

SPRINT 1
Project Deliverables (Code & Test Cases)

Date	04 November 2022
Team ID	PNT2022TMID00111
Project Name	Efficient Water Quality Analysis & Prediction using Machine Learning

HTML (Frontend):

In this sprint , we have created a html code that is going to display on user's screen

home.html - Water Quality - Visual Studio Code

File Edit Selection View Go Run Terminal Help

EXPLORER

- WATER QUALITY
 - static
 - templates
 - home.html 2
 - app.py
 - model.pkl
 - my_scaler.save
 - Procfile
 - requirements.txt
 - water_potability.csv
 - Water_quality.ipynb

home.html

```
1 <!doctype html>
2 <html>
3 <head>
4
5 <title> Water Quality </title>
6 <!-- Bootstrap -->
7 <link href="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.5/css/bootstrap.min.css"
8     rel="stylesheet"
9     integrity="sha256-MfvZlkHCEqatNoGiOxveE8FIwMzZg4W85qrfIFBFYc= sha512-dTfge/zgoMYpP7QbHy4gWMEGsbdsZeCXz7irItjcC3sPUFtf0kuFbDz/ixG7ArTxdJLXDMezHubENikyKGVyQ=="
10    crossorigin="anonymous">
11
12 <style>
13 input{
14     text-align: center;
15     width: 20%;
16     height: 70px;
17     font-size: 14px;
18     padding-top:0px ;
19 }
20 .thick {
21     text-decoration-line: underline;
22     text-decoration-style: solid;
23     text-decoration-color: blue
24     text-decoration-thickness: 2px;
25 }
26 </style>
27 </head>
28 <body style="background-color: powderblue;">
29
30
31 <div class="login">
32
33 <form action="{{ url_for('predict')}}" method="post">
```

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.

- * Running on http://127.0.0.1:5000
- Press CTRL+C to quit
- * Restarting with stat

C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator StandardScaler from version 0.24.0 when using version 1.1.3. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to: https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations

warnings.warn(

- * Debugger is active!
- * Debugger PIN: 873-600-839

127.0.0.1 - - [11/Nov/2022 13:57:12] "GET / HTTP/1.1" 200 -

C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:443: UserWarning: X has feature names, but StandardScaler was fitted without feature names

warnings.warn(

127.0.0.1 - - [11/Nov/2022 13:57:43] "POST /predict HTTP/1.1" 200 -

Ln 77, Col 90 Spaces: 2 UTF-8 CRLF HTML

02:48 PM 11-11-2022

Type here to search

The Marketplace has extensions that can help with 'save' files

Activate Windows

Search Marketplace to find more extensions

Don't Show Again for 'save' files

127.0.0.1:5000/predict

Water Quality_prediction

Enter values

pH value : Hardness : Solids :

Chloramines : Sulfate : Conductivity :

Organic_carbon : Trihalomethanes : Turbidity :

Water quality Test

water is safe for human consumption

Activate Windows
Go to Settings to activate Windows.

Type here to search

04:59 PM
11-11-2022

Data Collection and Preprocessing:

- We have collected data from various sources like sample readings from our college chemistry lab , Kaggle and sources from google
- We have also wrote a preprocessing code to clean , transform , remove the duplicates and intergrate (consolidate) into a single dataset to feed it into the ML model to train the model

IBM-Project-12969-1659503743

IBM-Project-34997-1660280431

IBM-Project-18872-1659690963

Water Quality

IBM

Project Deliverables Submission

Water Quality Prediction (7 mo

←

→

↺

kaggle.com/code/imakash3011/water-quality-prediction-7-model/data

⌵

☆

⚙

⬇

🖨

👤

⋮

Gmail

YouTube

Maps

perfect Plan B # 20...

My Courses | Perfect...

Python for Everybo...

Home | Codewars

Certifications | InfyIQ

AICTE Internship En...

Online Courses - Le...

Striver's SDE Sheet...

TCS NQT 2023 Batc...

Last 20 days of TCS...

Bubble Sort Algorit...

≡

kaggle

+

Create

🏠

Home

🏆

Competitions

📁

Datasets

<>

Code

💬

Discussions

🎓

Learn

▼

More

🔍

Search

Sign In

Register

Water Quality Prediction (7 model)

Notebook

Data

Logs

Comments (45)

50

Copy & Edit

136

⚙

⋮

water_potability.csv (525.19 kB)

⬇

➤

Detail

Compact

Column

10 of 10 columns ▼

About this file

ppm: parts per million
µg/L: microgram per litre
mg/L: milligram per litre
Column description:

# ph	# Hardness	# Solids	# Chloramines	# Sulfate	# Condu
pH of water	Capacity of water to precipitate soap in mg/L	Total dissolved solids in ppm	Amount of Chloramines in ppm	Amount of Sulfates dissolved in mg/L	Electrical water in µ
0	47.4	321	0.35	129	181
	204.8904554713363	20791.318980747026	7.300211873184757	368.51644134980336	564.3080
3.71608007538699	129.42292051494425	18630.057857970347	6.635245883862		592.8850
8.099124189298397	224.23625939355776	19909.541732292393	9.275883602694089		418.6060
8.316765884214679	214.37339408562252	22018.417440775294	8.05933237743854	356.88613564305666	363.2660
9.092223456290965	181.10150923612525	17978.98633892625	6.546599974207941	310.13573752420444	398.4100
5.584086638456089	188.3133237696164	28748.68773904612	7.54486878877965	326.6783629116736	280.4670

Input (525.19 kB)

📁 Data Sources

▼ Water Quality

📄 water_potability.csv

Activate Windows

Go to Settings to activate Windows.

We use cookies on Kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using Kaggle, you agree to our use of cookies.

Got it

Learn more

🏠

🔍 Type here to search

ENG

02:55 PM

11-11-2022

File Edit Selection View Go Run Terminal Help

water_potability.csv - Water Quality - Visual Studio Code

EXPLORER

app.py Water_quality.ipynb home.html 2 water_potability.csv

WATER QUALITYstatictemplateshome.html2app.pymodel.pklmy_scaler.saverequirements.txtwater_potability.csvWater_quality.ipynb

1

ph,Hardness,Solids,Chloramines,Sulfate,Conductivity,Organic_carbon,Trihalomethanes,Turbidity,Potability

2

,204.8904554713363,20791.318980747026,7.300211873184757,368.51644134980336,564.3086541722439,10.3797830780847,86.9909704615088,2.9631353806316407,0

3

3.71608007538699,129.42292051494425,18630.057857970347,6.635245883862,592.8853591348523,15.180013116357259,56.32907628451764,4.500656274942488,0

4

8.009124189298397,224.23625939355776,19909.541732202393,0.275883602694089,418.6062130644815,16.868636929550973,66.420009215176368,3.0559337496641685,0

5

8.316765884214679,214.37339408562252,22018.417440775294,8.05933237743854,356.88613564305666,363.2665161642437,18.436524495493302,100.34167436508008,4.628770536837084,0

6

9.092223456290965,181.10150923612525,17978.98633892625,6.546599974207941,310.13573752420444,398.41081338184466,11.558279443446395,31.997992727424737,4.075075425430034,0

7

5.584086638456089,188.3133237696164,28748.68773904612,7.54486878877965,326.6783629116736,280.4679159334877,8.399734640152758,54.917861841994466,2.5597082275565217,0

8

10.223862164528773,248.07173527013992,28749.716543528233,7.5134084658313025,393.66339551509645,283.6516335078445,13.789695317519886,84.60355617402357,2.672988736934779,0

9

8.635848718500734,203.36152258457054,13672.091763901635,4.563008685599703,303.3097711592812,474.60764494244853,12.36381669870525,62.798308962925155,4.401424715445482,0

10

,118.98857909025189,14285.583854224515,7.804173553073094,268.646940746221,389.3755658712614,12.70604896865791,53.928845767512236,3.5950171809576155,0

11

11.180284470721592,227.23146923797458,25484.50849098786,9.077200016914393,404.04163468408996,563.8854814810949,17.92780641128502,71.97660103221915,4.370561936655497,0

12

7.360640105838258,165.52079725952862,32452.6144009143884,7.550700906704114,326.62435345550164,425.38341949538733,15.586810438033126,78.74001566430479,3.6622917828524573,0

13

7.974521648923869,218.69330048866644,18767.65668181348,8.110384501123875,364.09823046204866,14.525745697593209,76.48591117065157,4.011718108339787,0

14

7.119824384264552,156.70499334039215,18730.813653342713,3.6060360905057203,282.3440504739606,347.71502726194376,15.929535908825699,79.5007783369744,3.445756223321899,0

15

,150.1749233951362,27331.361961927756,6.838223470687509,299.41578134685847,379.76183482577244,19.370807181232124,76.5099955279583,4.413974182974902,0

16

7.49623220797336,205.34498215818513,28388.00488673697,5.072557773840631,444.6453523327066,13.228311099224527,70.30021264692436,4.777382337225378,0

17

6.347271760539316,186.73288066057614,41065.23476453935,9.629596276480854,364.4876872467604,516.743281803657,11.530781191539419,75.07161728663777,4.376348290691898,0

18

7.051785800016845,211.04940606054578,30980.600786788862,10.094796011661426,315.1412672443021,20.39702184072246,56.65160378979331,4.268428857506186,0

19

9.181560007151536,273.81380665980095,24041.32628006128,6.904989726470096,398.3505168222779,477.9746418621779,13.387340780225543,71.4573622129516,4.503660796179122,0

20

8.975464347533963,279.35716677009236,19460.398131232112,6.204320858892474,431.44398999034894,12.88875905430399,63.82123709666397,2.4360855903052734,0

21

7.371050302429531,214.49661045715658,25630.320036999725,4.432669290371213,335.75443859606526,469.91455147923585,12.509163940498695,62.79727715266126,2.5602991476149146,0

22

,227.43504835115596,22305.56741374141,10.333917888218679,554.8200864605433,16.33169328269446,45.382815177870924,4.13342264357917,0

23

6.660212026118103,168.28374685651832,30944.363591242687,5.858769130547582,310.93085831787846,523.6712975009444,17.88423519296481,77.0423180517003,3.7497012410996176,0

24

,215.97785868806778,17107.224225827616,5.607060453087125,326.943977743867,436.25619397264916,14.189062206123708,59.85547582615388,4.559250956028731,0

25

3.902475685915096,196.9032467083208,21167.500098968772,6.996311586298768,444.47888250689795,16.609033155789916,90.1816758847452,4.528522696326911,0

26

5.400301780729467,140.73906225113961,17266.503421923077,10.05685248403495,328.35824069086835,472.8740732754293,11.256381166900478,56.9319064457562,4.824786389767524,0

27

6.514415093251676,198.76735125945606,21218.702871190195,8.67093691991312,323.5963490101317,413.2904500885347,14.89999566696977,79.84784281372556,5.20088507639757,0

28

3.4450618643852127,207.92626018799376,33424.7686784948,8.782147480773485,384.00700580172116,441.7858756739387,13.80590221127079,30.284597198002704,4.184396969028851,0

29

,145.76818060217258,13224.935638976958,7.906444720606137,304.0019927974152,298.99066649993244,12.729524720542258,49.536848802021865,4.004871127571583,0

30

,266.4210180681174,26362.965012309312,7.700063460729127,395.38949034184554,364.48010670373776,10.34895075743782,53.00838135392041,3.9915642477993067,0

31

,148.15306144508662,15193.41347396722,9.04683270725723,307.01179262256534,563.8047433250861,16.56865556799744,52.67618503420983,6.0381849531835865,0

32

7.181448580829175,209.62560053629045,15196.229987483843,5.994678646449973,338.3364310774872,342.1112862851926,7.92259833302262,71.5379532557936,5.088859989138795,0

33

9.82548990813439,190.7566182870043,19677.892465552013,6.757540731413941,452.8362348772383,16.8990378000164,47.081971185777654,2.857472426051184,0

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.

* Running on http://127.0.0.1:5000

Press CTRL+C to quit

* Restarting with stat

C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator StandardScaler from version 0.24.0 when using version 1.1.3. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to: https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations

warnings.warn(

* Debugger is active!

* Debugger PIN: 873-600-839

127.0.0.1 - - [11/Nov/2022 13:57:12] "GET / HTTP/1.1" 200 -

C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:443: UserWarning: X has feature names, but StandardScaler was fitted without feature names

warnings.warn(

127.0.0.1 - - [11/Nov/2022 13:57:43] "POST /predict HTTP/1.1" 200 -

Activate Windows

Go to Settings to activate Windows.

Ln 1, Col 1 Spaces: 4 UTF-8 CRLF Plain Text

2 0

02:55 PM 11-11-2022

File Edit Selection View Go Run Terminal Help

Water_quality.ipynb - Water Quality - Visual Studio Code

EXPLORER

Water_quality.ipynb

Water_quality.ipynb

static

templates

home.html

app.py

model.pkl

my_scaler.save

Profile

requirements.txt

water_potability.csv

Water_quality.ipynb

Water_quality.ipynb

Problem Statement

Hyperparameter Tuning with Support vector Machine

Pickle

+ Code

+ Markdown

Run All

Clear Outputs of All Cells

Restart

Variables

Outline

Python 3.9.1 64-bit

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

[3] ✓ 6.6s

Python

df = pd.read_csv("water_potability.csv")

df.head()

[4] ✓ 0.3s

Python

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
0	NaN	204.890455	20791.318981	7.300212	368.516441	564.308654	10.379783	86.990970	2.963135	0
1	3.716080	129.422921	18630.057858	6.635246	NaN	592.885359	15.180013	56.329076	4.500656	0
2	8.099124	224.236259	19909.541732	9.275884	NaN	418.606213	16.868637	66.420093	3.055934	0
3	8.316766	214.373394	22018.417441	8.059332	356.886136	363.266516	18.436524	100.341674	4.628771	0
4	9.092223	181.101509	17978.986339	6.546600	310.135738	398.410813	11.558279	31.997993	4.075075	0

[5] ✓ 0.1s

Python

df.shape

[5] ✓ 0.1s

Python

(3276, 10)

df.info()

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

JUPYTER

WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.

* Running on http://127.0.0.1:5000

Press CTRL+C to quit

* Restarting with stat

C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator StandardScaler from version 0.24.0 when using version 1.1.3. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to: https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations

warnings.warn(

* Debugger is active!

* Debugger PIN: 873-600-839

127.0.0.1 - - [11/Nov/2022 13:57:12] "GET / HTTP/1.1" 200 -

C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:443: UserWarning: X has feature names, but StandardScaler was fitted without feature names

warnings.warn(

127.0.0.1 - - [11/Nov/2022 13:57:43] "POST /predict HTTP/1.1" 200 -

Activate Windows

Go to Settings to activate Windows.

Jupyter Server: Local

Ln 1, Col 1

Spaces: 4

CRLF

Cell 51 of 52

Type here to search

02:56 PM

11-11-2022

File Edit Selection View Go Run Terminal Help

Water_quality.ipynb - Water Quality - Visual Studio Code

EXPLORER

WATER_QUALITY

static

templates

home.html

app.py

model.pkl

my_scaler.save

Profile

requirements.txt

water_potability.csv

Water_quality.ipynb

Water_quality.ipynb x home.html 2 water_potability.csv

Water_quality.ipynb > Problem Statement > Hyperparameter Tuning with Support vector Machine > Pickle

+ Code + Markdown | Run All | Clear Outputs of All Cells | Restart | Variables | Outline |

Python 3.9.1 64-bit

df.info()

[6] ✓ 0.1s

Python

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 3276 entries, 0 to 3275

Data columns (total 10 columns):

Column Non-Null Count Dtype

0 ph 2785 non-null float64

1 Hardness 3276 non-null float64

2 Solids 3276 non-null float64

3 Chloramines 3276 non-null float64

4 Sulfate 2495 non-null float64

5 Conductivity 3276 non-null float64

6 Organic_carbon 3276 non-null float64

7 Trihalomethanes 3114 non-null float64

8 Turbidity 3276 non-null float64

9 Potability 3276 non-null int64

dtypes: float64(9), int64(1)

memory usage: 256.1 KB

df.describe()

[7] ✓ 0.2s

Python

ph Hardness Solids Chloramines Sulfate Conductivity Organic_carbon Trihalomethanes Turbidity Potability

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.

* Running on http://127.0.0.1:5000

Press CTRL+C to quit

* Restarting with stat

C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator StandardScaler from version 0.24.0 when using version 1.1.3. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to: https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations

warnings.warn(

* Debugger is active!

* Debugger PIN: 873-600-839

127.0.0.1 - - [11/Nov/2022 13:57:12] "GET / HTTP/1.1" 200 -

C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:443: UserWarning: X has feature names, but StandardScaler was fitted without feature names

warnings.warn(

127.0.0.1 - - [11/Nov/2022 13:57:43] "POST /predict HTTP/1.1" 200 -

Activate Windows

Go to Settings to activate Windows.

Jupyter Server: Local Ln 1, Col 1 Spaces: 4 CRLF Cell 51 of 52

Type here to search

02:56 PM 11-11-2022

File Edit Selection View Go Run Terminal Help

Water_quality.ipynb - Water Quality - Visual Studio Code

EXPLORER

WATER QUALITY

static

templates

home.html

app.py

model.pkl

my_scaler.save

Profile

requirements.txt

water_potability.csv

Water_quality.ipynb

Water_quality.ipynb x home.html 2 water_potability.csv

Water_quality.ipynb > Problem Statement > Hyperparameter Tuning with Support vector Machine > ## Pickle

+ Code + Markdown | Run All Clear Outputs of All Cells Restart Variables Outline

Python 3.9.1 64-bit

df.describe()

[7] ✓ 0.2s

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
count	2785.000000	3276.000000	3276.000000	3276.000000	2495.000000	3276.000000	3276.000000	3114.000000	3276.000000	3276.000000
mean	7.080795	196.369496	22014.092526	7.122277	333.775777	426.205111	14.284970	66.396293	3.966786	0.390110
std	1.594320	32.879761	8768.570828	1.583085	41.416840	80.824064	3.308162	16.175008	0.780382	0.487849
min	0.000000	47.432000	320.942611	0.352000	129.000000	181.483754	2.200000	0.738000	1.450000	0.000000
25%	6.093092	176.850538	15666.690297	6.127421	307.699498	365.734414	12.065801	55.844536	3.439711	0.000000
50%	7.036752	196.967627	20927.833607	7.130299	333.073546	421.884968	14.218338	66.622485	3.955028	0.000000
75%	8.062066	216.667456	27332.762127	8.114887	359.950170	481.792304	16.557652	77.337473	4.500320	1.000000
max	14.000000	323.124000	61227.196008	13.127000	481.030642	753.342620	28.300000	124.000000	6.739000	1.000000

sns.countplot(x='Potability',data=df)

[8] ✓ 0.3s

<AxesSubplot: xlabel='Potability', ylabel='count'>

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
* Restarting with stat
C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator StandardScaler from version 0.24.0 when using version 1.1.3. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to: https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
warnings.warn(
* Debugger is active!
* Debugger PIN: 873-600-839
127.0.0.1 - - [11/Nov/2022 13:57:12] "GET / HTTP/1.1" 200 -
C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:443: UserWarning: X has feature names, but StandardScaler was fitted without feature names
warnings.warn(
127.0.0.1 - - [11/Nov/2022 13:57:43] "POST /predict HTTP/1.1" 200 -

Activate Windows
Go to Settings to activate Windows.

2 0

Jupyter Server: Local Ln 1, Col 1 Spaces: 4 CRLF Cell 51 of 52

Type here to search

02:56 PM 11-11-2022

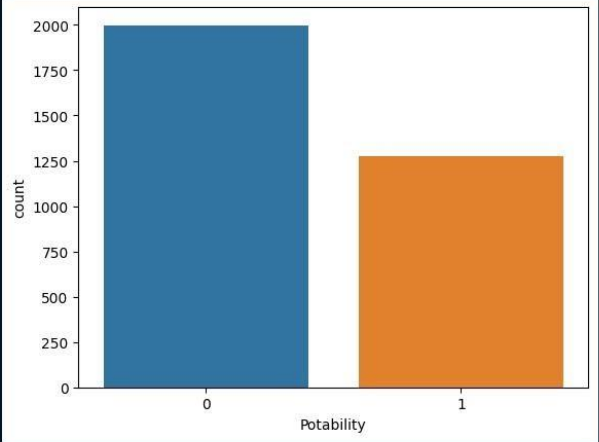
Visual Studio Code interface showing a Jupyter Notebook titled "Water_quality.ipynb" with a bar chart and a terminal window.

Explorer Panel:

- WATER_QUALITY
 - static
 - templates
 - home.html
 - app.py
 - model.pkl
 - my_scaler.save
 - Profile
 - requirements.txt
 - water_potability.csv
 - Water_quality.ipynb

Code Editor:

```
<AxesSubplot: xlabel='Potability', ylabel='count'>
```



Potability	count
0	2000
1	1300

Terminal:

```
df["Potability"].value_counts()
```

WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.

- * Running on http://127.0.0.1:5000
- Press CTRL+C to quit
- * Restarting with stat

C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator StandardScaler from version 0.24.0 when using version 1.1.3. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to: https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations

```
warnings.warn(
* Debugger is active!
* Debugger PIN: 873-600-839
127.0.0.1 - - [11/Nov/2022 13:57:12] "GET / HTTP/1.1" 200 -
C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:443: UserWarning: X has feature names, but StandardScaler was fitted without feature names
warnings.warn(
127.0.0.1 - - [11/Nov/2022 13:57:43] "POST /predict HTTP/1.1" 200 -
```

Activate Windows
Go to Settings to activate Windows.