

In [1]:

```
%%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))
```

```

### xgboost-dm/train/facies_train.csv
### s3://mutest-01/xgboost-dm/train/facies_train.csv
### Writing to {} s3://mutest-01/xgboost-dm/train/facies_train.csv

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

```

CPU times: user 132 ms, sys: 15.7 ms, total: 147 ms

Wall time: 150 ms

In [ ]:

In [2]:

```

### These are utilities
#####
### KEEP
session = boto3.Session()
credentials = session.get_credentials()
credentials = credentials.get_frozen_credentials()
access_key = credentials.access_key
secret_key = credentials.secret_key

```

In [ ]:

```

### KEEP
import boto3
s3_client = boto3.client('s3')
s3_resource = boto3.resource('s3')

```

In [ ]:

```

### KEEP
!aws s3 ls s3://'mutest-01'/xgboost-dm/train/

```

In [ ]:

```

### KEEP
!aws s3 ls s3://'mutest-01'/xgboost-dm/

```

In [ ]:

```
### KEEP
#### 'mutest-01'
import boto
import boto.s3

conn = boto.s3.connect_to_region('eu-west-1')
bucket = conn.get_bucket('mutest-01')

folders = bucket.list("", "/")
for folder in folders:
    print(folder.name)
```

In [ ]:

```
#### KEEP
import boto3
s3 = boto3.resource('s3')
for bucket in s3.buckets.all():
    print(bucket.name)
```

In [ ]:

```
import boto3
s3 = boto3.resource('s3')

my_bucket = s3.Bucket('mutest-01')

for file in my_bucket.objects.all():
    print(file.key)
```

In [ ]:

```
s3 = boto3.client('s3')
s3.list_objects_v2(Bucket='mutest-01')

##### END OF UTILITIES
```

In [ ]:

In [3]:

```
%%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 717 ms, sys: 63.8 ms, total: 781 ms  
Wall time: 1.08 s

In [ ]:

In [4]:

```
%%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 63.8 ms, sys: 7.64 ms, total: 71.4 ms

Wall time: 493 ms

In [5]:

```

%%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 119 ms, sys: 549  $\mu$ s, total: 120 ms  
 Wall time: 344 ms

In [ ]:

In [6]:

```
%%time
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)
```

CPU times: user 29.8 ms, sys: 54  $\mu$ s, total: 29.8 ms  
Wall time: 28.7 ms

In [ ]:

In [7]:

```
%%time
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)
```

```
xgboost-dm/train/facies_train.csv
s3://mutest-01/xgboost-dm/train/facies_train.csv
Writing to {} s3://mutest-01/xgboost-dm/train/facies_train.csv
xgboost-dm/validation/facies_validation.csv
s3://mutest-01/xgboost-dm/validation/facies_validation.csv
Writing to {} s3://mutest-01/xgboost-dm/validation/facies_validati
on.csv
xgboost-dm/test/facies_test.csv
s3://mutest-01/xgboost-dm/test/facies_test.csv
Writing to {} s3://mutest-01/xgboost-dm/test/facies_test.csv
CPU times: user 184 ms, sys: 8.59 ms, total: 193 ms
Wall time: 391 ms
```

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 [ ]:

```
## instantiate the estimator
```

In [8]:

```
%%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())

print('s3://{}/{}'/output'.format(bucket_name,prefix),)
```

```
s3://mutest-01/xgboost-dm/output/
CPU times: user 39.6 ms, sys: 3.66 ms, total: 43.3 ms
Wall time: 45.5 ms
```

In [9]:

```
%%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 µs, sys: 0 ns, total: 10 µs
Wall time: 13.6 µs
```

In [10]:

```
%%time
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')
```

```
CPU times: user 11 µs, sys: 1 µs, total: 12 µs
Wall time: 16.5 µs
```

In [ ]:

```
##### TRAIN #####
```



In [11]:

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

```
2019-09-07 12:01:08 Starting - Starting the training job...
2019-09-07 12:01:10 Starting - Launching requested ML instances.
..
2019-09-07 12:02:05 Starting - Preparing the instances for training.....
2019-09-07 12:03:03 Downloading - Downloading input data
2019-09-07 12:03:03 Training - Downloading the training image..
Arguments: train
[2019-09-07:12:03:22:INFO] Running standalone xgboost training.
[2019-09-07:12:03:22:INFO] File size need to be processed in the
node: 0.07mb. Available memory size in the node: 79.92mb
[2019-09-07:12:03:22:INFO] Determined delimiter of CSV input is
','
[12:03:22] S3DistributionType set as FullyReplicated
[12:03:22] 2312x3 matrix with 6936 entries loaded from /opt/ml/i
nput/data/train?format=csv&label_column=0&delimiter=,
[2019-09-07:12:03:22:INFO] Determined delimiter of CSV input is
','
[12:03:22] S3DistributionType set as FullyReplicated
[12:03:22] 471x3 matrix with 1413 entries loaded from /opt/ml/in
put/data/validation?format=csv&label_column=0&delimiter=,
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 roo
ts, 40 extra nodes, 0 pruned nodes, max_depth=5
[0]#011train-rmse:12.3719#011validation-rmse:10.3818
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 roo
ts, 42 extra nodes, 0 pruned nodes, max_depth=5
[1]#011train-rmse:10.3721#011validation-rmse:8.47349
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 roo
ts, 44 extra nodes, 0 pruned nodes, max_depth=5
[2]#011train-rmse:8.82776#011validation-rmse:6.98323
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 roo
ts, 40 extra nodes, 2 pruned nodes, max_depth=5
[3]#011train-rmse:7.63675#011validation-rmse:5.82869
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 roo
ts, 42 extra nodes, 0 pruned nodes, max_depth=5
[4]#011train-rmse:6.74381#011validation-rmse:4.96028
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 roo
ts, 40 extra nodes, 2 pruned nodes, max_depth=5
[5]#011train-rmse:6.08023#011validation-rmse:4.35484
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 roo
ts, 46 extra nodes, 4 pruned nodes, max_depth=5
[6]#011train-rmse:5.61219#011validation-rmse:3.9177
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 roo
ts, 52 extra nodes, 0 pruned nodes, max_depth=5
[7]#011train-rmse:5.27297#011validation-rmse:3.61474
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 roo
ts, 50 extra nodes, 0 pruned nodes, max_depth=5
[8]#011train-rmse:5.00942#011validation-rmse:3.44339
```

```
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
```

```
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 roots, 40 extra nodes, 2 pruned nodes, max_depth=5
```

```
ts, 34 extra nodes, 2 pruned nodes, max_depth=5
[62]#011train-rmse:3.32113#011validation-rmse:3.58667
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 38 extra nodes, 0 pruned nodes, max_depth=5
[63]#011train-rmse:3.30855#011validation-rmse:3.58817
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 28 extra nodes, 6 pruned nodes, max_depth=5
[64]#011train-rmse:3.30458#011validation-rmse:3.58932
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 30 extra nodes, 0 pruned nodes, max_depth=5
[65]#011train-rmse:3.29097#011validation-rmse:3.59932
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 32 extra nodes, 0 pruned nodes, max_depth=5
[66]#011train-rmse:3.27622#011validation-rmse:3.59747
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 26 extra nodes, 0 pruned nodes, max_depth=5
[67]#011train-rmse:3.27408#011validation-rmse:3.60109
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 26 extra nodes, 0 pruned nodes, max_depth=5
[68]#011train-rmse:3.26468#011validation-rmse:3.60828
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 34 extra nodes, 0 pruned nodes, max_depth=5
[69]#011train-rmse:3.24976#011validation-rmse:3.60718
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 48 extra nodes, 0 pruned nodes, max_depth=5
[70]#011train-rmse:3.2204#011validation-rmse:3.63519
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 24 extra nodes, 0 pruned nodes, max_depth=5
[71]#011train-rmse:3.21416#011validation-rmse:3.64633
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 48 extra nodes, 0 pruned nodes, max_depth=5
[72]#011train-rmse:3.19875#011validation-rmse:3.64252
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 26 extra nodes, 2 pruned nodes, max_depth=5
[73]#011train-rmse:3.18893#011validation-rmse:3.64365
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 52 extra nodes, 0 pruned nodes, max_depth=5
[74]#011train-rmse:3.16515#011validation-rmse:3.6367
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 18 extra nodes, 0 pruned nodes, max_depth=5
[75]#011train-rmse:3.15783#011validation-rmse:3.64633
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 32 extra nodes, 0 pruned nodes, max_depth=5
[76]#011train-rmse:3.14598#011validation-rmse:3.66603
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 26 extra nodes, 2 pruned nodes, max_depth=5
[77]#011train-rmse:3.13756#011validation-rmse:3.67388
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 44 extra nodes, 0 pruned nodes, max_depth=5
[78]#011train-rmse:3.11554#011validation-rmse:3.66371
[79]#011train-rmse:3.0873#011validation-rmse:3.66093
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 root
ts, 40 extra nodes, 0 pruned nodes, max_depth=5
```



```
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] 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
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 roots, 36 extra nodes, 2 pruned nodes, max_depth=5
```

```
[97]#011train-rmse:2.91525#011validation-rmse:3.68276
[12:03:22] src/tree/updater_prune.cc:74: tree pruning end, 1 roots, 32 extra nodes, 0 pruned nodes, max_depth=5
[98]#011train-rmse:2.90508#011validation-rmse:3.7062
[12:03:22] 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
```

In [12]:

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

```
-----
-----!CPU times: user 475 ms, sys: 39.1
ms, total: 514 ms
Wall time: 8min 14s
```

In [ ]:

In [ ]:

In [49]:

```
%%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:])
```

CPU times: user 8.32 ms, sys: 4.67 ms, total: 13 ms

Wall time: 39.7 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 [52]:

```
%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 [53]:

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

Out[53]:

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

In [129]:

```
xerror = Blindwell['PHIND'] - Blindwell['PPoro']
```

In [132]:

```
import seaborn as sns
sns.set()
from matplotlib import pyplot as plt
%matplotlib inline
```

In [277]:

```
# same plotting code as above!

ztop = Blindwell.Depth.min(); zbot=Blindwell.Depth.max()
```

In [267]:

```
### DO THIS ONCE ###
Blindwell = Blindwell.reset_index()
```

In [ ]:

In [285]:

```
import matplotlib.pyplot as plt

### DO THIS ONCE Blindwell = Blindwell.reset_index()

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

d1= Blindwell['PPoro']

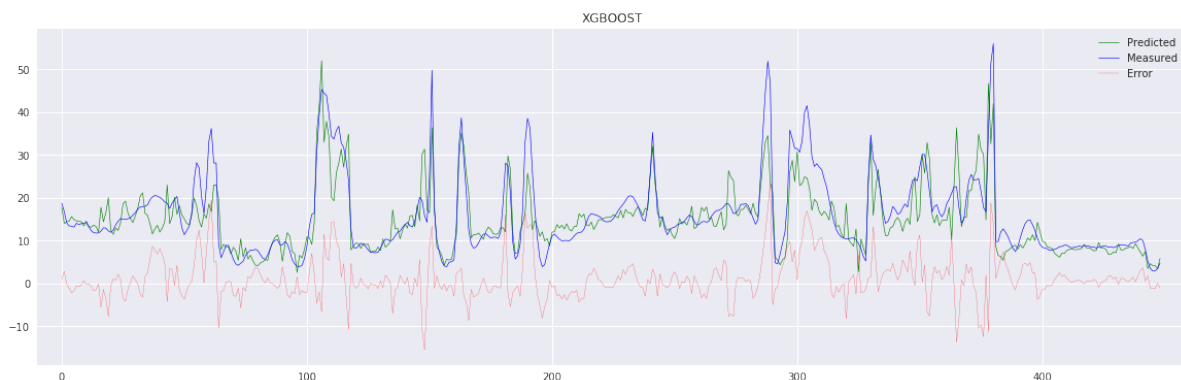
d2 = Blindwell['PHIND']

d3 = d2 - d1

def plotter():
    plt.title('XGBOOST')
    plt.plot(d1, 'g-', linewidth=0.6, label='Predicted')
    plt.plot(d2, 'b-', linewidth=0.6, label='Measured')
    plt.plot(d3, 'r-', linewidth=0.2, label='Error')
    plt.xlim(-10, len(Blindwell) + 10)
    plt.legend()
    plt.show()

plt.figure(figsize=(20,6))

plotter()
```



In [ ]:

```
### now take the varience and std of the error
```

In [75]:

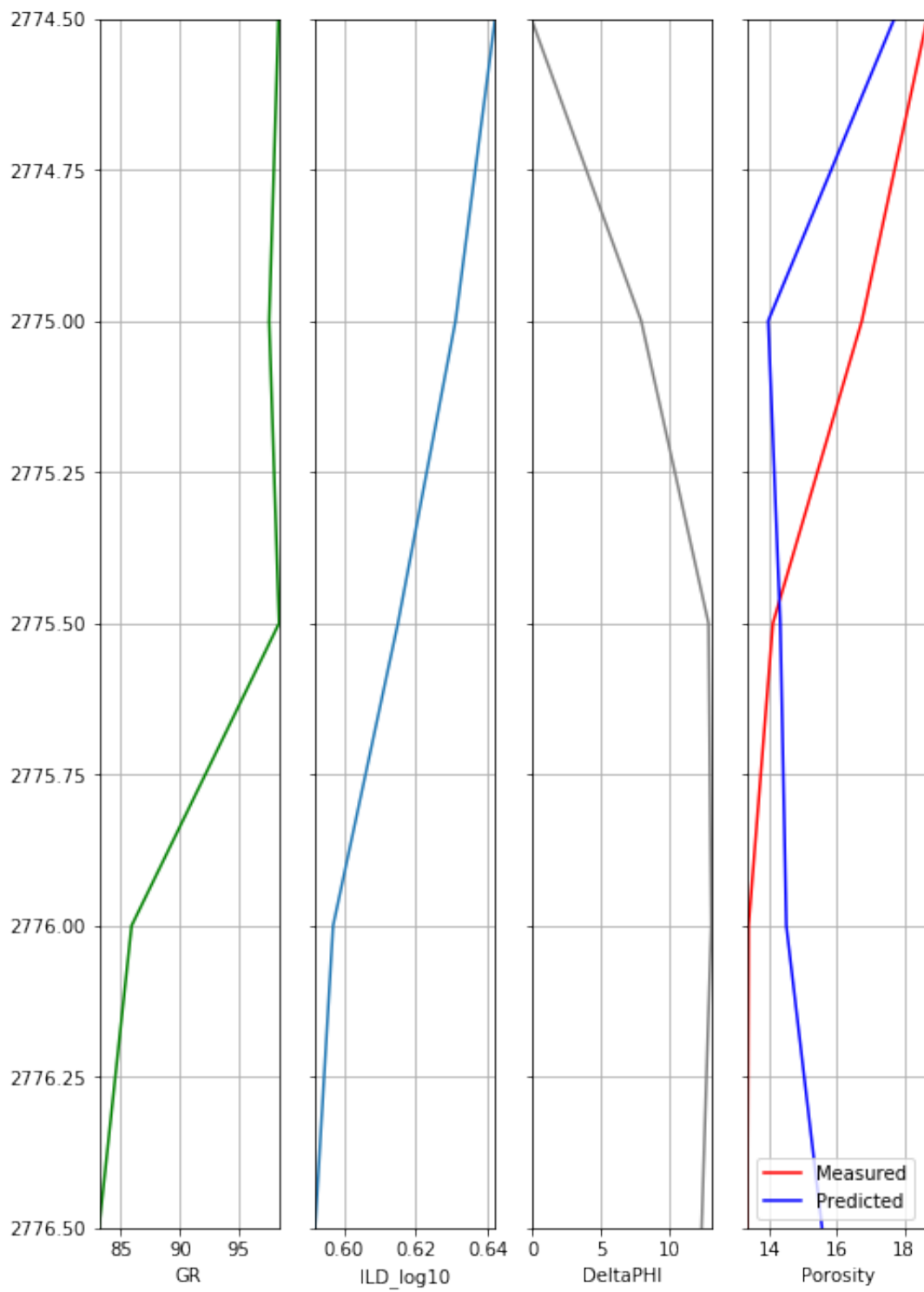
```
data = pd.read_csv('facies_vectors.csv')
Blindwell = data[data['Well Name'] == Blind_Well_Name]
Blindwell['PPoro'] = predictions
display(Blindwell.head())
###log_plot(Blindwell.head())
log_plot(Blindwell)
#print('Correlation coeficient = {0:.5f} \nMean Squared Error = {1:.5f}'.format(np.corrcoef(predictions,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

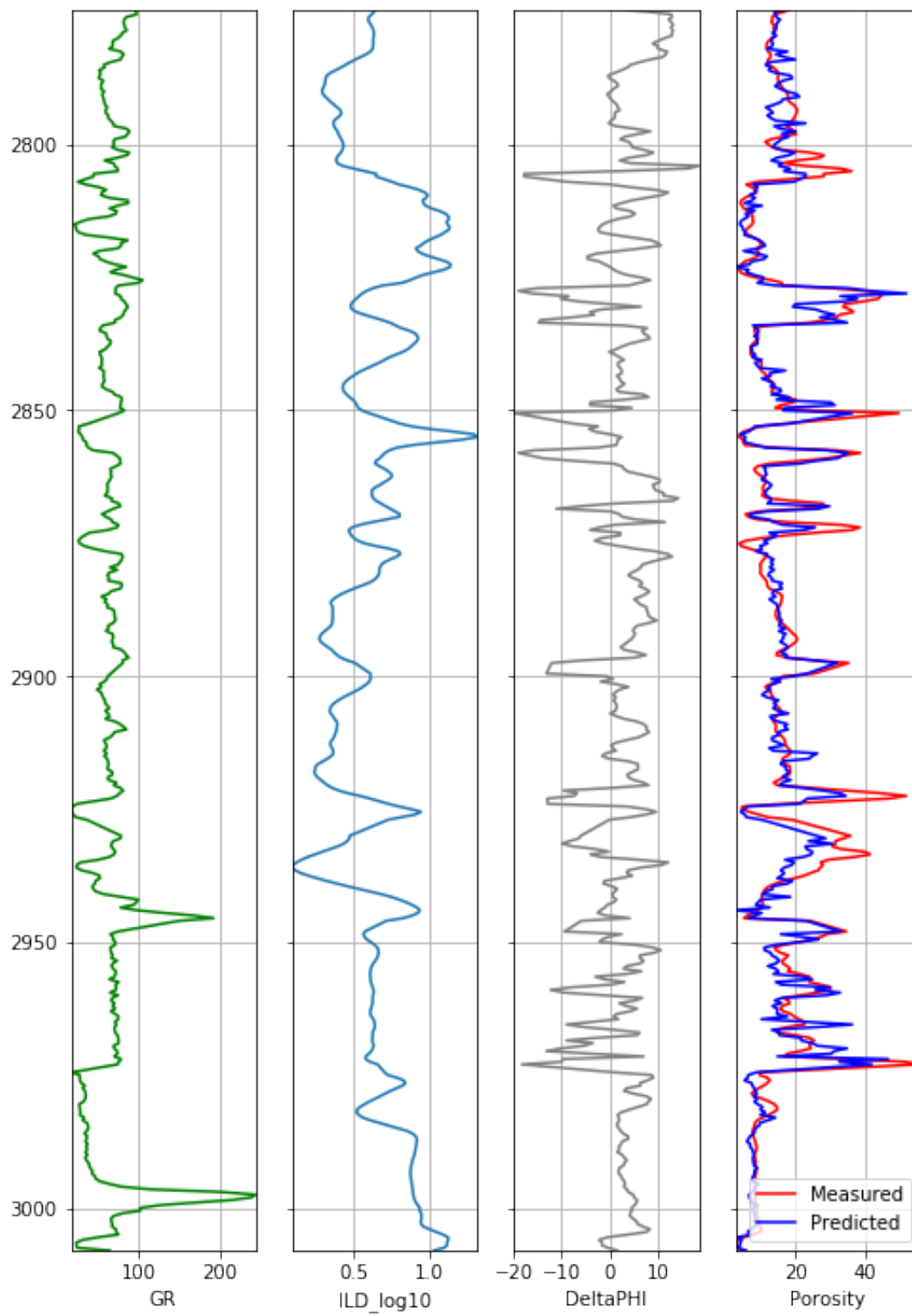
See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>  
app.launch\_new\_instance()

| 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 [50]:

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

xgboost-2019-09-07-12-01-07-834

In [ ]:

```
!ls
```

In [ ]: