```
Importing All The Necessary Stuff
In [ ]: import pandas as pd
        from sklearn.tree import DecisionTreeClassifier
        from sklearn.model_selection import train_test_split
        from sklearn.metrics import accuracy_score
        Loading The Data to prepare a Model
In [ ]: # Load the data from CSV file
        # df = pd.read_csv(r"D:\Stuff\CyberSec\archive\02-14-2018.csv")
        df = pd.read_csv(r"D:\Stuff\CyberSec\archive\03-02-2018.csv")
        # Remove any rows with missing values
        df = df.dropna()
        # Drop columns where all values are 0
        df = df.loc[:, (df != 0).any(axis=0)]
        columns = list(df.columns)
        Display the Imported Model
In [ ]: # display the dataframe
        df.head()
                               Timestamp Flow Duration Tot Fwd Pkts Tot Bwd Pkts TotLen Fwd Pkts TotLen Bwd Pkts Fwd Pkt Len Max Fwd Pkt Len Min ... Fwd Seg Size Min Active Mean Active Max Active Max Active Min Idle Mean Idle Max Idle Min Label
Out[]:
          Dst Port Protocol
                       6 02/03/2018 08:47:38
                                              141385
                                                                                    553
                                                                                               3773.0
                                                                                                                              0 ...
                                                                                                                                                        0.0
                                                                                                                                                                          0.0
                                                                                                                                                                                           0.0
                                                                                                                                                                                                  0.0
                                                                                                                                                                                                                 0.0 Benign
        1 49684
                       6 02/03/2018 08:47:38
                                                281
                                                                                                 0.0
                                                                                                                38
                                                                                                                              0 ...
                                                                                                                                              20
                                                                                                                                                        0.0
                                                                                    38
                                                                                                                                                                 0.0
                                                                                                                                                                          0.0
                                                                                                                                                                                   0.0
                                                                                                                                                                                           0.0
                                                                                                                                                                                                  0.0
                                                                                                                                                                                                          0.0
                                                                                                                                                                                                                 0.0 Benign
             443
                       6 02/03/2018 08:47:40
                                              279824
                                                            11
                                                                       15
                                                                                   1086
                                                                                              10527.0
                                                                                                               385
                                                                                                                              0 ...
                                                                                                                                              20
                                                                                                                                                        0.0
                                                                                                                                                                 0.0
                                                                                                                                                                          0.0
                                                                                                                                                                                   0.0
                                                                                                                                                                                           0.0
                                                                                                                                                                                                  0.0
                                                                                                                                                                                                          0.0
                                                                                                                                                                                                                 0.0 Benign
                                                                                                 0.0
                                                                                                                              0 ...
                                                                                                                                              20
                                                                                                                                                                                           0.0
             443
                       6 02/03/2018 08:47:40
                                                132
                                                                                                                                                        0.0
                                                                                                                                                                                   0.0
                                                                                                                                                                                                  0.0
                                                                                                                                                                                                          0.0
                                                                                                                                                                                                                 0.0 Benign
        3
                                                                                                                                                                 0.0
                                                                                                                                                                          0.0
                                                                                   1285
                                                                                                               517
                                                                                                                                              20
                                                                                                                                                        0.0
              443
                       6 02/03/2018 08:47:41
                                             274016
                                                                                               6141.0
                                                                                                                              0 ...
                                                                                                                                                                          0.0
                                                                                                                                                                                   0.0
                                                                                                                                                                                           0.0
                                                                                                                                                                                                  0.0
                                                                                                                                                                                                                 0.0 Benign
       5 rows × 70 columns
        Printing all the Features available in our Model
In [ ]: # Print column names with their respective column numbers
        for i, col_name in enumerate(df.columns):
           print(f"Feature {i+1}:\t\"{col_name}\"")
        Feature 1:
                        "Dst Port"
                       "Protocol"
        Feature 2:
        Feature 3:
                       "Timestamp"
        Feature 4:
                       "Flow Duration"
        Feature 5:
                       "Tot Fwd Pkts"
        Feature 6:
                        "Tot Bwd Pkts"
        Feature 7:
                        "TotLen Fwd Pkts"
                        "TotLen Bwd Pkts"
        Feature 8:
        Feature 9:
                       "Fwd Pkt Len Max"
        Feature 10:
                       "Fwd Pkt Len Min"
        Feature 11:
                       "Fwd Pkt Len Mean"
                       "Fwd Pkt Len Std"
        Feature 12:
                       "Bwd Pkt Len Max"
        Feature 13:
                        "Bwd Pkt Len Min"
        Feature 14:
        Feature 15:
                        "Bwd Pkt Len Mean"
        Feature 16:
                       "Bwd Pkt Len Std"
                        "Flow Byts/s
        Feature 17:
        Feature 18:
                       "Flow Pkts/s"
        Feature 19:
                       "Flow IAT Mean'
        Feature 20:
                       "Flow IAT Std"
        Feature 21:
                       "Flow IAT Max"
                       "Flow IAT Min"
        Feature 22:
                       "Fwd IAT Tot"
        Feature 23:
        Feature 24:
                       "Fwd IAT Mean"
        Feature 25:
                       "Fwd IAT Std"
        Feature 26:
                       "Fwd IAT Max"
        Feature 27:
                       "Fwd IAT Min"
                        "Bwd IAT Tot"
        Feature 28:
                        "Bwd IAT Mean'
        Feature 29:
        Feature 30:
                        "Bwd IAT Std"
        Feature 31:
                        "Bwd IAT Max"
        Feature 32:
                       "Bwd IAT Min"
                       "Fwd PSH Flags"
        Feature 33:
        Feature 34:
                       "Fwd Header Len"
                       "Bwd Header Len"
        Feature 35:
        Feature 36:
                       "Fwd Pkts/s"
                        "Bwd Pkts/s"
        Feature 37:
                        "Pkt Len Min"
        Feature 38:
        Feature 39:
                        "Pkt Len Max"
        Feature 40:
                        "Pkt Len Mean"
        Feature 41:
                       "Pkt Len Std"
        Feature 42:
                       "Pkt Len Var"
        Feature 43:
                       "FIN Flag Cnt"
        Feature 44:
                        "SYN Flag Cnt"
        Feature 45:
                        "RST Flag Cnt"
        Feature 46:
                        "PSH Flag Cnt"
        Feature 47:
                        "ACK Flag Cnt"
        Feature 48:
                        "URG Flag Cnt"
        Feature 49:
                       "ECE Flag Cnt"
        Feature 50:
                       "Down/Up Ratio"
        Feature 51:
                       "Pkt Size Avg"
                       "Fwd Seg Size Avg"
        Feature 52:
        Feature 53:
                       "Bwd Seg Size Avg"
                        "Subflow Fwd Pkts"
        Feature 54:
        Feature 55:
                        "Subflow Fwd Byts"
        Feature 56:
                        "Subflow Bwd Pkts"
        Feature 57:
                       "Subflow Bwd Byts"
        Feature 58:
                       "Init Fwd Win Byts'
        Feature 59:
                       "Init Bwd Win Byts"
                       "Fwd Act Data Pkts"
        Feature 60:
                       "Fwd Seg Size Min"
        Feature 61:
        Feature 62:
                        "Active Mean"
        Feature 63:
                        "Active Std"
        Feature 64:
                        "Active Max"
        Feature 65:
                        "Active Min'
        Feature 66:
                       "Idle Mean"
        Feature 67:
                       "Idle Std"
                       "Idle Max"
        Feature 68:
        Feature 69:
                       "Idle Min"
                       "Label"
        Feature 70:
        Preparing the Classification model using Decision Tree Classifier
        Training and Obtaining accuracy for a range of features individually
In []: # to display individual features with a given range
        startC = 1
        endC = 10
        for start_col in range(startC-1, endC, 1):
            end_col = start_col + 1
               X = df.iloc[:, start_col:end_col].values
               y = df.iloc[:, -1].values
                # Split the data into training and testing sets
               X_train, X_test, y_train, y_test = train_test_split(
                   X, y, test_size=0.2, random_state=42)
               # Train the decision tree classifier
               clf = DecisionTreeClassifier()
               clf.fit(X_train, y_train)
               # Test the classifier
               accuracy = clf.score(X_test, y_test)
               print(f"Accuracy for column '{columns[end_col-1]}' : {accuracy}")
                print(f"'{columns[end_col-1]}' isn't a feature column")
        Accuracy for column 'Dst Port' : 0.9956549587961989
        Accuracy for column 'Protocol' : 0.7262337240205732
        'Timestamp' isn't a feature column
        Accuracy for column 'Flow Duration': 0.9491596718991989
        Accuracy for column 'Tot Fwd Pkts' : 0.7796026844611001
        Accuracy for column 'Tot Bwd Pkts' : 0.8112177587426627
        Accuracy for column 'TotLen Fwd Pkts' : 0.861603984627445
        Accuracy for column 'TotLen Bwd Pkts' : 0.8618429857937707
        Accuracy for column 'Fwd Pkt Len Max': 0.8614319037876905
        Accuracy for column 'Fwd Pkt Len Min': 0.7262337240205732
        Training and Obtaining accuracy for a group of features together
In [ ]: # Select the 6th column as the feature and the last column as the label
        # example:
        # from 4th col to 16th col
        bCol = 4
        eCol = 10
        X = df.iloc[:, bCol-1:eCol].values
        \# X = df.iloc[:, 3:4].values
       y = df.iloc[:, -1].values
        # Split the data into training and testing sets
        X_train, X_test, y_train, y_test = train_test_split(
           X, y, test_size=0.2, random_state=42)
        # Train the decision tree classifier
        clf = DecisionTreeClassifier()
        clf.fit(X_train, y_train)
        # Test the classifier
        accuracy = clf.score(X_test, y_test)
        selected_cols = df.iloc[:,bCol-1:eCol]
        print("Accuracy for the following features combined", selected_cols.columns.tolist(), "is: ", accuracy)
        Accuracy for the following features combined ['Flow Duration', 'Tot Fwd Pkts', 'TotLen Fwd Pkts', 'TotLen Bwd Pkts', 'Fwd Pkt Len Max', 'Fwd Pkt Len Min'] is: 0.9985468729087398
        Obtaining group accuracy with Feature names as Input
In [ ]: # specify column names to select
        selected_cols = ['Fwd Pkts/s', 'Bwd Pkts/s']
        # select columns by name using loc
        X = df.loc[:, selected_cols].values
        y = df.iloc[:, -1].values
        # Split the data into training and testing sets
        X_train, X_test, y_train, y_test = train_test_split(
           X, y, test_size=0.2, random_state=42)
        # Train the decision tree classifier
        clf = DecisionTreeClassifier()
        clf.fit(X_train, y_train)
        # Test the classifier
        accuracy = clf.score(X_test, y_test)
        print("Accuracy for the following features combined", selected_cols, "is: ", accuracy)
        Accuracy for the following features combined ['Fwd Pkts/s', 'Bwd Pkts/s'] is: 0.994632033804325
        Obtaining individual accuracy with Feature names as Input
In [ ]: # Define the list of column names to iterate over
        col_names = ['Fwd Pkts/s', 'Bwd Pkts/s']
        for col_name in col_names:
               X = df[[col_name]].values
               y = df.iloc[:, -1].values
               # Split the data into training and testing sets
               X_train, X_test, y_train, y_test = train_test_split(
                   X, y, test_size=0.2, random_state=42)
               # Train the decision tree classifier
               clf = DecisionTreeClassifier()
               clf.fit(X_train, y_train)
               # Test the classifier
               accuracy = clf.score(X_test, y_test)
               print(f"Accuracy for column '{col_name}' : {accuracy}")
               print(f"'{col_name}' isn't a feature column")
        Accuracy for column 'Fwd Pkts/s' : 0.9521997667348616
        Accuracy for column 'Bwd Pkts/s' : 0.848635781342613
        Obtaining individual accuracy with Feature numbers as Input
In [ ]: cols = [36,37] # Example list of column indices to test
        for start_col in cols:
           start_col-=1
            end_col = start_col + 1
               X = df.iloc[:, start_col:end_col].values
               y = df.iloc[:, -1].values
               # Split the data into training and testing sets
               X_train, X_test, y_train, y_test = train_test_split(
                   X, y, test_size=0.2, random_state=42)
               # Train the decision tree classifier
               clf = DecisionTreeClassifier()
               clf.fit(X_train, y_train)
               # Test the classifier
               accuracy = clf.score(X_test, y_test)
               print(f"Accuracy for column '{columns[end_col-1]}' : {accuracy}")
               print(f"'{columns[end_col-1]}' isn't a feature column")
        Accuracy for column 'Fwd Pkts/s' : 0.9521997667348616
        Accuracy for column 'Bwd Pkts/s' : 0.848635781342613
        Obtaining group accuracy with Feature numbers as Input
In [ ]: # specify column indexes to select
        selected\_cols\_idx = [36, 37]
        selected_cols_idx = [x - 1 for x in selected_cols_idx]
        # select columns by index using iloc
        X = df.iloc[:, selected_cols_idx].values
        y = df.iloc[:, -1].values
        # Split the data into training and testing sets
        X_train, X_test, y_train, y_test = train_test_split(
           X, y, test_size=0.2, random_state=42)
        # Train the decision tree classifier
        clf = DecisionTreeClassifier()
        clf.fit(X_train, y_train)
        # Test the classifier
        accuracy = clf.score(X_test, y_test)
        # Get the names of the selected columns
```

selected_cols = list(df.columns[selected_cols_idx])

print("Accuracy for the following features combined", selected_cols, "is:", accuracy)

Accuracy for the following features combined ['Fwd Pkts/s', 'Bwd Pkts/s'] is: 0.9946463738743045