**VERZEO MAJOR PROJECT**

****

**Project Title**

**Indian Human Image Classification Model**

**Mentor:** Verzeo

**Submitted by-**

Members of Group(B9) :

1. Anshul Kumar
2. Harshita Anand
3. Yash Tandon
4. Ankita Chand
5. Akash Joy
6. Sudhamoy Dey
7. Madhumitha V
8. Anushka singh
9. Sameer Sharma Chippada

10. Srilekha.ch

11.Ruchitha.G

12. Pakshan Badhniwalla

13. Tanishka Kolse

**STEP 1: Data collection**

**Strategy used:** Web scraping

**Tools used:** Selenium

**Installation of selenium:-**

1. **For Ubuntu**

**A1 -** pip3 install selenium

**A2 -** Download the selenium firefox webdriver (geckodriver):

Link: <https://github.com/mozilla/geckodriver/releases>

**A3** - Move the firefox webdriver to your bin folder:

Command : sudo mv geckodriver /usr/local/bin

**A4** - Set the PATH:

export PATH=$PATH:/usr/local/bin/geckodriver

**References** : <https://www.youtube.com/watch?v=xgsFwaw9W4Q>

1. **For Windows**

**References:** <https://www.youtube.com/watch?v=VYHxIUnvmpQ>

**Script for scraping:**

1. **For Ubuntu**

**A1-** importing what is necessary

from selenium import webdriver

from selenium.webdriver.common.keys import Keys

import time

**A2-** Setting webdriver

driver = webdriver.Firefox() // browser can be changed

driver.get('<https://www.google.com/>') // this link can be of any site

**A3-** Using xpath to add input (link in xpath will vary)

box=driver.find\_element\_by\_xpath('/html/body/div[1]/div[3]/form/div[1]/div[1]/div[1]/div/div[2]/input')

box.send\_keys('human')

box.send\_keys(Keys.ENTER)

driver.find\_element\_by\_xpath('/html/body/div[7]/div/div[4]/div/div[1]/div/div[1]/div/div[2]/a').click()

**A4-** Keep scrolling (not that important.code May vary for sites)

*Will keep scrolling down the webpage until it cannot scroll no more*

last\_height = driver.execute\_script('return document.body.scrollHeight')

while True:

driver.execute\_script('window.scrollTo(0,document.body.scrollHeight)')

time.sleep(2)

new\_height = driver.execute\_script('return document.body.scrollHeight')

Try:

driver.find\_element\_by\_xpath('//\*[@id="islmp"]/div/div/div/div/div[5]/input').click()

time.sleep(2)

except:

pass

if new\_height == last\_height:

break

last\_height = new\_height

**A5**- Download images(taking screenshots)

for i in range(1, 11):

Try:

driver.find\_element\_by\_xpath('/html/body/div[2]/c-wiz/div[3]/div[1]/div/div/div/div/div[1]/div[1]/div['+str(i)+']/a[1]/div[1]/img ').screenshot('human ('+str(i)+').png')

except:

pass

**References:**

Github: <https://gist.github.com/zita9999/e57e64a8eb550935c0b64d9c1763693c>

1. **For windows**

References:

Youtube:

<https://www.youtube.com/watch?v=V_MV5EsdKRc&t=704s>

Github:

<https://gist.github.com/zita9999/e57e64a8eb550935c0b64d9c1763693c>

**Summary**

We will divide aur work in downloading images(screenshots) from different sites using selenium and will save them on drive under appropriate directory.