

PROJECT REPORT

Done by: Amit Kumar

OBJECTIVE: To build a classification model on the given data using any Deep Learning approach.

APPROACH:

1. Importing important modules and reading the given Datasets.
2. Visualizing the loaded datasets. We find that there are a total of 6598 unique datasets, each having 166 features excluding names and ids.

```
In [3]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6598 entries, 0 to 6597
Columns: 170 entries, ID to class
dtypes: int64(168), object(2)
memory usage: 8.6+ MB
```

3. We checked if there are null values or blank values in the columns. We find that there are no any null values.

```
In [4]: df.isna().any().value_counts()

Out[4]: False    170
dtype: int64
```

4. Then, we checked if our data is imbalanced or not. Our data was found to be slightly imbalanced but can be ignored.

```
In [14]: Y.value_counts()

Out[14]: 0    5581
         1    1017
         Name: class, dtype: int64
```

5. We then drop the non-features columns in our dataframe.
6. No of features for each dataset is 166
7. Since our data class was not distributed, we shuffled the data frame so that 0's and 1's are distributed.
8. Splitting dataframe into Input X and Output Y.

```
In [17]: X.shape, Y.shape

Out[17]: ((6598, 167), (6598,))
```

9. We then Scale the input between [0,1] for every column using sklearn min-max_scaler.
10. Our Data is fully pre-processed now and we are ready to build model.
11. Importing the import modules and libraries for Modelling the Artificial Neural Network.
12. Parameters defining for saving the modle_weights into h5.
13. Defining the model as given.

PROJECT REPORT

14. Using Dropout to perform regularization.

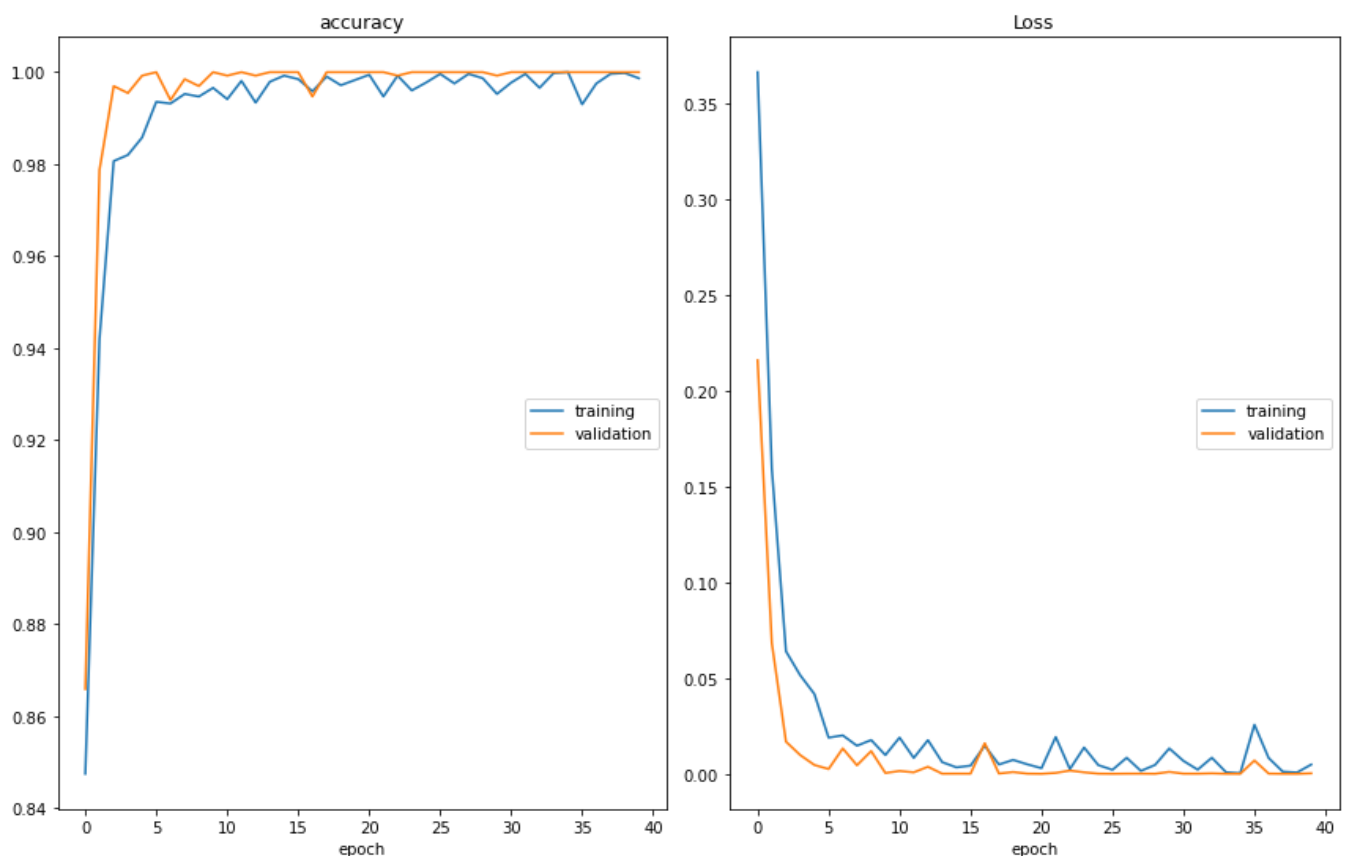
15. Summarizing the Model.

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 167)	28056
dropout (Dropout)	(None, 167)	0
dense_1 (Dense)	(None, 83)	13944
dropout_1 (Dropout)	(None, 83)	0
dense_2 (Dense)	(None, 41)	3444
dropout_2 (Dropout)	(None, 41)	0
dense_3 (Dense)	(None, 20)	840
dropout_3 (Dropout)	(None, 20)	0
dense_4 (Dense)	(None, 1)	21
Total params: 46,305		
Trainable params: 46,305		
Non-trainable params: 0		

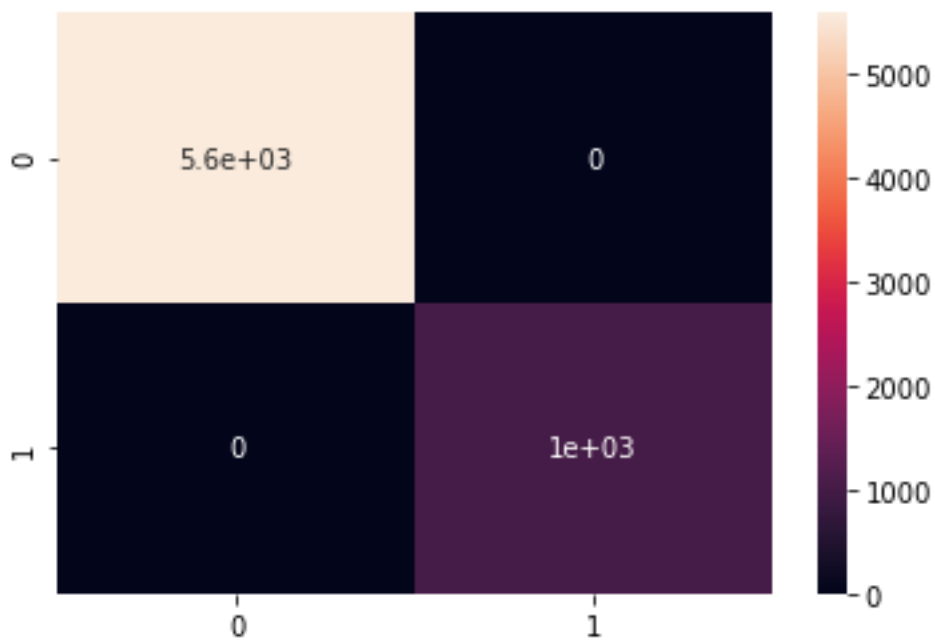
16. Fitting the model with the input and output and defining validation size to be 20%.

17. We get the following Loss and Accuracy Plot.



PROJECT REPORT

18. Confusion Matrix and Classification report is given below.



```
In [29]: print(classification_report(Y,Y_pred))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	5581
1	1.00	1.00	1.00	1017
accuracy			1.00	6598
macro avg	1.00	1.00	1.00	6598
weighted avg	1.00	1.00	1.00	6598

INFERENCE: We get an accuracy of 100% and f1-score of 100%. Our model is booming.

Name: Amit Kumar

Email : amitk29amitk@gmail.com / amitbtee@iitk.ac.in

Mob: 9113348549