**EMR Cluster Code:**

## Start the EMR cluster

* 1. Create an EMR cluster with name “wineTesting”
  2. In Applications, choose Spark
  3. Choose number of Instances as 5 (1 Master and 4 cores)
  4. Enable SSH inbould rule for port 22 in security group of EMR Master and roles to create the cluster.
  5. Wait till the cluster reaches in “Waiting” State

1. Create a S3 Bucket with name “winedataset” and add the files to the bucket.
2. Check the versions of the all the installed components.
   1. To provide EMR an access to the AWS CLI, open the SSH Role in the master and then execute the SSH command provided by the AWS
   2. Run the command, to download all the files in the EC2 locally

aws s3 cp s3://winedataset . --recursive

**Model Code (in Spark):**

1. Establish a connection to Spark using SparkContext
2. Import pySpark and initialize Spark Session.
3. To load the Validation Data set file in Memory.read the file Validation Data set, write the following commands.
4. To create a Vector of features to predict the values execute the following Lines
5. To perform RandomForest Classification on the given data and develop the model.  
   Import regression module from ml.regression and provide the constructor with parameters.

It calibrates the model for maximum accuracy by checking its RMS error

Save the model for prediction.

**Docker Container Setup:**

1. Create docker file using nano file-name command.
2. Sign up on docker hub
3. start docker using command: sudo service docker start
4. Build docker on EC2 instance and create docker image using command: sudo docker build . -f docker-file-name -t image-name
5. Build docker using command: sudo docker build . -f docker-file-name -t image-name
6. Run docker image using command: sudo docker run image-name
7. Create image in docker hub account using command: sudo docker build . -f docker-id-file -t docker-hub-id/image-name-on-docker-hub
8. Login to Docker hub account using following command and enter password: sudo docker login -u docker-hub-account-id/docker-id
9. Push files to docker using command: sudo docker push docker-account-id/docker-image-name
10. Run the image of the Docker within the Docker Hub on EC2 Instance sudo docker run -t docker-hub-account-id/docker-image-name

**Github Link for Models:**

**Github :**

<https://github.com/manishptl/cloudcomputing>

**Docker Hub:**

https://hub.docker.com/repository/docker/manishptl/ccassignment