

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).

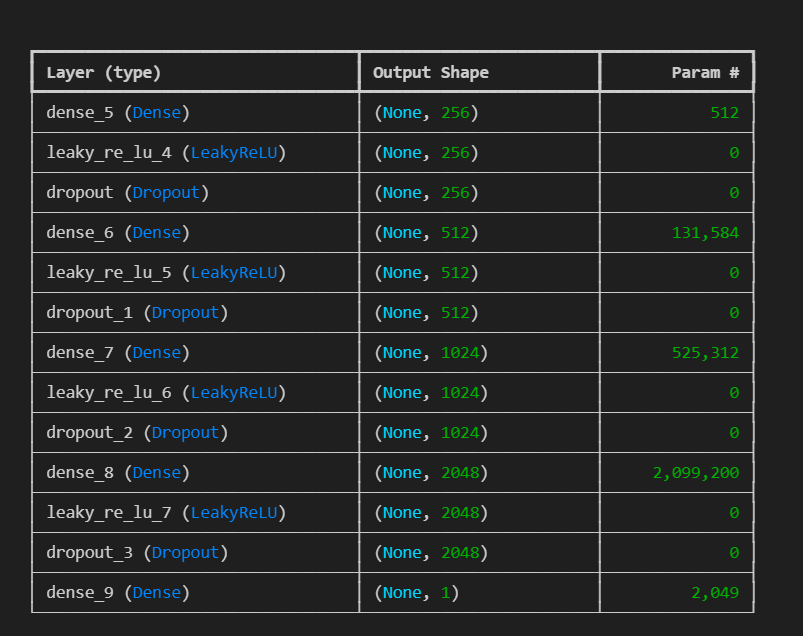
**Data Preprocessing and Outlier Detection:**

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:**

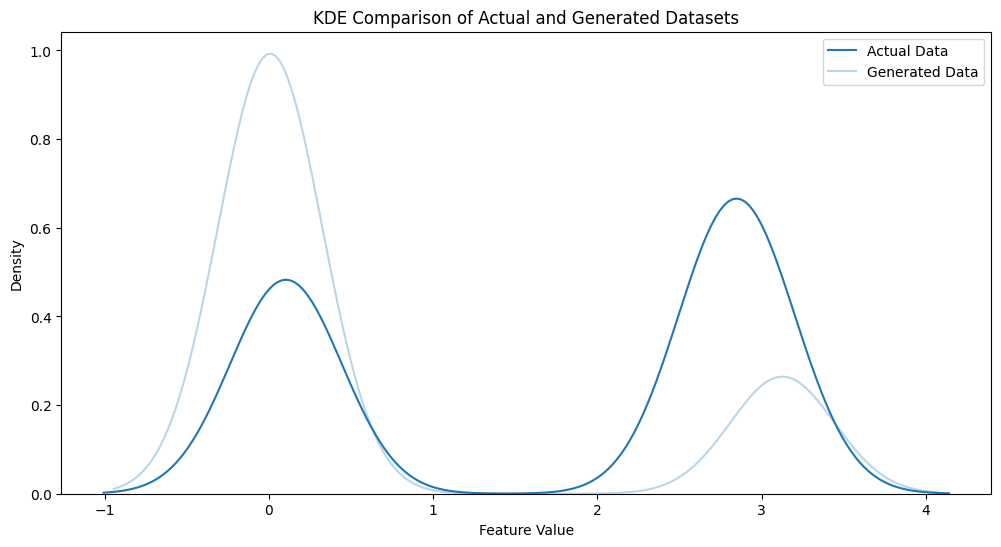
**Generator Architecture Used:**

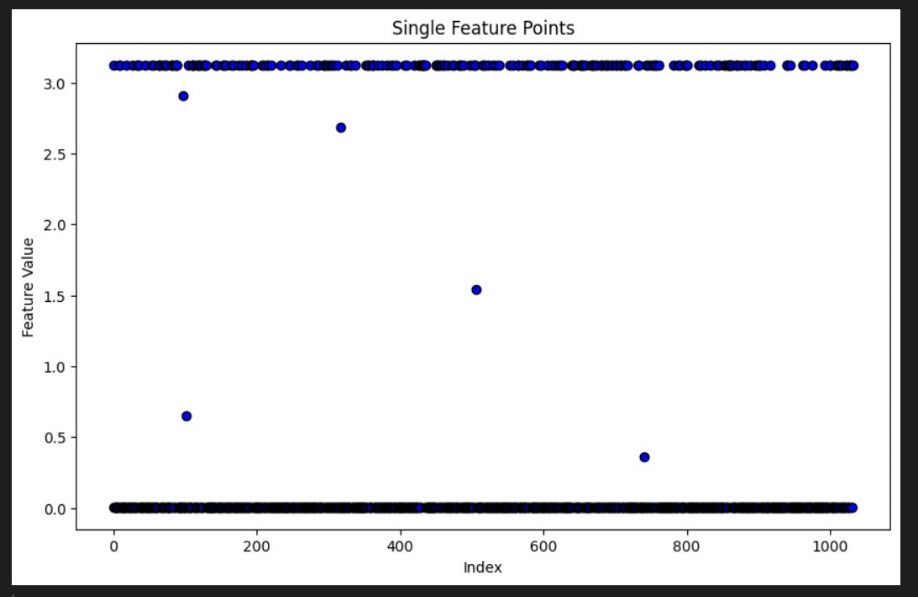
**Discriminator Architecture:**



* The GAN was trained on 25000-30000 iterations.

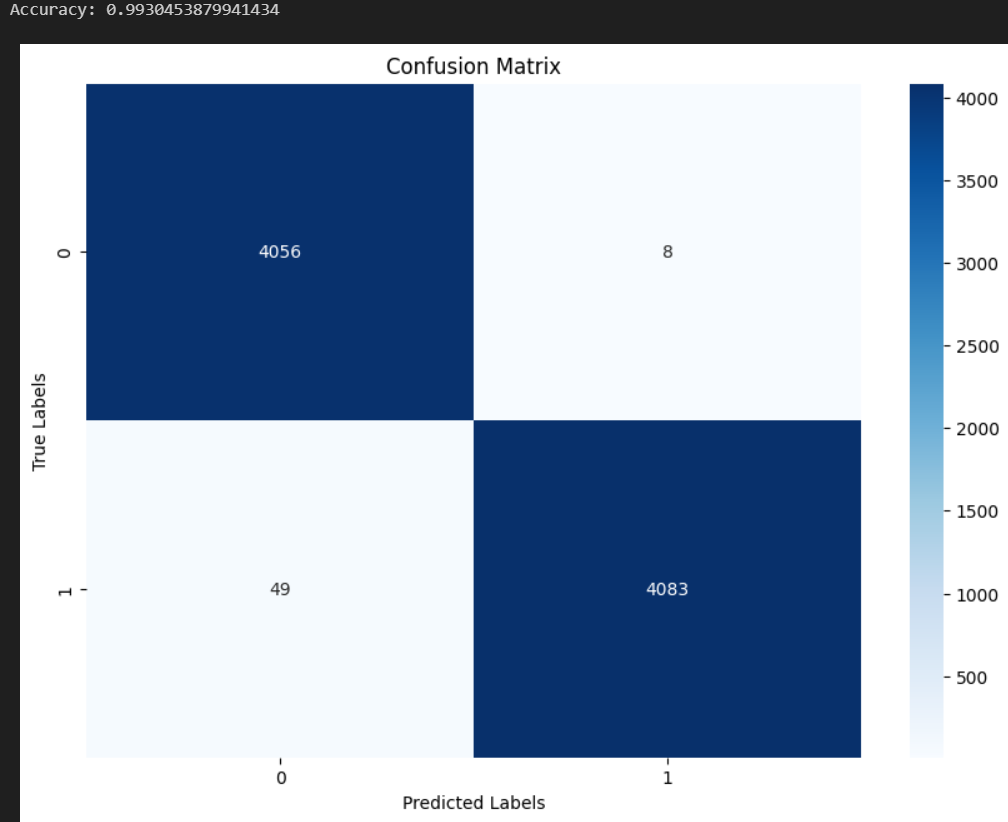
**Noise Generated by GAN:**



* 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.

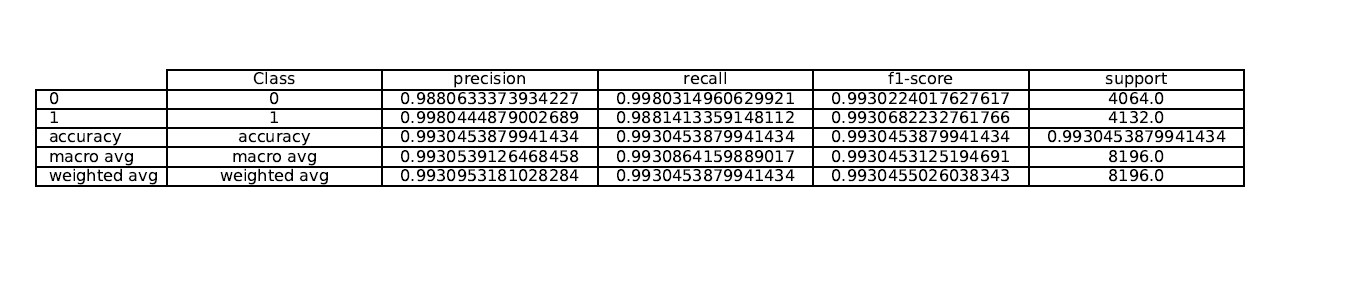
Final Classifier for Distinguishing between Original Class 1 samples

**Confusion Matrix:**



**Classification Report:**

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



* 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.