```
In [44]: #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 [45]: # Read in the Data
MW_Data = pd.read_csv("Malware-staDyn-data.csv")
```

In [46]: # Plots the Pandas Data Frame as a check.
MW\_Data

Out[46]:		Virtual	Offset	loc	Import	Imports	var	Forwarder	UINT	LONG	BOOL	•••	count_file_written	count_file_exists	count_file_deleted
	0	0	0	1	0	0	0	0	0	0	0		0	0	0
	1	0	0	0	0	0	0	0	0	0	0		0	0	0
	2	0	0	1	0	0	0	0	0	0	0		0	0	0
	3	0	0	1	0	0	0	0	0	0	0		0	0	0
	4	0	0	0	0	0	0	0	0	0	0		0	0	0
	•••												<del></del>		<b></b>
	6243	0	0	0	0	0	0	0	0	0	0		0	0	0
	6244	0	0	58	0	0	0	0	0	0	0		0	4	0
	6245	0	0	0	0	0	0	0	0	0	0		0	0	0
	6246	0	0	0	0	0	0	0	0	0	0		0	0	0
	6247	0	0	0	0	0	0	0	0	0	0		0	0	0

6248 rows × 1085 columns

```
In [47]: # 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)

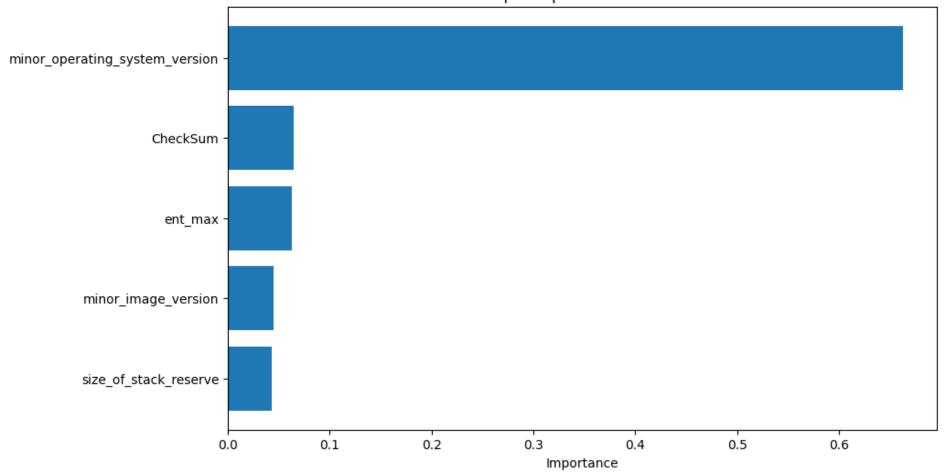
# Initializes the Decision Tree and fits the model.
    dtc = DecisionTreeClassifier()
    dTree1 = DecisionTreeClassifier(class_weight= None)
    dTree1.fit(x_train, y_train)

feature_importances = dTree1.feature_importances_
    sorted_idx = feature_importances.argsort()[-5:][::-1]

plt.figure(figsize=(10, 6))
    plt.barh(x.columns[sorted_idx], feature_importances[sorted_idx], align='center')
    plt.xlabel('Importance')
    plt.title('Top 5 Important Features')
```

```
plt.gca().invert_yaxis() # To display the most important feature at the top
plt.show()
```

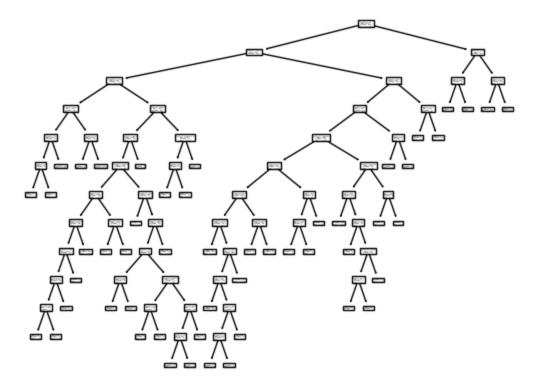




```
In [55]: # Drops all Missing values in Label and selects the Feature Column
    x = MW_Data.drop('label', axis = 1)
    y = MW_Data["label"]

# 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)

# Initializes the Decision Tree and fits the model.
    dTree2 = DecisionTreeClassifier(class_weight= {1:2, 0:1})
    dTree2.fit(x_train, y_train)
    tree.plot_tree(dTree2);
    plt.figure(figsize=(20, 10))
    plot_tree(dTree2, filled=True, feature_names=MW_Data["label"], proportion=True, fontsize=10)
    plt.show()
```



```
gini = 0.093
                                                                                                                                                         samples = 100.0%
                                                                                                                                                        value = [0.049, 0.951]
                                                                                                            1 <= 294912.0
                                                                                                             gini = 0.028
                                                                                                                                                                                                             aini = 0.2
                                                                                                          samples = 92.9%
                                                                                                                                                                                                          samples = 7.1\%
                                                                                                         value = [0.014, 0.986]
                                                                                                                                                                                                        value = [0.887, 0.113]
                                                 1 <= 2.776
                                                                                                                                                                         1 <= 460.5
                                                 gini = 0.181
                                                                                                                                                                        gini = 0.003
                                                                                                                                                                                                   gini = 0.048
                                                                                                                                                                                                                    gini = 0.041
                                               samples = 12.9\%
                                                                                                                                                                      samples = 79.9\%
                                                                                                                                                                                                  samples = 0
                                                                                                                                                                                                                  samples = 6.8%
                                             value = [0.101, 0.899]
                                                                                                                                                                    value = [0.001, 0.999]
                                                                                                                                                                                               value = [0.024, 0]
                              1 \le 12418.0
                               gini = 0.444
                                                                   gini = 0.059
                                                                                                                                                          qini = 0.002
                                                                                                                                                                                             sample sample sample
                                                                                                                                                                                                                      samples = 0.1\%
                             samples = 3.6\%
                                                                  samples = 9.3\%
                                                                                                                                                        samples = 79.6%
                                                               value = [0.031, 0.969]
                                                                                                                                                                                                      valuė =
                                                                                                                                                                                                                      value = [0.0, 1.0]
                                                                                                                                                                                             value =
                                                                                                                                                      /alue = [0.001, 0.999]
                          value = [0.333, 0.667]
                                                                                                                                                                                   /alue = [0
                                                                                                                                        <= 1258293
                      gini = 0.01
                                                       qini = 0.042
                                                                               qini = 0.444
                                                                                                                                         gini = 0.002
                                                                                                                                                                                         samples = 0.3\%
                                     samples = 2
                                                      samples = 8.9\%
                                                                             samples = 0.4\%
                                                                                                                                       samples = 79.5%
                                                                                                                                                                                        value = [0.0, 1.0]
                                                                          value = [0.333, 0.667]
                  value = [0.007, 0]
                                  value = [0.652, 0]
                                                   value = [0.021, 0.979]
                                                                                                                                     /alue = [0.001, 0.999]
                                                                                                                      1 <= 559.5
                                                                                                                     gini = 0.001
                                                                                                                                                             gini
                                                           samples
                                                                                  samples = 0.2\%
                                                                                                                                                                            samples = 0.0%
                                                                                                                   samples = 76.7\%
                                                  sampl
                                                                                                                                                           sample
                                                                                 value = [1.0, 0.0
                                                                                                                                                                            value = [1.0, 0.0]
                        value =
                                value =
                                                          value = [1.
                                                                                                                                                                    value =
                                                                                                                 value = [0.001, 0.999]
                                                                                                                                                          value =
                                                                                                      gini = 0.001
                                                                                                                                     gini = 0.03
                                                                                                                                                     gini = 0.013
                                                                                                                                                                      gini = 0.245
                                                                             samples = 0.2\%
                    samples = 0.0^{\circ}
                                                                     sample
                                       samples = 6.9\%
                                                            sample
                                                                                                     amples = 76.1%
                                                                                                                                  samples = 0.6
                                                                                                                                                   samples = 2.
                                                                                                                                                                    samples = 0.1\%
                    value = [1.0, 0.0]
                                                                             value = [0.0, 1.0]
                                     value = [0.004, 0.99d value = [0
                                                                                                     value = [0.0, 1.0]
                                                                                                                                value = [0.015, d value = [0.007, 0
                                                                                                                                                                 value = [0.143, 0.857]
                                                                                            gini = 0.001 gini = 0.021 gini = gini | gample | sample | sample | sample | sample | sample | value = [0.011, 0 value = [0]
                                                                  gini = 0.123
                                                 gini =
                                                                                                                                                         gini =
                                                                                                                                                                        samples = 0.1\%
                               samples = 6.
                                                sample
                                                                 samples = 2.0\%
                                                                                                                                                                        value = [0.0, 1.0]
                                                              value = [0.066, 0.934]
                             value = [0.003, 0 value = [0
                                                                                                                                                             qini = 0.062
                                   sample
                                                                     samples = 1.0\%
                                                                                                                                  samples = 0.0\%
                                                            sample
                                                                                                                                                           samples = 0.3%
                                                   value
                                                                     value = [0.0, 1.0]
                                                                                                                 value = value =
                                                                                                                                  value = [1.0, 0.0
                                                         value = [0]
                                                                                                                                                         value = [0.032, 0.968]
                                                                                                                                                                  gini = 0.0
                                                    gini = 0.346
                                                                         gini = 0.151
                               samples = 0.0?
                                                                                                    samples = 12.6\%
                                                                                                                                                                samples = 0.2%
                                                  samples = 0.1\%
                                                                       samples = 0.8\%
                                                                                                    value = [0.0, 1.0]
                                                                                                                                                               value = [0.0, 1.0]
                                                value = [0.778, 0.222 value = [0.082, 0.918]
                                                                gini = 0.494
                                                                                                   gini = 0.32
                           samples = 0.6\%
                                                                                                                                                           samples = 0.0\%
                                                              samples = 0.3
                                                                                                 samples = 0.2\%
                value = [0] value = [0.0, 1.0]
                                                                                                                                                   value = value = [1.0, 0.0]
                                                            value = [0.556, 0]
                                                                             value = [0
                                                                                                value = [0.2, 0.8]
                                                                                                     samples = 0.0\%
                       samples = 0.0\%
                       valuė = [1.0, 0.0]
                                                                                                     value = [1.0, 0.0]
                                                                               sample
                                                                                        sampl
                                                                                                samples = 0.1\%
                                                                                                 value = [0.0, 1.0]
In [ ]: | ros = RandomOverSampler(random_state=42)
              x_train_resampled, y_train_resampled = ros.fit_resample(x_train, y_train)
              MW_Data = pd.concat([pd.DataFrame(x_train_resampled, columns=MW_Data.columns[:-1]), pd.DataFrame(y_train_resampled, columns=['labe
```

1 <= 1.5

```
x_train_resampled, y_train_resampled = ros.fit_resample(x_train, y_train)

MW_Data = pd.concat([pd.DataFrame(x_train_resampled, columns=MW_Data.columns[:-1]), pd.DataFrame(y_train_resampled, columns=['labe

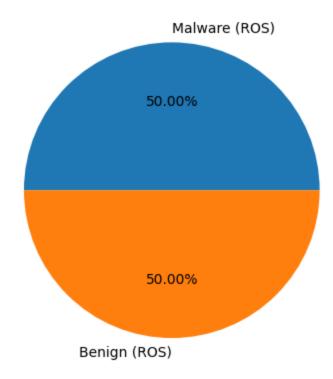
# Counts number of 0 or 1, occurences within the data. Will be used in Pie chart

MW_Data_count = MW_Data["label"].value_counts()
print(MW_Data_count)
Not_Malware_count = MW_Data_count[1]
Malware_count = MW_Data_count[0]

# Code for Pie chart.
pie_labels = "Malware (ROS)", "Benign (ROS)"
pie_sizing = [int(Malware_count), int(Not_Malware_count)]

fig, ax = plt.subplots()
ax.pie(pie_sizing, labels=pie_labels, autopct='%1.2f%%');
```

```
label
1   5101
0   5101
Name: count, dtype: int64
```



```
In [56]: dTree2 = DecisionTreeClassifier(random_state=42)
    dTree2.fit(x_train_resampled, y_train_resampled)
    plt.figure(figsize=(20, 10))
    plot_tree(dTree2, filled=True, feature_names=MW_Data["label"], proportion=True, fontsize=10)
    plt.show()
```

```
samples = 100.0%
                                                                                                                                                  1 <= 2.966
                                                                                                                                                                                                                      value = [0.5, 0.5]
                                                                                                                                                                                                                                                                                                  1 \le 3072.0
                                                                                                                                                  gini = 0.023
                                                                                                                                                                                                                                                                                                  gini = 0.209
                                                                                                                                             samples = 43.8\%
                                                                                                                                                                                                                                                                                              samples = 56.2%
                                                                                                           1 <= 7.5
                                                                                                                                          value = [0.012, 0.988]
                                                                                                                                                                                           1 <= 127.0
                                                                                                                                                                                                                                                     1 \le 4200.0
                                                                                                                                                                                                                                                                                       value = [0.881, 0.119]
                                                                                                                                                                                                                                                                                                                                               1 <= 2.755
                                                                                                         gini = 0.019
                                                                                                                                                                                          qini = 0.426
                                                                                                                                                                                                                                                      gini = 0.087
                                                                                                                                                                                                                                                                                                                                              gini = 0.112
                                                                                                     samples = 43.7%
                                                                                                                                                                                       samples = 0.1%
                                                                                                                                                                                                                                                 samples = 51.6%
                                                                                                                                                                                                                                                                                                                                           samples = 4.6%
                                                                                                                                                                                                                                                                                                                                1 <= value = [0.06, 0.94] 0.0
                                                                     1 <= 0.073
                                                                                                   value = [0.01, 0.99]
                                                                                                                                               1 <= 18.5
                                                                                                                                                                                                                                               value = [0.955, 0.045] |5
                                                                                                                                                                                   value = [0.692, 0.308]
                                                                                                                                                                                                                                                                    <del>gini = 0.</del>036
                                                                    gini = 0.015
                                                                                                                                              gini = 0.461
                                                                                                                                                                                                                                                                                                                               gini = 0.454
                                                                                                                                                                                                                                  samples = 1.4
                                                                                                                                                                         samples = 0.0% | samples = 0.1%
                                                                                                                                                                                                                                                                                                                                                          samples = 4.2\%
                                                                                                                                                                                                                                                                                                                            samples = 0.4%
                                                                samples = 43.5\%
                                                                                                                                           samples = 0.2\%
                                                                                                                                                                                                                                                                samples = 50.1%
                                                                                                                                                                        value = [0.0, 1.0] value = [1.0, 0.0] value = [0.0, 1.
                                                                                                                                                                                                                                                                                                                                                         value = [0.0, 1.0]
                                      1 <= 2560.0 value = [0.008, 0.992] | 1 <= 0.075
                                                                                                                                        value = [0.36, 0.64
                                                                                                                                                                                                                                                              alue = [0.981, 0.019] 5
                                                                                                                                                                                                                                                                                                                        value = [0.651, 0.34
                                       gini = 0.01
                                                                                                  gini = 0.329
                                                                                                                             samples = 0.2\% | samples = 0.1\%
                                                                                                                                                                                                                                                   samples = 0.2
                                                                                                                                                                                                                                                                                                                                           samples = 0.3%
                                  samples = 42.9\%
                                                                                              samples = 0.5\%
                                                                                                                                                                                                                                                                               samples = 49.9\%
                                                                                                                           value = [0.0, 1.0] value = [1.0, 0.0]
                                                                                                                                                                                                                                                                                                             value = [0.0, 1.0] value = [1.0, 0.0]
                                                                                                                                                                                                                                                 value = [0.0, 1]
                               value = [0.005, 0.995] | 5
                                                                                          value = [0.208, 0.79
                                                                                                                                                                                                                                                                            value = [0.986, 0.0]
                                                                                                                                                                                                                                                                                                                        qini = 0.436
                                                                                                                                                                                                                                               gini = 0.02
                                                     gini = 0.052
                    samples = 34.4
                                                                                samples = 0.1\% | samples = 0.4\%
                                                  samples = 8.5\%
                                                                                                                                                                                                                                          samples = 49.6%
                                                                                                                                                                                                                                                                                                                    samples = 0.3\%
                    value = [0.0, 1]
                                                                                value = [1.0, 0.0] value = [0.0, 1.0]
                                             value = [0.027, 0.97
                                                                                                                                                                                                                                         value = [0.99, 0.01]
                                                                                                                                                                                                                                                                            0 <= 11.0
                                                                                                                                                                                                                                                                                                                value = [0.321, 0.679]
                                                                    <del>gm – v.</del>211
                                                                                                                                                                                                                gini = 0.201
                                                                                                                                                                                                                                                                            gini = 0.005
                                    samples = 6.6\%
                                                                                                                                                                                                                                                                                                      samples = 0.2% | samples = 0.1%
                                                                samples = 1.9\%
                                                                                                                                                                                                             samples = 3.5\%
                                                                                                                                                                                                                                                                       samples = 46.29
                                   value = [0.0, 1.0
                                                                                                                                                                                                                                                                                                     value = [0.0, 1.0] value = [1.0, 0.0]
                                                               value = [0.12, 0.88]
                                                                                                                                                                                                        value = [0.887, 0.113]
                                                                                                                                                                                                                                                          1 <= value = [0.997, 0.00
                                                  samples = 0.3% samples = 1.6%
                                                                                                                                                                                                                            samples = 0.3
                                                                                                                                                                                                                                                                                       samples = 0.0\%
                                                                                                                                                                                              samples = 3.2\%
                                                                                                                                                                                                                                                         samples = 46.1\%
                                                                              value = [0.0, 1.0]
                                                                                                                                                                                                                           value = [0.0, 1]
                                                                                                                                                                                                                                                                                      value = [0.0, 1.0]
                                              value = [0.885, 0.11
                                                                                                                                                                                           value = [0.969, 0.03
                                                                                                                                                                                                                                                      value = [0.998, 0.00
                                                                                                                                                                                                                <del>gını = 0.0</del>31
                                    samples = 0.0% samples = 0.2%
                                                                                                                                                                                                                                           samples = 0.0
                                                                                                                                                                                                             samples = 3.1\%
                                                                                                                                                                                                                                                                       samples = 46.1%
                                                                                                                                                                              value = [0.0, 1.
                                                                                                                                                                                                                                          value = [0.0, 1
                                  value = [0.0, 1.0] value = [1.0, 0.0]
                                                                                                                                                                                                                                                                    value = [0.999, 0.001]
                                                                                                                                                                                                         value = [0.984, 0.01
                                                                                                                                                                                                 gin = 0.013
                                                                                                                                                                                                                                                             ginl = 0.002
                                                                                                                                                                                                                            samples = 0.0
                                                                                                                                                                                                                                                                                       samples = 0.0\%
                                                                                                                                                                                              samples = 3.1\%
                                                                                                                                                                                                                                                         samples = 46.1%
                                                                                                                                                                                                                           value = [0.0, 1]
                                                                                                                                                                                                                                                                                      value = [0.0, 1.0]
                                                                                                                                                                                   1 < value = [0.994, 0.00
                                                                                                                                                                                                                                                      value 😑 [0.999, 0.00
                                                                                                                                                                                   gin<del>l = 0.000</del>
                                                                                                                                                                                                                                              gin<del>l = 0.002</del>
                                                                                                                                                                                                             samples = 0.0
                                                                                                                                                                                                                                                                        samples = 0.0\%
                                                                                                                                                                                samples = 3.1\%
                                                                                                                                                                                                                                         samples = 46.1%
                                                                                                                                                                                                            value = [0.0, 1]
                                                                                                                                                                                                                                                                        value = [0.0, 1.0]
                                                                                                                                                                                                                                       value = [0.999, 0.00
                                                                                                                                                                                                                                                            <del>gını = 0.</del>018
                                                                                                                                                                  samples = 3.1\% samples = 0.0\% samples = 41.
                                                                                                                                                                                                                                                         samples = 4.4\%
                                                                                                                                                                value = [1.0, 0.0] value = [0.0, 1.0] value = [1.0, 0
                                                                                                                                                                                                                                                     value = [0.991, 0.009] |.5
                                                                                                                                                                                                                                                                           \frac{1}{1} \frac{1}
                                                                                                                                                                                                                                                                        samples = 4.4\%
                                                                                                                                                                                                                                           value = [0.0, 1]
                                                                                                                                                                                                                                                                    value = [0.993, 0.007] | 01
                                                                                                                                                                                                                                                          samples = 0.0
                                                                                                                                                                                                                                                                                      samples = 4.4\%
                                                                                                                                                                                                                                                         value = [0.0, 1
                                                                                                                                                                                                                                                                                   value = [0.996, 0.004]
                                                                                                                                                                                                                                                                           gin<del>l = 0.231</del>
                                                                                                                                                                                                                                                                                                     samples = 4.3%
                                                                                                                                                                                                                                                                        samples = 0.1%
                                                                                                                                                                                                                                                                                                     value = [1.0, 0.0]
                                                                                                                                                                                                                                                                    value = [0.867, 0.13]
                                                                                                                                                                                                                                                          samples = 0.1\% samples = 0.0\%
                                                                                                                                                                                                                                                          value = [1.0, 0.0] value = [0.0, 1.0]
            |: | y pred = dTree2.predict(x test)
             ]: |y_pred_proba = dTree2.predict_proba(x_test)[:, 1] # Assuming model is your trained classifier
In [ ]: # 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))
```

1 <= 0.5 gini = 0.5 Accuracy: 0.9952

AUC-ROC: 0.9913390907286083 F1 Score: 0.9972801450589303 Precision: 0.9981851179673321 Recall: 0.9963768115942029