

In [29]:

```
%%time

import io
import boto3
import random

def data_split(FILE_DATA, FILE_TRAIN, FILE_VALIDATION, FILE_TEST, Valid_Well_Name, Blind_Well_Name, TARGET_VAR):

    ### FILE_DATA = 'facies_num.csv'
    ### TARGET_VAR = 'PHIND'
    ### FILE_TRAIN = 'facies_train.csv'
    ### FILE_VALIDATION = 'facies_validation.csv'
    ### FILE_TEST = 'facies_test.csv'
    ### Valid_Well_Name = 'SHRIMPLIN'
    ### Blind_Well_Name = 'SHANKLE'

    data = pd.read_csv(FILE_DATA)
    n = data.shape[0]
    ### make first col the target feature
    cols = data.columns.tolist()
    target_pos = data.columns.get_loc(TARGET_VAR)
    cols.pop(target_pos)
    cols = [TARGET_VAR] + cols
    data = data.loc[:,cols]

    ### remove target col from test set

    ### split data into Well names

    test_data = data[data['Well Name'] == Blind_Well_Name]
    test_data = test_data.drop(['Well Name'], axis=1)

    valid_data = data[data['Well Name'] == Valid_Well_Name]
    valid_data = valid_data.drop(['Well Name'], axis=1)

    train_data = data[data['Well Name'] != Blind_Well_Name]
    train_data = train_data[train_data['Well Name'] != Valid_Well_Name]

    train_data = train_data.drop(['Well Name'], axis = 1)

    train_data.to_csv(FILE_TRAIN, index=False, header=False)
    valid_data.to_csv(FILE_VALIDATION, index=False, header=False)
    test_data.to_csv(FILE_TEST, index=False, header=False)

def write_to_s3(fobj, bucket, key):
    return(boto3.Session().resource('s3').Bucket(bucket).Object(key).upload_fileobj(fobj))
```

```
def upload_to_s3(bucket,channel, filename):
    fobj=open(filename, 'rb')
    key = prefix+'/' +channel+'/' +filename
    url = 's3://{}/{}'.format(bucket, key)
    print("XXXXXX")
    print(url)
    print('Writing to {}'.format(url))
    write_to_s3(fobj, bucket, key)
    return(url)
```

CPU times: user 19  $\mu$ s, sys: 0 ns, total: 19  $\mu$ s  
 Wall time: 23.8  $\mu$ s

In [30]:

```
%%time
import boto3, re, sys, math, json, os, sagemaker, urllib.request
from sagemaker import get_execution_role

from time import gmtime, strftime

# Define IAM role

# Each region has its XGBoost container

role = get_execution_role()
prefix = 'xgboost-dm'
containers = {'us-west-2': '433757028032.dkr.ecr.us-west-2.amazonaws.com/xgboost:latest',
              'us-east-1': '811284229777.dkr.ecr.us-east-1.amazonaws.com/xgboost:latest',
              'us-east-2': '825641698319.dkr.ecr.us-east-2.amazonaws.com/xgboost:latest',
              'eu-west-1': '685385470294.dkr.ecr.eu-west-1.amazonaws.com/xgboost:latest'}

my_region = boto3.session.Session().region_name # set the region of the instance

print("Success - the MySageMakerInstance is in the " + my_region + " region. You will use the " + containers[my_region] + " container for your SageMaker endpoint.")
```

Success - the MySageMakerInstance is in the eu-west-1 region. You will use the 685385470294.dkr.ecr.eu-west-1.amazonaws.com/xgboost:latest container for your SageMaker endpoint.  
 CPU times: user 72.9 ms, sys: 121  $\mu$ s, total: 73 ms  
 Wall time: 373 ms

In [31]:

```
%%time
bucket_name = 'mutest-01' # <--- CHANGE THIS VARIABLE TO A UNIQUE NAME FOR YOUR BUCKET

s3 = boto3.resource('s3')
try:
    if my_region == 'us-east-1':
        s3.create_bucket(Bucket=bucket_name)
        print("create bucket")
    else:
        s3.create_bucket(Bucket=bucket_name, CreateBucketConfiguration={ 'LocationConstraint': my_region })
        print('S3 bucket created successfully')
except Exception as e:
    print('S3 error: ',e)
```

S3 error: An error occurred (BucketAlreadyOwnedByYou) when calling the CreateBucket operation: Your previous request to create the named bucket succeeded and you already own it.

CPU times: user 21.6 ms, sys: 172  $\mu$ s, total: 21.8 ms

Wall time: 404 ms

In [42]:

```

%%time
### remove non numeric cols
import pandas as pd

### get from https
### Geology example Oil & Gas

url = 'https://raw.githubusercontent.com/seg/tutorials-2016/master/1610_Facies_classification/facies_vectors.csv'

data = pd.read_csv(url)

### put to local storage

data.to_csv('facies_vectors.csv')

#### remove rows with missing values
data.dropna(inplace=True)

display(data.head())

data['Well Name'].unique()

data = data.loc[:, ['Well Name', 'GR', 'ILD_log10', 'DeltaPHI', 'PHIND']]

### write to disk
data.to_csv('facies_num.csv', index=False)

```

	Facies	Formation	Well Name	Depth	GR	ILD_log10	DeltaPHI	PHIND	PE	NM_M
0	3	A1 SH	SHRIMPLIN	2793.0	77.45	0.664	9.9	11.915	4.6	1
1	3	A1 SH	SHRIMPLIN	2793.5	78.26	0.661	14.2	12.565	4.1	1
2	3	A1 SH	SHRIMPLIN	2794.0	79.05	0.658	14.8	13.050	3.6	1
3	3	A1 SH	SHRIMPLIN	2794.5	86.10	0.655	13.9	13.115	3.5	1
4	3	A1 SH	SHRIMPLIN	2795.0	74.58	0.647	13.5	13.300	3.4	1

CPU times: user 92.8 ms, sys: 4.22 ms, total: 97 ms  
 Wall time: 194 ms

In [85]:

In [86]:

Out[86]:

	PHIND	Well Name	GR	ILD_log10	DeltaPHI
0	11.915	SHRIMPLIN	77.450	0.664	9.900
1	12.565	SHRIMPLIN	78.260	0.661	14.200
2	13.050	SHRIMPLIN	79.050	0.658	14.800
3	13.115	SHRIMPLIN	86.100	0.655	13.900
4	13.300	SHRIMPLIN	74.580	0.647	13.500
5	13.385	SHRIMPLIN	73.970	0.636	14.000
6	13.930	SHRIMPLIN	73.720	0.630	15.600
7	13.920	SHRIMPLIN	75.650	0.625	16.500
8	13.980	SHRIMPLIN	73.790	0.624	16.200
9	14.220	SHRIMPLIN	76.890	0.615	16.900
10	13.375	SHRIMPLIN	76.110	0.600	14.800
11	12.690	SHRIMPLIN	74.950	0.583	13.300
12	12.475	SHRIMPLIN	71.870	0.561	11.300
13	14.930	SHRIMPLIN	83.420	0.537	13.300
14	16.555	SHRIMPLIN	90.100	0.519	14.300
15	15.960	SHRIMPLIN	78.150	0.467	11.800
16	15.120	SHRIMPLIN	69.300	0.438	9.500
17	15.190	SHRIMPLIN	63.540	0.418	8.800
18	15.390	SHRIMPLIN	63.870	0.401	7.200
19	14.885	SHRIMPLIN	58.320	0.386	6.600
20	14.800	SHRIMPLIN	56.610	0.369	5.500
21	14.460	SHRIMPLIN	55.970	0.352	6.100
22	14.745	SHRIMPLIN	63.670	0.344	6.000
23	15.135	SHRIMPLIN	66.200	0.342	6.800
24	15.480	SHRIMPLIN	61.270	0.346	6.100
25	14.675	SHRIMPLIN	69.480	0.354	5.800
26	13.635	SHRIMPLIN	76.370	0.354	5.200
27	15.055	SHRIMPLIN	82.200	0.348	7.400
28	20.230	SHRIMPLIN	90.250	0.346	11.500
29	24.015	SHRIMPLIN	94.380	0.358	14.200
...	...	...	...	...	...
4119	5.382	CHURCHMAN BIBLE	30.734	0.991	1.552

4120	5.055	CHURCHMAN BIBLE	32.219	1.013	1.342
4121	4.739	CHURCHMAN BIBLE	37.688	1.040	0.681
4122	3.533	CHURCHMAN BIBLE	35.844	1.044	0.960
4123	3.337	CHURCHMAN BIBLE	42.156	1.051	1.448
4124	4.051	CHURCHMAN BIBLE	42.094	1.057	2.736
4125	5.893	CHURCHMAN BIBLE	49.719	1.060	3.092
4126	6.503	CHURCHMAN BIBLE	46.219	1.062	3.018
4127	5.958	CHURCHMAN BIBLE	42.313	1.050	2.245
4128	5.936	CHURCHMAN BIBLE	36.031	1.028	1.193
4129	5.978	CHURCHMAN BIBLE	32.594	1.014	0.662
4130	6.605	CHURCHMAN BIBLE	37.094	1.005	0.377
4131	6.270	CHURCHMAN BIBLE	40.031	1.027	0.615
4132	5.871	CHURCHMAN BIBLE	42.500	1.057	0.672
4133	4.479	CHURCHMAN BIBLE	39.719	1.087	0.648
4134	2.686	CHURCHMAN BIBLE	38.844	1.109	1.025
4135	2.320	CHURCHMAN BIBLE	41.719	1.107	0.659
4136	2.937	CHURCHMAN BIBLE	44.750	1.085	1.165
4137	5.013	CHURCHMAN BIBLE	46.469	1.070	1.872
4138	6.445	CHURCHMAN BIBLE	51.000	1.061	3.760
4139	7.325	CHURCHMAN BIBLE	55.563	1.052	4.296
4140	7.465	CHURCHMAN BIBLE	58.313	1.034	3.863
4141	7.541	CHURCHMAN BIBLE	55.344	1.003	2.225
4142	7.295	CHURCHMAN BIBLE	53.313	0.972	1.640
4143	7.149	CHURCHMAN BIBLE	49.594	0.954	1.494
4144	7.254	CHURCHMAN BIBLE	46.719	0.947	1.828
4145	8.013	CHURCHMAN BIBLE	44.563	0.953	2.241
4146	8.013	CHURCHMAN BIBLE	49.719	0.964	2.925
4147	7.708	CHURCHMAN BIBLE	51.469	0.965	3.083
4148	6.668	CHURCHMAN BIBLE	50.031	0.970	2.609

3232 rows × 5 columns

In [77]:

In [87]:

```
FILE_DATA = 'facies_num.csv'
TARGET_VAR = 'PHIND'
FILE_TRAIN = 'facies_train.csv'
FILE_VALIDATION = 'facies_validation.csv'
FILE_TEST = 'facies_test.csv'
Valid_Well_NAME = 'SHRIMPLIN'
Blind_Well_Name = 'SHANKLE'

data_split(FILE_DATA, FILE_TRAIN, FILE_VALIDATION, FILE_TEST, Valid_Well_NAME,
Blind_Well_Name, TARGET_VAR)
```

In [89]:

```
test_data = data[data['Well Name'] == Blind_Well_Name]
```

In [91]:

```
test_data = test_data.drop(['Well Name'], axis=1)
```

In [92]:

```
test_data
```

Out[92]:

	PHIND	GR	ILD_log10	DeltaPHI
937	18.685	98.36	0.642	-0.1
938	16.745	97.57	0.631	7.9
939	14.105	98.41	0.615	12.8
940	13.385	85.92	0.597	13.0
941	13.345	83.16	0.592	12.3
942	13.105	82.56	0.599	12.9
943	14.000	83.12	0.611	13.2
944	13.685	84.26	0.620	12.8
945	13.920	86.29	0.627	13.2
946	13.975	80.95	0.629	13.0
947	14.235	79.07	0.629	13.0
948	13.170	75.31	0.626	11.4
949	12.320	69.30	0.625	10.1
950	11.805	69.16	0.626	9.0
951	11.720	69.03	0.624	10.0

952	11.750	64.95	0.622	10.1
953	12.155	78.95	0.607	12.0
954	12.955	82.36	0.580	11.7
955	12.865	80.27	0.543	11.1
956	12.230	78.87	0.507	8.5
957	11.890	66.38	0.473	6.2
958	12.440	58.63	0.444	4.4
959	13.640	59.16	0.413	3.7
960	14.490	52.86	0.378	2.5
961	15.085	54.32	0.344	1.1
962	14.985	53.27	0.320	-0.3
963	14.945	52.34	0.312	-0.1
964	15.055	58.01	0.310	0.9
965	15.280	55.38	0.307	1.6
966	15.905	55.18	0.301	1.7
...	...	...	...	...
1356	8.715	43.87	0.870	1.9
1357	8.375	47.40	0.872	2.4
1358	8.470	48.75	0.875	2.6
1359	8.200	53.77	0.877	3.1
1360	8.555	65.61	0.880	3.2
1361	8.435	79.98	0.884	3.0
1362	8.760	115.60	0.885	3.6
1363	8.645	201.75	0.887	3.9
1364	8.420	242.75	0.889	4.9
1365	8.515	238.02	0.893	5.1
1366	8.310	215.23	0.899	5.7
1367	8.545	168.12	0.907	5.6
1368	8.855	124.80	0.919	5.2
1369	8.915	102.74	0.932	4.6
1370	8.915	102.74	0.932	4.6
1371	8.860	81.37	0.941	4.3
1372	8.675	70.46	0.943	4.2
1373	9.365	67.29	0.943	4.9



1374	9.555	67.61	0.939	5.6
1375	9.585	68.63	0.941	5.8
1376	9.950	68.45	0.957	6.9
1377	10.390	73.25	0.996	8.4
1378	9.950	73.21	1.039	7.7
1379	7.900	76.24	1.084	4.0
1380	5.215	52.57	1.118	-0.4
1381	3.455	32.84	1.120	-2.2
1382	2.890	25.16	1.112	-1.6
1383	2.890	25.16	1.112	-1.6
1384	3.400	26.22	1.092	-0.4
1385	4.715	65.36	1.026	1.6

449 rows × 4 columns

In [6]:

```
s3_train_loc = upload_to_s3(bucket = bucket_name, channel = 'train', filename
= FILE_TRAIN)
s3_valid_loc = upload_to_s3(bucket = bucket_name, channel = 'validation', file
name = FILE_VALIDATION)
s3_test_loc = upload_to_s3(bucket = bucket_name, channel = 'test', filename =
FILE_TEST)
```

XXXXXX

s3://mutest-01/xgboost-dm/train/facies\_train.csv

Writing to {} s3://mutest-01/xgboost-dm/train/facies\_train.csv

XXXXXX

s3://mutest-01/xgboost-dm/validation/facies\_validation.csv

Writing to {} s3://mutest-01/xgboost-dm/validation/facies\_validati  
on.csv

XXXXXX

s3://mutest-01/xgboost-dm/test/facies\_test.csv

Writing to {} s3://mutest-01/xgboost-dm/test/facies\_test.csv

In [ ]:

```
### For csv the algorithm assumes that the target variable is in teh first col
umn and that the csv does not
### have a header record nor lable cols
```

In [ ]:

```
## instatiate the estimator
```

In [7]:

```
%%time
import sagemaker
xgboost = sagemaker.estimator.Estimator(containers[my_region],
    role,
    train_instance_count=1,
    train_instance_type='ml.c4.xlarge',
    output_path='s3://{}/{}/output'.format(bucket_name,prefix),
    sagemaker_session=sagemaker.Session())
```

CPU times: user 44.9 ms, sys: 492  $\mu$ s, total: 45.4 ms

Wall time: 48.9 ms

In [8]:

```
%%time
xgboost.set_hyperparameters(max_depth=5,
    eta=0.2,
    gamma=4,
    min_child_weight=6,
    subsample=0.8,
    silent=0,
    objective='reg:linear',
    num_round=100)
```

CPU times: user 10  $\mu$ s, sys: 0 ns, total: 10  $\mu$ s

Wall time: 13.8  $\mu$ s

In [9]:

```
s3_input_train = sagemaker.s3_input(s3_data=s3_train_loc, content_type='csv')
s3_input_validation = sagemaker.s3_input(s3_data=s3_valid_loc, content_type='csv')
```

In [10]:

```
%%time
xgboost.fit({'train': s3_input_train, 'validation': s3_input_validation})
```

2019-09-04 14:13:44 Starting - Starting the training job...

2019-09-04 14:13:46 Starting - Launching requested ML instances...

2019-09-04 14:14:43 Starting - Preparing the instances for training  
g.....

2019-09-04 14:15:33 Downloading - Downloading input data...

2019-09-04 14:16:11 Training - Training image download completed.

Training in progress.

2019-09-04 14:16:11 Uploading - Uploading generated training model  
.

Arguments: train

[2019-09-04:14:16:06:INFO] Running standalone xgboost training.

[2019-09-04:14:16:06:INFO] File size need to be processed in the node: 0.07mb. Available memory size in the node: 92.71mb

```
[2019-09-04:14:16:06:INFO] Determined delimiter of CSV input is ',
'

[14:16:06] S3DistributionType set as FullyReplicated
[14:16:06] 2312x3 matrix with 6936 entries loaded from /opt/ml/input/data/train?format=csv&label_column=0&delimiter=,
[2019-09-04:14:16:06:INFO] Determined delimiter of CSV input is ',
'

[14:16:06] S3DistributionType set as FullyReplicated
[14:16:06] 471x3 matrix with 1413 entries loaded from /opt/ml/input/data/validation?format=csv&label_column=0&delimiter=,
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 40 extra nodes, 0 pruned nodes, max_depth=5
[0]#011train-rmse:12.3719#011validation-rmse:10.3818
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 42 extra nodes, 0 pruned nodes, max_depth=5
[1]#011train-rmse:10.3721#011validation-rmse:8.47349
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 44 extra nodes, 0 pruned nodes, max_depth=5
[2]#011train-rmse:8.82776#011validation-rmse:6.98323
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 40 extra nodes, 2 pruned nodes, max_depth=5
[3]#011train-rmse:7.63675#011validation-rmse:5.82869
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 42 extra nodes, 0 pruned nodes, max_depth=5
[4]#011train-rmse:6.74381#011validation-rmse:4.96028
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 40 extra nodes, 2 pruned nodes, max_depth=5
[5]#011train-rmse:6.08023#011validation-rmse:4.35484
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 46 extra nodes, 4 pruned nodes, max_depth=5
[6]#011train-rmse:5.61219#011validation-rmse:3.9177
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 52 extra nodes, 0 pruned nodes, max_depth=5
[7]#011train-rmse:5.27297#011validation-rmse:3.61474
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 50 extra nodes, 0 pruned nodes, max_depth=5
[8]#011train-rmse:5.00942#011validation-rmse:3.44339
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 30 extra nodes, 0 pruned nodes, max_depth=5
[9]#011train-rmse:4.82428#011validation-rmse:3.34229
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 42 extra nodes, 0 pruned nodes, max_depth=5
[10]#011train-rmse:4.67474#011validation-rmse:3.22938
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 42 extra nodes, 0 pruned nodes, max_depth=5
[11]#011train-rmse:4.56737#011validation-rmse:3.17778
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 32 extra nodes, 0 pruned nodes, max_depth=5
[12]#011train-rmse:4.4664#011validation-rmse:3.23029
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 38 extra nodes, 4 pruned nodes, max_depth=5
[13]#011train-rmse:4.38755#011validation-rmse:3.23371
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 44 extra nodes, 0 pruned nodes, max_depth=5
```

```
[14]#011train-rmse:4.32834#011validation-rmse:3.1915
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 38 extra nodes, 0 pruned nodes, max_depth=5
[15]#011train-rmse:4.27737#011validation-rmse:3.19302
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 44 extra nodes, 4 pruned nodes, max_depth=5
[16]#011train-rmse:4.23489#011validation-rmse:3.19369
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 40 extra nodes, 0 pruned nodes, max_depth=5
[17]#011train-rmse:4.19031#011validation-rmse:3.23641
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 22 extra nodes, 2 pruned nodes, max_depth=5
[18]#011train-rmse:4.1648#011validation-rmse:3.22666
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 30 extra nodes, 0 pruned nodes, max_depth=5
[19]#011train-rmse:4.14957#011validation-rmse:3.2572
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 16 extra nodes, 2 pruned nodes, max_depth=5
[20]#011train-rmse:4.13589#011validation-rmse:3.26127
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 32 extra nodes, 0 pruned nodes, max_depth=5
[21]#011train-rmse:4.11156#011validation-rmse:3.27684
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 40 extra nodes, 0 pruned nodes, max_depth=5
[22]#011train-rmse:4.07263#011validation-rmse:3.32018
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 34 extra nodes, 0 pruned nodes, max_depth=5
[23]#011train-rmse:4.04385#011validation-rmse:3.32525
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 36 extra nodes, 0 pruned nodes, max_depth=5
[24]#011train-rmse:4.02604#011validation-rmse:3.3235
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 28 extra nodes, 0 pruned nodes, max_depth=5
[25]#011train-rmse:4.01215#011validation-rmse:3.32214
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 32 extra nodes, 0 pruned nodes, max_depth=5
[26]#011train-rmse:3.95189#011validation-rmse:3.3115
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 26 extra nodes, 0 pruned nodes, max_depth=5
[27]#011train-rmse:3.93089#011validation-rmse:3.31405
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 42 extra nodes, 0 pruned nodes, max_depth=5
[28]#011train-rmse:3.87182#011validation-rmse:3.33876
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 30 extra nodes, 2 pruned nodes, max_depth=5
[29]#011train-rmse:3.85152#011validation-rmse:3.35466
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 42 extra nodes, 0 pruned nodes, max_depth=5
[30]#011train-rmse:3.82793#011validation-rmse:3.38057
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 34 extra nodes, 0 pruned nodes, max_depth=5
[31]#011train-rmse:3.79872#011validation-rmse:3.37546
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 32 extra nodes, 0 pruned nodes, max_depth=5
```

```
[32]#011train-rmse:3.79049#011validation-rmse:3.36703
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 26 extra nodes, 0 pruned nodes, max_depth=5
[33]#011train-rmse:3.76905#011validation-rmse:3.35526
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 32 extra nodes, 2 pruned nodes, max_depth=5
[34]#011train-rmse:3.7456#011validation-rmse:3.37119
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 48 extra nodes, 0 pruned nodes, max_depth=5
[35]#011train-rmse:3.71931#011validation-rmse:3.37605
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 22 extra nodes, 0 pruned nodes, max_depth=5
[36]#011train-rmse:3.70776#011validation-rmse:3.38158
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 36 extra nodes, 2 pruned nodes, max_depth=5
[37]#011train-rmse:3.68145#011validation-rmse:3.37278
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 28 extra nodes, 0 pruned nodes, max_depth=5
[38]#011train-rmse:3.66684#011validation-rmse:3.38404
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 46 extra nodes, 0 pruned nodes, max_depth=5
[39]#011train-rmse:3.64325#011validation-rmse:3.41733
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 30 extra nodes, 0 pruned nodes, max_depth=5
[40]#011train-rmse:3.62092#011validation-rmse:3.40612
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 44 extra nodes, 2 pruned nodes, max_depth=5
[41]#011train-rmse:3.60335#011validation-rmse:3.40994
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 28 extra nodes, 2 pruned nodes, max_depth=5
[42]#011train-rmse:3.59428#011validation-rmse:3.42439
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 34 extra nodes, 4 pruned nodes, max_depth=5
[43]#011train-rmse:3.57535#011validation-rmse:3.42187
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 34 extra nodes, 6 pruned nodes, max_depth=5
[44]#011train-rmse:3.56948#011validation-rmse:3.42592
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 22 extra nodes, 0 pruned nodes, max_depth=5
[45]#011train-rmse:3.55858#011validation-rmse:3.41791
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 30 extra nodes, 4 pruned nodes, max_depth=5
[46]#011train-rmse:3.55071#011validation-rmse:3.41552
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 32 extra nodes, 4 pruned nodes, max_depth=5
[47]#011train-rmse:3.52518#011validation-rmse:3.42209
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 32 extra nodes, 0 pruned nodes, max_depth=5
[48]#011train-rmse:3.51305#011validation-rmse:3.42725
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 18 extra nodes, 0 pruned nodes, max_depth=5
[49]#011train-rmse:3.50516#011validation-rmse:3.42721
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 44 extra nodes, 0 pruned nodes, max_depth=5
```

```
[50]#011train-rmse:3.49263#011validation-rmse:3.45409
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 26 extra nodes, 0 pruned nodes, max_depth=5
[51]#011train-rmse:3.4856#011validation-rmse:3.45858
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 40 extra nodes, 0 pruned nodes, max_depth=5
[52]#011train-rmse:3.46071#011validation-rmse:3.5059
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 56 extra nodes, 0 pruned nodes, max_depth=5
[53]#011train-rmse:3.43032#011validation-rmse:3.52595
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 46 extra nodes, 0 pruned nodes, max_depth=5
[54]#011train-rmse:3.40992#011validation-rmse:3.53671
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 26 extra nodes, 0 pruned nodes, max_depth=5
[55]#011train-rmse:3.40086#011validation-rmse:3.53023
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 34 extra nodes, 2 pruned nodes, max_depth=5
[56]#011train-rmse:3.39504#011validation-rmse:3.56338
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 46 extra nodes, 0 pruned nodes, max_depth=5
[57]#011train-rmse:3.381#011validation-rmse:3.56348
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 28 extra nodes, 6 pruned nodes, max_depth=5
[58]#011train-rmse:3.37513#011validation-rmse:3.56838
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 28 extra nodes, 0 pruned nodes, max_depth=5
[59]#011train-rmse:3.36361#011validation-rmse:3.58151
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 32 extra nodes, 2 pruned nodes, max_depth=5
[60]#011train-rmse:3.35814#011validation-rmse:3.58854
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 40 extra nodes, 2 pruned nodes, max_depth=5
[61]#011train-rmse:3.33342#011validation-rmse:3.58408
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 34 extra nodes, 2 pruned nodes, max_depth=5
[62]#011train-rmse:3.32113#011validation-rmse:3.58667
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 38 extra nodes, 0 pruned nodes, max_depth=5
[63]#011train-rmse:3.30855#011validation-rmse:3.58817
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 28 extra nodes, 6 pruned nodes, max_depth=5
[64]#011train-rmse:3.30458#011validation-rmse:3.58932
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 30 extra nodes, 0 pruned nodes, max_depth=5
[65]#011train-rmse:3.29097#011validation-rmse:3.59932
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 32 extra nodes, 0 pruned nodes, max_depth=5
[66]#011train-rmse:3.27622#011validation-rmse:3.59747
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 26 extra nodes, 0 pruned nodes, max_depth=5
[67]#011train-rmse:3.27408#011validation-rmse:3.60109
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 26 extra nodes, 0 pruned nodes, max_depth=5
```



```
[68]#011train-rmse:3.26468#011validation-rmse:3.60828
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 34 extra nodes, 0 pruned nodes, max_depth=5
[69]#011train-rmse:3.24976#011validation-rmse:3.60718
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 48 extra nodes, 0 pruned nodes, max_depth=5
[70]#011train-rmse:3.2204#011validation-rmse:3.63519
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 24 extra nodes, 0 pruned nodes, max_depth=5
[71]#011train-rmse:3.21416#011validation-rmse:3.64633
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 48 extra nodes, 0 pruned nodes, max_depth=5
[72]#011train-rmse:3.19875#011validation-rmse:3.64252
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 26 extra nodes, 2 pruned nodes, max_depth=5
[73]#011train-rmse:3.18893#011validation-rmse:3.64365
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 52 extra nodes, 0 pruned nodes, max_depth=5
[74]#011train-rmse:3.16515#011validation-rmse:3.6367
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 18 extra nodes, 0 pruned nodes, max_depth=5
[75]#011train-rmse:3.15783#011validation-rmse:3.64633
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 32 extra nodes, 0 pruned nodes, max_depth=5
[76]#011train-rmse:3.14598#011validation-rmse:3.66603
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 26 extra nodes, 2 pruned nodes, max_depth=5
[77]#011train-rmse:3.13756#011validation-rmse:3.67388
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 44 extra nodes, 0 pruned nodes, max_depth=5
[78]#011train-rmse:3.11554#011validation-rmse:3.66371
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 40 extra nodes, 0 pruned nodes, max_depth=5
[79]#011train-rmse:3.0873#011validation-rmse:3.66093
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 40 extra nodes, 6 pruned nodes, max_depth=5
[80]#011train-rmse:3.07078#011validation-rmse:3.65492
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 36 extra nodes, 0 pruned nodes, max_depth=5
[81]#011train-rmse:3.05598#011validation-rmse:3.67403
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 40 extra nodes, 0 pruned nodes, max_depth=5
[82]#011train-rmse:3.04689#011validation-rmse:3.68914
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 24 extra nodes, 2 pruned nodes, max_depth=5
[83]#011train-rmse:3.04088#011validation-rmse:3.68579
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 24 extra nodes, 0 pruned nodes, max_depth=5
[84]#011train-rmse:3.03631#011validation-rmse:3.69059
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 46 extra nodes, 0 pruned nodes, max_depth=5
[85]#011train-rmse:3.01702#011validation-rmse:3.69377
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 30 extra nodes, 0 pruned nodes, max_depth=5
```

```
[86]#011train-rmse:3.00679#011validation-rmse:3.69508
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 42 extra nodes, 0 pruned nodes, max_depth=5
[87]#011train-rmse:2.99212#011validation-rmse:3.68192
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 30 extra nodes, 0 pruned nodes, max_depth=5
[88]#011train-rmse:2.98216#011validation-rmse:3.67763
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 28 extra nodes, 0 pruned nodes, max_depth=5
[89]#011train-rmse:2.97839#011validation-rmse:3.67361
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 26 extra nodes, 0 pruned nodes, max_depth=5
[90]#011train-rmse:2.97103#011validation-rmse:3.66288
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 24 extra nodes, 4 pruned nodes, max_depth=5
[91]#011train-rmse:2.96733#011validation-rmse:3.66188
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 32 extra nodes, 0 pruned nodes, max_depth=5
[92]#011train-rmse:2.95412#011validation-rmse:3.67431
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 44 extra nodes, 0 pruned nodes, max_depth=5
[93]#011train-rmse:2.94002#011validation-rmse:3.67096
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 40 extra nodes, 0 pruned nodes, max_depth=5
[94]#011train-rmse:2.9275#011validation-rmse:3.68211
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 16 extra nodes, 2 pruned nodes, max_depth=5
[95]#011train-rmse:2.92416#011validation-rmse:3.68685
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 26 extra nodes, 2 pruned nodes, max_depth=5
[96]#011train-rmse:2.91891#011validation-rmse:3.6884
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 36 extra nodes, 2 pruned nodes, max_depth=5
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 32 extra nodes, 0 pruned nodes, max_depth=5
[97]#011train-rmse:2.91525#011validation-rmse:3.68276
[98]#011train-rmse:2.90508#011validation-rmse:3.7062
[14:16:06] src/tree/updater_prune.cc:74: tree pruning end, 1 roots
, 40 extra nodes, 4 pruned nodes, max_depth=5
[99]#011train-rmse:2.89767#011validation-rmse:3.71381
```

2019-09-04 14:16:18 Completed - Training job completed

Billable seconds: 46

CPU times: user 388 ms, sys: 28.2 ms, total: 416 ms

Wall time: 3min 11s



In [11]:

```
%%time
xgboost_predictor = xgboost.deploy(initial_instance_count=1, instance_type='ml
.m4.xlarge')
```

```
-----
-----!CPU times: user 480 ms, sys: 31 m
s, total: 511 ms
Wall time: 8min 14s
```

In [ ]:

```
from sagemaker.predictor import csv_serializer
```

In [12]:

```
%%time
import numpy as np

from sagemaker.predictor import csv_serializer

### def predict
def predict(data):
    rows=len(data)
    xgboost_predictor.content_type = 'text/csv'
    xgboost_predictor.serializer = csv_serializer
    xgboost_predictor.deserializer = None

    split_array = np.array_split(data, int(data.shape[0] / float(rows) + 1))
    predictions = ''
    for array in split_array:
        predictions = ','.join([predictions, xgboost_predictor.predict(array).
decode('utf-8')])
    return np.fromstring(predictions[1:], sep=',')

test_data = pd.read_csv(FILE_TEST, header=None)
labels = test_data.iloc[:,0]
predictions = predict(test_data.as_matrix()[1:,1:])
```

```
CPU times: user 24 ms, sys: 0 ns, total: 24 ms
Wall time: 185 ms
```

```
/home/ec2-user/anaconda3/envs/python3/lib/python3.6/site-packages/
ipykernel/__main__.py:20: FutureWarning: Method .as_matrix will be
removed in a future version. Use .values instead.
```

In [ ]:

In [23]:

```

%matplotlib inline
import pandas as pd
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
import matplotlib.colors as colors
##import mpl_toolkits.axes_grid1
###import make_axes_locatable
from mpl_toolkits.axes_grid1 import make_axes_locatable, axes_size

def log_plot(logs):
    logs = logs.sort_values(by='Depth')

    ztop = logs.Depth.min(); zbot=logs.Depth.max()

    f, ax = plt.subplots(nrows=1, ncols=4,figsize=(8,12))
    ax[0].plot(logs.GR, logs.Depth, '-g')
    ax[1].plot(logs.ILD_log10, logs.Depth, '-')
    ax[2].plot(logs.DeltaPHI, logs.Depth, '-', color='0.5')
    ax[3].plot(logs.PHIND, logs.Depth, '-', color='r')
    ax[3].plot(logs.PPor, logs.Depth, '-', color='blue')
    ax[3].legend(['Measured', 'Predicted'])

    for i in range(len(ax)):
        ax[i].set_ylim(ztop, zbot)
        ax[i].invert_yaxis()
        ax[i].grid()
        ax[i].locator_params(axis='x', nbins=3)

    ax[0].set_xlabel("GR")
    ax[0].set_xlim(logs.GR.min(), logs.GR.max())
    ax[1].set_xlabel("ILD_log10")
    ax[1].set_xlim(logs.ILD_log10.min(), logs.ILD_log10.max())
    ax[2].set_xlabel("DeltaPHI")
    ax[2].set_xlim(logs.DeltaPHI.min(), logs.DeltaPHI.max())
    ax[3].set_xlabel("Porosity")
    ax[3].set_xlim(logs.PHIND.min(), logs.PHIND.max())

    ax[1].set_yticklabels([]); ax[2].set_yticklabels([]); ax[3].set_yticklabels([])
    f.suptitle('Well: %s'%logs.iloc[0]['Well Name'], fontsize=14, y=0.94)

```

In [26]:

```
np.corrcoef(predictions, labels)
```

Out[26]:

```
array([[1.          , 0.83889661],
       [0.83889661, 1.          ]])
```

In [28]:

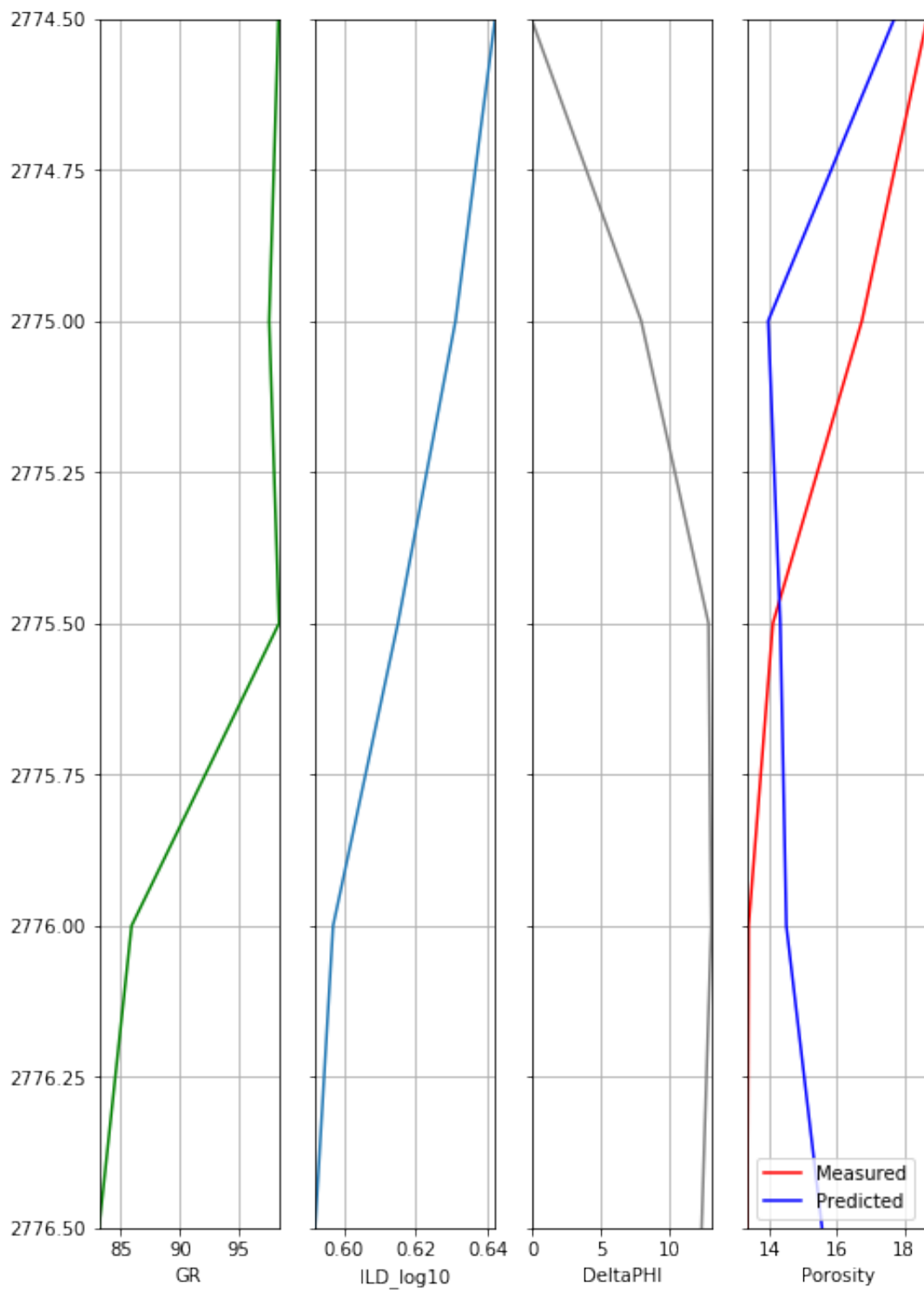
```
data = pd.read_csv('facies_vectors.csv')
Blindwell = data[data['Well Name'] == Blind_Well_Name]
Blindwell['PPoro'] = predictions
display(Blindwell.head())
log_plot(Blindwell.head())
log_plot(Blindwell)
##print('Correlation coeficient = {0:.5f} \nMean Squared Error = {1:.5f}'.format(
    np.corrcoef(predictions, labels)))
```

/home/ec2-user/anaconda3/envs/python3/lib/python3.6/site-packages/  
ipykernel/\_\_main\_\_.py:3: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

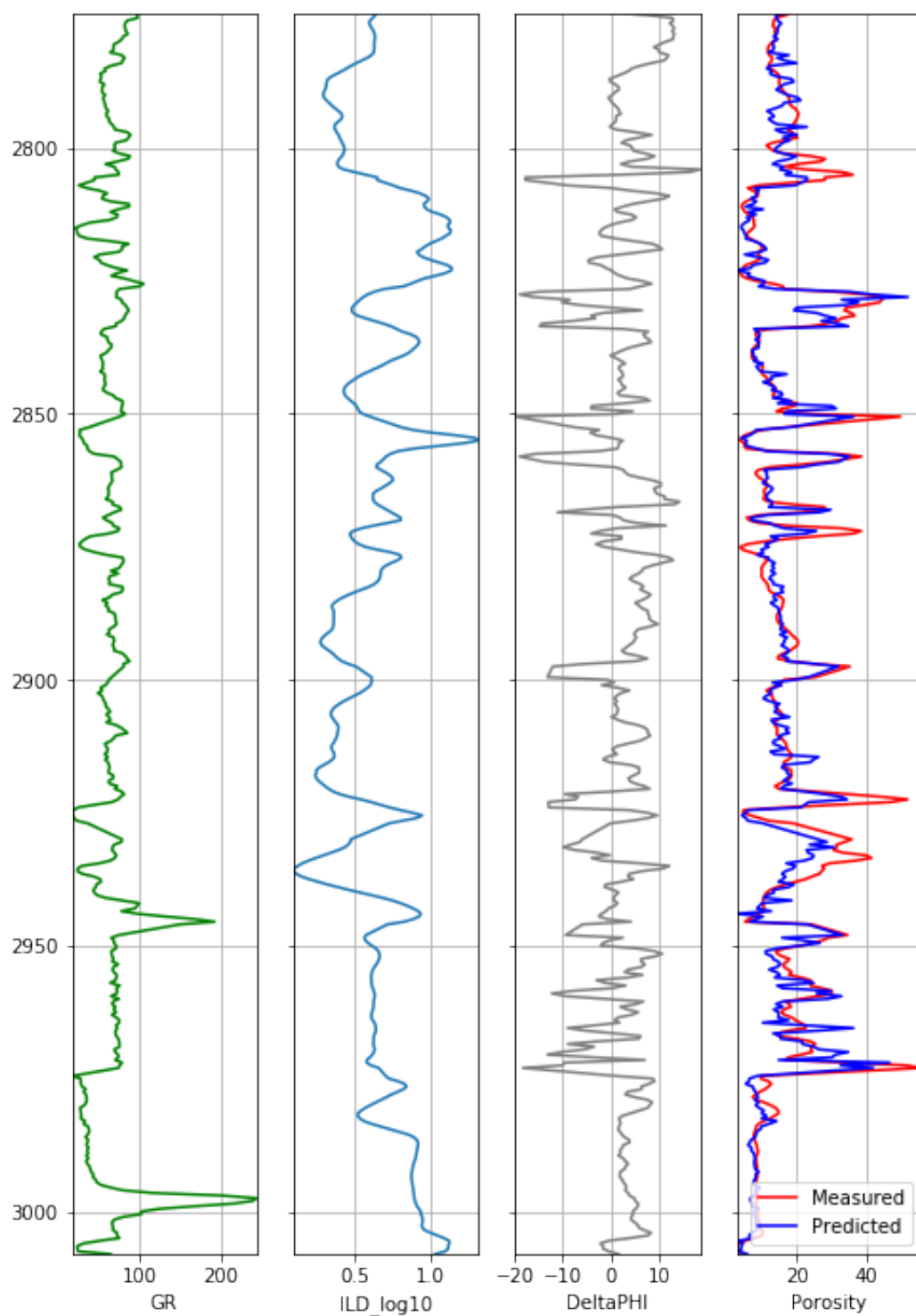
See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>  
app.launch\_new\_instance()

Unnamed: 0	Facies	Formation	Well Name	Depth	GR	ILD_log10	DeltaPHI	PHIND
937	937	2	A1 SH SHANKLE	2774.5	98.36	0.642	-0.1	18.685
938	938	2	A1 SH SHANKLE	2775.0	97.57	0.631	7.9	16.745
939	939	2	A1 SH SHANKLE	2775.5	98.41	0.615	12.8	14.105
940	940	2	A1 SH SHANKLE	2776.0	85.92	0.597	13.0	13.385
941	941	2	A1 SH SHANKLE	2776.5	83.16	0.592	12.3	13.345

# Well: SHANKLE



# Well: SHANKLE



In [ ]:

In [27]:

```
print(xgboost_predictor.endpoint)
import sagemaker
sagemaker.Session().delete_endpoint(xgboost_predictor.endpoint)
```

xgboost-2019-09-04-14-13-44-678

In [ ]:

```
!ls
```

In [ ]: