Before Attacks Results

**The output of “Evaluation Script for each device in the network”**  
  
In this case we trained the model with 100% normal data without any poisoned data and test the Federated learning on each test device data and analysed the output. The data that was fed while training (UNSW-NB15\_1.csv & UNSW-NB15\_2.csv) and testing (UNSW-NB15\_3 & UNSW-NB15\_4.csv) are different datasets.

### **Label:**

0 represents normal data

1 represents attacked data

### **Training and Testing Cases:**

1. Algorithm was trained and test on 100% normal data where there are no instances of attacks 0% attacked data.
2. Algorithm trained with 100% normal data and tested with 50% normal data and 50%backdoor data.
3. Algorithm trained with 100% normal data and tested with 80% backdoor data and 20% normal data.
4. Algorithm trained with 100% normal data and tested with 100% backdoor data

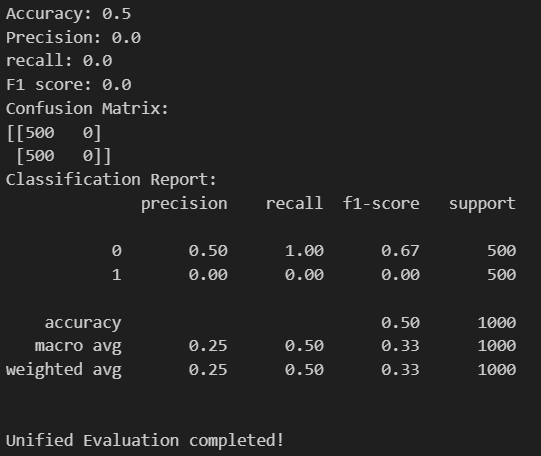
### **Result Analysis:**

#### Case 1: Trained and tested on 100% normal data

In this case the Federated Learning algorithm is trained on 100% normal data and tested on unknown data which does not have any incidences of any kind of attacks.

So, the evaluation metrics shows that the accuracy is 100% which means that the algorithm can correctly analyse all the normal data. Since we have not introduced any kind of attacked/compromised data as our test data, its expected that precision, recall and F1 score will be zero because these evaluation metrics are designed to measure the model’s performance in identifying positive instances(attacks) too.

#### Case 2: Trained with 100% normal data and tested with 50% normal data and 50%backdoor data



In this case the Federated Learning algorithm is trained on 100% normal data and tested on unknown data which has 50%backdoor attack and 50% normal data.

So, the evaluation metrics shows that the accuracy is 50% which means that the algorithm can correctly analyse all the normal data but also recognises the backdoor data as normal data.

#### Case3: Algorithm trained with 100% normal data and tested with 80% backdoor data and 20% normal data. A screenshot of a computer program Description automatically generated

#### Case 4: Algorithm trained with 100% normal data and tested with 100% backdoor data

A screenshot of a computer program

Description automatically generated

### Conclusion:

Since the model was trained only on normal data, it has learned to predict all instances as normal. It hasn’t seen any attack data during training, so it doesn’t recognise the patterns associated with attacks.

Without exposure to attack data during training the model cannot generalize to detect attacks in the test data.