

README file for PA2 by Pranav Subramanian

Running the file on EC2 instances involves

- setting up on EC2 instances,
- Running the script
- running with Docker file

Please follow the steps below to setup and run the Python script

Setup on EC2

1. Install spark – Can run the command `sudo wget apache`
<http://archive.apache.org/dist/spark/spark-2.0.0/spark-2.0.0-bin-hadoop2.7.tgz>
2. Do `sudo tar -zxvf spark-2.0.0-bin-hadoop2.7.tgz`
3. Install Hadoop –
run the command `sudo tar -zxvf hadoop` to unzip tar file if needed
4. Install Python
Run `python 3 get-pip.py`
5. Edit bashrc
Do `ls -la` and set `HADOOP_HOME` to be the location of the Hadoop folder
Modify `PATH` to include Hadoop Home

```
# User specific environment
PATH="$HOME/.local/bin:$HOME/bin:$PATH"
export PATH=$PATH:$HADOOP_HOME/bin/
export HADOOP_HOME=/home/ec2-user/hadoop-2.7.3
```
6. Uploading to cloud (if needed)

For example, I uploaded the file using the command:

Sudo scip -I "instanconnect.pem"

mnt/c/pranav/CS643_Cloud_Computing/TraininDataset.csv EC2-user@EC2-3-89-57-96.compute-1.amazonaws.com:~/.

7. Run the python script – you may need to edit lines 3 & 4 for os.environ and sys.path.append to your location for spark folder

Can run python3 PredictionModel.py pathtocsv

8. While running program may see output like this:

```
Setting default log level to 'WARN'.
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
DataFrames: fixed acidity: double, volatile acidity: double, citric acid: double, residual sugar: double, chlorides: double, free sulfur dioxide: double, total sulfur dioxide: double, density: double, pH: double, sulphates: double, alcohol: double, quality: bigint]
root
|-- fixed acidity: double (nullable = true)
|-- volatile acidity: double (nullable = true)
|-- citric acid: double (nullable = true)
|-- residual sugar: double (nullable = true)
|-- chlorides: double (nullable = true)
|-- free sulfur dioxide: double (nullable = true)
|-- total sulfur dioxide: double (nullable = true)
|-- density: double (nullable = true)
|-- pH: double (nullable = true)
|-- sulphates: double (nullable = true)
|-- alcohol: double (nullable = true)
|-- quality: long (nullable = true)

Rowing predictions
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|fixed acidity|volatile acidity|citric acid|residual sugar|chlorides|free sulfur dioxide|total sulfur dioxide|density| pH|sulphates|alcohol|quality|features|rawPrediction|probability|prediction|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|5.2|0.32|0.25|1.8|0.10|13.0|50.0|0.9957|3.38|0.55|9.2|5|[5.2,0.32,0.25,1...|-2446.4517654991...|[0.0,0.0,0.0,2.50...|5.0|
|5.4|0.85|0.08|1.2|0.04|13.0|93.0|0.9924|3.57|0.85|13.0|7|[5.4,0.85,0.08,1...|-2439.1489983386...|[0.0,0.0,0.0,1.85...|7.0|
|5.6|0.5|0.09|2.3|0.04|17.0|99.0|0.9937|3.63|0.63|13.0|0|[5.6,0.5,0.09,2.3...|-2442.1984513980...|[0.0,0.0,0.0,7.61...|5.0|
|5.5|0.5|0.09|2.3|0.04|17.0|99.0|0.9937|3.63|0.63|13.0|5|[5.5,0.5,0.09,2.3...|-2442.1984513980...|[0.0,0.0,0.0,7.61...|5.0|
|6.3|0.3|0.48|1.8|0.06|18.0|61.0|0.9959|3.44|0.78|10.3|6|[6.3,0.3,0.48,1.8...|-2447.2531316292...|[0.0,0.0,0.0,1.09...|6.0|
|6.6|0.5|0.01|1.5|0.06|17.0|26.0|0.9952|3.4|0.58|9.8|0|[6.6,0.5,0.01,1.5...|-2445.3996789802...|[0.0,0.0,0.0,7.24...|6.0|
|6.5|0.5|0.04|2.1|0.06|18.0|34.0|0.9955|3.39|0.64|9.4|0|[6.5,0.5,0.04,2.1...|-2446.8983649147...|[0.0,0.0,0.0,0.00...|6.0|
|6.8|0.64|0.1|2.1|0.08|18.0|101.0|0.9956|3.34|0.52|10.2|5|[6.8,0.64,0.1,2.1...|-2446.3248822971...|[0.0,0.0,0.0,7.88...|5.0|
|6.8|0.67|0.02|1.8|0.05|5.0|11.0|0.9962|3.48|0.52|9.5|5|[6.8,0.67,0.02,1...|-2447.8443274836...|[0.0,0.0,0.0,0.00...|5.0|
|6.9|0.4|0.14|2.4|0.08|21.0|40.0|0.9968|3.43|0.63|9.7|0|[6.9,0.4,0.14,2.4...|-2449.3838870038...|[0.0,0.0,0.0,5.05...|6.0|
|7.0|0.75|0.05|2.0|0.08|13.0|54.0|0.9966|3.39|0.57|9.8|5|[7.0,0.75,0.05,2...|-2448.8032421349...|[0.0,0.0,0.0,0.00...|5.0|
|7.3|0.39|0.31|2.4|0.07|9.0|46.0|0.9962|3.43|0.54|9.4|6|[7.3,0.39,0.31,2...|-2447.8589987934...|[0.0,0.0,0.0,4.41...|6.0|
|7.3|0.38|0.3|2.4|0.07|15.0|55.0|0.9968|3.46|0.59|10.2|5|[7.3,0.38,0.3,2.4...|-2449.416359985...|[0.0,0.0,0.0,0.00...|5.0|
|7.3|0.59|0.26|7.2|0.07|35.0|121.0|0.9981|3.37|0.49|9.4|5|[7.3,0.59,0.26,7...|-2452.5851685571...|[0.0,0.0,0.0,0.00...|5.0|
|7.5|0.52|0.42|2.3|0.08|8.0|38.0|0.9972|3.58|0.61|10.5|6|[7.5,0.52,0.42,2...|-2450.6262921511...|[0.0,0.0,0.0,0.00...|6.0|
|7.5|0.63|0.32|5.1|0.11|50.0|110.0|0.9981|3.26|0.77|9.4|5|[7.5,0.63,0.32,5...|-2452.9328438724...|[0.0,0.0,0.0,0.00...|5.0|
|7.6|0.08|0.02|1.3|0.07|9.0|20.0|0.9965|3.37|1.08|9.2|4|[7.6,0.08,0.02,1...|-2448.2786735279...|[0.0,0.0,0.0,1.92...|5.0|
|7.7|0.09|0.49|1.8|0.11|20.0|112.0|0.9968|3.21|0.71|9.3|5|[7.7,0.09,0.49,1...|-2449.1456412769...|[0.0,0.0,0.0,0.98...|5.0|
|7.7|0.35|0.43|2.2|0.11|22.0|114.0|0.9971|3.25|0.73|9.2|5|[7.7,0.35,0.43,2...|-2449.7164785453...|[0.0,0.0,0.0,0.00...|5.0|
|7.8|0.43|0.7|1.9|0.46|22.0|67.0|0.9974|3.13|1.28|9.4|5|[7.8,0.43,0.7,1.9...|-2450.7246688105...|[0.0,0.0,0.0,0.02...|5.0|

only showing top 20 rows

rawPrediction[prediction] probability
-2446.4517654991... 5.0|[0.0,0.0,0.0,2.50...|
-2439.1489983386... 7.0|[0.0,0.0,0.0,1.85...|
-2442.1984513980... 5.0|[0.0,0.0,0.0,7.61...|
-2442.1984513980... 5.0|[0.0,0.0,0.0,7.61...|
-2447.2531316292... 6.0|[0.0,0.0,0.0,1.09...|
-2445.3996789802... 6.0|[0.0,0.0,0.0,7.24...|
-2446.8983649147... 6.0|[0.0,0.0,0.0,0.00...|
-2446.3248822971... 5.0|[0.0,0.0,0.0,7.88...|
-2447.8443274836... 5.0|[0.0,0.0,0.0,0.00...|
-2449.3838870038... 6.0|[0.0,0.0,0.0,5.05...|
-2448.8032421349... 5.0|[0.0,0.0,0.0,0.00...|
-2447.8589987934... 6.0|[0.0,0.0,0.0,4.41...|
-2449.416359985... 5.0|[0.0,0.0,0.0,0.00...|
-2452.5851685571... 5.0|[0.0,0.0,0.0,0.00...|
-2450.6262921511... 6.0|[0.0,0.0,0.0,0.00...|
-2452.9328438724... 5.0|[0.0,0.0,0.0,0.00...|
-2448.2786735279... 5.0|[0.0,0.0,0.0,1.92...|
-2449.1456412769... 5.0|[0.0,0.0,0.0,0.98...|
-2449.7164785453... 5.0|[0.0,0.0,0.0,0.00...|
-2450.7246688105... 5.0|[0.0,0.0,0.0,0.02...|
```

9. Run with Docker

Using the dockerfile can create image by running a command like docker build -t image name .

Can run the docker file by running a command like docker run imagename

Link to Github

<https://github.com/GitDeveloperLite/Models>

Link to DockerHUB

<https://hub.docker.com/repository/docker/ps789/models>

Thank you for taking the time to read the README file for PA2!