CODE

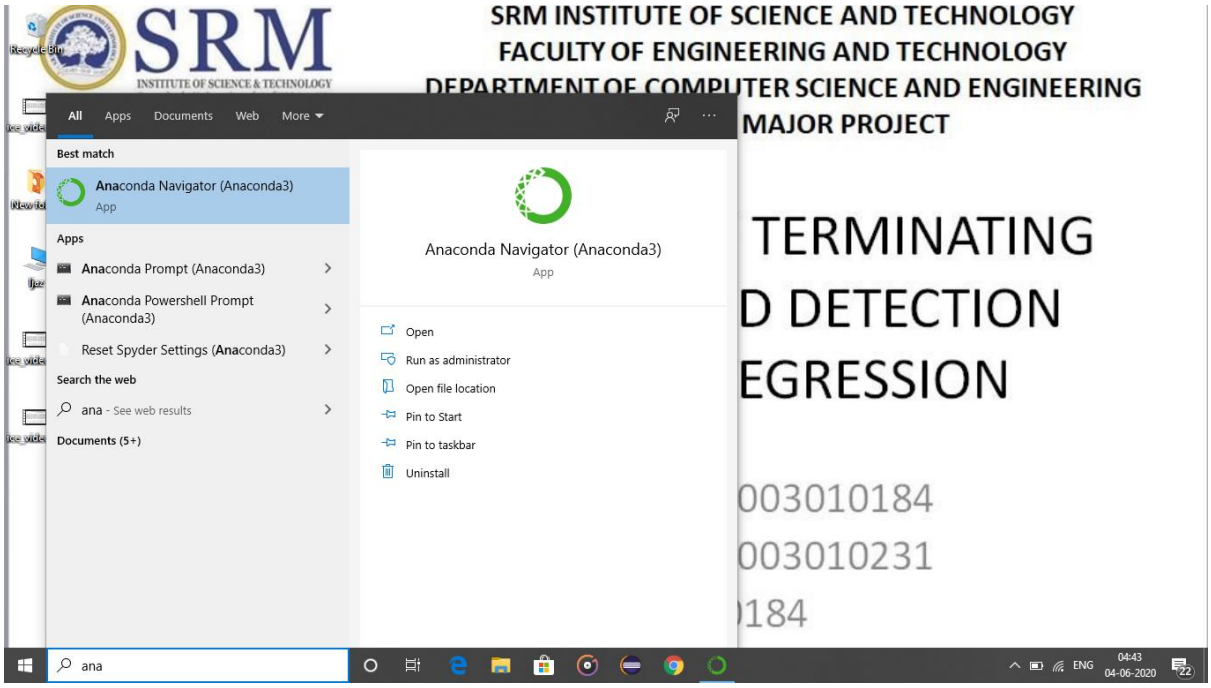
```
import pandas as pd
import numpy as np
import random
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split

urls_data = pd.read_csv("urldata.csv")

type(urls_data)

urls_data.head()

def makeTokens(f):
    tkns_BySlash = str(f.encode('utf-8')).split('/')  //make tokens after splitting by slash
    total Tokens = []
    for i in tkns_BySlash:
        tokens = str(i).split('-')     //make tokens after splitting by dash
        tkns ByDot = []
        for j in range(0,len(tokens)):
            temp_Tokens = str(tokens[j]).split('.')    //make tokens after splitting by dot
            tkns ByDot = tkns ByDot + temp Tokens
        total Tokens = total Tokens + tokens + tkns ByDot
    total_Tokens = list(set(total_Tokens))  //remove redundant tokens
    if 'com' in total_Tokens:
        total Tokens.remove('com')   #removing .com because it occurs a lot of times
    return total_Tokens

y = urls_data["label"]

vectorizer = CountVectorizer(tokenizer=makeTokens)

X = vectorizer.fit_transform(url_list)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2
, random_state=42)

logit = LogisticRegression()

logit.fit(X_train, y_train)

print("Accuracy ",logit.score(X_test, y_test))

import tkinter as tk
root = tk.Tk()
```

```
canvas1 = tk.Canvas(root,width = 800, height = 600)
canvas1.pack()

label1 = tk.Label(root, text='Phishing Website Prediction')
label1.config(font=('helvetica', 24))
canvas1.create_window(400, 50, window=label1)

label2 = tk.Label(root, text='Paste Here:')
label2.config(font=('helvetica', 20))
canvas1.create_window(400, 200, window=label2)


entry1 = tk.Entry (root)
canvas1.create_window(400, 280, window=entry1)


def getSquareRoot():
    x1 = entry1.get()

    label3 = tk.Label(root, text= 'The Result For ' + x1 + ' is:',font=
('helvetica', 20))
    canvas1.create_window(400, 420, window=label3)

    res = vectorizer.transform([x1])
    prediction = logit.predict(res)

    label4 = tk.Label(root, text= prediction,font=('helvetica', 20, 'bo
ld'))
    canvas1.create_window(400, 460, window=label4)

button1 = tk.Button(text='Find', command=getSquareRoot, bg='brown', fg=
'white', font=('helvetica', 18, 'bold'))
canvas1.create_window(400, 350, window=button1)

root.mainloop()
```

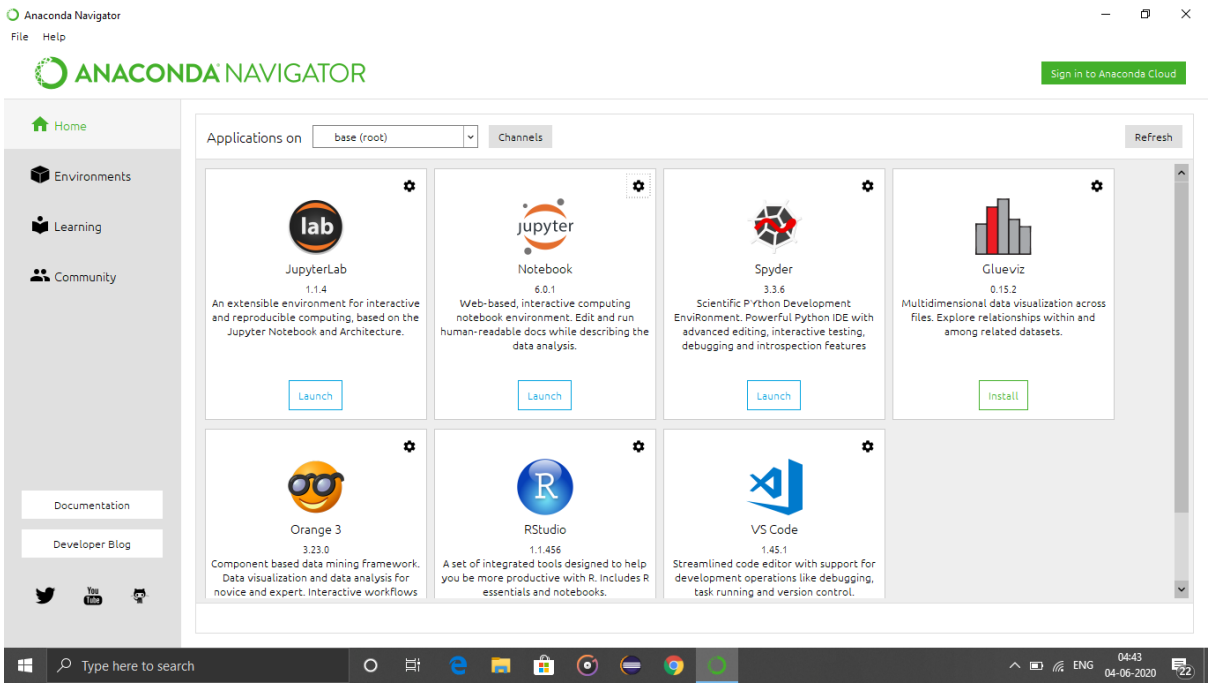
SOFTWARE USED: Anaconda navigator

IMPLIMENTATION

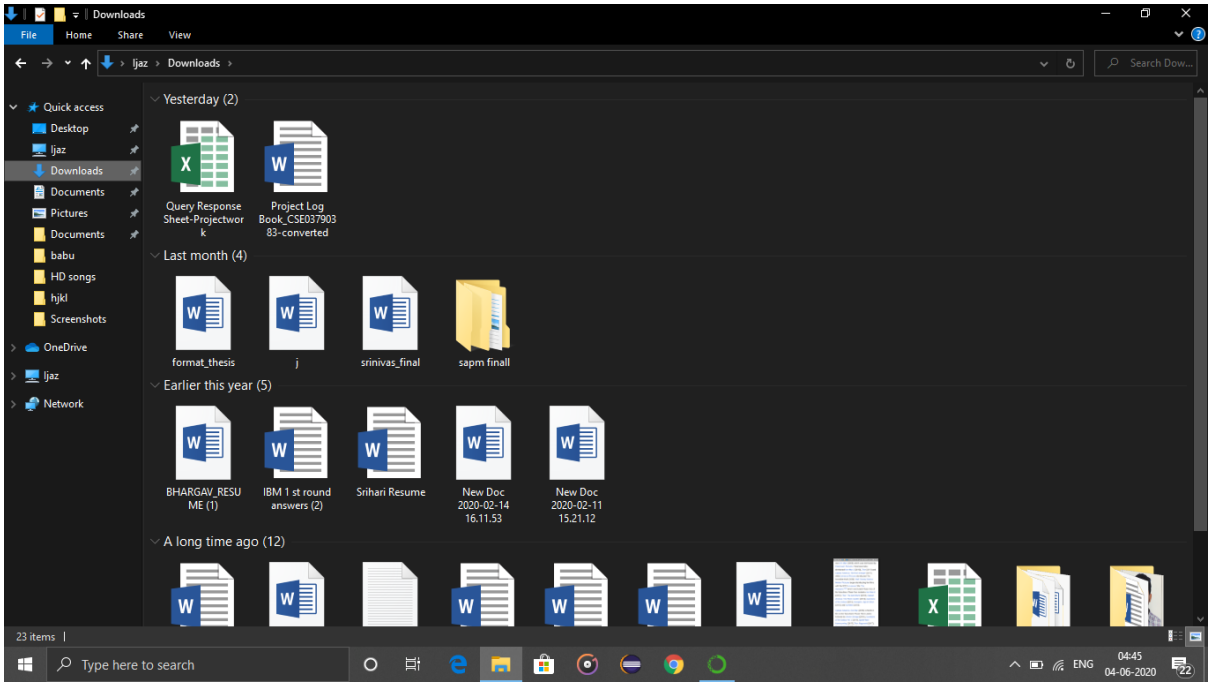
Step 1:Download and install anaconda navigator



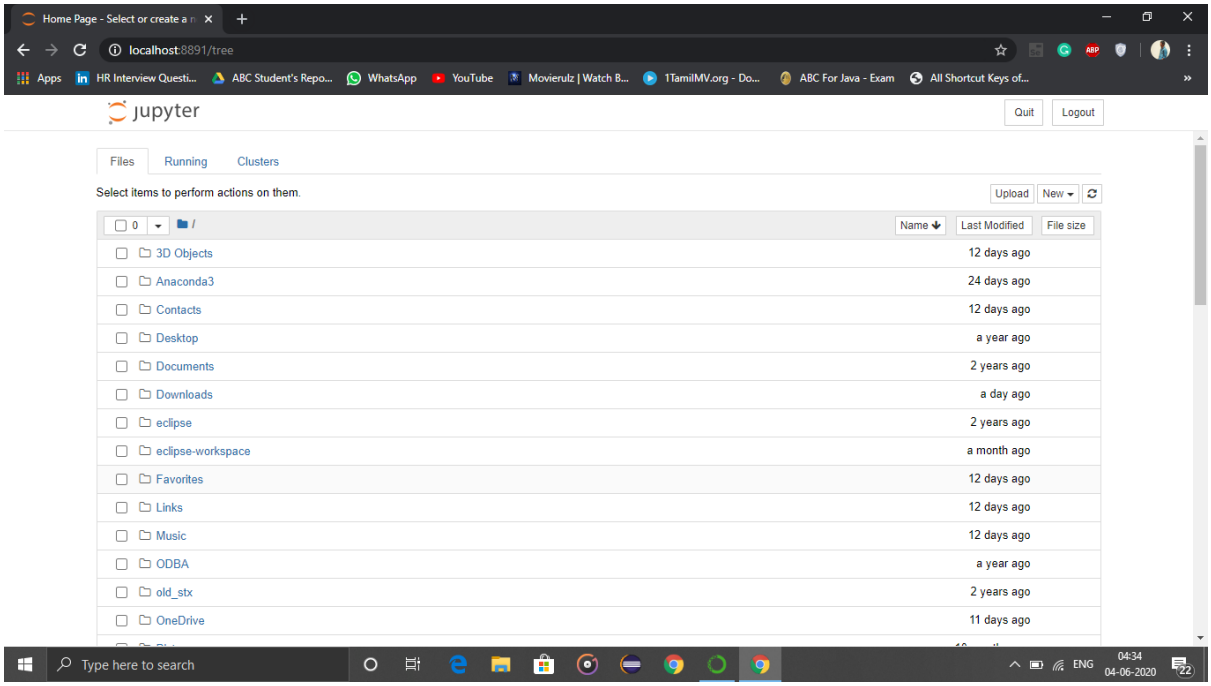
Step 2:Install Jupiter notebook and launch

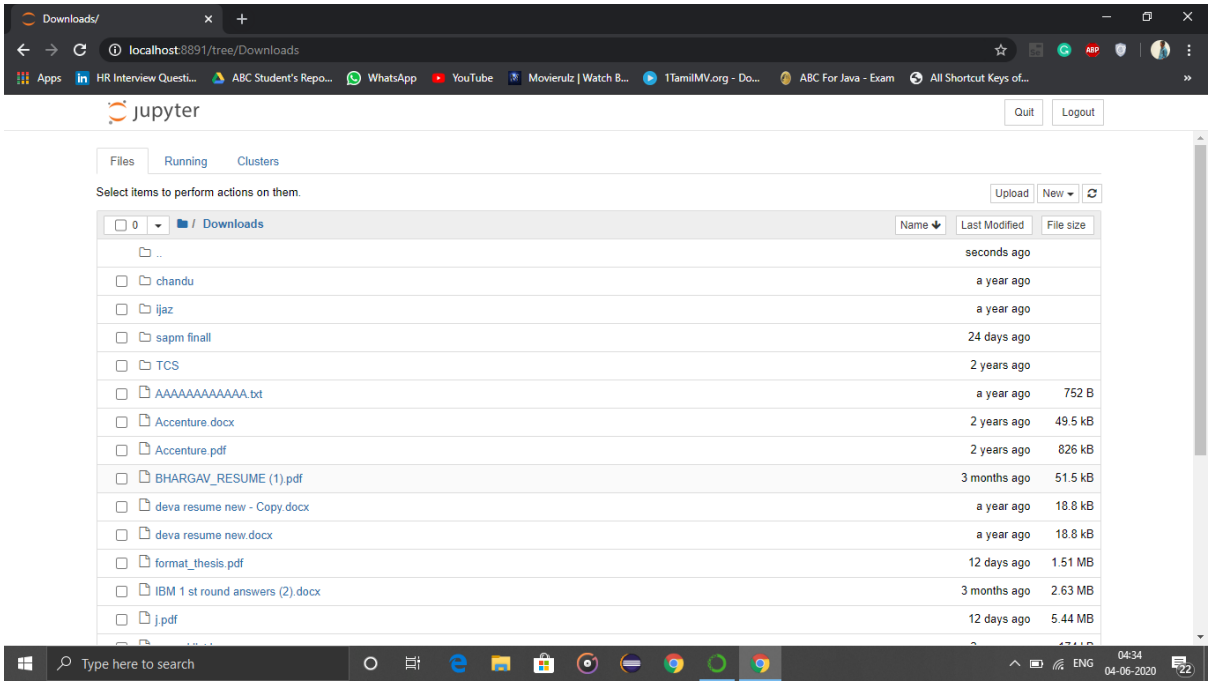


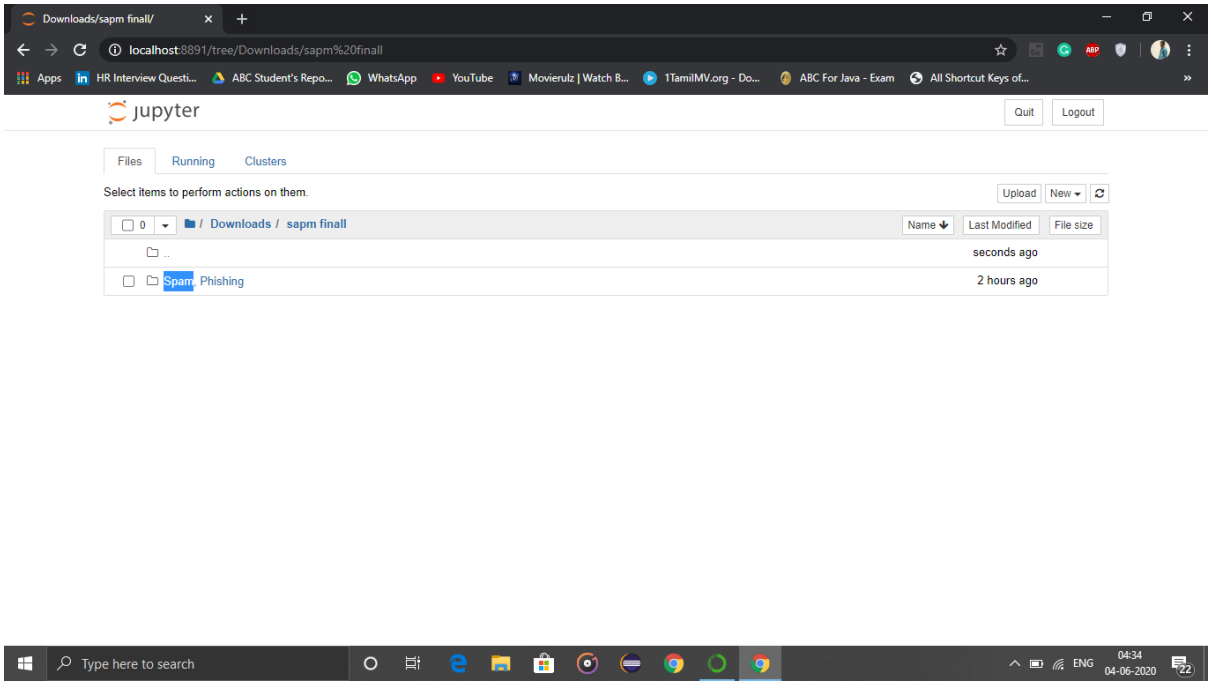
Step 3:Save the file of code and data set in your desktop here it is saved as sapm final

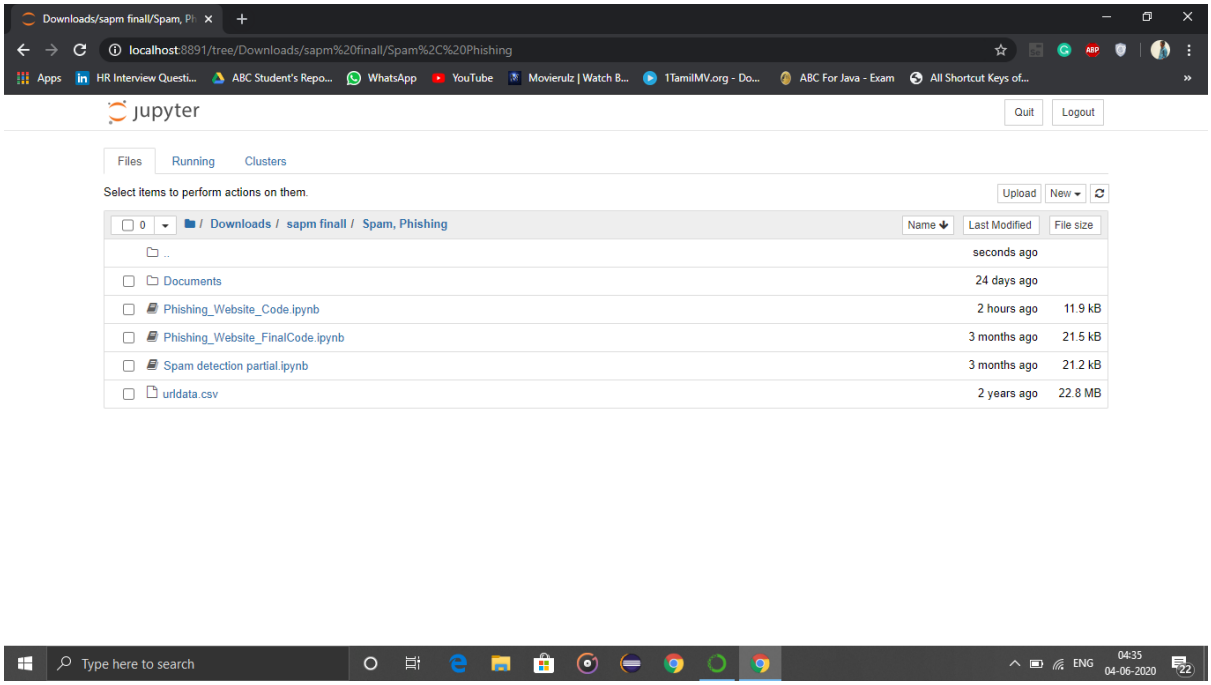


Step 4:When the chrome open a localhost link go to the file stored location and open the code file

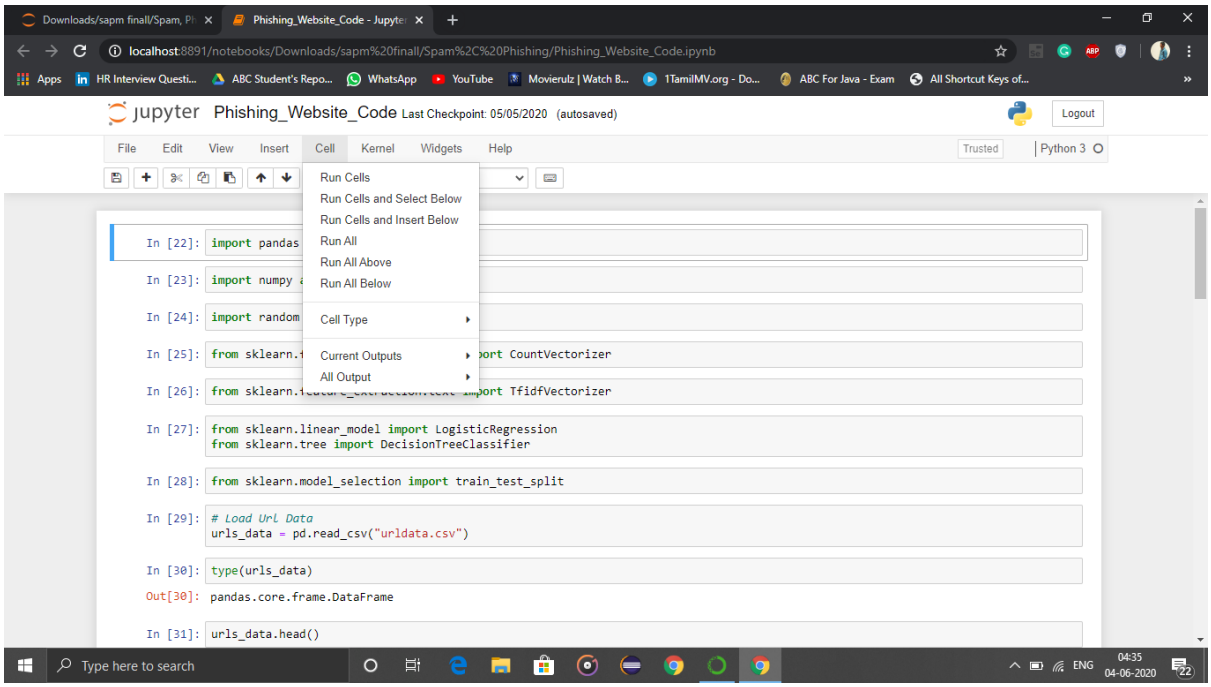



Downloads/
localhost:8891/tree/Downloads
jupyter
Quit
Logout
Files
Running
Clusters
Select items to perform actions on them.
Upload
New
/ Downloads
Name
Last Modified
File size
seconds ago
chandu a year ago
ijaz a year ago
sapm finall 24 days ago
TCS 2 years ago
AAAAAAAAAAA.txt a year ago 752 B
Accenture.docx 2 years ago 49.5 kB
Accenture.pdf 2 years ago 826 kB
BHARGAV_RESUME (1).pdf 3 months ago 51.5 kB
deva resume new - Copy.docx a year ago 18.8 kB
deva resume new.docx a year ago 18.8 kB
format_thesis.pdf 12 days ago 1.51 MB
IBM 1 st round answers (2).docx 3 months ago 2.63 MB
j.pdf 12 days ago 5.44 MB
Type here to search
ENG
04:34
04-06-2020

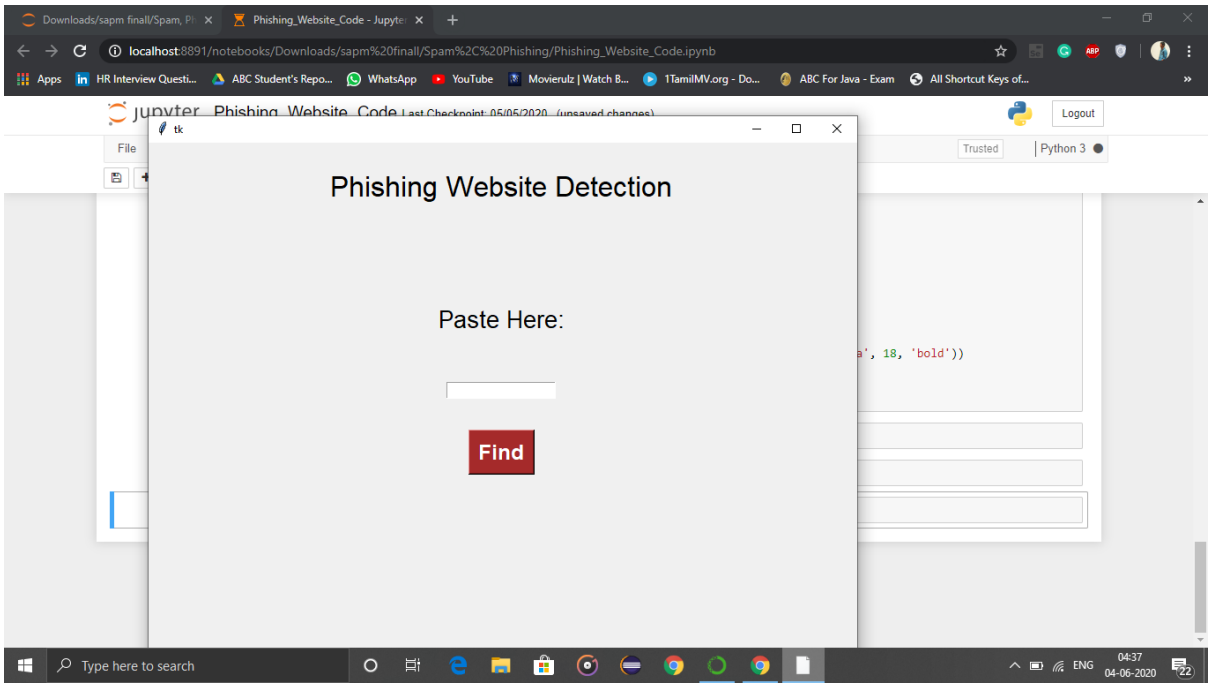

Downloads/sapm finall/
localhost:8891/tree/Downloads/sapm%20finall
jupyter
Quit
Logout
Files
Running
Clusters
Select items to perform actions on them.
Upload
New
/ Downloads / sapm finall
Name
Last Modified
File size
seconds ago
Spam_Phishing 2 hours ago
Type here to search
ENG
04:34
04-06-2020



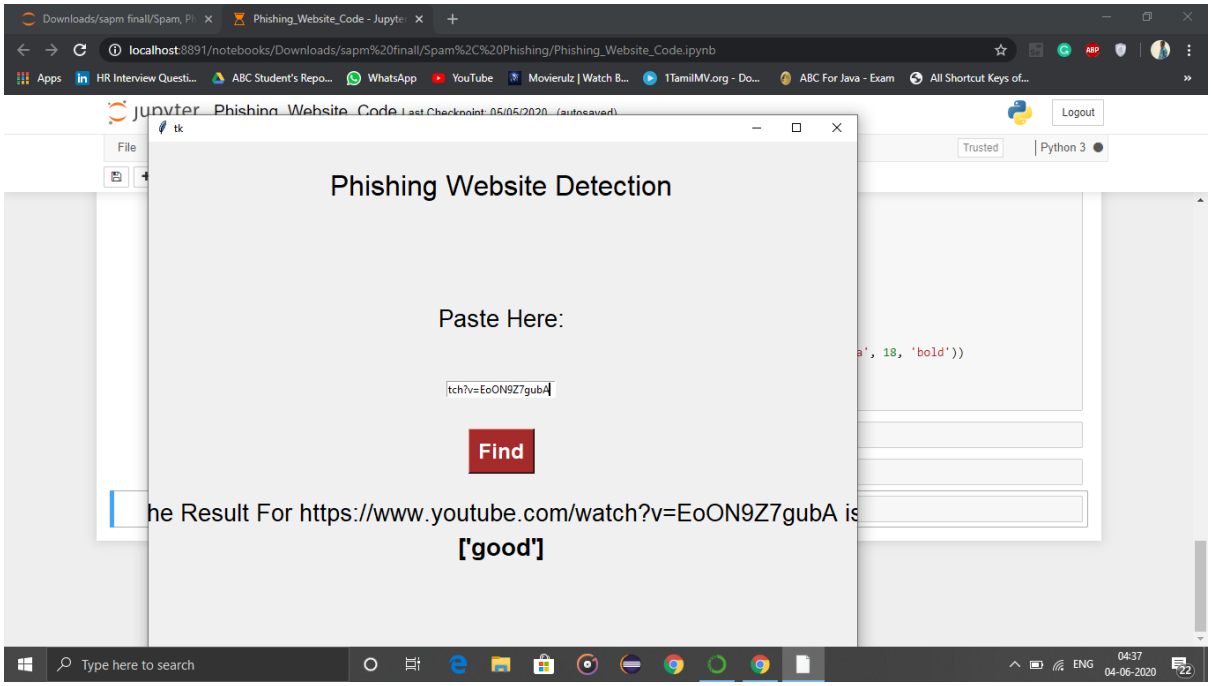
Step 5:When the code opens at the top the you can see “cell” option click on that choose run all



Step 6:Wait until a popup webpage it takes time



Step 7:Paste the link you wanted to check in the given blank box and click the below button



Step 8:You can check the result there it self.to find for the other links close the webpage and repeat from step 5.