

Wine Quality Prediction ML Model User-Guide

a) Set up AWS EC2 Instances:

- Launch 4 EC2 instances for model training and use 1 instance for the prediction application.

<input checked="" type="checkbox"/>	Name	Instance ID	Instance state	Instance type
<input checked="" type="checkbox"/>	WinePred_D	i-067256253cec67d5b	Running	t2.large
<input type="checkbox"/>	WinePred_A	i-0d49bab18aadd4ec6	Running	t2.large
<input type="checkbox"/>	WinePred_B	i-09def3b9435f12728	Running	t2.large
<input type="checkbox"/>	WinePred_C	i-0821de9bf5229384d	Running	t2.large

b) Prepare the Environment:

- Install Python and Apache Spark on all EC2 instances.
- Clone the repository to each instance or transfer the project files.

c) Train the Model:

- Execute the Spark application on the 4 instances to train the model using TrainingDataset.csv.

```
ubuntu@ip-10-0-1-10: ~/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 [redacted]
ubuntu@ip-10-0-1-10: ~$ export SPARK_HOME=/home/ubuntu/spark-3.5.0-bin-hadoop3
ubuntu@ip-10-0-1-10: ~$ export PATH=$PATH:$SPARK_HOME/bin
ubuntu@ip-10-0-1-10: ~$ cd $SPARK_HOME/sbin
ubuntu@ip-10-0-1-10: ~/spark-3.5.0-bin-hadoop3/sbin$ ./start-worker.sh spark://ip-10-0-1-10:7077
starting org.apache.spark.deploy.worker.Worker, logging to /home/ubuntu/spark-3.5.0-bin-hadoop3/logs/spark-ubuntu-0-0-1-10.out
ubuntu@ip-10-0-1-10: ~/spark-3.5.0-bin-hadoop3/sbin$
```

```
Last login: Mon Dec  4 01:01:49 2023 from [redacted]
ubuntu@ip-10-0-1-10: ~$ export SPARK_HOME=/home/ubuntu/spark-3.5.0-bin-hadoop3
ubuntu@ip-10-0-1-10: ~$ export PATH=$PATH:$SPARK_HOME/bin
ubuntu@ip-10-0-1-10: ~$ cd $SPARK_HOME/sbin
ubuntu@ip-10-0-1-10: ~/spark-3.5.0-bin-hadoop3/sbin$ ./start-worker.sh spark://ip-10-0-1-10:7077
starting org.apache.spark.deploy.worker.Worker, logging to /home/ubuntu/spark-3.5.0-bin-hadoop3/logs/spark-ubuntu-0-0-1-10.out
ubuntu@ip-10-0-1-10: ~/spark-3.5.0-bin-hadoop3/sbin$
```

```
ubuntu@ip-10-0-1-10: ~/winePrediction$ ls
Dockerfile  README.md  TrainingDataset.csv  ValidationDataset.csv  testData  wine_quality_prediction.py
```

d) Validate and Optimize the Model:

- i) Use ValidationDataset.csv to validate and fine-tune the model's parameters.

```
ubuntu@ip-172-31-1-100:~/winePrediction$ ls
Dockerfile TrainingDataset.csv ValidationDataset.csv wine_quality_prediction.py
ubuntu@ip-172-31-1-100:~/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 04:37:13 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin
-java classes where applicable
23/12/04 04:37:25 WARN InstanceBuilder: Failed to load implementation from:dev.ludovic.netlib.blas.JNIBLAS
F1 Score: 0.5672726692311375
ubuntu@ip-172-31-1-100:~/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.

Search				Delete		Space to be reclaimed 2.03 GB	
	Name	Tag	Status	Created	Size	Actions	
<input checked="" type="checkbox"/>	nanab11/wine-prediction-app	latest	Unused	58 minutes ag	2.03 GB		
	45a9f4c59f79						