1. Choose the type of prediction model you will use (e.g., linear regression, deep neural network, etc. )

* We are planning to use Random Forest, decision tree and extra tree classifier.

1. Justify the type of prediction model that you chose.

Random Forest Classifier is the most accurate supervised learning method. In this each observation is considered as the one decision tree.

1. Choose the software library that you will use to implement the model.

* from sklearn.tree import DecisionTreeClassifier
* from sklearn.ensemble import RandomForestClassifier
* from sklearn.ensemble import ExtraTreesClassifier

1. Assuming your predictive model is trained, describe the input data that your model will use during the inference phase.

* The input to our models will be trained data set, X\_train and Y\_train.

1. Assuming your predictive model is trained, describe the output of your model during the inference phase.

* The output of the models will consist of Accuracy, precision, recall andf1-score.

1. Describe the parameters of your model (e.g., how many hidden layers, how many features, etc).

* There are four main classification features in the data set that is Flow bytes, Packet length, Packet Time, Inter-packet delay. So on these main four features we will be classifying the malicious DoH, Bengie DoH, and Non-DoH.
* We will be using max\_features and n\_estimators , n\_jobs and random\_state Random Forest.
* Criterion and max\_depth in extra tree classifier.