ML for Cyber Security

Lab2 Report

Name: Naveen Mallemala

Net Id: nm3937

→ I have uploaded a jupyter notebook on colab that contains the code for the pruning of networks, saving models for the pruned networks for x=2,4,10,30%.

→ I have saved these models as 'B1_2,B1_4,B1_10 and B1_30.h5' respectively.

→ The backdoor accuracies for the pruned networks are as follows:

→ no pruning : 100%

→ x = 2 : 100%

→ x = 4 : 99.99%

→ x = 10 : 77.02%

→ x = 30 : 15.87%

→ I have created the good network 'G' by combining both the backdoored (B) and pruned models (B1)(followed by fine tuning).

→ The network 'G' predicts N+1th (1283) class if the network B, B1 outputs are different and predicts the class predicted by B/B1 if both are the same.

→ G has a backdoor detection prediction accuracy of 74.5%.

→ The clean dataset accuracy for a good network 'G' is around 88%.

→ This is my github link: https://github.com/naveenmallemala5/MLCyberSecurity