In [1]:
```
#Created by Rami ALmehdawi

# Import the necessary Libraries

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import sklearn as sk
from sklearn import tree
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix, classification_report, roc_
from imblearn.over_sampling import RandomOverSampler, SMOTE
```

In [2]:
```
# Read in the Data
MW_Data = pd.read_csv("Malware-staDyn-data.csv")
```

In [3]:
```
# Drops all Missing values in Label and selects the Feature Column
x = MW_Data.drop('label', axis = 1)
y = MW_Data["label"]

# Split into Training and Test Sets
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.1, random_state=42)
```

In [9]:
```
# Creates the random forest. n_Estimators is the number of trees.
rf = RandomForestClassifier(n_estimators=10, random_state=42, class_weight= {1:2, 0:1})
```

In [10]:
```
# Trains the Random Forest
rf.fit(x_train, y_train)
```

Out[10]:
```
▾                      RandomForestClassifier

RandomForestClassifier(class_weight={0: 1, 1: 2}, n_estimators=10,
                       random_state=42)
```

In [11]:
```
y_pred = rf.predict(x_test)

y_pred_proba = rf.predict_proba(x_test)[:, 1]
```

In [12]:
```
# Prints out Relevant Metrics
print("Accuracy: ", accuracy_score(y_test, y_pred))
```

```
print("AUC-ROC:", roc_auc_score(y_test, y_pred_proba))

print("F1 Score: ", f1_score(y_test, y_pred))

print("Precision:", precision_score(y_test, y_pred))

print("Recall: ", recall_score(y_test, y_pred))
```

```
Accuracy:  0.9904
AUC-ROC: 0.9997766527695056
F1 Score:  0.9945750452079566
Precision: 0.9927797833935018
Recall:  0.9963768115942029
```