In [1]:
```
#In this project, we are going to predict if the customer will purchase an iPhone
#gender, age and salary. (100 points)

#Steps for training a model with random forest algorithm.
#• Step 1: Load Data; Split data into training set and test set.(30 points)
#• Step 2: Building the Random Forest Classifier.(40 points)
#• Step 3: Calculate Model Accuracy.(Please show your confusion matrix)(30 points

import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sn

from sklearn.preprocessing import LabelEncoder
from sklearn import tree
from sklearn.datasets import load_digits
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix


df = pd.read_csv('OneDrive\Desktop\iphone_sale.csv')
df.head()

df = df.rename(columns={'Purchased Iphone': 'Purchased'})
df.head()

le_Gender = LabelEncoder()
df['Gender'] = le_Gender.fit_transform(df['Gender'])
df.head()
```

Out[1]:

	Gender	Age	Salary	Purchased
0	1	19	19000	0
1	1	35	20000	0
2	0	26	43000	0
3	0	27	57000	0
4	1	19	76000	0

In [2]:
```
X = df.drop('Purchased', axis='columns')
y = df.Purchased

X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.2)

model = RandomForestClassifier(n_estimators=15, criterion='entropy')
model.fit(X_train, y_train)
```

Out[2]:
```
RandomForestClassifier(criterion='entropy', n_estimators=15)
```

In [3]:
```
model.score(X_test, y_test)
```

Out[3]:
```
1.0
```

In [4]:
```
y_predicted = model.predict(X_test)
y_predicted
```

Out[4]:
```
array([1, 0, 0, 0], dtype=int64)
```

In [5]:
```
cm = confusion_matrix(y_test, y_predicted)
cm
```

Out[5]:
```
array([[3, 0],
       [0, 1]], dtype=int64)
```

In [6]:
```
plt.figure(figsize=(10,7))
sn.heatmap(cm, annot=True)
plt.xlabel('Predicted')
plt.ylabel('Truth')
```

Out[6]:
```
Text(69.0, 0.5, 'Truth')
```