AERO 5 – HANDS ON MACHINE LERANING FOR CYBERSECURITY (2023/2024)

By Leila GHARSALLI

Instructions:

Provide:

- A GitHub repository containing the properly commented Python code.
- A file in .pdf format presenting the work carried out. This involves explaining the problem, the possible solutions and those you have chosen by presenting the strengths and weaknesses. Particular attention should be paid to the explanation of the considered algorithms.

Please note: Projects that do not comply with one of these rules will receive a 20% penalty.

You can send **one solution per group of 4** to: Leila.gharsalli@ipsa.fr by January the 10 at the latest.

Proposition 1: Securing user authentication

As a securing user authentication, Keystroke dynamic information could be used to verify or even try to determine the identity of the person who is producing the keystrokes. The techniques used to do this vary widely in sophistication and range from statistical techniques to artificial intelligence approaches like neural networks. Propose a small tutorial about the Keystroke dynamic information and how it can be a way for protecting sensitive information and assets as well as an anomaly detection tool. Then, based on the dataset 'StrongPasswordData.csv', analyze the latter and propose different classifier models for keystroke detection. Describe every used Machine Learning technique and the motivation behind its use.

Proposition 2: Generative adversarial networks (GANs)

GANs represent the most advanced example of neural networks that deep learning makes available to us in the context of cybersecurity. GANs can be used for legitimate purposes such as authentication procedures, but they can also be exploited to violate these procedures. Propose a small tutorial to getting to know GANs especially their use in attack and defense scenarios then give an example of attacks against intrusion detection systems (IDS) via GANs. The choice of the dataset is free.

Proposition 3: From images to malware

Based on the article 'Towards Building an Intelligent Anti-Malware System: A Deep Learning Approach using Support Vector Machine (SVM) for Malware Classification', show how malware detection could take advantage of the typical skills of CNNs in image recognition. Explain why we should use images for malware detection and how this could be done from images with CNNs? 'images_malware.npz' is a dataset of images of malware codes. Convert your malware codes to grayscale images then apply your classification algorithm to detect malwares. A great attention should be paid to the description and analysis of the solutions.