**PROGRAMMING ASSIGNMENT -02**

**Wine Quality Analysis**

Email-Id: [vb53@njit.edu](mailto:vb53@njit.edu)

Git-Hub: <https://github.com/vinayb004/wine-quality>

DockerHub: <https://hub.docker.com/repository/docker/vinaybasavaraja/winequal>

**Parallel training on four ec2 Instances**

**TASK 1: Parallel training on four ec2 Instances**

**Create the Cluster using flintrock.**

* Create the ec2 instance in AWS console.
* Open the ec2 instance and run it.
* Configure the ec2 credentials for the instance.
* Run the command (**pip2 install flintrock**) which will download flintrock files.
* Configure the flintrock using the **flintrock configure**.
* Then it will show the path of config.yml file and using vi text editor change the key name and path of key pair.
* Change the number of slaves to 3 which will create one master and 3 slave instances.
* Launch the cluster using the command **Flintrock launch wine-quality-cluster.**
* Login into the cluster using the command **Flintrock login wine-quality-cluster.**
* Configure the master instance security settings that is change permission of the ports 22 and 7077 to anywhere IPv4.

Upload files to Cluster Master node

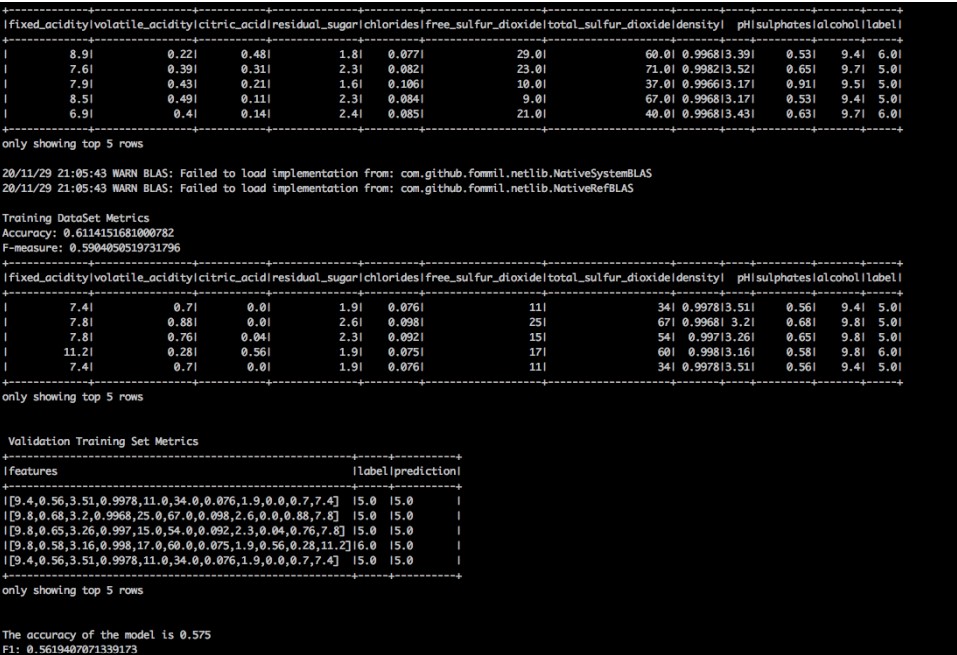
* Once Cluster goes into waiting state, copy master node dns address and open command prompt on local machine.
* Upload TrainingDataSet.csv, ValidationDataset.csv and winequalityanalysis-1.0.jar to master node using FileZilla.

**Launch Model-Trainer application:**

Now everything is finished, I want to run Apache-spark application on cluster.

Execute following command to run application

**spark-submit winequalityanalysis-1.0.jar**



* This generates the training model and now the generated training model file is converted into zip file.
* Send the training model zip file into new ec2 instance and unzip the file in that ec2 instance.

**Ec2 instance Create:**

* After logging into AWS console,
* Go to EC2 ->run instance
* Choose key-pairs and run it.

**Ec2 instance pre configuration:**

• Install SCALA:

o wget http://downloads.typesafe.com/scala/2.11.6/scala-2.11.6.tgz o tar -xzvf scala-2.11.6.tgz o Update PATH environment variable:

▪ vim ~/.bashrc

* copy following lines into file and then save it
* export SCALA\_HOME=/home/ec2-user/scala-2.11.6
* export PATH=$PATH:/home/ec2-user/scala-2.11.6/bin
* source ~/.bashrc

**Install SPARK:**

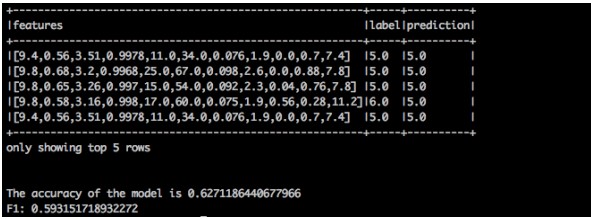
* wget https://archive.apache.org/dist/spark/spark-2.4.5/spark-2.4.5-binhadoop2.7.tgz
* sudo tar xvf spark-2.4.5-bin-hadoop2.7.tgz -C /opt
* sudo chown -R ec2-user:ec2-user /opt/spark-2.4.5-bin-hadoop2.7
* sudo ln -fs spark-2.4.5-bin-hadoop2.7 /opt/spark o Update PATH Environment
  + vim ~/.bash\_profile
  + copy following lines into file and then save it

export SPARK\_HOME=/opt/spark

PATH=$PATH:$SPARK\_HOME/bin

export PATH

* source ~/.bash\_profile
* Run the command winequalityanalysis-1.0.jar



**Predict wine-quality using docker**

Test filename has to be TestDataset.csv and file has to be placed under data/ directory of container. To do this use -v parameter to map volumes.

Docker run instructions

**docker pull vinaybasavaraja/winequal:v1**

**docker run -v [fullLocalPath of TestDataset.csv: data/TestDataset.csv ] vinaybasavaraja/winequal:v1**

for e.g.

sudo docker run -v /home/ec2-user/data/ValidationDataset.csv vinaybasavaraja/winequal:v1

