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:

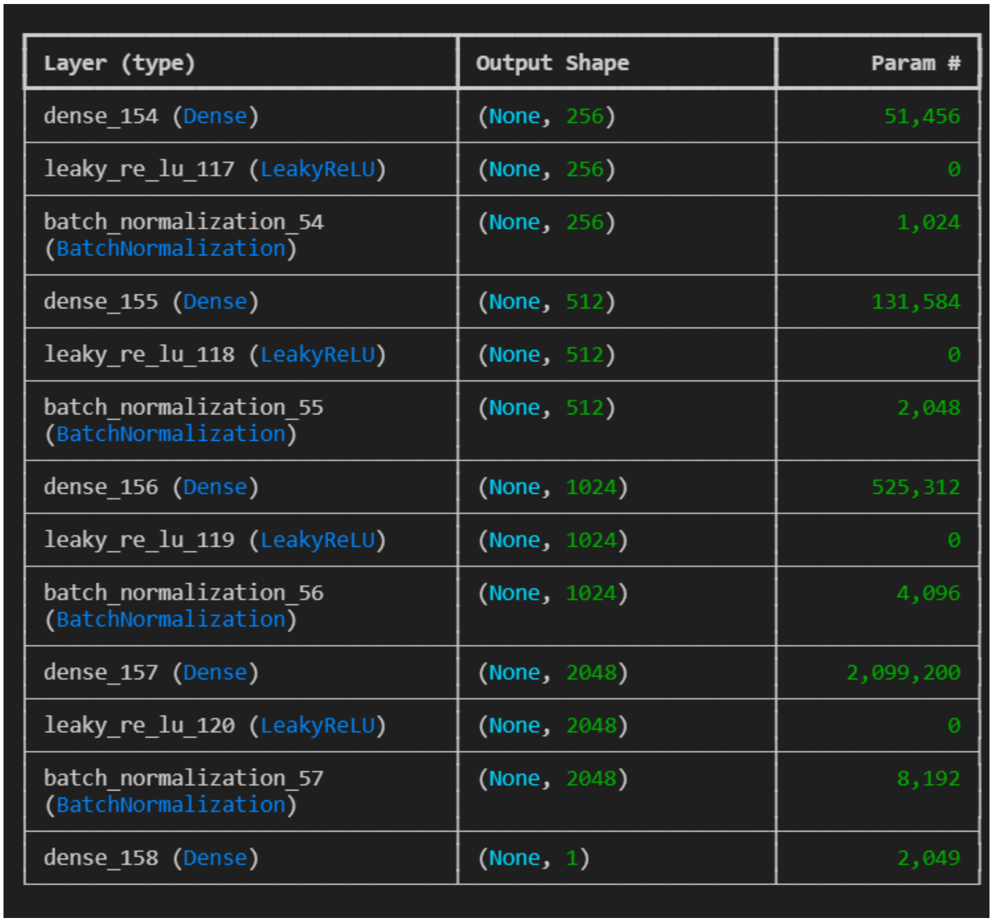
The dataset contains 20,468 entries of data containing single column feature (SOPAS) and one output labels (only of Class 1).

Data Preprocessing and Outlier Detection:

Since the GAN had to be trained for Anomaly detection, an Algorithm for Outlier Detection had to be trained. In this regard **Isolation Forest** was used to detect Outliers. The outliers were then Scaled between -1 and 1 and feed to the GAN.

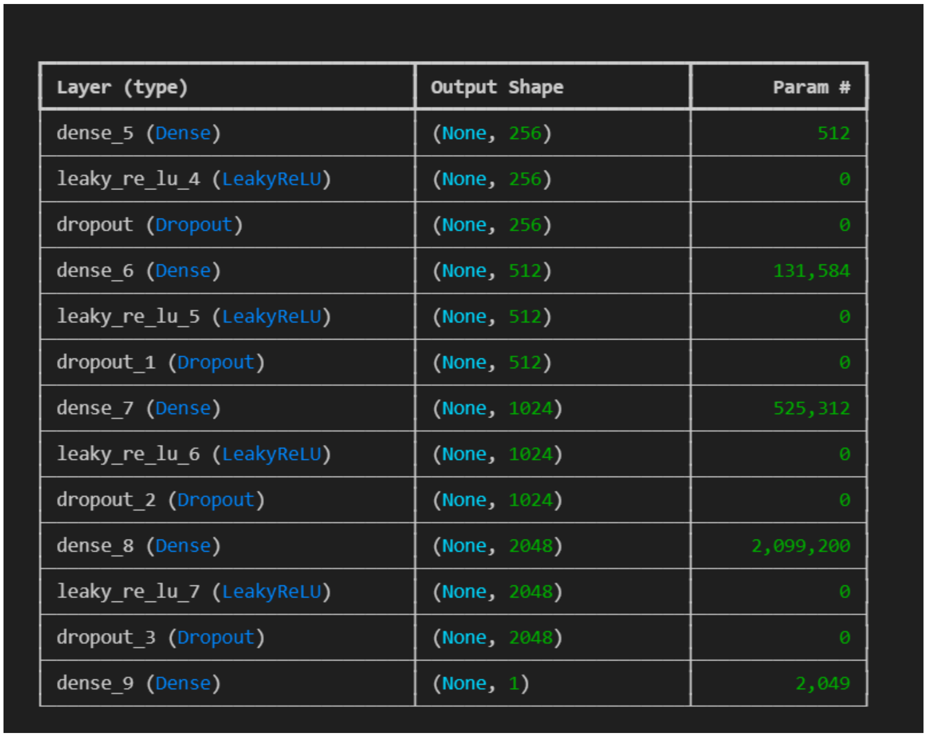
GAN Training:

Generator Architecture Used:



Layer (type)	Output Shape	Param #
dense_154 (Dense)	(None, 256)	51,456
leaky_re_lu_117 (LeakyReLU)	(None, 256)	0
batch_normalization_54 (BatchNormalization)	(None, 256)	1,024
dense_155 (Dense)	(None, 512)	131,584
leaky_re_lu_118 (LeakyReLU)	(None, 512)	0
batch_normalization_55 (BatchNormalization)	(None, 512)	2,048
dense_156 (Dense)	(None, 1024)	525,312
leaky_re_lu_119 (LeakyReLU)	(None, 1024)	0
batch_normalization_56 (BatchNormalization)	(None, 1024)	4,096
dense_157 (Dense)	(None, 2048)	2,099,200
leaky_re_lu_120 (LeakyReLU)	(None, 2048)	0
batch_normalization_57 (BatchNormalization)	(None, 2048)	8,192
dense_158 (Dense)	(None, 1)	2,049

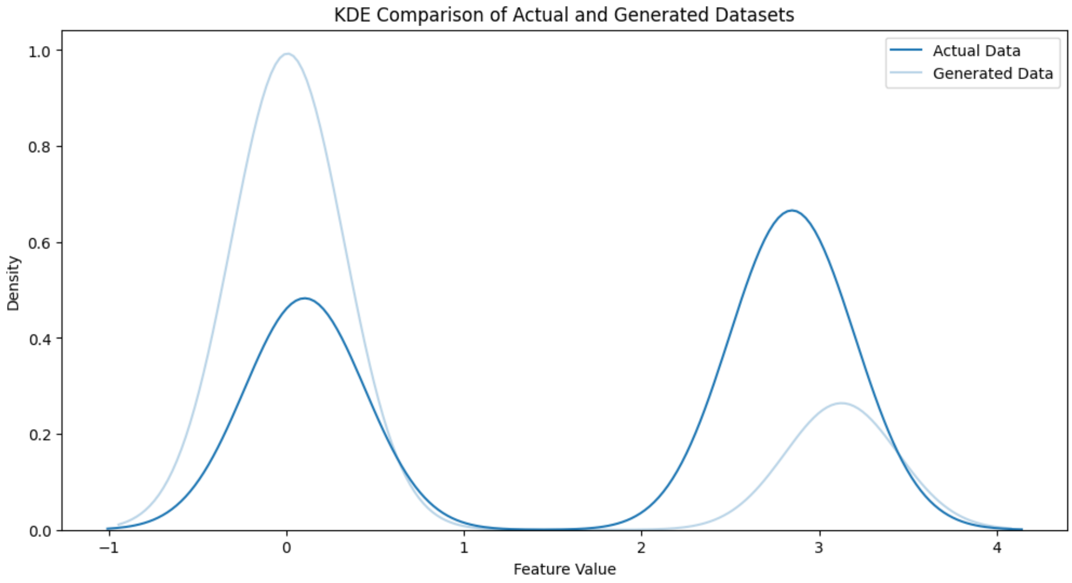
Discriminator Architecture:

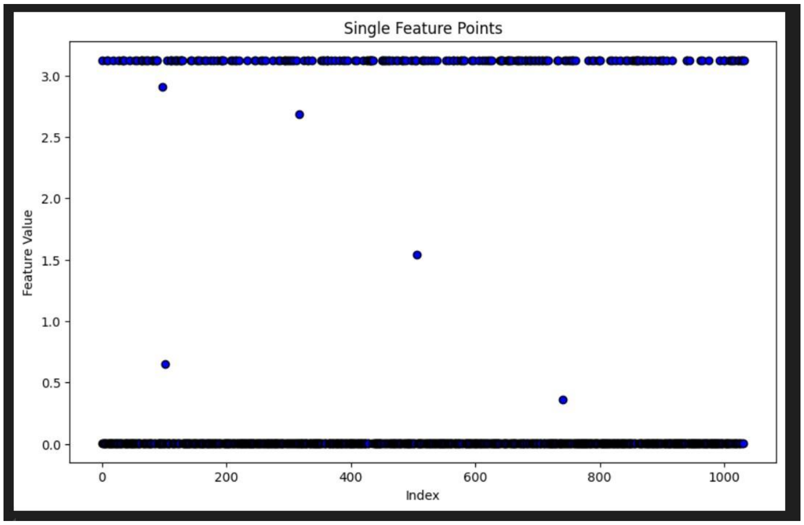


```
| Layer (type)                   | Output Shape           |       Param # |
|--------------------------------|------------------------|---------------|
| dense_5 (Dense)                | (None, 256)            |           512 |
| leaky_re_lu_4 (LeakyReLU)      | (None, 256)            |             0 |
| dropout (Dropout)              | (None, 256)            |             0 |
| dense_6 (Dense)                | (None, 512)            |       131,584 |
| leaky_re_lu_5 (LeakyReLU)      | (None, 512)            |             0 |
| dropout_1 (Dropout)            | (None, 512)            |             0 |
| dense_7 (Dense)                | (None, 1024)           |       525,312 |
| leaky_re_lu_6 (LeakyReLU)      | (None, 1024)           |             0 |
| dropout_2 (Dropout)            | (None, 1024)           |             0 |
| dense_8 (Dense)                | (None, 2048)           |     2,099,200 |
| leaky_re_lu_7 (LeakyReLU)      | (None, 2048)           |             0 |
| dropout_3 (Dropout)            | (None, 2048)           |             0 |
| dense_9 (Dense)                | (None, 1)              |         2,049 |
```

- The GAN was trained on 25000-30000 iterations.

<u>Noise Generated by GAN:</u>

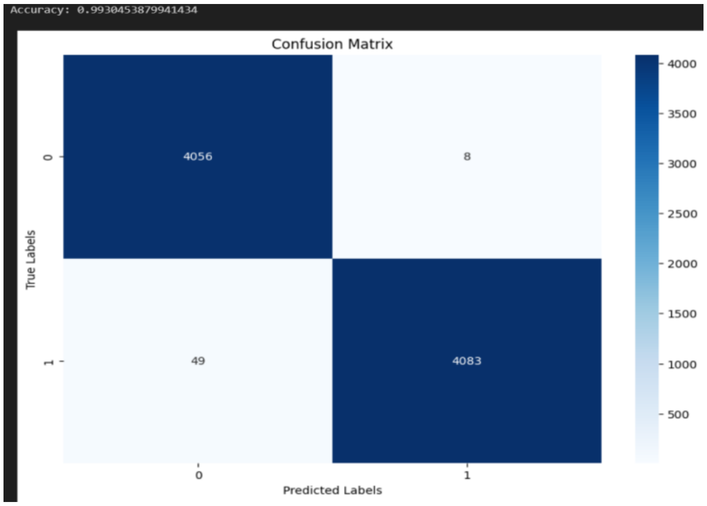


- It can be seen from KDE comparison , that the Actual and Generated data Approximately mimic each other.
- In the Jupyter-Notebook it is shown that the Standard-Deviation of the actual noise-data is about 1.35321 while the standard deviation of GAN generated data is 1.30797 which shows it is performing quite well.
- 

- The black strips at the top and bottom show the generated noise while only some noise is in the between.

Final Classifier for Distinguishing between Original Class 1 samples

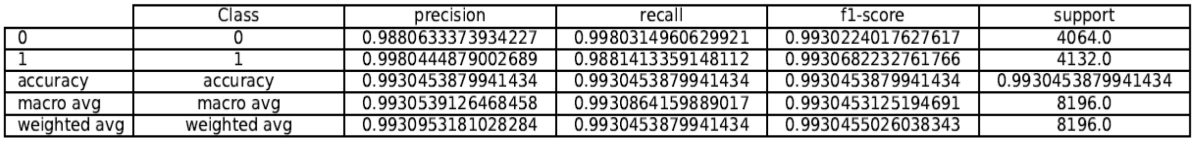
Confusion Matrix:



```
Accuracy: 0.9930453879941434
```

Classification Report:

A separate PDF has been attached containing the classification Report but a screenshot has been pasted below:



	Class	precision	recall	f1-score	support
0	0	0.9880633373934227	0.9980314960629921	0.9930224017627617	4064.0
1	1	0.9980444879002689	0.9881413359148112	0.9930682232761766	4132.0
accuracy	accuracy	0.9930453879941434	0.9930453879941434	0.9930453879941434	0.9930453879941434
macro avg	macro avg	0.9930539126468458	0.9930864159889017	0.9930453125194691	8196.0
weighted avg	weighted avg	0.9930953181028284	0.9930453879941434	0.9930455026038343	8196.0

- An intermediate dataset for consisting of noise generated by GAN and entries of Class one that was used for Final Classifier training has also been attached.
- Both Generator Model and Classifier are attached.