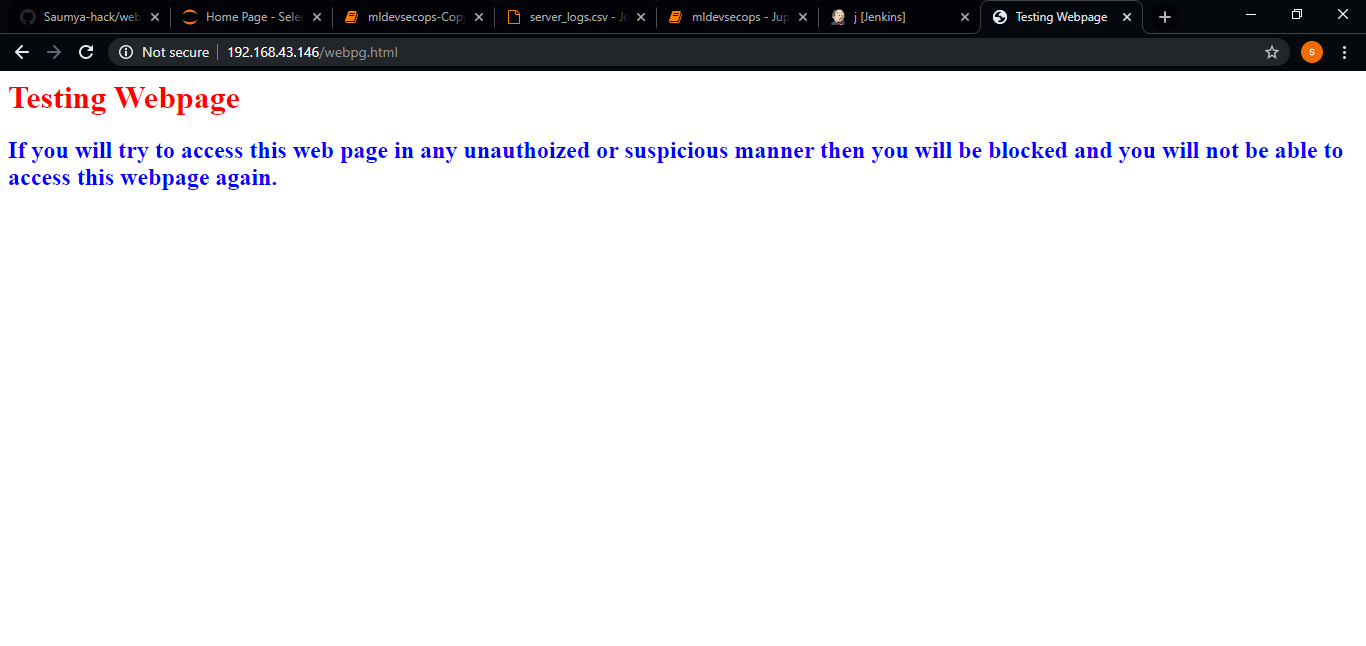
**Creating an automated system which blocks an IP in case of some suspicious activity from that IP (MLDevSecOps)**

***This project is completed by* ‘Saumya Sharma’ *and* ‘Rahul Pareek’.**

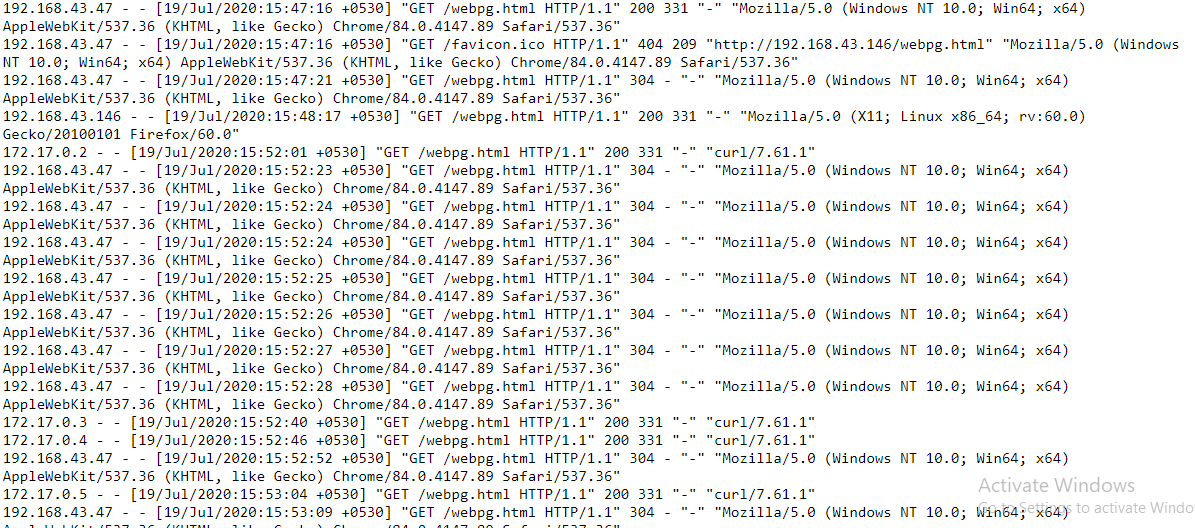
**Our Web page:**

Initially our webpage is working properly.



**Logs generated:**

These are the starting rows of our logs generated in ‘var/log/httpd/access\_log’ file.



**Dockerfile Of Docker Container:**

This is the dockerfile of our docker container in which we will run our ML model program. We have installed python and some required libraries of Python required for our purpose.



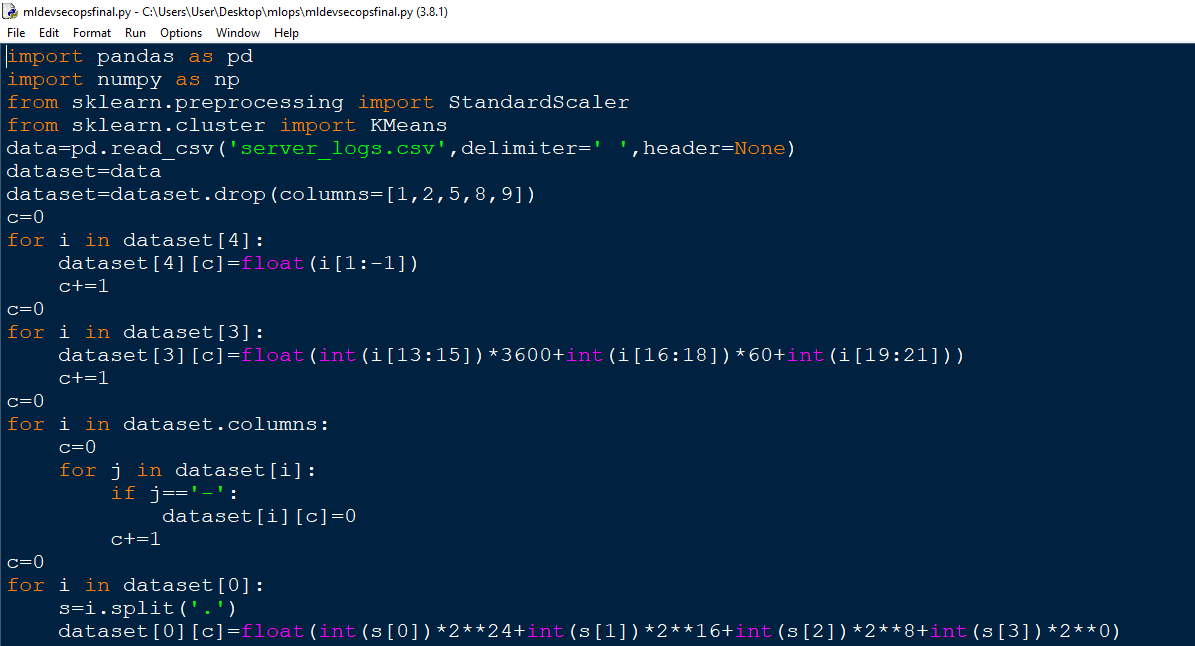
We will build the image using ‘docker build’ command.

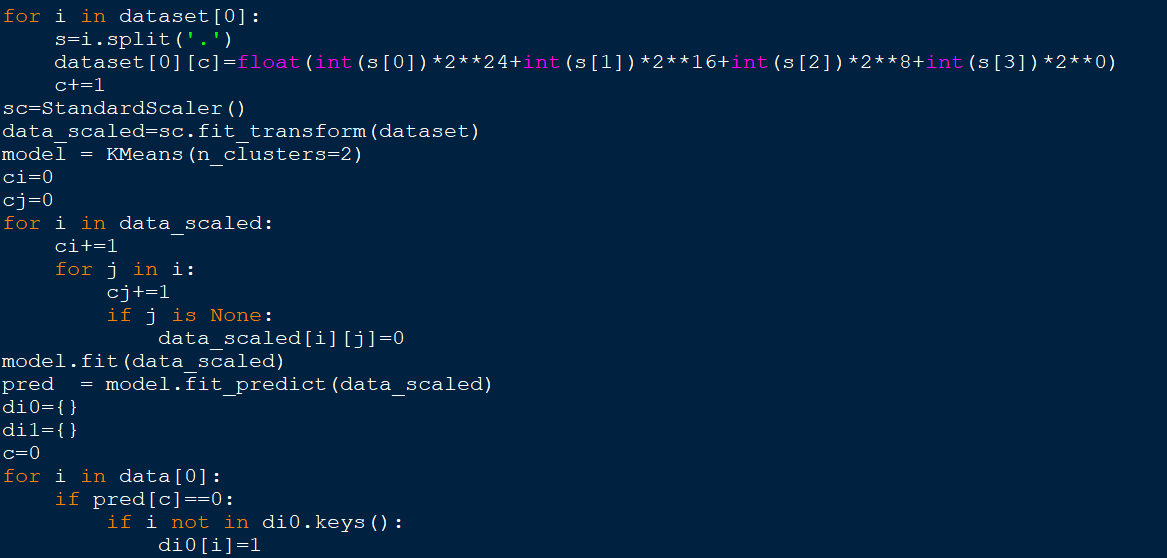


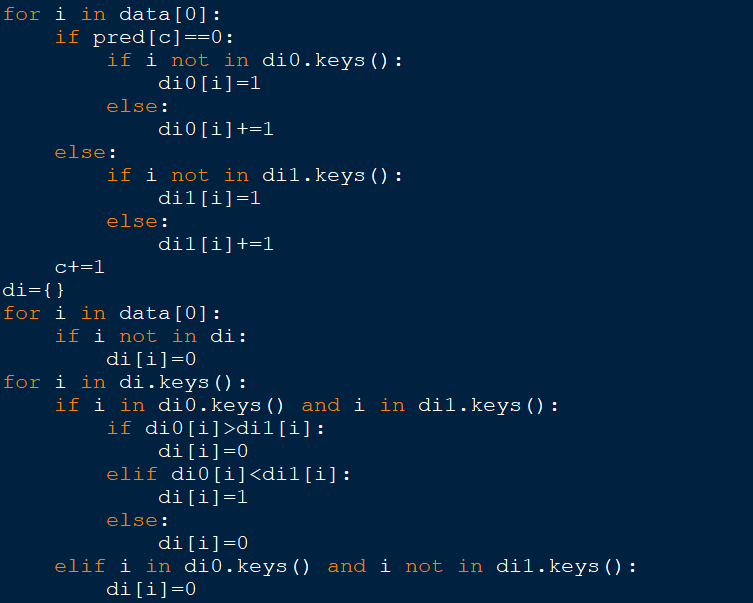
Now, till here, we have created a webpage, we have accessed that webpage with different IP’s and we have now generated logs for that webpage. Also, we have created a dockerfile and created it’s image which we will use in Jenkins job. Now we will see our ML model code.

**ML model Code:**

This is the code of our ML model.



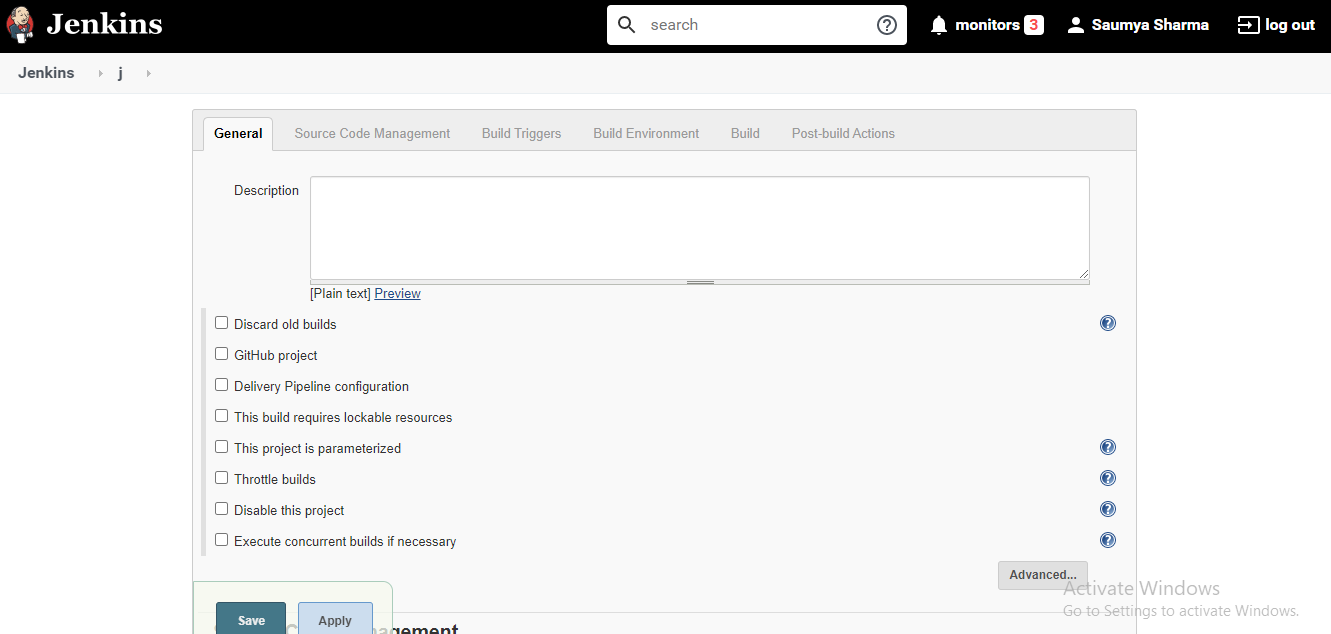


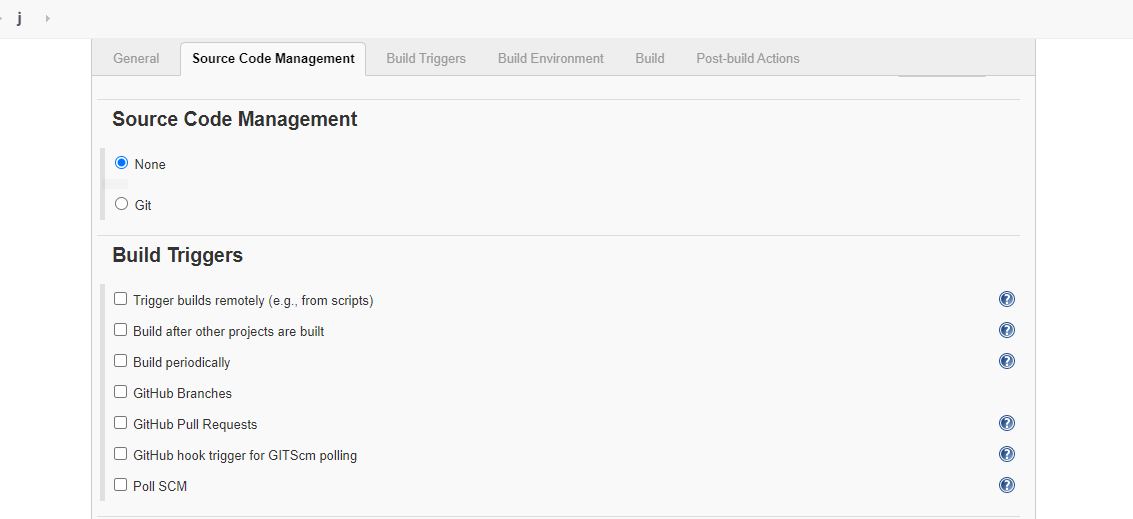


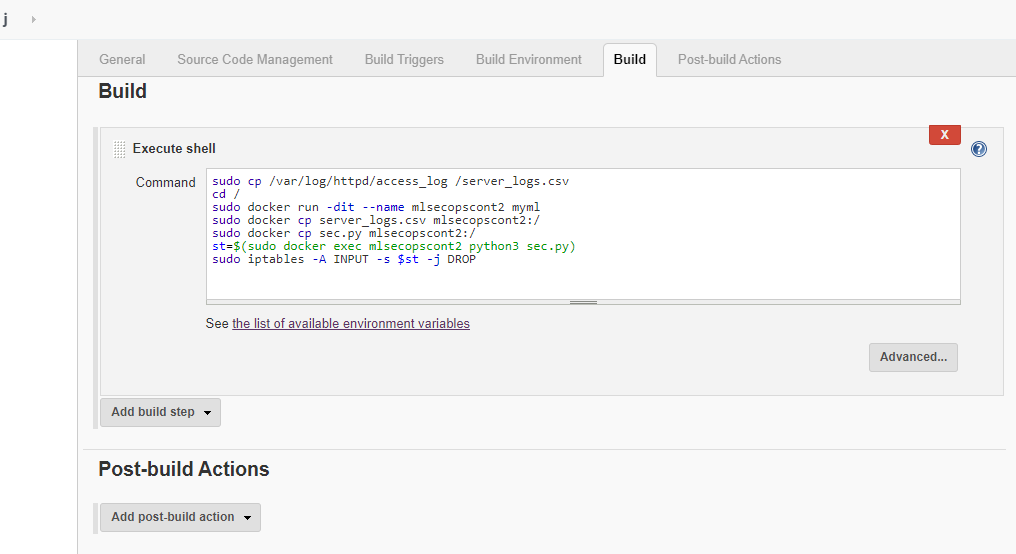


**Jenkins job:**

This is the only jenkins job required for our project. We have named this job as “j”.







**sudo cp /var/log/httpd/access\_log /server\_logs.csv**

This will copy ‘access\_log’ file of our webpage to the ‘/’ directory with ‘server\_logs.csv’ name

**cd /**

By this command we will now move to ‘/’ directory

**sudo docker run -dit –name mlsecopscont2 myml**

This will launch a docker container wih name ’mlsecopscont2’ of ‘myml’ image which we have created above

**sudo docker cp server\_logs.csv mlsecopscont2:/**

This will copy our ‘server\_logs.csv’ file to the ‘/’ directory of our docker container

**sudo docker cp sec.py mlsecopscont2:/**

This will copy our ‘sec.py’ file to the ‘/’ directory of our docker container. Actually our ML model file is with name ‘sec.py’ which we have created in our Redhat system. It contains same code as we have stated above.

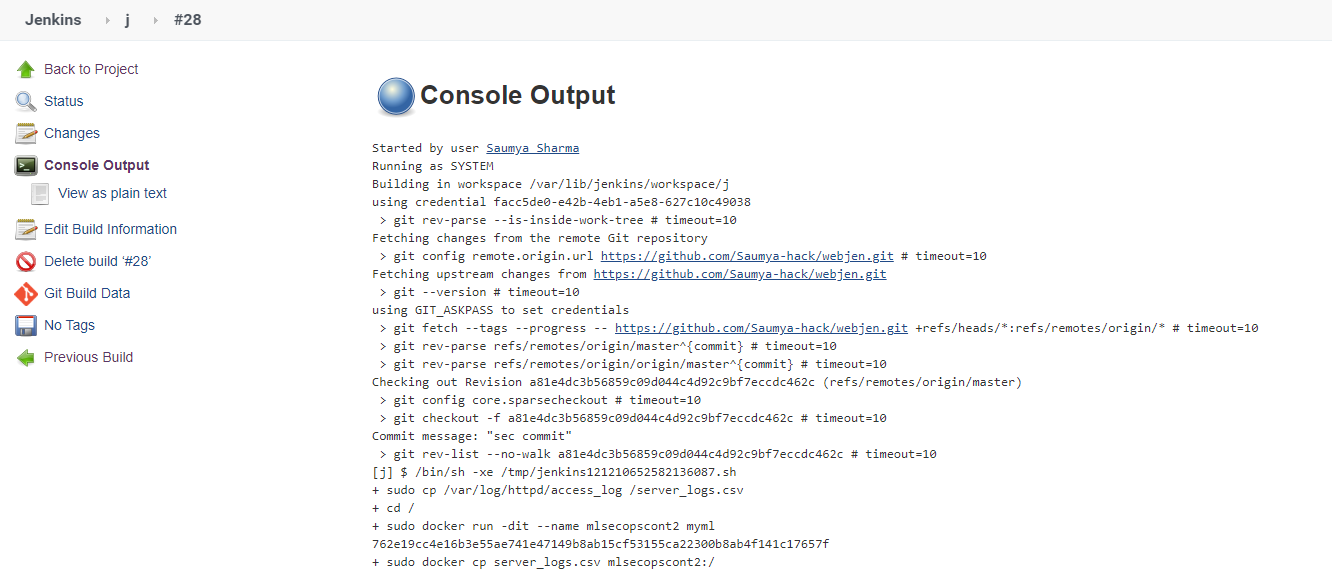
**st=$(sudo docker exec mlsecopscont2 python3 sec.py)**

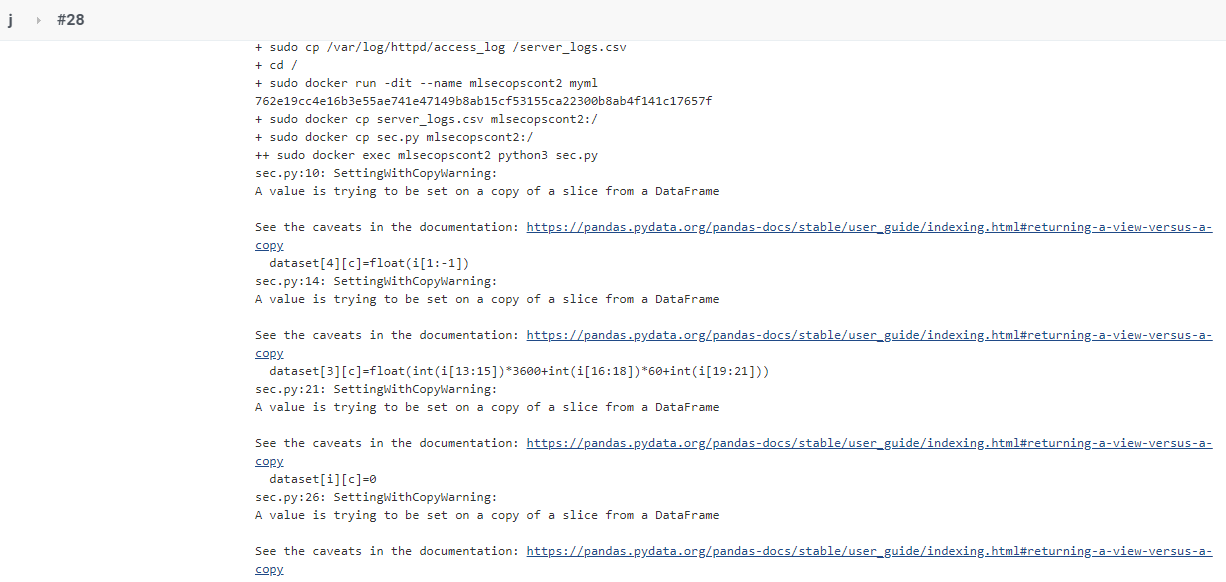
This will run the ‘sec.py’ file in the container and then store it’s output in ‘st’ variable.

**sudo iptables –A INPUT –s $st –j DROP**

This will block the IP which is triggering our webpage again and again in suspicious way. Actually ‘$st’ contains output of ‘sec.py’ file and the output of ‘sec.py’ file is the IP’s which we need to block.

**Jenkins Job Console Output:**

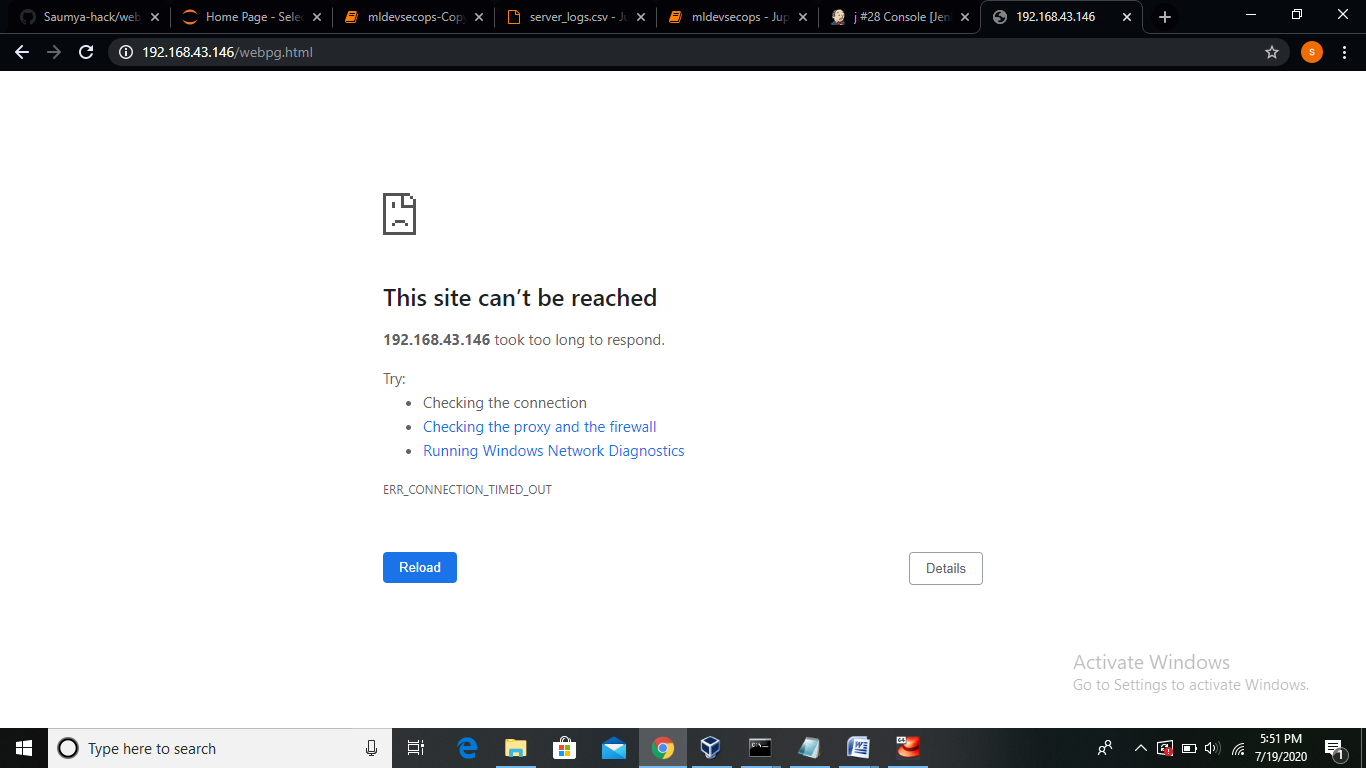






We can see in the above image that our job is executed successfully.

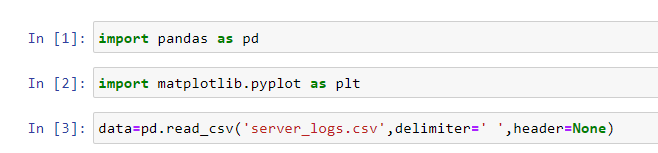
In following image we can see that now, we are not able to access that webpage, which we were able to access before.



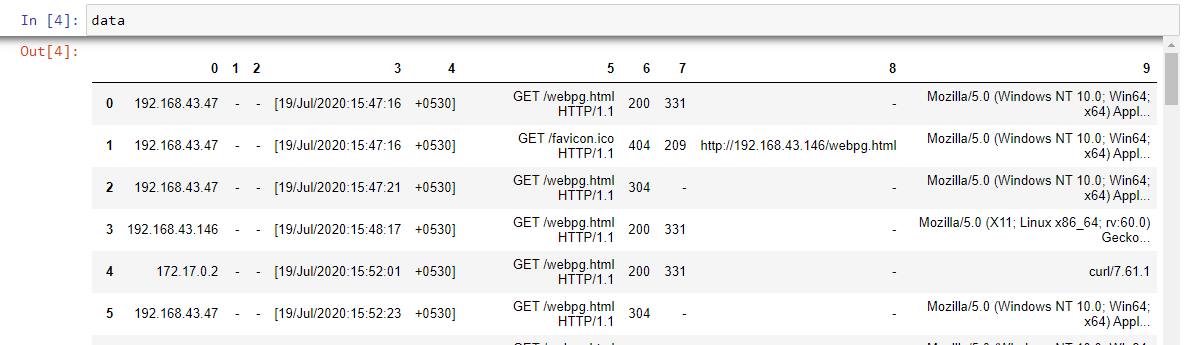
**ML model code explanation:**

This is the “jupyter notebook” code. This code is almost similar to the code of the above python file. But here, we have added little bit more code to explain the concept properly.

Firstly, we have imported ‘pandas’ and ‘matplotlib.pyplot’ library of python. Then we have stored the ‘server\_logs.csv’ file in the data variable delimited by space(‘ ’).

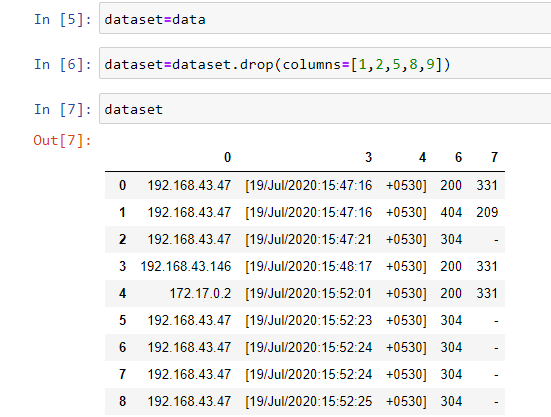


We can see here that our ‘server\_logs.csv’ data is stored in ‘data’ variable.



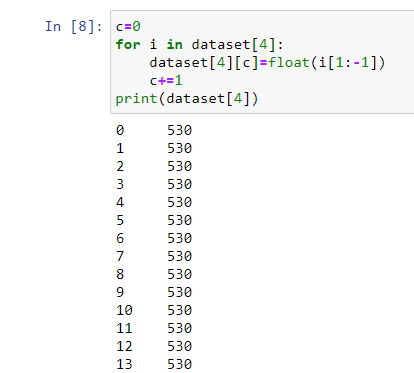
Now in the next image we can see that, here we have copied our ‘data’ variable to ‘dataset’ variable because from here we will manipulate our ‘dataset’ variable and our original data will remain stored in ‘data’ variable.

Then we will do some feature engineering and perform some data preparation. We will find the features which we can eliminate from ‘dataset’. As in column ‘1’,’2’,’8’, we have many unknown values. So we can eliminate these columns. Also in columns ‘5’ we have same data values in all entries, it means it will not make any effect on our ML model, so we can also eliminate it. Now, in column ‘9’ also we have that type of that which will not affect much our ML model. So we can remove this column. So our final dataset is as shown here.

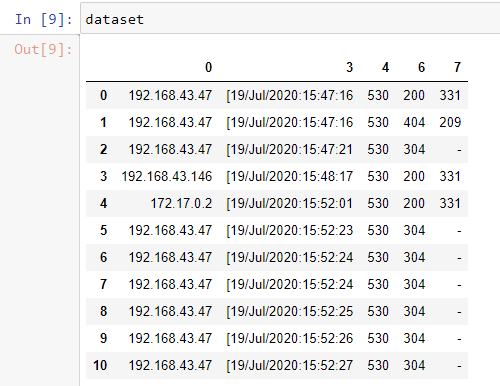


We know that our ML model will be very intelligent if we will have integers only integers in our dataset. So now we have to remove symbols in the data from our dataset or we have to convert some data into some numbers.

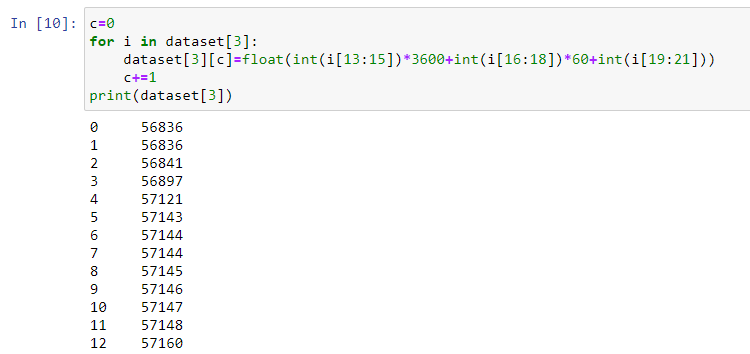
Firstly we are considering column ‘4’. We have data like ‘+0530]’ which helps us to know abou the time. So we have removed ‘]’ and ‘+’. So we will remain with ‘530’ which is now float after conversion.



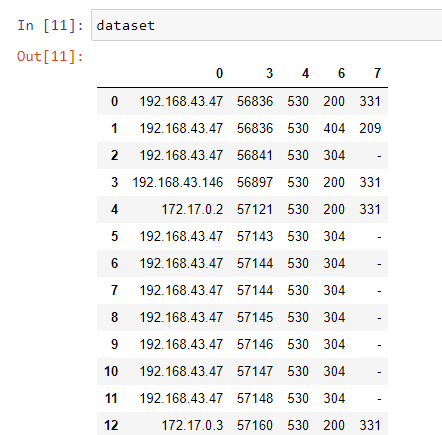
After now our dataset will look like the following.



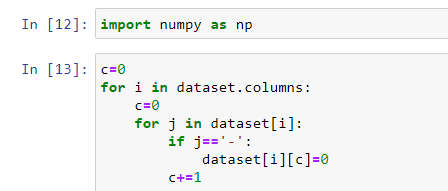
Now, we will pick column ‘3’. Since this log dataset contains contain data of same day. Now, we can convert time in seconds. So we have removed date and then converted time in seconds.



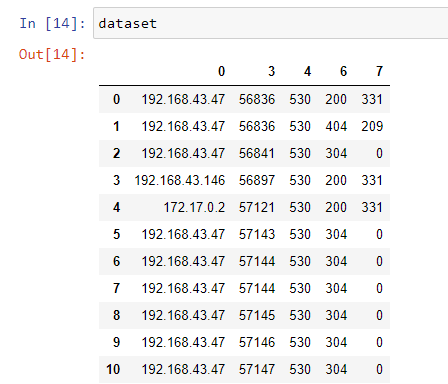
Now our dataset will look like this.



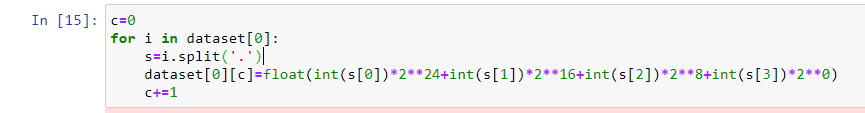
Now wherever we have ‘-‘ in dataset, there we will put ‘0’.



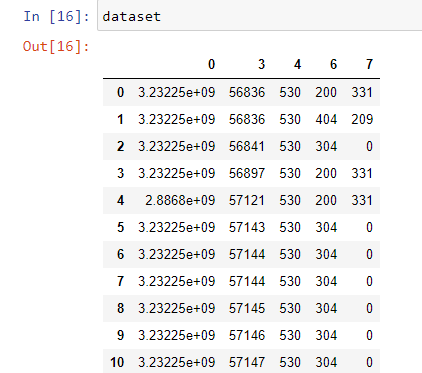
Now our dataset will look like this



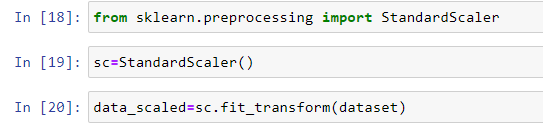
Now we will convert the IP into a number. For eg. - if we have ‘192.168.43.47’, then it will be converted to “ (2^24)\*192+(2^16)\*168+(2^8)\*43+(2^0)\*47 ”

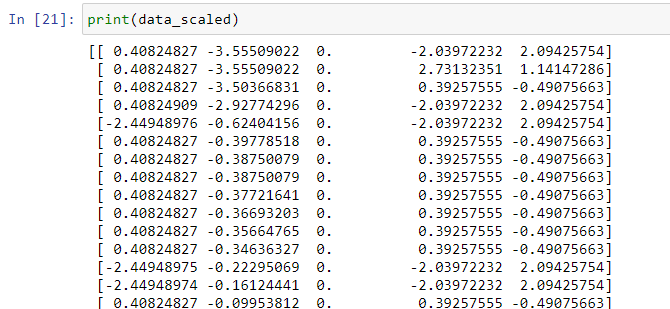


Now, we have dataset as follow.

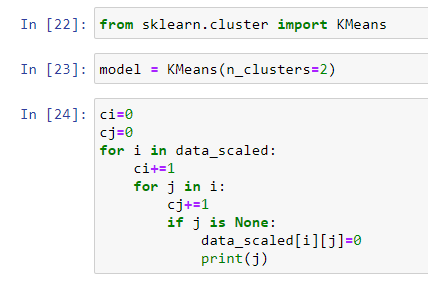


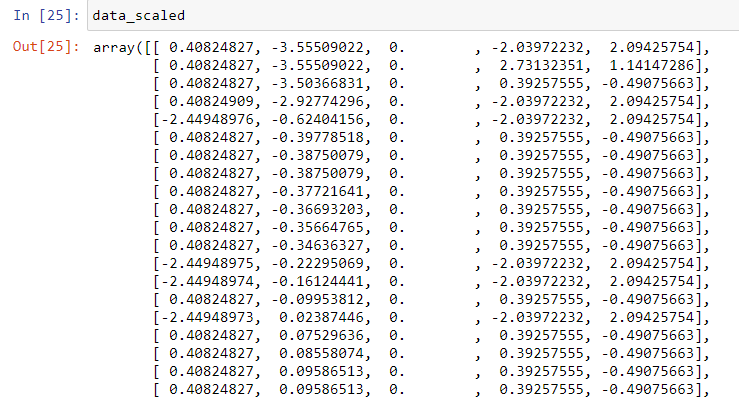
Now we will perform standard scaling, so that all the data will be treated with same priority.



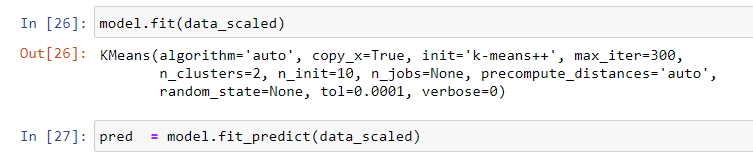


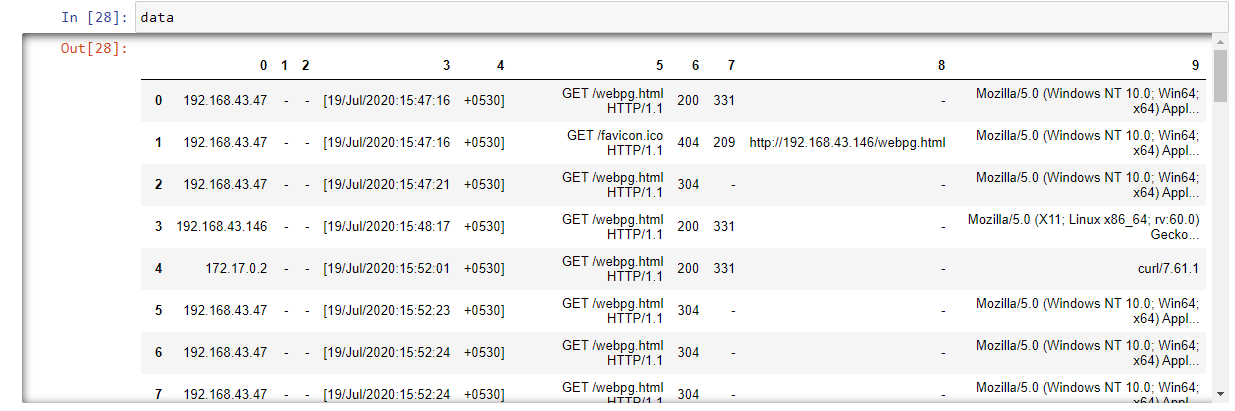
Now we will import ‘KMeans’ and provide our model ‘n\_clusters=2’ to make 2 clusters.



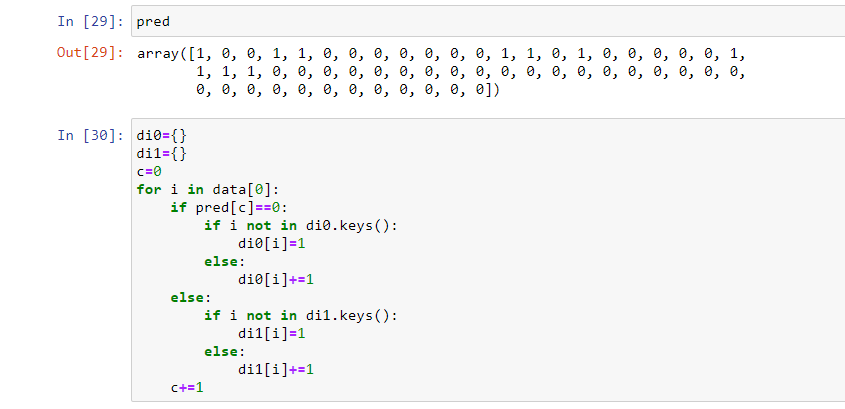


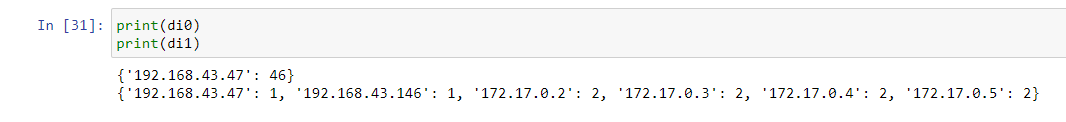
Now we will fit scaled data to our model. Here, we have made a ‘pred’ variable to predict cluster.



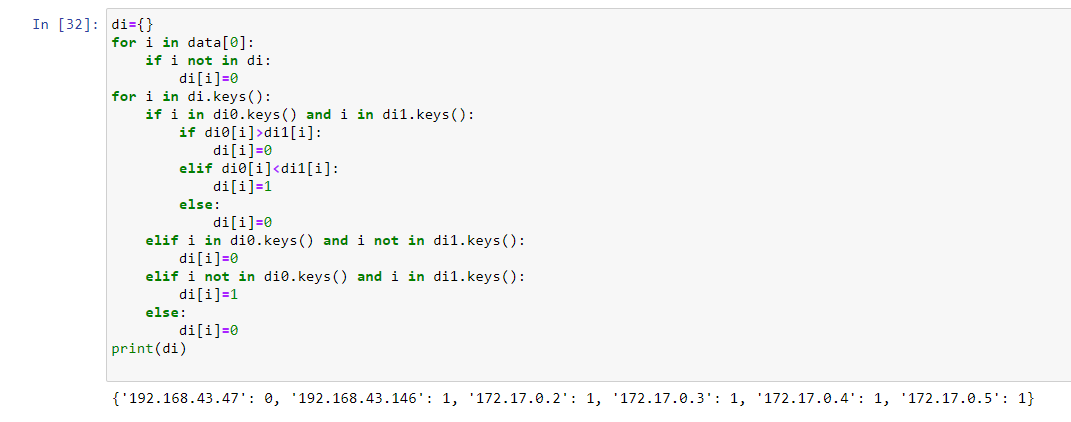


Following ‘pred’ shows the predicted clusters. Now we will make two dictionaries and classify the data according to their clusters.

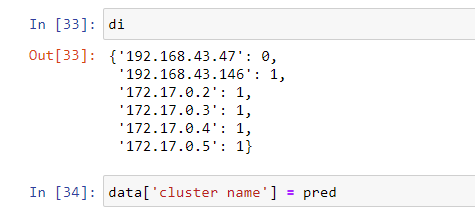




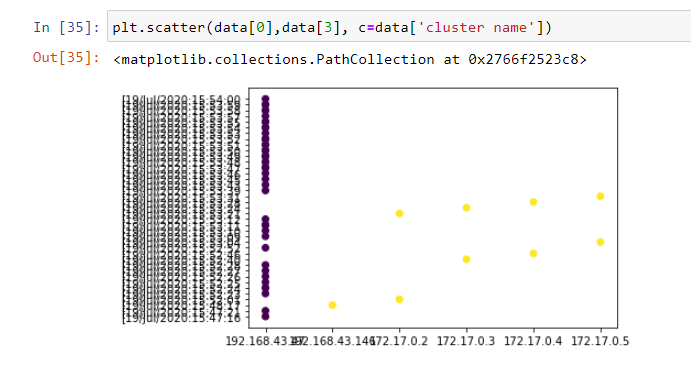
Now when a single IP is lying in both dictionaries then that data will be considered where it’s value is maximum.



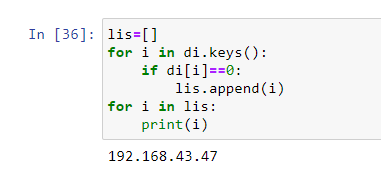
This ‘di’ shows the cluster in which each IP is lying.



The following graph is showing the different clusters



Here, we got the IP which we are required to block as it is behaving differently and triggering our webpage again and again.



**We have learnt these things in MLOps training organized by Linux World under the mentorship of Mr. Vimal Daga Sir.**

**Thanks for reading.**