AWS(Amazon Web Service)

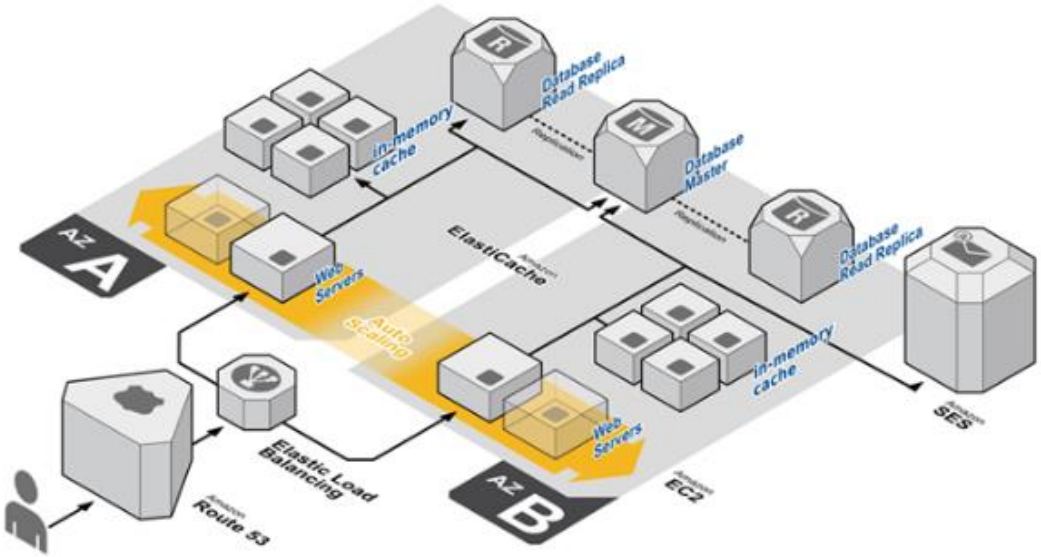
빅데이터팀 강선아
2017. 3. 28

목차

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



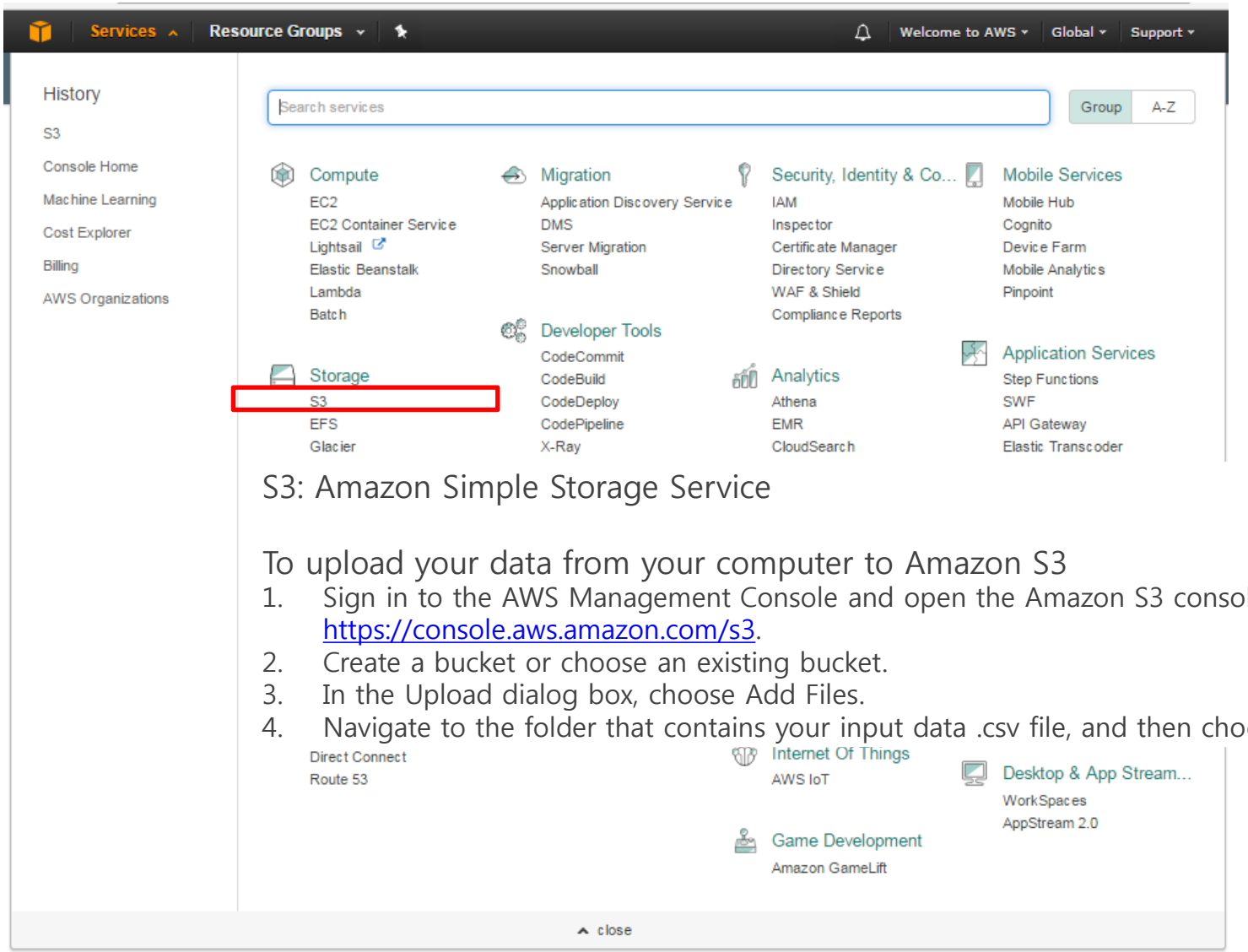
- EC2: 가상 서버
- CloudWatch: 리소스 모니터링 서비스
- S3: 스토리지 서비스
- Glacier: 매우 저렴한 데이터 저장 서비스
- CloudFront: CDN 서비스
- RDS: MySQL 등의 관계형 DB 서비스
- DynamoDB: NoSQL 서비스
- ElastiCache: 인 메모리 캐시 서비스
- Route 53: DNS 서비스

- 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

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.

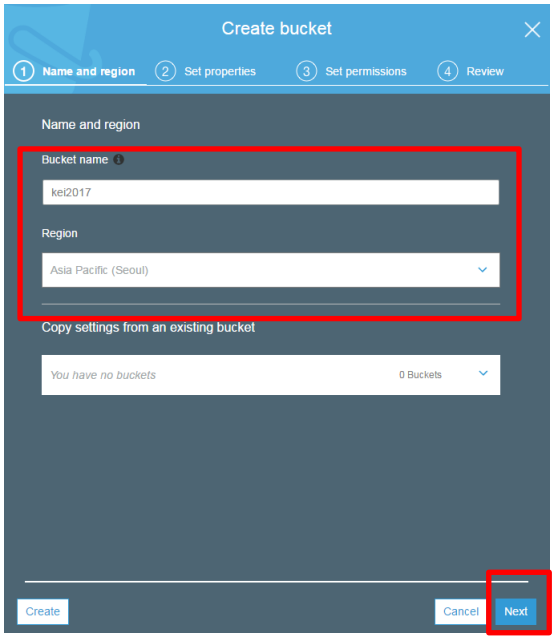
3. AWS 저장공간에 데이터 업로드

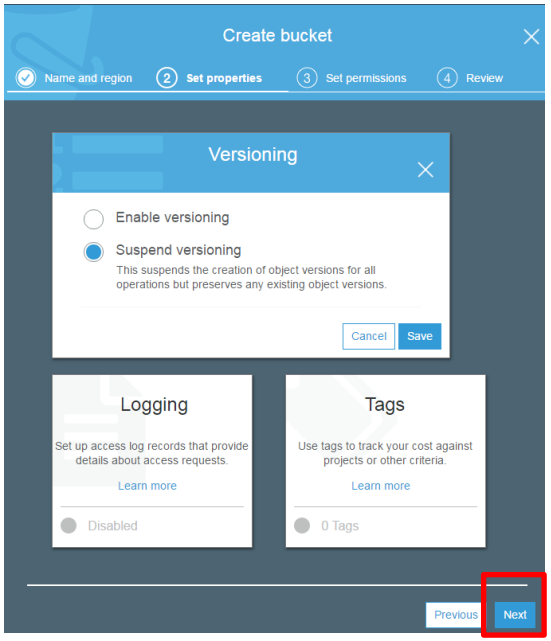
* Data CSV 파일로 변환

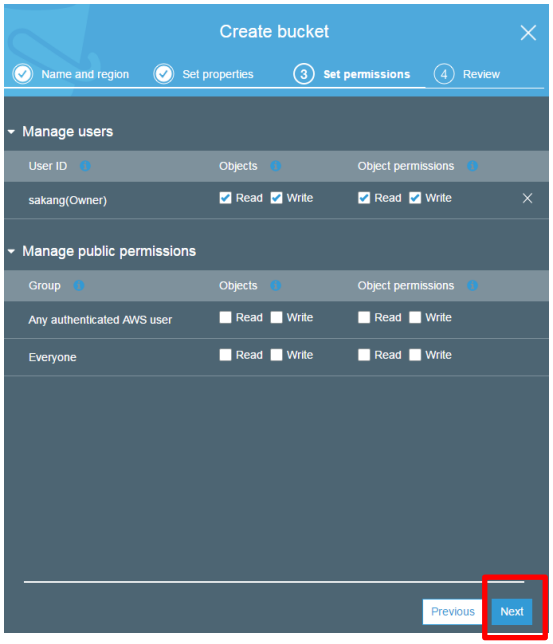
```
load("D:/업무/빅데이터팀/세미나/세미나/machine learning platform/Binary.rdata")
write.csv(D_cat, "D:/업무/빅데이터팀/세미나/세미나/machine learning platform/Binary.csv")

load("D:/업무/빅데이터팀/세미나/세미나/machine learning platform/Linear.rdata")
write.csv(D_lin, "D:/업무/빅데이터팀/세미나/세미나/machine learning platform/Linear.csv")
```

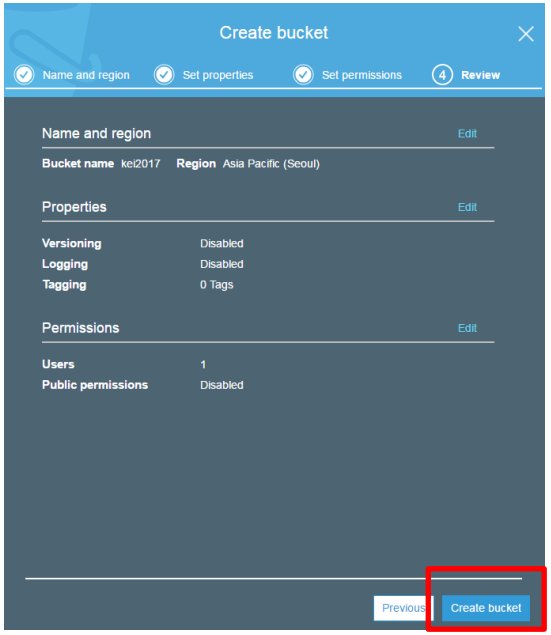
```
> str(binary)
'data.frame':   999951 obs. of  6 variables:
 $ X    : int  1 2 3 4 5 6 7 8 9 10 ...
 $ ystar: int  0 0 1 1 0 1 0 1 0 1 ...
 $ X1   : num  0.378 -0.853 0.748 0.933 -0.933 ...
 $ X2   : num  -0.0128 -1.5188 0.8686 0.0796 -0.3775 ...
 $ X3   : num  0.3596 0.0454 -0.6697 -1.184 1.6786 ...
 $ 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  
>
```

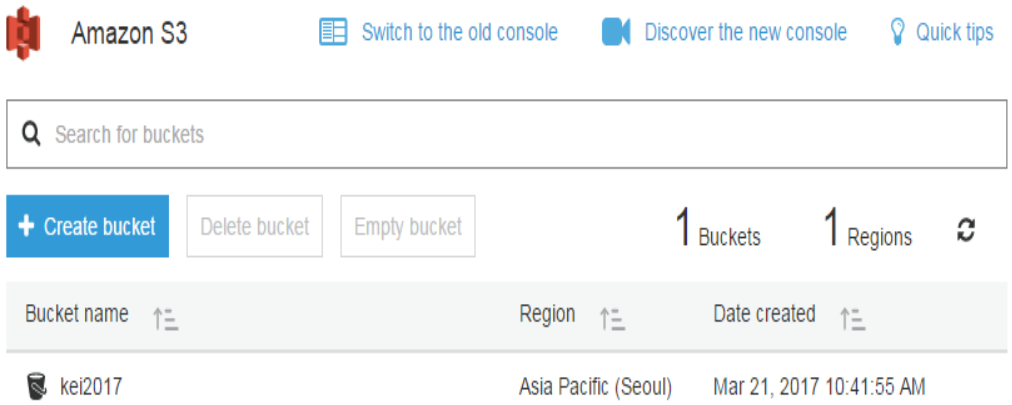
```
> str(linear)
'data.frame':   1000000 obs. of  6 variables:
 $ X : int  1 2 3 4 5 6 7 8 9 10 ...
 $ y : num  -2.042 -2.06 0.145 5.921 -7.543 ...
 $ X1: num  0.378 -0.853 0.748 0.933 -0.933 ...
 $ X2: num  -0.0128 -1.5188 0.8686 0.0796 -0.3775 ...
 $ X3: num  0.3596 0.0454 -0.6697 -1.184 1.6786 ...
 $ 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  
>
```

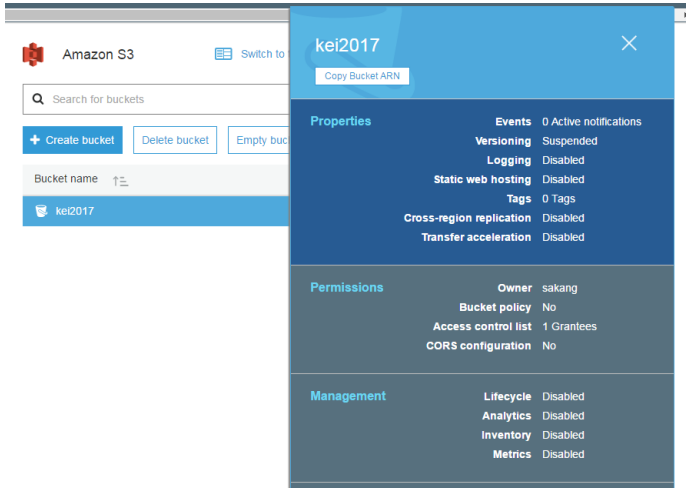

Create bucket
1 Name and region
2 Set properties
3 Set permissions
4 Review
Name and region
Bucket name
kei2017
Region
Asia Pacific (Seoul)
Copy settings from an existing bucket
You have no buckets
0 Buckets
Create
Cancel
Next

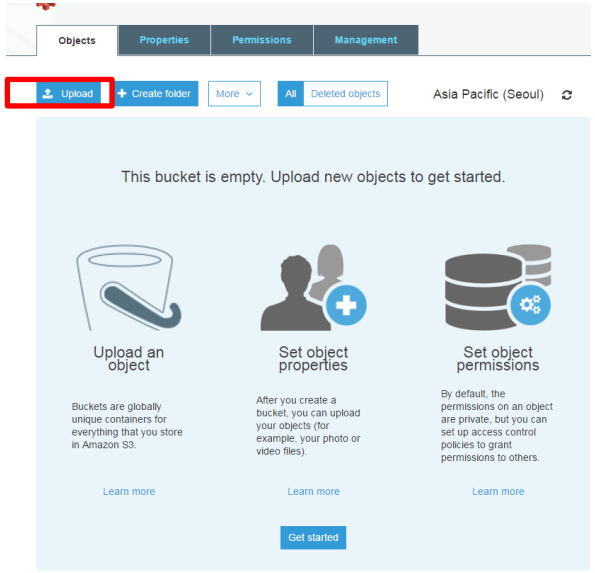

Create bucket
Name and region
2 Set properties
3 Set permissions
4 Review
Versioning
Enable versioning
Suspend versioning
This suspends the creation of object versions for all operations but preserves any existing object versions.
Cancel
Save
Logging
Set up access log records that provide details about access requests.
Learn more
Disabled
Tags
Use tags to track your cost against projects or other criteria.
Learn more
0 Tags
Previous
Next

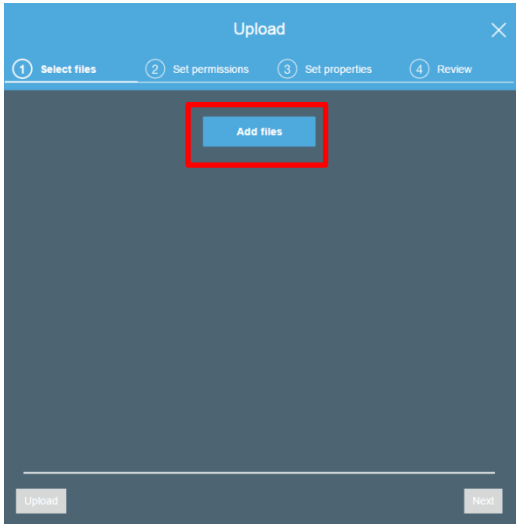

Create bucket
Name and region
Set properties
3 Set permissions
4 Review
Manage users
User ID
Objects
Object permissions
sakang(Owner)
Read
Write
Read
Write
Manage public permissions
Group
Objects
Object permissions
Any authenticated AWS user
Everyone
Previous
Next

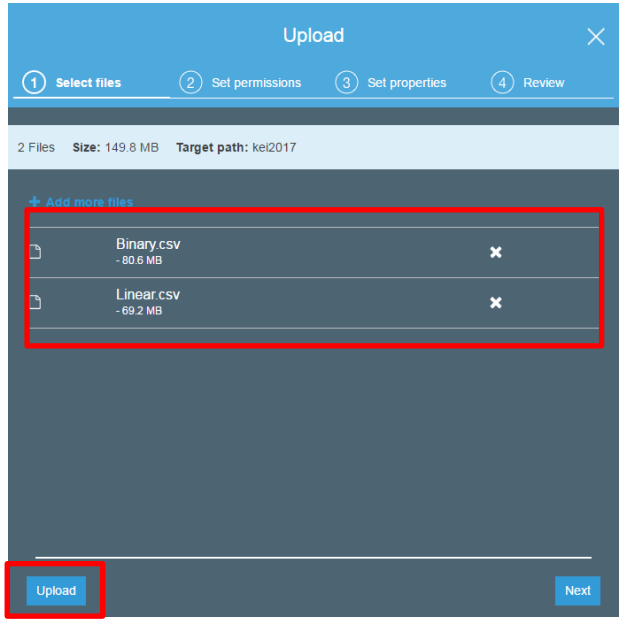

Create bucket
Name and region
Set properties
Set permissions
4 Review
Name and region
Edit
Bucket name kei2017 Region Asia Pacific (Seoul)
Properties
Edit
Versioning Disabled
Logging Disabled
Tagging 0 Tags
Permissions
Edit
Users 1
Public permissions Disabled
Previous
Create bucket

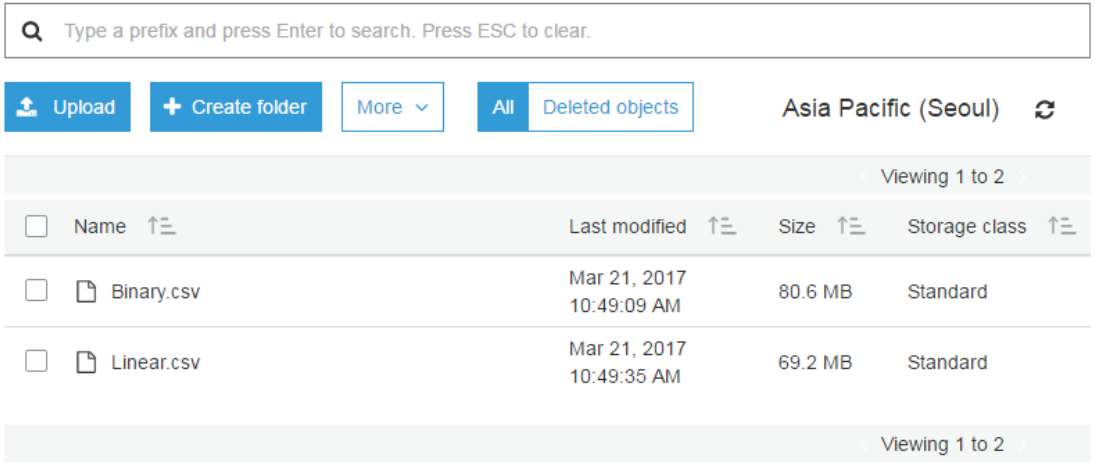

Amazon S3
Switch to the old console
Discover the new console
Quick tips
Search for buckets
Create bucket
Delete bucket
Empty bucket
1 Buckets
1 Regions
Bucket name
Region
Date created
kei2017
Asia Pacific (Seoul)
Mar 21, 2017 10:41:55 AM


Amazon S3
Switch to
Search for buckets
+ Create bucket
Delete bucket
Empty buc
Bucket name
kei2017
kei2017
Copy Bucket ARN
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

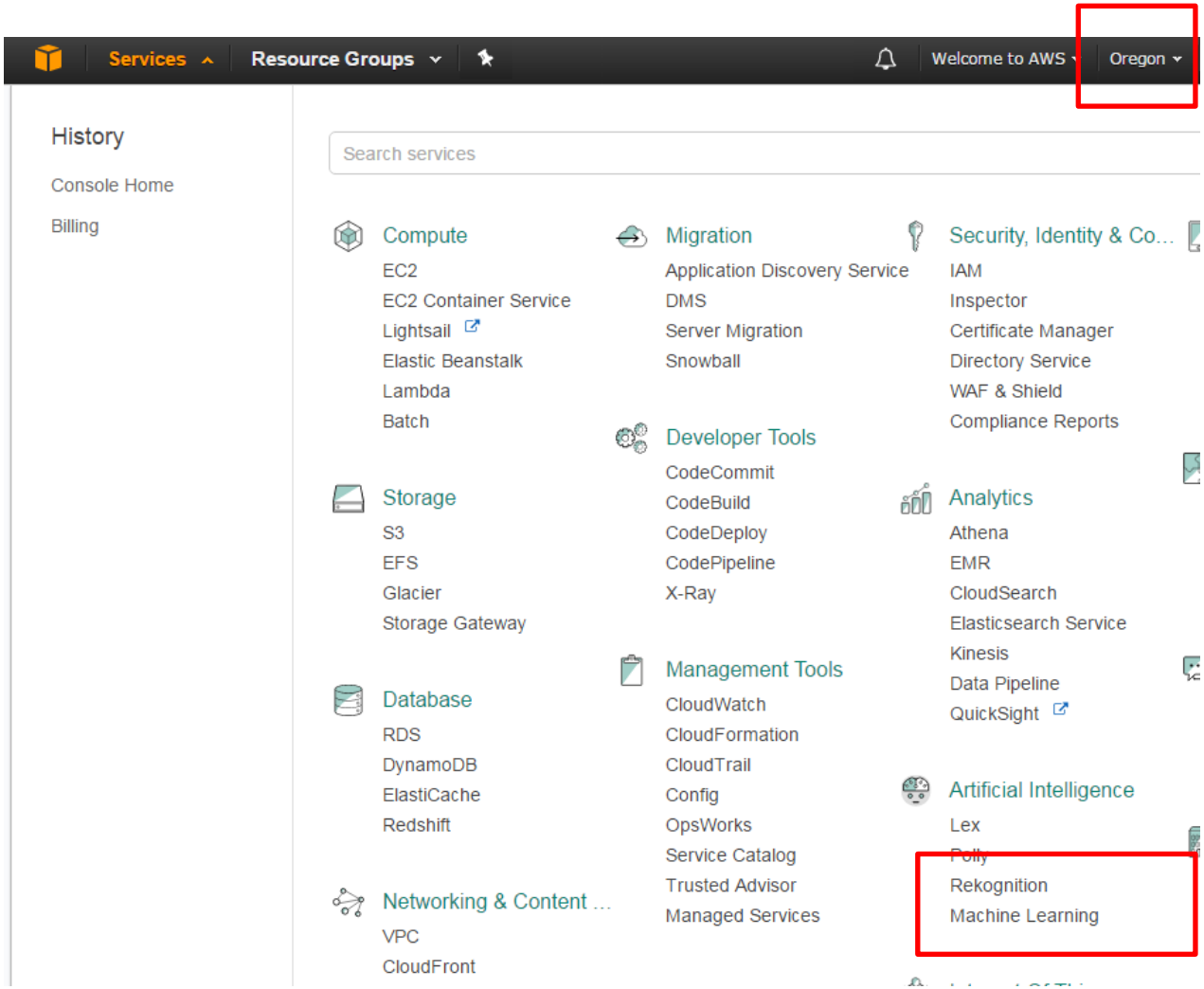

Objects
Properties
Permissions
Management
Upload
+ Create folder
More
All
Deleted objects
Asia Pacific (Seoul)
This bucket is empty. Upload new objects to get started.
Upload an object
Buckets are globally unique containers for everything that you store in Amazon S3.
Learn more
Set object properties
After you create a bucket, you can upload your objects (for example, your photo or video files).
Learn more
Set object permissions
By default, the permissions on an object are private, but you can set up access control policies to grant permissions to others.
Learn more
Get started


Upload
1 Select files
2 Set permissions
3 Set properties
4 Review
Add files
Upload
Next

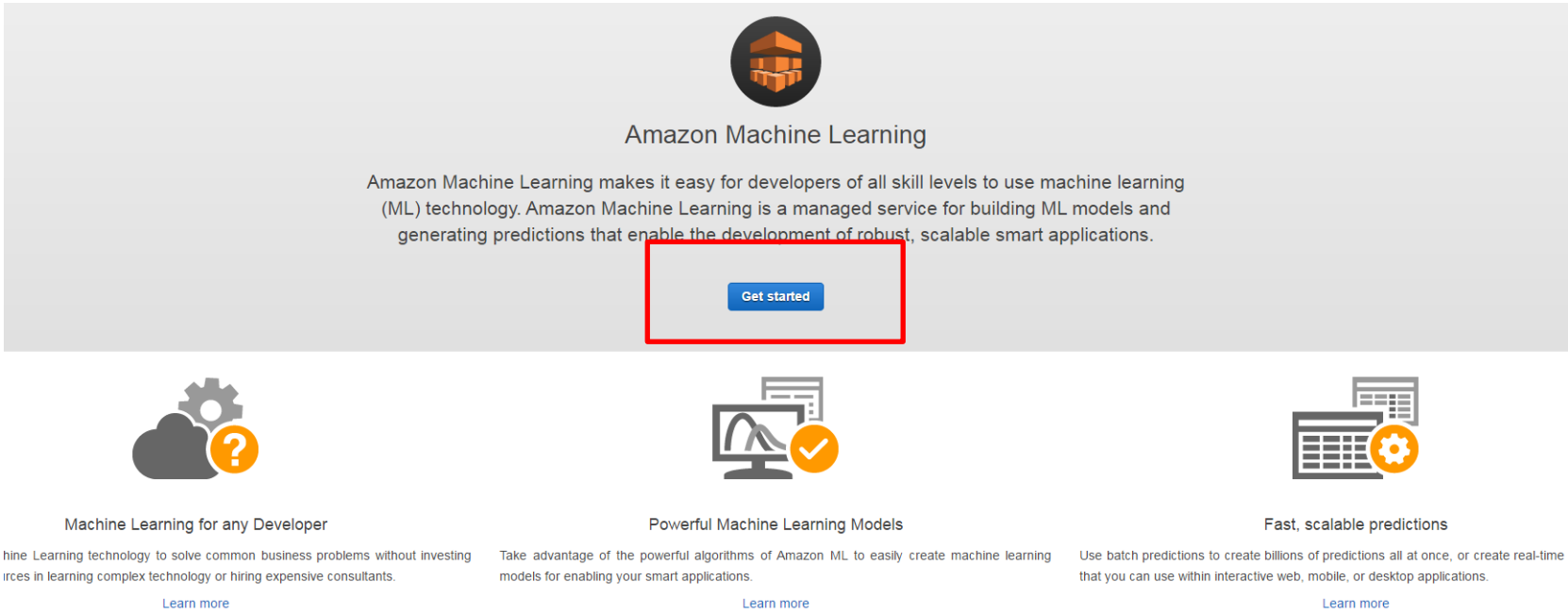

Upload
1 Select files
2 Set permissions
3 Set properties
4 Review
2 Files Size: 149.8 MB Target path: kei2017
+ Add more files
Binary.csv
- 80.6 MB
Linear.csv
- 69.2 MB
Upload
Next


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

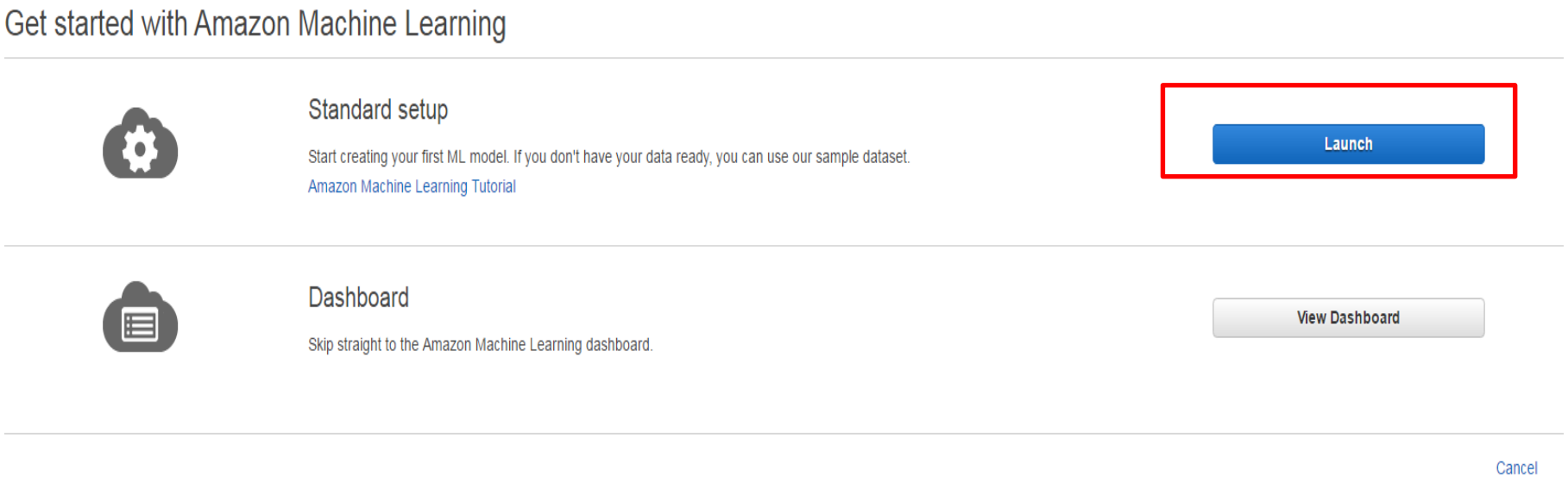
4. AWS를 이용하여 예측모형 구축(binary)



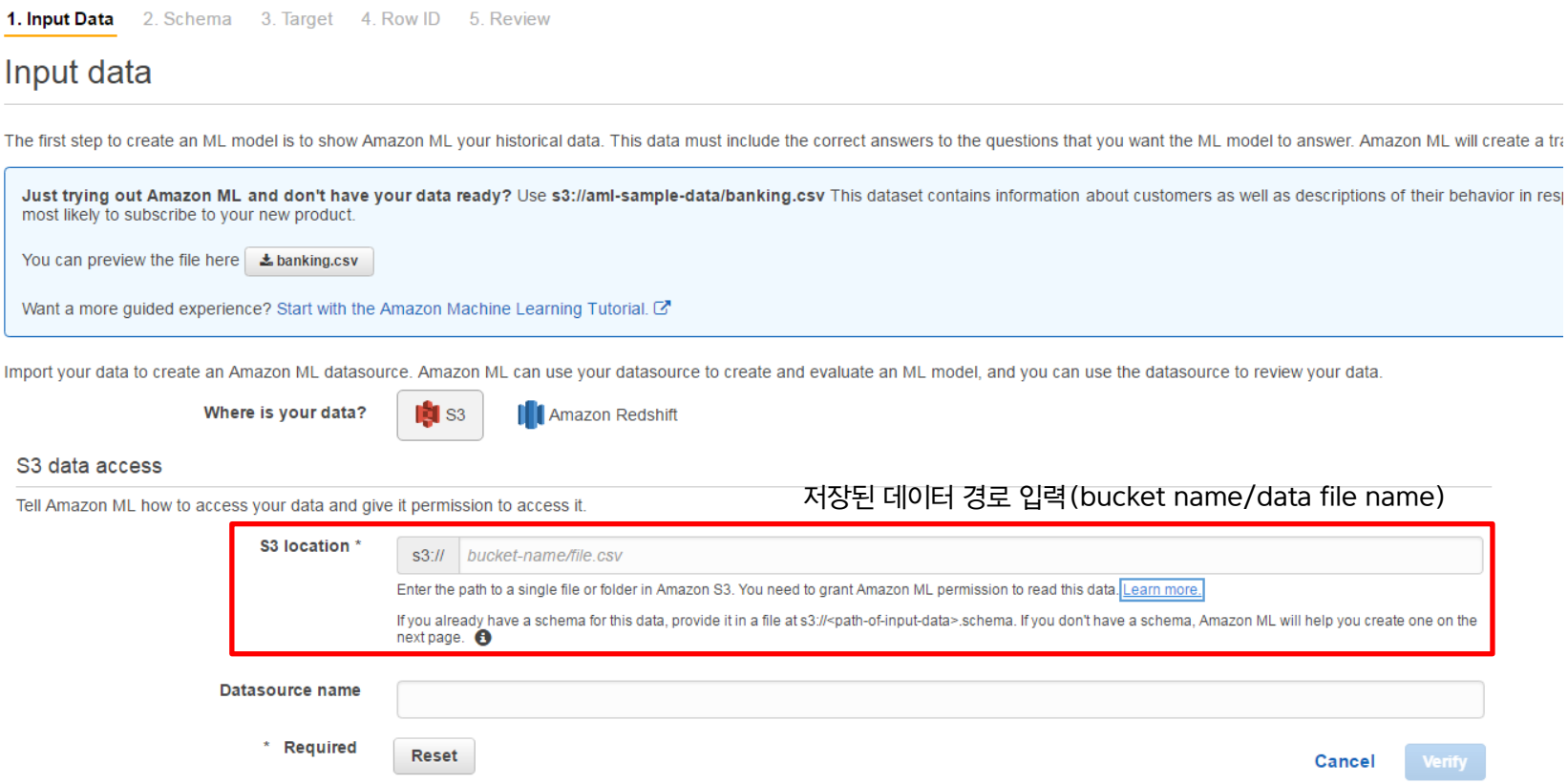
4. AWS를 이용하여 예측모형 구축(binary)



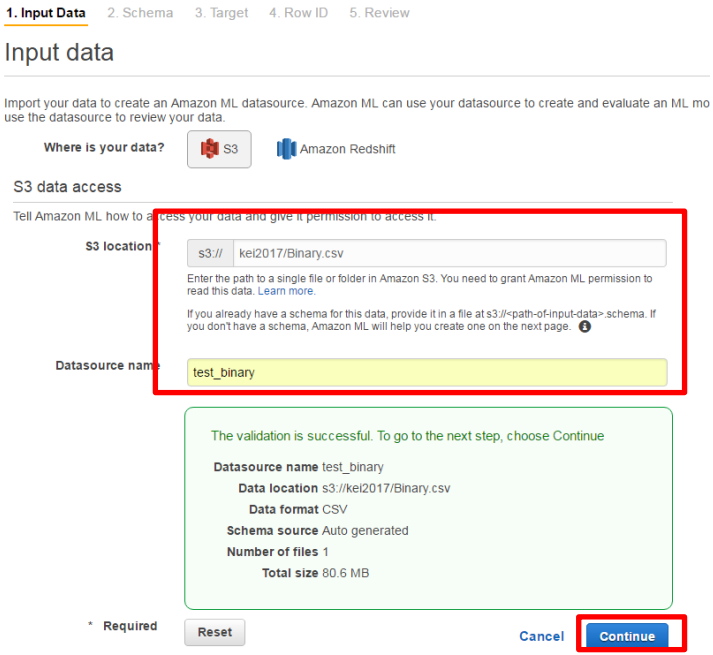
4. AWS를 이용하여 예측모형 구축(binary)

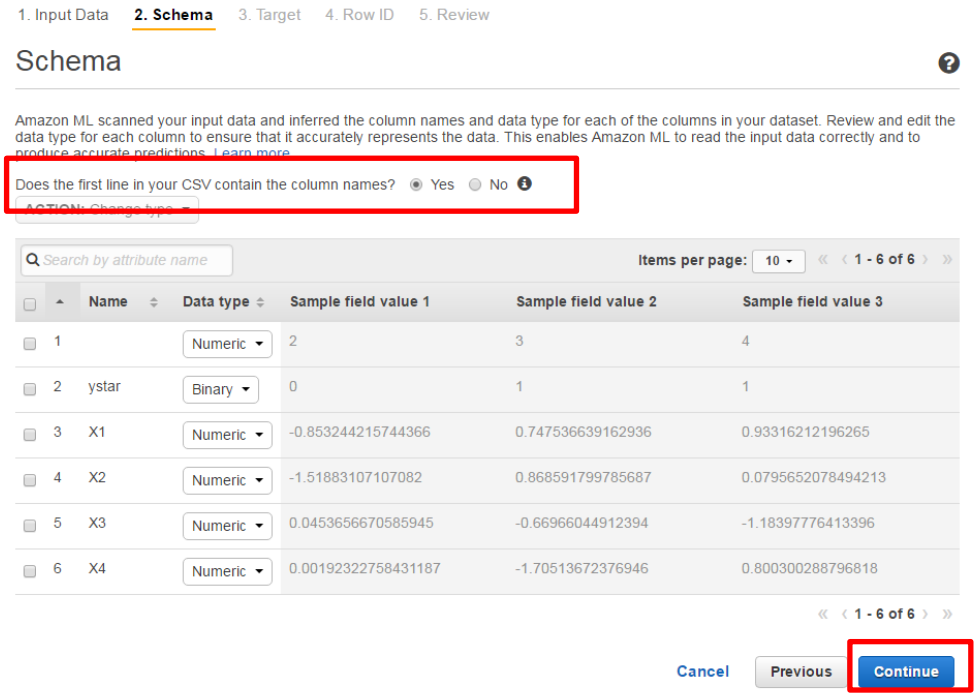


4. AWS를 이용하여 예측모형 구축(binary)

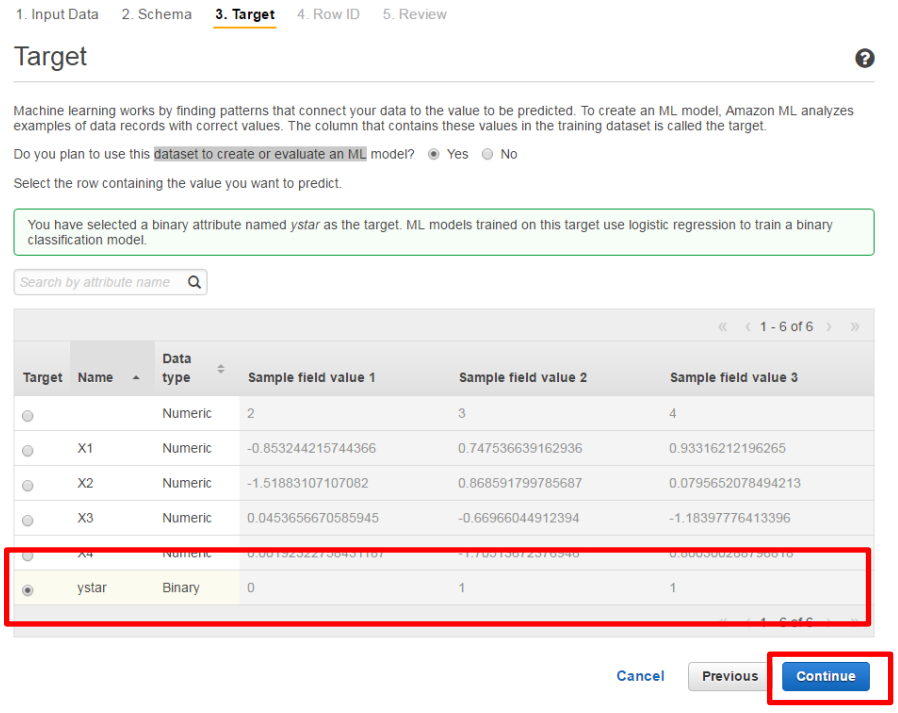


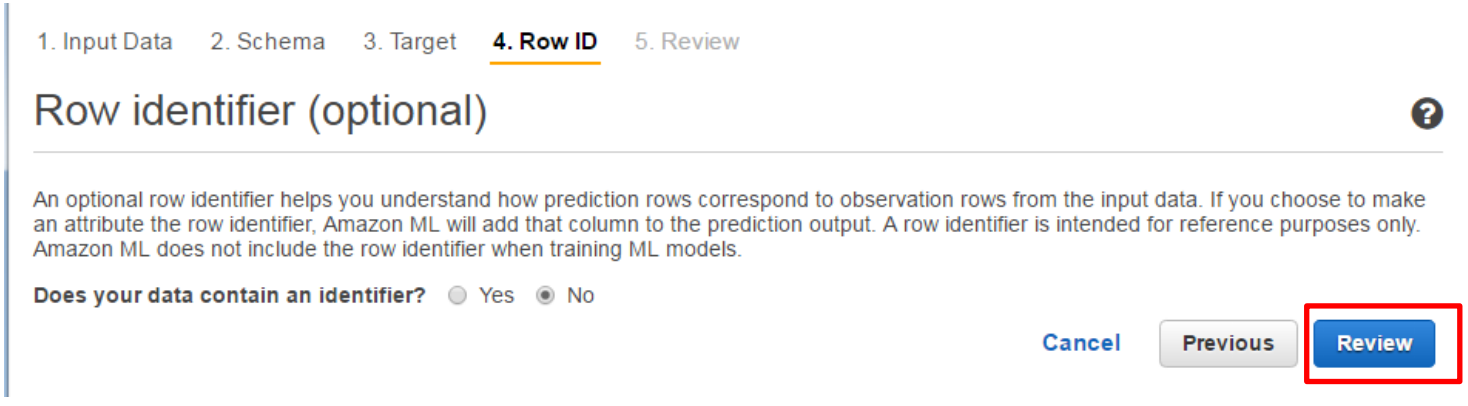
4. AWS를 이용하여 예측모형 구축(binary)



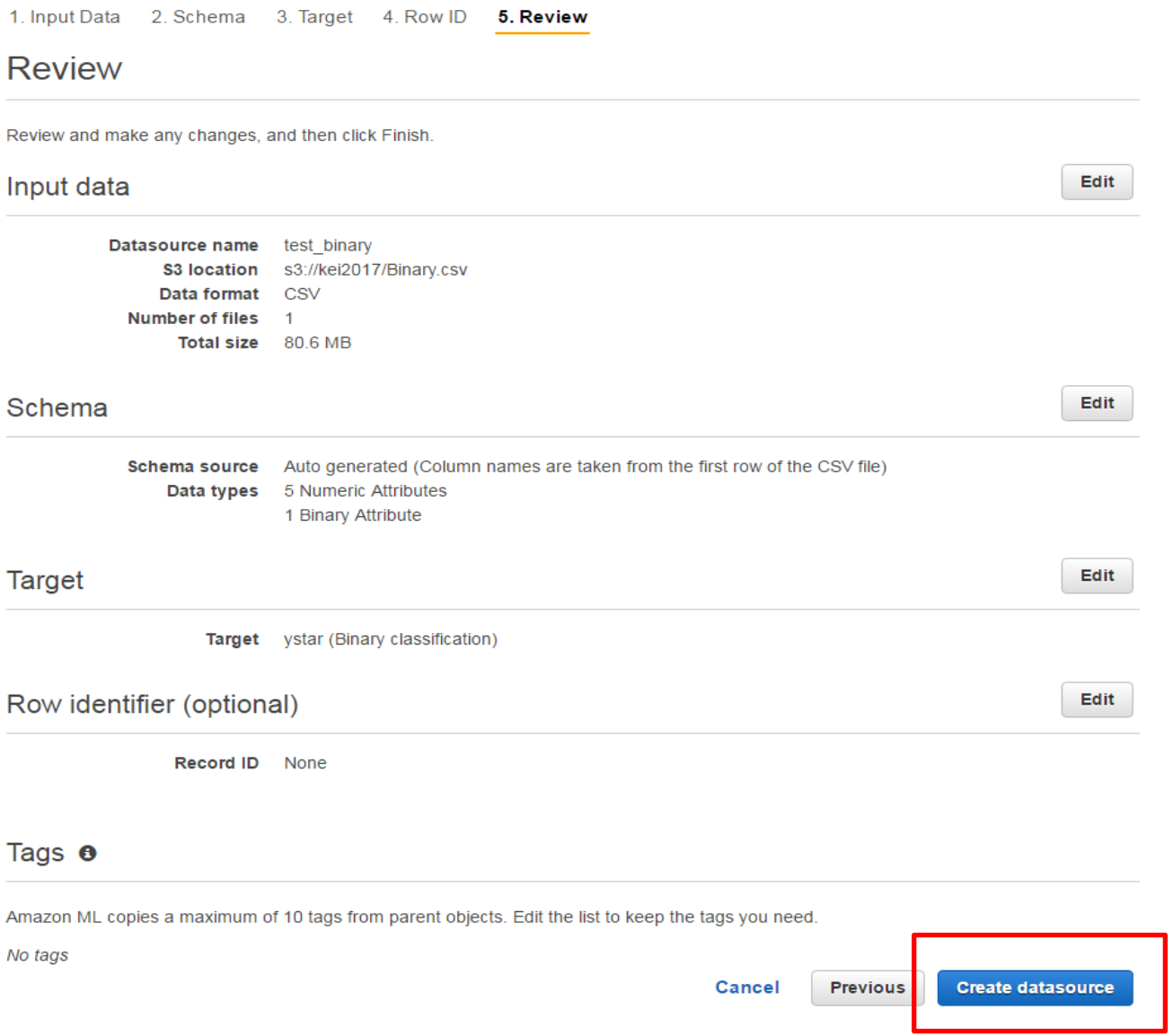


4. AWS를 이용하여 예측모형 구축(binary)

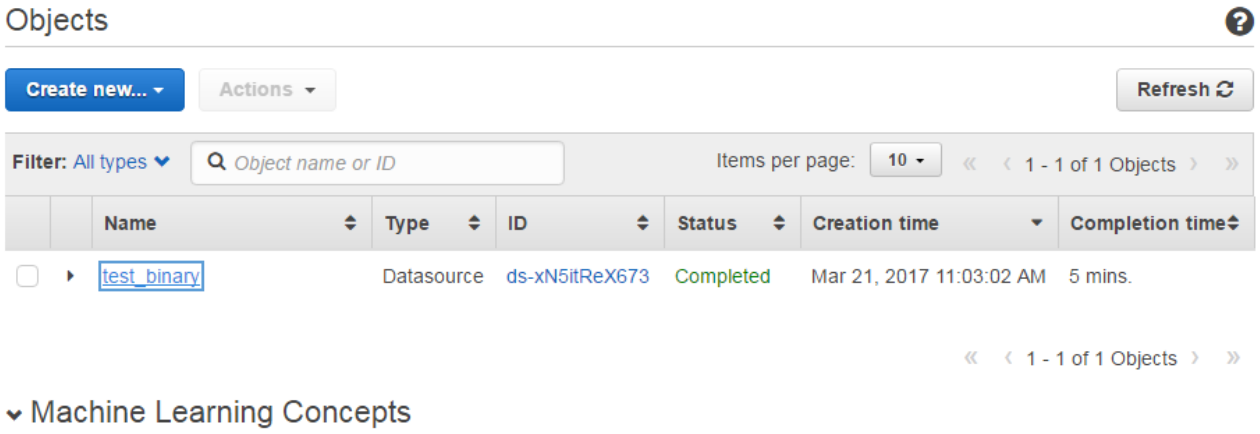




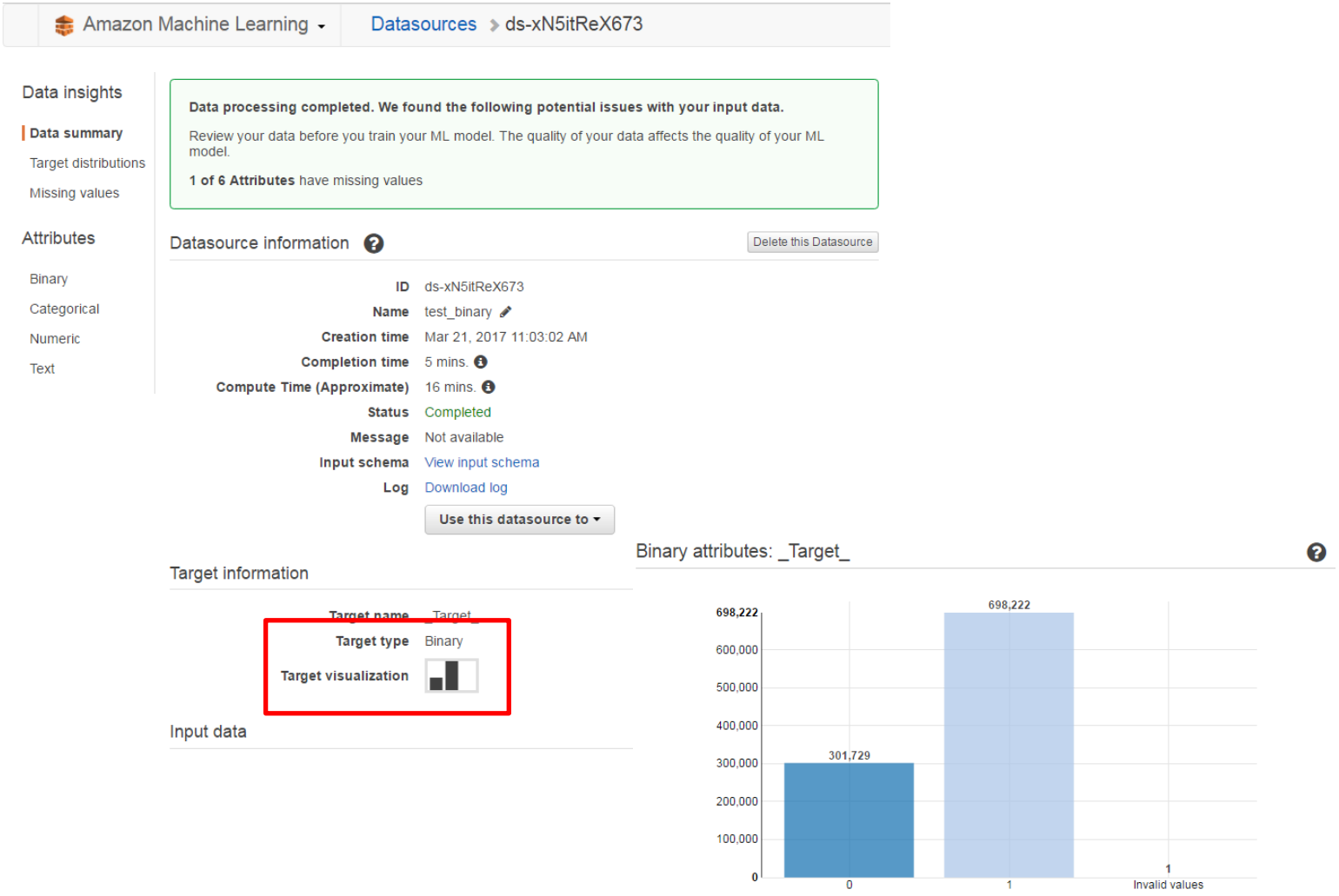
4. AWS를 이용하여 예측모형 구축(binary)



4. AWS를 이용하여 예측모형 구축(binary)



4. AWS를 이용하여 예측모형 구축(binary)



4. AWS를 이용하여 예측모형 구축(binary)

Data insights

Data summary
Target distributions
Missing values

Attributes

Binary
Categorical
Numeric
Text

Global attribute name Search

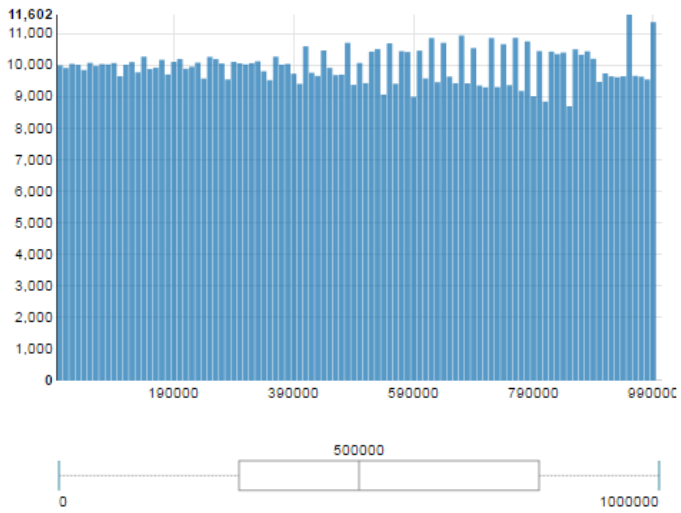
Numeric attributes

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	0.00075077291628656 1	

1 - 5 of 5

Numeric attributes: Var1

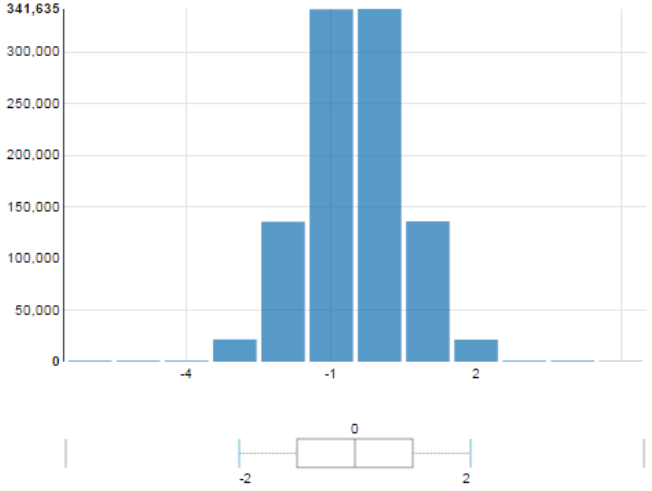
Select bin width: 1e+5 | 5e+4 | 20000 | 10000 | 5000



Min: 0 **Mean:** 500000 **Max:** 1000000

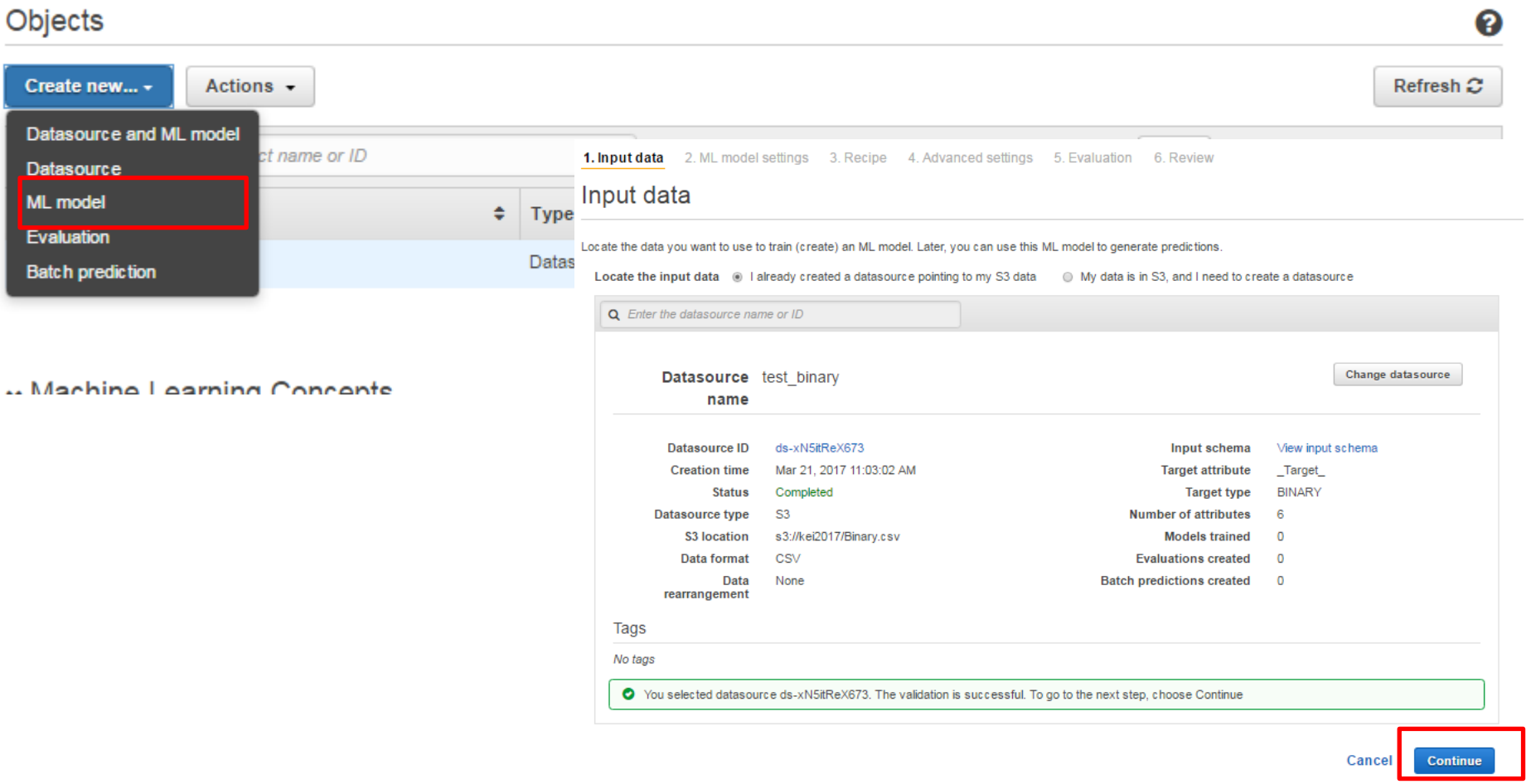
Numeric attributes: Var2

Select bin width: 1 | 0.5 | 0.2 | 0.1 | 0.05

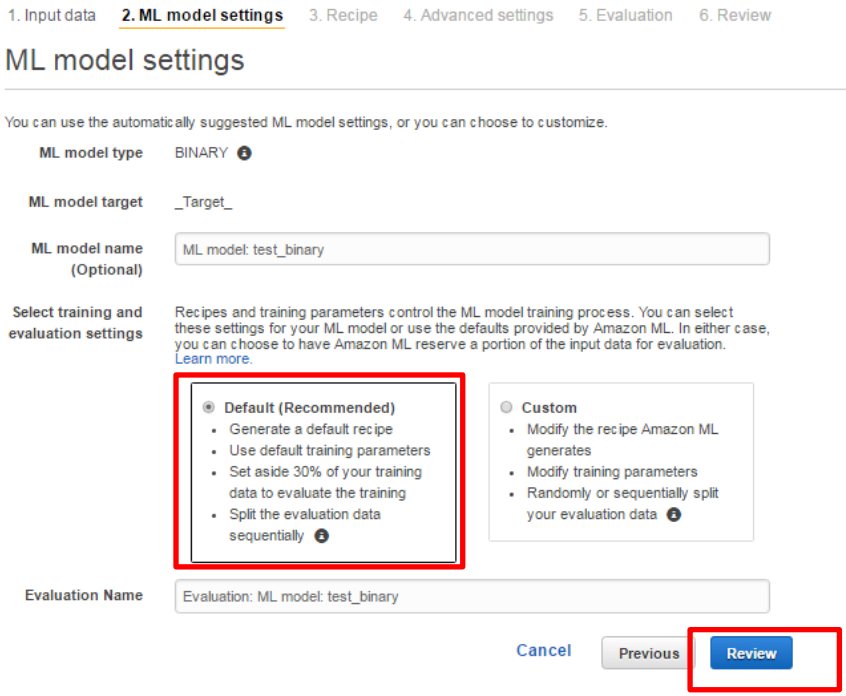


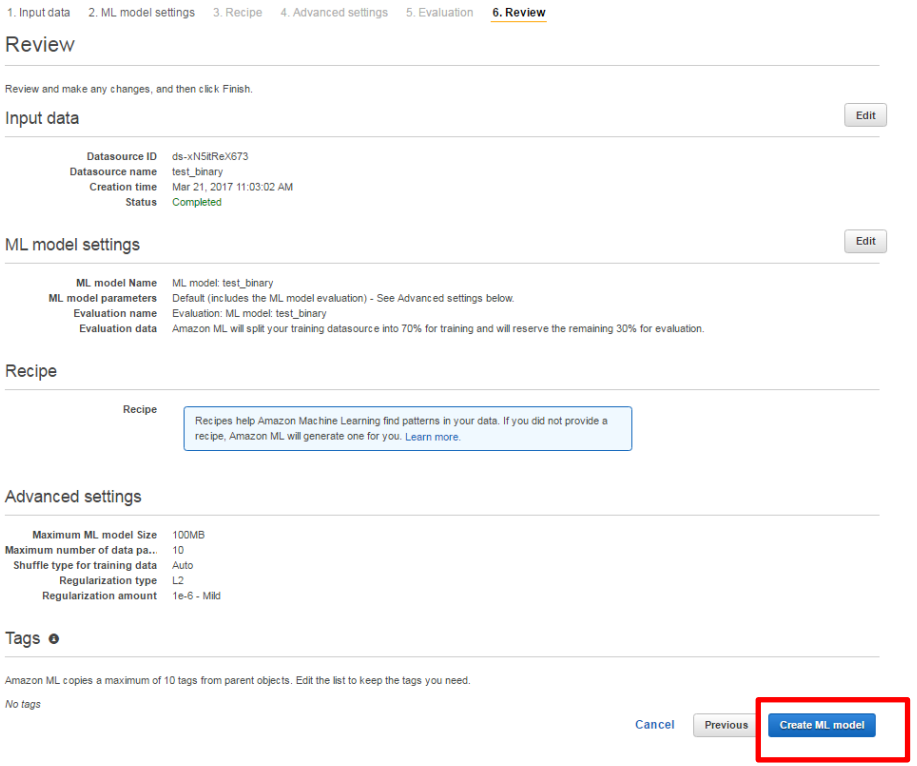
Min: -5.101 **Mean:** 0.001 **Max:** 4.611

4. AWS를 이용하여 예측모형 구축(binary)

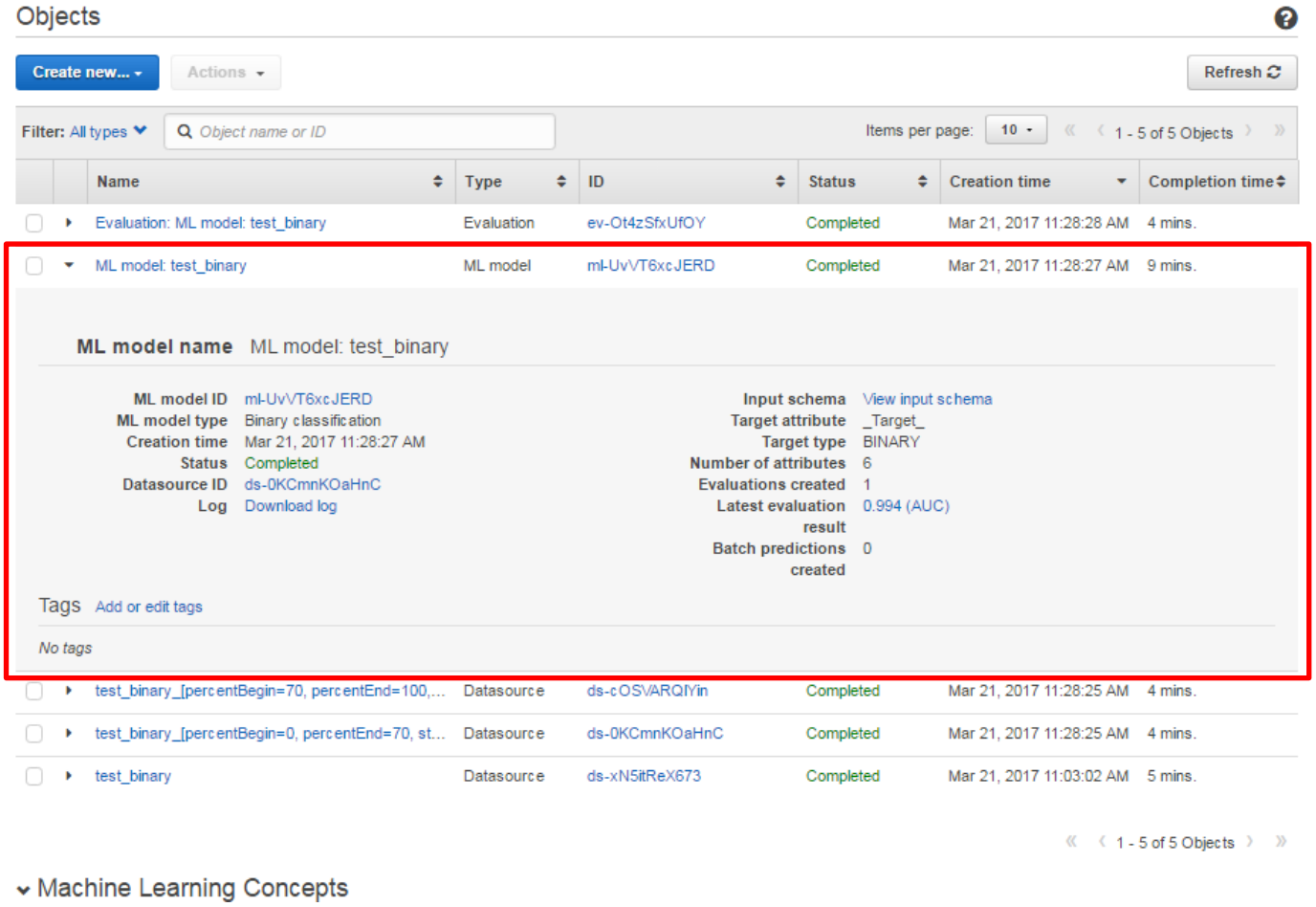


4. AWS를 이용하여 예측모형 구축(binary)

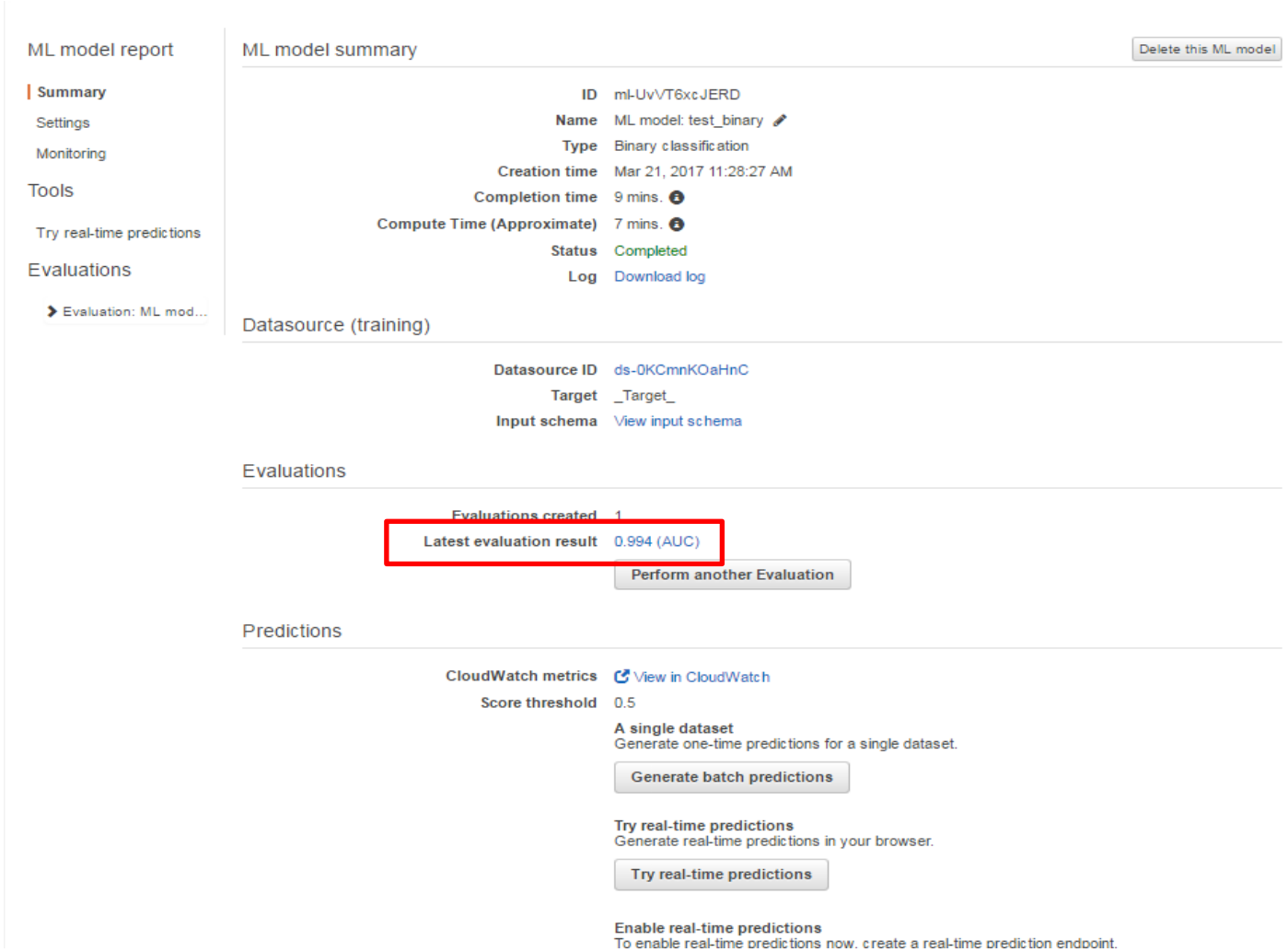




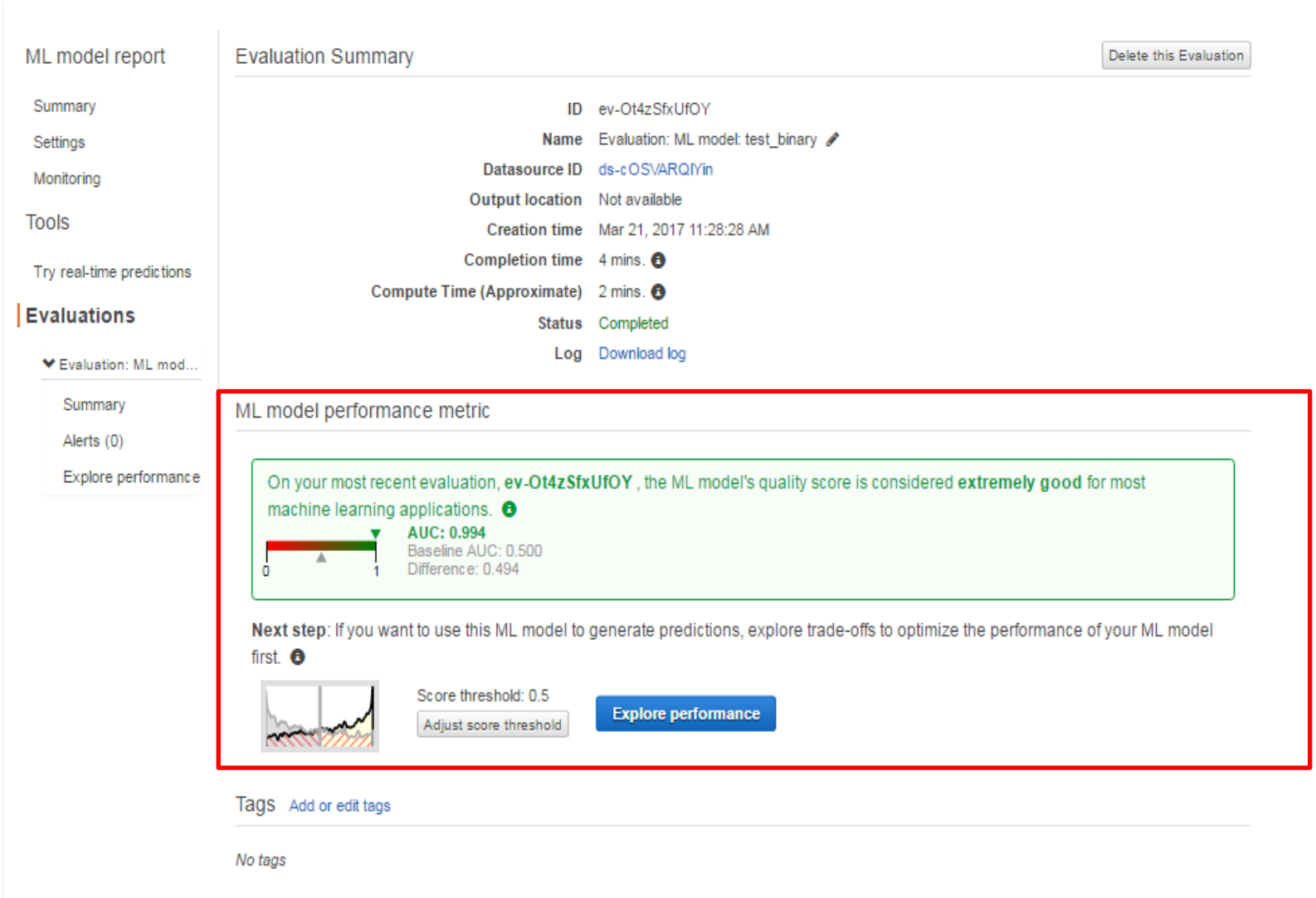
4. AWS를 이용하여 예측모형 구축(binary)-성능확인



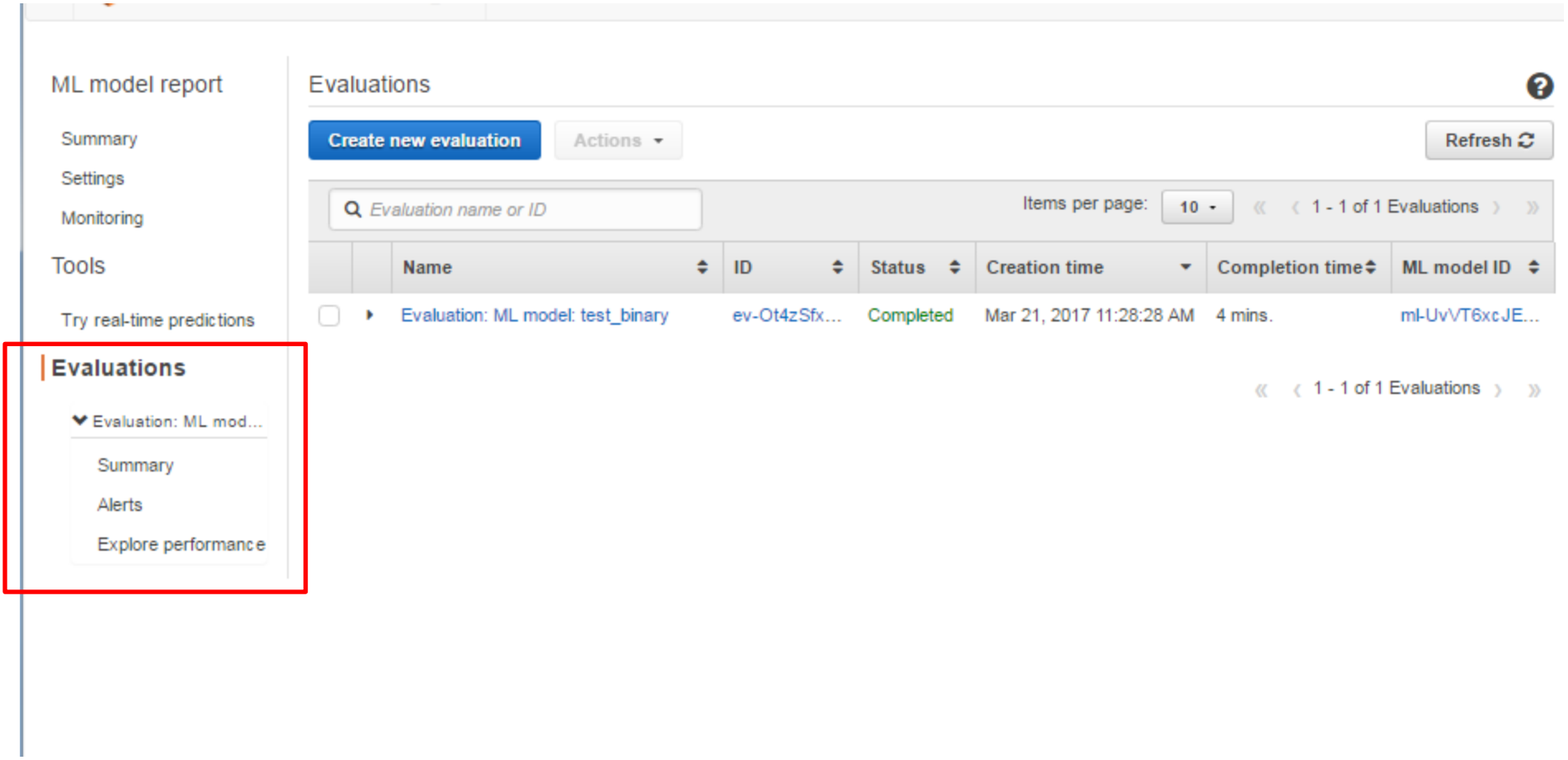
4. AWS를 이용하여 예측모형 구축(binary)-성능확인



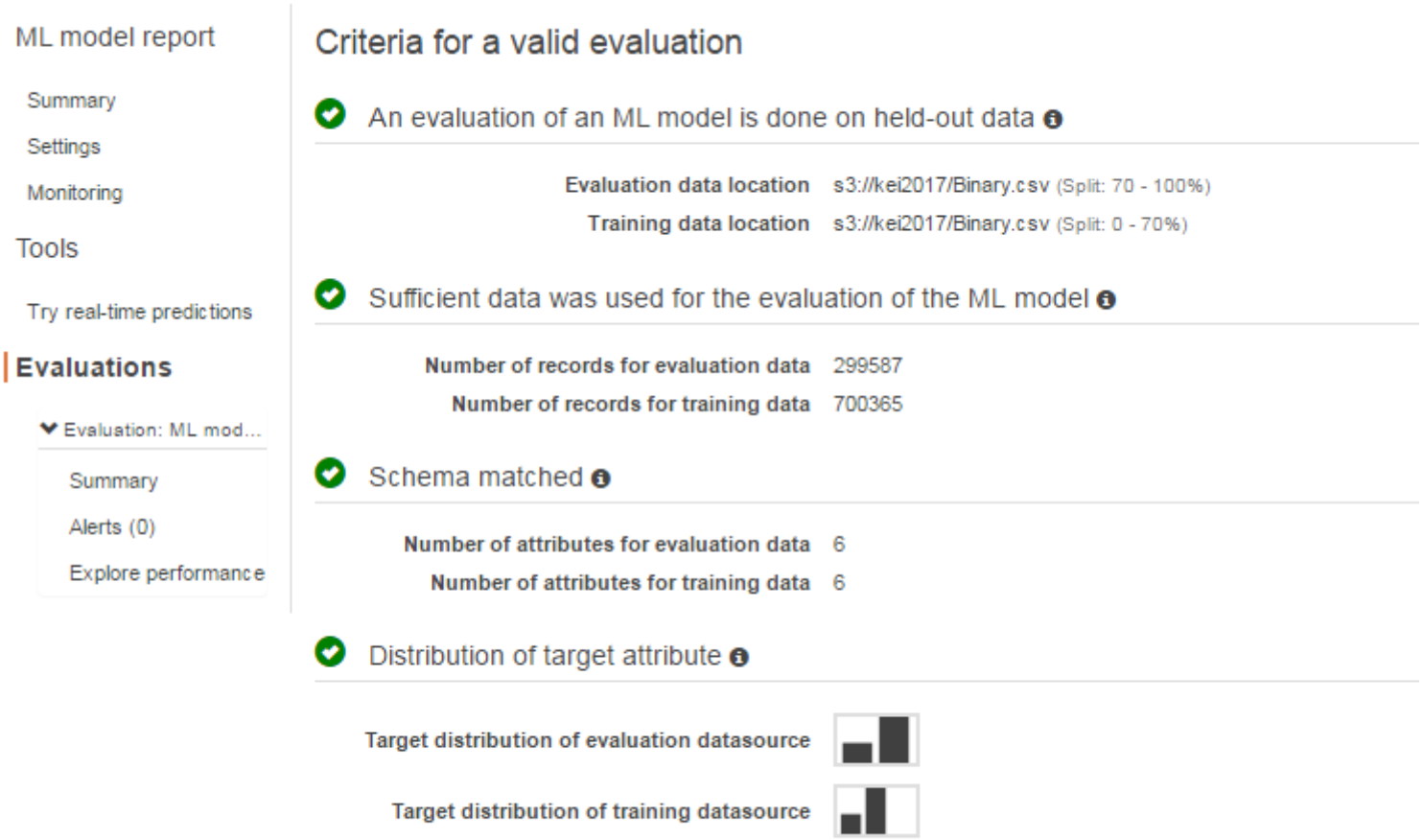
4. AWS를 이용하여 예측모형 구축(binary)-성능확인



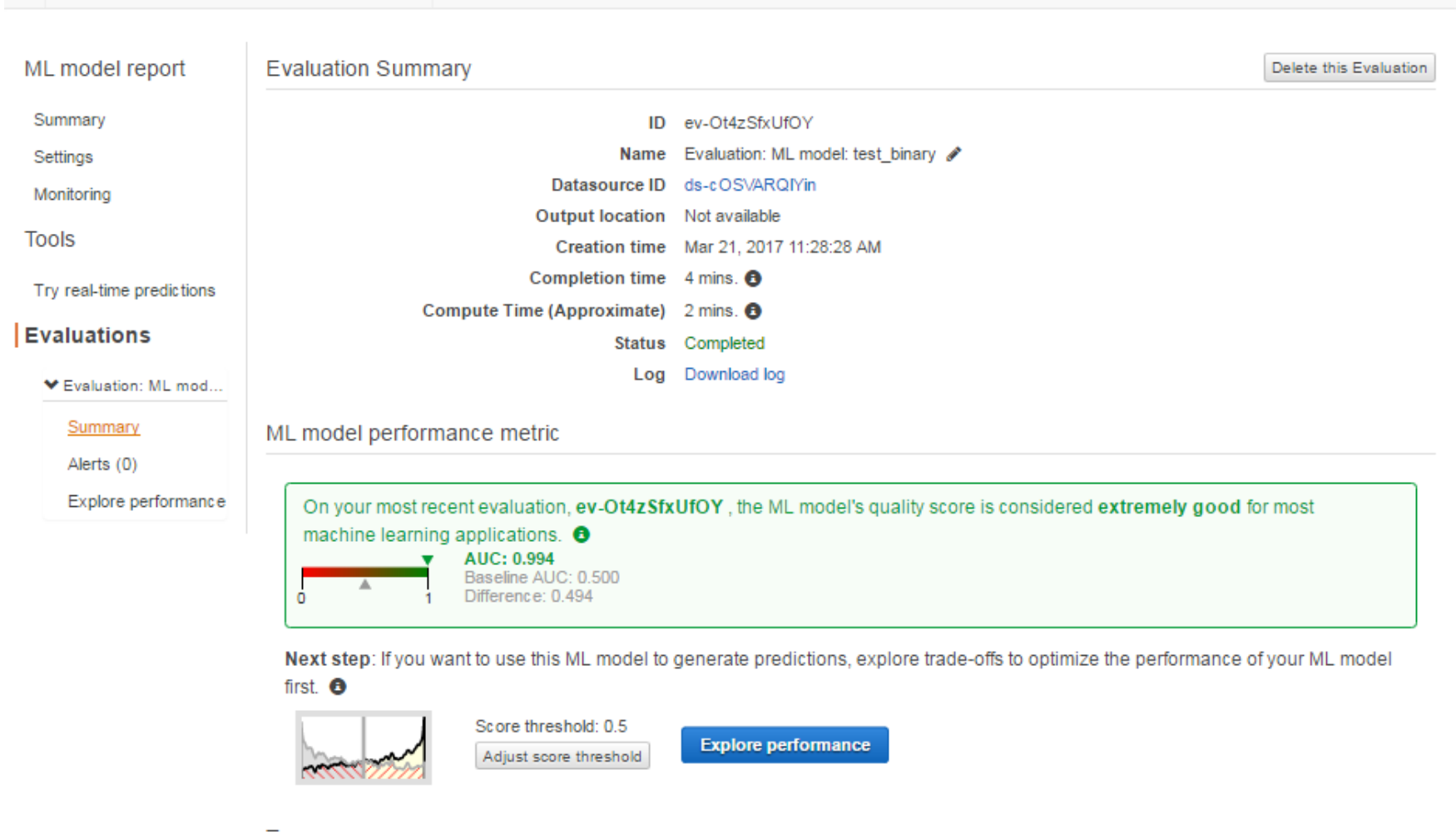
4. AWS를 이용하여 예측모형 구축(binary)-성능확인



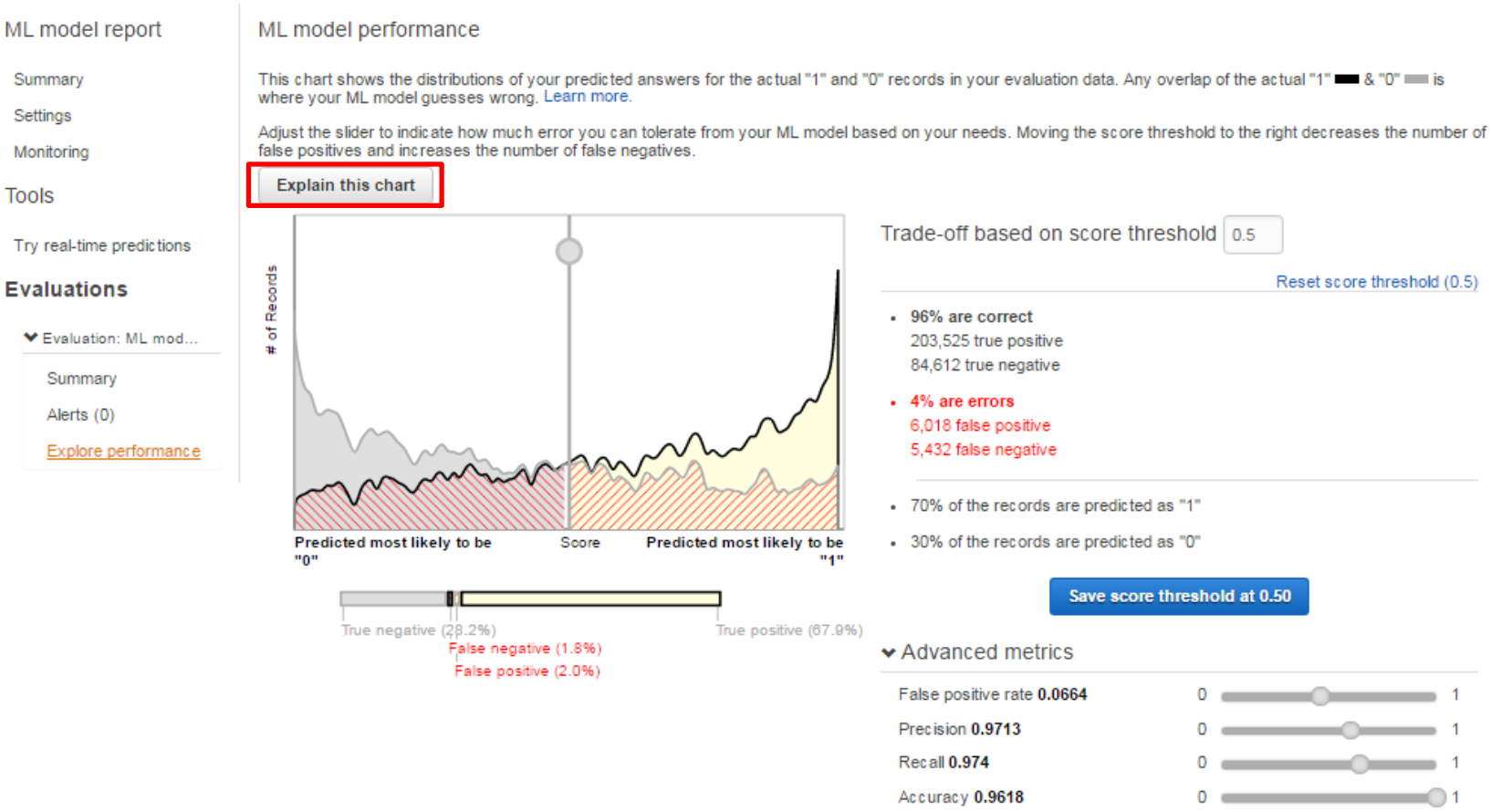
4. AWS를 이용하여 예측모형 구축(binary)-성능확인



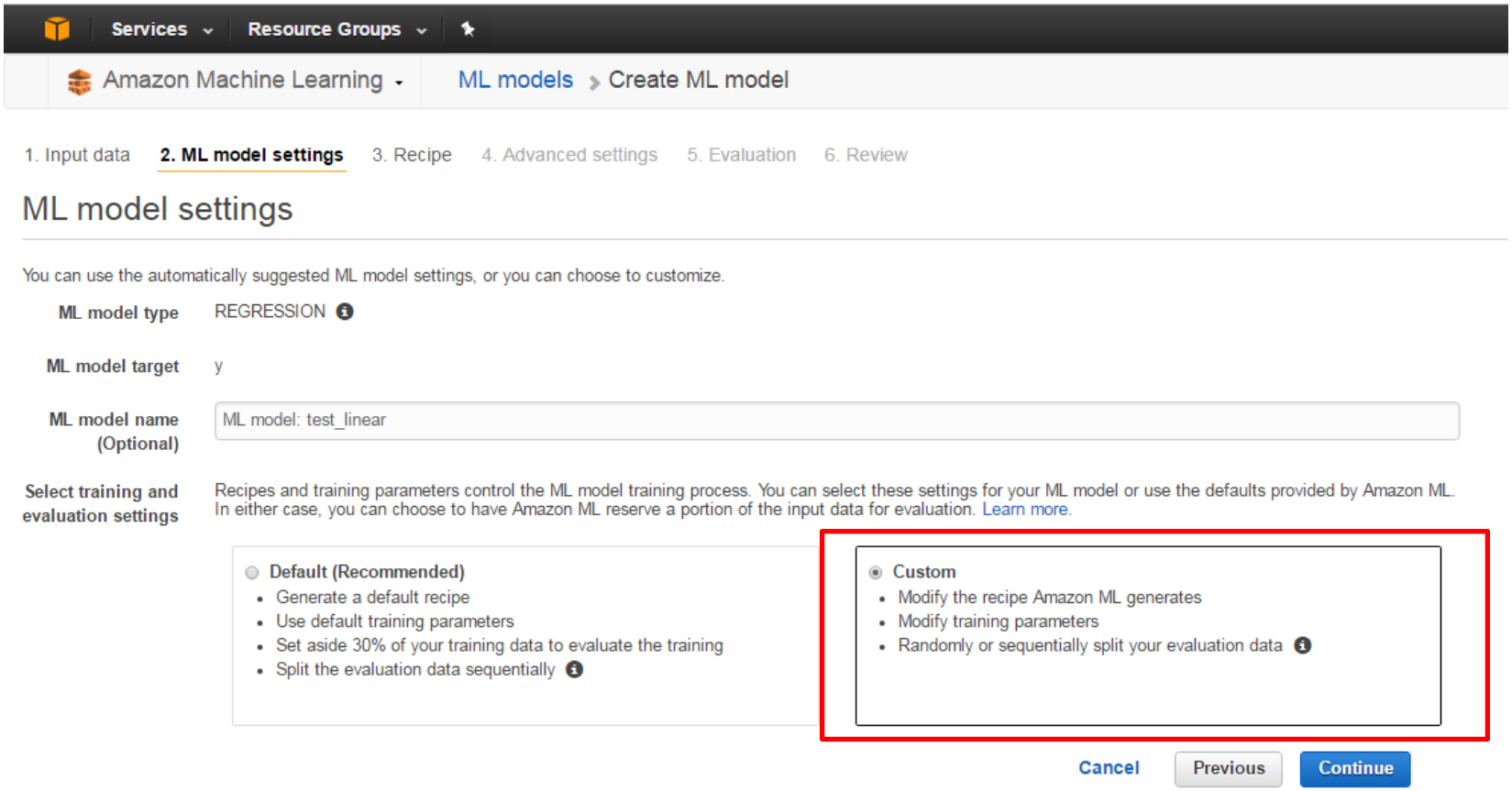
4. AWS를 이용하여 예측모형 구축(binary)-성능확인



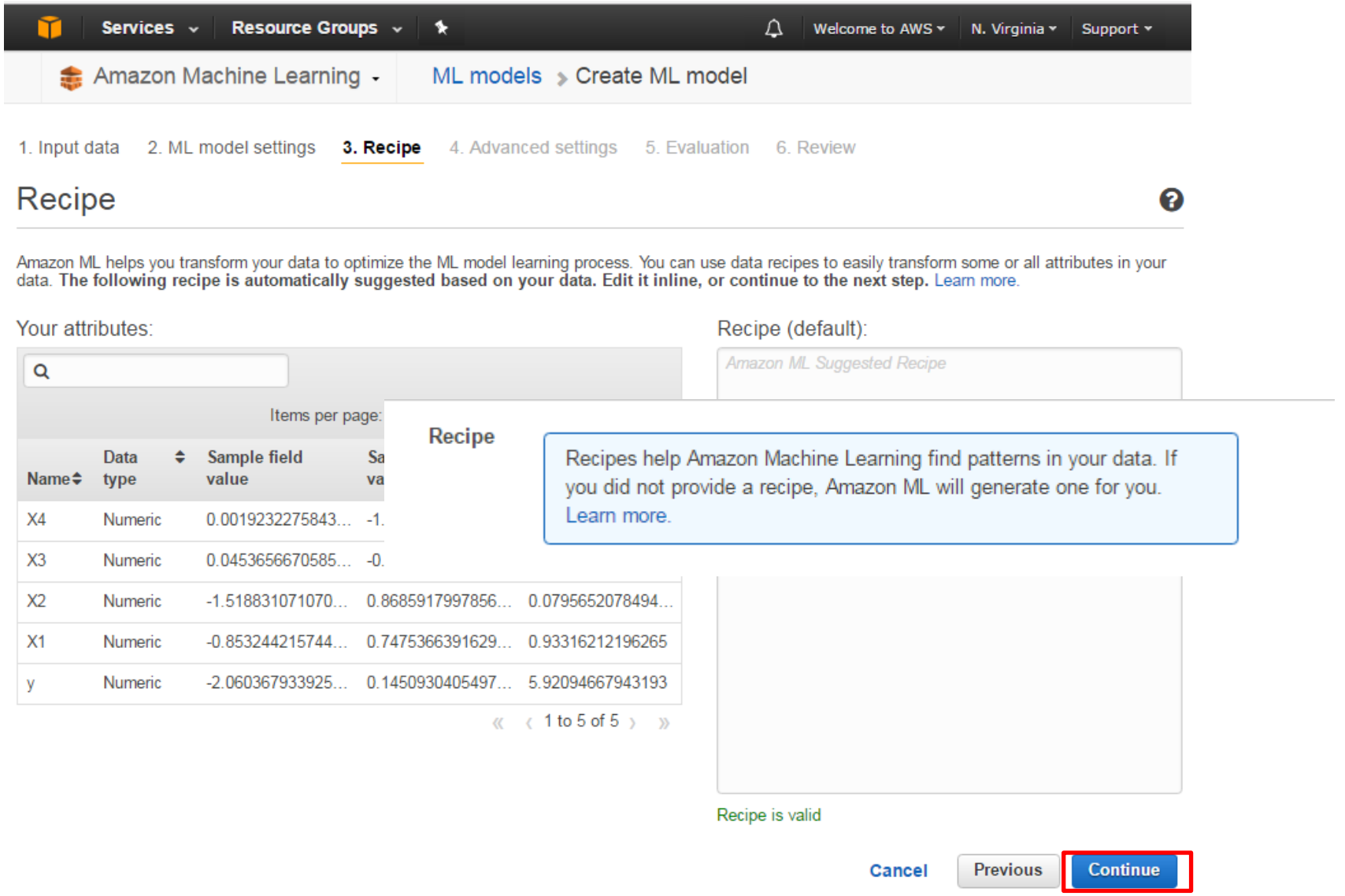
4. AWS를 이용하여 예측모형 구축(binary)-성능확인



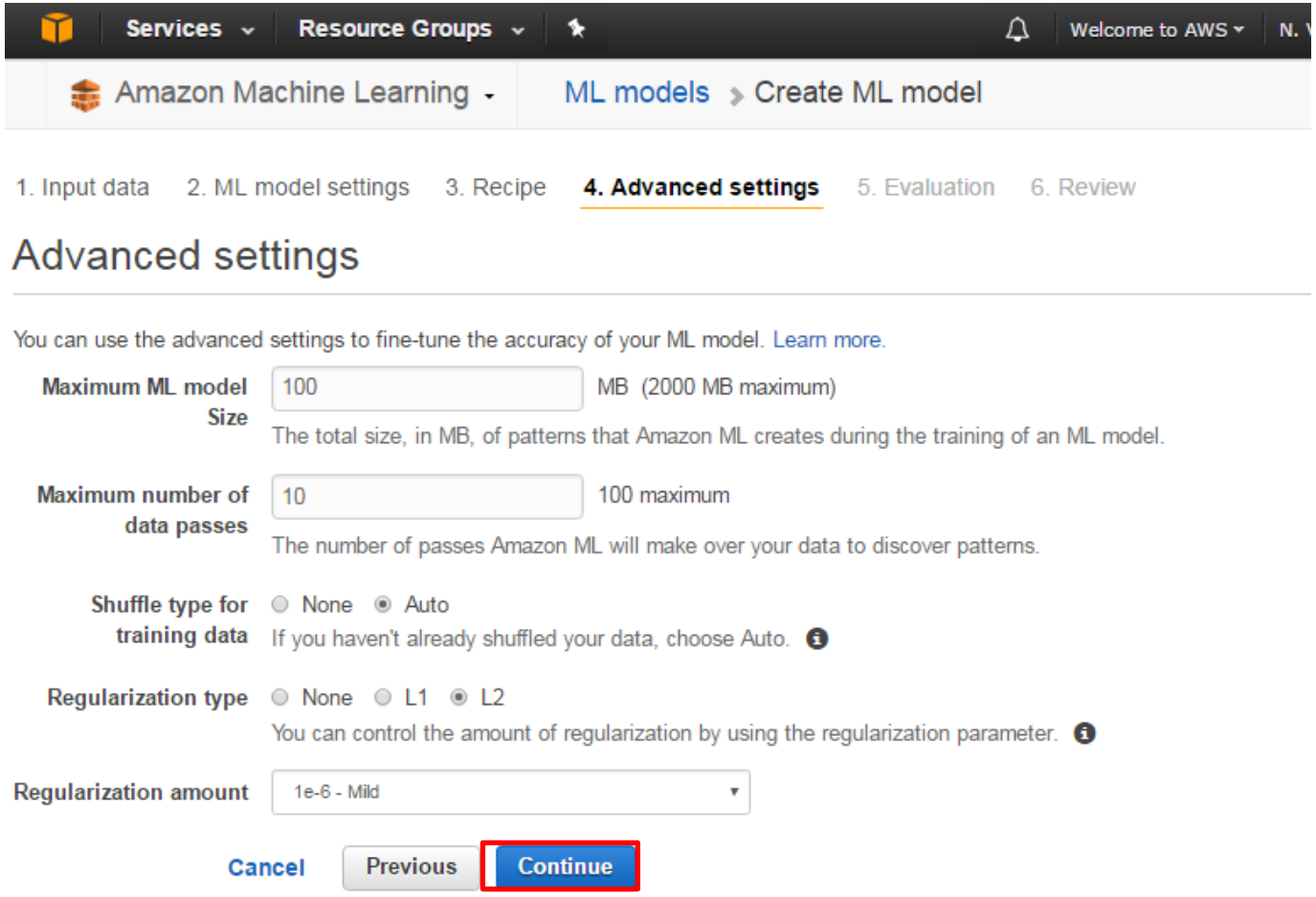
4. AWS를 이용하여 예측모형 구축(linear)



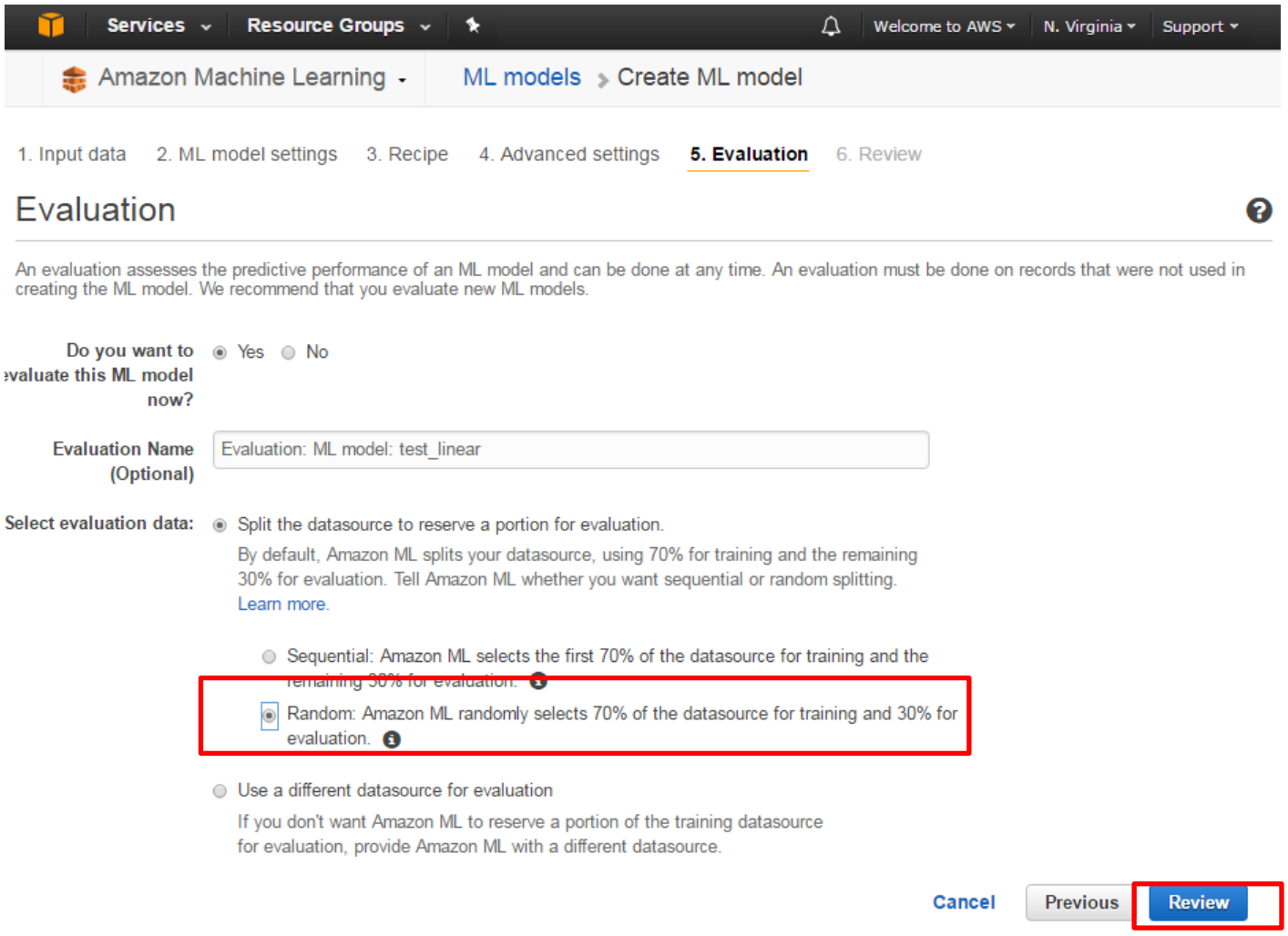
4. AWS를 이용하여 예측모형 구축(linear)



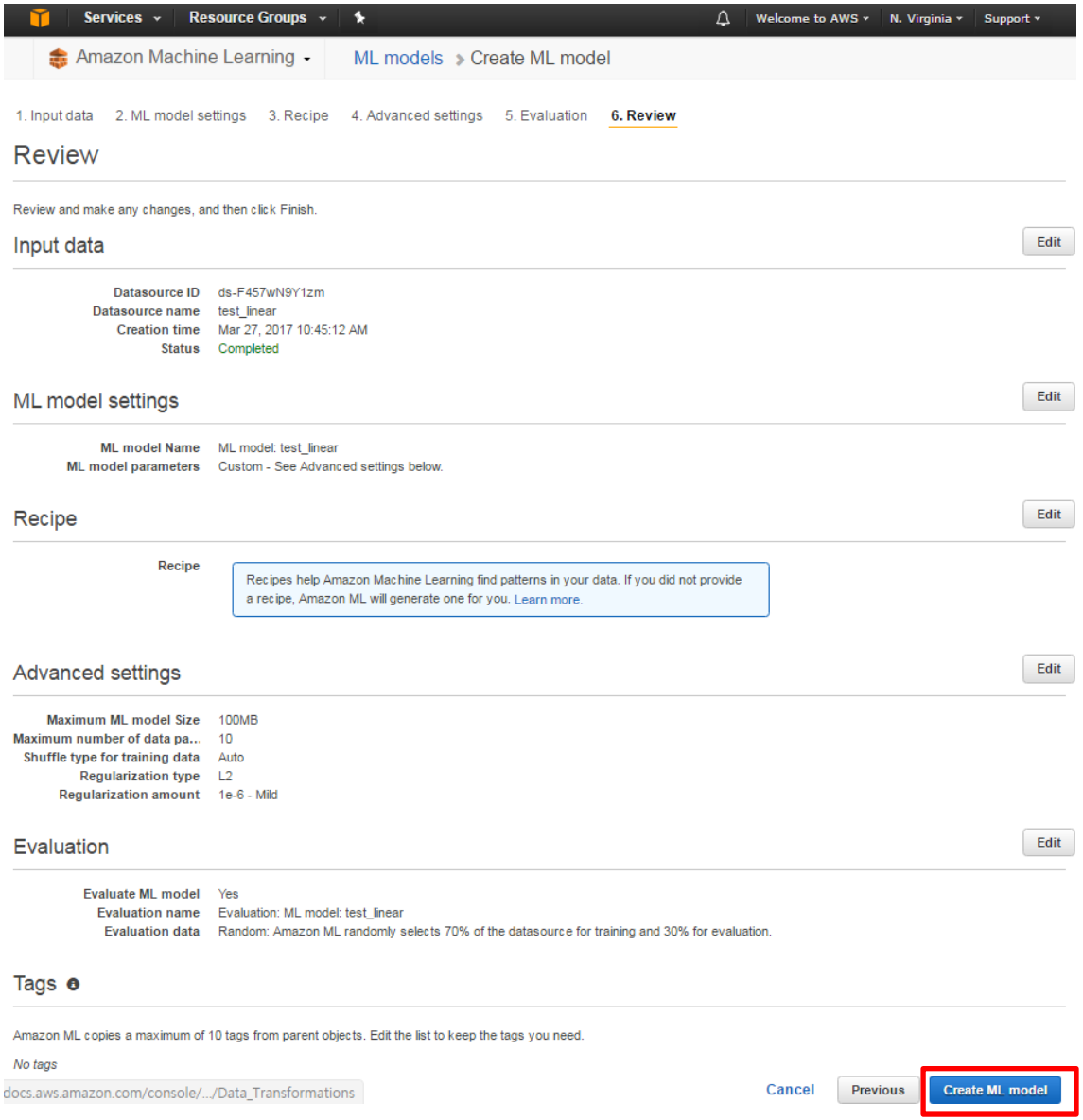
4. AWS를 이용하여 예측모형 구축(linear)



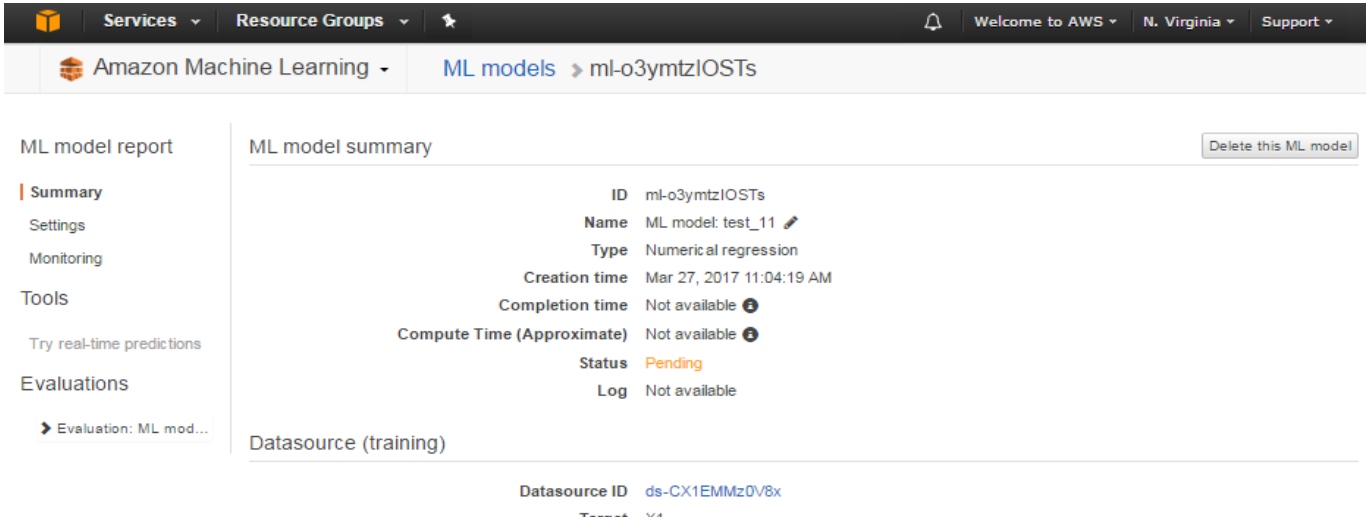
4. AWS를 이용하여 예측모형 구축(linear)



4. AWS를 이용하여 예측모형 구축(linear)



4. AWS를 이용하여 예측모형 구축(linear)



Objects

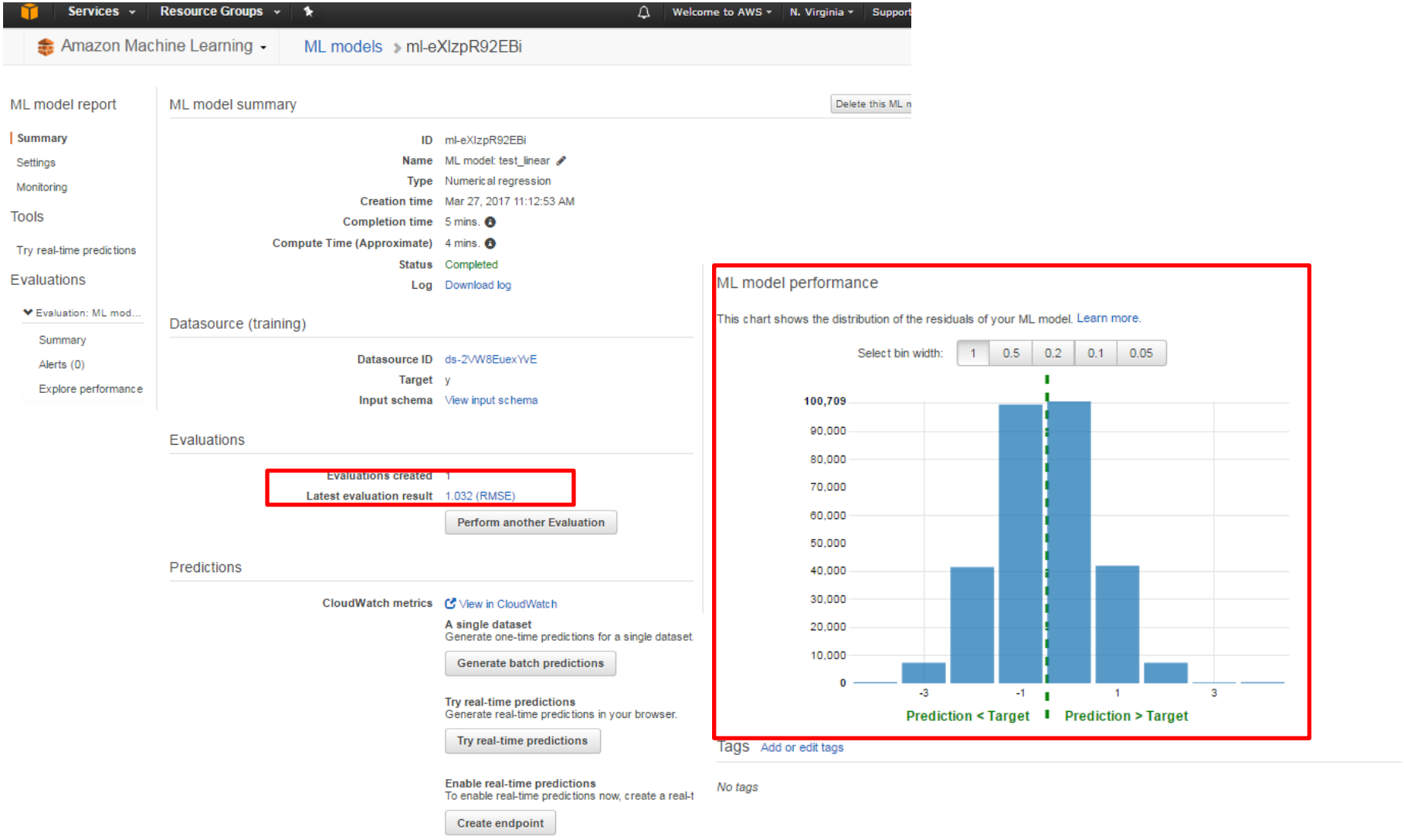
Create new... | Actions | Refresh

Filter: All types | Object name or ID | Items per page: 10 | 1 - 5 of 5 Objects

Name	Type	ID	Status	Creation time	Completion time
Evaluation: ML model: test_linear	Evaluation	ev-ncVMlt2e3DP	Pending	Mar 27, 2017 11:12:54 AM	Not available
ML model: test_linear	ML model	ml-eXlzpR92EBi	Pending	Mar 27, 2017 11:12:53 AM	Not available
test_linear_[percentBegin=0, percentEnd=70, str...	Datasource	ds-2VW8EuexYvE	In progress	Mar 27, 2017 11:12:52 AM	Not available
test_linear_[percentBegin=70, percentEnd=100, ...	Datasource	ds-EQJzZs96cxv	In progress	Mar 27, 2017 11:12:52 AM	Not available
test_linear	Datasource	ds-pTjopEnbbcL	In progress	Mar 27, 2017 11:11:01 AM	Not available

1 - 5 of 5 Objects

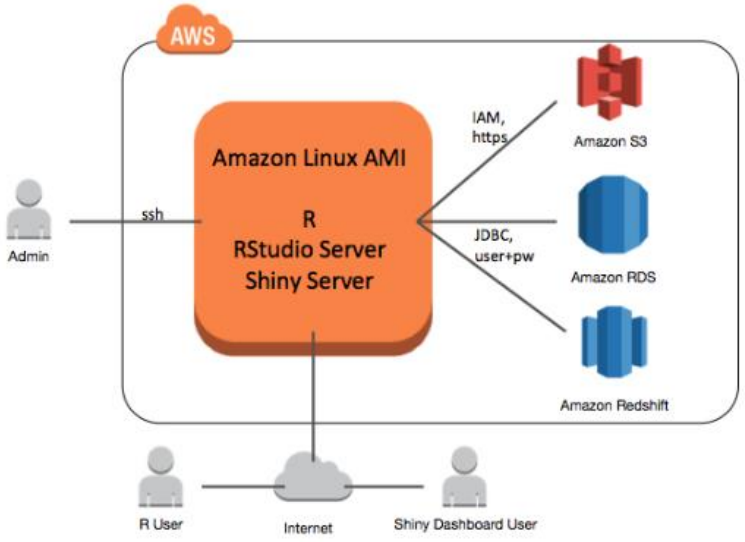
4. AWS를 이용하여 예측모형 구축(linear)



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

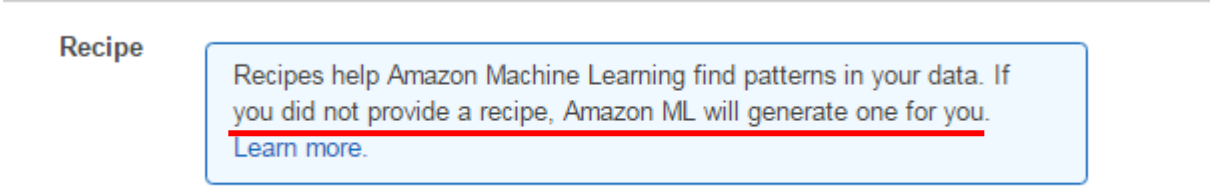
	AWS	Azure	GCE	OpenStack
Virtual network	Yes	Yes	No	Yes
API	Yes	Yes	Yes	No
Machine-learning Tools	Yes	Yes	Yes	No
Deployment Tools	Yes	Yes	Yes	No
On-premises Data-center integration	Yes	Yes	Yes	No
Number of Service	Most	Many	Enough	Few
Global Region	12 (Gov/베이징 포함)	13	3	N/A
Compliance	Common standards, IT Grundschutz(Germany), G-Cloud(UK)	Common standards, ISO27018, G-Cloud(UK)	Common Standards	Depends on the OpenStack provider
SDK	Android, Javascirpt(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 Dtatabase, 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

-



Thank you