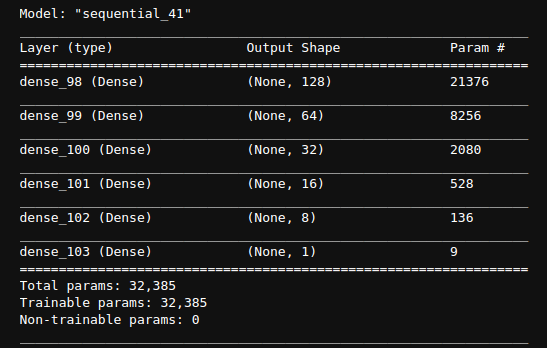
**REPORT**

Preprocessing:

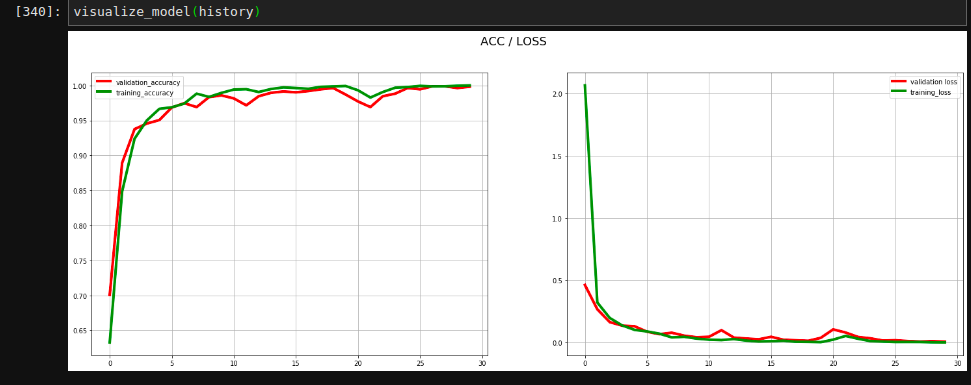
1. Given Data set is a binary classification dataset. i.e. Musk vs. Non-Musk.
2. There are no null values in any column.
3. The Dataset is imbalanced so I have performed upsampling of minority class.
4. Features like ID, molecule\_name, conformation\_name have no effect on output class. So we can ignore them.
5. Features f1-f166 shows some bond length of compound, so they are important.
6. Separated features and labels and stored it in diff. variables.
7. Implemented shuffling because classes are in particular order.
8. Splitted my dataset into two parts i.e. test and validation having test size=80%.

**Model:**

1. I have used a simple Deep Neural Network for Prediction having architecture:



1. Various Performance measures:
   1. **Accuracy: 0.9981**
   2. **Loss: 0.0071**
   3. **Recall: 0.9972**
   4. **F1: 0.9966**
   5. **Precision: 0.9963**
2. Optimizer used is Adam along with binary\_crossentropy for loss.
3. The graphical Representation are:



1. The epochs I have used are 30 with batch size of 128.
2. The other model I have used is CNN which is getting good accuracy of about 0.99 but precision and f1 scores are not good. They are in range of 0.4 - 0.5.So I have used simple DNN for prediction.