Week 5 Assignment

Name: Enias Vontas

Batch Code: LISUM01

Submission Date: 11th of July, 2021

Submitted to:

https://github.com/EniasVontas/Assignments/upload/main/Week5

We will consider the same dataset, as the one used for Week 4, the 'insurance.csv' dataset, which was obtained from kaggle (https://www.kaggle.com/mirichoi0218/insurance/) and has 1338 rows and 7 features: age, gender, BMI, number of children, whether the person is a smoker or not, the region that they live and the insurance cost.

We will try to predict a person's insurance cost based on 5 of the above features, as the region that someone lives will not be considered.

The first five rows of our dataset and the types of features we have:

					smoker	region	charges
Θ	19	female	27.900	Θ	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	Θ	no	northwest	21984.47061
4	32	male	28.880	Θ	no	northwest	3866.85520

```
age int64
sex object
bmi float64
children int64
smoker object
region object
charges float64
dtype: object
```

We observe that the 'bmi' features is a float, as well as the 'charges' feature. The 'sex', 'smoker' and 'region' features are objects. As we said, for simplicity, the feature 'region' will not be considered so we create dummy variables for the other two features. Our new dataset with the independent variables and our dependent ('charges') variable are:

	age	bmi	children	sex_male	smoker_yes
Θ	19	27.900	Θ	Θ	- 1
1	18	33.770	1	1	Θ
2	28	33.000	3	1	Θ
3	33	22.705	Θ	1	Θ
4	32	28.880	Θ	1	Θ

```
charges
0 16884.92400
1 1725.55230
2 4449.46200
3 21984.47061
4 3866.85520
```

We would like to create our linear regression model and then use 'pickle' library to save it on our disk. So we create a .py file where we import the necessary libraries for our regression and the pickling that will follow. We load the data and create the data frame 'X' for our independent variables and the data frame 'Y' for our dependent variable. Then we modify the categorical features of our 'X' data frame to dummy variables. Lastly, we fit the linear regression model.

```
import pandas as pd
import numpy as np
import pickle
from sklearn.linear_model import LinearRegression

dataset = pd.read_csv("C:\Program Files\Git\DataSets\insurance.csv")

X = dataset[["age", "sex", "bmi", "children", "smoker"]]
Y = dataset[["charges"]]

X = pd.get_dummies(data=X, drop_first=True)

model = LinearRegression()
model.fit(X,Y)
```

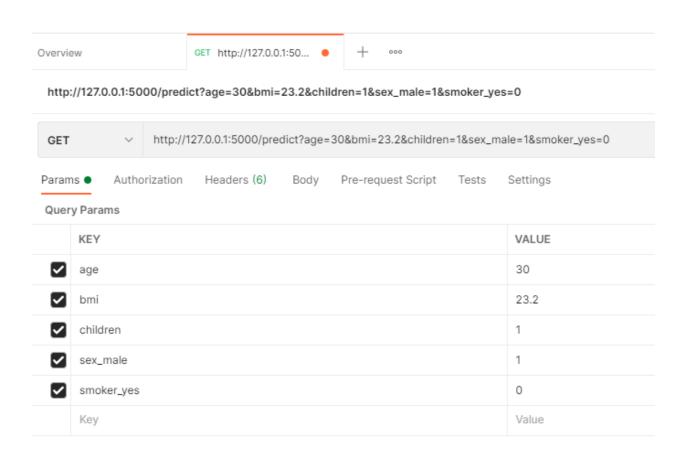
We then pickle the model in order to save it to the disk and then load it again when we run the .py script in the command line.

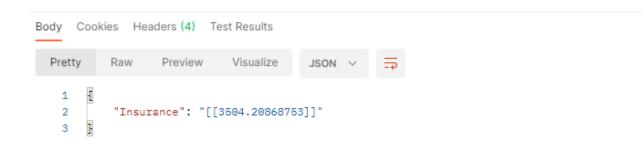
```
pickle.dump(model,open('model.pkl','wb'))
model = pickle.load(open('model.pkl','rb'))
```

Next we create the app.py file in order to initialize the flask app and to render the template of the html file. We load our 'pickled' model and then create two routes, one for the template, and the other for the prediction, where we superimpose floating point numbers, as we can have BMIs with decimal points. Then we predict the insurance cost based on the parameters that we will enter on the Postman application.

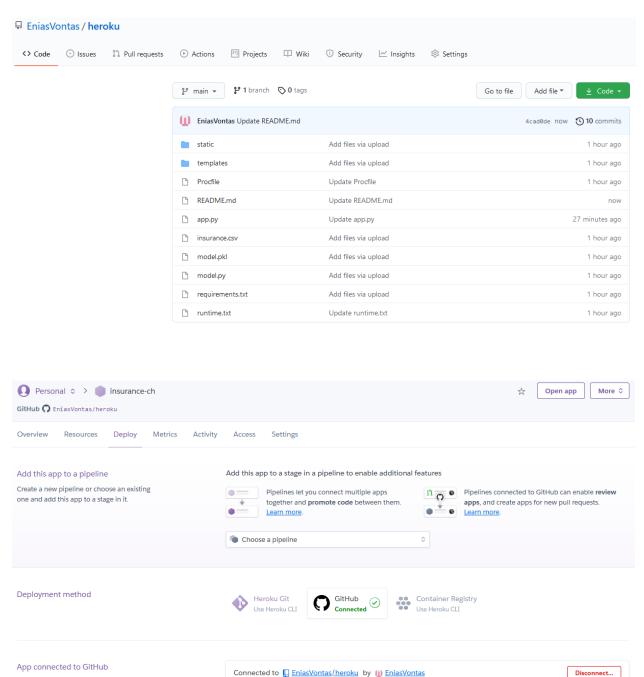
```
import numpy as np
from flask import Flask, request, jsonify
import pickle
import pandas as pd
app = Flask(__name__)
@app.route('/', methods = ["GET", "POST"])
def home():
    if(request.method == "GET"):
        data = "hello world"
        return jsonify({'data':data})
@app.route('/predict')
def predict():
    model = pickle.load(open('model.pkl', 'rb'))
    age = request.args.get('age')
    bmi = request.args.get('bmi')
    children = request.args.get('children')
    sex = request.args.get('sex male')
    smoker = request.args.get('smoker yes')
    test_df = pd.DataFrame({'Age':[age],'BMI':[bmi],
                             "No of Children":[children],
                             "Sex":[sex],"Smoker":[smoker]})
    pred_price = model.predict(test_df)
    return jsonify({'Insurance':str(pred_price)})
if __name__ == "__main__":
app.run(debug=True)
```

And the result from Postman is as shown below, for a male 30 year old individual, with 23.2 BMI, who has 1 child and is a non-smoker. The charges are 3504.21 \$.





Next, we would like to create an application through Heroku, where we can predict the amount of insurance someone will be charged. In order to do that, we create a Heorku account, and connect our GitHub repository (named 'heroku'), with Heorku.

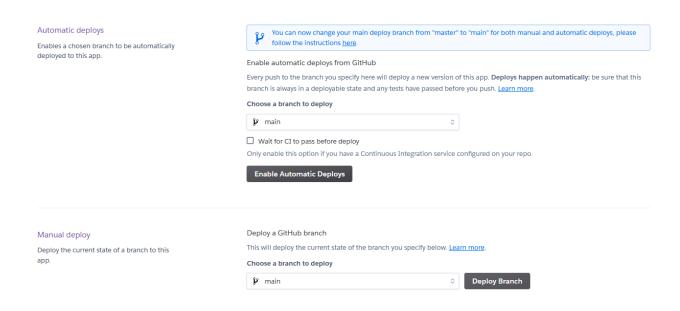


- Releases in the activity feed link to GitHub to view commit diffs

Code diffs, manual and auto deploys are

available for this app.

And then we deploy the main branch of this respository:



The application has been created with url 'https://insurance-ch.herokuapp.com' . We would like to again compute the insurance charges for a 30 year old male individual, with 23.2 BMI, who has 1 child and is a non-smoker.



I would fail to get a result through the Heroku app and unfortunately, I was not able to find the solution in time, before the end of the deadline.