**SAI KARTHIK KASUMURTHY (sk3374@njit.edu)  
Programming Assignment 2  
  
GitHub Link:** <https://github.com/karthik984/WinePredictionAnalysis> **Docker Image Link:** <https://hub.docker.com/r/karthikkk999/wine-prediction>  
 **Step-by-step process on how to set-up the cloud environment, run the model training and the application prediction:**

1. **Create a Key Pair:**
   1. Go to EC2 feature > “Key Pairs” and click on “Create key pair” as  
      A screenshot of a computer

      Description automatically generated  
      A screenshot of a computer

      Description automatically generated
   2. Give it a name and select “.pem” if you have MAC OS or “.ppk” if you have Windows and click on “Create Key pair”  
      A screenshot of a computer

      Description automatically generated
   3. This will also download the “.ppk” or “.pem” file to your local machine. Make sure you save it. You will require it later.
   4. Once Created it should appear in your key pairs list:  
      A screenshot of a computer

      Description automatically generated
2. **Create EMR Cluster:**
   1. Go to “Amazon EMR > EMR on EC: Clusters > Create cluster” and give it a name and select the latest emr version. Also check the Hadoop, Spark , Zeppelin which are required by our code to train the model  
      A screenshot of a computer

      Description automatically generated
   2. Keep Cluster Configuration as it is and scroll until “Cluster scaling and provisioning”.
   3. In “Cluster scaling and provisioning” select 1 Core Instance and 3 Task – 1 instances. (Total 4)  
      A screenshot of a computer

      Description automatically generated
   4. Select the Cluster termination as per your convenience. I had selected 3 hours  
      A screenshot of a computer

      Description automatically generated
   5. Keep rest as it is and scroll down to “Security configuration and “EC2 key pair”. Select the key that you have created in Step 1.   
      A screenshot of a computer program

      Description automatically generated
   6. Select the EMR\_DefaultRole as the service role and EMR\_EC2\_DefaultRole as the instance profile and the rest as it is.  
      A screenshot of a computer

      Description automatically generated
   7. Verify the summary and create the cluster  
      A screenshot of a computer

      Description automatically generated
3. **Creating S3 Bucket:**
   1. Go Amazon S3 > Buckets and create an S3 bucket with a name as shown below  
      A screenshot of a computer

      Description automatically generated
   2. Select the AWS region, give a name to the S3 bucket, keep the rest settings as it is  
      A screenshot of a bucket

      Description automatically generated  
      A screenshot of a computer

      Description automatically generated
   3. Once created it should appear in the list of s3 buckets  
      A screenshot of a computer

      Description automatically generated
   4. Upload the TrainingDataset.csv and ValidationDataset.csv and the src/wine\_prediction\_train.py train and A screenshot of a computer

      Description automatically generated
   5. The folder “src/” would have the code that trains and creates the model.
4. **Parallelly Training the Model in the EMR Cluster**
   1. Connect to the EMR EC2 Master node through putty using the “winepred.ppk” file from your local as shown below:  
      A computer screen shot of a computer

      Description automatically generated A screenshot of a computer

      Description automatically generated
   2. Use ec2-user as the login user  
      A screen shot of a computer

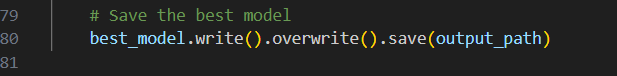
      Description automatically generated
   3. After successful you should see something like this  
      A screenshot of a computer

      Description automatically generated
   4. Now run the command “sudo spark-submit s3://sk3374-winepred/src/wine\_prediction\_train.py” to start and train the model  
      A screen shot of a computer

      Description automatically generated
   5. You will see something like this it the code is running and the model is getting trained:  
      A screen shot of a computer

      Description automatically generated  
      A black screen with white text

      Description automatically generated
   6. Once the code completed executed it should create a model and store in in the AWS S3 Bucket that you had created “s3://sk3374-winepred” earlier as per the login in the code:  
      A computer screen with text and red line

      Description automatically generated 
   7. As shown below, the model would be created in the s3 bucket in the folder “trained.model/”  
      A screenshot of a computer

      Description automatically generated
5. **Testing the Trained Model on EC2 Instance and Locally**
   1. Now can you can get the model from your s3 bucket to your EC2 instance node using the below command  
      “aws s3 sync s3://sk3374-winepred/trained.model ./trained.model/”  
      A black screen with a white rectangular object

      Description automatically generated  
      Now when you do ls you can see the downloaded model into your instance  
      A screen shot of a computer program

      Description automatically generated
   2. Now using WinSCP connect to the EC2 instances to test the trained model using a script. Create a testdata.csv using the previous training data or the validation data just to test the model  
      A screenshot of a computer screen

      Description automatically generated
   3. Now download the trained.model folder from the EC2 instance to your local and transfer the testdata.csv and wine\_prediction\_test.py (the script to test the accuracy of the trained model) to the EC2 instance  
      A screenshot of a computer

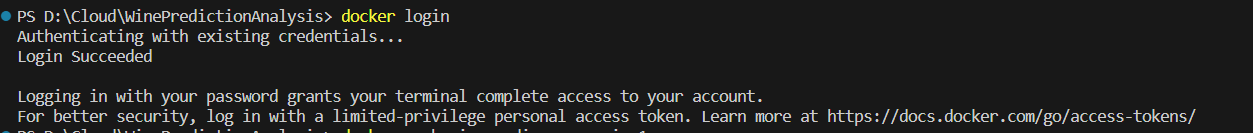
      Description automatically generated
   4. One done make sure all the files needed for testing the trained model are on your EC2 instance.  
      A black screen with a white background

      Description automatically generated
   5. Now go ahead and run the wine\_prediction\_test.py using the below command  
      python wine\_prediction\_test.py  
      A screen shot of a computer

      Description automatically generated
   6. Also you can run the wine\_prediction\_test.py in your local once you get the trained.model from the S3 as shown below  
      A black screen with white and blue squares

      Description automatically generated with medium confidence
6. **Create a Docker Image and Push it to DockerHub**
   1. Now create a Dockerfile with all the required configurations to run the wine\_prediction\_test.py in your repo
   2. Build an image using Dockerfile. Below is the command that is used create and build a docker image.  
      Command: “docker build -t winepredimage:version1 .”  
      A screenshot of a computer program

      Description automatically generated
   3. Once built, test it on your local using the docker run command as shown below   
      Command: “docker run winepredimage:version1” (This command is only for running on local)  
      A screenshot of a computer

      Description automatically generated
   4. Now use docker login to login to your DockerHub account. (I have already logged in and that’s why it said Login Succeeded)  
      
   5. Now go to DockerHub on a web browser and create a repository called wine-prediction  
      A screenshot of a web browser

      Description automatically generated
   6. Now tag the local image to your repository that you have created using the below command  
      “docker tag winepredimage:version1 karthikkk999/wine-prediction:version1”
   7. And push the docker image to the DockerHub using the command  
      “docker push karthikkk999/wine-prediction:version1”
   8. Once successfully pushed, You can see this on your DockerHub account.  
      A screenshot of a computer

      Description automatically generated
7. **Pulling the Docker Image and Running it on an EC2 Instance**
   1. SSH into any of the EC2 instances that you want to run the docker image and install docker  
      “sudo yum install docker”  
      A screenshot of a computer

      Description automatically generated  
      A computer screen with white text

      Description automatically generated
   2. Start docker and pull the wine-prediction image that you have built  
      “sudo service docker start”  
      “sudo docker pull karthikkk999/wine-prediction:version1”  
      A screen shot of a computer

      Description automatically generated
   3. Once the docker image is pulled run it using the docker run command  
      “sudo docker run karthikkk999/wine-prediction:version1” A screenshot of a computer screen

      Description automatically generated
   4. If you want to specify a custom dataset file that you want, you can do that by persisting volume outside the container by using a flag -v as shown below  
      “docker run -v D:/Cloud/WinePred\_files/src:/winepred/src karthikkk999/wine-prediction:version1 ValidationDataset.csv”  
      A computer screen with text on it

      Description automatically generated