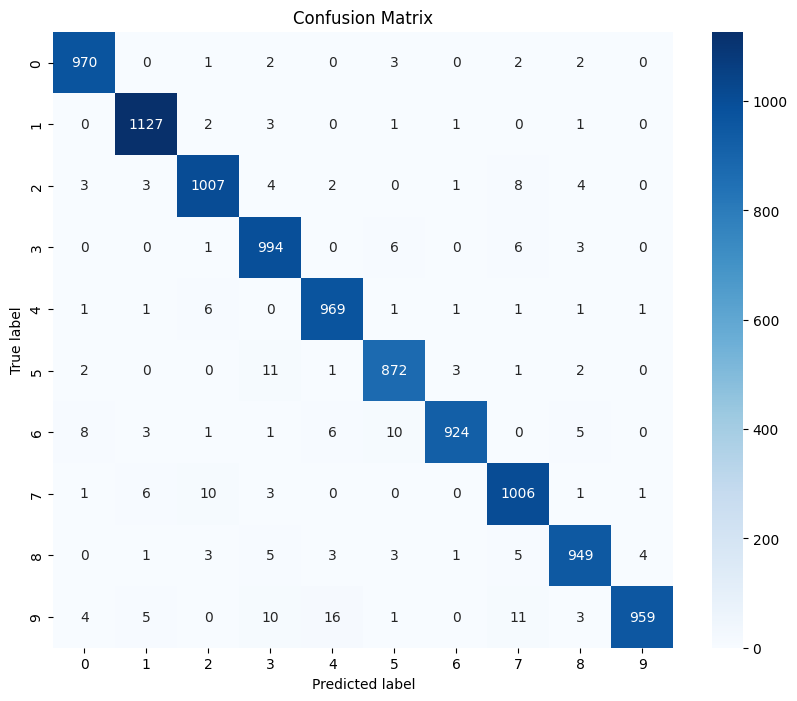
**Overall Conclusion on selection of Hyper parameters->**

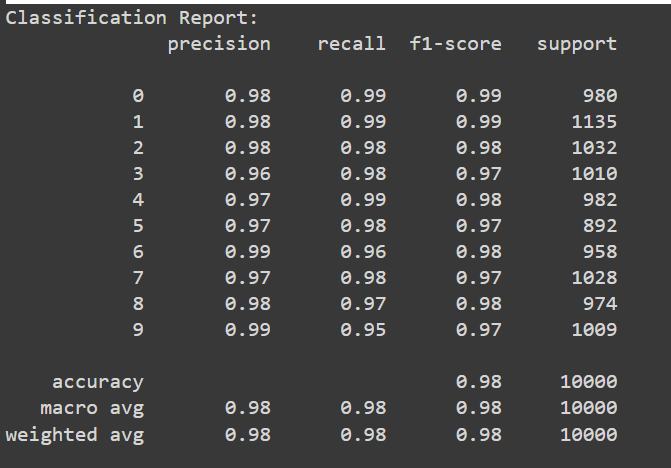
* In terms of Learning rate 0.001 seems to be the best suited one with some good speed and early convergence towards the most accurate result.
* In terms of Batch size 128 is selected to be the best as with lower sizes like 64 or 32 the training process gets somewhat slowed and whereas for higher sizes like 256 the learning process becomes too fast and it also hampers the overall accuracy of the result.
* The number of neurons in the hidden layers are chosen to be 512, 128 after multiple tests for different combinations among 64, 128, 256, 512.
* The Dropout factor for regularization is chosen to be 0.3 as it’s performing better then 0.5 in almost every model.
* The use of cuda or the GPU in the google colab helped a lot in speeding up the training time and decreasing it by 90% as compared to usual CPU.

**The major differences between the implementation on CNN and that on MLP are->**

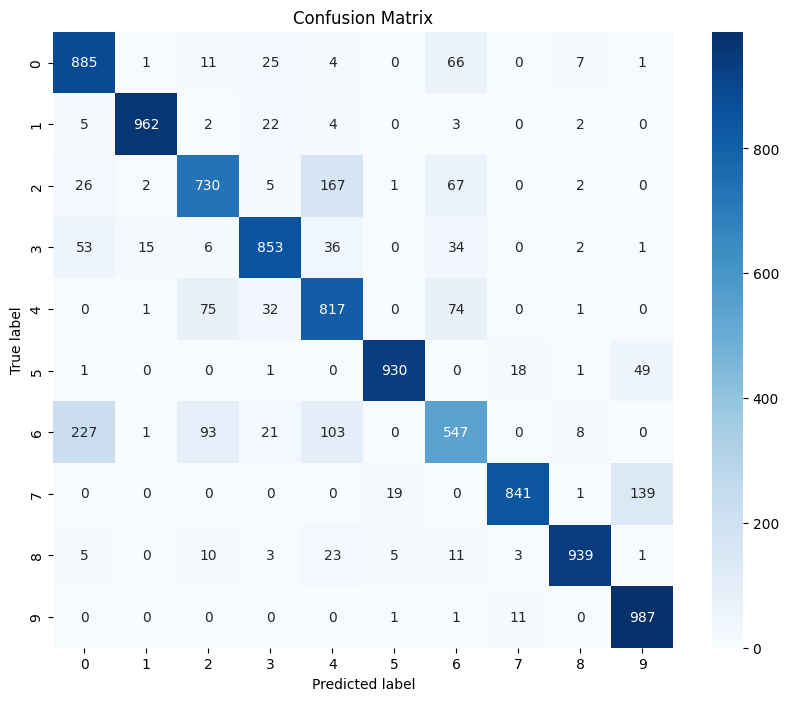
* The most important and visible difference can be seen in the accuracy of both models. The CNN model clearly given better accuracy than MLP model by 3-4% which is clearly depicted in the Overall accuracies obtained on training the models on CNN and MLP respectively.
* CNN uses less no of parameters as compared to that for MLP.
* CNN takes somewhat more training time than MLP when we run it on CPU. Whereas if we run it on GPU they more or less take equal time as here CNN is not so much dense and complicated.

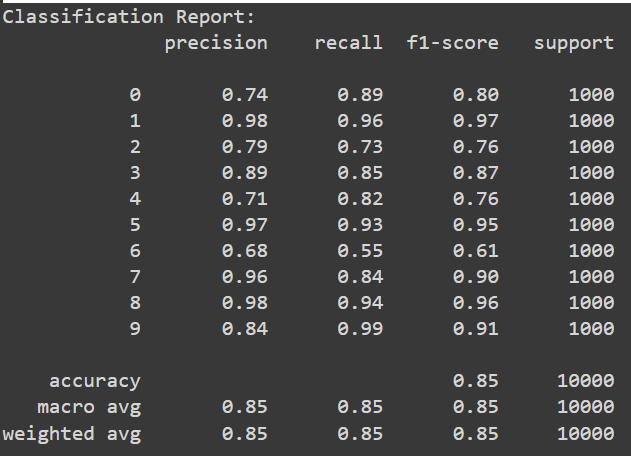
**The Confusion Matrix of MLP model on MNIST dataset**



****

**The Confusion Matrix of MLP model on Fashion\_MNIST dataset**



****