# README FILE

Github Link: https://github.com/SwapnilDhavale/SparkWinePrediction/tree/main

Docker hub: docker pull swappy7/cloudwinepred:v3

docker run -v ~/Desktop/TestDataset.csv:/app/TestDataset.csv swappy7/cloudwinepred:v3

~/Desktop/TestDataset.csv is the file location in local system.

STEPS:

1. Create EMR Cluster.

<https://us-east-2.console.aws.amazon.com/elasticmapreduce/home?region=us-east-2#quick-create>:

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

1. Update Security group of Master.

Under Inbound add new rule type: SSH and source: Anywhere.

1. Now copy all the files including RFModelTrainer.py ValidationDataset.csv and TrainingDataset.cvs into EMR.

sftp -i ~/Desktop/Jars/frstEc2.pem [hadoop@ec2-18-218-239-94.us-east-2.compute.amazonaws.com](mailto:hadoop@ec2-18-218-239-94.us-east-2.compute.amazonaws.com)

put RFModelTrainer.py and similarly remaining two file in EMR.

Text

Description automatically generated

1. Now login into the cluster using following:

ssh -i frstEc2.pem [hadoop@ec2-18-218-239-94.us-east-2.compute.amazonaws.com](mailto:hadoop@ec2-18-218-239-94.us-east-2.compute.amazonaws.com)

A picture containing graphical user interface, chart

Description automatically generated

1. Now all your files will be in home/Hadoop.

You need to send or place files in Hadoop file system (fs) for parallel execution using following command.

hadoop fs -put TrainingDataset.csv /user/hadoop/TrainingDataset.csv

hadoop fs -put ValidationDataset.csv

/user/hadoop/ ValidationDataset.csv

Text

Description automatically generated

1. Once all the files are placed in fs we can start the execution by using following command:

RUN spark-submit RFModelTrainer.py

A picture containing shape

Description automatically generated

Graphical user interface, text, application

Description automatically generated

1. Our Model is saved in Hadoop file system in order to copy it locally we need to use following command

hdfs dfs -copyToLocal ModelV1 /home/hadoop/

Text

Description automatically generated

1. Now we finally have our ModelV1. Copy the model (ModelV1) to our local machine.
2. Create a container using docker following files are required:

Graphical user interface

Description automatically generated

1. Now run following command to build a docker

docker build -t cloudwinepred .

docker tag 8ec0a32ca8bf swappy7/cloudwinepred:v3

docker push swappy7/cloudwinepred

1. Now Image is pushed on docker hub.

Graphical user interface, text, application, email, website

Description automatically generated

1. Fetch the image using:

docker pull swappy7/cloudwinepred:v3

1. In order to run use

docker run -v ~/Desktop/TestDataset.csv:/app/TestDataset.csv swappy7/cloudwinepred:v3

~/Desktop/TestDataset.csv is the file location local system.

1. Run Docker in EC2 instance:
2. Create a EC2 instance
3. Copy the TestDataset.csv into EC2 /home/ec2-user using following command:

scp -i ~/Desktop/Jars/frstEc2.pem ~/Desktop/TestDataset.csv [ec2-user@ec2-13-58-212-106.us-east-2.compute.amazonaws.com:/home/ec2-user](mailto:ec2-user@ec2-13-58-212-106.us-east-2.compute.amazonaws.com:/home/ec2-user)

1. Use ssh -i ~/Desktop/Jars/frstEc2.pem [ec2-user@ec2-13-58-212-106.us-east-2.compute.amazonaws.com](mailto:ec2-user@ec2-13-58-212-106.us-east-2.compute.amazonaws.com) to access EC2.
2. In EC2 install docker use following command:

sudo yum update -y

sudo yum install docker -y

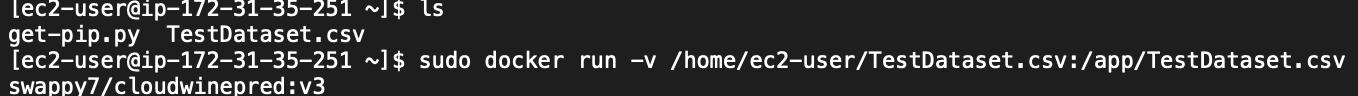
sudo service docker start

1. sudo docker pull swappy7/cloudwinepred:v3 to pull the Image.

Text

Description automatically generated

1. sudo docker run -v /home/ec2-user/TestDataset.csv:/app/TestDataset.csv swappy7/cloudwinepred:v3



Text

Description automatically generated

You will get F1-score and Accuracy