Importing Required Libraries

```
import pandas as pd
from sklearn.tree import DecisionTreeClassifier # Import Decision Tree Classifier
from sklearn.model_selection import train_test_split # Import train_test_split function
from sklearn import metrics #Import scikit-learn metrics module for accuracy calculation
```

Loading Data

Let's first load the required 'Pima Indian Diabetes dataset' using pandas read CSV function.

```
# load dataset
pima = pd.read_csv('diabetes.csv')
```

```
pima.head()
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigree
0	6	148	72	35	0	33.6	
1	1	85	66	29	0	26.6	
2	8	183	64	0	0	23.3	
3	1	89	66	23	94	28.1	
4	0	137	40	35	168	43.1	

```
pima.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Pregnancies               768 non-null    int64  
 1   Glucose                   768 non-null    int64  
 2   BloodPressure             768 non-null    int64  
 3   SkinThickness             768 non-null    int64  
 4   Insulin                   768 non-null    int64  
 5   BMI                       768 non-null    float64
 6   DiabetesPedigreeFunction  768 non-null    float64
 7   Age                       768 non-null    int64  
 8   Outcome                   768 non-null    int64  
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

```
pima.describe()
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000

Feature Selection

Here, you need to divide given columns into two types of variables dependent(or target variable) and independent variable(or feature variables).

```
#split dataset in features and target variable
feature_cols = ['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness','Insulin','BMI','DiabetesPedigreeFunction','Age']
```

```
X = pima[feature_cols] # Features
y = pima.Outcome # Target variable
```

Splitting Data

To understand model performance, dividing the dataset into a training set and a test set is a good strategy.

Let's split the dataset by using function train_test_split(). Ratio is 80:20

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2) # 80% training and 20% test
```

Building Decision Tree Model

Let's create a Decision Tree Model using Scikit-learn.

```
# Create Decision Tree classifer object
clf = DecisionTreeClassifier(criterion='entropy')

# Train Decision Tree Classifer
clf = clf.fit(X_train,y_train)

#Predict the response for test dataset
y_pred = clf.predict(X_test)
```

Evaluating Model

Let's estimate, how accurately the classifier or model can predict.

```
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

```
Accuracy: 0.6883116883116883
```

Plotting the Tree

```
clf = clf.fit(X,y)
```

```
from sklearn import tree
tree.plot_tree(clf)
```

```
[Text(0.43926299778761063, 0.9705882352941176, 'X[1] <= 127.5\nentropy = 0.933\nsamples = 768\nvalue = [500, 268]'),
 Text(0.160329092920354, 0.9117647058823529, 'X[7] <= 28.5\nentropy = 0.709\nsamples = 485\nvalue = [391, 94]'),
 Text(0.05641592920353982, 0.8529411764705882, 'X[5] <= 30.95\nentropy = 0.419\nsamples = 271\nvalue = [248, 23]'),
 Text(0.02654867256637168, 0.7941176470588235, 'X[0] <= 7.5\nentropy = 0.102\nsamples = 151\nvalue = [149, 2]'),
 Text(0.017699115044247787, 0.7352941176470589, 'X[6] <= 0.672\nentropy = 0.058\nsamples = 150\nvalue = [149, 1]'),
 Text(0.008849557522123894, 0.6764705882352942, 'entropy = 0.0\nsamples = 131\nvalue = [131, 0]'),
 Text(0.02654867256637168, 0.6764705882352942, 'X[6] <= 0.686\nentropy = 0.297\nsamples = 19\nvalue = [18, 1]'),
 Text(0.017699115044247787, 0.6176470588235294, 'entropy = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.035398230088495575, 0.6176470588235294, 'entropy = 0.0\nsamples = 18\nvalue = [18, 0]'),
 Text(0.035398230088495575, 0.7352941176470589, 'entropy = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.08628318584070796, 0.7941176470588235, 'X[2] <= 37.0\nentropy = 0.669\nsamples = 120\nvalue = [99, 21]'),
 Text(0.07743362831858407, 0.7352941176470589, 'entropy = 0.0\nsamples = 2\nvalue = [0, 2]'),
 Text(0.09513274336283185, 0.7352941176470589, 'X[6] <= 0.501\nentropy = 0.637\nsamples = 118\nvalue = [99, 19]'),
 Text(0.061946902654867256, 0.6764705882352942, 'X[5] <= 45.35\nentropy = 0.456\nsamples = 73\nvalue = [66, 7]'),
 Text(0.05309734513274336, 0.6176470588235294, 'X[6] <= 0.279\nentropy = 0.367\nsamples = 71\nvalue = [66, 5]'),
 Text(0.04424778761061947, 0.5588235294117647, 'X[4] <= 89.0\nentropy = 0.562\nsamples = 38\nvalue = [33, 5]'),
 Text(0.035398230088495575, 0.5, 'X[2] <= 55.0\nentropy = 0.755\nsamples = 23\nvalue = [18, 5]'),
 Text(0.02654867256637168, 0.4411764705882353, 'entropy = 0.0\nsamples = 2\nvalue = [0, 2]'),
 Text(0.04424778761061947, 0.4411764705882353, 'X[6] <= 0.273\nentropy = 0.592\nsamples = 21\nvalue = [18, 3]'),
 Text(0.035398230088495575, 0.38235294117647056, 'X[3] <= 30.5\nentropy = 0.469\nsamples = 20\nvalue = [18, 2]'),
 Text(0.02654867256637168, 0.3235294117647059, 'X[3] <= 29.0\nentropy = 0.684\nsamples = 11\nvalue = [9, 2]'),
 Text(0.017699115044247787, 0.2647058823529412, 'X[2] <= 81.5\nentropy = 0.469\nsamples = 10\nvalue = [9, 1]'),
 Text(0.008849557522123894, 0.20588235294117646, 'entropy = 0.0\nsamples = 8\nvalue = [8, 0]'),
 Text(0.02654867256637168, 0.20588235294117646, 'X[0] <= 2.0\nentropy = 1.0\nsamples = 2\nvalue = [1, 1]'),
 Text(0.017699115044247787, 0.14705882352941177, 'entropy = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.035398230088495575, 0.14705882352941177, 'entropy = 0.0\nsamples = 1\nvalue = [1, 0]'),
 Text(0.035398230088495575, 0.2647058823529412, 'entropy = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.04424778761061947, 0.3235294117647059, 'entropy = 0.0\nsamples = 9\nvalue = [9, 0]'),
 Text(0.05309734513274336, 0.38235294117647056, 'entropy = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.05309734513274336, 0.5, 'entropy = 0.0\nsamples = 15\nvalue = [15, 0]'),
 Text(0.061946902654867256, 0.5588235294117647, 'entropy = 0.0\nsamples = 33\nvalue = [33, 0]'),
 Text(0.07079646017699115, 0.6176470588235294, 'entropy = 0.0\nsamples = 2\nvalue = [0, 2]'),
 Text(0.12831858407079647, 0.6764705882352942, 'X[5] <= 38.65\nentropy = 0.837\nsamples = 45\nvalue = [33, 12]'),
 Text(0.09734513274336283, 0.6176470588235294, 'X[3] <= 39.5\nentropy = 0.928\nsamples = 32\nvalue = [21, 11]'),
 Text(0.07964601769911504, 0.5588235294117647, 'X[5] <= 32.2\nentropy = 0.826\nsamples = 27\nvalue = [20, 7]'),
 Text(0.07079646017699115, 0.5, 'entropy = 0.0\nsamples = 5\nvalue = [5, 0]'),
 Text(0.08849557522123894, 0.5, 'X[2] <= 61.0\nentropy = 0.902\nsamples = 22\nvalue = [15, 7]'),
 Text(0.07964601769911504, 0.4411764705882353, 'entropy = 0.0\nsamples = 4\nvalue = [4, 0]'),
 Text(0.09734513274336283, 0.4411764705882353, 'X[5] <= 34.3\nentropy = 0.964\nsamples = 18\nvalue = [11, 7]'),
 Text(0.07964601769911504, 0.38235294117647056, 'X[5] <= 32.7\nentropy = 0.954\nsamples = 8\nvalue = [3, 5]'),
 Text(0.07079646017699115, 0.3235294117647059, 'X[1] <= 117.5\nentropy = 0.811\nsamples = 4\nvalue = [3, 1]'),
 Text(0.061946902654867256, 0.2647058823529412, 'entropy = 0.0\nsamples = 3\nvalue = [3, 0]'),
 Text(0.07964601769911504, 0.2647058823529412, 'entropy = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.08849557522123894, 0.3235294117647059, 'entropy = 0.0\nsamples = 4\nvalue = [0, 4]'),
 Text(0.11504424778761062, 0.38235294117647056, 'X[5] <= 36.45\nentropy = 0.722\nsamples = 10\nvalue = [8, 2]'),
 Text(0.10619469026548672, 0.3235294117647059, 'entropy = 0.0\nsamples = 6\nvalue = [6, 0]'),
 Text(0.12389380530973451, 0.3235294117647059, 'X[6] <= 0.652\nentropy = 1.0\nsamples = 4\nvalue = [2, 2]'),
 Text(0.11504424778761062, 0.2647058823529412, 'entropy = 0.0\nsamples = 2\nvalue = [2, 0]'),
 Text(0.13274336283185842, 0.2647058823529412, 'entropy = 0.0\nsamples = 2\nvalue = [0, 2]'),
 Text(0.11504424778761062, 0.5588235294117647, 'X[4] <= 143.5\nentropy = 0.722\nsamples = 5\nvalue = [1, 4]'),
 Text(0.10619469026548672, 0.5, 'entropy = 0.0\nsamples = 3\nvalue = [0, 3]'),
 Text(0.12389380530973451, 0.5, 'X[0] <= 1.5\nentropy = 1.0\nsamples = 2\nvalue = [1, 1]'),
 Text(0.11504424778761062, 0.4411764705882353, 'entropy = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.13274336283185842, 0.4411764705882353, 'entropy = 0.0\nsamples = 1\nvalue = [1, 0]'),
 Text(0.1592920353982301, 0.6176470588235294, 'X[7] <= 21.5\nentropy = 0.391\nsamples = 13\nvalue = [12, 1]'),
 Text(0.1504424778761062, 0.5588235294117647, 'X[1] <= 104.0\nentropy = 1.0\nsamples = 2\nvalue = [1, 1]'),
 Text(0.1415929203539823, 0.5, 'entropy = 0.0\nsamples = 1\nvalue = [1, 0]'),
 Text(0.1592920353982301, 0.5, 'entropy = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.168141592920354, 0.5588235294117647, 'entropy = 0.0\nsamples = 11\nvalue = [11, 0]'),
 Text(0.26424225663716816, 0.8529411764705882, 'X[5] <= 26.35\nentropy = 0.917\nsamples = 214\nvalue = [143, 71]'),
 Text(0.20353982300884957, 0.7941176470588235, 'X[5] <= 9.65\nentropy = 0.281\nsamples = 41\nvalue = [39, 2]'),
 Text(0.19469026548672566, 0.7352941176470589, 'entropy = 0.0\nsamples = 2\nvalue = [0, 2]'),
 Text(0.21238938053097345, 0.7352941176470589, 'entropy = 0.0\nsamples = 39\nvalue = [39, 0]'),
 Text(0.3249446902654867, 0.7941176470588235, 'X[1] <= 99.5\nentropy = 0.97\nsamples = 173\nvalue = [104, 69]'),
 Text(0.23008849557522124, 0.7352941176470589, 'X[1] <= 28.5\nentropy = 0.684\nsamples = 55\nvalue = [45, 10]'),
 Text(0.22123893805309736, 0.6764705882352942, 'entropy = 0.0\nsamples = 2\nvalue = [0, 2]'),
 Text(0.23893805309734514, 0.6764705882352942, 'X[6] <= 0.796\nentropy = 0.612\nsamples = 53\nvalue = [45, 8]'),
 Text(0.21238938053097345, 0.6176470588235294, 'X[1] <= 94.5\nentropy = 0.433\nsamples = 45\nvalue = [41, 4]'),
 Text(0.18584070796460178, 0.5588235294117647, 'X[0] <= 11.5\nentropy = 0.196\nsamples = 33\nvalue = [32, 1]'),
 Text(0.17699115044247787, 0.5, 'entropy = 0.0\nsamples = 30\nvalue = [30, 0]'),
 Text(0.19469026548672566, 0.5, 'X[5] <= 31.25\nentropy = 0.918\nsamples = 3\nvalue = [2, 1]'),
 Text(0.18584070796460178, 0.4411764705882353, 'entropy = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.20353982300884957, 0.4411764705882353, 'entropy = 0.0\nsamples = 2\nvalue = [2, 0]'),
 Text(0.23893805309734514, 0.5588235294117647, 'X[1] <= 97.5\nentropy = 0.811\nsamples = 12\nvalue = [9, 3]'),
 Text(0.23008849557522124, 0.5, 'X[5] <= 36.2\nentropy = 1.0\nsamples = 6\nvalue = [3, 3]'),
 Text(0.22123893805309736, 0.4411764705882353, 'entropy = 0.0\nsamples = 3\nvalue = [0, 3]'),
 Text(0.23893805309734514, 0.4411764705882353, 'entropy = 0.0\nsamples = 3\nvalue = [3, 0]'),
 Text(0.24778761061946902, 0.5, 'entropy = 0.0\nsamples = 6\nvalue = [6, 0]'),
 Text(0.26548672566371684, 0.6176470588235294, 'X[0] <= 3.0\nentropy = 1.0\nsamples = 8\nvalue = [4, 4]'),
 Text(0.25663716814159293, 0.5588235294117647, 'entropy = 0.0\nsamples = 3\nvalue = [3, 0]'),
 Text(0.2743362831858407, 0.5588235294117647, 'X[1] <= 94.0\nentropy = 0.722\nsamples = 5\nvalue = [1, 4]'),
 Text(0.26548672566371684, 0.5, 'entropy = 0.0\nsamples = 3\nvalue = [0, 3]'),
 Text(0.2831858407079646, 0.5, 'X[7] <= 37.5\nentropy = 1.0\nsamples = 2\nvalue = [1, 1]'),
 Text(0.2743362831858407, 0.4411764705882353, 'entropy = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.2920353982300885, 0.4411764705882353, 'entropy = 0.0\nsamples = 1\nvalue = [1, 0]'),
 Text(0.4198008849557522, 0.7352941176470589, 'X[6] <= 0.561\nentropy = 1.0\nsamples = 118\nvalue = [59, 59]'),
 Text(0.3528761061946903, 0.6764705882352942, 'X[7] <= 54.5\nentropy = 0.974\nsamples = 84\nvalue = [50, 34]'),
 Text(0.3440265486725664, 0.6176470588235294, 'X[5] <= 27.55\nentropy = 0.99\nsamples = 77\nvalue = [43, 34]'),
 Text(0.33517699115044247, 0.5588235294117647, 'entropy = 0.0\nsamples = 4\nvalue = [0, 4]'),
 Text(0.3528761061946903, 0.5588235294117647, 'X[6] <= 0.2\nentropy = 0.977\nsamples = 73\nvalue = [43, 30]'),
```