

Name : Keval Kavle
Assignment : Programming Assignment-2
Subject : Cloud Computing CS643

Docker Image :
<https://hub.docker.com/repository/docker/kkdockernet/wine-quality-prediction>

Github Link to Code :
<https://github.com/KKGITHUBNET/Cloud-Computing>

➤ **Installation process for cloud setup and running application without docker :**

1. Create EMR cluster on AWS using following configurations.

EMR version 5.20.0

In advance option :

- 1.1 Check spark, hadoop and uncheck others
- 1.2 Paste below code in configurations :

```
[
  {
    "Classification": "spark-env",
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "PYSPARK_PYTHON": "/usr/bin/python3"
        }
      }
    ]
  }
]
```

2. EMR Master configurations :

Add following in bound traffic rule :
Protocol : ssh port : 22

3. Copy code from github on EMR master using ssh client(winscp / filezilla)

TrainingDataset.csv
traintest.py
dockerfile

Add TestDataset.csv file to this EMR(for training use ValidationDataset.csv)

4. EMR setup :

Run following commands on EMR master.

```
sudo yum update -y
```

```
sudo nano /etc/sudoers
```

Update this line -> Defaults secure_path = /sbin:/bin:/usr/sbin:/usr/bin
add "/usr/local/bin" to this PATH

```
# Install pip, python, and other libraries
```

```
sudo easy_install pip
sudo pip install --upgrade pip
sudo pip install wheel
sudo pip install findspark
sudo pip install pyspark
sudo pip install numpy
```

```
##### Following commands were run on ec2 – not needed on EMR #####
sudo yum install java-1.8.0
export JAVA_HOME=/etc/alternatives/jre
#JAVA_HOME=/etc/alternatives/jre
pwd
export PATH=$PATH:<pwd>
```

5. Run Program :

```
# Program for validation dataset (training)
> python train.py
# Program to run for testdataset (predictions)
> python traintest.py
```

➤ Docker setup and running application with docker :

1. Run following command on EC2(master) or EMR(master) :

1.1 Install docker and start docker service

```
sudo yum install -y docker
sudo service docker start
```

1.2 Following commands are run for creating and pushing docker image to dockerhub (not to be run for running the application) :

```
sudo docker login -u kkdockernet
sudo docker build . -f dockerfile -t kkdockernet/wine-quality-prediction
sudo docker run -t kkdockernet/wine-quality-prediction
sudo docker push kkdockernet/wine-quality-prediction
```

1.3 To run from anywhere(linux or docker desktop) where docker is installed :

```
sudo docker pull kkdockernet/wine-quality-prediction
```

Please check following guideline to run the docker image :

sudo docker run -v <Testfile> -t <DockerImage> <path for testfile (only path and not the name.)>

```
sudo docker run -v TestDataset.csv -t kkdockernet/dockertrial ./
```

Following screenshot demonstrates running the docker image on ec2(not on emr.) This is only for reference. Docker image could be run anywhere where docker is installed.(EMR/ linux machine etc.)

```
ec2-user@ip-172-31-80-94:~
e0c2e1de5e10: Layer already exists
18f8c5db909c: Layer already exists
94c3580a488c: Layer already exists
b9c9b5de16d1: Layer already exists
0e5e06178d52: Layer already exists
3fc64803ca2d: Layer already exists
latest: digest: sha256:db888a902e3928cf17533d3279d13e9f33478647a3d02ea0dd1a80e641609c1a size: 3677
[ec2-user@ip-172-31-80-94 ~]$ sudo docker pull kkdockernet/wine-quality-prediction
Using default tag: latest
latest: Pulling from kkdockernet/wine-quality-prediction
Digest: sha256:db888a902e3928cf17533d3279d13e9f33478647a3d02ea0dd1a80e641609c1a
Status: Image is up to date for kkdockernet/wine-quality-prediction:latest
docker.io/kkdockernet/wine-quality-prediction:latest
[ec2-user@ip-172-31-80-94 ~]$ sudo docker pull kkdockernet/wine-quality-prediction
Using default tag: latest
latest: Pulling from kkdockernet/wine-quality-prediction
Digest: sha256:db888a902e3928cf17533d3279d13e9f33478647a3d02ea0dd1a80e641609c1a
Status: Image is up to date for kkdockernet/wine-quality-prediction:latest
docker.io/kkdockernet/wine-quality-prediction:latest
[ec2-user@ip-172-31-80-94 ~]$ sudo docker run -v TestDataset.csv -t kkdockernet/dockertrial ./
2020-12-09 21:24:09 WARN NativeCodeLoader:62 - Unable to load native-hadoop library for your platfor
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
2020-12-09 21:24:34 WARN WeightedLeastSquares:66 - regParam is zero, which might cause numerical ins
2020-12-09 21:24:34 WARN BLAS:61 - Failed to load implementation from: com.github.fommil.netlib.Nati
2020-12-09 21:24:34 WARN BLAS:61 - Failed to load implementation from: com.github.fommil.netlib.Nati
2020-12-09 21:24:34 WARN LAPACK:61 - Failed to load implementation from: com.github.fommil.netlib.Na
2020-12-09 21:24:34 WARN LAPACK:61 - Failed to load implementation from: com.github.fommil.netlib.Na
#### Testdataset Accuracy ####
('Accuracy :', 60.0, '%')
('f1 Score :', 0.5846181534031478)
[ec2-user@ip-172-31-80-94 ~]$
```