**Machine Learning Planning**

1. **Dataset:**

Existing Ones:

* [API Call based Malware Dataset](https://www.kaggle.com/datasets/focatak/malapi2019)
* [Windows\_Malware\_Detection\_Dataset](https://www.kaggle.com/datasets/radwahashiesh/windows-malware-detection-dataset?rvi=1)
* [VHS-22](https://www.kaggle.com/datasets/f3a18ca1e8f028aee166b12128e9532dc96d1005f3eb33e18db92c145939bc55)
* [SDN Dataset for DDoS Flooding Attack Detection](https://www.kaggle.com/datasets/mobayodeakinsolu/sdn-dataset-for-ddos-flooding-attack-scenarios)

More:

<https://www.unb.ca/cic/datasets/index.html>

<https://registry.opendata.aws/sorel-20m/>

Create Alone or to make new ones when the ML is Working : <https://cuckoo.sh/blog/>

1. **Model:**

Random Forests:

Type: Ensemble Learning

Description: Random Forest is an ensemble learning method that constructs a multitude of decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. It is robust, handles high-dimensional data well, and is less prone to overfitting.