Sedat CAN

Bach Code: LISUM14

Submission Date: 28/10/2022

Submitted to: Data Glacier

Step 1:

I use adult.csv data

The data

```
28, Private,338409, Bachelors,13, Married-civ-spouse, Prof-specialty, Wife, Black, Female,0,0,40, Cuba, <=50K
37, Private,284582, Masters,14, Married-civ-spouse, Exec-managerial, Wife, White, Female,0,0,40, United-States, <=50K
49, Private,160187, 9th,5, Married-spouse-absent, Other-service, Not-in-family, Black, Female,0,0,16, Jamaica, <=50K
52, Self-emp-not-inc,209642, HS-grad,9, Married-civ-spouse, Exec-managerial, Husband, White, Male,0,0,45, United-States, >50K
31, Private,45781, Masters,14, Never-married, Prof-specialty, Not-in-family, White, Female,14084,0,50, United-States, >50K
42, Private,159449, Bachelors,13, Married-civ-spouse, Exec-managerial, Husband, White, Male,5178,0,40, United-States, >50K
37, Private,280464, Some-college,10, Married-civ-spouse, Exec-managerial, Husband, Black, Male,0,0,80, United-States, >50K
30, State-gov,141297, Bachelors,13, Married-civ-spouse, Prof-specialty, Husband, Asian-Pac-Islander, Male,0,0,40, India, >50K
23, Private,122272, Bachelors,13, Never-married, Adm-clerical, Own-child, White, Female,0,0,30, United-States, <=50K
32, Private,205019, Assoc-acdm,12, Never-married, Sales, Not-in-family, Black, Male,0,0,50, United-States, <=50K
40, Private,121772, Assoc-voc,11, Married-civ-spouse, Craft-repair, Husband, Asian-Pac-Islander, Male,0,0,40, ?, >50K
34, Private,245487, 7th-8th,4, Married-civ-spouse, Transport-moving, Husband, Amer-Indian-Eskimo, Male,0,0,45, Mexico, <=50K
25, Self-emp-not-inc,176756, HS-grad,9, Never-married, Farming-fishing, Own-child, White, Male,0,0,35, United-States, <=50K
32, Private,186824, HS-grad,9, Never-married, Machine-op-inspct, Unmarried, White, Male,0,0,40, United-States, <=50K
38, Private,28887, 11th,7, Married-civ-spouse, Sales, Husband, White, Male,0,0,50, United-States, <=50K
43, Self-emp-not-inc,292175, Masters,14, Divorced, Exec-managerial, Unmarried, White, Female,0,0,45, United-States, >50K
40, Private,193524, Doctorate,16, Married-civ-spouse, Prof-specialty, Husband, White, Male,0,0,60, United-States, >50K
54, Private,302146, HS-grad,9, Separated, Other-service, Unmarried, Black, Female,0,0,20, United-States, <=50K
35, Federal-gov,76845, 9th,5, Married-civ-spouse, Farming-fishing, Husband, Black, Male,0,0,40, United-States, <=50K
43, Private,117037, 11th,7, Married-civ-spouse, Transport-moving, Husband, White, Male,0,2042,40, United-States, <=50K
59, Private,109015, HS-grad,9, Divorced, Tech-support, Unmarried, White, Female,0,0,40, United-States, <=50K
56, Local-gov,216851, Bachelors,13, Married-civ-spouse, Tech-support, Husband, White, Male,0,0,40, United-States, >50K
19, Private,168294, HS-grad,9, Never-married, Craft-repair, Own-child, White, Male,0,0,40, United-States, <=50K
54, ?,180211, Some-college,10, Married-civ-spouse, ?, Husband, Asian-Pac-Islander, Male,0,0,60, South, >50K
39, Private,367260, HS-grad,9, Divorced, Exec-managerial, Not-in-family, White, Male,0,0,80, United-States, <=50K
49, Private,193366, HS-grad,9, Married-civ-spouse, Craft-repair, Husband, White, Male,0,0,40, United-States, <=50K
23, Local-gov,190709, Assoc-acdm,12, Never-married, Protective-serv, Not-in-family, White, Male,0,0,52, United-States, <=50K
20, Private,266015, Some-college,10, Never-married, Sales, Own-child, Black, Male,0,0,44, United-States, <=50K
45, Private,386940, Bachelors,13, Divorced, Exec-managerial, Own-child, White, Male,0,1408,40, United-States, <=50K
30, Federal-gov,59951, Some-college,10, Married-civ-spouse, Adm-clerical, Own-child, White, Male,0,0,40, United-States, <=50K
22, State-gov,311512, Some-college,10, Married-civ-spouse, Other-service, Husband, Black, Male,0,0,15, United-States, <=50K
48, Private,242406, 11th,7, Never-married, Machine-op-inspct, Unmarried, White, Male,0,0,40, Puerto-Rico, <=50K
21, Private,197200, Some-college,10, Never-married, Machine-op-inspct, Own-child, White, Male,0,0,40, United-States, <=50K
19, Private,544091, HS-grad,9, Married-AF-spouse, Adm-clerical, Wife, White, Female,0,0,25, United-States, <=50K
31, Private,84154, Some-college,10, Married-civ-spouse, Sales, Husband, White, Male,0,0,38, ?, >50K
48, Self-emp-not-inc,265477, Assoc-acdm,12, Married-civ-spouse, Prof-specialty, Husband, White, Male,0,0,40, United-States, <=50K
31, Private,507875, 9th,5, Married-civ-spouse, Machine-op-inspct, Husband, White, Male,0,0,43, United-States, <=50K
53, Self-emp-not-inc,88506, Bachelors,13, Married-civ-spouse, Prof-specialty, Husband, White, Male,0,0,40, United-States, <=50K
24, Private,172987, Bachelors,13, Married-civ-spouse, Tech-support, Husband, White, Male,0,0,50, United-States, <=50K
49, Private,94638, HS-grad,9, Separated, Adm-clerical, Unmarried, White, Female,0,0,40, United-States, <=50K
25, Private,289980, HS-grad,9, Never-married, Handlers-cleaners, Not-in-family, White, Male,0,0,35, United-States, <=50K
57, Federal-gov,337895, Bachelors,13, Married-civ-spouse, Prof-specialty, Husband, Black, Male,0,0,40, United-States, >50K
53, Private,144361, HS-grad,9, Married-civ-spouse, Machine-op-inspct, Husband, White, Male,0,0,38, United-States, <=50K
44, Private,128354, Masters,14, Divorced, Exec-managerial, Unmarried, White, Female,0,0,40, United-States, <=50K
41, State-gov,101603, Assoc-voc,11, Married-civ-spouse, Craft-repair, Husband, White, Male,0,0,40, United-States, <=50K
29, Private,271466, Assoc-voc,11, Never-married, Prof-specialty, Not-in-family, White, Male,0,0,43, United-States, <=50K
25, Private,32275, Some-college,10, Married-civ-spouse, Exec-managerial, Wife, Other, Female,0,0,40, United-States, <=50K
18, Private,226956, HS-grad,9, Never-married, Other-service, Own-child, White, Female,0,0,30, ?, <=50K
```

Step 2:

The model uses Decision Tree Classifier which predict whether real income more 50K or not. The model form a classidier and record it in modevl.pkl file

```
Users > elifh > OneDrive > Masaüstü > mag.py > ...
import pandas
import numpy
from sklearn import preprocessing
import pickle

df = pandas.read_csv('c://Users//elifh//OneDrive//Masaüstü//adult.csv')
df.head()

df = df.drop(['fnlwgt', 'educational-num'], axis=1)

col_names = df.columns

for c in col_names:
    df = df.replace("?", numpy.NaN)
df = df.apply(lambda x: x.fillna(x.value_counts().index[0]))


df.replace(['Divorced', 'Married-AF-spouse',
            'Married-civ-spouse', 'Married-spouse-absent',
            'Never-married', 'Separated', 'Widowed'],
           ['divorced', 'married', 'married', 'married',
            'not married', 'not married', 'not married'], inplace=True)

category_col = ['workclass', 'race', 'education', 'marital-status', 'occupation',
                'relationship', 'gender', 'native-country', 'income']
labelEncoder = preprocessing.LabelEncoder()

mapping_dict = {}
for col in category_col:
    df[col] = labelEncoder.fit_transform(df[col])

    le_name_mapping = dict(zip(labelEncoder.classes_,
                               labelEncoder.transform(labelEncoder.classes_)))

    mapping_dict[col] = le_name_mapping
print(mapping_dict)

from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score

X = df.values[:, 0:12]
Y = df.values[:, 12]


X_train, X_test, y_train, y_test = train_test_split(
            X, Y, test_size = 0.3, random_state = 100)

dt_clf_gini = DecisionTreeClassifier(criterion = "gini",
                                     random_state = 100,
                                     max_depth = 5,
                                     min_samples_leaf = 5)

dt_clf_gini.fit(X_train, y_train)
y_pred_gini = dt_clf_gini.predict(X_test)

print ("Decision Tree using Gini Index\nAccuracy is ",
            accuracy_score(y_test, y_pred_gini)*100 )

pickle.dump(dt_clf_gini, open('model.pkl','wb'))
```

UNLAR ÇIKIŞ HATA AYIKLAMA KONSOLU TERMİNAL JUPYTER

Step 3:

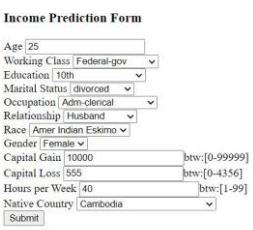
I deploy the model with help of flask.

```python
import numpy as np
from flask import Flask, request, jsonify, render_template
import pickle


app = Flask(__name__)

@app.route('/')
def home():
    return render_template('vis.html')

# prediction function
def ValuePredictor(to_predict_list):
    to_predict = np.array(to_predict_list).reshape(1, 12)
    loaded_model = pickle.load(open("model.pkl", "rb"))
    result = loaded_model.predict(to_predict)
    return result[0]

@app.route('/result', methods = ['POST'])
def result():
    if request.method == 'POST':
        to_predict_list = request.form.to_dict()
        to_predict_list = list(to_predict_list.values())
        to_predict_list = list(map(int, to_predict_list))
        result = ValuePredictor(to_predict_list)
        if int(result)== 1:
            prediction ='Income more than 50K'
        else:
            prediction ='Income less that 50K'
        return render_template("result.html", prediction = prediction)


if __name__ == "__main__":
    app.run(debug=True)
```

Step 4:

I create vis.html file to webpage where users can enter data

```html
<body>
    <h3>Income Prediction Form</h3>

<div>
<form action="/result" method="POST">
    <label for="age">Age</label>
    <input type="text" id="age" name="age">
    <br>
    <label for="w_class">Working Class</label>
    <select id="w_class" name="w_class">
    <option value="0">Federal-gov</option>
    <option value="1">Local-gov</option>
    <option value="2">Never-worked</option>
    <option value="3">Private</option>
    <option value="4">Self-emp-inc</option>
    <option value="5">Self-emp-not-inc</option>
    <option value="6">State-gov</option>
    <option value="7">Without-pay</option>
    </select>
    <br>
    <label for="edu">Education</label>
    <select id="edu" name="edu">
    <option value="0">10th</option>
    <option value="1">11th</option>
    <option value="2">12th</option>
    <option value="3">1st-4th</option>
    <option value="4">5th-6th</option>
    <option value="5">7th-8th</option>
    <option value="6">9th</option>
    <option value="7">Assoc-acdm</option>
    <option value="8">Assoc-voc</option>
    <option value="9">Bachelors</option>
    <option value="10">Doctorate</option>
    <option value="11">HS-grad</option>
    <option value="12">Masters</option>
    <option value="13">Preschool</option>
    <option value="14">Prof-school</option>
    <option value="15">16 - Some-college</option>
    </select>
    <br>
    <label for="martial_stat">Marital Status</lab
    <select id="martial_stat" name="martial_stat"
```



Step 5:

The model generates prediction by using ‘model.pkl’ file and show it with ‘result.html’ file

```
<!doctype html>
<html>
    <head>

    </head>
    <body>
    <link rel="stylesheet"  href="stylesheet.css">
        <h1> {{ prediction }}</h1>
    </body>
</html>
```

ncome more than 50K