ML ASSIGNMENT

Task: The given dataset contains details about organic chemical compounds including their chemical features, isomeric conformation, names and the classes in which they are classified. The compounds are classified as either ‘Musk’ or ‘Non-Musk’ compounds. Your task is to build a classification model on the given data using any Deep Learning approach that you deem appropriate viz. Multi-Layer Perceptron, CNN, RNN, etc. or you could also use transfer learning approaches through selection of appropriate pre-trained model. The data has to be split in a 80:20 ratio for training and validation datasets. You can perform whatever preprocessing and post-processing operations on the data that may help you improve the performance of your model. You are required to report the performance measures of the model viz. Accuracy( Training and Validation) and Loss(Training and Validation) graphs, F1 score, precision, recall, etc. along with a well detailed report of what models, pre-processing, post-processing approaches you have used and why you chose to use these approaches.

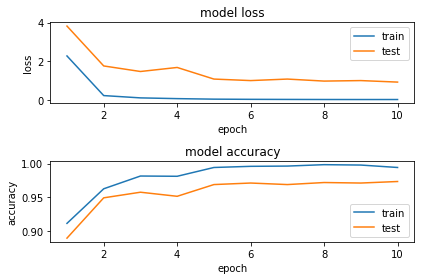
Output:

Note: Based on the language of the task assigned, I have assumed testing set to be synomous with validation set, though in reality they are different.

1. A detailed report of the model

I used a 3 hidden-layer Multi-Layer Perceptron with 500 neurons in each layer. I used this model because the I felt that it would be able to track the various orientations of the musk and non-musk compounds [ as the model is based on the human brain which can pick on orientations with ease].

1. The pickle model that I have trained is attached with code to use it at the end of ml\_musk.py.
2. The Loss and Accuracy graphs in proper format are as follows:



1. Final performance measures of your model including validation accuracy, loss, precision, recall, F1 score.

Precision: 0.99

Recall: 0.98

F1-score: 0.98

Accuracy:0.97

Balanced Accuracy:0.97

A more detailed report is as follows:

