Wine Quality Prediction ML Model User-Guide

a) Set up AWS EC2 Instances:
  i)   Launch 4 EC2 instances for model training and use 1 instance for the prediction
       application.

	Name ✎ ▽	Instance ID	Instance state ▽	Instance type ▽
☑	WinePred_D	i-067256253cec67d5b	⊘ Running ⊕ ⊖	t2.large
☐	WinePred_A	i-0d49bab18aadd4ec6	⊘ Running ⊕ ⊖	t2.large
☐	WinePred_B	i-09def3b9435f12728	⊘ Running ⊕ ⊖	t2.large
☐	WinePred_C	i-0821de9bf5229384d	⊘ Running ⊕ ⊖	t2.large

b) Prepare the Environment:
  i)   Install Python and Apache Spark on all EC2 instances.
  ii) Clone the repository to each instance or transfer the project files.
c) Train the Model:
  i)   Execute the Spark application on the 4 instances to train the model using
       TrainingDataset.csv.

```
ubuntu@ip-172-31-51-235:~/winePrediction$ ps
  PID TTY             TIME CMD
 1621 pts/0       00:00:00 bash
 2820 pts/0       00:00:15 java
 4931 pts/0       00:00:00 ps
```

```
Last login: Mon Dec  4 02:05:20 2023 from 108.53.196.128
ubuntu@ip-172-31-52-53:~$ export SPARK_HOME=/home/ubuntu/spark-3.5.0-bin-hadoop3
ubuntu@ip-172-31-52-53:~$ export PATH=$PATH:$SPARK_HOME/bin
ubuntu@ip-172-31-52-53:~$ cd $SPARK_HOME/sbin
ubuntu@ip-172-31-52-53:~/spark-3.5.0-bin-hadoop3/sbin$ ./start-worker.sh spark://ip-172-31-51-235:7077
starting org.apache.spark.deploy.worker.Worker, logging to /home/ubuntu/spark-3.5.0-bin-hadoop3/logs/spark-ubunt
u-org.apache.spark.deploy.worker.Worker-1-ip-172-31-52-53.out
ubuntu@ip-172-31-52-53:~/spark-3.5.0-bin-hadoop3/sbin$ ▯
```

```
LINUX_PACKAGES_LICENSES              PYTHON_PACKAGES_LICENSES  spark-3.5.0-bin-hadoop3
ubuntu@ip-172-31-55-111:~$ export SPARK_HOME=/home/ubuntu/spark-3.5.0-bin-hadoop3
ubuntu@ip-172-31-55-111:~$ export PATH=$PATH:$SPARK_HOME/bin
ubuntu@ip-172-31-55-111:~$ cd $SPARK_HOME/sbin
ubuntu@ip-172-31-55-111:~/spark-3.5.0-bin-hadoop3/sbin$ ./start-worker.sh spark://ip-172-31-51-235:7077
starting org.apache.spark.deploy.worker.Worker, logging to /home/ubuntu/spark-3.5.0-bin-hadoop3/logs/spark-ubunt
u-org.apache.spark.deploy.worker.Worker-1-ip-172-31-55-111.out
ubuntu@ip-172-31-55-111:~/spark-3.5.0-bin-hadoop3/sbin$ ▯
```

```
Last login: Mon Dec  4 01:01:49 2023 from 108.53.196.128
ubuntu@ip-172-31-62-254:~$ export SPARK_HOME=/home/ubuntu/spark-3.5.0-bin-hadoop3
ubuntu@ip-172-31-62-254:~$ export PATH=$PATH:$SPARK_HOME/bin
ubuntu@ip-172-31-62-254:~$ cd $SPARK_HOME/sbin
ubuntu@ip-172-31-62-254:~/spark-3.5.0-bin-hadoop3/sbin$ ./start-worker.sh spark://ip-172-31-51-235:7077
starting org.apache.spark.deploy.worker.Worker, logging to /home/ubuntu/spark-3.5.0-bin-hadoop3/logs/spark-ubunt
u-org.apache.spark.deploy.worker.Worker-1-ip-172-31-62-254.out
ubuntu@ip-172-31-62-254:~/spark-3.5.0-bin-hadoop3/sbin$ ▮
```

d) Validate and Optimize the Model:
   i) Use ValidationDataset.csv to validate and fine-tune the model's parameters.
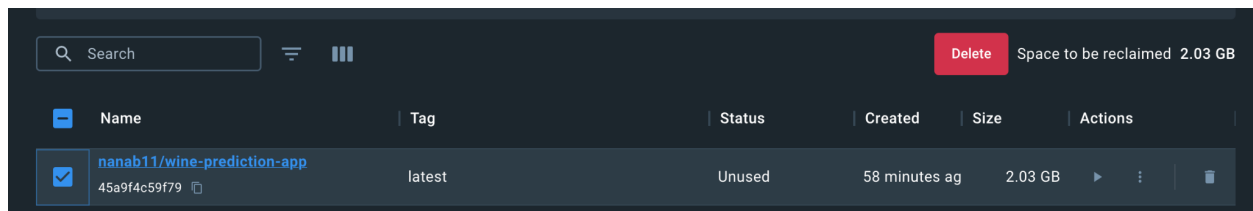
```
ubuntu@ip-172-31-51-235:~/winePrediction$ nano wine_quality_prediction.py
ubuntu@ip-172-31-51-235:~/winePrediction$ python3 wine_quality_prediction.py
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/12/04 04:18:31 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes
where applicable
23/12/04 04:18:42 WARN InstanceBuilder: Failed to load implementation from:dev.ludovic.netlib.blas.JNIBLAS
F1 Score: 0.5672726692311375
ubuntu@ip-172-31-51-235:~/winePrediction$
```

e) Build and Deploy the Docker Container:
   i) Build the Docker container for the prediction application.
   ii) Deploy the container to the prediction EC2 instance.
f) Run the Prediction Application:
   i) Execute the prediction application within Docker, using the test dataset.

Name	Tag	Status	Created	Size	Actions
nanab11/wine-prediction-app 45a9f4c59f79	latest	Unused	58 minutes ag	2.03 GB	▶ ⋮ 🗑