# Training from EMR 5 ec2

# Prediction from 1 EMR ec2

Clone cc643-winepred\_dock EMR cluster

Docker login – enter credentials

Mkdir app

Cd app

Nano am3329\_test.py

Pip install findspark

python am3329\_test.py s3://am3329-cc648-wineapp/ValidationDataset.csv

In local terminal - scp -i cc643-wineapp-prgass.pem ValidationDataset.csv hadoop@ec2-18-234-163-139.compute-1.amazonaws.com:~/app

scp -r cc643-wineapp-prgass.pem trainmodel.model hadoop@ec2-18-234-163-139.compute-1.amazonaws.com:~/app

## docker steps

sudo systemctl enable docker.service

sudo systemctl start docker.service

sudo systemctl status docker.service – should so active

sudo chmod 666 /var/run/docker.sock

docker pull (container)

docker run -it masroorhasan/pyspark:latest bash

pip install findspark

cat > /data/am3329\_test.py

python am3329\_test.py

exit

nano Dockerfile

docker build . -t cc643winepred

docker images – to verify the image

docker run cc643winepred ValidationDataset.csv

docker build . -t am3329ahmeda/cc643winepred

docker system prune – to clear docker image from the instance/device

docker run am3329ahmeda/cc643winepred ValidationDataset.csv

## Docker run from a new ec2 instance

Install docker using the link - <https://www.simplilearn.com/tutorials/docker-tutorial/how-to-install-docker-on-ubuntu>

Mkdir app

Cd app

scp -i cc643-ec2.pem ValidationDataset.csv [ubuntu@ec2-44-202-33-73.compute-1.amazonaws.com:~/app](mailto:ubuntu@ec2-44-202-33-73.compute-1.amazonaws.com:~/app)

sudo chmod 666 /var/run/docker.sock

docker login

docker run am3329ahmeda/cc643winepred ValidationDataset.csv