NATIONAL UNIVERSITY OF SCIENCES & TECHNOLOGY
PAKISTAN

Name: **Muhammad Nabeel**

Assignment Type: **Internship Task**

Supervisor/Instructor: **Dr. Fehmida Usmani**

CMS: **413522**

Task: **Train a GAN to accurately mimic Anomalies and then train a classifier to distinguish between those anomalies and real Values.**

Data Description:

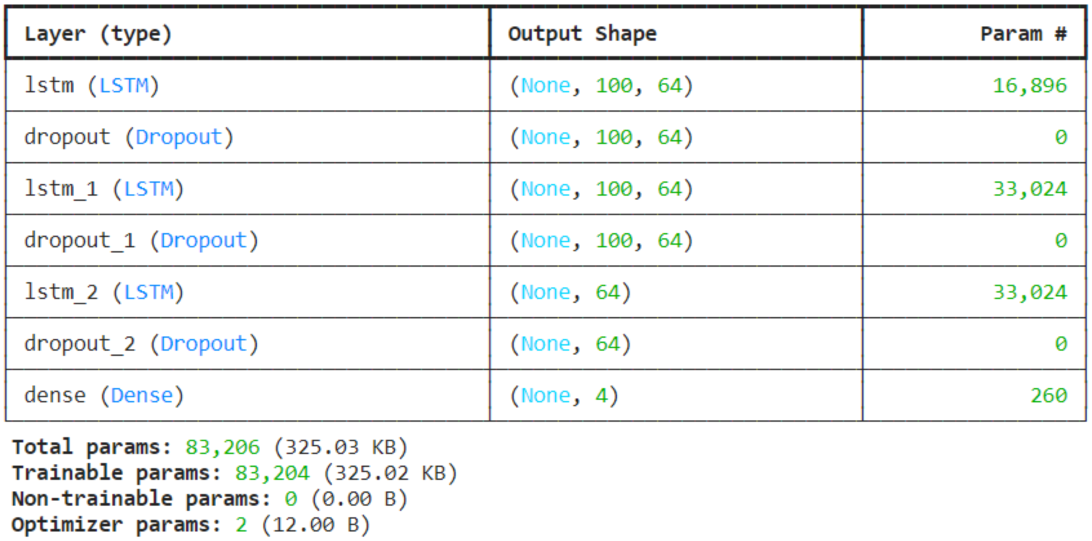
The dataset contains 204686 entries of data containing Univariate Time series with single column feature (SOPAS) and one output label (of Class 0,1,2 and 3).

Data Preprocessing and Outlier Detection:

The data was standardized using built in Scikit learn library. Furthermore, to feed data to a RNN, it has to be converted to a special Data-structure which is for nth row/entries, it should has k entries previous to n as input. The size of **K (Window size)** was chosen to be **100** and produced optimum Results. The data was also checked if any null or empty cells are present.

Model Choice and Architecture:

A simple RNN consisting of LSTM Neural Layers was sufficient for the following task. The Architecture of the Network is as follows:

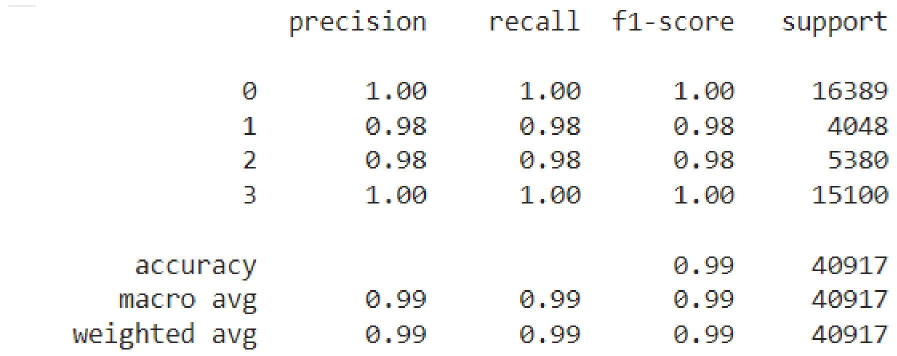


Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 100, 64)	16,896
dropout (Dropout)	(None, 100, 64)	0
lstm_1 (LSTM)	(None, 100, 64)	33,024
dropout_1 (Dropout)	(None, 100, 64)	0
lstm_2 (LSTM)	(None, 64)	33,024
dropout_2 (Dropout)	(None, 64)	0
dense (Dense)	(None, 4)	260

```
Total params: 83,206 (325.03 KB)
Trainable params: 83,204 (325.02 KB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 2 (12.00 B)
```

Precision and Accuracy Score:

The model was run for 50 epochs and Early Stopping was also used to prevent Overfitting. The Confusion Matrix and Metric Results are Attached in the zip and their respective Screen Shots are below:



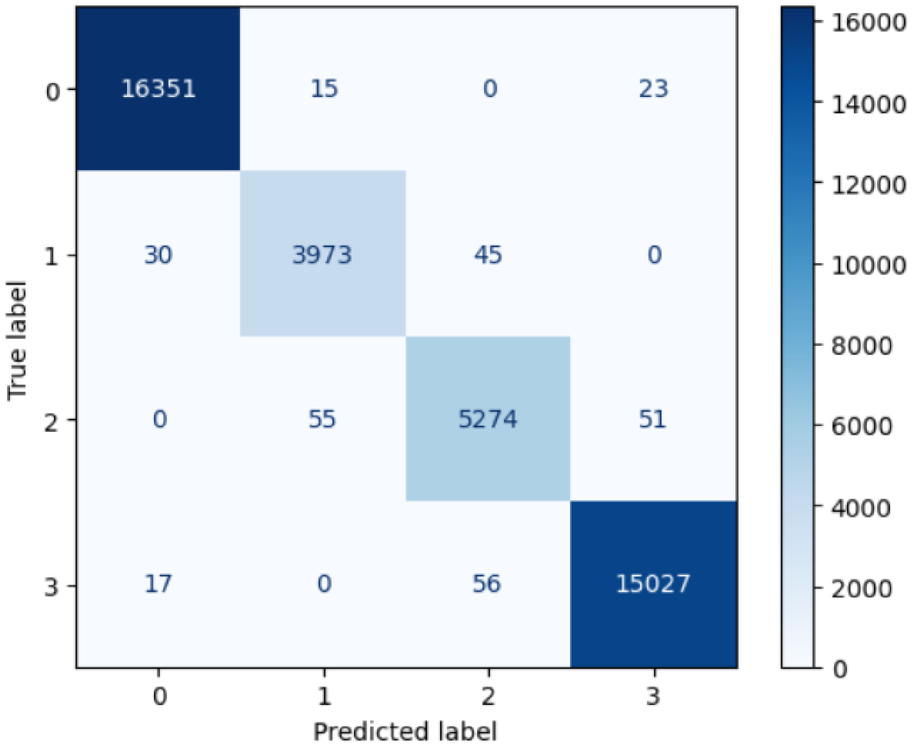
```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     16389
           1       0.98      0.98      0.98      4048
           2       0.98      0.98      0.98      5380
           3       1.00      1.00      1.00     15100

    accuracy                           0.99     40917
   macro avg       0.99      0.99      0.99     40917
weighted avg       0.99      0.99      0.99     40917
```

- From the above classification Report, we can see that the Precision and Recall of **Class 1** is **98** percent.

Confusion Matrix:



- <u>**Further the Trained model and Jupyter Notebook for code base is int the zip.**</u>