BUILDING A MACHINE LEARNING MODEL

what is a machine learning model?

A machine learning model can be a mathematical representation of a real world processi
steps to build a machine learning model:

- gathering the data
- data preprocessing
- researching the model that will be best for type of data
- training and testing the model
- evaluation

DATA PRE-PROCESSING

➤ **it is a process of cleaning the raw data i.e, the data collected in**

steps to be followed to process our data

- import the libraries:

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

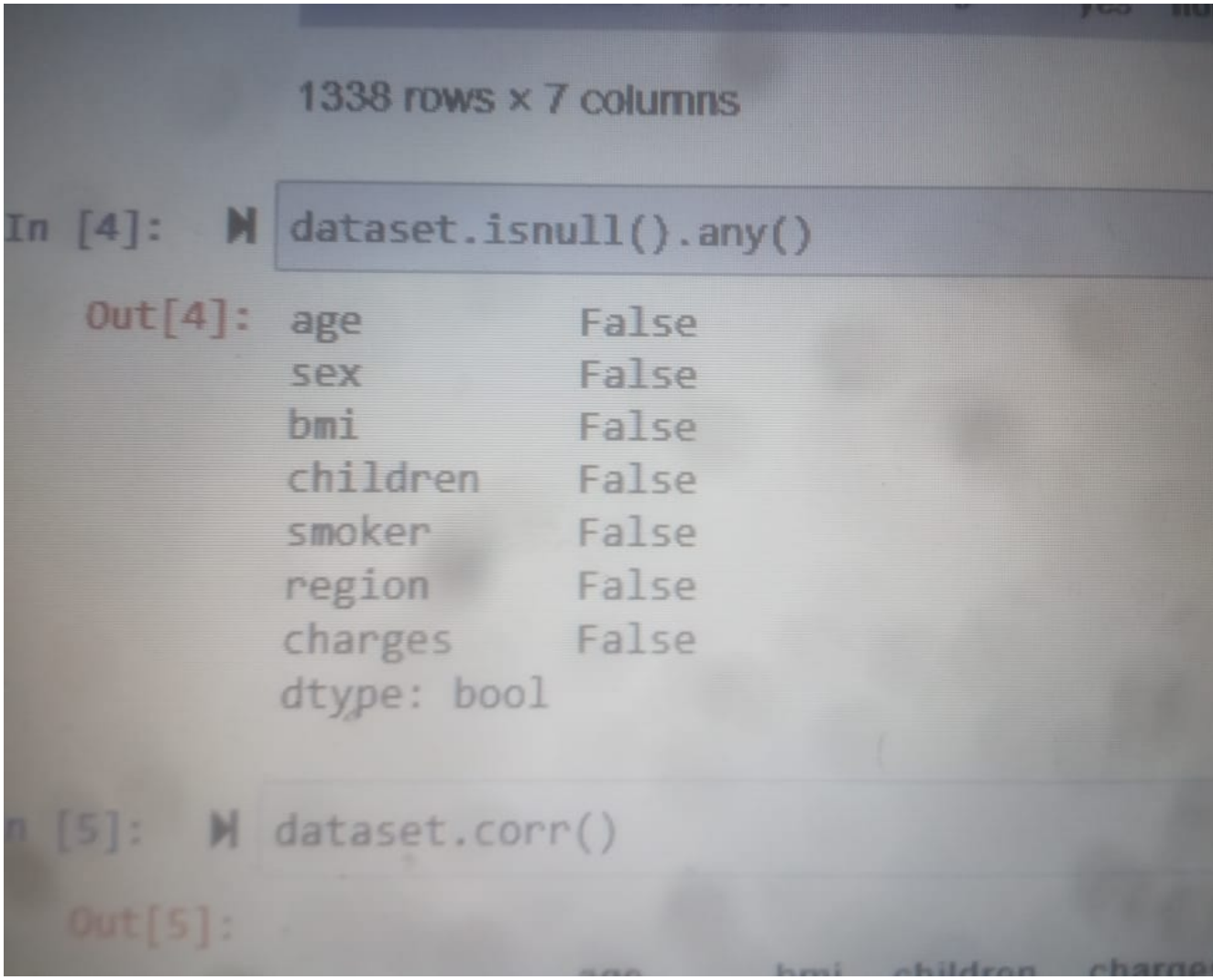
- import the data set:
- WE **WILL NEED TO LOACTE THE DIRECTORY OF CSV FILE AT FIRST AND READ IT USING A METHOD CALLED read__csv which can be found in the libraries called pandas**

```
dataset = pd.read_csv(
r'C:\Users\Hari Chandan\Desktop\Data_Preprocessing\Data.csv
```

- taking care of missing data:

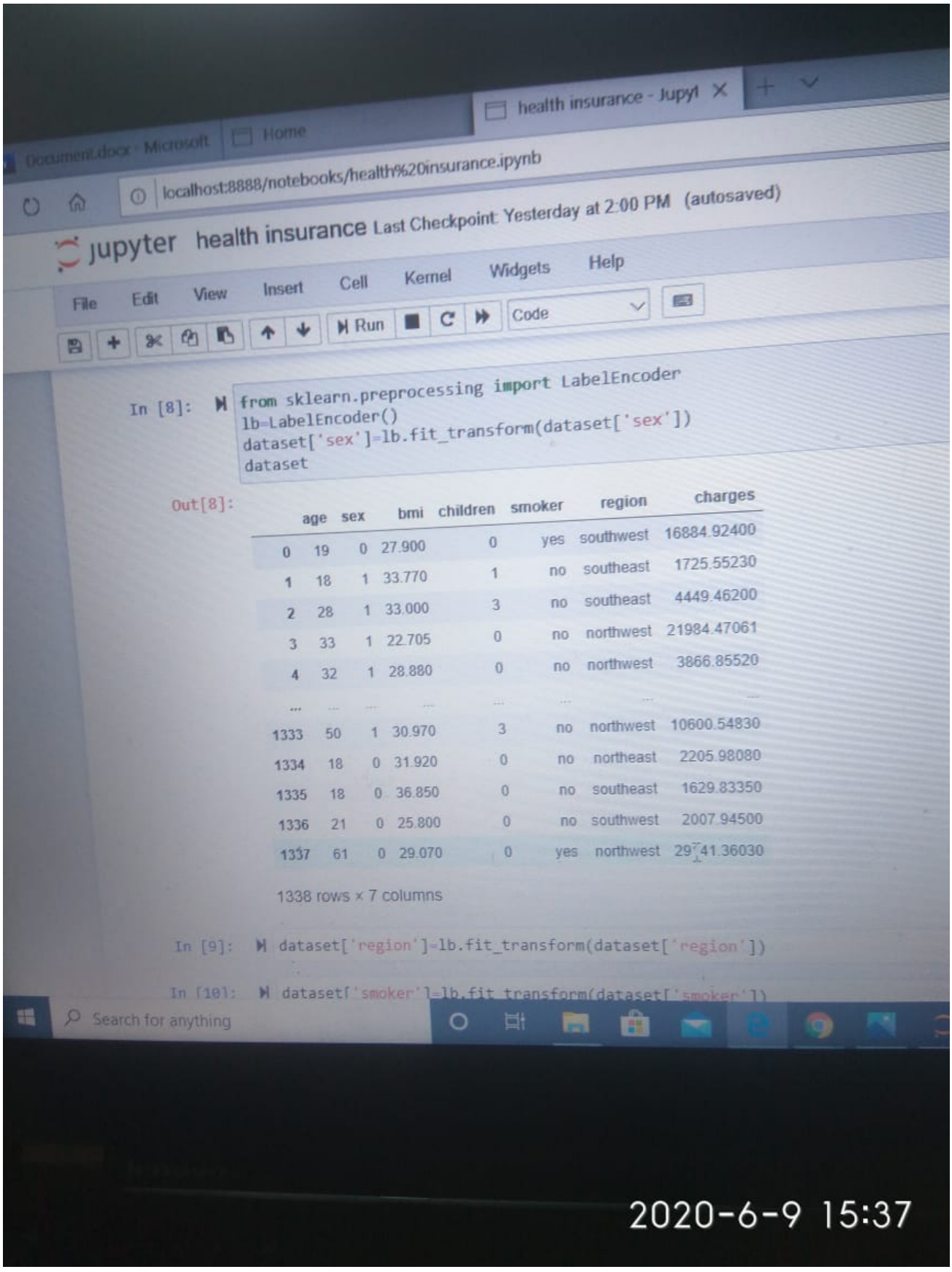
SOME TIMES YOU MAY FIND DATA ARE MISSING IN THE DATA SET.

- WE WILL BE SUING dataset.isnull().any()method to see which column has missing values
- we can replace the missing values by mean,median,mode by using fillna



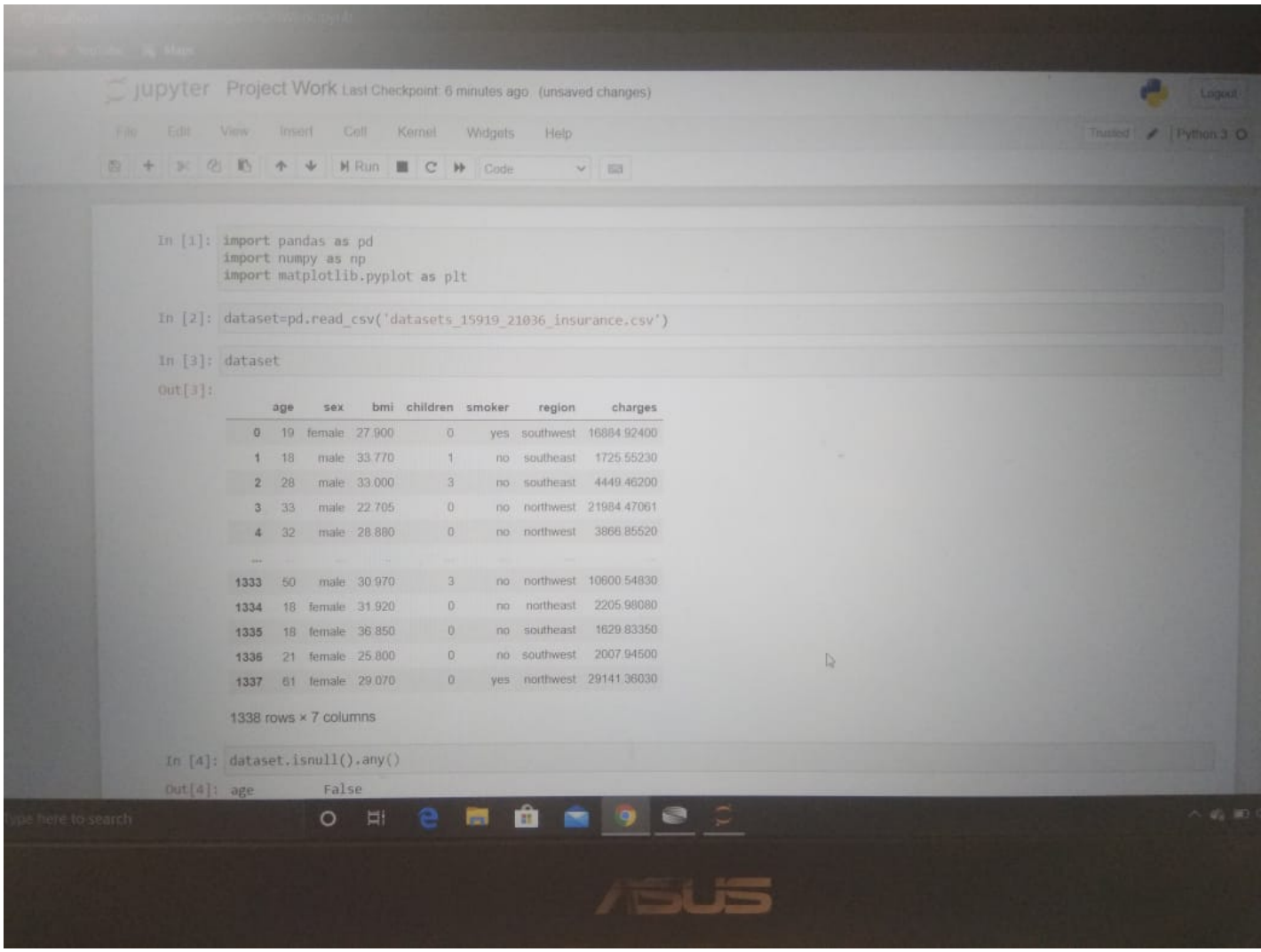
- label encoding:

➤ **some times in dataset we will find textual data like names,countries,states then the machine cannot do mathematical operations or cannot understand textual data.so the textual data will be converted into numerical format called LABEL ENCODING**


Document.docx - Microsoft
Home
health insurance - Jupyt
localhost:8888/notebooks/health%20insurance.ipynb
jupyter health insurance Last Checkpoint: Yesterday at 2:00 PM (autosaved)
File Edit View Insert Cell Kernel Widgets Help
Run
Code
In [8]:
from sklearn.preprocessing import LabelEncoder
lb=LabelEncoder()
dataset['sex']=lb.fit_transform(dataset['sex'])
dataset
Out[8]:
age sex bmi children smoker region charges
0 19 0 27.900 0 yes southwest 16884.92400
1 18 1 33.770 1 no southeast 1725.55230
2 28 1 33.000 3 no southeast 4449.46200
3 33 1 22.705 0 no northwest 21984.47061
4 32 1 28.880 0 no northwest 3866.85520
... ...
1333 50 1 30.970 3 no northwest 10600.54830
1334 18 0 31.920 0 no northeast 2205.98080
1335 18 0 36.850 0 no southeast 1629.83350
1336 21 0 25.800 0 no southwest 2007.94500
1337 61 0 29.070 0 yes northwest 29141.36030
1338 rows × 7 columns
In [9]:
dataset['region']=lb.fit_transform(dataset['region'])
In [10]:
dataset['smoker']=lb.fit_transform(dataset['smoker'])
Search for anything
2020-6-9 15:37

- one hot enoding
- **in the above figure extra three columns are created.Based on the catagories those many columns will be appended to x variable.to accomplish task we will import another library called ONE HOT ENCODING**
- feature scaling:
- THE **FINAL STEP OF DATA PREPROCESSING IS TO APPLY A VERY IMPORTANT FEATURESCALING**
- **IT IS A METHOD USED TO STANDARDIZE THE RANGE OF INDEPENDENT VARIABLES OR FEATURES OF DATA**
- splitting data into train and test
- TO READ **THE COLUMNS,WE WILL USE ILOC OF PANDAS(used to fix indexes for selection)**

```
x = dataset.iloc[:, 0:3].values
y = dataset.iloc[:, 3].values
```



THE DATA IS PRE-PROCESSED BY USING JUPYTOR NOTE BOOK

FLASK FRAME WORK WITH MACHINE LEARNING MODEL

- **in these section we will be building a web application that is to be integrated with model we built**
- **we are using a machine learning model which is built for predicting the health insurance premiums and saved these**

file as insurance.pkl

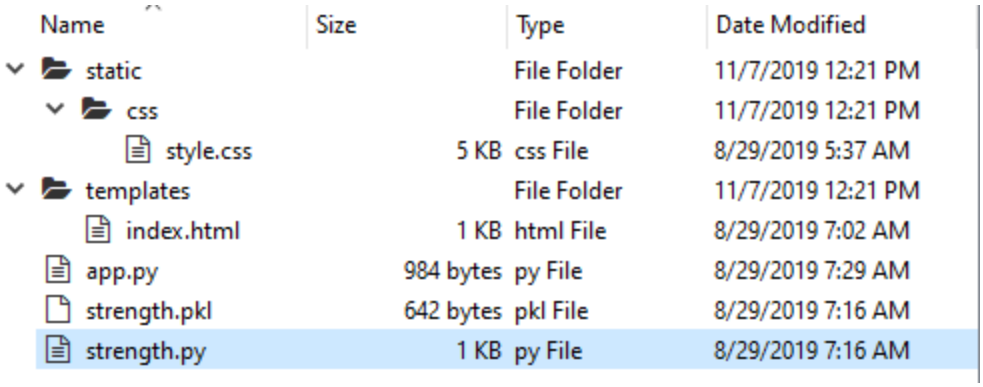
✈To build these you should know basics of html,css boot strap,flask frame work and python

➤TO BUILD A PYTHON CODE

- ⭐ **IMPORTING LIBRARIES**
- ⭐ **ROUTING TO HTML PAGE**
- ⭐ **SHOW CASING PREDICTION ON UI**
- ⭐ **RUN THE APP IN LOCAL BROWSER**

CREATE A PROJECT FOLDER WHICH SHOULD CONTAIN

- AN PYTHON FILL CALLED APP.Y
- YOUR MACHINE LEARNING ALGORITHM FILE
- MODEL FILE
- TEMPLATES FOLDER WHICH SHOULD CONTAIN INDEX.HTML FILE
- STATIC FOLDER WHICH CONTAINS CSS FOLDER WHICH CONTAINS STYLES.CSS



-

IMPORTING LIBRARIES:

IMPORTING FLASK MODULE IN THE PROJECT IS MANDATORY.AN OBJECT OF FLASK CLASS IS OUR WSGI APPLICATION.FLASK CONSTRUCTOR TAKES THE NAME OF CURRENT MODULE(__name__)as ARGUMENT.PICK LIBRARY TO LOAD THE MODEL FILE

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

app=FLASK(__NAME__)
model=pickle.load(open('strength.pkl','rb'))
@app.route('/')
def home():
```

return render__template('index.html')

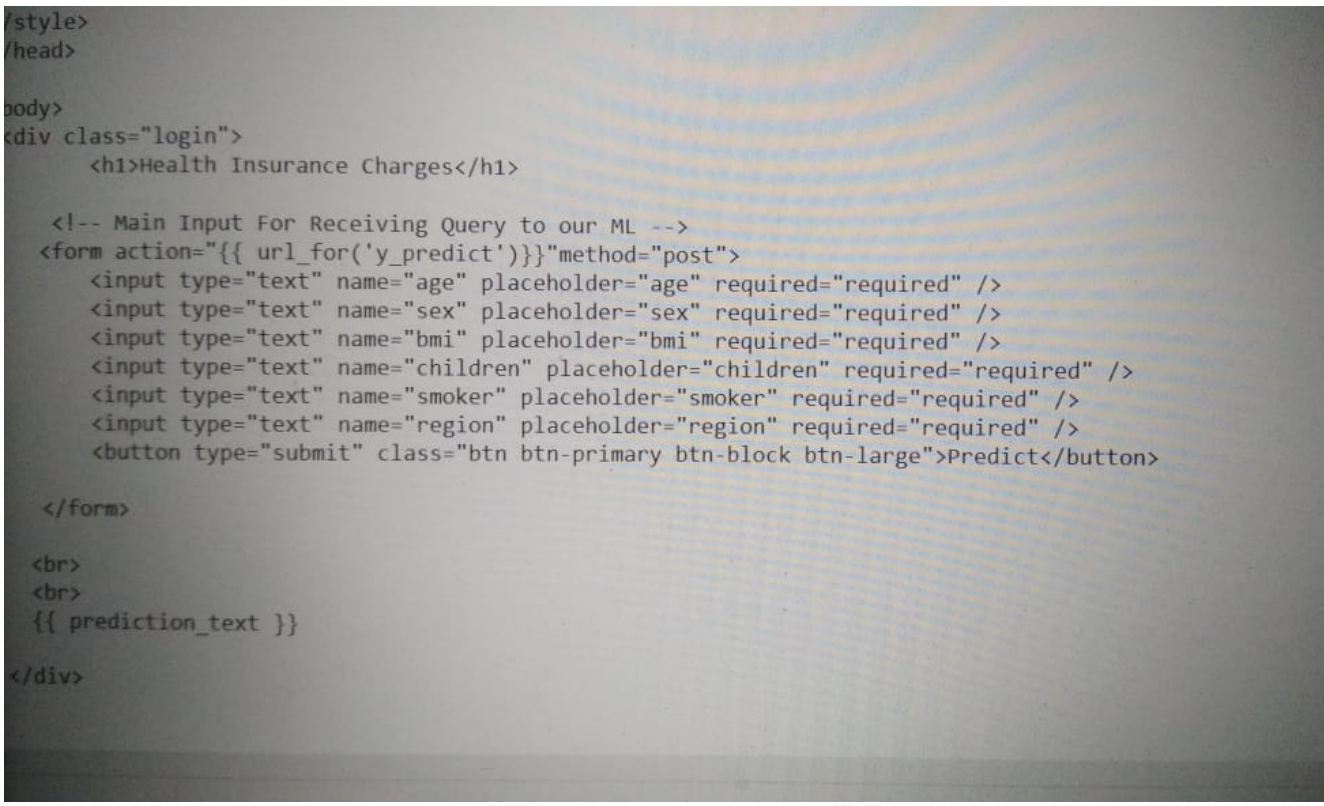
- MAIN FUNCTION

THIS IS USED TO RUN THE APPLICATION IN LOCAL HOST

if__name__=="__main__":

app.run(debug==true)

- **BUILDING AN INDEX.HTML FILE**



```
/style>
/head>

body>
<div class="login">
      <h1>Health Insurance Charges</h1>

   <!-- Main Input For Receiving Query to our ML -->
   <form action="{{ url_for('y_predict')}}"method="post">
      <input type="text" name="age" placeholder="age" required="required" />
      <input type="text" name="sex" placeholder="sex" required="required" />
      <input type="text" name="bmi" placeholder="bmi" required="required" />
      <input type="text" name="children" placeholder="children" required="required" />
      <input type="text" name="smoker" placeholder="smoker" required="required" />
      <input type="text" name="region" placeholder="region" required="required" />
      <button type="submit" class="btn btn-primary btn-block btn-large">Predict</button>

   </form>

  <br>
  <br>
  {{ prediction_text }}

</div>
```

-

YOUR OUTPUT LOOKS LIKE

THE OUTPUT IS LIKE WHATEVER WE ARE GIVING TO THE MODEL SHOULD BE SAME AS HOW THE MODEL ACCEPTS FOR ANY KIND OF DATA TYPE LIKE INT,FLOAT ,CHAR & EVEN STRINGS