**Dissertation Project Plan**

Meta-Heuristics and Machine Learning for Detecting Cyber Attacks

Christos Christou Mavropoulos

**Description:**

Much work has been done in intrusion/anomaly detection systems. In 2015 the UNSW-NB15 was generated providing a chance to train, test and experiment with newer algorithms on a database with more features. The purpose of this project is to firstly process the dataset making it fitter for the machine learning algorithms used (i.e. dimensionality reduction) and subsequently to apply and compare different anomaly detection strategies in order to create an Intrusion Detection System.

**Preliminary preparation:**

* An understanding of the python programming language and data processing and machine learning libraries such as pytorch, sklearn, numpy and pandas.
* Familiarity with statistics, machine learning and networking concepts.

**Objectives:**

**Basic:**

* Encode categorical data.
* Balance dataset classes.
* Apply and compare dimensionality reduction methods.
* Apply and compare anomaly detection methods.

**Intermediate:**

* Use different combinations of dimensionality reduction and anomaly detection methods.
* Visualise results in as few dimensions as possible while not losing too much information.

**Advanced:**

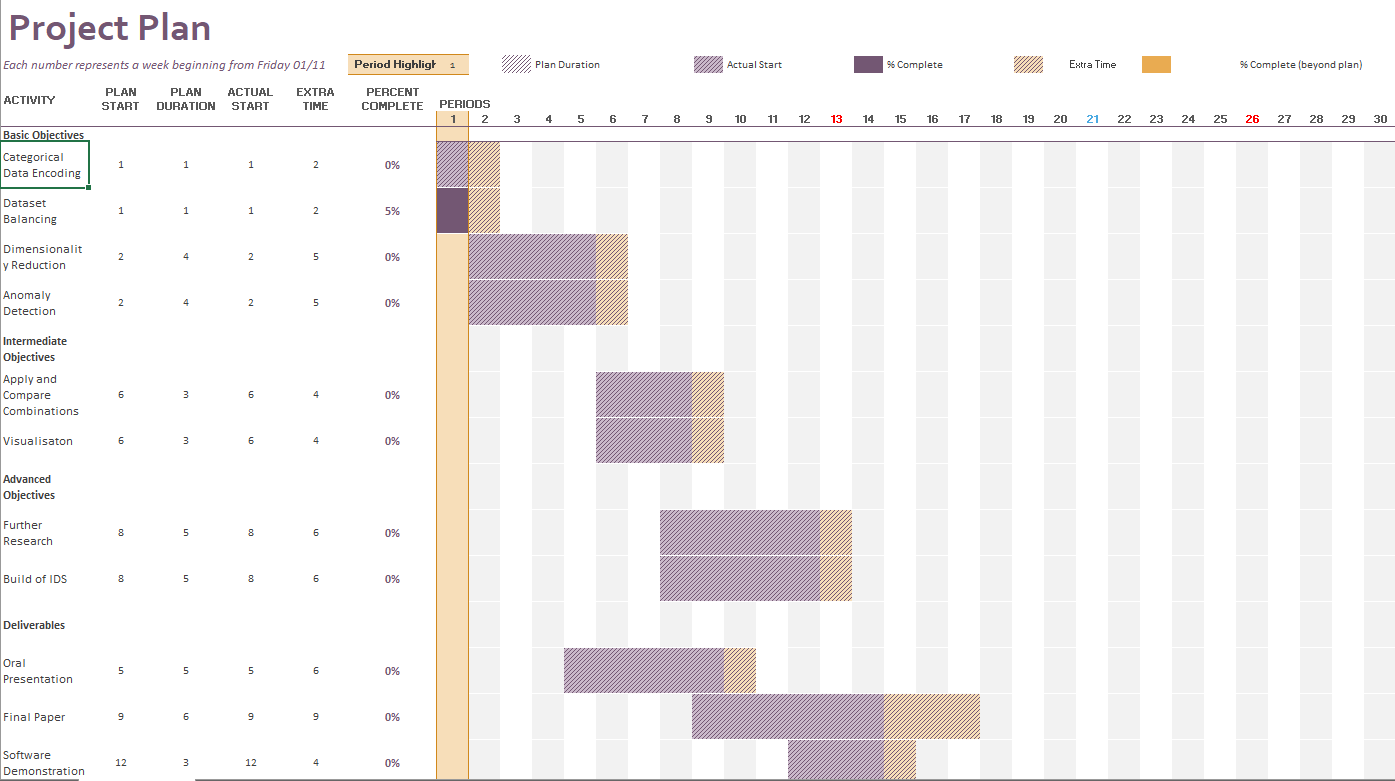
* Further research into newer and more advanced techniques.
* Build of an intrusion detection system (IDS) that can work in real time, using the above methods.

**Project Plan**

The project plan is broken down in the three above sections. Before any other operations the categorical data should be encoded, and the dataset should be balanced. Some algorithms will work without some of these pre-processing steps however its better if they are done for all of them so that all the algorithms are tested on similarly structured data.

Furthermore, in terms of the intermediate objectives some work will be done for them in the basic objectives section. The work done for these should extend the work previously done, for example while applying some dimensionality reduction technique such as t-SNE, one can easily visualise the results.

Lastly, the advanced objectives are additional objectives if the previous one has been completed to a satisfactory degree. Their purpose is to build a system that checks network traffic and categorizes normal and abnormal behaviour. Because the purpose of the system is to work in real time new or alternative methods might be required.

****