National College of Ireland

Configuration Manual

MSc Research Project
MSc in Cloud Computing

Tanuj Kumar
Student ID: X20228821

School of Computing
National College of Ireland

Supervisor: Dr. Shivani Jaswal

National College of Ireland
MSc Project Submission Sheet
School of Computing

Student Name: ……. ……Tanuj Kumar……
Student ID: ……X20228821……
Program: ……Msc in Cloud Computing…… **Year:** 2023……..
Module: ……MSc Research Project……
Supervisor: ……Dr Shivani Jaswal……
Submission Due Date: ……25th May2023……
Project Title: Use of Machine Learning Model for Improving Cardiovascular Condition Using Cloud Computing
Word Count ……1275…… **Page Count**……23……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……Tanuj Kumar……
Date: ……25th April 2023……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

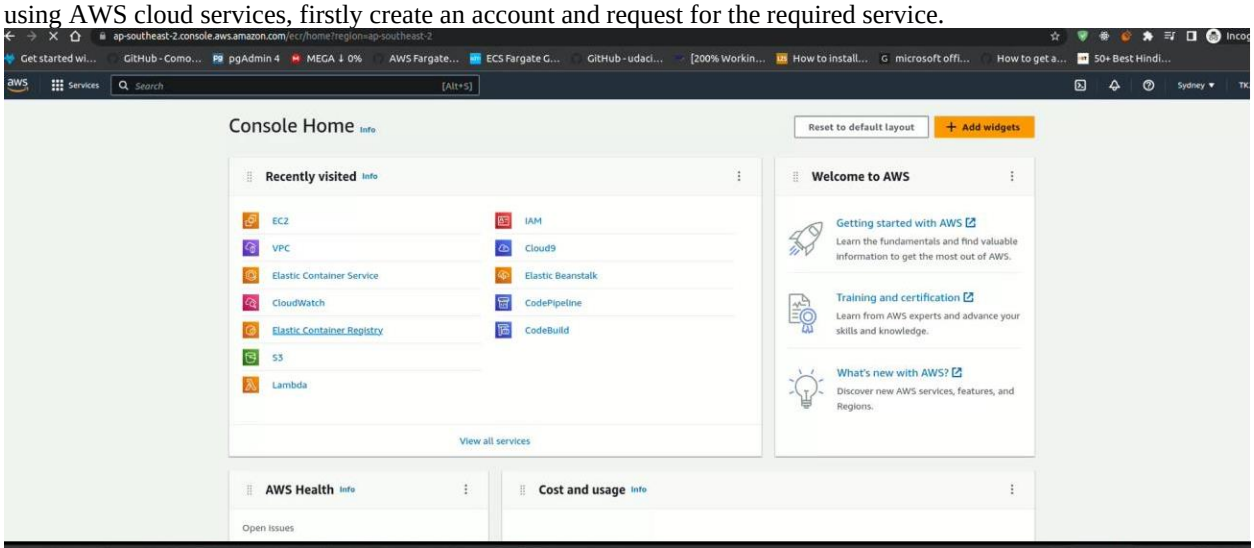
Configuration Manual

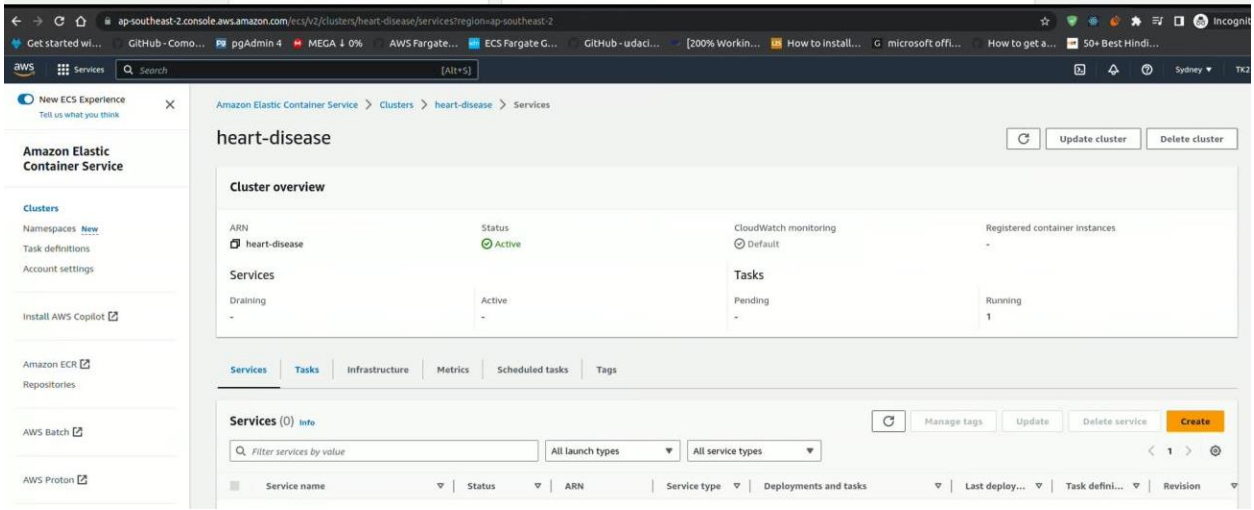
1 Introduction

The configuration manual includes the different aspects of machine learning that consist of the different machine learning aspects. There are different tools and techniques are included for performing the machine learning algorithm. In order to describe to develop a complete understanding of the complete experiment for the heart prediction machine learning system, we designed the configuration manually. Prerequisite

1.1 AWS

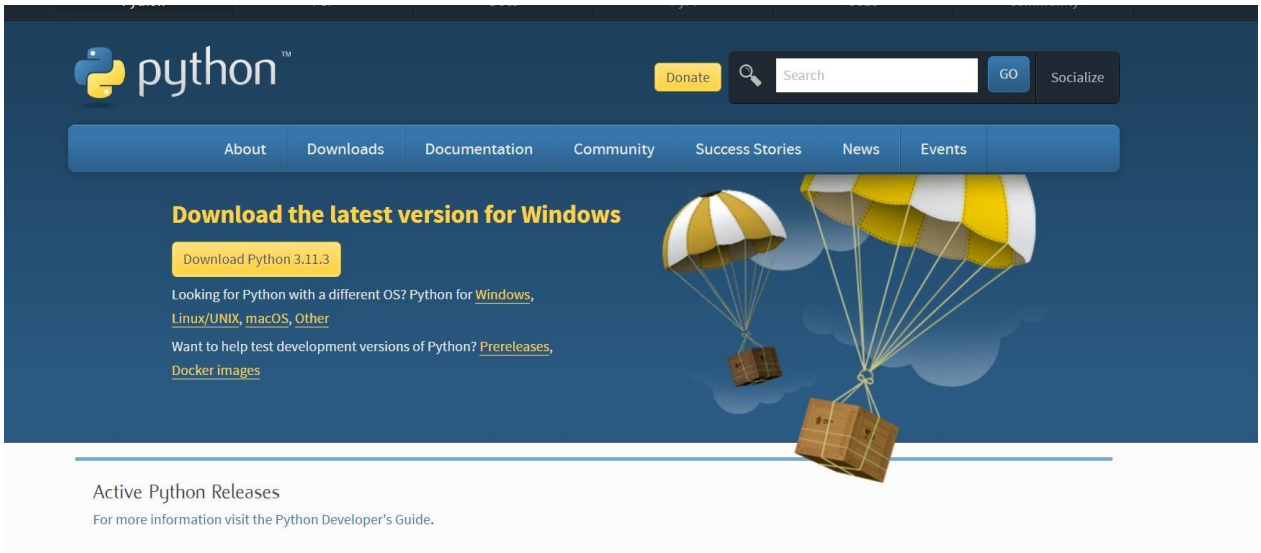
In order to deploy the machine learning model, we adapted the docker container image for deploying the machine learning model. AWS allows users to host their applications in a secure manner. Container images provide the ready-to-use template and then the code of an application can be run in the proper manner. For using AWS cloud services, firstly create an account and request for the required service.

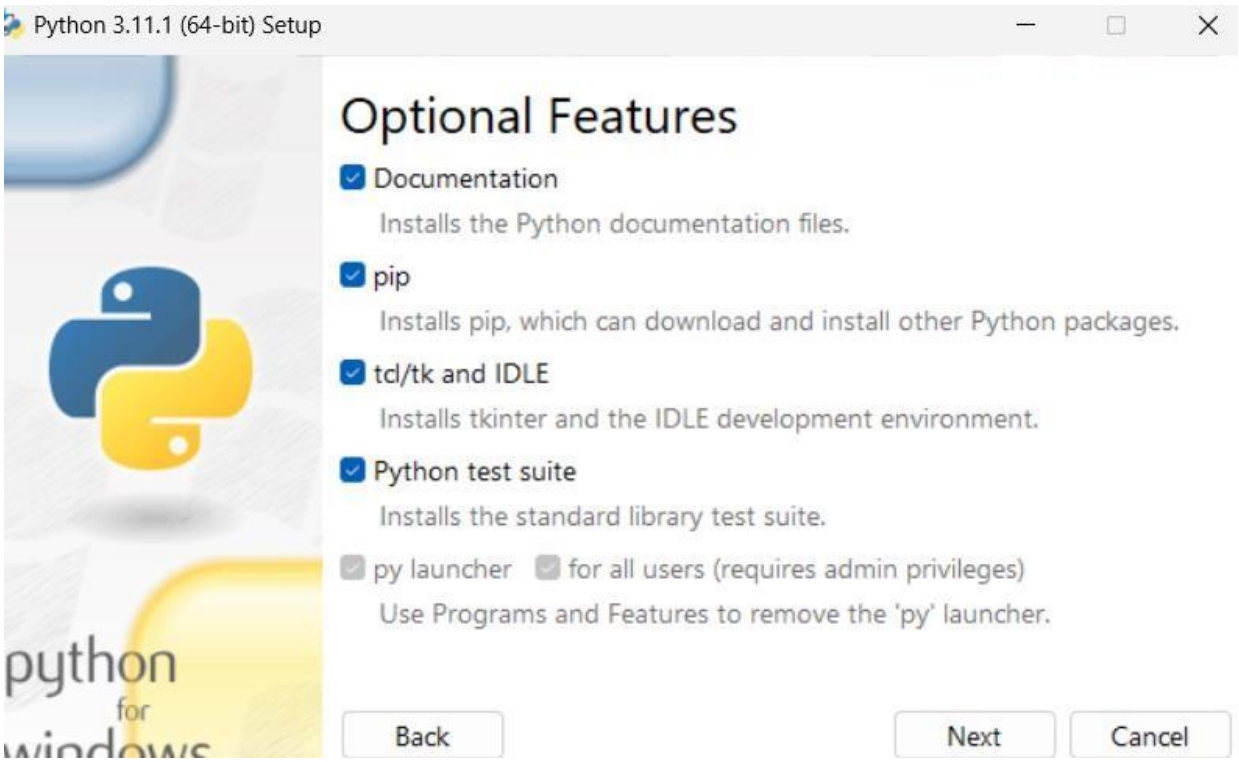


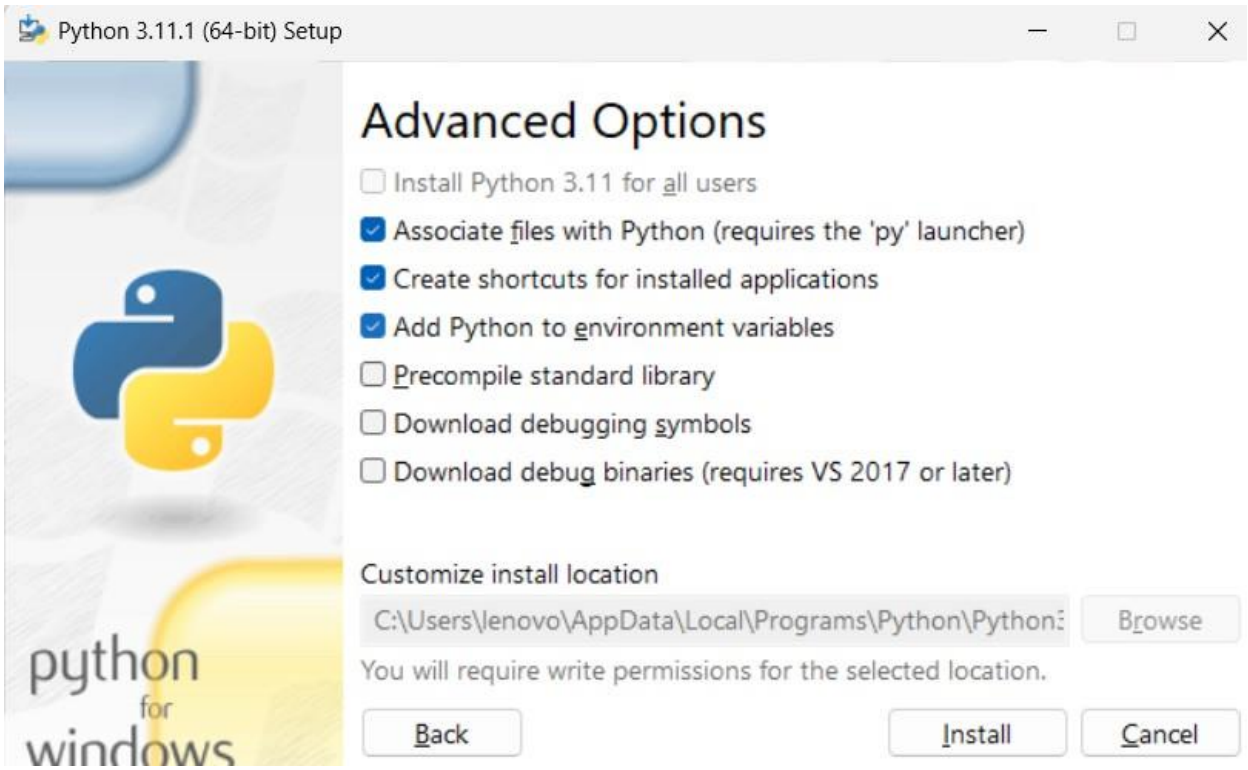


1.2 Python

Python is defined as the programming language that can be used for both scripting and programming tasks. The Python language can be used for different tasks such as the machine learning algorithm, web development, and security of applications. The research is associated with the different aims of Python such as machine learning algorithms and web development tasks (Python, 2023).
Python can be download from its official site.

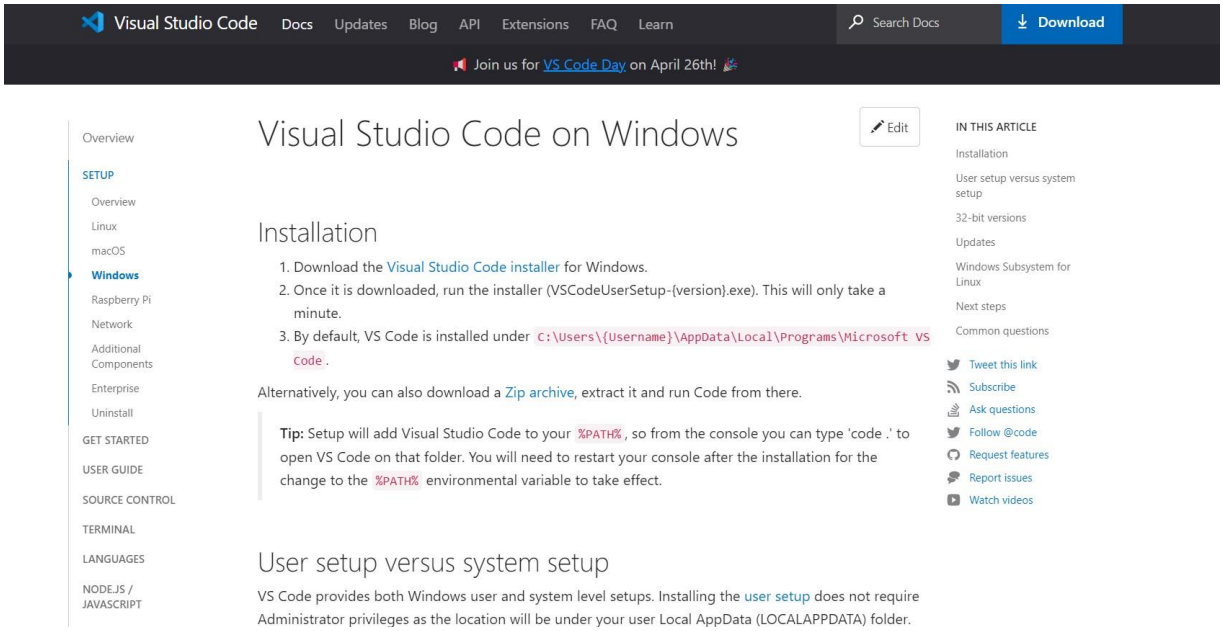

python
Donate
Search
GO
Socialize
About
Downloads
Documentation
Community
Success Stories
News
Events
Download the latest version for Windows
Download Python 3.11.3
Looking for Python with a different OS? Python for Windows, Linux/UNIX, macOS, Other
Want to help test development versions of Python? Prereleases, Docker images
Active Python Releases
For more information visit the Python Developer's Guide.

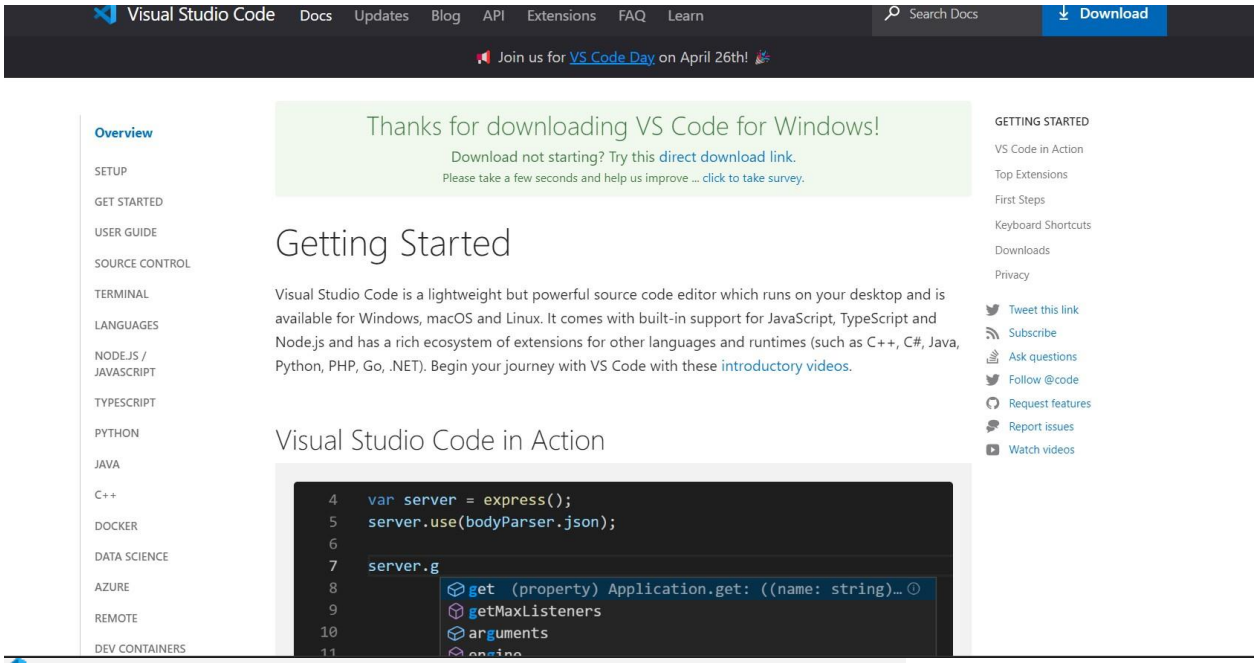

Python 3.11.1 (64-bit) Setup
Optional Features
Documentation
Installs the Python documentation files.
pip
Installs pip, which can download and install other Python packages.
tcl/tk and IDLE
Installs tkinter and the IDLE development environment.
Python test suite
Installs the standard library test suite.
py launcher
for all users (requires admin privileges)
Use Programs and Features to remove the 'py' launcher.
python
for
Back
Next
Cancel

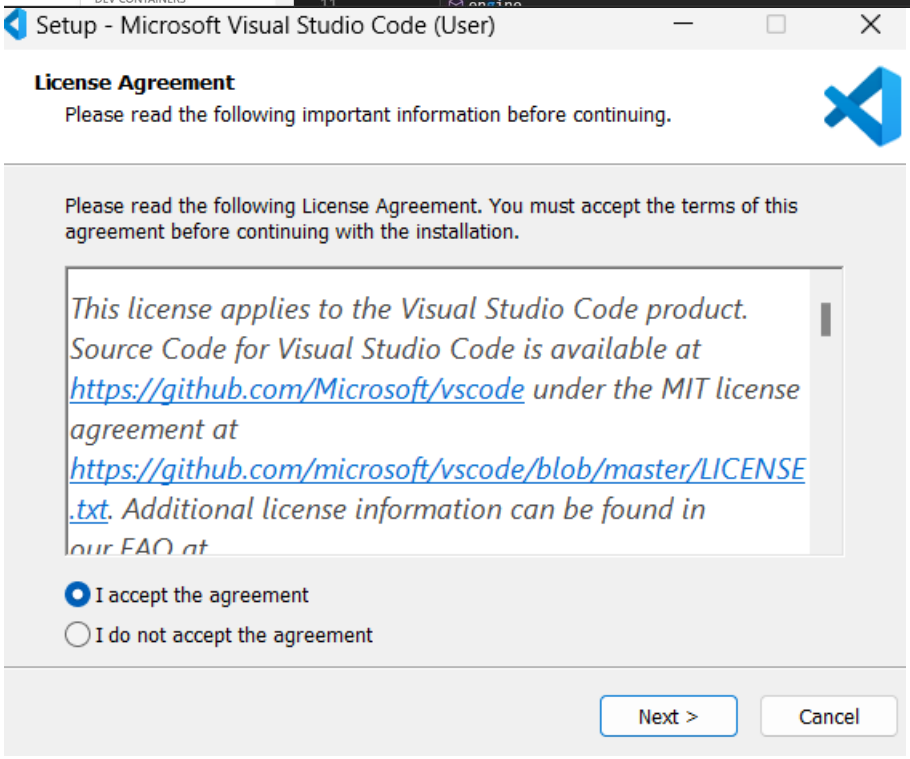


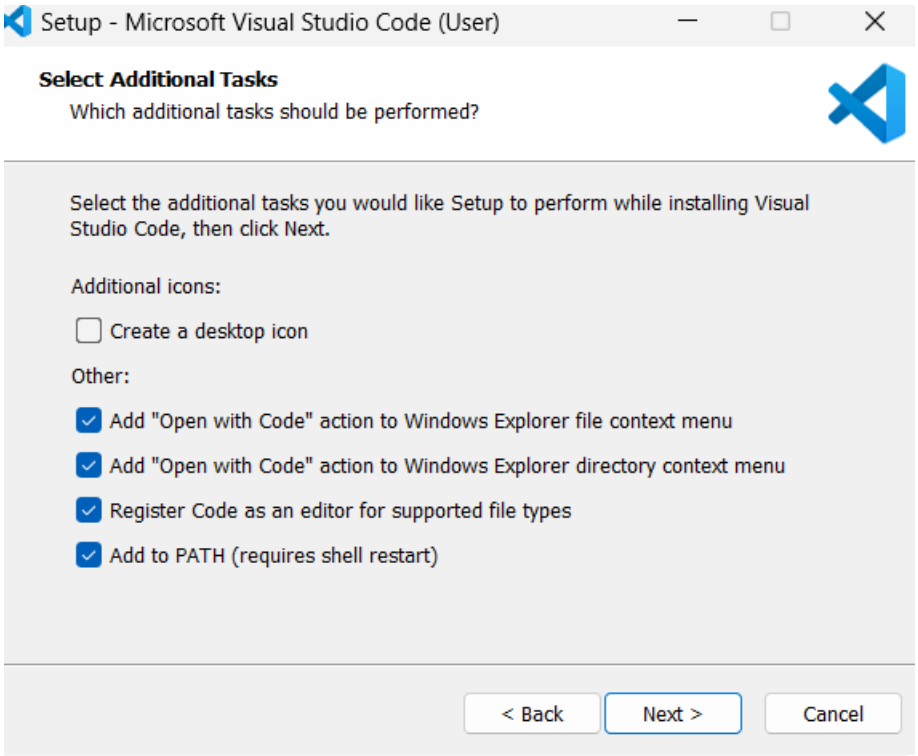
1.3 Visual Studio Code

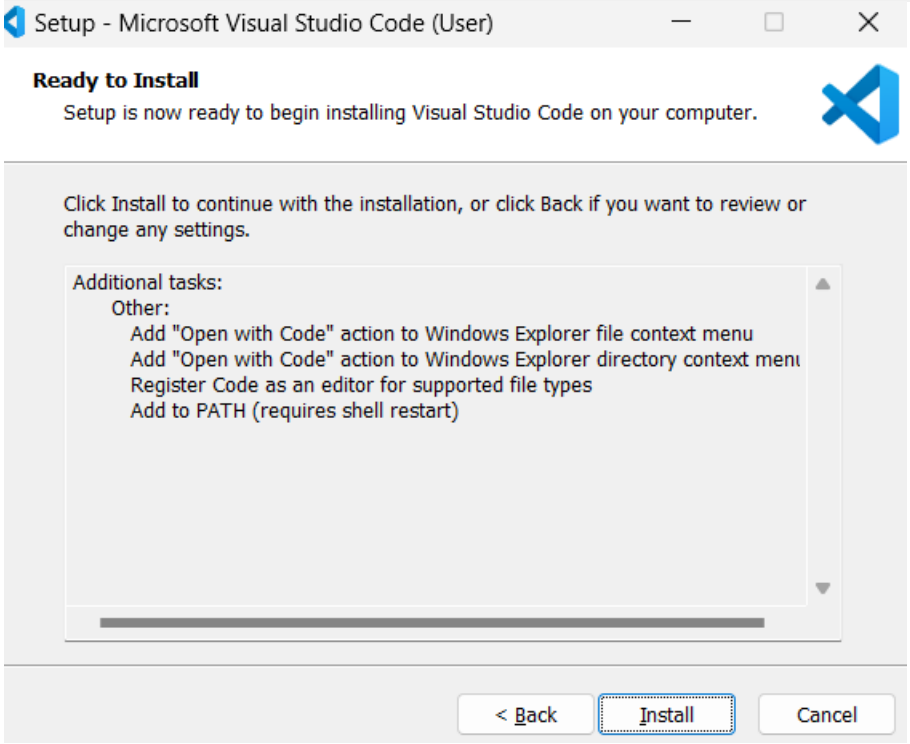
In order to write the machine learning algorithms, the research is associated with the Visual Studio code that is integrated with the Python environment and Jupyter Notebook. The HTML, CSS, and Flask can also be performed with the help of VS code. There are different features included in the VS code such as syntax highlighting, debugging, code completion, embedding git, code refactoring, and snippets.
For using the vs code download vs code from its official site and launch.




Visual Studio Code
Docs
Updates
Blog
API
Extensions
FAQ
Learn
Search Docs
Download
Join us for VS Code Day on April 26th!
Overview
SETUP
GET STARTED
USER GUIDE
SOURCE CONTROL
TERMINAL
LANGUAGES
NODE.JS / JAVASCRIPT
TYPESCRIPT
PYTHON
JAVA
C++
DOCKER
DATA SCIENCE
AZURE
REMOTE
DEV CONTAINERS
Thanks for downloading VS Code for Windows!
Download not starting? Try this direct download link.
Please take a few seconds and help us improve ... click to take survey.
Getting Started
Visual Studio Code is a lightweight but powerful source code editor which runs on your desktop and is available for Windows, macOS and Linux. It comes with built-in support for JavaScript, TypeScript and Node.js and has a rich ecosystem of extensions for other languages and runtimes (such as C++, C#, Java, Python, PHP, Go, .NET). Begin your journey with VS Code with these introductory videos.
Visual Studio Code in Action
GETTING STARTED
VS Code in Action
Top Extensions
First Steps
Keyboard Shortcuts
Downloads
Privacy
Tweet this link
Subscribe
Ask questions
Follow @code
Request features
Report issues
Watch videos
var server = express();
server.use(bodyParser.json);
server.g
get (property) Application.get: ((name: string)…
getMaxListeners
arguments

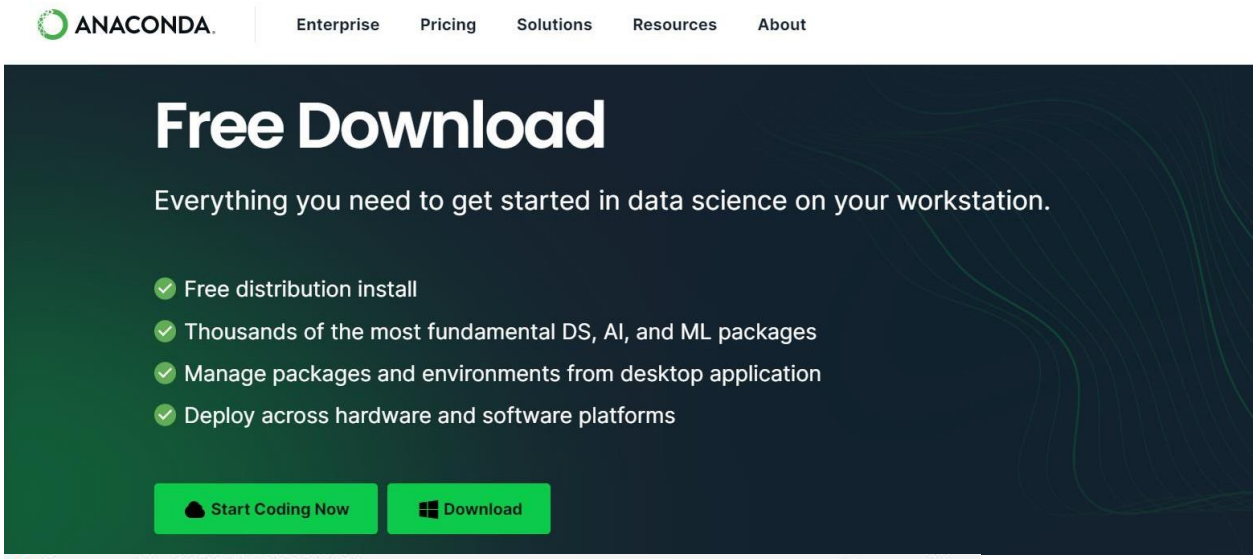

Setup - Microsoft Visual Studio Code (User)
License Agreement
Please read the following important information before continuing.
Please read the following License Agreement. You must accept the terms of this agreement before continuing with the installation.
This license applies to the Visual Studio Code product. Source Code for Visual Studio Code is available at https://github.com/Microsoft/vscode under the MIT license agreement at https://github.com/microsoft/vscode/blob/master/LICENSE.txt. Additional license information can be found in
I accept the agreement
I do not accept the agreement
Next >
Cancel

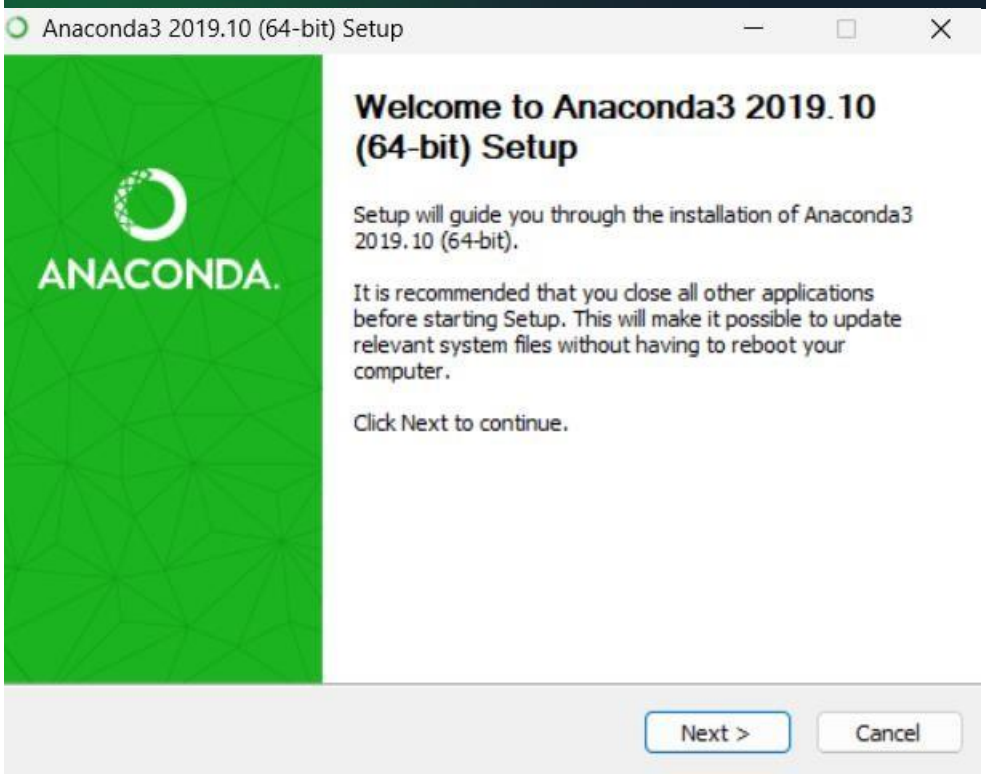


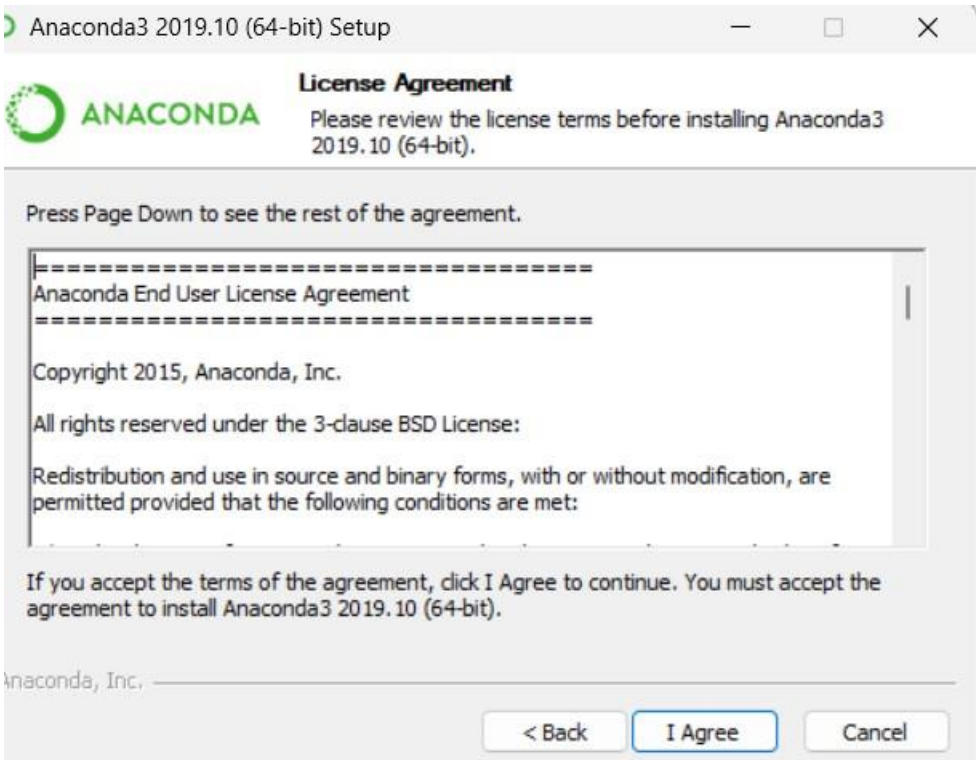


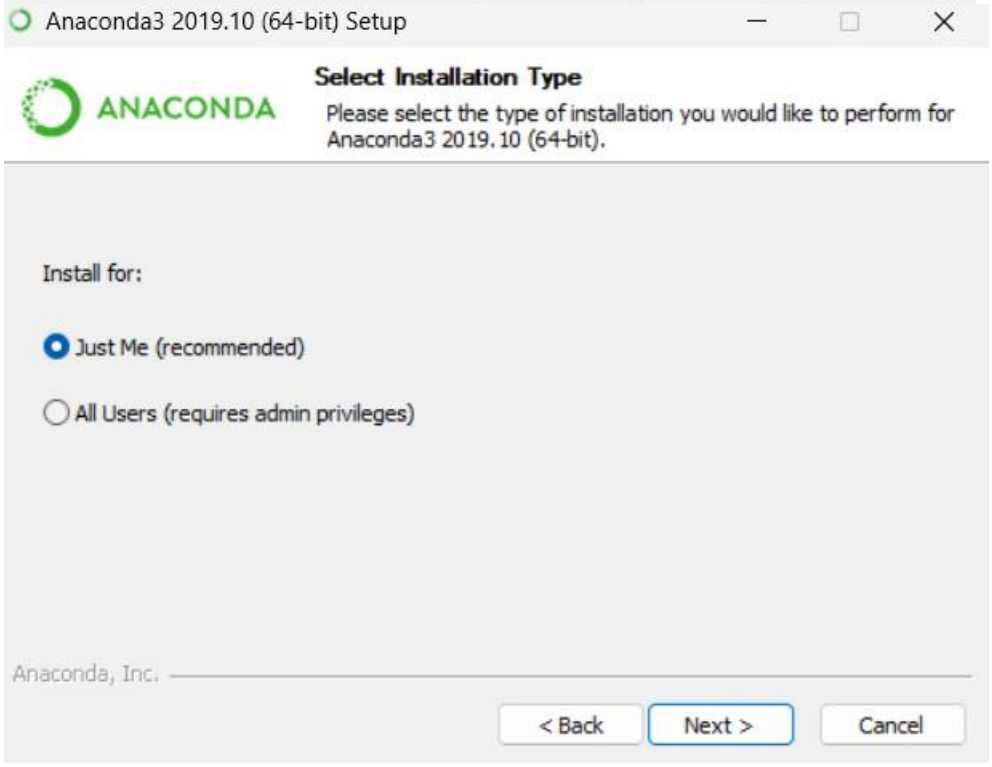
1.4 Anaconda:

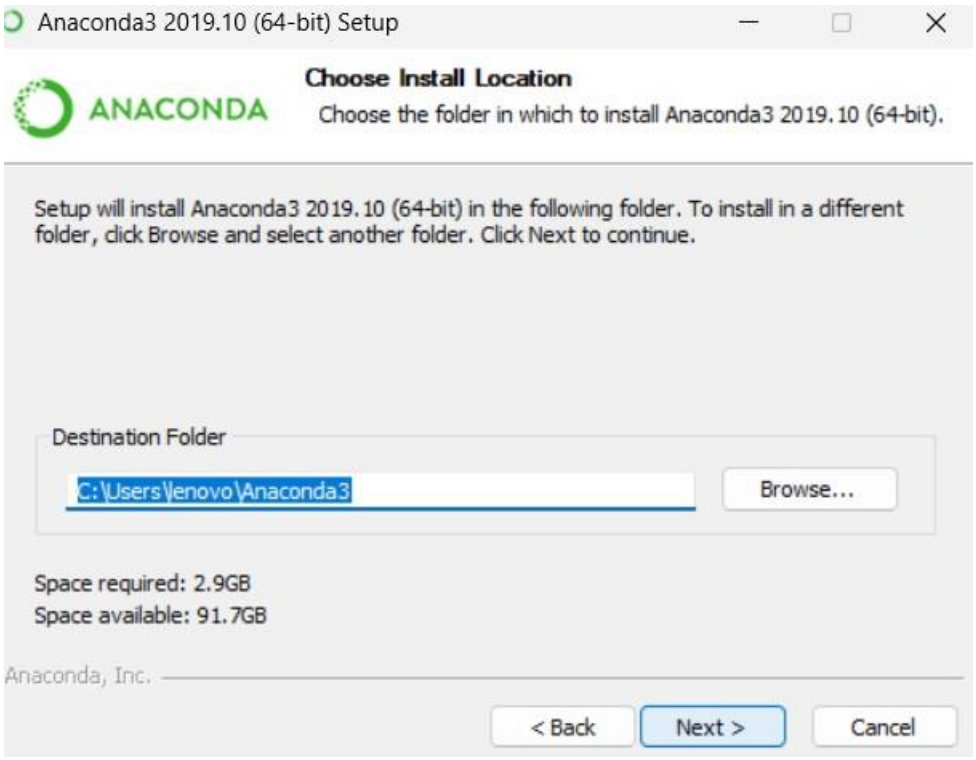
Anaconda is a platform that is compatible with R and Python programming languages which can be used for scientific computing. The main aim behind the development of Anaconda is to develop a smooth process for data science. There are different platforms are integrated with the Anaconda such as the Jupyter Notebook, VS code, Spyder, orange, Pycharm, and so on. This can be download from the official site of anaconda.


ANACONDA.
Enterprise
Pricing
Solutions
Resources
About
Free Download
Everything you need to get started in data science on your workstation.
Free distribution install
Thousands of the most fundamental DS, AI, and ML packages
Manage packages and environments from desktop application
Deploy across hardware and software platforms
Start Coding Now
Download


Anaconda3 2019.10 (64-bit) Setup
ANACONDA.
Welcome to Anaconda3 2019.10 (64-bit) Setup
Setup will guide you through the installation of Anaconda3 2019.10 (64-bit).
It is recommended that you close all other applications before starting Setup. This will make it possible to update relevant system files without having to reboot your computer.
Click Next to continue.
Next >
Cancel


Anaconda3 2019.10 (64-bit) Setup
ANACONDA
License Agreement
Please review the license terms before installing Anaconda3 2019.10 (64-bit).
Press Page Down to see the rest of the agreement.
Anaconda End User License Agreement
Copyright 2015, Anaconda, Inc.
All rights reserved under the 3-clause BSD License:
Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:
If you accept the terms of the agreement, click I Agree to continue. You must accept the agreement to install Anaconda3 2019.10 (64-bit).
Anaconda, Inc.
< Back
I Agree
Cancel

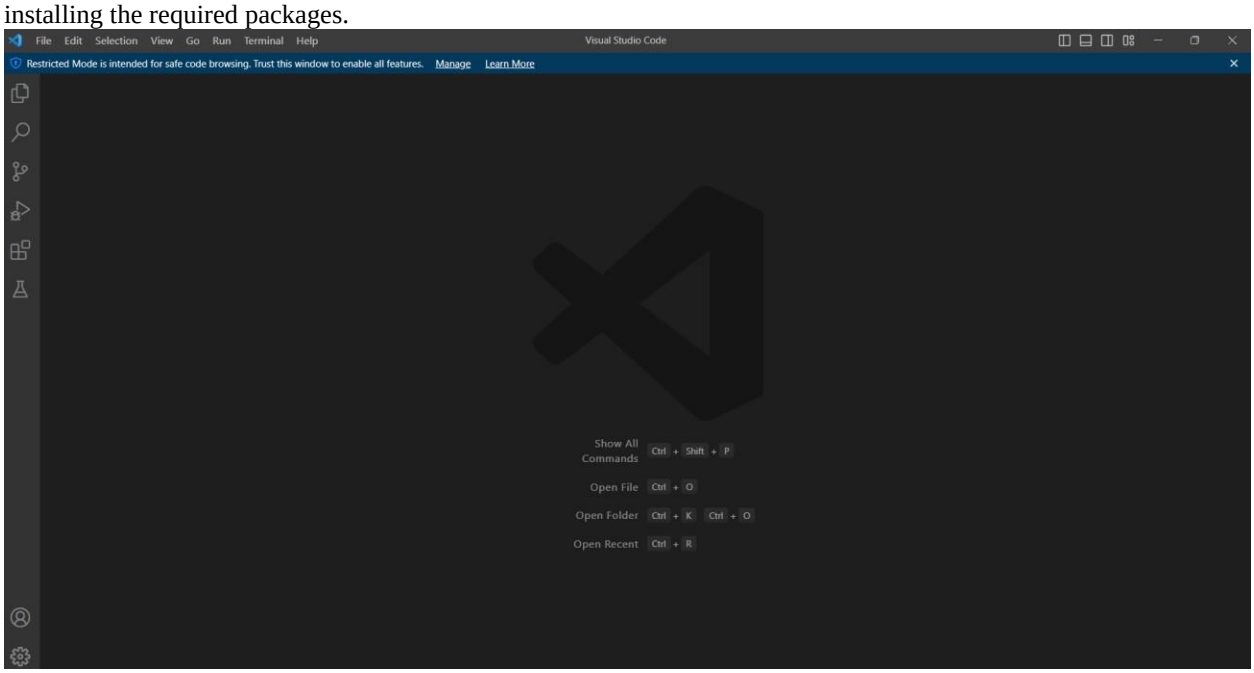

Anaconda3 2019.10 (64-bit) Setup
ANACONDA
Select Installation Type
Please select the type of installation you would like to perform for Anaconda3 2019.10 (64-bit).
Install for:
Just Me (recommended)
All Users (requires admin privileges)
Anaconda, Inc.
< Back
Next >
Cancel



2 Libraries/Packages used

Python is associated with the different open-source frameworks which can be used for the different tasks that are listed below.

For use the libraries of python need to install the libraries. For this open vs-code and follow below steps for installing the required packages.

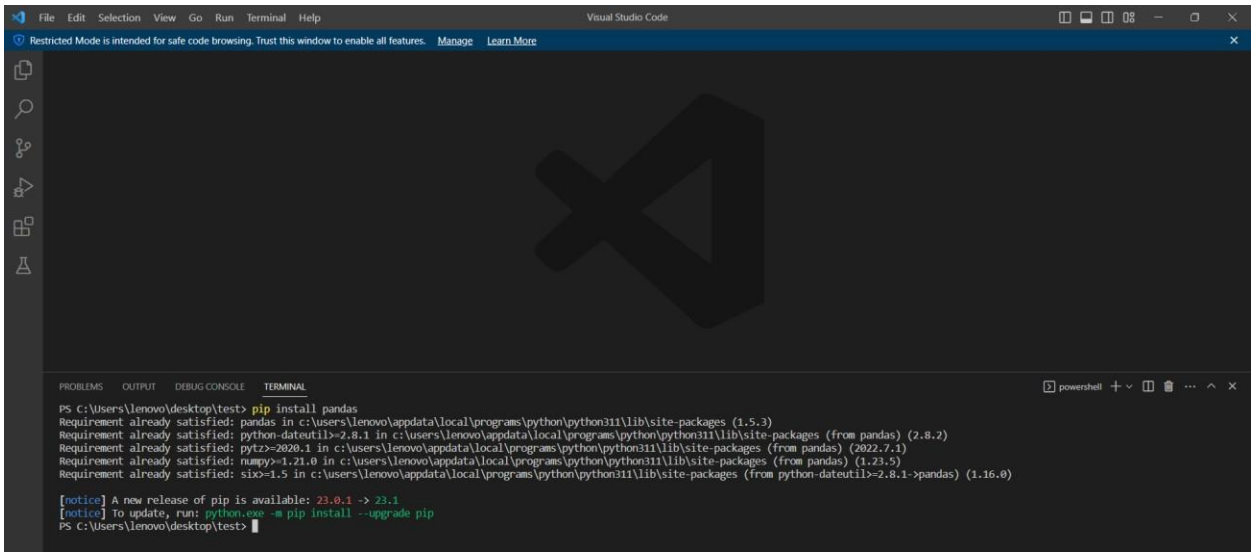


2.1 Pandas

Pandas is a Python-based library that is used for dealing with the data and it is associated with a different type of statistical analysis. It can also be integrated with different Python-based functions such as Sklearn, matplotlib, seaborn, and so on.

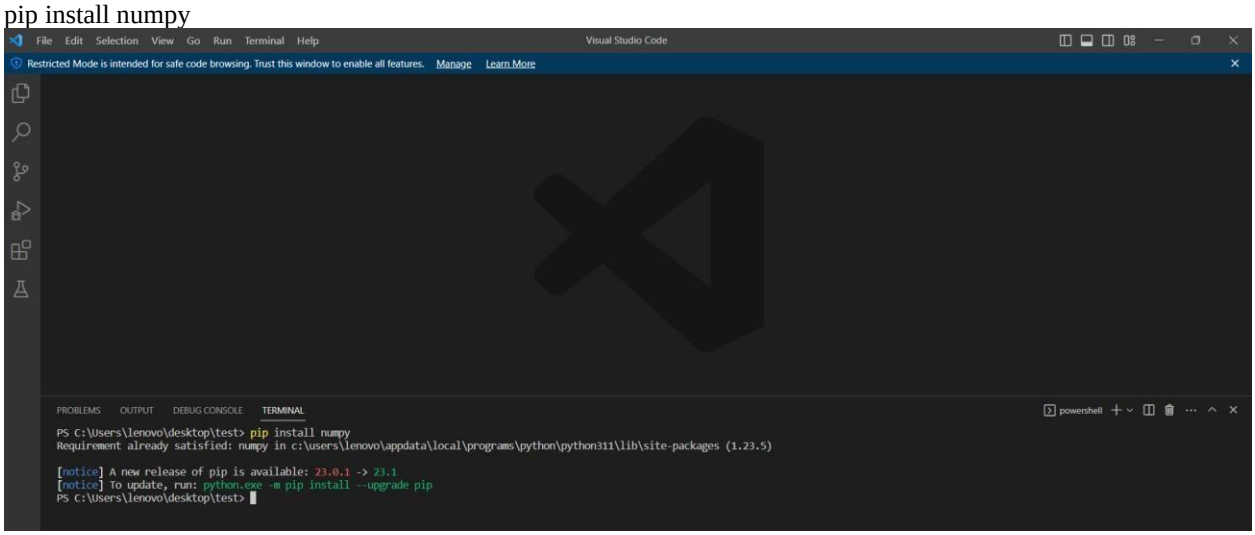
For installing pandas, open the vs-code and click on terminal and type this command:

pip install pandas



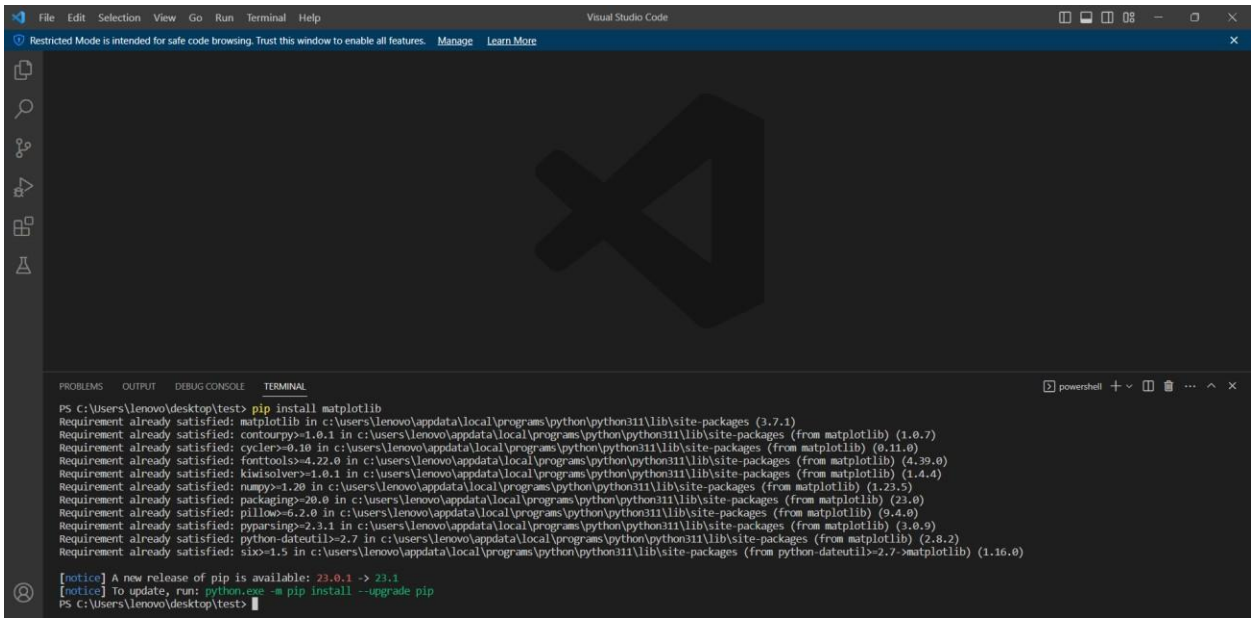
2.2 NumPy

NumPy is a computation library that can be used for different statistical functions and it supports the use of an array which is the concept of Python programming language so it can be stated that it is faster than the python data types such as the list and tuple.
For installing Numpy, open the vs-code and click on terminal and type this command:
pip install numpy



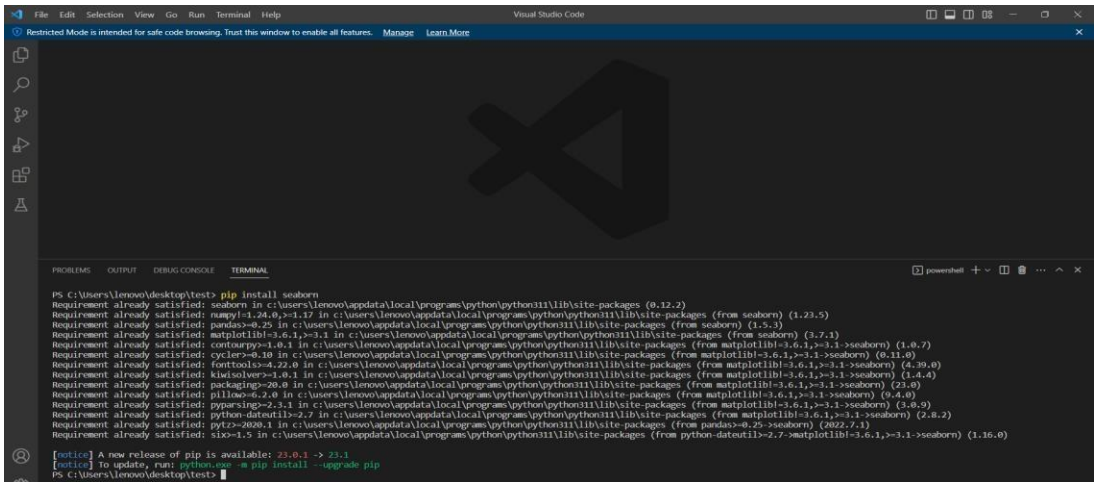
2.3 Matplotlib

It is considered as the graphical extension of the NumPy library that provides the GUI for plotting the data into different formats. It is designed with the help of different toolkits such as GTK, Qt, and Tkinter. It also facilities the functions for plotting the graph.
For installing matplotlib, open the vs-code and click on terminal and type this command:
pip install matplotlib



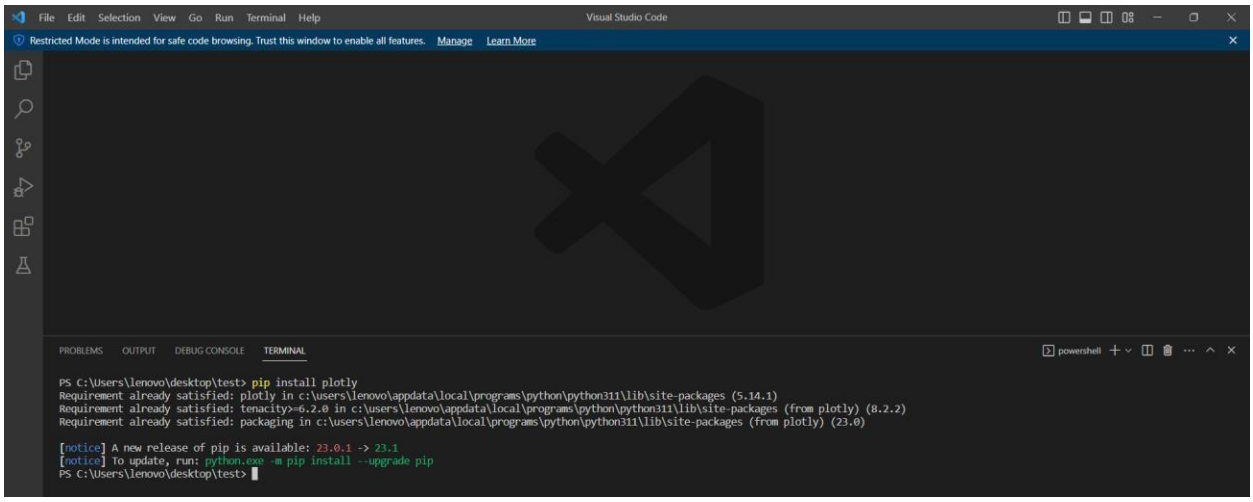
2.4 Seaborn

Seaborn is a Python-based library that works for visualizing the results in an interesting way. There is a range of more methods are considered by the seaborn in comparison with the matplotlib for plotting the graphs. For installing seaborn, open the vs-code and click on terminal and type this command:
pip install seaborn



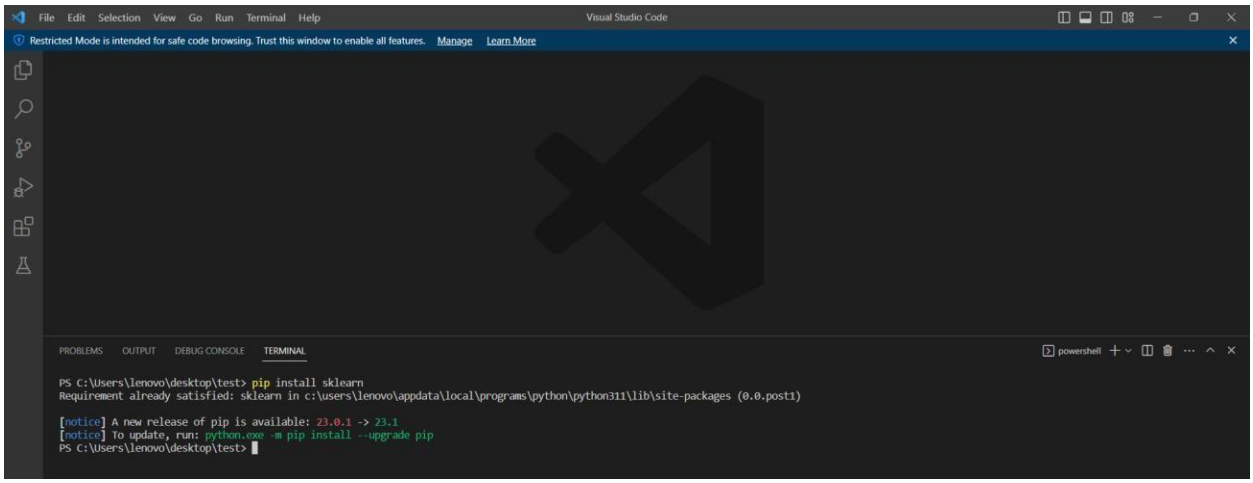
2.5 Plotly

Plotly is a visualization library that facilitates the user for plotting the interactive chart. There is the range of graphs is supported by this library.
For installing plotly, open the vs-code and click on terminal and type this command:
pip install plotly



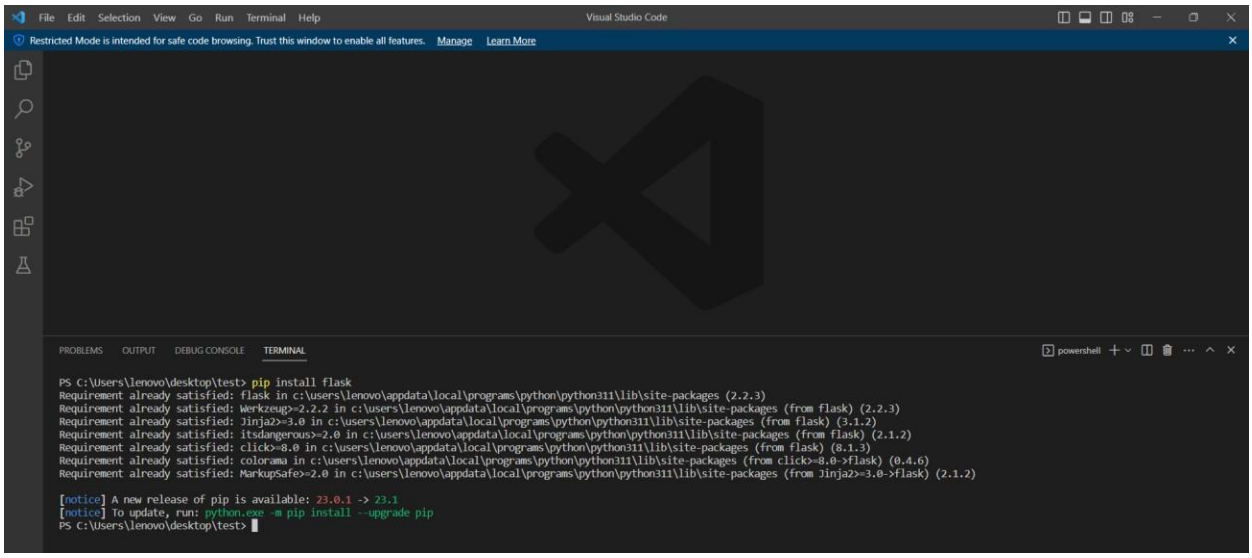
2.6 Sklearn

It is also known as the scikit-learn stands for scientific python and provides a range of facilities in order to build the machine learning model. It is helpful for splitting the dataset in a different manner, model building, and fitting because it is associated with different algorithms such as linear regression, logistic regression, decision trees, random forest, and so on (Scikit-learn, 2023).
For installing sklearn, open the vs-code and click on terminal and type this command:
pip install sklearn



2.7 Flask

Flask is also an open-source framework that provides the API in order to build web applications. It is considered the micro-framework for developing the web app in a quick and simplest manner (Flask, 2023).
For installing flask, open the vs-code and click on terminal and type this command:
pip install flask



3 Configuration steps

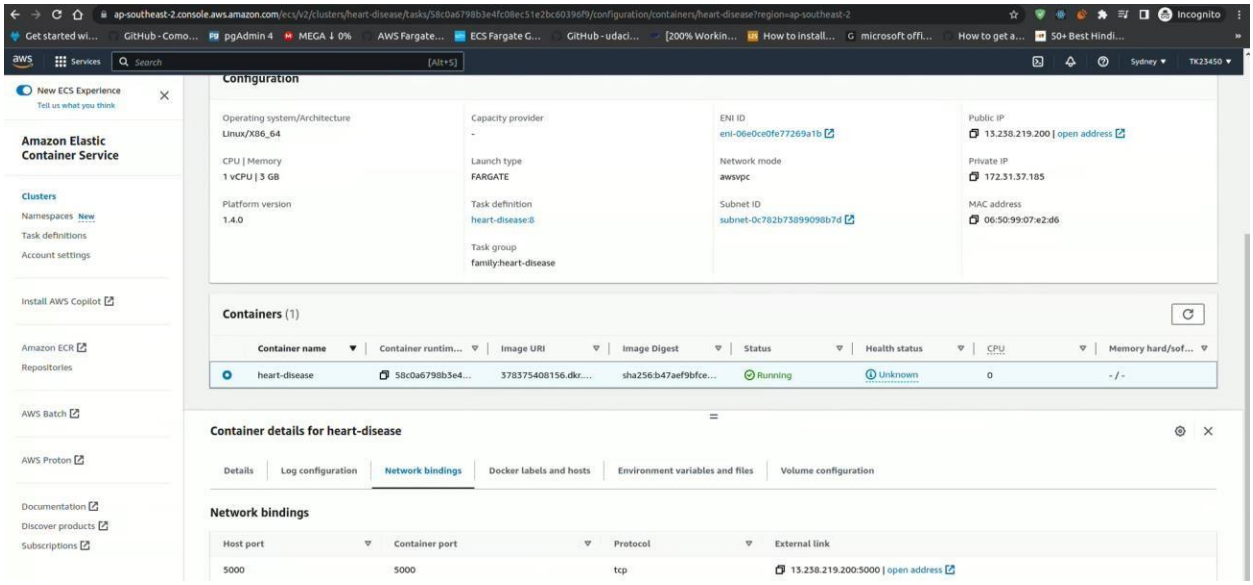
- Developing the Machine learning model

```
44
45  # Train and evaluate model
46  def fit_eval_model(model, train_features, y_train, test_features, y_test):
47      results = {}
48
49      # Train the model
50      model.fit(train_features, y_train)
51
52      # Test the model
53      train_predicted = model.predict(train_features)
54      test_predicted = model.predict(test_features)
55
56      # Classification report and Confusion Matrix
57      results['classification_report'] = classification_report(y_test, test_predicted)
58      results['confusion_matrix'] = confusion_matrix(y_test, test_predicted)
59
60      return results
61
62  # Initialize the models
63  sv = SVC(random_state = 1)
64  lr = LogisticRegression(random_state =1)
65  rf = RandomForestClassifier(random_state = 1)
66  ab = AdaBoostClassifier(random_state = 1)
67  gb = GradientBoostingClassifier(random_state = 1)
68
```

- Creating an API with Flask

```python
@app.route('/')
def home():
    return render_template('Heart Disease Classifier.html')

# Bind predict function to URL
@app.route('/predict', methods =['POST'])
def predict():

    # Put all form entries values in a list
    features = [float(i) for i in request.form.values()]
    # Convert features to array
    array_features = [np.array(features)]
    # Predict features
    prediction = model.predict(array_features)

    output = prediction

    # Check the output values and retrive the result with html tag based on the value
    if output == 1:
        return render_template('Heart Disease Classifier.html',
                               result = 'The patient is not likely to have heart disease!')
    else:
        return render_template('Heart Disease Classifier.html',
                               result = 'The patient is likely to have heart disease!')

if __name__ == '__main__':
#Run the application
    app.run(host='0.0.0.0', debug= True)
```

- Deployment on cloud using EC2.



4 Implementation steps

4.1 Data Processing

```
# Import needed libraries
import re
import pickle
import numpy as np
import plotly as py
import pandas as pd
import seaborn as sns
import plotly.tools as tls
from sklearn.svm import SVC
import matplotlib.pyplot as plt
from plotly.offline import iplot
from sklearn.tree import DecisionTreeClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report, make_scorer
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier, GradientBoostingClassifier
```

[1]

```
# Read data from csv file
dataset = pd.read_csv('data.csv')
dataset.head()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	70	1	4	130	322	0	2	109	0	2.4	2	3	3	2
1	67	0	3	115	564	0	2	160	0	1.6	2	0	7	1
2	57	1	2	124	261	0	0	141	0	0.3	1	0	7	2
3	64	1	4	128	263	0	0	105	1	0.2	2	1	7	1
4	74	0	2	120	269	0	2	121	1	0.2	1	1	3	1

```
dataset.shape
```

```
(270, 14)
```

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270 entries, 0 to 269
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   age       270 non-null    int64  
 1   sex       270 non-null    int64  
 2   cp        270 non-null    int64  
 3   trestbps  270 non-null    int64  
 4   chol      270 non-null    int64  
 5   fbs       270 non-null    int64  
 6   restecg   270 non-null    int64  
 7   thalach   270 non-null    int64  
 8   exang     270 non-null    int64  
 9   oldpeak   270 non-null    float64
 10  slope     270 non-null    int64  
 11  ca        270 non-null    int64  
 12  thal      270 non-null    int64  
 13  target    270 non-null    int64  
dtypes: float64(1), int64(13)
memory usage: 29.7 KB
```

```
#check for the null values
dataset.isnull().sum()
```

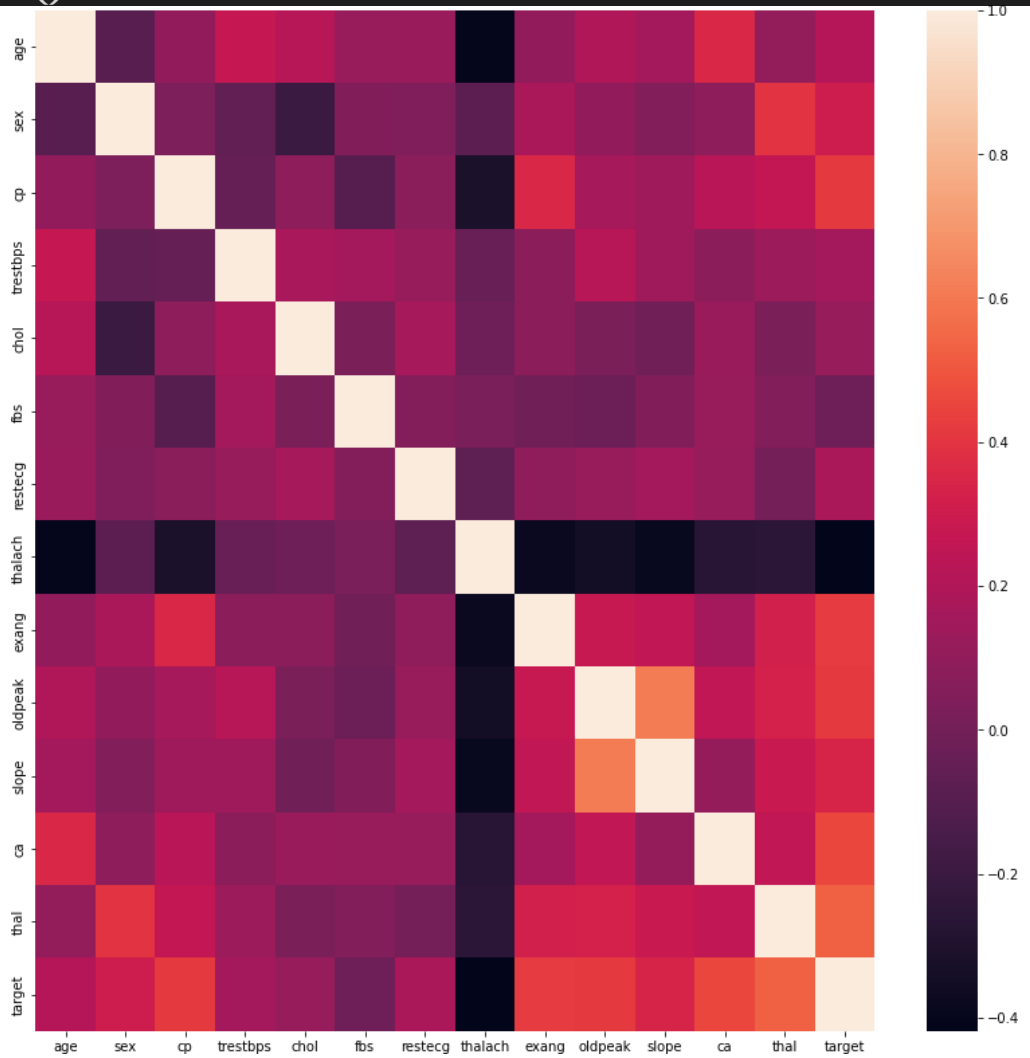
```
age         0
sex         0
cp          0
trestbps    0
chol        0
fbs         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
target      0
dtype: int64
```

```
dataset['target'].value_counts()
dataset['target'].value_counts()/dataset.shape[0]*100
```

```
1    55.555556
2    44.444444
Name: target, dtype: float64
```

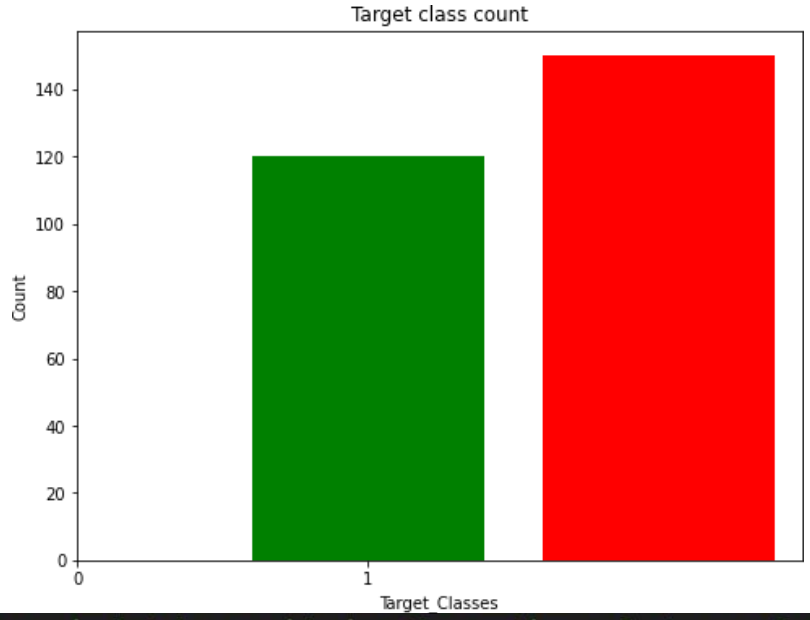
4.2 Data visualization

```
# Correlation Heatmap
plt.figure(figsize=(14, 14))
sns.heatmap(dataset.corr())
plt.show()
```

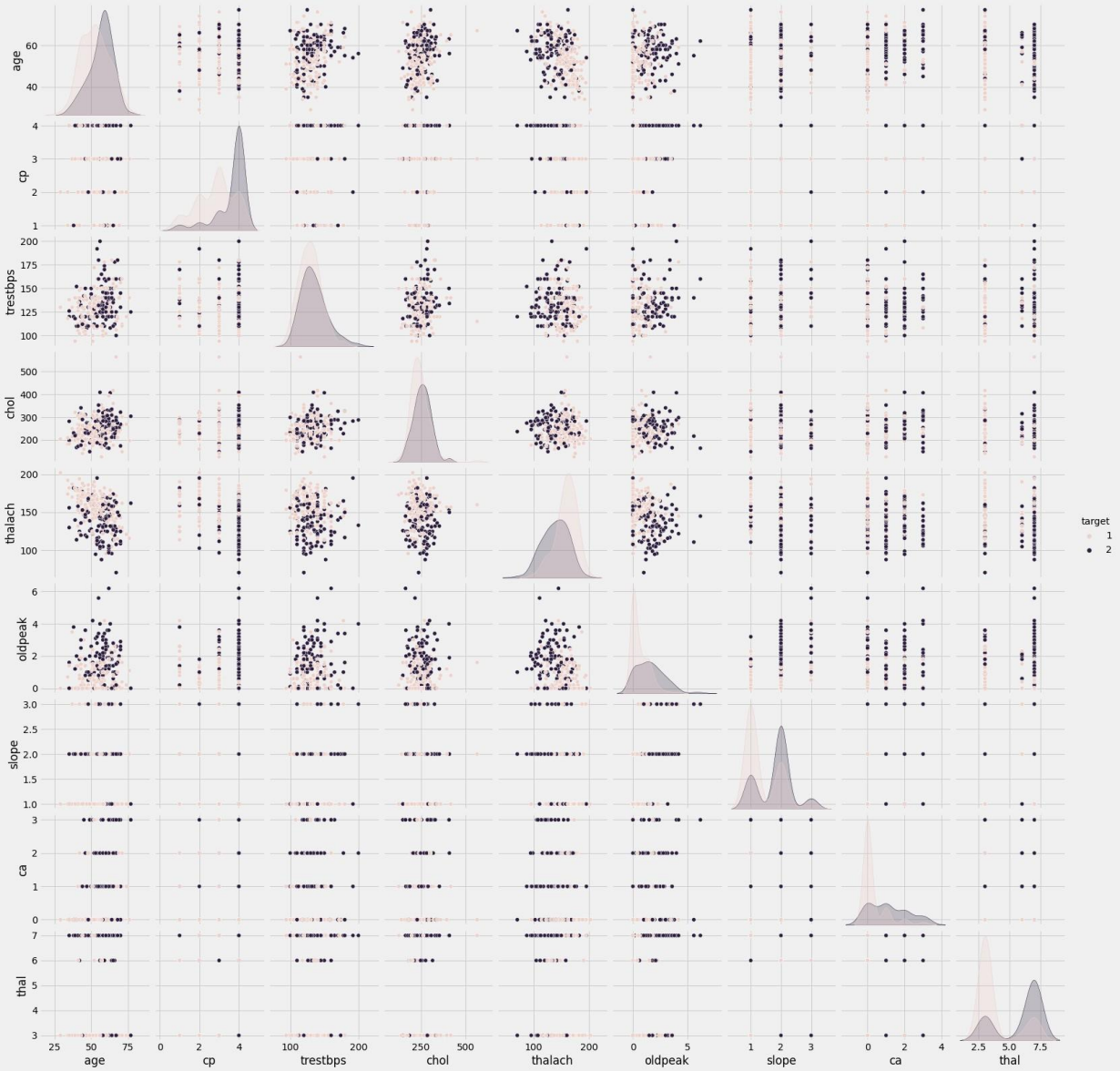



```
# histogram of all all variable for understanding the distribution of data
from matplotlib import rcParams

rcParams['figure.figsize'] = 25, 14
dataset.hist()
```

```
# plotting the count of target class
rcParams['figure.figsize'] = 8,6
plt.bar(dataset['target'].unique(), dataset['target'].value_counts(), color =
['red', 'green'])
plt.xticks([0,1])
plt.xlabel('Target Classes')
plt.ylabel('Count')
plt.title('Target class count')
```

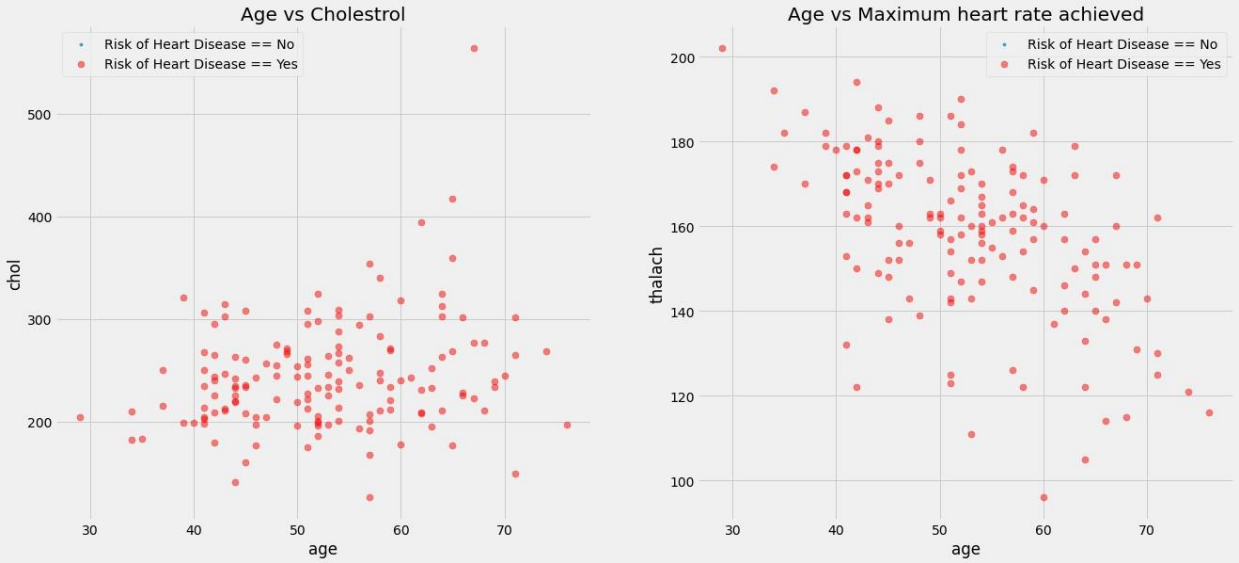



```
# plotting the pairplot by considering the continous features of dataset
plt.style.use('fivethirtyeight')

sns.pairplot(dataset , hue = 'target',
             vars = ['age' , 'cp' , 'trestbps', 'chol' ,'thalach' , 'oldpeak'
,  'slope', 'ca','thal'])
plt.show()
```


age
cp
trestbps
chol
thalach
oldpeak
slope
ca
thal
target
1
2

```
# function for scatter plot
def scatter_plot(x , y , hue , label1 , label2 , s1 = 50 , s2 = 100):

    plt.scatter(x = x ,y = y,s = s1,data = dataset[dataset[hue] == 0] ,label =
label1 , alpha = 0.8 )
    plt.scatter(x = x , y = y , s = s2 , data = dataset[dataset[hue] ==
1],label = label2 , color = 'red',alpha = 0.5 )

    plt.xlabel(x)
    plt.ylabel(y)
    plt.legend()

plt.figure(1 , figsize = (20 ,9))

plt.subplot(1 , 2 , 1)
scatter_plot(x ='age',y ='cp',hue = 'target' , label1 = 'Risk by Heart Disease
== No',label2 = 'Risk by Heart Disease == Yes',s1=10,s2 = 50)
```


```
plt.figure(1 , figsize = (20 ,9))

plt.subplot(1 , 2 , 1)
scatter_plot(x = 'age' , y = 'chol' ,hue = 'target' , label1 = 'Risk of Heart
Disease == No' ,
             label2 = 'Risk of Heart Disease == Yes' ,s1 = 10 , s2 = 50)
plt.title('Age vs Cholestrol')
```



5 Results

Heart Disease Test Form

Age

Sex

-- Select an Option --

Chest Pain Type

-- Select an Option --

Resting Blood Pressure in mm Hg

Serum Cholestoral in mg/dl

Fasting Blood Sugar > 120 mg/dl

-- Select an Option --

Resting ECG Results

-- Select an Option --

Maximum Heart Rate

ST Depression Induced

Exercise Induced Angina

-- Select an Option --

Slope of the Peak Exercise ST Segment

-- Select an Option --

Number of Vessels Colored by Flourosopy

-- Select an Option --

Thalassemia

-- Select an Option --

Result

Result

The patient is not likely to have heart disease!

6 Conclusion

The research represents the manual of configuration in order to execute the machine learning system. The study represents each stage and tool that is required for the development of a machine learning system. It also represents every aspect of running the model in the local path. This manual also defines about the different technologies which were used in the research.

7 References

Python. (2023). *Welcome to Python.org*. [online] Available at: https://www.python.org/ [Accessed 17 Apr. 2023].

Scikit-learn. (2023). *scikit-learn: machine learning in Python — scikit-learn 1.2.2 documentation*. [online] Available at: https://scikit-learn.org/stable/ [Accessed 17 Apr. 2023].

Flask. (2023). *Welcome to Flask — Flask Documentation (2.2.x)*. [online] Available at: https://flask.palletsprojects.com/en/2.2.x/ [Accessed 17 Apr. 2023].