

AWS (Amazon Web Service)

빅데이터팀 강선아
2017. 3. 28

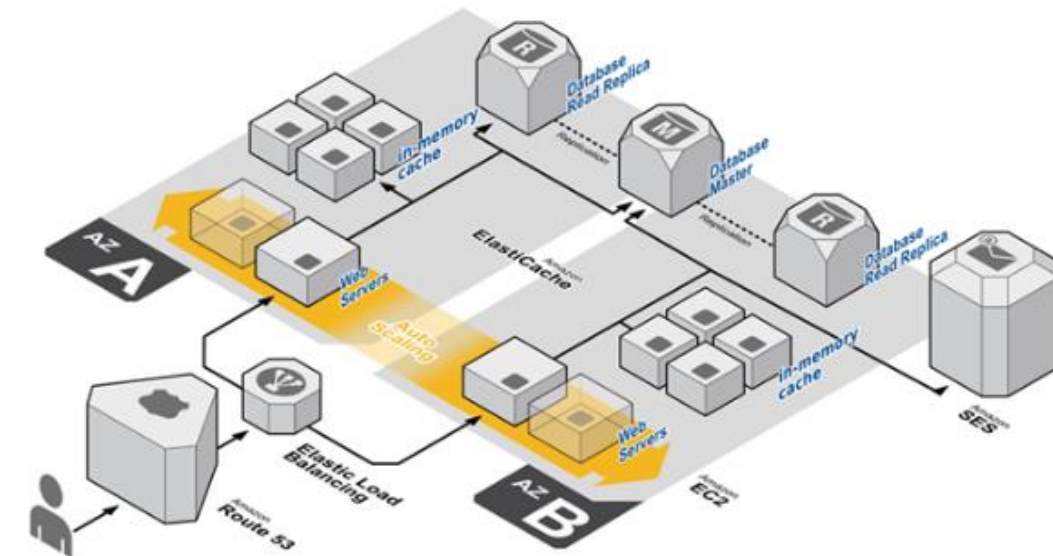
목차

- I. AWS란?
- II. AWS 계정 만들기
- III. AWS 저장공간에 데이터 업로드
- IV. AWS를 이용하여 예측모형 구축
- V. AWS의 장점 및 단점

1. AWS란?

아마존 웹 서비스(Amazon Web Service; AWS)

아마존에서 제공하는 서버 인프라를 온라인으로 이용할 수 있는 IaaS(Infrastructure as a service) 크게 컴퓨팅, storage, content delivery, database, networking 서비스를 제공



- EC2: 가상 서버
- CloudWatch: 리소스 모니터링 서비스
- S3: 스토리지 서비스
- Glacier: 매우 저렴한 데이터 저장 서비스
- CloudFront: CDN 서비스
- RDS: MySQL 등의 관계형 DB 서비스
- DynamoDB: NoSQL 서비스
- ElastiCache: 인 메모리 캐시 서비스
- Route 53: DNS 서비스
- Auto Scaling: 트래픽에 따른 횡적 확장(EC2 추가 및 삭제)
- CloudFormation: 서버 구성 자동화
- Elastic Beanstalk: PaaS
- OpsWorks: Chef를 기반으로 하는 PaaS
- CloudSearch: 검색 서비스
- SNS: 푸시 알림 서비스
- SES: 이메일 전송 서비스
- SQS: 메시지 큐 서비스
- Elastic Transcoder: 동영상 인코딩 서비스

2. AWS 계정 만들기

console.aws.amazon.com/에 접속



로그인 또는 AWS 계정 만들기

이메일 주소(모바일 계정의 경우 전화 번호)는 무엇입니까?

이메일 또는 휴대폰 번호:

sakang@kei.re.kr

- 새 사용자입니다.
- 기존 사용자이며
비밀번호는 다음과 같습니다.

보안 서버를 사용하여 로그인

비밀번호가 생각나지 않는 경우

Amazon Aurora | 1/10의 비용으로 사용할 수 있는 엔터프라이즈급 데이터베이스

"MySQL 환경보다 10배 빠른 속도 정말 효과가 있습니다!"
-Alfresco

자세히 알아보기

AWS 계정의 보안을 강화해 주는 [AWS Identity and Access Management](#) 및 AWS 멀티 팩터 인증 기능에 대해 자세히 알아보십시오.
[AWS 프리 티어 제한 약관 전체 내용 보기](#)

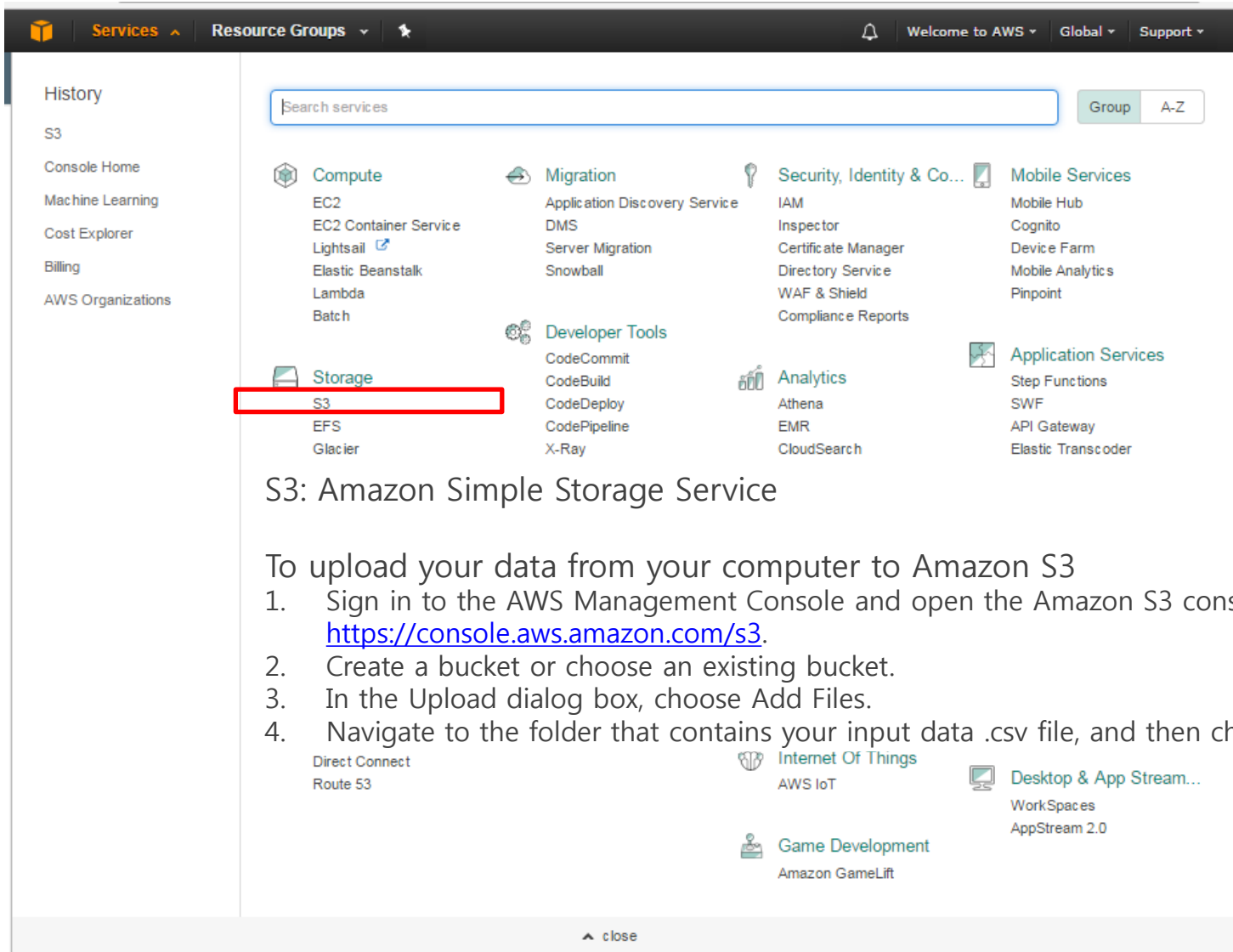
Amazon.com 로그인 정보

Amazon Web Services는 귀하의 Amazon.com 계정의 정보를 사용하여 자격 증명을 확인하고 Amazon Web Services에 대한 액세스를 허용합니다. 이 이트의 사용은 아래 링크된 이용 약관과 개인 정보 보호 정책을 따릅니다. 해당 제품 및 서비스를 AWS 부가 가치 대리점에서 구매한 경우를 제외하고 Amazon Web Services 제품 및 서비스의 사용은 아래 링크된 AWS 고객 계약에 따릅니다.

[이용 약관](#) [개인 정보 보호 정책](#) [AWS 고객 계약](#) © 1996-2017, Amazon.com, Inc. 또는 계열사

An [amazon.com](#) company

3. AWS 저장공간에 데이터 업로드



S3: Amazon Simple Storage Service

To upload your data from your computer to Amazon S3

1. Sign in to the AWS Management Console and open the Amazon S3 console at <https://console.aws.amazon.com/s3>.
2. Create a bucket or choose an existing bucket.
3. In the Upload dialog box, choose Add Files.
4. Navigate to the folder that contains your input data .csv file, and then choose Open.

3. AWS 저장공간에 데이터 업로드

* Data CSV 파일로 변환

```
1 load("D:/업무/빅데이터팀/세미나/세미나/machine learning platform/Binary.rdata")
2 write.csv(D_cat, "D:/업무/빅데이터팀/세미나/세미나/machine learning platform/Binary.csv")
3
4 load("D:/업무/빅데이터팀/세미나/세미나/machine learning platform/Linear.rdata")
5 write.csv(D_lin, "D:/업무/빅데이터팀/세미나/세미나/machine learning platform/Linear.csv")
6 |
```

```
> str(binary)
'data.frame': 999951 obs. of 6 variables:
 $ X : int 1 2 3 4 5 6 7 8 9 10 ...
 $ ystar: int 0 0 1 1 0 1 0 1 0 1 ...
 $ X1 : num 0.378 -0.853 0.748 0.933 -0.933 ...
 $ X2 : num -0.0128 -1.5188 0.8686 0.0796 -0.3775 ...
 $ X3 : num 0.3596 0.0454 -0.6697 -1.184 1.6786 ...
 $ X4 : num -1.79957 0.00192 -1.70514 0.8003 -2.11681 ...
> summary(binary)
  X          ystar          X1          X2          X3          X4
Min.   : 1   Min.   :0.0000   Min.   : -5.101359   Min.   : -5.182524   Min.   : -4.462652   Min.   : -4.689179
1st Qu.:249989 1st Qu. :0.0000   1st Qu. : -0.673561   1st Qu. : -0.672598   1st Qu. : -0.674882   1st Qu. : -0.672349
Median :500025 Median :1.0000   Median : 0.001334   Median : 0.002628   Median : 0.001086   Median : -0.000317
Mean   :500001 Mean  :0.6983   Mean   : 0.000692   Mean   : 0.001846   Mean   : 0.000153   Mean   : 0.000769
3rd Qu.:750013 3rd Qu. :1.0000   3rd Qu. : 0.675756   3rd Qu. : 0.675292   3rd Qu. : 0.674922   3rd Qu. : 0.674194
Max.   :1000000 Max.   :1.0000   Max.   : 4.611399   Max.   : 4.724537   Max.   : 5.142138   Max.   : 5.037569
```

```
> str(linear)
'data.frame': 1000000 obs. of 6 variables:
 $ X : int 1 2 3 4 5 6 7 8 9 10 ...
 $ y : num -2.042 -2.06 0.145 5.921 -7.543 ...
 $ X1 : num 0.378 -0.853 0.748 0.933 -0.933 ...
 $ X2 : num -0.0128 -1.5188 0.8686 0.0796 -0.3775 ...
 $ X3 : num 0.3596 0.0454 -0.6697 -1.184 1.6786 ...
 $ X4 : num -1.79957 0.00192 -1.70514 0.8003 -2.11681 ...
> summary(linear)
  X          y          X1          X2          X3          X4
Min.   : 1   Min.   : -16.8430   Min.   : -5.101359   Min.   : -5.182524   Min.   : -4.462652   Min.   : -4.689179
1st Qu.:250001 1st Qu. : -0.6853   1st Qu. : -0.673552   1st Qu. : -0.672573   1st Qu. : -0.674882   1st Qu. : -0.672355
Median :500001 Median : 2.0047   Median : 0.001346   Median : 0.002649   Median : 0.001071   Median : -0.000334
Mean   :500001 Mean  : 2.0068   Mean   : 0.000700   Mean   : 0.001853   Mean   : 0.000156   Mean   : 0.000762
3rd Qu.:750000 3rd Qu. : 4.6930   3rd Qu. : 0.675754   3rd Qu. : 0.675306   3rd Qu. : 0.674925   3rd Qu. : 0.674186
Max.   :1000000 Max.   :20.3576   Max.   : 4.611399   Max.   : 4.724537   Max.   : 5.142138   Max.   : 5.037569
```

Create bucket

1 Name and region 2 Set properties 3 Set permissions 4 Review

Name and region

Bucket name

Region

Copy settings from an existing bucket

You have no buckets 0 Buckets

Create Cancel **Next**

Create bucket

1 Name and region 2 Set properties 3 Set permissions 4 Review

Versioning

Enable versioning

Suspend versioning
This suspends the creation of object versions for all operations but preserves any existing object versions.

Cancel Save

Logging

Set up access log records that provide details about access requests.

Learn more

Disabled

Tags

Use tags to track your cost against projects or other criteria.

Learn more

0 Tags

Previous **Next**

Create bucket

1 Name and region 2 Set properties 3 Set permissions 4 Review

Manage users

User ID	Objects	Object permissions
sakang(Owner)	<input checked="" type="checkbox"/> Read <input checked="" type="checkbox"/> Write	<input checked="" type="checkbox"/> Read <input checked="" type="checkbox"/> Write

Manage public permissions

Group	Objects	Object permissions
Any authenticated AWS user	<input type="checkbox"/> Read <input type="checkbox"/> Write	<input type="checkbox"/> Read <input type="checkbox"/> Write
Everyone	<input type="checkbox"/> Read <input type="checkbox"/> Write	<input type="checkbox"/> Read <input type="checkbox"/> Write

Previous **Next**

Create bucket

1 Name and region 2 Set properties 3 Set permissions 4 Review

Name and region [Edit](#)

Bucket name kei2017 Region Asia Pacific (Seoul)

Properties [Edit](#)

Versioning Disabled

Logging Disabled

Tagging 0 Tags

Permissions [Edit](#)

Users 1

Public permissions Disabled

Previous **Create bucket**

Amazon S3 [Switch to the old console](#) [Discover the new console](#) [Quick tips](#)

Search for buckets

+ Create bucket Delete bucket Empty bucket 1 Buckets 1 Regions [Refresh](#)

Bucket name	Region	Date created
kei2017	Asia Pacific (Seoul)	Mar 21, 2017 10:41:55 AM

Amazon S3

kei2017

Copy Bucket ARN

Search for buckets

Bucket name kei2017

Properties

- Events: 0 Active notifications
- Versioning: Suspended
- Logging: Disabled
- Static web hosting: Disabled
- Tags: 0 Tags
- Cross-region replication: Disabled
- Transfer acceleration: Disabled

Permissions

- Owner: sakang
- Bucket policy: No
- Access control list: 1 Grantees
- CORS configuration: No

Management

- Lifecycle: Disabled
- Analytics: Disabled
- Inventory: Disabled
- Metrics: Disabled

Objects Properties Permissions Management

Upload Create folder More All Deleted objects Asia Pacific (Seoul)

This bucket is empty. Upload new objects to get started.

Upload an object

Set object properties

Set object permissions

Get started

Upload

Select files Set permissions Set properties Review

Add files

Upload Next

Upload

Select files Set permissions Set properties Review

2 Files Size: 149.8 MB Target path: kei2017

Binary.csv - 80.6 MB

Linear.csv - 69.2 MB

Upload Next

Type a prefix and press Enter to search. Press ESC to clear.

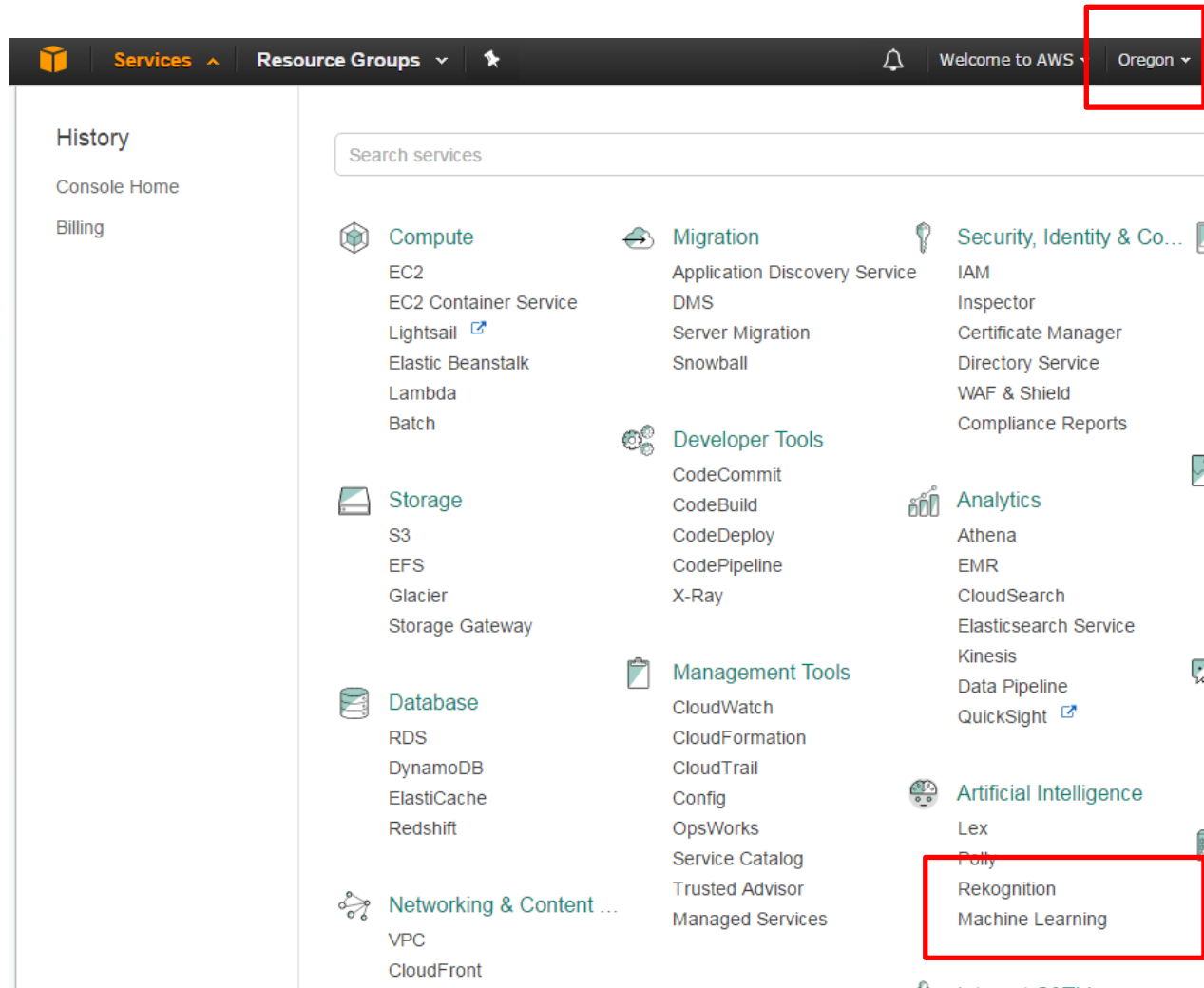
Upload Create folder More All Deleted objects Asia Pacific (Seoul)

Viewing 1 to 2

Name	Last modified	Size	Storage class
Binary.csv	Mar 21, 2017 10:49:09 AM	80.6 MB	Standard
Linear.csv	Mar 21, 2017 10:49:35 AM	69.2 MB	Standard

Viewing 1 to 2

4. AWS를 이용하여 예측모형 구축 (binary)



4. AWS를 이용하여 예측모형 구축 (binary)



Amazon Machine Learning

Amazon Machine Learning makes it easy for developers of all skill levels to use machine learning (ML) technology. Amazon Machine Learning is a managed service for building ML models and generating predictions that enable the development of robust, scalable smart applications.

[Get started](#)



Machine Learning for any Developer

Use Machine Learning technology to solve common business problems without investing in complex technology or hiring expensive consultants.

[Learn more](#)



Powerful Machine Learning Models

Take advantage of the powerful algorithms of Amazon ML to easily create machine learning models for enabling your smart applications.

[Learn more](#)



Fast, scalable predictions

Use batch predictions to create billions of predictions all at once, or create real-time predictions that you can use within interactive web, mobile, or desktop applications.

[Learn more](#)

4. AWS를 이용하여 예측모형 구축 (binary)

Get started with Amazon Machine Learning



Standard setup

Start creating your first ML model. If you don't have your data ready, you can use our sample dataset.

[Amazon Machine Learning Tutorial](#)

Launch



Dashboard

Skip straight to the Amazon Machine Learning dashboard.

View Dashboard

Cancel

4. AWS를 이용하여 예측모형 구축 (binary)

1. **Input Data** 2. Schema 3. Target 4. Row ID 5. Review

Input data

The first step to create an ML model is to show Amazon ML your historical data. This data must include the correct answers to the questions that you want the ML model to answer. Amazon ML will create a tr

Just trying out Amazon ML and don't have your data ready? Use `s3://aml-sample-data/banking.csv` This dataset contains information about customers as well as descriptions of their behavior in resp most likely to subscribe to your new product.

You can preview the file here [banking.csv](#)

Want a more guided experience? [Start with the Amazon Machine Learning Tutorial.](#)

Import your data to create an Amazon ML datasource. Amazon ML can use your datasource to create and evaluate an ML model, and you can use the datasource to review your data.

Where is your data?



Amazon Redshift

S3 data access

Tell Amazon ML how to access your data and give it permission to access it.

저장된 데이터 경로 입력 (bucket name/data file name)

S3 location *

s3:// bucket-name/file.csv

Enter the path to a single file or folder in Amazon S3. You need to grant Amazon ML permission to read this data. [Learn more.](#)

If you already have a schema for this data, provide it in a file at s3://<path-of-input-data>.schema. If you don't have a schema, Amazon ML will help you create one on the next page.

Datasource name

* Required

Reset

Cancel

Verify

4. AWS를 이용하여 예측모형 구축 (binary)

1. Input Data 2. Schema 3. Target 4. Row ID 5. Review

Input data

Import your data to create an Amazon ML datasource. Amazon ML can use your datasource to create and evaluate an ML model. Use the datasource to review your data.

Where is your data?



S3 data access

Tell Amazon ML how to access your data and give it permission to access it.

S3 location

s3://kei2017/Binary.csv

Enter the path to a single file or folder in Amazon S3. You need to grant Amazon ML permission to read this data. [Learn more](#).

If you already have a schema for this data, provide it in a file at s3://<path-of-input-data>/schema. If you don't have a schema, Amazon ML will help you create one on the next page. [?](#)

Datasource name

test_binary

The validation is successful. To go to the next step, choose Continue

Datasource name test_binary
Data location s3://kei2017/Binary.csv
Data format CSV
Schema source Auto generated
Number of files 1
Total size 80.6 MB

* Required

Reset

Cancel

Continue

1. Input Data 2. Schema 3. Target 4. Row ID 5. Review

Schema



Amazon ML scanned your input data and inferred the column names and data type for each of the columns in your dataset. Review and edit the data type for each column to ensure that it accurately represents the data. This enables Amazon ML to read the input data correctly and to produce accurate predictions. [Learn more](#)

Does the first line in your CSV contain the column names? Yes No [?](#)

ACTION: Change type

Search by attribute name

Items per page: 10 << 1 - 6 of 6 >>

	Name	Data type	Sample field value 1	Sample field value 2	Sample field value 3
<input type="checkbox"/>	1	Numeric	2	3	4
<input type="checkbox"/>	2 ystar	Binary	0	1	1
<input type="checkbox"/>	3 X1	Numeric	-0.853244215744366	0.747536639162936	0.93316212196265
<input type="checkbox"/>	4 X2	Numeric	-1.51883107107082	0.868591799785687	0.0795652078494213
<input type="checkbox"/>	5 X3	Numeric	0.0453656670585945	-0.66966044912394	-1.18397776413396
<input type="checkbox"/>	6 X4	Numeric	0.00192322758431187	-1.70513672376946	0.800300288796818

<< 1 - 6 of 6 >>

Cancel

Previous

Continue

4. AWS를 이용하여 예측모형 구축 (binary)

1. Input Data 2. Schema **3. Target** 4. Row ID 5. Review

Target ?

Machine learning works by finding patterns that connect your data to the value to be predicted. To create an ML model, Amazon ML analyzes examples of data records with correct values. The column that contains these values in the training dataset is called the target.

Do you plan to use this dataset to create or evaluate an ML model? Yes No

Select the row containing the value you want to predict.

You have selected a binary attribute named *ystar* as the target. ML models trained on this target use logistic regression to train a binary classification model.

Search by attribute name

Target	Name	Data type	Sample field value 1	Sample field value 2	Sample field value 3
<input type="radio"/>		Numeric	2	3	4
<input type="radio"/>	X1	Numeric	-0.853244215744366	0.747536639162936	0.93316212196265
<input type="radio"/>	X2	Numeric	-1.51883107107082	0.868591799785687	0.0795652078494213
<input type="radio"/>	X3	Numeric	0.0453656670585945	-0.66966044912394	-1.18397776413396
<input type="radio"/>	X4	Numeric	0.00132322758451167	1.70513672370546	0.000306266758676
<input checked="" type="radio"/>	ystar	Binary	0	1	1

Cancel

Previous

Continue

1. Input Data 2. Schema 3. Target **4. Row ID** 5. Review

Row identifier (optional) ?

An optional row identifier helps you understand how prediction rows correspond to observation rows from the input data. If you choose to make an attribute the row identifier, Amazon ML will add that column to the prediction output. A row identifier is intended for reference purposes only. Amazon ML does not include the row identifier when training ML models.

Does your data contain an identifier? Yes No

Cancel

Previous

Review

4. AWS를 이용하여 예측모형 구축 (binary)

1. Input Data 2. Schema 3. Target 4. Row ID **5. Review**

Review

Review and make any changes, and then click Finish.

Input data

Edit

Datasource name test_binary
S3 location s3://kei2017/Binary.csv
Data format CSV
Number of files 1
Total size 80.6 MB

Schema

Edit

Schema source Auto generated (Column names are taken from the first row of the CSV file)
Data types 5 Numeric Attributes
1 Binary Attribute

Target

Edit

Target ystar (Binary classification)

Row identifier (optional)

Edit

Record ID None

Tags ⓘ

Amazon ML copies a maximum of 10 tags from parent objects. Edit the list to keep the tags you need.

No tags

Cancel

Previous

Create datasource

4. AWS를 이용하여 예측모형 구축 (binary)

Objects ?

Create new... Actions Refresh

Filter: All types Items per page: 10 << < 1 - 1 of 1 Objects > >>

	Name	Type	ID	Status	Creation time	Completion time
<input type="checkbox"/>	test_binary	Datasource	ds-xN5itReX673	Completed	Mar 21, 2017 11:03:02 AM	5 mins.

<< < 1 - 1 of 1 Objects > >>

Machine Learning Concepts

4. AWS를 이용하여 예측모형 구축 (binary)

Amazon Machine Learning ▾ Datasources ▸ ds-xN5itReX673

Data insights

- Data summary
 - Target distributions
 - Missing values

Attributes

- Binary
- Categorical
- Numeric
- Text

Data processing completed. We found the following potential issues with your input data.

Review your data before you train your ML model. The quality of your data affects the quality of your ML model.

1 of 6 Attributes have missing values

Datasource information ? Delete this Datasource

ID	ds-xN5itReX673
Name	test_binary ✎
Creation time	Mar 21, 2017 11:03:02 AM
Completion time	5 mins. ?
Compute Time (Approximate)	16 mins. ?
Status	Completed
Message	Not available
Input schema	View input schema
Log	Download log

Use this datasource to ▾

Target information

Target name	Target
Target type	Binary
Target visualization	

Input data

Binary attributes: _Target_ ?

Category	Count
0	301,729
1	698,222
Invalid values	1

4. AWS를 이용하여 예측모형 구축 (binary)

Data insights

Data summary

Target distributions

Missing values

Attributes

Binary

Categorical

Numeric

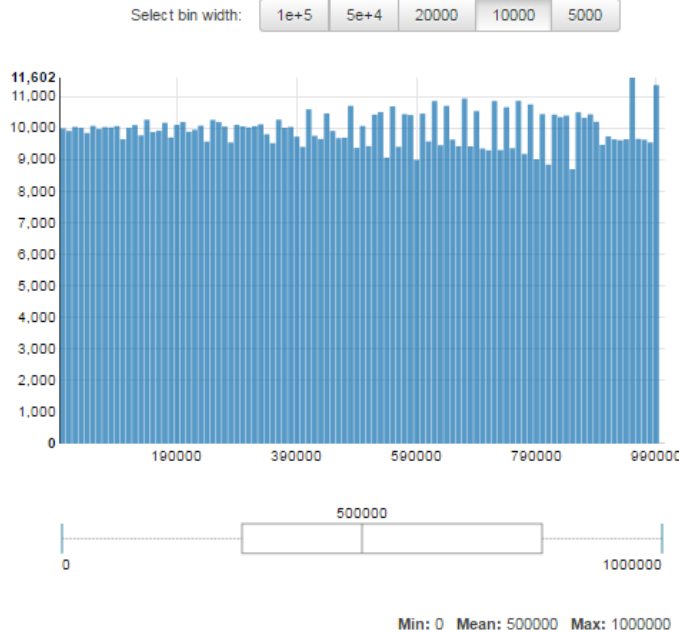
Text

Global attribute name Search

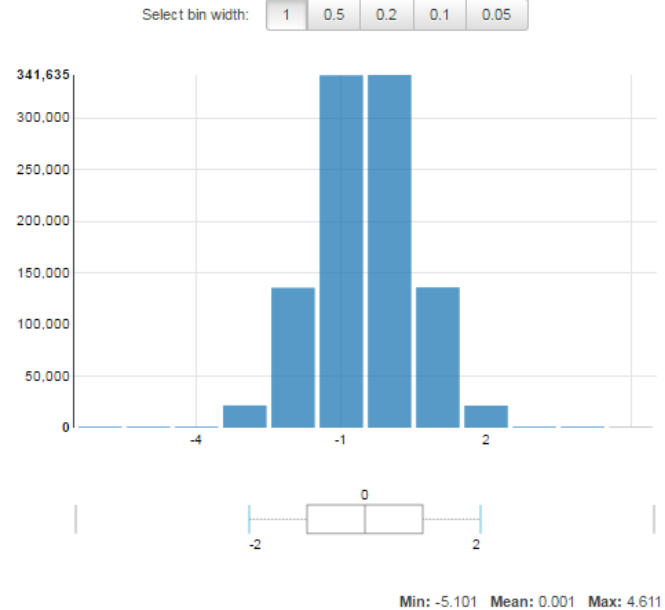
Numeric attributes

Attributes	Correlations to target	Missing values	Invalid values	Range	Mean	Median	Preview
Var1	0.00066	1 (1%)	0 (0%)	1 - 1000000	500001.19909275556	500208	
Var2	0.01522	0 (0%)	1 (1%)	-5.10135946519737 - 4.61139928832948	0.0006921853518343294	0.00273827425286016	
Var3	0.06531	0 (0%)	1 (1%)	-5.18252377904155 - 4.7245368350171	0.001845639192874637	0.00322693139741236	
Var4	0.01433	0 (0%)	1 (1%)	-4.46265163087911 - 5.14213774621773	0.0001528381517160616	0.00211395710309124	
Var5	0.17017	0 (0%)	1 (1%)	-4.68917907660572 - 5.03756935364921	0.0007693948385802102	0.000750772916286561	

Numeric attributes: Var1



Numeric attributes: Var2



4. AWS를 이용하여 예측모형 구축 (binary)

Objects ?

Create new... ▾ Actions ▾ Refresh ↻

Datasource and ML model
Datasource
ML model
Evaluation
Batch prediction

1. **Input data** 2. ML model settings 3. Recipe 4. Advanced settings 5. Evaluation 6. Review

Input data

Locate the data you want to use to train (create) an ML model. Later, you can use this ML model to generate predictions.

Locate the input data I already created a datasource pointing to my S3 data My data is in S3, and I need to create a datasource

Q Enter the datasource name or ID

Datasource name test_binary Change datasource

Datasource ID	ds-xN5itReX673	Input schema	View input schema
Creation time	Mar 21, 2017 11:03:02 AM	Target attribute	__Target__
Status	Completed	Target type	BINARY
Datasource type	S3	Number of attributes	6
S3 location	s3://kei2017/Binary.csv	Models trained	0
Data format	CSV	Evaluations created	0
Data rearrangement	None	Batch predictions created	0

Tags

No tags

✔ You selected datasource ds-xN5itReX673. The validation is successful. To go to the next step, choose Continue

Cancel **Continue**

4. AWS를 이용하여 예측모형 구축 (binary)

1. Input data **2. ML model settings** 3. Recipe 4. Advanced settings 5. Evaluation 6. Review

ML model settings

You can use the automatically suggested ML model settings, or you can choose to customize.

ML model type BINARY ⓘ

ML model target _Target_

ML model name (Optional)

Select training and evaluation settings Recipes and training parameters control the ML model training process. You can select these settings for your ML model or use the defaults provided by Amazon ML. In either case, you can choose to have Amazon ML reserve a portion of the input data for evaluation. [Learn more.](#)

- Default (Recommended)**
 - Generate a default recipe
 - Use default training parameters
 - Set aside 30% of your training data to evaluate the training
 - Split the evaluation data sequentially ⓘ
- Custom**
 - Modify the recipe Amazon ML generates
 - Modify training parameters
 - Randomly or sequentially split your evaluation data ⓘ

Evaluation Name

[Cancel](#) [Previous](#) [Review](#)

1. Input data 2. ML model settings 3. Recipe 4. Advanced settings 5. Evaluation **6. Review**

Review

Review and make any changes, and then click Finish.

Input data [Edit](#)

Datasource ID	ds-xN5fReX673
Datasource name	test_binary
Creation time	Mar 21, 2017 11:03:02 AM
Status	Completed

ML model settings [Edit](#)

ML model Name	ML model: test_binary
ML model parameters	Default (includes the ML model evaluation) - See Advanced settings below.
Evaluation name	Evaluation: ML model: test_binary
Evaluation data	Amazon ML will split your training datasource into 70% for training and will reserve the remaining 30% for evaluation.

Recipe

Recipe

Advanced settings

Maximum ML model Size	100MB
Maximum number of data pa...	10
Shuffle type for training data	Auto
Regularization type	L2
Regularization amount	1e-6 - Mild

Tags ⓘ

Amazon ML copies a maximum of 10 tags from parent objects. Edit the list to keep the tags you need.
No tags

[Cancel](#) [Previous](#) [Create ML model](#)

4. AWS를 이용하여 예측모형 구축 (binary) -성능확인

Objects ?

Create new... Actions Refresh ↻

Filter: All types Items per page: 10 « < 1 - 5 of 5 Objects > »

	Name	Type	ID	Status	Creation time	Completion time																												
<input type="checkbox"/>	▶ Evaluation: ML model: test_binary	Evaluation	ev-0T4zSfxUfOY	Completed	Mar 21, 2017 11:28:28 AM	4 mins.																												
<input type="checkbox"/>	▼ ML model: test_binary	ML model	ml-UvVT6xcJERD	Completed	Mar 21, 2017 11:28:27 AM	9 mins.																												
<div style="border: 2px solid red; padding: 10px;"><p>ML model name ML model: test_binary</p><hr/><table><tbody><tr><td>ML model ID</td><td>ml-UvVT6xcJERD</td><td>Input schema</td><td>View input schema</td></tr><tr><td>ML model type</td><td>Binary classification</td><td>Target attribute</td><td>_Target_</td></tr><tr><td>Creation time</td><td>Mar 21, 2017 11:28:27 AM</td><td>Target type</td><td>BINARY</td></tr><tr><td>Status</td><td>Completed</td><td>Number of attributes</td><td>6</td></tr><tr><td>Datasource ID</td><td>ds-0KCmnKOaHnC</td><td>Evaluations created</td><td>1</td></tr><tr><td>Log</td><td>Download log</td><td>Latest evaluation result</td><td>0.994 (AUC)</td></tr><tr><td></td><td></td><td>Batch predictions created</td><td>0</td></tr></tbody></table><p>Tags Add or edit tags</p><p>No tags</p></div>							ML model ID	ml-UvVT6xcJERD	Input schema	View input schema	ML model type	Binary classification	Target attribute	_Target_	Creation time	Mar 21, 2017 11:28:27 AM	Target type	BINARY	Status	Completed	Number of attributes	6	Datasource ID	ds-0KCmnKOaHnC	Evaluations created	1	Log	Download log	Latest evaluation result	0.994 (AUC)			Batch predictions created	0
ML model ID	ml-UvVT6xcJERD	Input schema	View input schema																															
ML model type	Binary classification	Target attribute	_Target_																															
Creation time	Mar 21, 2017 11:28:27 AM	Target type	BINARY																															
Status	Completed	Number of attributes	6																															
Datasource ID	ds-0KCmnKOaHnC	Evaluations created	1																															
Log	Download log	Latest evaluation result	0.994 (AUC)																															
		Batch predictions created	0																															
<input type="checkbox"/>	▶ test_binary_[percentBegin=70, percentEnd=100,...	Datasource	ds-c0SvARQIYin	Completed	Mar 21, 2017 11:28:25 AM	4 mins.																												
<input type="checkbox"/>	▶ test_binary_[percentBegin=0, percentEnd=70, st...	Datasource	ds-0KCmnKOaHnC	Completed	Mar 21, 2017 11:28:25 AM	4 mins.																												
<input type="checkbox"/>	▶ test_binary	Datasource	ds-xN5itReX673	Completed	Mar 21, 2017 11:03:02 AM	5 mins.																												

« < 1 - 5 of 5 Objects > »

▼ Machine Learning Concepts

4. AWS를 이용하여 예측모형 구축 (binary) - 성능확인

ML model report

- Summary
- Settings
- Monitoring

Tools

- Try real-time predictions

Evaluations

- ▶ Evaluation: ML mod...

ML model summary Delete this ML model

ID	mL-UvVT6xcJERD
Name	ML model: test_binary
Type	Binary classification
Creation time	Mar 21, 2017 11:28:27 AM
Completion time	9 mins.
Compute Time (Approximate)	7 mins.
Status	Completed
Log	Download log

Datasource (training)

Datasource ID	ds-0KCmnKOaHnC
Target	<code>_Target_</code>
Input schema	View input schema

Evaluations

Evaluations created: 1

Latest evaluation result	0.994 (AUC)
--------------------------	-------------

[Perform another Evaluation](#)

Predictions

CloudWatch metrics [View in CloudWatch](#)

Score threshold: 0.5

A single dataset
Generate one-time predictions for a single dataset.

[Generate batch predictions](#)

Try real-time predictions
Generate real-time predictions in your browser.

[Try real-time predictions](#)

Enable real-time predictions
To enable real-time predictions now, create a real-time prediction endpoint.

4. AWS를 이용하여 예측모형 구축 (binary) -성능확인

ML model report

- Summary
- Settings
- Monitoring

Tools

- Try real-time predictions

Evaluations

- ▼ Evaluation: ML mod...
- Summary
- Alerts (0)
- Explore performance

Evaluation Summary Delete this Evaluation

ID	ev-Ot4zSfxUfOY
Name	Evaluation: ML model: test_binary
Datasource ID	ds-cOSVARQIYin
Output location	Not available
Creation time	Mar 21, 2017 11:28:28 AM
Completion time	4 mins.
Compute Time (Approximate)	2 mins.
Status	Completed
Log	Download log

ML model performance metric

On your most recent evaluation, **ev-Ot4zSfxUfOY**, the ML model's quality score is considered **extremely good** for most machine learning applications.

AUC: 0.994
Baseline AUC: 0.500
Difference: 0.494

Next step: If you want to use this ML model to generate predictions, explore trade-offs to optimize the performance of your ML model first.

Score threshold: 0.5 Adjust score threshold Explore performance

Tags [Add or edit tags](#)

No tags

4. AWS를 이용하여 예측모형 구축 (binary) - 성능확인

ML model report

Summary

Settings

Monitoring

Tools

Try real-time predictions

Evaluations

▼ Evaluation: ML mod...

Summary

Alerts

Explore performance

Evaluations

Create new evaluation

Actions ▾

Refresh ↻

🔍 Evaluation name or ID

Items per page: 10 - << < 1 - 1 of 1 Evaluations > >>

	Name	ID	Status	Creation time	Completion time	ML model ID
<input type="checkbox"/>	▶ Evaluation: ML model: test_binary	ev-Ot4zSfx...	Completed	Mar 21, 2017 11:28:28 AM	4 mins.	ml-UvVT6xcJE...

<< < 1 - 1 of 1 Evaluations > >>

4. AWS를 이용하여 예측모형 구축 (binary) - 성능확인

ML model report

Summary

Settings

Monitoring

Tools

Try real-time predictions

Evaluations

▼ Evaluation: ML mod...

Summary

Alerts (0)

Explore performance

Criteria for a valid evaluation

✔ An evaluation of an ML model is done on held-out data ⓘ

Evaluation data location s3://kei2017/Binary.csv (Split: 70 - 100%)

Training data location s3://kei2017/Binary.csv (Split: 0 - 70%)

✔ Sufficient data was used for the evaluation of the ML model ⓘ

Number of records for evaluation data 299587

Number of records for training data 700365

✔ Schema matched ⓘ

Number of attributes for evaluation data 6

Number of attributes for training data 6

✔ Distribution of target attribute ⓘ

Target distribution of evaluation datasource



Target distribution of training datasource



4. AWS를 이용하여 예측모형 구축 (binary) - 성능확인

ML model report

Summary

Settings

Monitoring

Tools

Try real-time predictions

Evaluations

▼ Evaluation: ML mod...

[Summary](#)

Alerts (0)

[Explore performance](#)

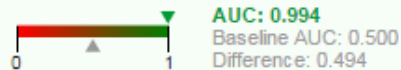
Evaluation Summary

Delete this Evaluation

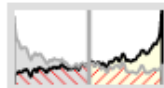
ID	ev-Ot4zSfxUfOY
Name	Evaluation: ML model: test_binary
Datasource ID	ds-cOSVARQIYin
Output location	Not available
Creation time	Mar 21, 2017 11:28:28 AM
Completion time	4 mins.
Compute Time (Approximate)	2 mins.
Status	Completed
Log	Download log

ML model performance metric

On your most recent evaluation, **ev-Ot4zSfxUfOY**, the ML model's quality score is considered **extremely good** for most machine learning applications.



Next step: If you want to use this ML model to generate predictions, explore trade-offs to optimize the performance of your ML model first.



Score threshold: 0.5

[Adjust score threshold](#)

[Explore performance](#)

4. AWS를 이용하여 예측모형 구축 (binary) -성능확인

ML model report

- Summary
- Settings
- Monitoring

Tools

- Try real-time predictions

Evaluations

▼ Evaluation: ML mod...

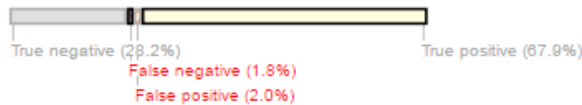
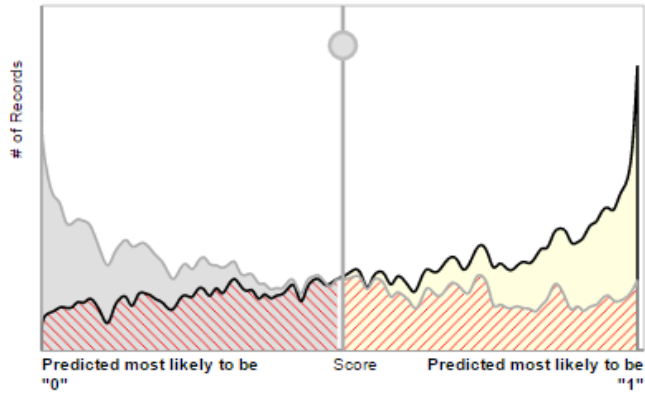
- Summary
- Alerts (0)
- [Explore performance](#)

ML model performance

This chart shows the distributions of your predicted answers for the actual "1" and "0" records in your evaluation data. Any overlap of the actual "1" and "0" is where your ML model guesses wrong. [Learn more.](#)

Adjust the slider to indicate how much error you can tolerate from your ML model based on your needs. Moving the score threshold to the right decreases the number of false positives and increases the number of false negatives.

[Explain this chart](#)



Trade-off based on score threshold

[Reset score threshold \(0.5\)](#)

- 96% are correct
203,525 true positive
84,612 true negative
- 4% are errors
6,018 false positive
5,432 false negative

- 70% of the records are predicted as "1"
- 30% of the records are predicted as "0"

[Save score threshold at 0.50](#)

▼ Advanced metrics

- | | | | | |
|---------------------|--------|---|-----------------------|---|
| False positive rate | 0.0664 | 0 | <input type="range"/> | 1 |
| Precision | 0.9713 | 0 | <input type="range"/> | 1 |
| Recall | 0.974 | 0 | <input type="range"/> | 1 |
| Accuracy | 0.9618 | 0 | <input type="range"/> | 1 |

4. AWS를 이용하여 예측모형 구축 (linear)

The screenshot shows the AWS ML console interface for creating a new ML model. The breadcrumb navigation indicates the path: Amazon Machine Learning > ML models > Create ML model. The progress bar shows six steps: 1. Input data, 2. ML model settings (current step), 3. Recipe, 4. Advanced settings, 5. Evaluation, and 6. Review.

ML model settings

You can use the automatically suggested ML model settings, or you can choose to customize.

ML model type REGRESSION ⓘ

ML model target y

ML model name (Optional) ML model: test_linear

Select training and evaluation settings

Recipes and training parameters control the ML model training process. You can select these settings for your ML model or use the defaults provided by Amazon ML. In either case, you can choose to have Amazon ML reserve a portion of the input data for evaluation. [Learn more.](#)

- Default (Recommended)**
 - Generate a default recipe
 - Use default training parameters
 - Set aside 30% of your training data to evaluate the training
 - Split the evaluation data sequentially ⓘ
- Custom**
 - Modify the recipe Amazon ML generates
 - Modify training parameters
 - Randomly or sequentially split your evaluation data ⓘ

[Cancel](#) [Previous](#) [Continue](#)

4. AWS를 이용하여 예측모형 구축 (linear)

Services ▾ Resource Groups ▾

Welcome to AWS ▾ N. Virginia ▾ Support ▾

Amazon Machine Learning ▾ ML models > Create ML model

1. Input data 2. ML model settings **3. Recipe** 4. Advanced settings 5. Evaluation 6. Review

Recipe ?

Amazon ML helps you transform your data to optimize the ML model learning process. You can use data recipes to easily transform some or all attributes in your data. The following recipe is automatically suggested based on your data. Edit it inline, or continue to the next step. [Learn more.](#)

Your attributes:

Name	Data type	Sample field value	Sample value
X4	Numeric	0.0019232275843...	-1.
X3	Numeric	0.0453656670585...	-0.
X2	Numeric	-1.518831071070...	0.8685917997856... 0.0795652078494...
X1	Numeric	-0.853244215744...	0.7475366391629... 0.93316212196265
y	Numeric	-2.060367933925...	0.1450930405497... 5.92094667943193

Items per page: 1 to 5 of 5

Recipe (default):
Amazon ML Suggested Recipe

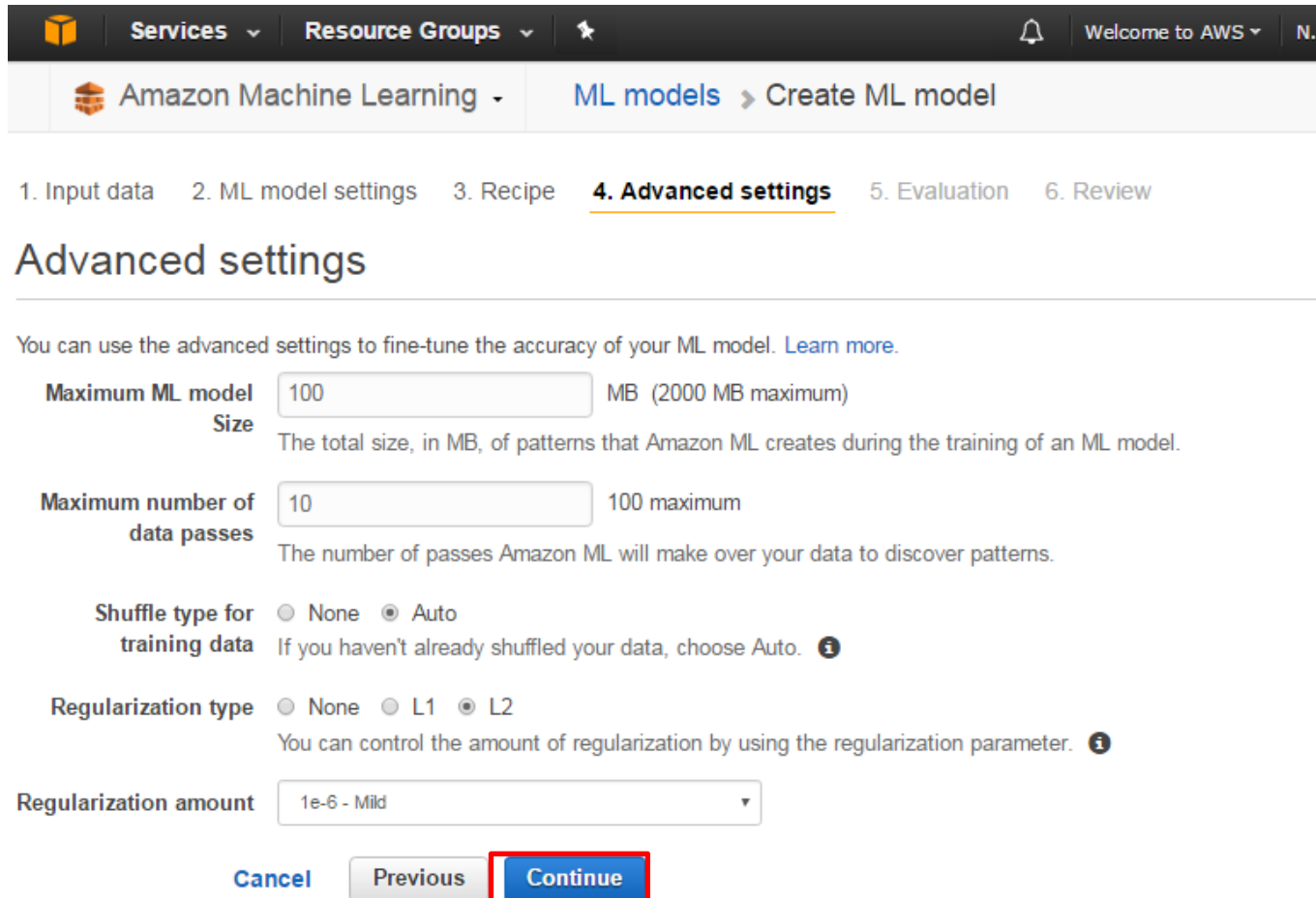
Recipe

Recipes help Amazon Machine Learning find patterns in your data. If you did not provide a recipe, Amazon ML will generate one for you. [Learn more.](#)

Recipe is valid

[Cancel](#) [Previous](#) [Continue](#)

4. AWS를 이용하여 예측모형 구축 (linear)



The screenshot shows the AWS Management Console interface for creating an ML model. The breadcrumb navigation is: Amazon Machine Learning > ML models > Create ML model. The progress bar indicates the current step is '4. Advanced settings', with previous steps being '1. Input data', '2. ML model settings', '3. Recipe', and subsequent steps being '5. Evaluation' and '6. Review'. The 'Advanced settings' section includes several configuration options:

- Maximum ML model Size:** A text input field containing '100' MB (2000 MB maximum). Description: 'The total size, in MB, of patterns that Amazon ML creates during the training of an ML model.'
- Maximum number of data passes:** A text input field containing '10' (100 maximum). Description: 'The number of passes Amazon ML will make over your data to discover patterns.'
- Shuffle type for training data:** Radio buttons for 'None' and 'Auto' (selected). Description: 'If you haven't already shuffled your data, choose Auto. ⓘ'
- Regularization type:** Radio buttons for 'None', 'L1', and 'L2' (selected). Description: 'You can control the amount of regularization by using the regularization parameter. ⓘ'
- Regularization amount:** A dropdown menu currently showing '1e-6 - Mild'.

At the bottom of the settings, there are three buttons: 'Cancel', 'Previous', and 'Continue'. The 'Continue' button is highlighted with a red rectangular border.

4. AWS를 이용하여 예측모형 구축 (linear)

Services Resource Groups

Welcome to AWS N. Virginia Support

Amazon Machine Learning ML models Create ML model

1. Input data 2. ML model settings 3. Recipe 4. Advanced settings **5. Evaluation** 6. Review

Evaluation

An evaluation assesses the predictive performance of an ML model and can be done at any time. An evaluation must be done on records that were not used in creating the ML model. We recommend that you evaluate new ML models.

Do you want to evaluate this ML model now? Yes No

Evaluation Name (Optional) Evaluation: ML model: test_linear

Select evaluation data: Split the datasource to reserve a portion for evaluation.
By default, Amazon ML splits your datasource, using 70% for training and the remaining 30% for evaluation. Tell Amazon ML whether you want sequential or random splitting. [Learn more.](#)

- Sequential: Amazon ML selects the first 70% of the datasource for training and the remaining 30% for evaluation.
- Random: Amazon ML randomly selects 70% of the datasource for training and 30% for evaluation.
- Use a different datasource for evaluation
If you don't want Amazon ML to reserve a portion of the training datasource for evaluation, provide Amazon ML with a different datasource.

Cancel Previous **Review**

4. AWS를 이용하여 예측모형 구축 (linear)

The screenshot shows the 'Review' step of the 'Create ML model' process in the AWS ML console. The breadcrumb trail is 'Amazon Machine Learning > ML models > Create ML model'. The progress bar indicates the current step is '6. Review'.

Review

Review and make any changes, and then click Finish.

Input data Edit

Datasource ID	ds-F457wN9Y1zm
Datasource name	test_linear
Creation time	Mar 27, 2017 10:45:12 AM
Status	Completed

ML model settings Edit

ML model Name	ML model: test_linear
ML model parameters	Custom - See Advanced settings below.

Recipe Edit

Recipe

Recipes help Amazon Machine Learning find patterns in your data. If you did not provide a recipe, Amazon ML will generate one for you. [Learn more.](#)

Advanced settings Edit

Maximum ML model Size	100MB
Maximum number of data pa...	10
Shuffle type for training data	Auto
Regularization type	L2
Regularization amount	1e-6 - Mild

Evaluation Edit

Evaluate ML model	Yes
Evaluation name	Evaluation: ML model: test_linear
Evaluation data	Random: Amazon ML randomly selects 70% of the datasource for training and 30% for evaluation.

Tags +

Amazon ML copies a maximum of 10 tags from parent objects. Edit the list to keep the tags you need.

No tags

docs.aws.amazon.com/console/.../Data_Transformations

Cancel Previous Create ML model

4. AWS를 이용하여 예측모형 구축 (linear)

Services ▾ Resource Groups ▾ ✎

Welcome to AWS ▾ N. Virginia ▾ Support ▾

Amazon Machine Learning ▾ ML models ▶ ml-o3ymtzIOSTs

ML model report

- Summary
- Settings
- Monitoring
- Tools
- Try real-time predictions
- Evaluations
 - ▶ Evaluation: ML mod...

ML model summary Delete this ML model

ID ml-o3ymtzIOSTs
Name ML model: test_11 ✎
Type Numerical regression
Creation time Mar 27, 2017 11:04:19 AM
Completion time Not available ⓘ
Compute Time (Approximate) Not available ⓘ
Status Pending
Log Not available

Datasource (training)

Datasource ID ds-CX1EMMz0V8x

Objects

Create new... ▾ Actions ▾ Refresh ↻

Filter: All types ▾ Items per page: 10 ▾ « < 1 - 5 of 5 Objects > »

	Name	Type	ID	Status	Creation time	Completion time
<input type="checkbox"/>	▶ Evaluation: ML model: test_linear	Evaluation	ev-ncVMIt2e3DP	Pending	Mar 27, 2017 11:12:54 AM	Not available
<input type="checkbox"/>	▶ ML model: test_linear	ML model	ml-eXlzpR92EBi	Pending	Mar 27, 2017 11:12:53 AM	Not available
<input type="checkbox"/>	▶ test_linear_[percentBegin=0, percentEnd=70, str...	Datasource	ds-2VW8EuexYvE	In progress	Mar 27, 2017 11:12:52 AM	Not available
<input type="checkbox"/>	▶ test_linear_[percentBegin=70, percentEnd=100, ...	Datasource	ds-EQJzZs96cxv	In progress	Mar 27, 2017 11:12:52 AM	Not available
<input type="checkbox"/>	▶ test_linear	Datasource	ds-pTjopEnbbcL	In progress	Mar 27, 2017 11:11:01 AM	Not available

« < 1 - 5 of 5 Objects > »

4. AWS를 이용하여 예측모형 구축 (linear)

Services Resource Groups Welcome to AWS N. Virginia Support

Amazon Machine Learning ML models ml-eXlzpR92EBi

ML model report ML model summary Delete this ML model

Summary
Settings
Monitoring
Tools
Try real-time predictions
Evaluations

▼ Evaluation: ML model
Summary
Alerts (0)
Explore performance

ML model summary

ID ml-eXlzpR92EBi
Name ML model: test_linear
Type Numerical regression
Creation time Mar 27, 2017 11:12:53 AM
Completion time 5 mins.
Compute Time (Approximate) 4 mins.
Status Completed
Log Download log

Datasource (training)

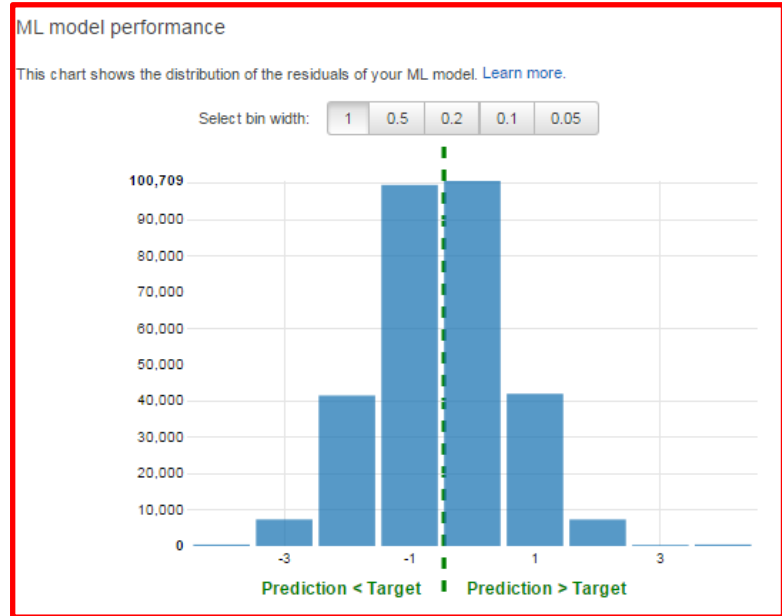
Datasource ID ds-2VW8EuexYvE
Target y
Input schema View input schema

Evaluations

Evaluations created 1
Latest evaluation result 1.032 (RMSE)
Perform another Evaluation

Predictions

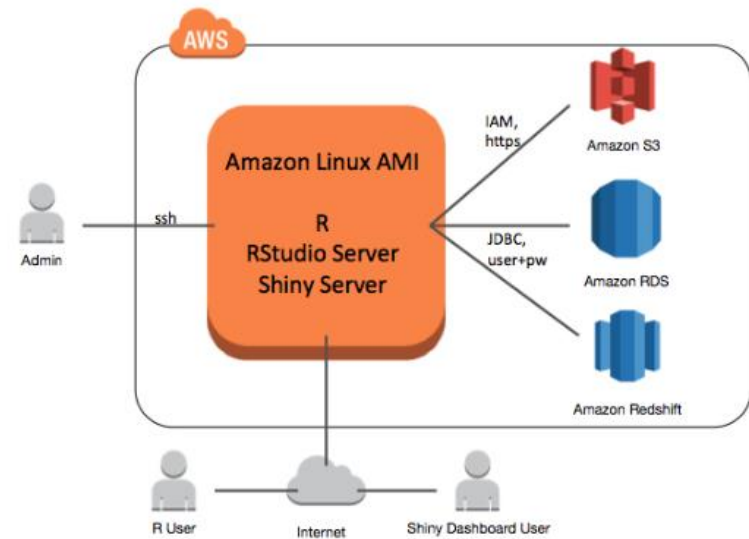
CloudWatch metrics View in CloudWatch
A single dataset
Generate one-time predictions for a single dataset.
Generate batch predictions
Try real-time predictions
Generate real-time predictions in your browser.
Try real-time predictions
Enable real-time predictions
To enable real-time predictions now, create a real-time endpoint.
Create endpoint



5. AWS 장점

AWS Machine Learning	
data sources	text file uploaded into S3 AWS RDS AWS Redshift AWS S3 table
data formats	csv file S3 or Redshift database
dataset maximum size	100 GB
data types	boolean categorical numeric string

- 확장가능성
Python - boto 3을이용하여 AWS 활용 가능
AWS EC2에 R server 설치



- 서버 사용의 용이함
- S3(Simple Storage Service) -클라우드 스토리지 보유
- 시각화의 우수함
- 서울에 리전이 생겨서 빠른 응답속도를 보여줌(응답속도 0.08초)

5. AWS 장점

	AWS	Azure	GCE	OpenStack
Virtual network	Yes	Yes	No	Yes
API	Yes	Yes	Yes	No
Machine-learning Tools	Yes	Yes	Yes	No
Deployment Tools	Yes	Yes	Yes	No
On-premises Data-center integration	Yes	Yes	Yes	No
Number of Service	Most	Many	Enough	Few
Global Region	12 (Gov/베이징 포함)	13	3	N/A
Compliance	Common standards, IT Grundschutz(Germany), G-Cloud(UK)	Common standards, ISO27018, G-Cloud(UK)	Common Standards	Depends on the OpenStack provider
SDK	Android, Javascript(node.js), iOS, Java, .NET, PHP, Python, Ruby, Go	Android, iOS, Java, .NET, Javascript(node.js), PHP, Python, Ruby	Java, Javascript(node.js), .NET, PHP, Python	N/A
Integration into development process	Medium, not linked to specific ecosystems	High, lined to the Microsoft ecosystem(for example, .NET development)	High, linked to the Google ecosystem(for example, Android)	N/A
Relational database	Yes(MySQL, PostgreSQL, Oracle Ddatabase, MS-SQL, Aurora)	Yes(Azure SQ Database, MS-SQL)	Yes(MySQL)	Depends on the OpenStack provider
Block-level storage(attached via network)	Yes	Yes	No	Yes
NoSQL database	Yes	Yes	Yes	No
DNS	Yes	No	Yes	No

출처: Amazon Web Service in Action (2016)

5. AWS 단점

- 요금부과 방법의 복잡함: 데이터 업로드, 저장, 분석 등 비용이 모두 각각 청구되어 요금 책정 방식이 복잡함

- **Recipe**

Recipes help Amazon Machine Learning find patterns in your data. If you did not provide a recipe, Amazon ML will generate one for you.
[Learn more.](#)

Thank you
