**MoM AML Project**

The following is the assigned stages:

1. Pre-Processing - **Alberto**
2. Selecting Features - **Dave**
3. Exploring and selecting ML algorithms - **Adam**
4. Refining algorithm - **Ali**
5. Evaluating model and analyzing results – **Andy**

Next Meeting will take place:

**@ 6:30pm Thursday the 21st of November**

Code format:

**Jupyter Notebook**

Schedule:

|  |  |  |  |
| --- | --- | --- | --- |
| **Stage** | **Assignment** | **Deadline for Plan** | **Deadline completion** |
| Pre-Processing | **Alberto** | **17.11.19** | **21.11.19** |
| Selecting Features | **Dave** | **24.11.19** | **28.11.19** |
| Exploring and selecting ML algorithms | **Adam** | **01.12.19** | **05.12.19** |
| Refining algorithm | **Ali** | **8.12.19** | **12.12.19** |
| Evaluating model and analyzing results | **Andy** | **To be confirmed** | **To be confirmed** |

References:

1. All references should be from journals and publication that is **accepted** in our project.
2. Take a note of each reference along the way in **order** of reference
3. Have all the references **saved on a word doc** and sent to **ALI**

Extra responsibilities:

1. Creating a Gantt Chart – **Adam**
2. Referencing and appendix – **Ali**
3. Format of code and upload (checking whether markdown works for us) – **Dave**
4. Writing out the plan at the end and integrating it to our report – **Andy**

Advice and recommendation from publications:

**Best classifiers:**

1. J48 (c4.5) decision tree
2. Random Forest
3. Naïve Bayes

**Optimization:**

1. Learning speed is commonly rejected in Intrusion Detection Systems (IDS)
2. Try to achieve extremely low False Positive Rate (FPR of 1% is bad)
3. Make model highly adaptive to network behavior
4. Be able to detect novel attacks
5. Have independence from supervised learning (labelling data is not practical when data are in the multi-Gigabytes level.