

Lab 9

In lab task 1:

```
# Load 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
from sklearn import metrics #Import scikit-learn metrics module for accuracy ca
```

Loading Data:

```
# Load Dataset
col_names = ['pregnant', 'glucose', 'bp', 'skin', 'insulin', 'bmi', 'pedigree', 'age', 'Label' ]
# load dataset
pima = pd.read_csv("diabetes.csv", header=None, names=col_names)
pima_df= pima.head()
print (pima_df)
```

Output:

	pregnant	glucose	bp	skin	insulin	bmi	pedigree	age	Label
0	pregnant	glucose	bp	skin	insulin	bmi	pedigree	age	label
1	6	148	72	35	0	33.6	0.627	50	1
2	1	85	66	29	0	26.6	0.351	31	0
3	8	183	64	0	0	23.3	0.672	32	1
4	1	89	66	23	94	28.1	0.167	21	0

Feature Selection:

```
# Feature Selection
# split dataset into features and target variable
feature_cols = ['pregnant', 'insulin', 'bmi', 'age', 'glucose', 'bp', 'pedigree']
X = pima[feature_cols] # Features
```

```
y = pima.Label # Target variable
```

Output:

	pregnant	glucose	bp	skin	insulin	bmi	pedigree	age	Label
0	pregnant	glucose	bp	skin	insulin	bmi	pedigree	age	label
1	6	148	72	35	0	33.6	0.627	50	1
2	1	85	66	29	0	26.6	0.351	31	0
3	8	183	64	0	0	23.3	0.672	32	1
4	1	89	66	23	94	28.1	0.167	21	0

Splitting Data:

```
# Splitting Data
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=1)
```

Building Decision Tree Model:

```
# -----Building Decision Tree Model-----
-----
clf = DecisionTreeClassifier()
clf = clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
```

Evaluating the Model:

```
# -----Evaluating the Model-----
-----
print("-----\n")
```

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

Output:

	pregnant	glucose	bp	skin	insulin	bmi	pedigree	age	Label
0	pregnant	glucose	bp	skin	insulin	bmi	pedigree	age	label
1	6	148	72	35	0	33.6	0.627	50	1
2	1	85	66	29	0	26.6	0.351	31	0
3	8	183	64	0	0	23.3	0.672	32	1
4	1	89	66	23	94	28.1	0.167	21	0

```
-----
Accuracy: 0.7272727272727273
-----
```

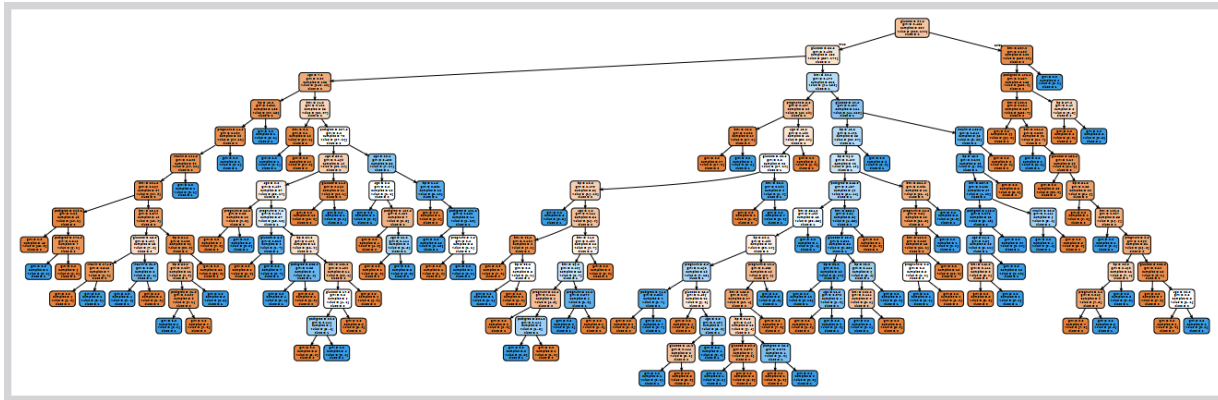
Visualizing Decision Trees:

```
# -----Visualizing Decision Trees -----
-----

from sklearn.tree import export_graphviz
import graphviz

dot_data = export_graphviz(clf, out_file=None,
                           feature_names=X_train.columns,
                           class_names= [str(x) for x in
clf.classes_],
                           filled=True, rounded=True,
                           special_characters=True)

graph = graphviz.Source(dot_data)
graph.render("decision_tree")
graph.view("decision_tree")
```

Output:**Optimizing Decision Tree Performance:**

```
# -----Optimizing Decision Tree Performance -----
-----
clf = DecisionTreeClassifier(criterion="entropy", max_depth=3)
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("-----\n")
print("Accuracy:", metrics.accuracy_score(y_test, y_pred))
print("-----\n")
-----\n")
```

Output:

Accuracy: 0.7445887445887446

