

# BOSTON AQI ANALYSIS

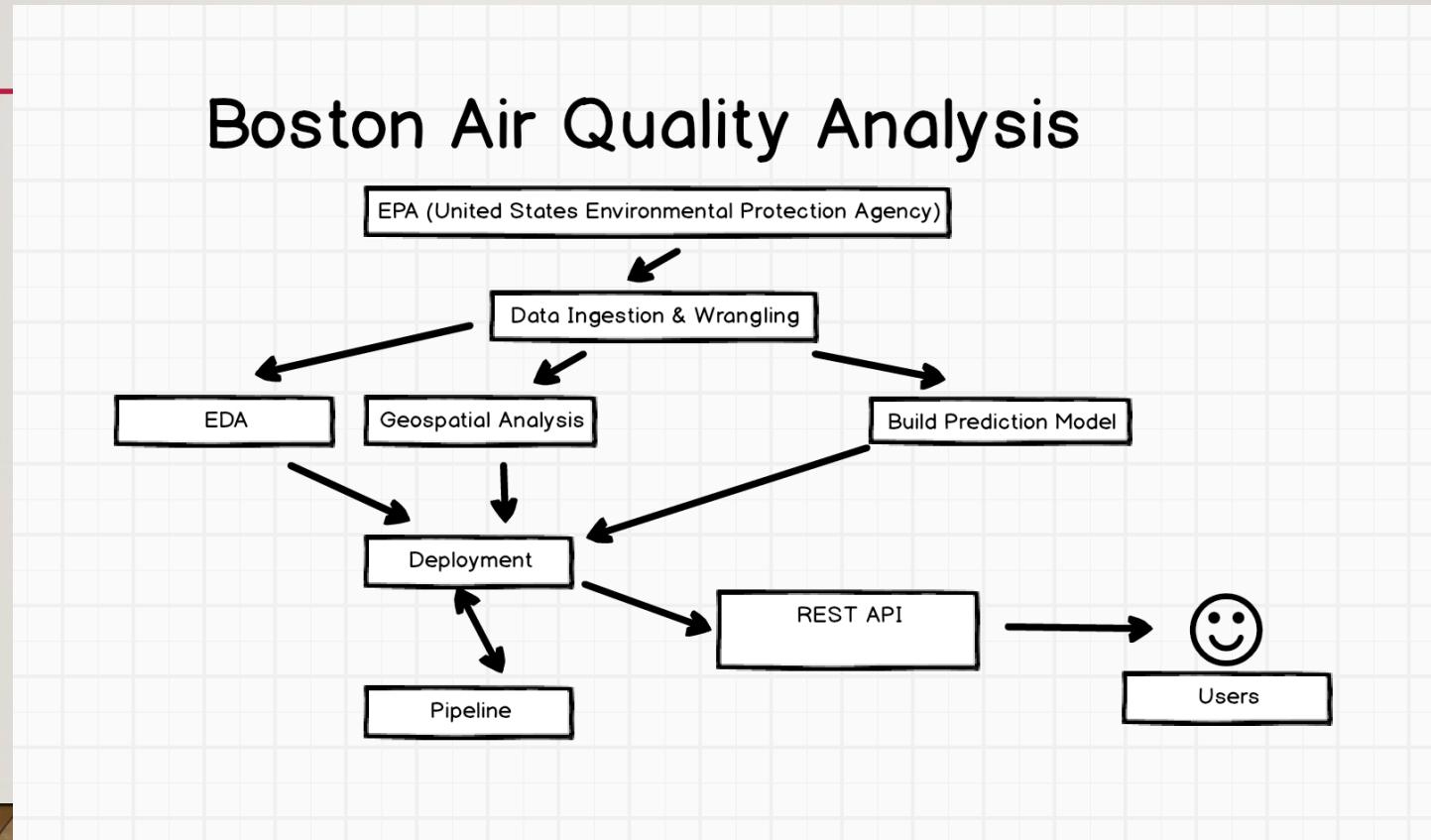
---

HONGWEI HU

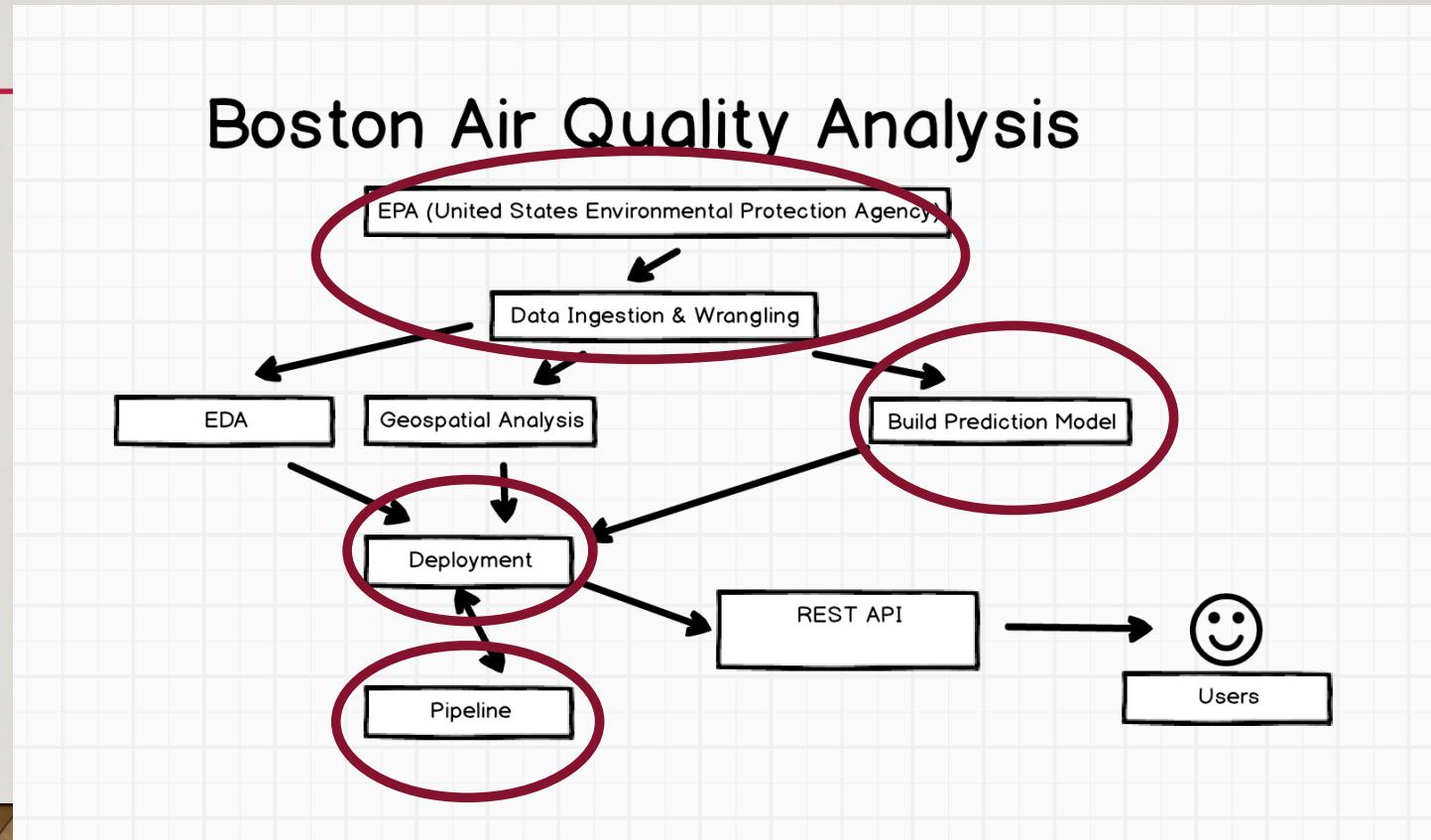
GUANGNAN LIANG



# WORKFLOW



# WORKFLOW



# WORKFLOW

---

- Data Ingestion & wrangling Automation
- Azure MI Model
- Pipeline for regular fetching & retraining
- Deployment



# I. DATA INGESTION & WRANGLING AUTOMATION

---

- Docker Hub: besimilar/advanced-data-analysis:aqi
- GitHub: <https://github.com/Besimilar/Air-Quality-Analysis/tree/master/I-Data-Ingestion>



# I. DATA INGESTION & WRANGLING AUTOMATION

---

- Prepare “env.list”



# I. DATA INGESTION & WRANGLING AUTOMATION

Branch: master ▾ [Air-Quality-Analysis](#) / env.list

 **Besimilar** Instruction Update ada1215 5 hours ago

1 contributor

12 lines (12 sloc) | 150 Bytes

[Raw](#) [Blame](#) [History](#)   

```
1 AWSACCESS=
2 AWSSECRET=
3 REGION=us-west-2
4 EPAUSERNAME=
5 EPAPASSWORD=
6 BDATE=20140101
7 EDATE=20170817
8 ACCOUNTNAME=
9 ACCOUNTKEY=
10 CONTAINERNAME=
11 APIKEY=
12 APIURL=
```

# I. DATA INGESTION & WRANGLING AUTOMATION

---

- Prepare “env.list”
  - 1. AWS params:
    - AWSACCESS: aws access key
    - AWSSECRET: aws private key
    - REGION: aws region
  - 2. EPA API params:
    - EPAUSERNAME: EPA API username
    - EPAPASSWORD: EPA API password
    - BDATE: the first day of dataset (better within 3 years, try to set around 20140101, its a bug-free choice)
    - EDATE: the last day of dataset: (default: current date)



# I. DATA INGESTION & WRANGLING AUTOMATION

---

- Prepare “env.list”
  - **BDATE:** the first day of dataset (better within 3 years, try to set around 20140101, its a bug-free choice)
    - You might get error, if you fetch large data from EPA api, because they seem to limit my account.
    - You also might get error, if you fetch small data, because the data might not contain some predictive features, either wrangling or model retraining could produce "no features" errors.



# I. DATA INGESTION & WRANGLING AUTOMATION

---

- Prepare “env.list”
  - 3.Azure params:
    - ACCOUNTNAME: azure storage account name
    - ACCOUNTKEY: azure storage primary key
    - CONTAINERNAME: azure storage container name
  - 4.Azure ML params:
    - APIKEY: predictive service key
    - APIURL: predictive service url



# I. DATA INGESTION & WRANGLING AUTOMATION

---

- Prepare “env.list”
  - 5. Other default params:
    - change them in "config.json" After starting container
    - You don't need to change them to do a demo



# DATA INGESTION & WARNING AUTOMATION

Branch: master ▾ Air-Quality-Analysis / 1-Data-Ingestion / src / config.json Find file Copy path

Besimilar Pipeline src Update 3f6431a 16 hours ago

1 contributor

18 lines (15 sloc) | 367 Bytes Raw Blame History

```
1  {
2    "name": "boston-aqi",
3    "assignNum": "final",
4    "team": 3,
5
6    "AWSAccess": "your aws accesskey",
7    "AWSSecret": "your aws secretkey",
8    "region": "your aws region",
9
10   "username": "your epa username",
11   "password": "your epa password",
12   "data_format": "DMCSV",
13   "parameter_class": "FORECAST",
14   "begin_date": "20130101",
15   "state": "25",
16   "county": "025"
17 }
```

# I. DATA INGESTION & WRANGLING AUTOMATION

---

- Command Lines:

- \$ docker run --env-file env.list -w /Final/I-Data-Ingestion-Wrangling/data-ingestion -it  
besimilar/advanced-data-analysis:aqi
- \$ ./run.sh



# I. DATA INGESTION & WRANGLING AUTOMATION

---

- Fetch rawdata from EPA API ("src/fetchdata.py"):



# I. DATA INGESTION & WRANGLING AUTOMATION

```
[root@bcd23214fbff: /Final/1-Data-Ingestion-Wrangling/data-ingestion] docker run --env-file env.list -w /Final/1-Data-Ingestion-Wrangling/data-ingestion -it besimilar/advanced-data-analysis:aqi
[Hongweis-MacBook-Pro:src Similar$ ls
awbservice.py config.json  env.list    fetchdata.py  main.py      run.sh      wrangling.py
[Hongweis-MacBook-Pro:src Similar$ vi env.list
[Hongweis-MacBook-Pro:src Similar$ docker run --env-file env.list -w /Final/1-Data-Ingestion-Wrangling/data-ingestion -it besimilar/advanced-data-analysis:aqi
[root@bcd23214fbff:/Final/1-Data-Ingestion-Wrangling/data-ingestion# ./run.sh
##### Fetch Data From EPA #####
```

# I. DATA INGESTION & WRANGLING AUTOMATION

---

- Clean rawdata and save both rawdata and cleandata to local ("src/wrangling.py")



# I. DATA INGESTION & WRANGLING AUTOMATION

```
[Hongweis-MacBook-Pro:src Similar$ vi env.list
[Hongweis-MacBook-Pro:src Similar$ ls
awservice.py config.json  env.list      fetchdata.py  main.py      run.sh      wrangling.py
[Hongweis-MacBook-Pro:src Similar$ docker run --env-file env.list -w /Final/1-Data-Ingestion-Wrangling/data-ingestion -it besimilar/advanced-data-analysis:aqi
[root@225973f74b3e:/Final/1-Data-Ingestion-Wrangling/data-ingestion# ./run.sh
##### Fetch Data From EPA #####
##### Data From EPA Download as rawdata-20140101-20170818.txt #####
##### Starting Cleansing Raw DataSet #####
main.py:76: DtypeWarning: Columns (0,22) have mixed types. Specify dtype option on import or set low_memory=False.
  main()
#####
DataSet Cleansing Completed #####
Skip: Bucket(team3assignmentfinal) Already Created.
Upload: Success
#####
Find logs in ddmmmyyHHMMSS.log #####
[root@225973f74b3e:/Final/1-Data-Ingestion-Wrangling/data-ingestion# ls
180817044008.log  __pycache__  awbservice.py  cleandata-20140101-20170818.csv  config.json  fetchdata.py  main.py  rawdata-20140101-20170818.txt  run.sh  wrangling.py
root@225973f74b3e:/Final/1-Data-Ingestion-Wrangling/data-ingestion# ]
```

# I. DATA INGESTION & WRANGLING AUTOMATION

---

- Upload cleandata to AWS S3 ("src/awsservice.py"):



# I. DATA INGESTION & WRANGLING AUTOMATION

The screenshot shows the Amazon S3 console interface. At the top, there's a breadcrumb navigation: 'Amazon S3 > team3assignmentfinal'. Below it is a navigation bar with four tabs: 'Overview' (highlighted), 'Properties', 'Permissions', and 'Management'. A search bar with placeholder text 'Type a prefix and press Enter to search. Press ESC to clear.' is positioned below the navigation bar. Underneath are buttons for 'Upload', '+ Create folder', and 'More'. To the right of these buttons is the region 'US West (Oregon)' and a refresh icon. A message 'Viewing 1 to 6' is displayed above the file list. The main area lists six CSV files:

Name	Last modified	Size	Storage class
cleandata-20130101-20140101.csv	Aug 16, 2017 11:28:57 PM	12.8 MB	Standard
cleandata-20130101-20170801.csv	Aug 16, 2017 12:31:16 AM	46.9 MB	Standard
cleandata-20130101-20170816.csv	Aug 16, 2017 9:01:15 PM	46.9 MB	Standard
cleandata-20140101-20170816.csv	Aug 18, 2017 12:31:07 AM	34.2 MB	Standard
cleandata-20140101-20170818.csv	Aug 18, 2017 12:49:07 AM	34.2 MB	Standard
cleandata-20150101-20150201.csv	Aug 16, 2017 12:17:06 AM	600.7 KB	Standard

At the bottom of the list, another 'Viewing 1 to 6' message is visible.

# I. DATA INGESTION & WRANGLING AUTOMATION

---

- Maintenance Issue:
  - The EPA API won't work on Sunday from 1800 to 2359. So avoid running this project during that time.



# I. DATA INGESTION & WRANGLING AUTOMATION

- Maintenance
- Test
- The

## HTTP status code 500

Internal server error. We have weekly scheduled maintenance Sundays from 1800 - 2359 eastern standard time (GMT-5). If you are getting this message outside that time window, we are experiencing unexpected problems. You may check our news feed at <http://www3.epa.gov/airquality/airdata/rssairdata.xml> or email aqsdatamart@epa.gov for additional information.

## HTTP status code 503

Our API is down. We have weekly scheduled maintenance Sundays from 1800 - 2359 eastern standard time (GMT-5). If you are getting this message outside that time window, we are experiencing unexpected problems. You may check our news feed at <http://www3.epa.gov/airquality/airdata/rssairdata.xml> or email aqsdatamart@epa.gov for additional information.

object during



## 2. AZURE ML MODEL

---

- Azure Gallery:
  - <https://gallery.cortanaintelligence.com/Experiment/Boosted-Decision-Tree>



## 2. AZURE ML MODEL

---

- Y: PM2.5 AQI



## 2. AZURE ML MODEL

- X

Barometric pressure  
Black carbon PM2.5 STP  
Carbon monoxide  
Nitric oxide (NO)  
Nitrogen dioxide (NO<sub>2</sub>)  
Outdoor Temperature  
Oxides of nitrogen (NO<sub>x</sub>)  
Ozone  
PM10 - LC  
PM10 Total 0-10um STP  
PM10-2.5 - Local Conditions  
Reactive oxides of nitrogen (NO<sub>y</sub>)  
Relative Humidity  
Solar radiation  
Sulfate (TSP) STP  
Sulfur dioxide  
Total NMOC (non-methane organic compound)  
Wind Direction - Resultant  
Wind Direction - Scalar  
Wind Speed - Resultant  
Wind Speed - Scalar

## 2. AZURE ML MODEL

---

- Algorithms:
  - Boosted Decision Tree
  - Neural Network
  - Decision Forest



## 2. AZURE ML MODEL

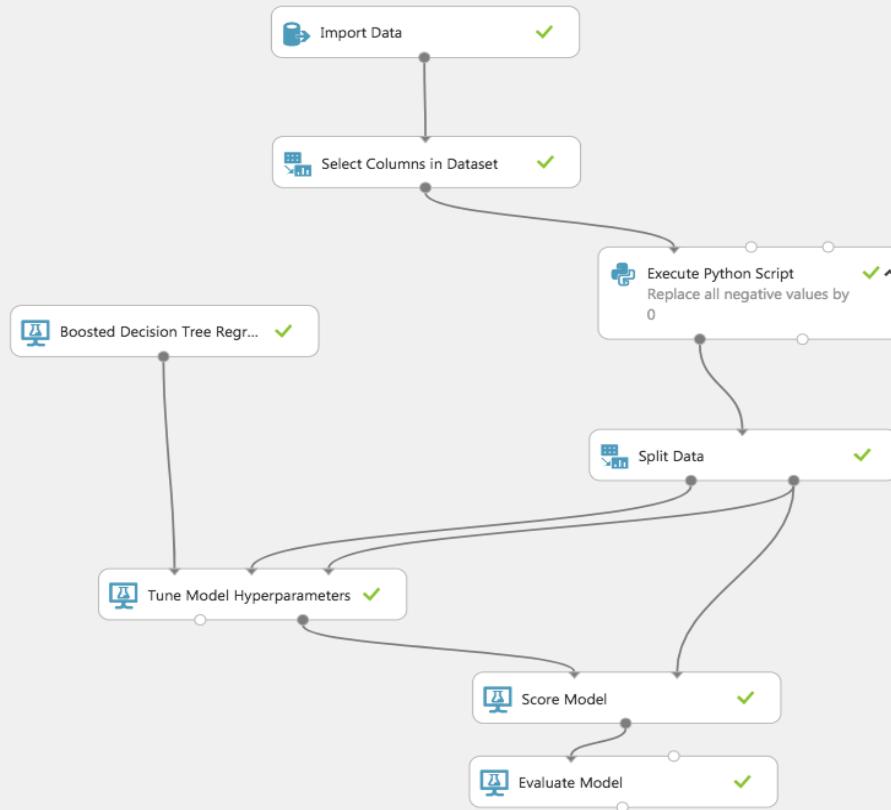
---

- Best model: **Boosted Decision Tree Regression**



## 2. A

- Be



Boosted-Decision-Tree ➤ Tune Model Hyperparameters ➤ Sweep results

rows    columns

180    9

Number of leaves	Minimum leaf instances	Learning rate	Number of trees	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error	Coefficient of Determination
128	10	0.1	500	0.904103	2.108509	0.292878	0.231865	0.768135
128	10	0.2	500	0.914603	2.12629	0.296279	0.235792	0.764208
128	50	0.1	500	0.919372	2.134582	0.297824	0.237635	0.762365
128	50	0.2	500	0.934542	2.140759	0.302738	0.239012	0.760988
128	1	0.2	500	0.924579	2.142636	0.299511	0.239431	0.760569

view as

Statistics and Visualizations

### 3. PIPELINE FOR REGULAR FETCHING & RETRAINING

---

- Docker Hub: besimilar/advanced-data-analysis:pipeline
- GitHub: <https://github.com/Besimilar/Air-Quality-Analysis/tree/master/4-Pipeline>



## 3. PIPELINE FOR REGULAR FETCHING & RETRAINING

---

- For demo:
  - the job will be run every 2 hour. If you need to run every day, please uncomment code in "src/Pipeline/celery.py"



### 3 PIPELINES FOR REGULAR FETCHING &

R

```
'run-every-2-hour': {
    'task': 'tasks.run',
    'schedule': crontab(minute=0, hour='*/2')
}

# Executes every Day morning at 8:30 a.m.
# 'run-every-day-morning': {
#     'task': 'tasks.run',
#     # 'schedule': crontab(hour=8, minute=30),
# },
```

ent code in

## 3. PIPELINE FOR REGULAR FETCHING & RETRAINING

---

- Prepare “env.list” (**same as Part I**)



### 3. PIPELINE FOR REGULAR FETCHING & RETRAINING

---

- Command Lines:
  - \$ docker run --env-file env.list -w /Final/2-Pipeline/ -it besimilar/advanced-data-analysis:pipeline /bin/bash -c "sudo rabbitmq-server -detached; /bin/sleep 5; celery -A Pipeline worker -B -l info"
  - View Results:
    - \$ docker start <containerID>
    - \$ docker exec -it <containerID> /bin/bash



## 3. PIPELINE FOR REGULAR FETCHING & RETRAINING

---

- Start Job (notice the time: 2:06)



```
[Hongwei-MacBook-Pro:2-Pipeline_Similar$ vi env.list
[Hongwei-MacBook-Pro:2-Pipeline_Similar$ docker run --env-file env.list -w /Final/2-Pipeline/ -it besimilar/advanced-data-analysis:pipeline /bin/bash -c "sudo rabbitmq-server -detac]
[ed; /bin/sleep 5; celery -A Pipeline worker -B -l info"
Warning: PID file not written; -detached was passed.
/usr/local/lib/python3.5/dist-packages/celery/platforms.py:795: RuntimeWarning: You're running the worker with superuser privileges: this is
absolutely not recommended!

Please specify a different user using the -u option.

User information: uid=0 euid=0 gid=0 egid=0

    uid=uid, euid=euid, gid=gid, egid=egid,
/usr/local/lib/python3.5/dist-packages/celery/backends/amqp.py:68: CPendingDeprecationWarning:
    The AMQP result backend is scheduled for deprecation in      version 4.0 and removal in version v5.0.      Please use RPC backend or a persistent backend.

    alternative='Please use RPC backend or a persistent backend.')

----- celery@5e249b0fd625 v4.1.0 (latentcall)
*** 
--- * -- Linux-4.9.36-moby-x86_64-with-Ubuntu-16.04-xenial 2017-08-18 02:06:49
-- * - *** 
-- ** ----- [config]
-- ** ----- .> app:      Pipeline:0x7f9ec9897240
-- ** ----- .> transport: amqp://guest:**@localhost:5672//
-- ** ----- .> results:   amqp://
-- *** - * --- .> concurrency: 4 (prefork)
-- ***** ---- .> task events: OFF (enable -E to monitor tasks in this worker)
-- **** ----- 

----- [queues]
-- > celery      exchange=celery(direct) key=celery

[tasks]
. tasks.run

[2017-08-18 02:06:49.482: INFO/MainProcess] Connected to amqp://guest:**@127.0.0.1:5672/
[2017-08-18 02:06:49.502: INFO/MainProcess] mingle: searching for neighbors
[2017-08-18 02:06:50.522: INFO/MainProcess] mingle: all alone
[2017-08-18 02:06:50.566: INFO/MainProcess] celery@5e249b0fd625 ready.
[2017-08-18 02:06:50.729: INFO/Beat] beat: Starting...
```

### **3. PIPELINE FOR REGULAR FETCHING & RETRAINING**

---

- 1<sup>st</sup> Job Running (4:00)



### 3. PIP RETR

---

- 1<sup>st</sup> Job R

```
[2017-08-18 02:06:50,522: INFO/MainProcess] mingle: all alone
[2017-08-18 02:06:50,566: INFO/MainProcess] celery@5e249b0fd625 ready.
[2017-08-18 02:06:50,729: INFO/Beat] beat: Starting...
[2017-08-18 04:00:00,107: INFO/Beat] Scheduler: Sending due task run-every-2-hour (tasks.run)
[2017-08-18 04:00:00,114: INFO/MainProcess] Received task: tasks.run[b1f14771d-82e9-42ad-a7a7-4f8957abb8fd]
##### Fetch Data From EPA #####
##### Data From EPA Download as rawdata-20140101-20170818.txt #####
##### Starting Cleansing Raw DataSet #####
main.py:96: DtypeWarning: Columns (0,22) have mixed types. Specify dtype option on import or set low_memory=False.
    main()
##### DataSet Cleansing Completed #####
Skip: Bucket(team3assignmentfinal) Already Created.
Upload: Success
Start Retrain Model...
Submitting the job...
Job ID: 70b0bf63071e40f8be905840d9f5eddf
Starting the job...
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf not yet started...
The Job has been running for 1min... You need to wait for almost 20min
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf running...
The Job has been running for 2min... You need to wait for almost 20min
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf running...
The Job has been running for 3min... You need to wait for almost 20min
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf running...
The Job has been running for 4min... You need to wait for almost 20min
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf running...
The Job has been running for 5min... You need to wait for almost 20min
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf running...
The Job has been running for 6min... You need to wait for almost 20min
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf running...
The Job has been running for 7min... You need to wait for almost 20min
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf running...
The Job has been running for 8min... You need to wait for almost 20min
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf running...
The Job has been running for 9min... You need to wait for almost 20min
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf running...
The Job has been running for 10min... You need to wait for almost 20min
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf running...
The Job has been running for 11min... You need to wait for almost 20min
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf running...
The Job has been running for 12min... You need to wait for almost 20min
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf running...
The Job has been running for 13min... You need to wait for almost 20min
```

### **3. PIPELINE FOR REGULAR FETCHING & RETRAINING**

---

- 2<sup>nd</sup> Job Running (6:00)



### 3.

RE

- 2

```
Job 70b0bf63071e40f8be905840d9f5eddf running...
The Job has been running for 16min... You need to wait for almost 20min
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf running...
The Job has been running for 17min... You need to wait for almost 20min
Checking the job status...
Job 70b0bf63071e40f8be905840d9f5eddf finished!
The results for output2 are available at the following Azure Storage location:
BaseLocation: https://info7390.blob.core.windows.net/
RelativeLocation: info7390/retrain/retrain-model-20140101-20170818.csv
SasBlobToken: ?sv=2015-02-21&sr=b&sig=fEjtWhLW8P5wgFj5zRxDs24Jd5b68W3tYdoTunoFc%3D&st=2017-08-18T08%3A21%3A51Z&se=2017-08-19T08%3A26%3A51Z&sp=r
Reading the result from https://info7390.blob.core.windows.net/info7390/retrain/retrain-model-20140101-20170818.csv?sv=2015-02-21&r=b&sig=fEjtWhLW8P5wgFj5zRxDs24Jd5b68W3tYdoTunoFc%3D&st=2017-08-18T08%3A21%3A51Z&se=2017-08-19T08%3A26%3A51Z&sp=r
The results for output2 have been written to the file retrain-model-20140101-20170818.csv
Retrain Model Completed.

##### Find logs in ddmmmyyHMMSS.log #####
[2017-08-18 04:26:52,069: INFO/ForkPoolWorker-2] Task tasks.run[bf14771d-82e9-42ad-a7a7-4f8957abb8fd] succeeded in 1611.9537109859994s: 8
[2017-08-18 06:00:00,047: INFO/Beat] Scheduler: Sending due task run-every-2-hour (tasks.run)
[2017-08-18 06:00:00,052: INFO/MainProcess] Received task: tasks.run[025c9b66-77ff-4a9e-8254-e544e2903e9f]
##### Fetch Data From EPA #####
##### Data From EPA Download as rawdata-20140101-20170818.txt #####
##### Starting Cleansing Raw DataSet #####
main.py:96: DtypeWarning: Columns (0,22) have mixed types. Specify dtype option on import or set low_memory=False.
    main()
##### DataSet Cleansing Completed #####
Skip: Bucket(team3assignmentfinal) Already Created.
Skip: File(cleandata-20140101-20170818.csv) Already Exists.
Start Retrain Model...
Submitting the job...
Job ID: e72a5f98c12b4e6fa41cd2a612772f84
Starting the job...
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 not yet started...
The Job has been running for 1min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 2min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 3min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 4min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 5min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 6min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 7min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 8min... You need to wait for almost 20min
Checking the job status...
```

## 3. PIPEL RETRAI

---

- Job End

```
The Job has been running for 2min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 3min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 4min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 5min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 6min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 7min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 8min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 9min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 10min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 11min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 12min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 13min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 14min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 15min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 running...
The Job has been running for 16min... You need to wait for almost 20min
Checking the job status...
Job e72a5f98c12b4e6fa41cd2a612772f84 finished!
The results for output2 are available at the following Azure Storage location:
BaseLocation: https://info7390.blob.core.windows.net/
RelativeLocation: info7390/retrain/retrain-model-20140101-20170818.csv
SasBlobToken: ?sv=2015-02-21&sr=b&sig=ijyWghPNKFng5F2ofucSK2FS1HeGuckT4FSzW5VxITAjs%3D&st=2017-08-18T10%3A18%3A40Z&se=2017-08-19T10%3A23%3A40Z&sp=r
Reading the result from https://info7390.blob.core.windows.net/info7390/retrain/retrain-model-20140101-20170818.csv?sv=2015-02-21&sr=b&sig=ijyWghPNKFng5F2ofucSK2FS1HeGuckT4FSzW5VxITAjs%3D&st=2017-08-18T10%3A18%3A40Z&se=2017-08-19T10%3A23%3A40Z&sp=r
The results for output2 have been written to the file retrain-model-20140101-20170818.csv
Retrain Model Completed.
##### Find logs in ddmmmyyHHMMSS.log #####
[2017-08-18 06:23:41,401: INFO/ForkPoolWorker-4] Task tasks.run[025c9b86-77ff-4a9e-8254-e544e2903e9f] succeeded in 1421.347720531001s: 0
```

### 3. PIPELINE FOR REGULAR FETCHING & RETRAINING

---

- Job End
  - It will download retrain result to container as "retrain-model-bdate-edate.csv"
  - You can also download retrain result from the link shown in terminal



```
Last login: Fri Aug 18 02:07:26 on ttys003
[Hongweis-MacBook-Pro:~ Similar$ docker ps
CONTAINER ID        IMAGE               COMMAND                  CREATED             STATUS              PORTS               NAMES
5e249b0fd625        besimilar/advanced-data-analysis:pipeline   "/bin/bash -c 'sud..."   5 hours ago       Up 5 hours          8888/tcp            nervous_ptolemy
[Hongweis-MacBook-Pro:~ Similar$ docker exec -it 5e249b0fd625 /bin/bash
[root@5e249b0fd625:/Final/2-Pipeline# ls
180817040000.log  Pipeline      awbservice.py    cleandata-20140101-20170818.csv  fetchdata.py  rawdata-20140101-20170818.txt  retrain.py
180817060000.log  __pycache__  celerybeat-schedule.db config.json           main.py      retrain-model-20140101-20170818.csv  wrangling.py
[root@5e249b0fd625:/Final/2-Pipeline# vi retrain-model-20140101-20170818.csv
[root@5e249b0fd625:/Final/2-Pipeline# vi 180817040000.log
[root@5e249b0fd625:/Final/2-Pipeline# ]
```



```
Similar — root@5e249b0fd625: /Final/2-Pipeline — docker exec -it 5e249b0fd625 /bin/bash — 175x24
Mean Absolute Error,Root Mean Squared Error,Relative Absolute Error,Relative Squared Error,Coefficient of Determination
0.85862998191252,2.08091186467001,0.321038411880295,0.276683316869996,0.723316683130004
~
~
```

"retrain-model-20140101-20170818.csv" [dos] 2L, 210C

1,1 All

Quickstart    Dashboard    **Batch Request Log**    Configure    Consume    Test    Swagger API

[← Boosted-Decision-Tree-Retrain](#)

## default

[View in Studio ↗](#)

Search

 CANCEL

JOB ID	STATUS	RUN START	RUN END	DURATION	Result
<input type="checkbox"/> 408be185cd8648fe8ce99ba09fda3c81	Running	08/18/2017 10:10 AM	--	0s	--
<input type="checkbox"/> 743f5cec38c34fa0a0cc5c4d274ede6e	Finished	08/18/2017 08:08 AM	08/18/2017 08:25 AM	17m	output2
<input type="checkbox"/> e72a5f98c12b4e6fa41cd2a612772f84	Finished	08/18/2017 06:06 AM	08/18/2017 06:22 AM	15m 51s	output2
<input type="checkbox"/> 70b0bf63071e40f8be905840d9f5eddf	Finished	08/18/2017 04:09 AM	08/18/2017 04:26 AM	16m 46s	output2

## 4. DEPLOYMENT

---

- Azure Model
  - Azure Gallery:
    - <https://gallery.cortanaintelligence.com/Experiment/Boosted-Decision-Tree>



## 4 DEPLOYMENT

EXPERIMENT



### Boosted-Decision-Tree

- 

Hongwei Hu • August 18, 2017

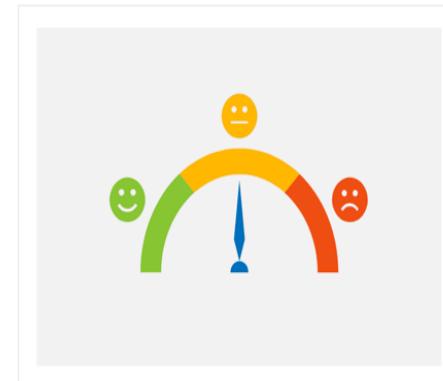
Be the first to like.



#### Summary

For Boston PM2.5 AQI Prediction.

#### Description



[Open in Studio](#)

[+ Add to Collection](#)

## 4. DEPLOYMENT

---

- For Docker images
  - You can use “docker pull”
  - Or You can download from links, and run your own demos
- Links:
  - Data Ingestion & Wrangling: <https://s3-us-west-2.amazonaws.com/team3assignmentfinal/boston-aqi.tar>
  - Pipelines: <https://s3-us-west-2.amazonaws.com/team3assignmentfinal/boston-aqi-pipeline.tar>
- Command lines: \$ docker load --input <dockerimages>.tar

