Büyük Veri Uygulamaları – Ders 5-6

Doç. Dr. Yuriy Mishchenko

Plan

- Amazon ML hizmetini kullanmaya pratik giriş
- Baze ek gereken makine öğrenme kavramları – genelleme, eğitim ve test veri kümeleri, makine öğrenme modellerinin performans ölçekleri

AWS Amazon ML


Amazon Machine Learnin
https://aws.amazon.com/machine-learning/
Menu
amazon web services
English
My Account
Create an AWS Account
PRODUCTS & SERVICES
Amazon Machine Learning
Product Details
Pricing
Getting Started
FAQs
Resources
Get Started with AWS for Free
Create Free Account
Amazon Machine Learning
Amazon Machine Learning is a service that makes it easy for developers of all skill levels to use machine learning technology. Amazon Machine Learning provides visualization tools and wizards that guide you through the process of creating machine learning (ML) models without having to learn complex ML algorithms and technology. Once your models are ready, Amazon Machine Learning makes it easy to obtain predictions for your application using simple APIs, without having to implement custom prediction generation code, or manage any infrastructure.
Amazon Machine Learning is based on the same proven, highly scalable, ML technology used for years by Amazon's internal data
Get started with Amazon Machine Learning
Create a Free Account
Receive twelve months of access to the AWS Free Tier and enjoy AWS Basic Support features including, 24x7x365 customer service, support forums, and more.
Please note that Amazon Machine Learning is not currently available on the AWS Free Tier.

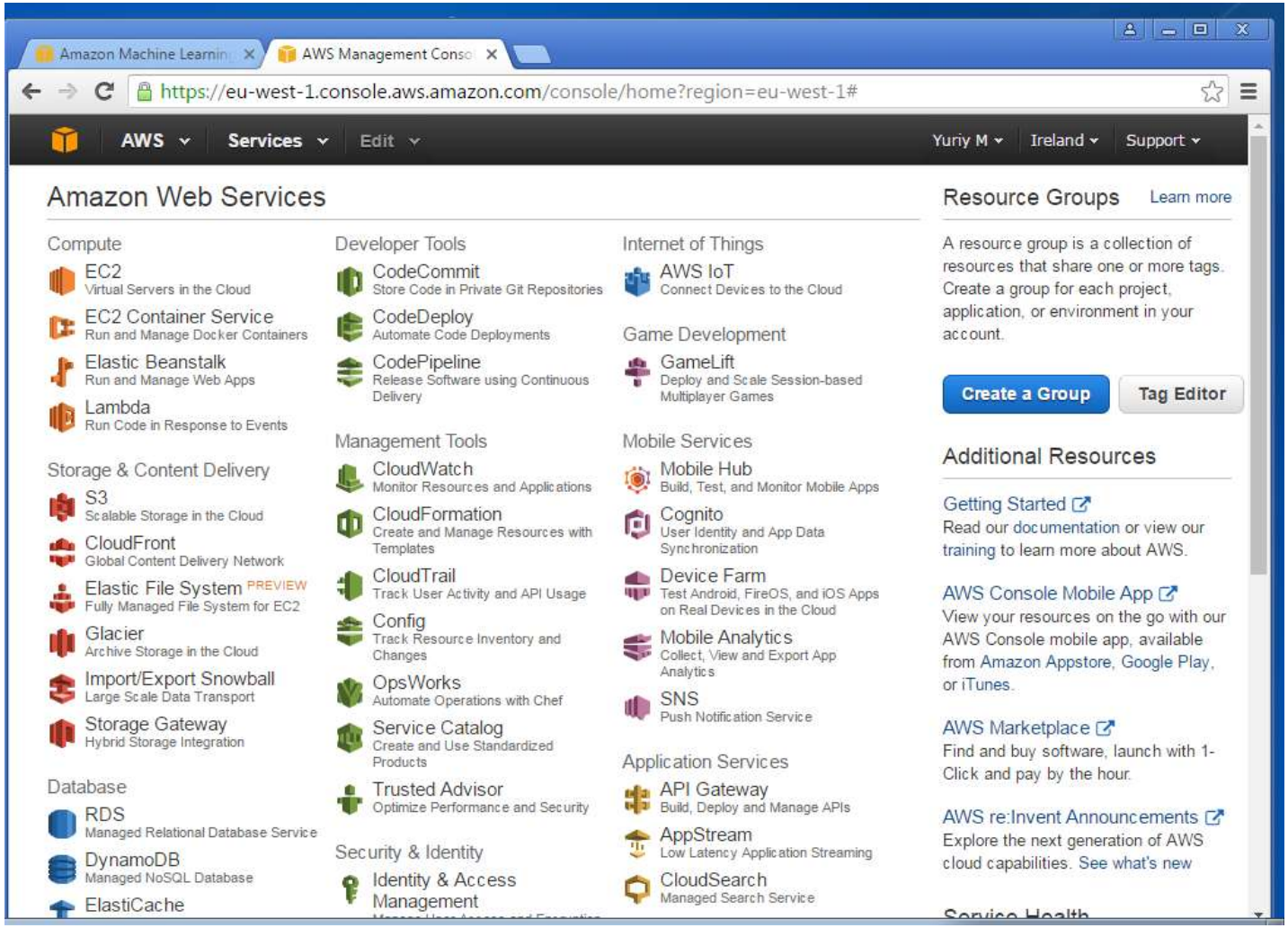
Amazon Web Services (AWS)

- AWS Amazon'ın bulut hesaplama hizmetleri
- Amazon şirketi sayesinde, ABD ve İrland'da yerleştirilen bilgisayar merkezlerinden müşterilerine sunulmaktadır
- Hosting hizmetleri (EC2), depolama hizmetleri (S3), veritabanı hizmetleri (RDS), Analytics hizmetleri (ML), IoT (İnternet of Things) hizmetleri, Mobile hizmetleri (...), Massive Multiplayer Online (MMO) bilgisayar oyunları için hizmetler (GameLift) ve sayre

Amazon Web Services (AWS)

- Makine öğrenme hizmeti (Amazon ML) AWS'nın "Analytics hizmetleri" kapsamında sunulmakta dır
- Amazon ML hizmeti, verilerinize göre ML modellerinin üretilmesi ve onlar kullanılarak offline (batch) veya online (API) tahminlerin üretilmesini sağlar
- Amazon ML hizmeti kulllanabilmek için, Amazon AWS'nın kayıtlı kullanıcı olması gerekiyor

Amazon Web Services (AWS)



Amazon Web Services (AWS)

- AWS hizmetlerinin kullanımı tam üçretsiz değil, fakat düşük maliyetle yeni kullanıcılar tarafından kullanılabilir
- AWS Free Tier (http://aws.amazon.com/free/) erişim, AWS hizmetlerinin tanıtımı olarak yeni kullanıcılar için açıktır, baze hizmetlere ücretsiz olarak 12 aylık erişim sağlar, diğer hizmetler ise aslı kullanımına bağlı ücretlendirme sistemi kullanıyor

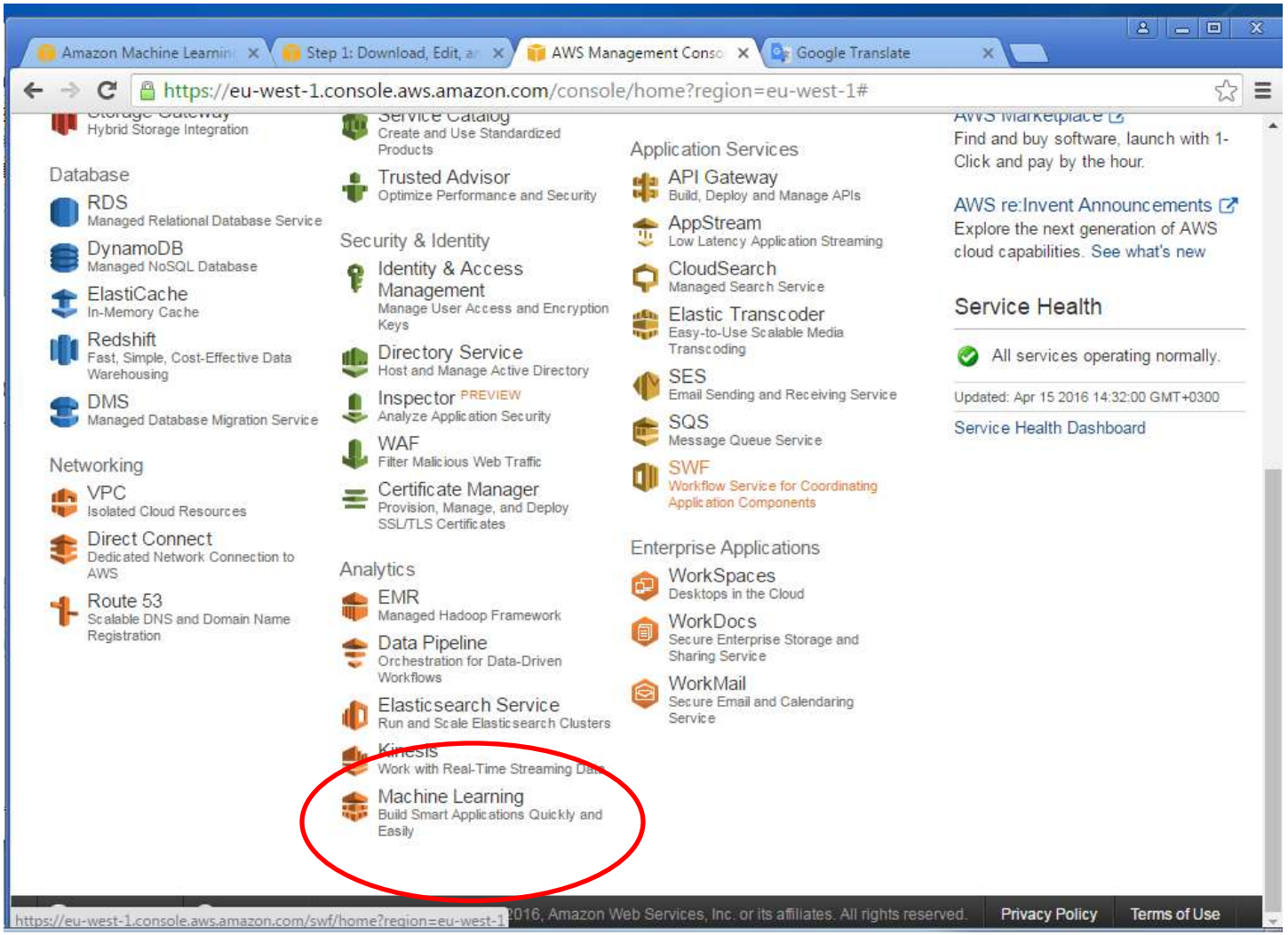
Amazon Web Services (AWS)

- Ücretsiz olarak sağlanan hizmetler (http://aws.amazon.com/free/)
  - EC2 temel hesaplama tesisi, aylık 750 saata kadar
  - S3 temel depolama tesisi, 5GB'a kadar
  - RDS veritabanı tesisi, 750 saat + 20GB'a kadar
  - IoT tesisi, 250K mesaj'a kadar
- ML tesisi, Free Tire hizmetine ait değil fakat düşük maaliyetle kullanılabilir (1 ML model ~1 USD ücrete neden olur)

Amazon ML tutorial

- Amazon ML hizmeti, AWS kontrol panelinin Analytics kısmından ulaşılabilir

AMAZON ML HIZMETI:



Amazon ML tutorial

- Amazon ML himetinde yeni modellerin oluşturulması üç adımdan oluşur
  1. Verilerin Amazon S3 depolama servisine yüklenmesi
  2. Amazon S3 bir veri dosyasından yeni ML modelin eğitilmesi
  3. Eğitilmiş modelin performansı değerlendirilmesi ve tahmin etme kullanımı için hazırlanması

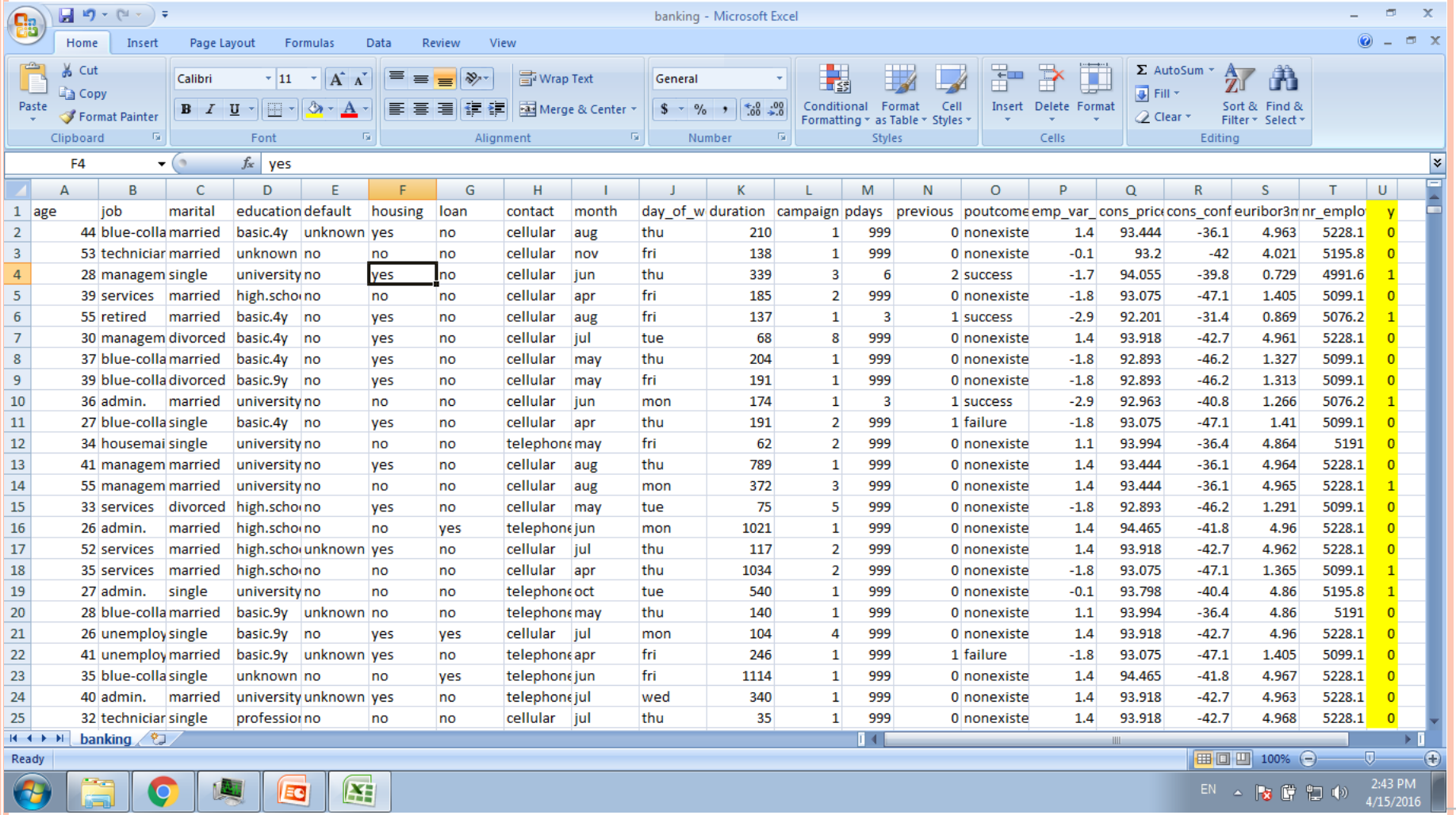
Amazon ML tutorial

- Verilerin yüklenmesi:
  - Amazon ML sadece Amazon S3 serverlerindeki veri kullanılarak yeni model üretebiliyor
  - Söz konusu tutorial için bir pazarlama (marketing) veri örneği https://s3.amazonaws.com/aml-sample-data/banking.csv adresinden indirilebilir, sonra S3 hizmetine kullanıcının adından yüklenmeli

Amazon ML tutorial

- Verilerin yüklenmesi:
  - ML modelleri için kullanılan veri, virgül kullanılarak ayıldırılmış CVS tabloları, böyle tablolar verilerinizin hepsi için Excel'de Save As/Kaydet Diğer menüsünden dosya formatı CVS olarak seçilerek oluşturulabilir
  - İlgili tabloda her satırlar verilerin örneklertir; sütünler şu örneklerin farklı özniteliklerdir; her bir satır bir tane veri örneği için ilgili özniteliklerin değerleri listeliyor
  - Öznitelikler sayısal veya metin şeklinde kaydedilebilir – genelde metin değerleri "kategorik" öznitelik olarak yorumlanacak

AMAZON ML VERI ÖRNEĞI

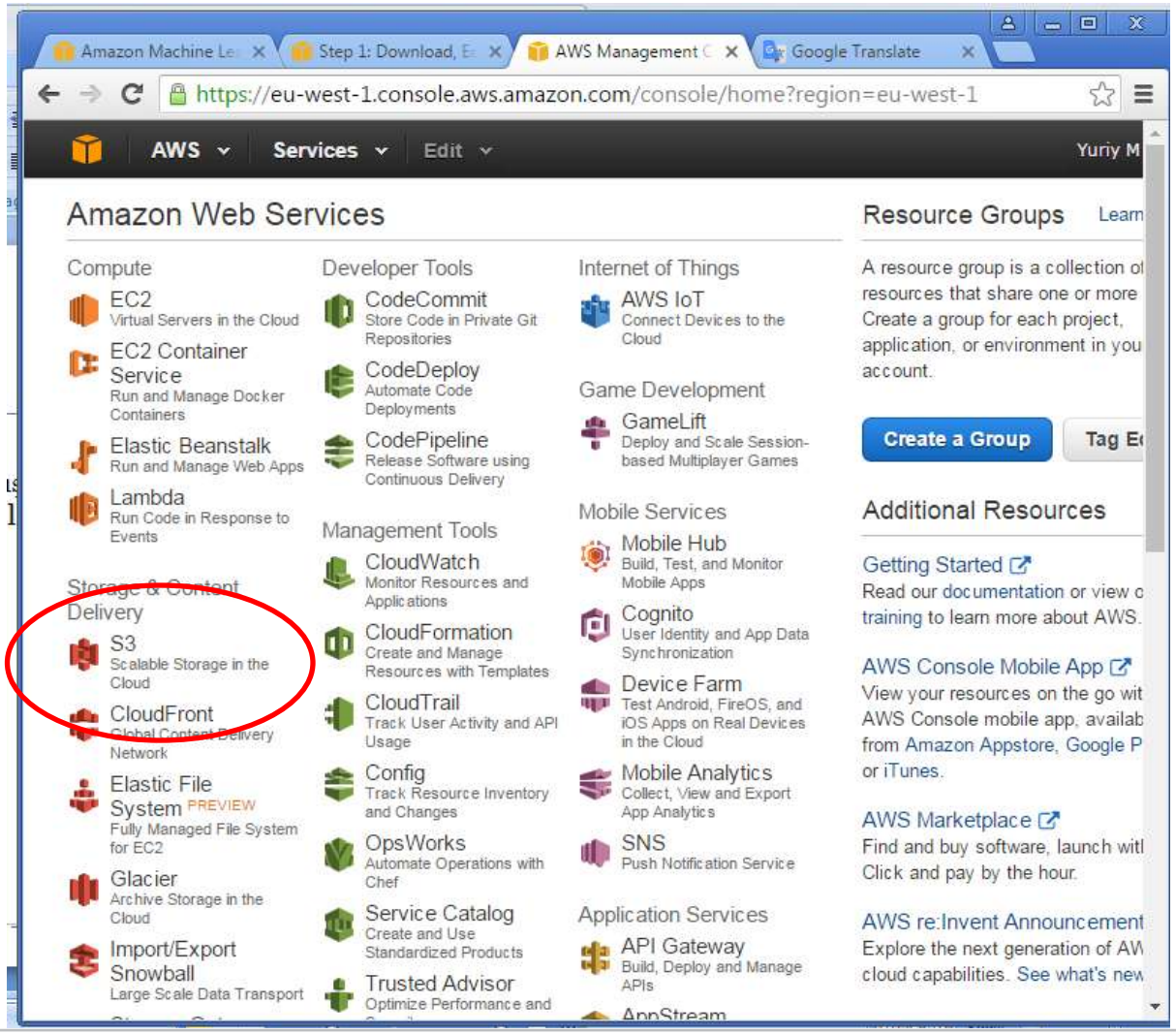


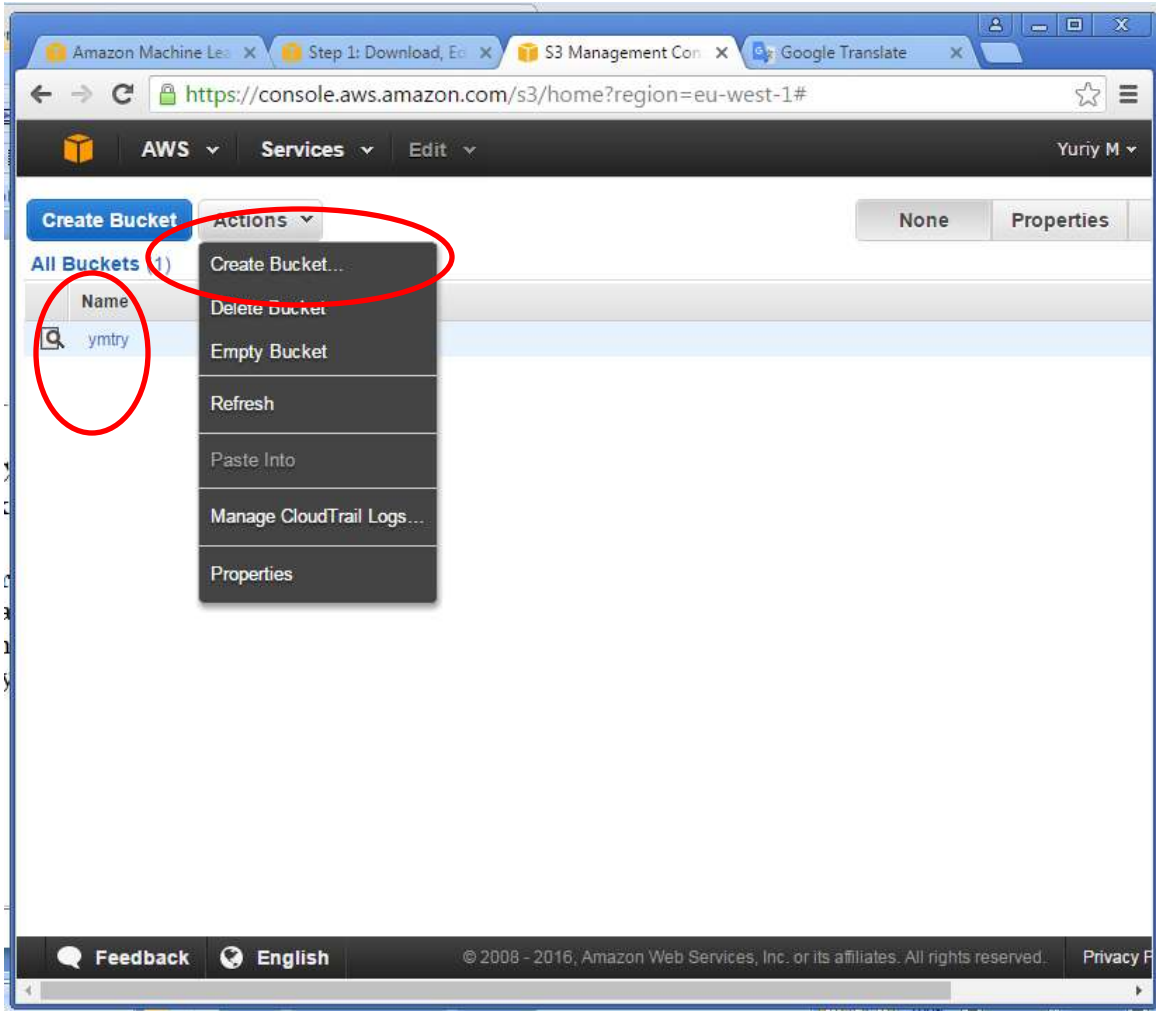
banking - Microsoft Excel

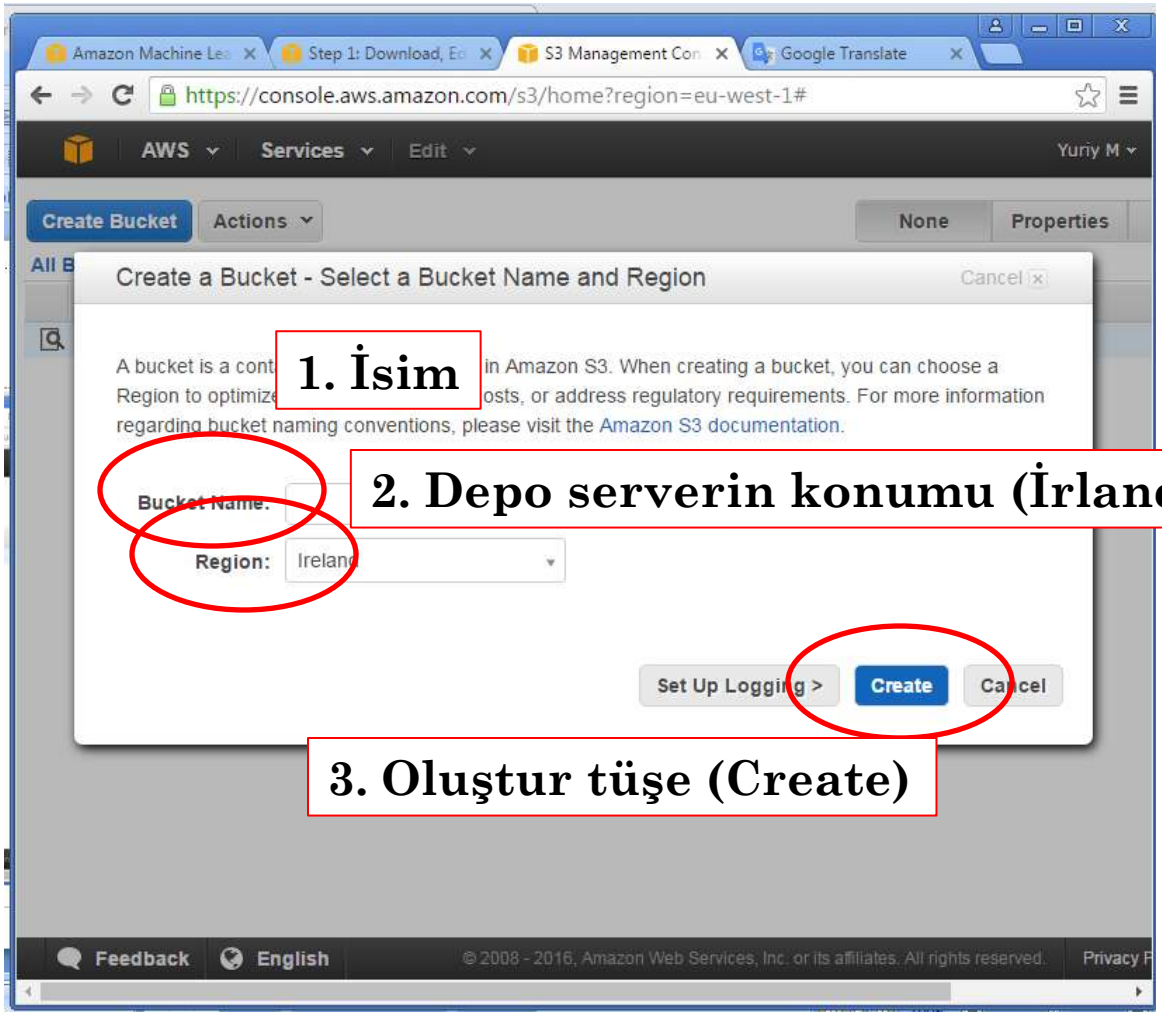
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	age	job	marital	education	default	housing	loan	contact	month	day_of_w	duration	campaign	pdays	previous	poutcome	emp_var_	cons_pric	cons_conf	euribor3m	nr_emplo	y
2	44	blue-colla	married	basic.4y	unknown	yes	no	cellular	aug	thu	210	1	999	0	nonexiste	1.4	93.444	-36.1	4.963	5228.1	0
3	53	technician	married	unknown	no	no	no	cellular	nov	fri	138	1	999	0	nonexiste	-0.1	93.2	-42	4.021	5195.8	0
4	28	managem	single	university	no	yes	no	cellular	jun	thu	339	3	6	2	success	-1.7	94.055	-39.8	0.729	4991.6	1
5	39	services	married	high.scho	no	no	no	cellular	apr	fri	185	2	999	0	nonexiste	-1.8	93.075	-47.1	1.405	5099.1	0
6	55	retired	married	basic.4y	no	yes	no	cellular	aug	fri	137	1	3	1	success	-2.9	92.201	-31.4	0.869	5076.2	1
7	30	managem	divorced	basic.4y	no	yes	no	cellular	jul	tue	68	8	999	0	nonexiste	1.4	93.918	-42.7	4.961	5228.1	0
8	37	blue-colla	married	basic.4y	no	yes	no	cellular	may	thu	204	1	999	0	nonexiste	-1.8	92.893	-46.2	1.327	5099.1	0
9	39	blue-colla	divorced	basic.9y	no	yes	no	cellular	may	fri	191	1	999	0	nonexiste	-1.8	92.893	-46.2	1.313	5099.1	0
10	36	admin.	married	university	no	no	no	cellular	jun	mon	174	1	3	1	success	-2.9	92.963	-40.8	1.266	5076.2	1
11	27	blue-colla	single	basic.4y	no	yes	no	cellular	apr	thu	191	2	999	1	failure	-1.8	93.075	-47.1	1.41	5099.1	0
12	34	housemai	single	university	no	no	no	telephone	may	fri	62	2	999	0	nonexiste	1.1	93.994	-36.4	4.864	5191	0
13	41	managem	married	university	no	yes	no	cellular	aug	thu	789	1	999	0	nonexiste	1.4	93.444	-36.1	4.964	5228.1	0
14	55	managem	married	university	no	no	no	cellular	aug	mon	372	3	999	0	nonexiste	1.4	93.444	-36.1	4.965	5228.1	1
15	33	services	divorced	high.scho	no	yes	no	cellular	may	tue	75	5	999	0	nonexiste	-1.8	92.893	-46.2	1.291	5099.1	0
16	26	admin.	married	high.scho	no	no	yes	telephone	jun	mon	1021	1	999	0	nonexiste	1.4	94.465	-41.8	4.96	5228.1	0
17	52	services	married	high.scho	unknown	yes	no	cellular	jul	thu	117	2	999	0	nonexiste	1.4	93.918	-42.7	4.962	5228.1	0
18	35	services	married	high.scho	no	no	no	cellular	apr	thu	1034	2	999	0	nonexiste	-1.8	93.075	-47.1	1.365	5099.1	1
19	27	admin.	single	university	no	no	no	telephone	oct	tue	540	1	999	0	nonexiste	-0.1	93.798	-40.4	4.86	5195.8	1
20	28	blue-colla	married	basic.9y	unknown	no	no	telephone	may	thu	140	1	999	0	nonexiste	1.1	93.994	-36.4	4.86	5191	0
21	26	unemploy	single	basic.9y	no	yes	yes	cellular	jul	mon	104	4	999	0	nonexiste	1.4	93.918	-42.7	4.96	5228.1	0
22	41	unemploy	married	basic.9y	unknown	yes	no	telephone	apr	fri	246	1	999	1	failure	-1.8	93.075	-47.1	1.405	5099.1	0
23	35	blue-colla	single	unknown	no	no	yes	telephone	jun	fri	1114	1	999	0	nonexiste	1.4	94.465	-41.8	4.967	5228.1	0
24	40	admin.	married	university	unknown	yes	no	telephone	jul	wed	340	1	999	0	nonexiste	1.4	93.918	-42.7	4.963	5228.1	0
25	32	technician	single	professio	no	no	no	cellular	jul	thu	35	1	999	0	nonexiste	1.4	93.918	-42.7	4.968	5228.1	0

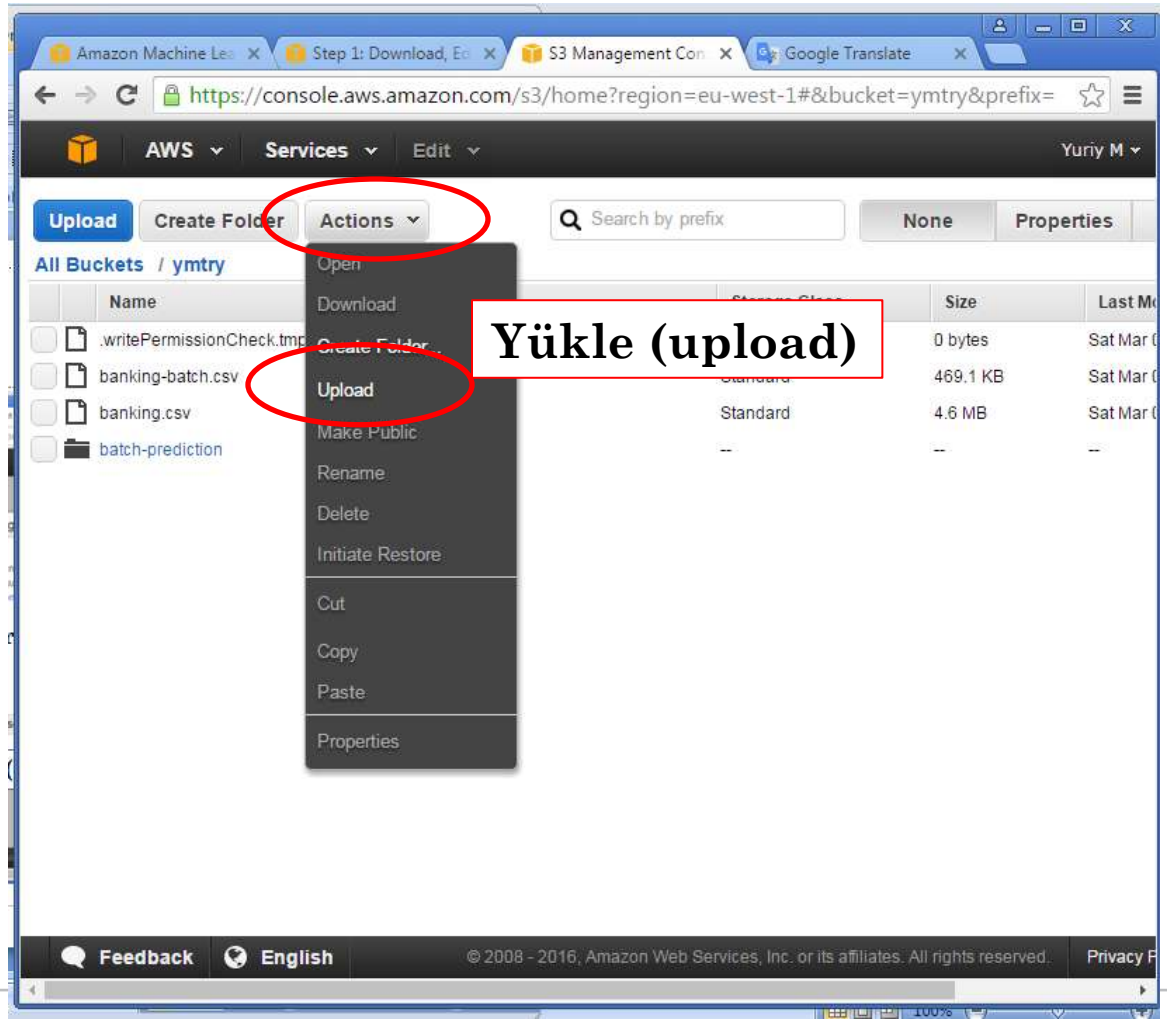
Amazon ML tutorial

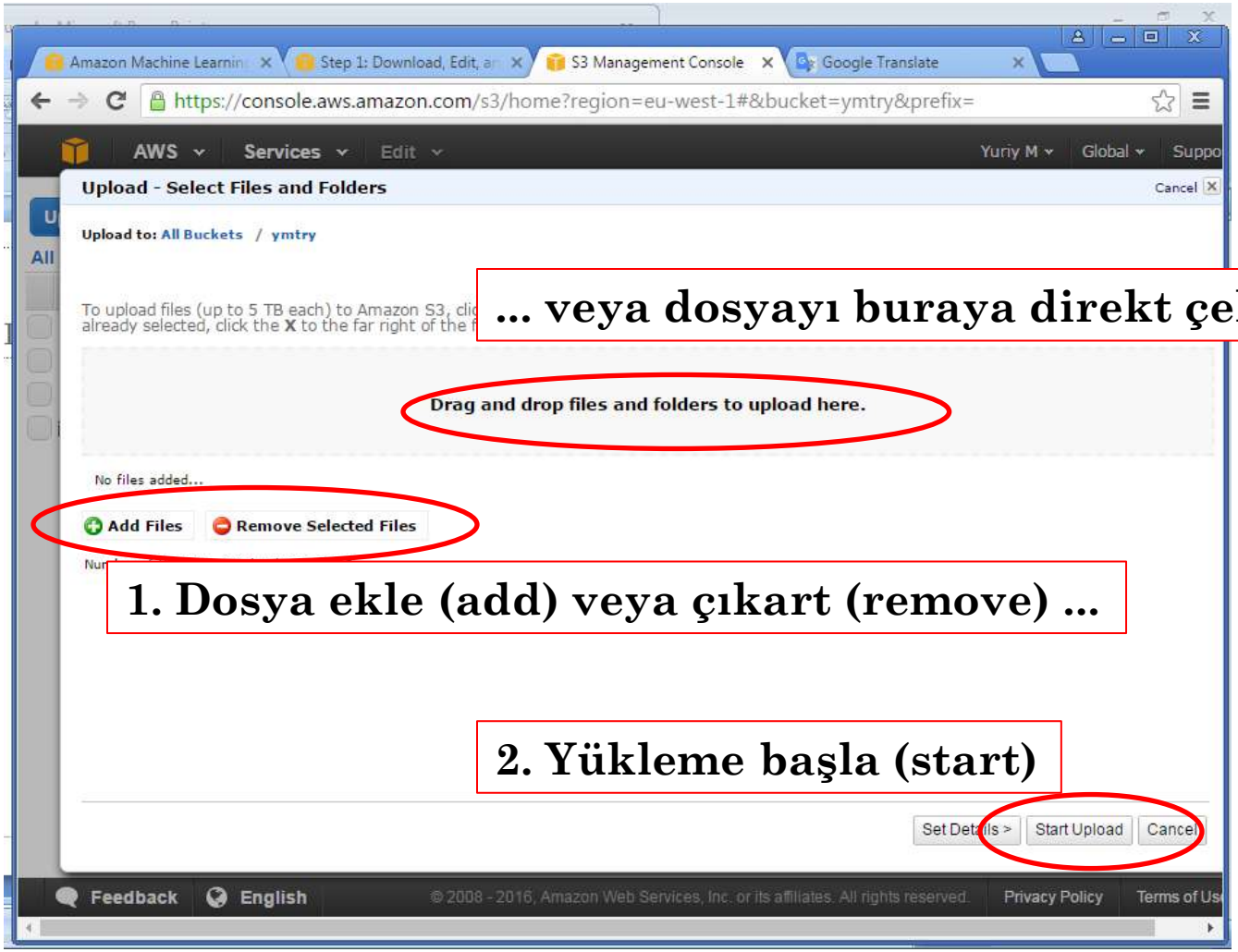
- Verilerin yüklenmesi:
  - Gereken şekilde hazırlanmış CVS dosyası Amazona yüklemek için, önceden AWS kontrol panelinden S3 hizmeti seçilmeli
  - S3'deki veri "bucket" (kova) lere organize edilir, bunlar aşağı yukarı normal klasörlere denk gelir; buna göre S3 kontrol panelinin Bucket kısmından yeni bir kova oluşturulmalı veya var olan bir kova seçilmeli

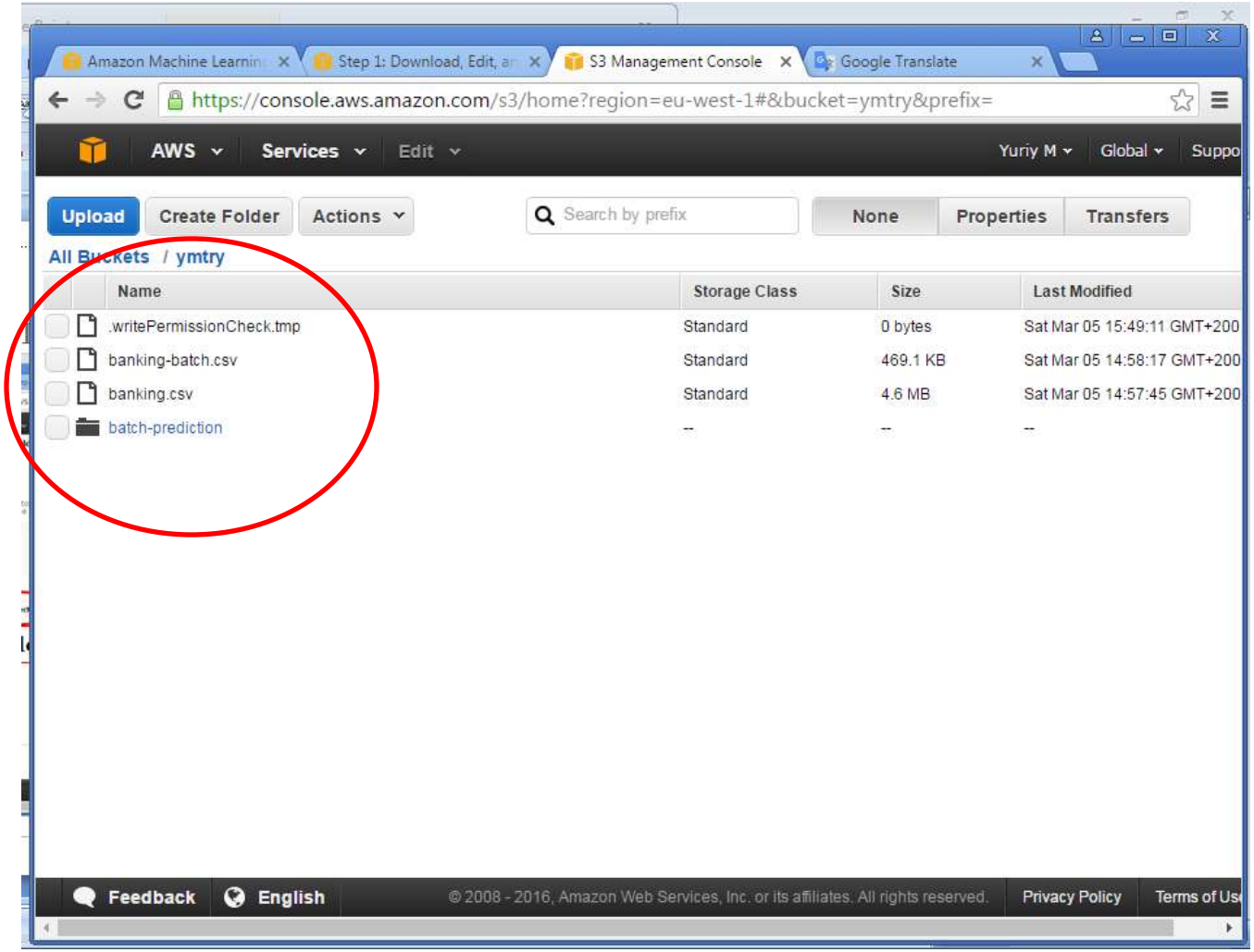

Amazon Machine Le
Step 1: Download, E
AWS Management C
Google Translate
https://eu-west-1.console.aws.amazon.com/console/home?region=eu-west-1
AWS
Services
Edit
Yuriy M
Amazon Web Services
Compute
EC2
Virtual Servers in the Cloud
EC2 Container Service
Run and Manage Docker Containers
Elastic Beanstalk
Run and Manage Web Apps
Lambda
Run Code in Response to Events
Storage & Content Delivery
S3
Scalable Storage in the Cloud
CloudFront
Global Content Delivery Network
Elastic File System PREVIEW
Fully Managed File System for EC2
Glacier
Archive Storage in the Cloud
Import/Export Snowball
Large Scale Data Transport
Developer Tools
CodeCommit
Store Code in Private Git Repositories
CodeDeploy
Automate Code Deployments
CodePipeline
Release Software using Continuous Delivery
Management Tools
CloudWatch
Monitor Resources and Applications
CloudFormation
Create and Manage Resources with Templates
CloudTrail
Track User Activity and API Usage
Config
Track Resource Inventory and Changes
OpsWorks
Automate Operations with Chef
Service Catalog
Create and Use Standardized Products
Trusted Advisor
Optimize Performance and
Internet of Things
AWS IoT
Connect Devices to the Cloud
Game Development
GameLift
Deploy and Scale Session-based Multiplayer Games
Mobile Services
Mobile Hub
Build, Test, and Monitor Mobile Apps
Cognito
User Identity and App Data Synchronization
Device Farm
Test Android, FireOS, and iOS Apps on Real Devices in the Cloud
Mobile Analytics
Collect, View and Export App Analytics
SNS
Push Notification Service
Application Services
API Gateway
Build, Deploy and Manage APIs
AppStream
Resource Groups
Learn
A resource group is a collection of resources that share one or more
Create a group for each project, application, or environment in you
account.
Create a Group
Tag E
Additional Resources
Getting Started
Read our documentation or view c
training to learn more about AWS.
AWS Console Mobile App
View your resources on the go wit
AWS Console mobile app, availab
from Amazon Appstore, Google P
or iTunes.
AWS Marketplace
Find and buy software, launch witl
Click and pay by the hour.
AWS re:Invent Announcement
Explore the next generation of AW
cloud capabilities. See what's new


Amazon Machine Le
Step 1: Download, Ed
S3 Management Con
Google Translate
https://console.aws.amazon.com/s3/home?region=eu-west-1#
AWS
Services
Edit
Yuriy M
Create Bucket
Actions
None
Properties
All Buckets (1)
Name
ymtry
Create Bucket...
Delete Bucket
Empty Bucket
Refresh
Paste Into
Manage CloudTrail Logs...
Properties
Feedback
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy


Amazon Machine Le
Step 1: Download, Ec
S3 Management Con
Google Translate
https://console.aws.amazon.com/s3/home?region=eu-west-1#
AWS
Services
Edit
Yuriy M
Create Bucket
Actions
None
Properties
All B
Create a Bucket - Select a Bucket Name and Region
Cancel
A bucket is a cont
in Amazon S3. When creating a bucket, you can choose a
Region to optimize
osts, or address regulatory requirements. For more information
regarding bucket naming conventions, please visit the Amazon S3 documentation.
1. İsim
Bucket Name:
Region:
Irelan
2. Depo serverin konumu (İrlanda veya ABD)
Set Up Logging >
Create
Cancel
3. Oluştur tüşe (Create)
Feedback
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy


Amazon Machine Le
Step 1: Download, Ec
S3 Management Con
Google Translate
https://console.aws.amazon.com/s3/home?region=eu-west-1#&bucket=ymtry&prefix=
AWS
Services
Edit
Yuriy M
Upload
Create Folder
Actions
Search by prefix
None
Properties
All Buckets / ymtry
Name
Size
Last M
.writePermissionCheck.tmp
0 bytes
Sat Mar
banking-batch.csv
469.1 KB
Sat Mar
banking.csv
Standard
4.6 MB
Sat Mar
batch-prediction
--
--
--
Open
Download
Upload
Make Public
Rename
Delete
Initiate Restore
Cut
Copy
Paste
Properties
Yükle (upload)
Feedback
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy


Amazon Machine Learnin
Step 1: Download, Edit, a
S3 Management Console
Google Translate
https://console.aws.amazon.com/s3/home?region=eu-west-1#&bucket=ymtry&prefix=
AWS
Services
Edit
Yuriy M
Global
Suppo
Upload - Select Files and Folders
Cancel
Upload to: All Buckets / ymtry
To upload files (up to 5 TB each) to Amazon S3, cli
already selected, click the X to the far right of the f
... veya dosyayı buraya direkt çek
Drag and drop files and folders to upload here.
No files added...
Add Files
Remove Selected Files
1. Dosya ekle (add) veya çıkart (remove) ...
2. Yükleme başla (start)
Set Details >
Start Upload
Cancel
Feedback
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Us


Amazon Machine Learnin
Step 1: Download, Edit, a
S3 Management Console
Google Translate
https://console.aws.amazon.com/s3/home?region=eu-west-1#&bucket=ymtry&prefix=
AWS
Services
Edit
Yuriy M
Global
Suppo
Upload
Create Folder
Actions
Search by prefix
None
Properties
Transfers
All Buckets / ymtry
Name
Storage Class
Size
Last Modified
.writePermissionCheck.tmp
Standard
0 bytes
Sat Mar 05 15:49:11 GMT+200
banking-batch.csv
Standard
469.1 KB
Sat Mar 05 14:58:17 GMT+200
banking.csv
Standard
4.6 MB
Sat Mar 05 14:57:45 GMT+200
batch-prediction
--
--
--
Feedback
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Us

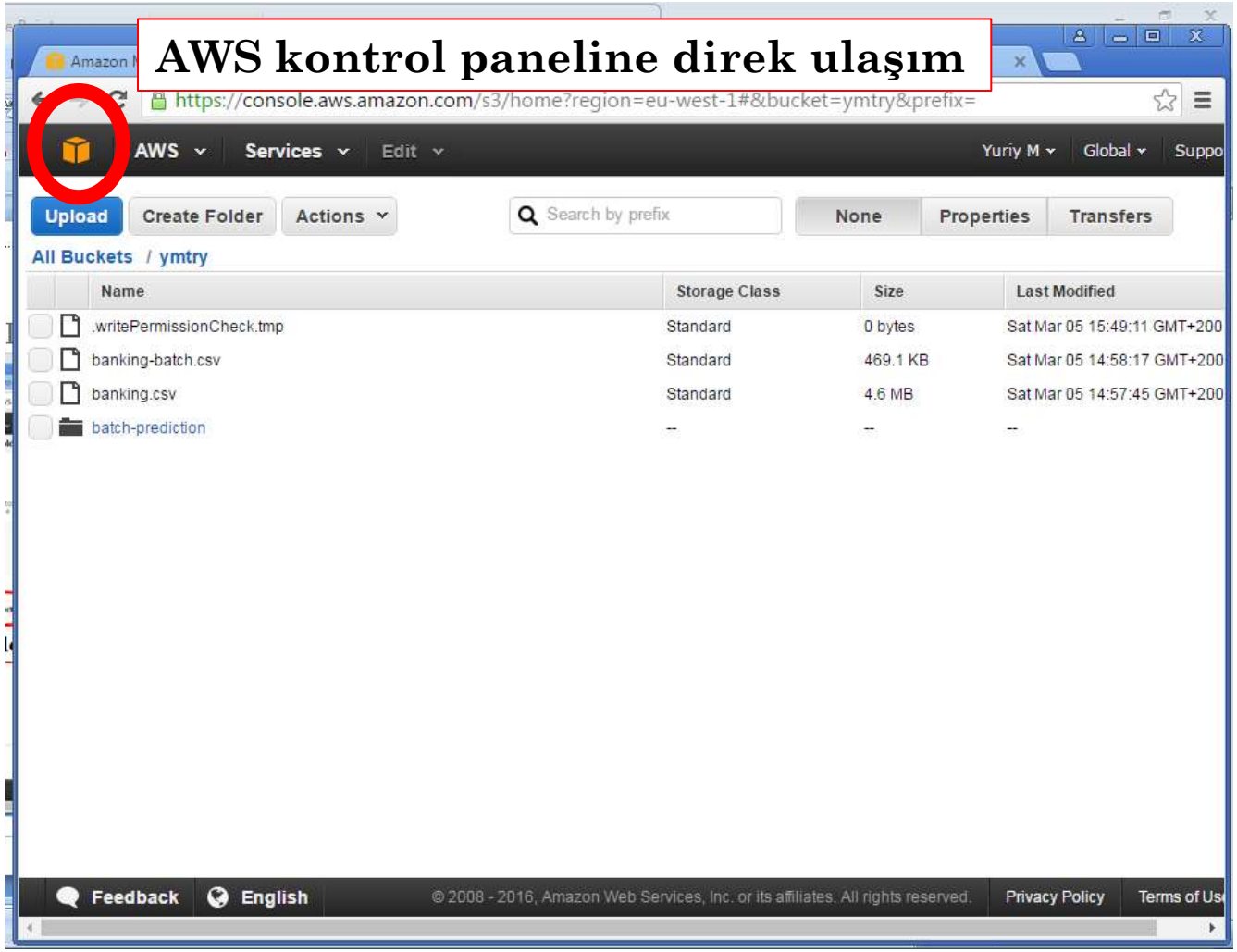
Amazon ML tutorial

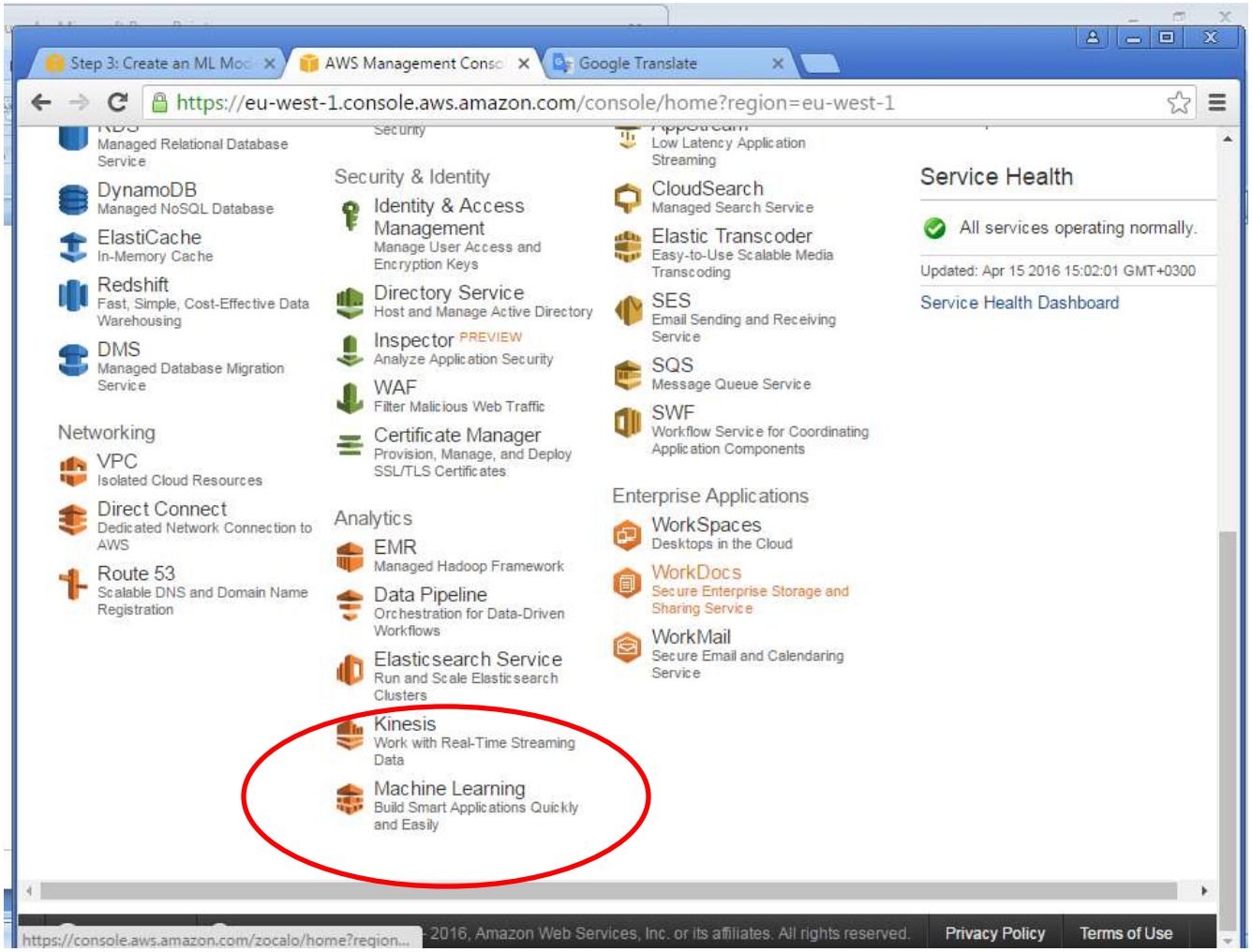
- Belli veri için ML modelinin üretilmesi:
  - Bu süreç kapsamında öncelikle S3'e yüklenmiş verilerden bir "data source" (veri kaynak) oluşturulması gerekiyor – data source diğer hizmetler için dosyanızdan veri sağlar
    (not: data source gerçekten verilerinizin bir kopyası değil, sadece S3'te zaten var olan dosyanıza ulaşım bir metodu veya daha doğru ifade ile arayüzü)
  - Oluşturulmuş bir veri kaynaktan ML model eğitiliyor
  - ML modelin performans raporu hazırlanıyor

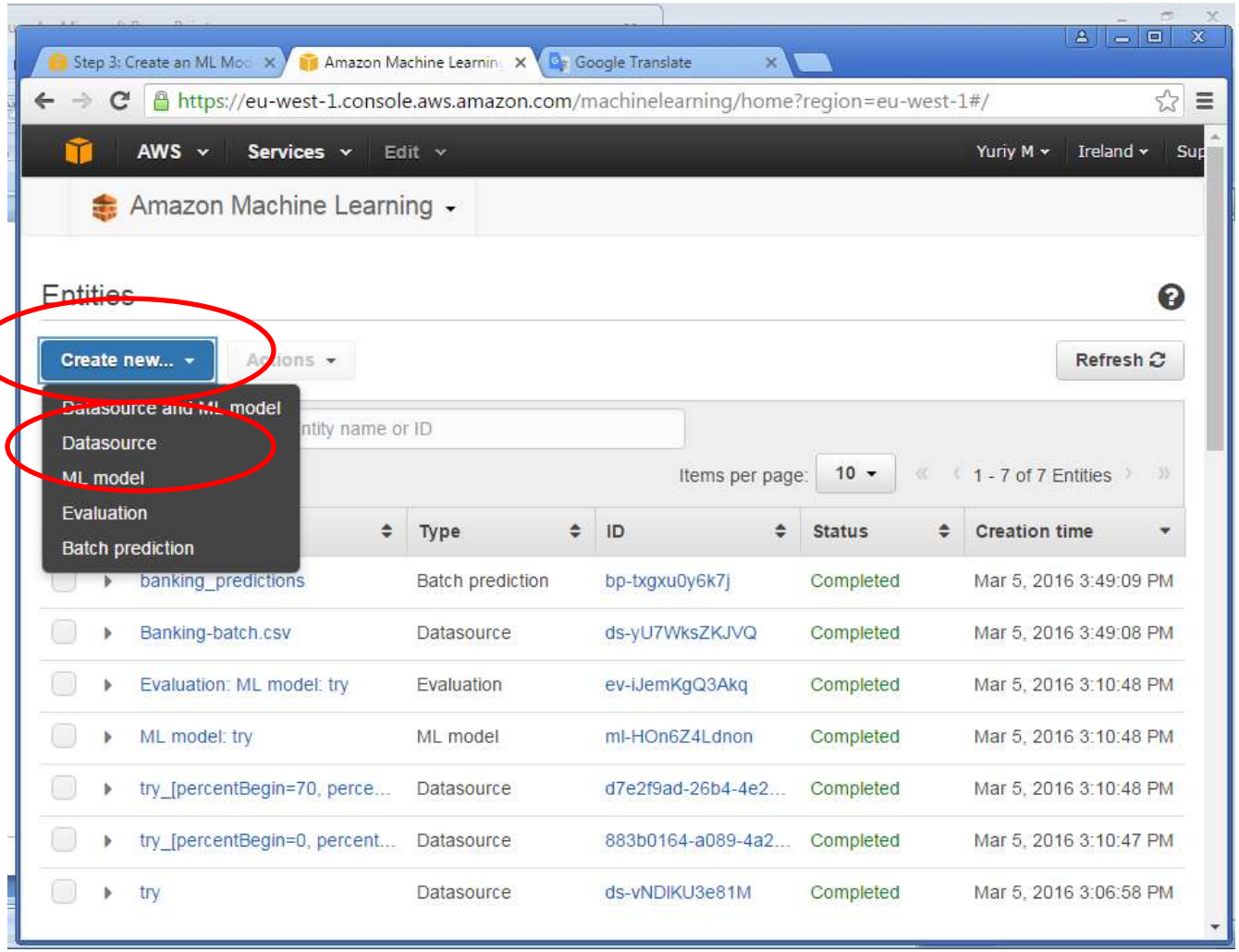
Amazon ML tutorial

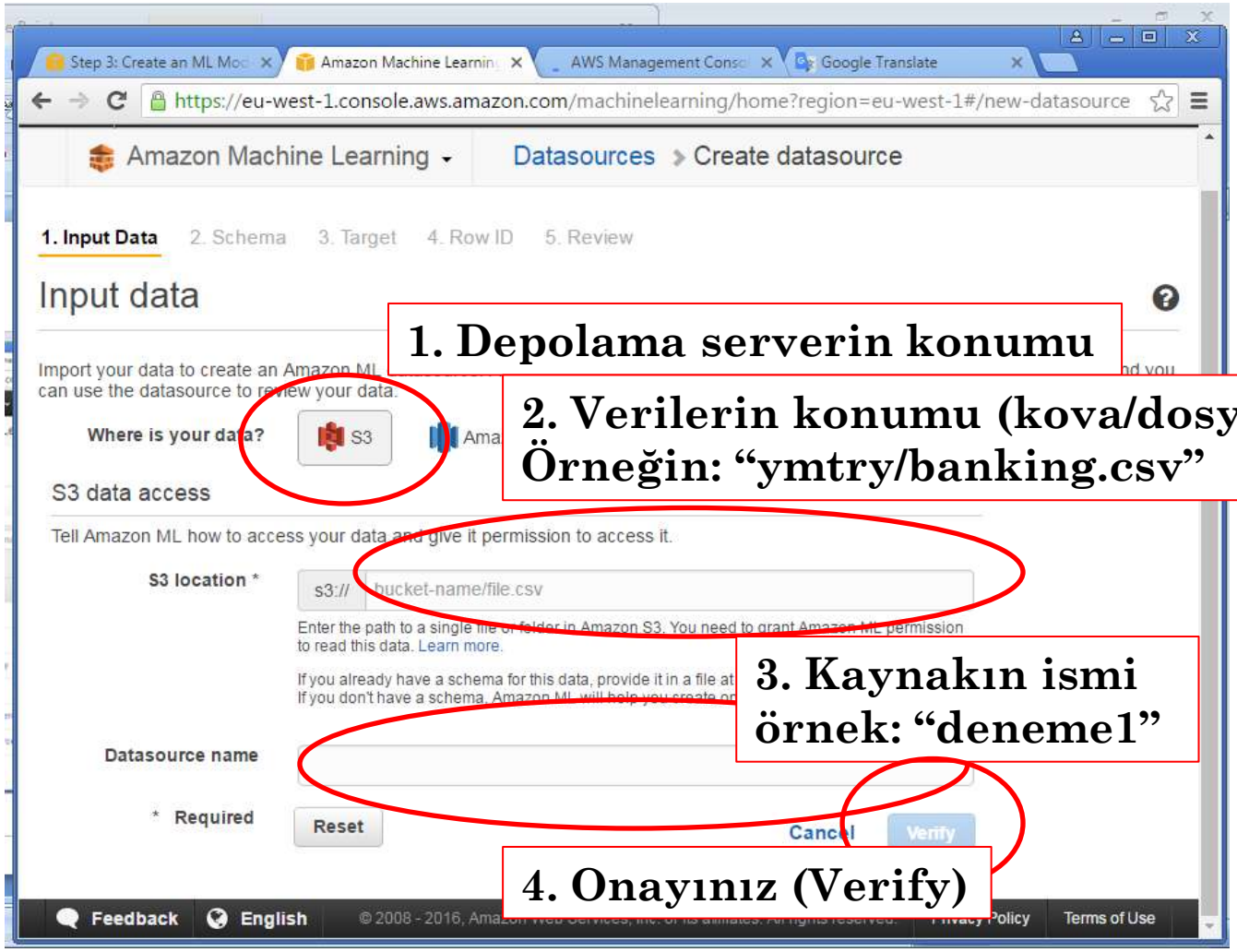
- S3'e yüklediğiniz veri için yeni bir veri kaynak oluşturmak için, AWS kontrol panelinden direk ML hizmetine gitmelisiniz
- Kaynağınız için S3'deki dosyası belirttikten sonra, bu dosya kontrol edilir ve veri kaynağa bağlanır
- Kaynak doğrulunca içindeki verilerin Amazon ML tarafından otomatik olarak analiz edilir ve özniteliklerin tipi tanımlanır

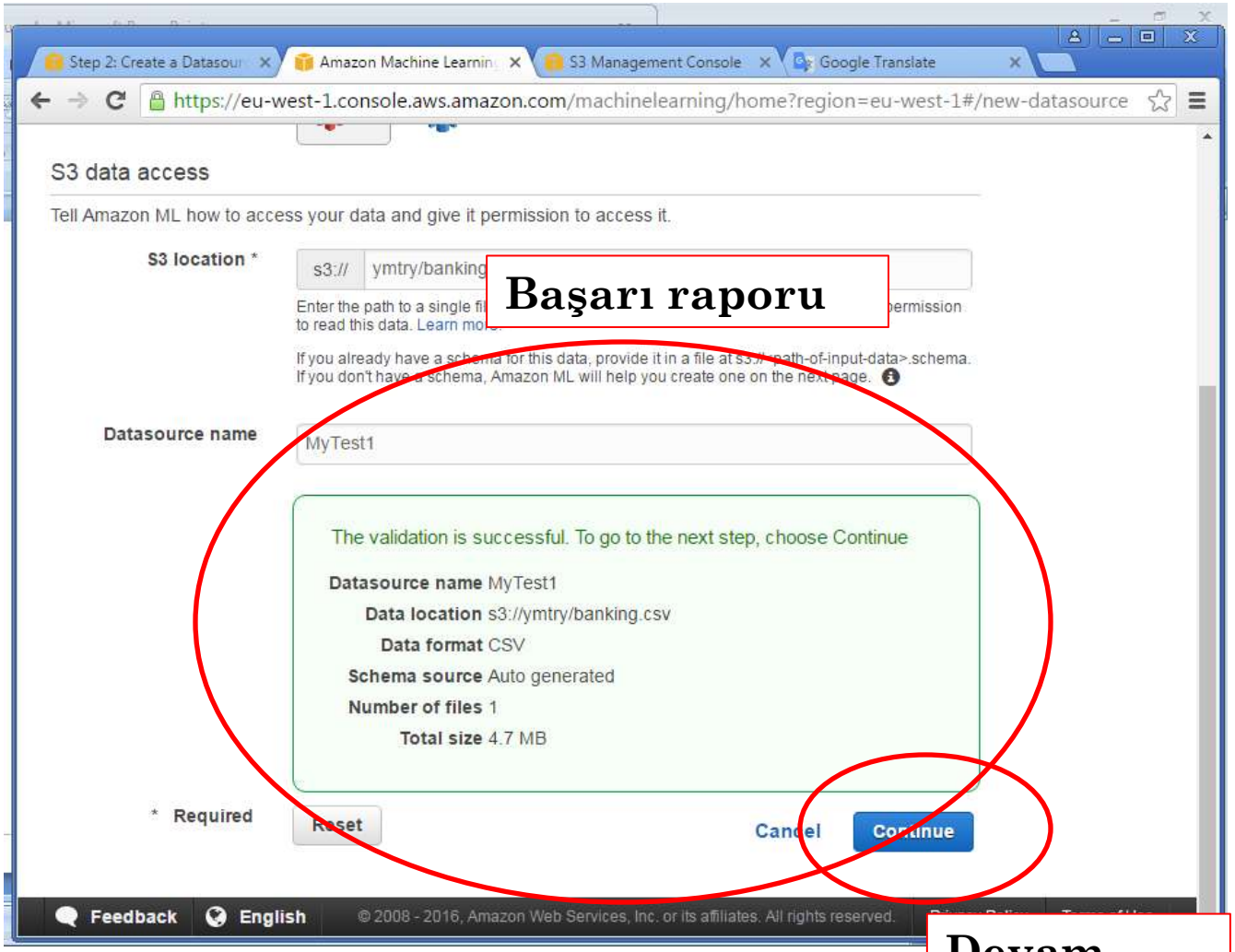
AWS kontrol paneline direk ulaşım



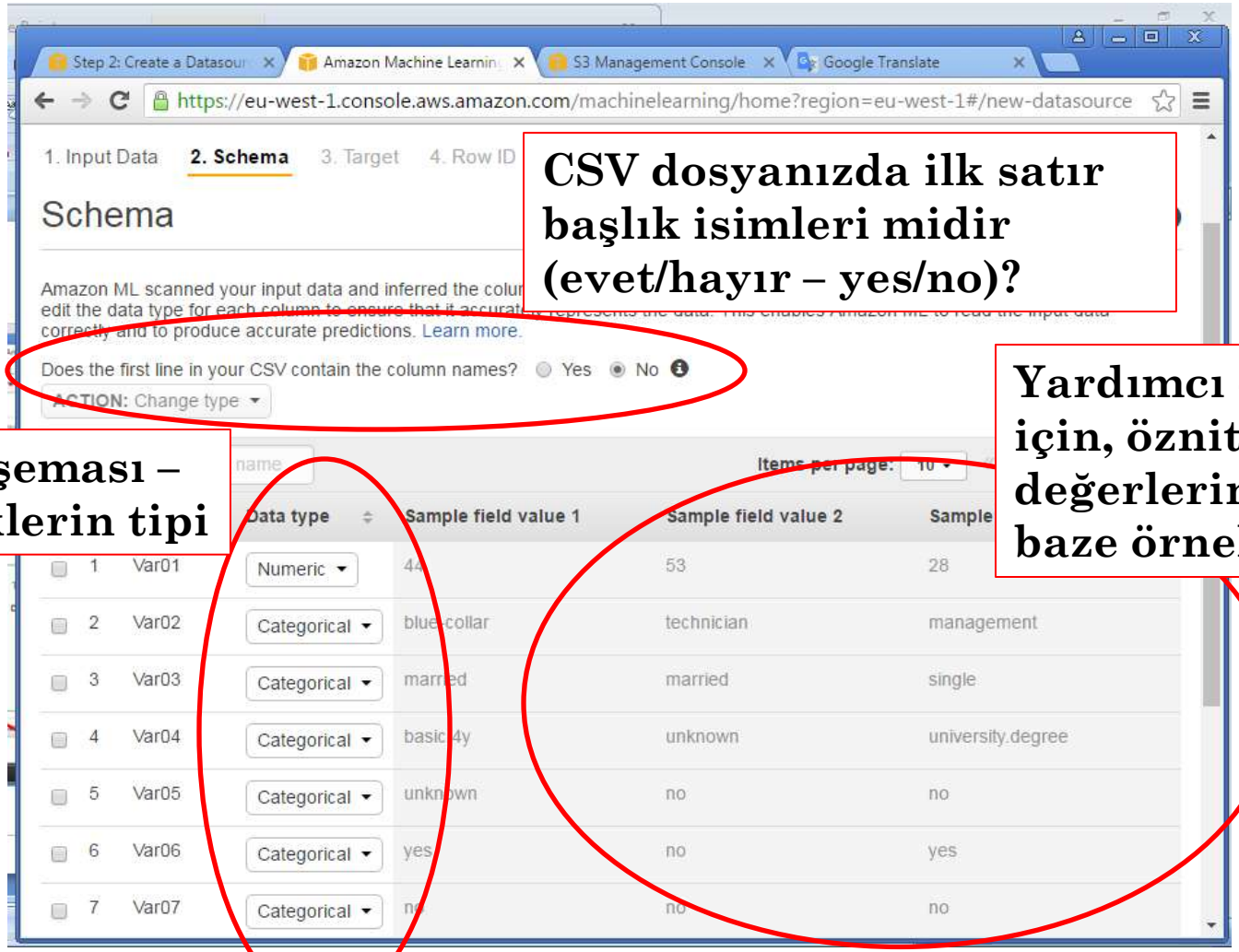

Step 3: Create an ML Mod
AWS Management Conso
Google Translate
https://eu-west-1.console.aws.amazon.com/console/home?region=eu-west-1
Managed Relational Database Service
DynamoDB
Managed NoSQL Database
ElastiCache
In-Memory Cache
Redshift
Fast, Simple, Cost-Effective Data Warehousing
DMS
Managed Database Migration Service
Networking
VPC
Isolated Cloud Resources
Direct Connect
Dedicated Network Connection to AWS
Route 53
Scalable DNS and Domain Name Registration
Security & Identity
Identity & Access Management
Manage User Access and Encryption Keys
Directory Service
Host and Manage Active Directory
Inspector PREVIEW
Analyze Application Security
WAF
Filter Malicious Web Traffic
Certificate Manager
Provision, Manage, and Deploy SSL/TLS Certificates
Analytics
EMR
Managed Hadoop Framework
Data Pipeline
Orchestration for Data-Driven Workflows
Elasticsearch Service
Run and Scale Elasticsearch Clusters
Kinesis
Work with Real-Time Streaming Data
Machine Learning
Build Smart Applications Quickly and Easily
Low Latency Application Streaming
CloudSearch
Managed Search Service
Elastic Transcoder
Easy-to-Use Scalable Media Transcoding
SES
Email Sending and Receiving Service
SQS
Message Queue Service
SWF
Workflow Service for Coordinating Application Components
Enterprise Applications
WorkSpaces
Desktops in the Cloud
WorkDocs
Secure Enterprise Storage and Sharing Service
WorkMail
Secure Email and Calendaring Service
Service Health
All services operating normally.
Updated: Apr 15 2016 15:02:01 GMT+0300
Service Health Dashboard
2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
https://console.aws.amazon.com/zocalo/home?region...

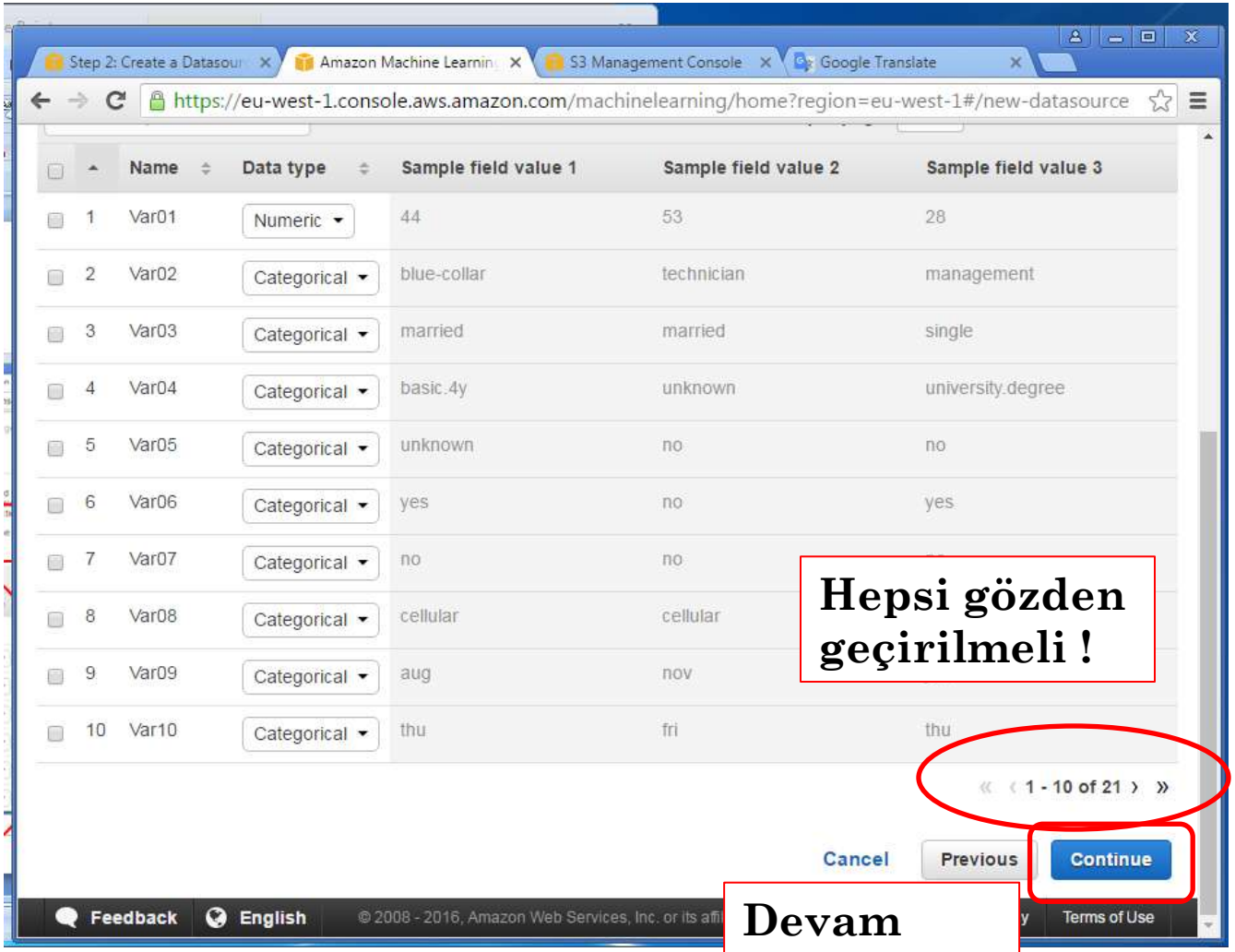

Step 3: Create an ML Mod
Amazon Machine Learnin
Google Translate
https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/
AWS
Services
Edit
Yuriy M
Ireland
Amazon Machine Learning
Entities
Create new...
Actions
Refresh
Datasource and ML model
Datasource
ML model
Evaluation
Batch prediction
ntity name or ID
Items per page: 10
1 - 7 of 7 Entities
Type
ID
Status
Creation time
banking_predictions
Batch prediction
bp-txgxu0y6k7j
Completed
Mar 5, 2016 3:49:09 PM
Banking-batch.csv
Datasource
ds-yU7WksZKJVQ
Completed
Mar 5, 2016 3:49:08 PM
Evaluation: ML model: try
Evaluation
ev-iJemKgQ3Akq
Completed
Mar 5, 2016 3:10:48 PM
ML model: try
ML model
ml-HOn6Z4Ldnon
Completed
Mar 5, 2016 3:10:48 PM
try_[percentBegin=70, perce...
Datasource
d7e2f9ad-26b4-4e2...
Completed
Mar 5, 2016 3:10:48 PM
try_[percentBegin=0, percent...
Datasource
883b0164-a089-4a2...
Completed
Mar 5, 2016 3:10:47 PM
try
Datasource
ds-vNDlKU3e81M
Completed
Mar 5, 2016 3:06:58 PM


Step 3: Create an ML Mod
Amazon Machine Learnin
AWS Management Conso
Google Translate
https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/new-datasource
Amazon Machine Learning
Datasources
Create datasource
1. Input Data
2. Schema
3. Target
4. Row ID
5. Review
Input data
Import your data to create an Amazon ML
can use the datasource to review your data.
Where is your data?
S3
Ama
S3 data access
Tell Amazon ML how to access your data and give it permission to access it.
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
If you don't have a schema, Amazon ML will help you create or
Datasource name
* Required
Reset
Cancel
Verify
Feedback
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
1. Depolama serverin konumu
2. Verilerin konumu (kova/dosya-ismi)
Örneğin: “ymtry/banking.csv”
3. Kaynakın ismi
örnek: “deneme1”
4. Onayınız (Verify)



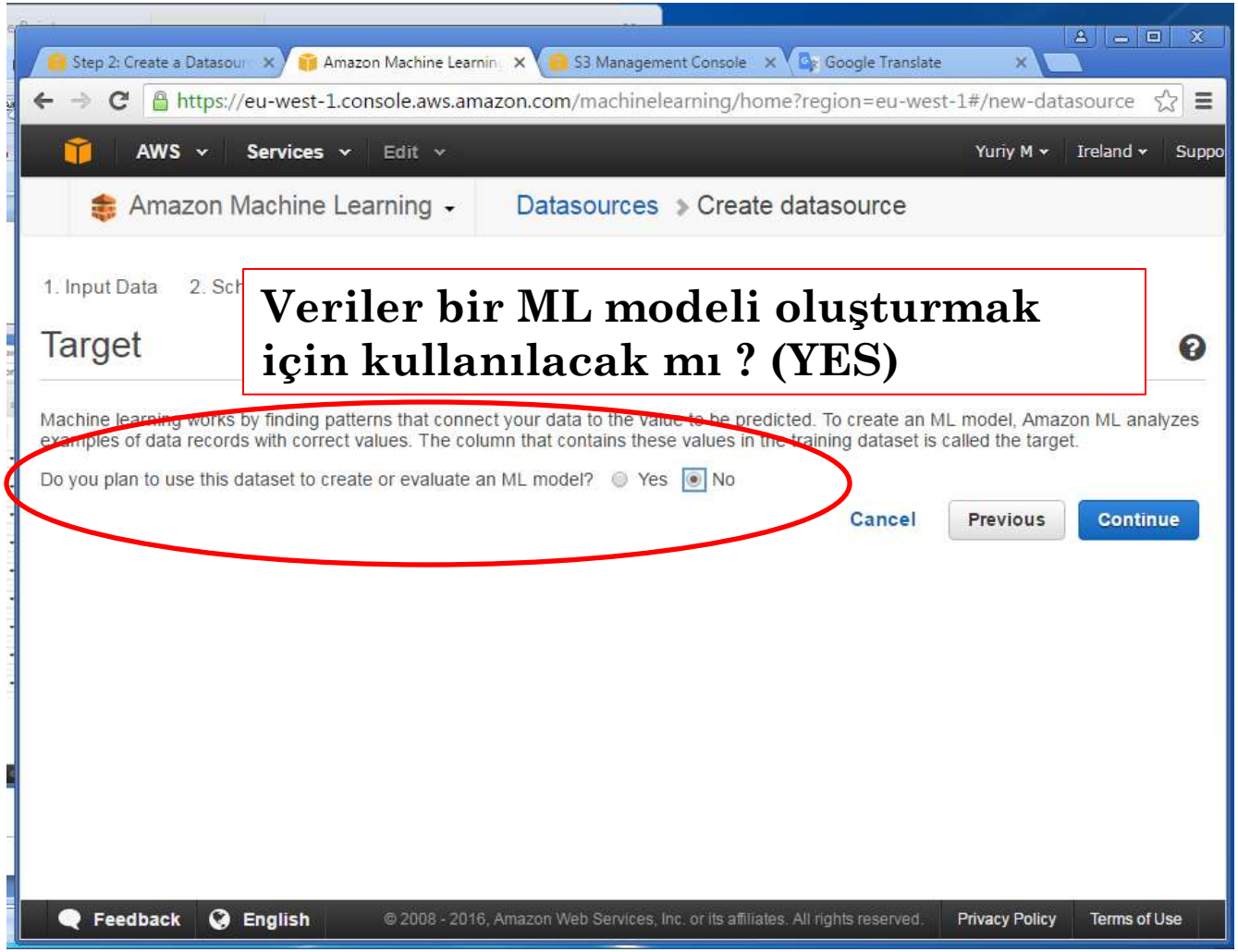
Devam (Continue)

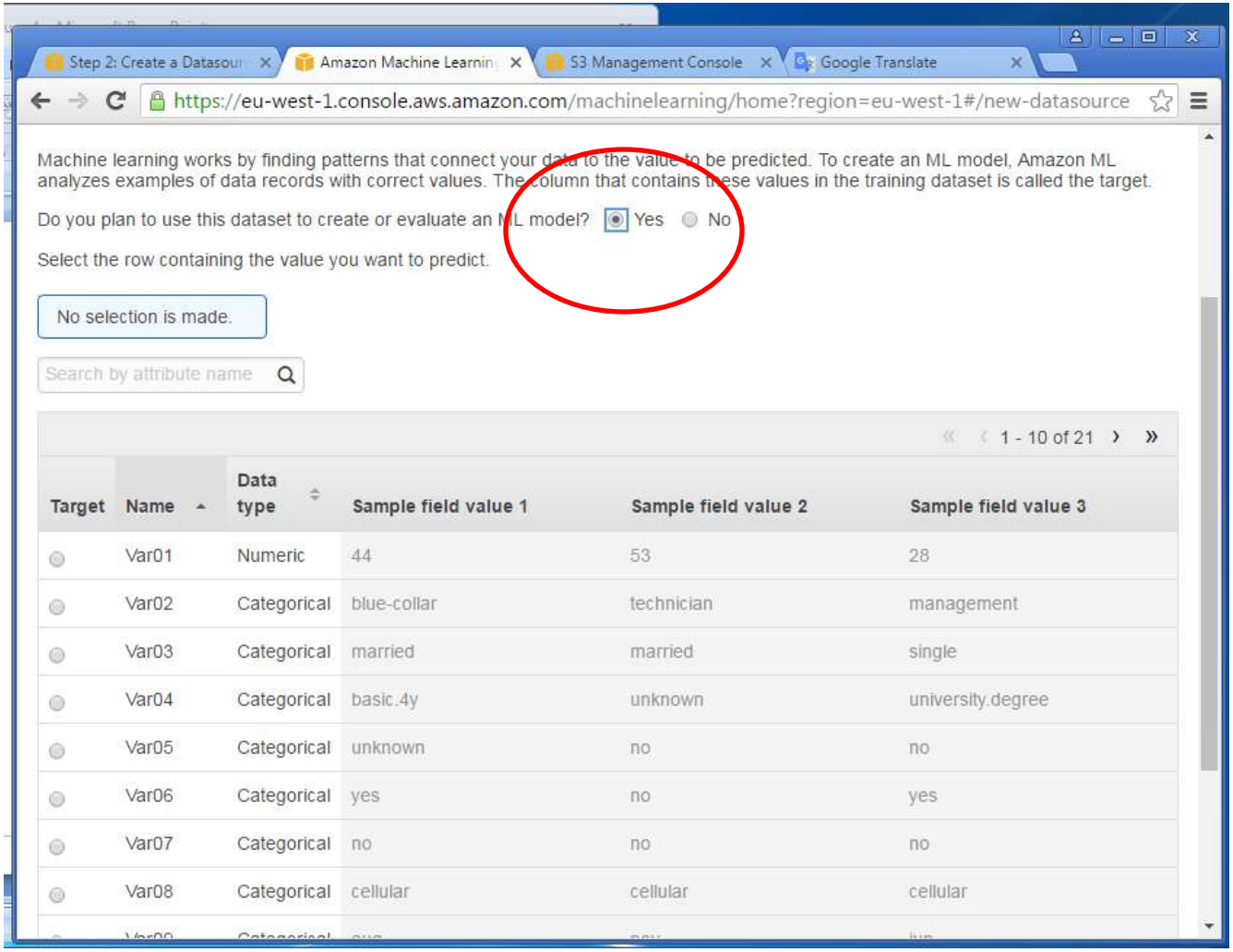

Step 2: Create a Datasour
Amazon Machine Learnin
S3 Management Console
Google Translate
https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/new-datasource
1. Input Data
2. Schema
3. Target
4. Row ID
Schema
Amazon ML scanned your input data and inferred the colu
edit the data type for each column to ensure that it accurat
correctly and to produce accurate predictions. Learn more.
Does the first line in your CSV contain the column names? Yes No
ACTION: Change type
CSV dosyanızda ilk satır başlık isimleri midir (evet/hayır – yes/no)?
Verinin şeması – özniteliklerin tipi
Yardımcı olmak için, öznitelik değerlerinin baze örnekleri
Items per page: 10
Data type
Sample field value 1
Sample field value 2
Sample
1 Var01 Numeric 44 53 28
2 Var02 Categorical blue-collar technician management
3 Var03 Categorical married married single
4 Var04 Categorical basic.4y unknown university.degree
5 Var05 Categorical unknown no no
6 Var06 Categorical yes no yes
7 Var07 Categorical no no



Hepsi gözden geçirilmeli !

Devam (continue)

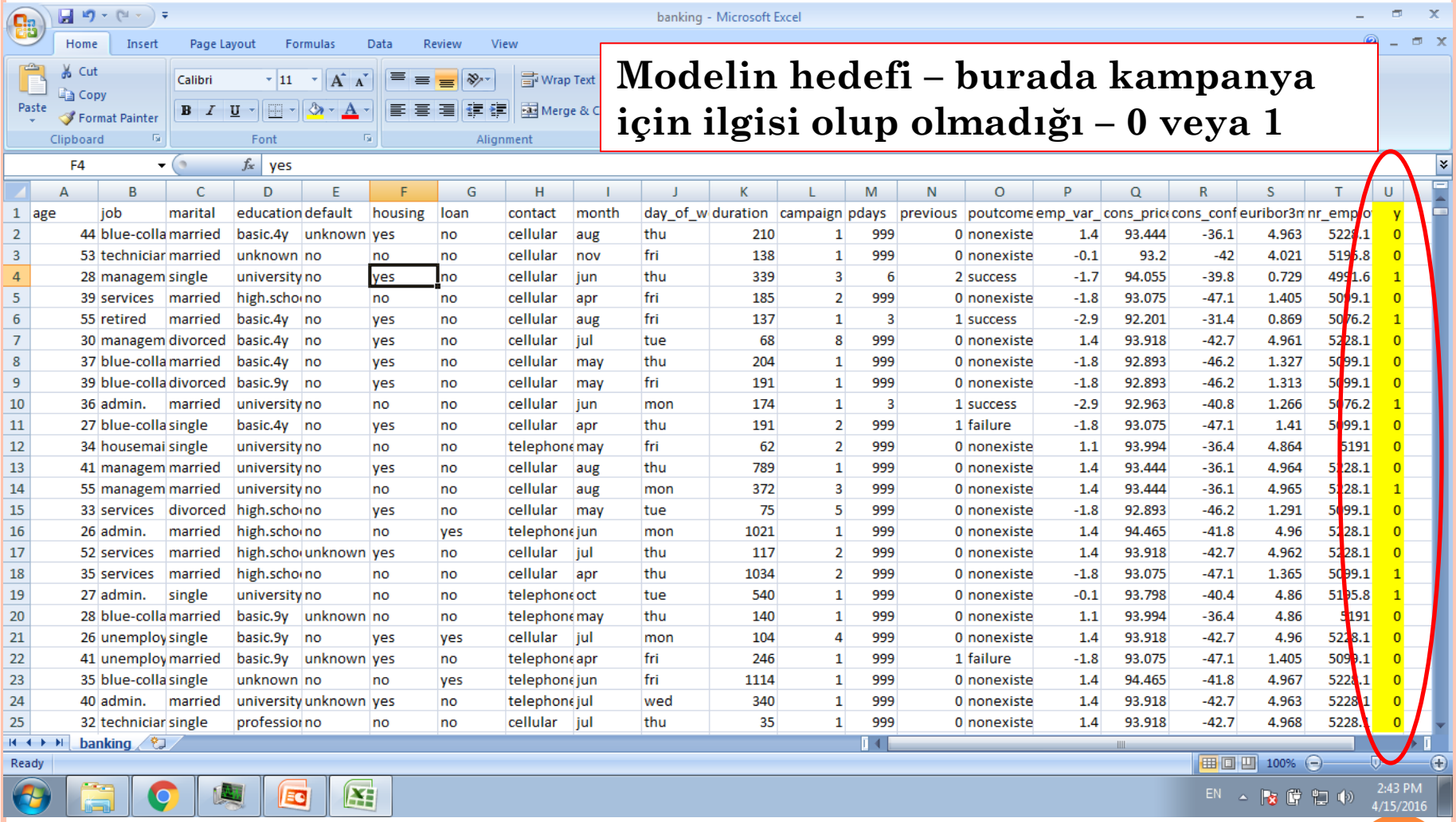

Step 2: Create a Datasour
Amazon Machine Learnin
S3 Management Console
Google Translate
https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/new-datasource
AWS
Services
Edit
Yuriy M
Ireland
Suppo
Amazon Machine Learning
Datasources
Create datasource
1. Input Data
2. Sch
Target
Veriler bir ML modeli oluşturmak için kullanılacak mı ? (YES)
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
Cancel
Previous
Continue
Feedback
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use


Step 2: Create a Datasour
Amazon Machine Learnin
S3 Management Console
Google Translate
https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/new-datasource
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.
No selection is made.
Search by attribute name
1 - 10 of 21
Target Name Data type Sample field value 1 Sample field value 2 Sample field value 3
Var01 Numeric 44 53 28
Var02 Categorical blue-collar technician management
Var03 Categorical married married single
Var04 Categorical basic.4y unknown university.degree
Var05 Categorical unknown no no
Var06 Categorical yes no yes
Var07 Categorical no no no
Var08 Categorical cellular cellular cellular

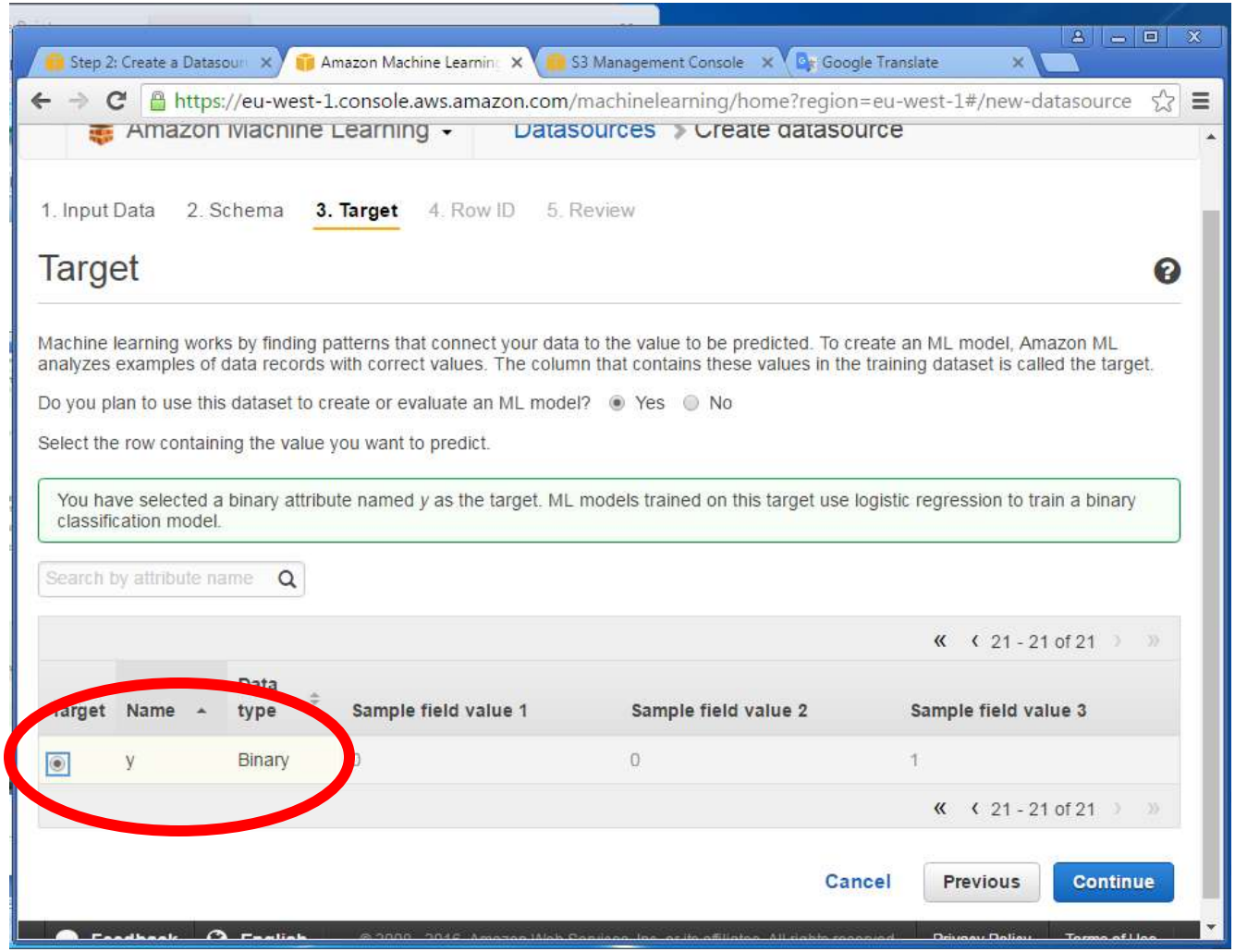
Amazon ML tutorial

- Amazon ML'deki ML modelleri, herhangi yüklediğiniz CVS dosyasında tüm diğer sütünleri kullanılarak bir sütünün diğer tahmin edilmesi için oluşturulur
- Buna göre CVS dosyanızda bir sütün ML modelin hedefi olarak tanımlanacak, ve bu sütünün Amazon ML için belirtilmesi lazım

Modelin hedefi – burada kampanya için ilgisi olup olmadığı – 0 veya 1

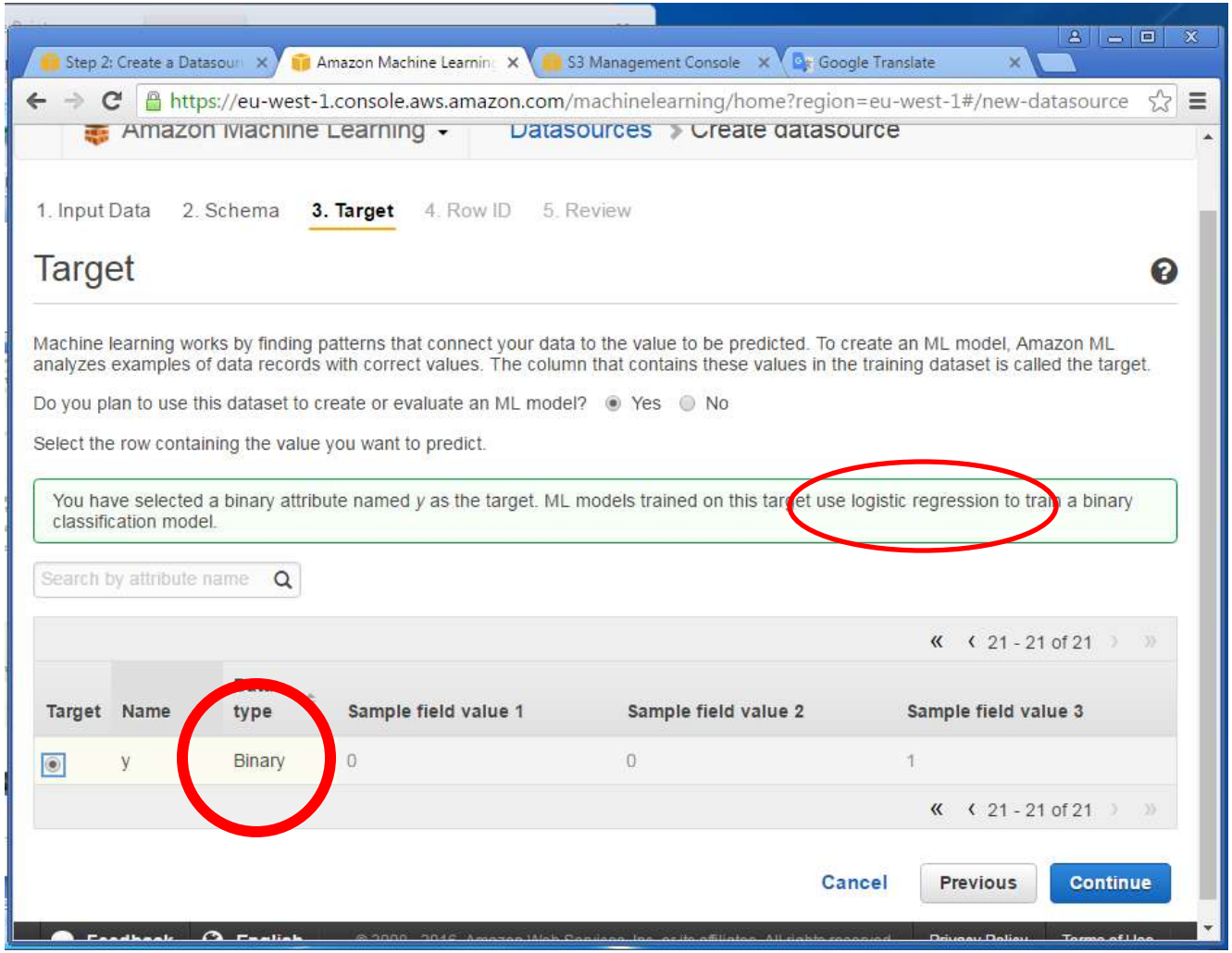


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	age	job	marital	education	default	housing	loan	contact	month	day_of_w	duration	campaign	pdays	previous	poutcome	emp_var_	cons_pric	cons_conf	euribor3m	nr_emplo	y
2	44	blue-colla	married	basic.4y	unknown	yes	no	cellular	aug	thu	210	1	999	0	nonexiste	1.4	93.444	-36.1	4.963	5228.1	0
3	53	technician	married	unknown	no	no	no	cellular	nov	fri	138	1	999	0	nonexiste	-0.1	93.2	-42	4.021	5195.8	0
4	28	managem	single	university	no	yes	no	cellular	jun	thu	339	3	6	2	success	-1.7	94.055	-39.8	0.729	4991.6	1
5	39	services	married	high.scho	no	no	no	cellular	apr	fri	185	2	999	0	nonexiste	-1.8	93.075	-47.1	1.405	5099.1	0
6	55	retired	married	basic.4y	no	yes	no	cellular	aug	fri	137	1	3	1	success	-2.9	92.201	-31.4	0.869	5076.2	1
7	30	managem	divorced	basic.4y	no	yes	no	cellular	jul	tue	68	8	999	0	nonexiste	1.4	93.918	-42.7	4.961	5228.1	0
8	37	blue-colla	married	basic.4y	no	yes	no	cellular	may	thu	204	1	999	0	nonexiste	-1.8	92.893	-46.2	1.327	5099.1	0
9	39	blue-colla	divorced	basic.9y	no	yes	no	cellular	may	fri	191	1	999	0	nonexiste	-1.8	92.893	-46.2	1.313	5099.1	0
10	36	admin.	married	university	no	no	no	cellular	jun	mon	174	1	3	1	success	-2.9	92.963	-40.8	1.266	5076.2	1
11	27	blue-colla	single	basic.4y	no	yes	no	cellular	apr	thu	191	2	999	1	failure	-1.8	93.075	-47.1	1.41	5099.1	0
12	34	housemai	single	university	no	no	no	telephone	may	fri	62	2	999	0	nonexiste	1.1	93.994	-36.4	4.864	5191	0
13	41	managem	married	university	no	yes	no	cellular	aug	thu	789	1	999	0	nonexiste	1.4	93.444	-36.1	4.964	5228.1	0
14	55	managem	married	university	no	no	no	cellular	aug	mon	372	3	999	0	nonexiste	1.4	93.444	-36.1	4.965	5228.1	1
15	33	services	divorced	high.scho	no	yes	no	cellular	may	tue	75	5	999	0	nonexiste	-1.8	92.893	-46.2	1.291	5099.1	0
16	26	admin.	married	high.scho	no	no	yes	telephone	jun	mon	1021	1	999	0	nonexiste	1.4	94.465	-41.8	4.96	5228.1	0
17	52	services	married	high.scho	unknown	yes	no	cellular	jul	thu	117	2	999	0	nonexiste	1.4	93.918	-42.7	4.962	5228.1	0
18	35	services	married	high.scho	no	no	no	cellular	apr	thu	1034	2	999	0	nonexiste	-1.8	93.075	-47.1	1.365	5099.1	1
19	27	admin.	single	university	no	no	no	telephone	oct	tue	540	1	999	0	nonexiste	-0.1	93.798	-40.4	4.86	5195.8	1
20	28	blue-colla	married	basic.9y	unknown	no	no	telephone	may	thu	140	1	999	0	nonexiste	1.1	93.994	-36.4	4.86	5191	0
21	26	unemploy	single	basic.9y	no	yes	yes	cellular	jul	mon	104	4	999	0	nonexiste	1.4	93.918	-42.7	4.96	5228.1	0
22	41	unemploy	married	basic.9y	unknown	yes	no	telephone	apr	fri	246	1	999	1	failure	-1.8	93.075	-47.1	1.405	5099.1	0
23	35	blue-colla	single	unknown	no	no	yes	telephone	jun	fri	1114	1	999	0	nonexiste	1.4	94.465	-41.8	4.967	5228.1	0
24	40	admin.	married	university	unknown	yes	no	telephone	jul	wed	340	1	999	0	nonexiste	1.4	93.918	-42.7	4.963	5228.1	0
25	32	technician	single	professio	no	no	no	cellular	jul	thu	35	1	999	0	nonexiste	1.4	93.918	-42.7	4.968	5228.1	0

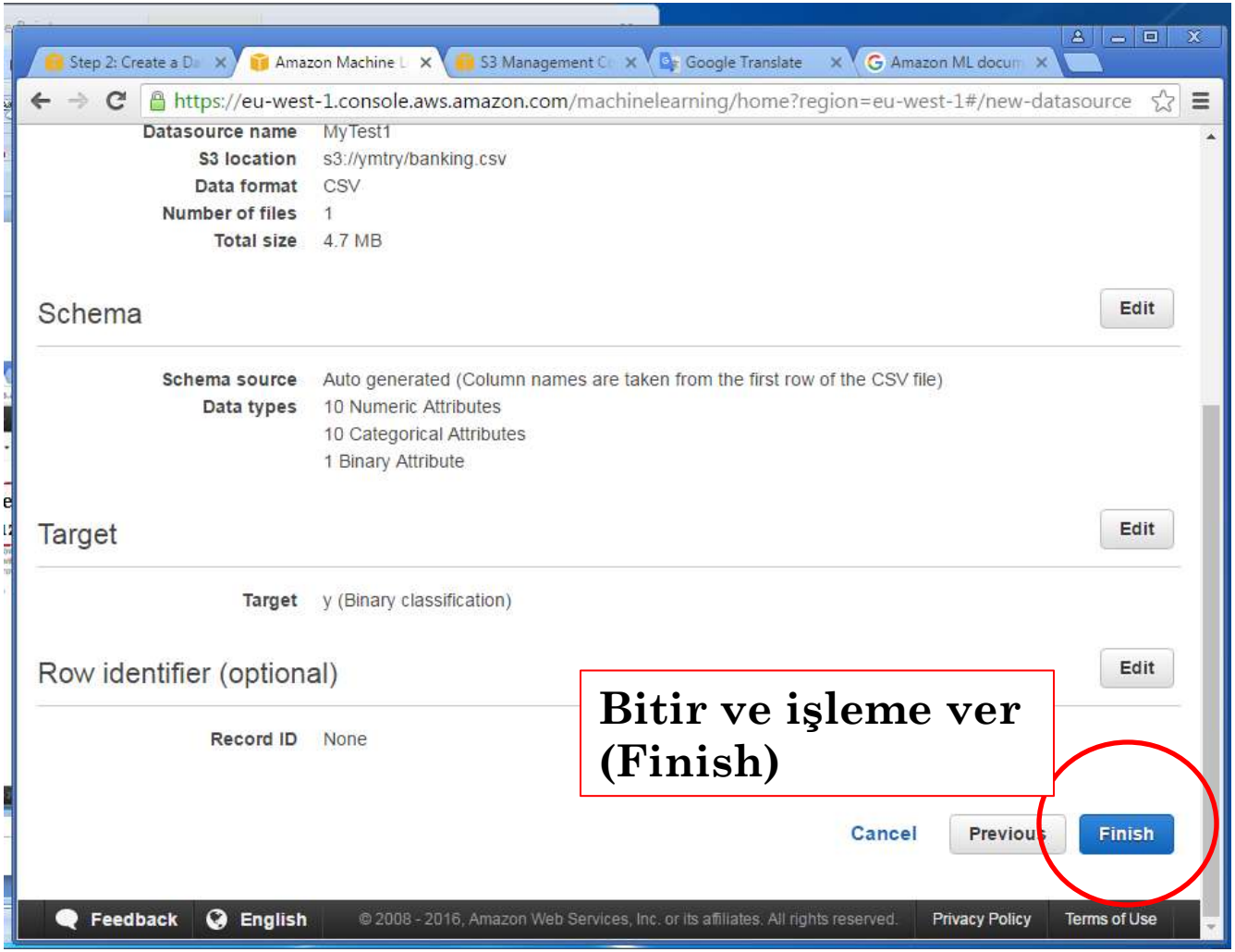

Step 2: Create a Datasour
Amazon Machine Learnin
S3 Management Console
Google Translate
https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/new-datasource
Amazon Machine Learning
Datasources
Create datasource
1. Input Data
2. Schema
3. Target
4. Row ID
5. Review
Target
Machine 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 y as the target. ML models trained on this target use logistic regression to train a binary classification model.
Search by attribute name
21 - 21 of 21
Target
Name
Data type
Sample field value 1
Sample field value 2
Sample field value 3
y
Binary
0
1
Cancel
Previous
Continue

Amazon ML tutorial

- Amazon ML ML modelleri için sadece, regresyon durumunda lineer model ve sınıflandırma durumunda lojistik regresyon diye adlandırılan lineer sınıflandırma modelleri oluşturur
- Bunları, hedef sütünündeki değerleri kullanılarak Amazon ML sizin için otomatik olarak belirtecektir

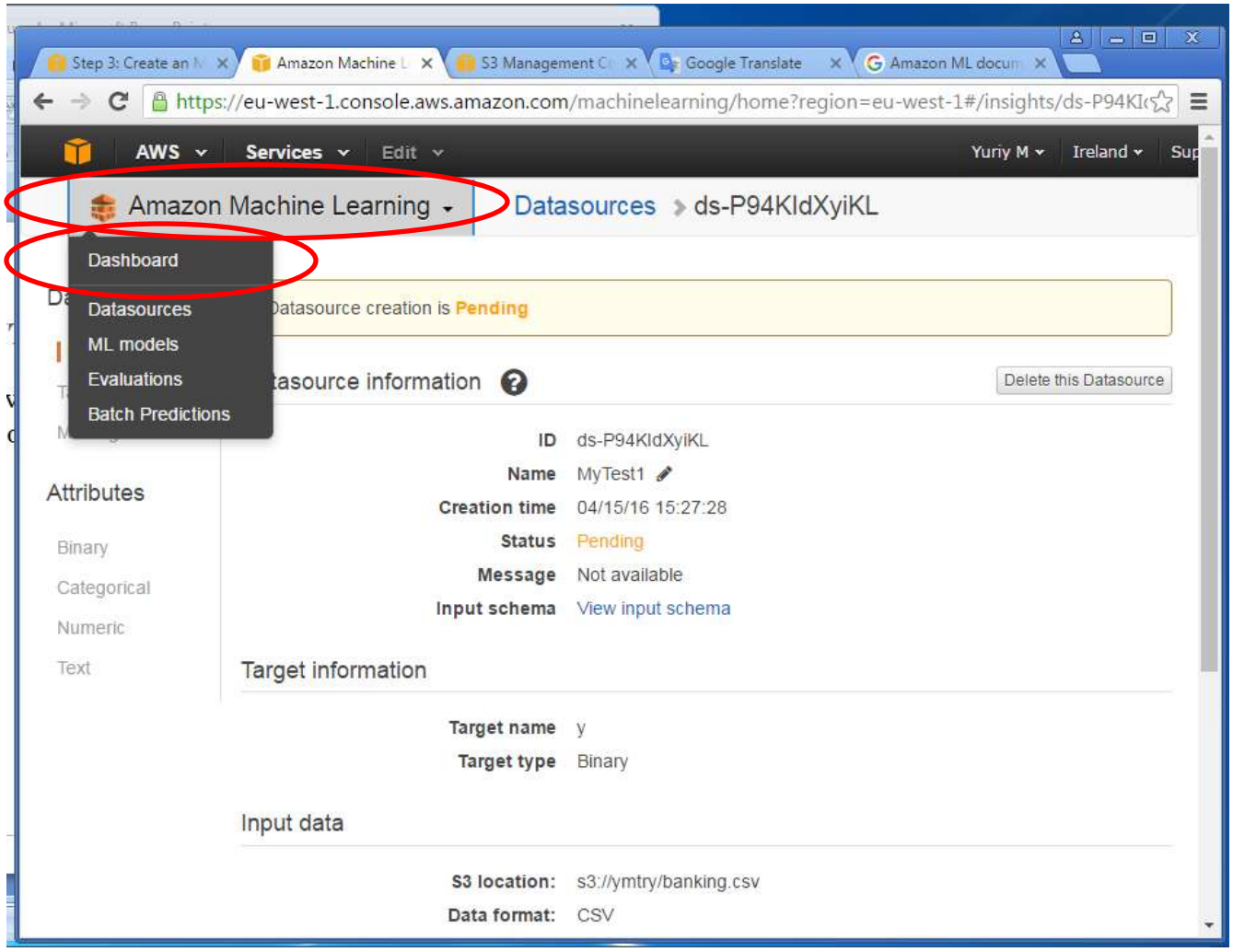

Step 2: Create a Datasour
Amazon Machine Learnin
S3 Management Console
Google Translate
https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/new-datasource
Amazon Machine Learning
Datasources
Create datasource
1. Input Data
2. Schema
3. Target
4. Row ID
5. Review
Target
Machine 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 y as the target. ML models trained on this target use logistic regression to train a binary classification model.
Search by attribute name
21 - 21 of 21
Target
Name
type
Sample field value 1
Sample field value 2
Sample field value 3
y
Binary
0
0
1
21 - 21 of 21
Cancel
Previous
Continue

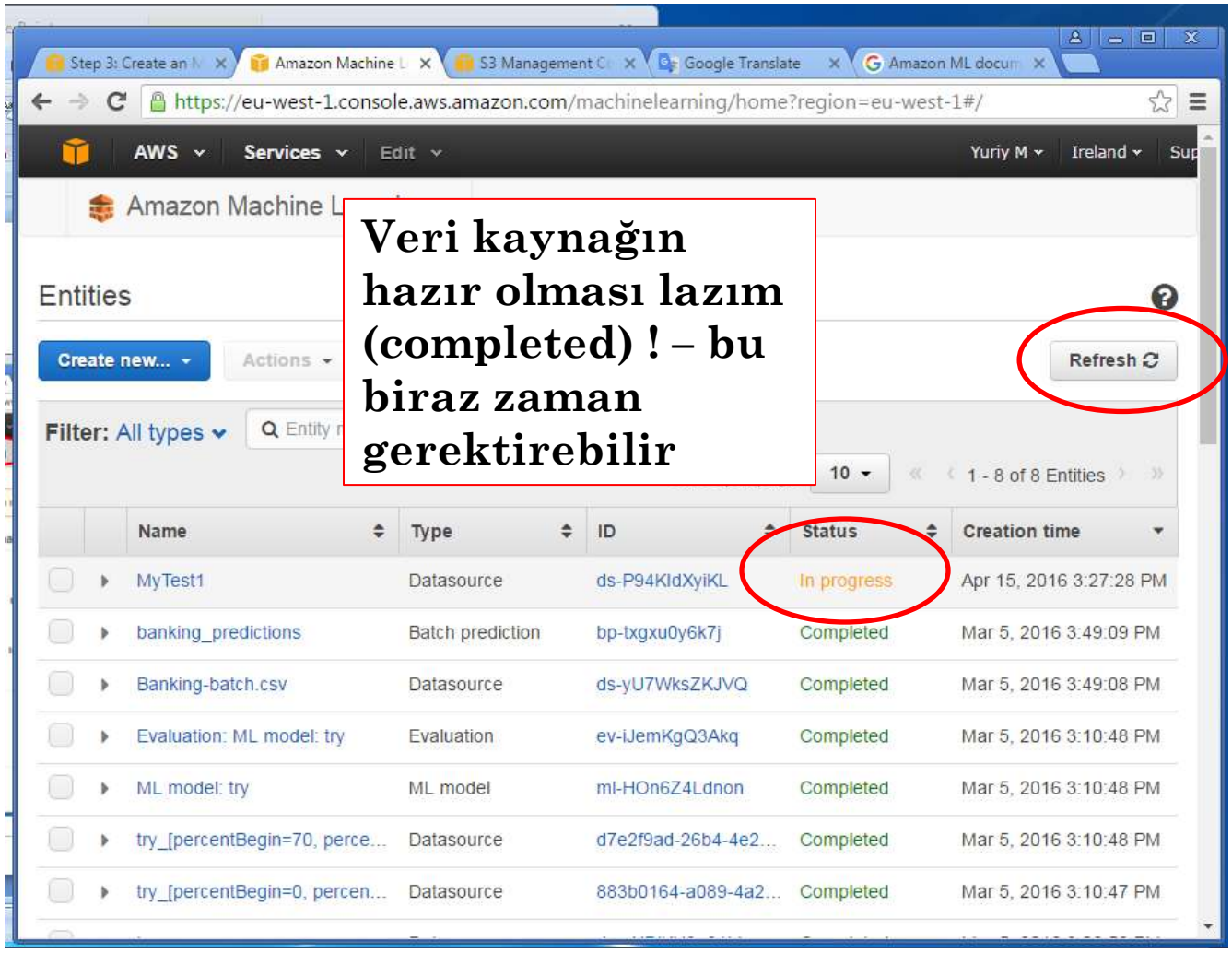

Step 2: Create a Da
Amazon Machine L
S3 Management Co
Google Translate
Amazon ML docum
https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/new-datasource
AWS
Services
Edit
Yuriy M
Ireland
Suppo
Amazon Machine Learning
Datasources
Create datasource
1. Input Data
2. Sc
Row identifi
Sonraki ekrandan direk Review yapıyoruz
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
Feedback
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use


Step 2: Create a D
Amazon Machine L
S3 Management C
Google Translate
Amazon ML docum
https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/new-datasource
Datasource name MyTest1
S3 location s3://ymtry/banking.csv
Data format CSV
Number of files 1
Total size 4.7 MB
Schema
Edit
Schema source Auto generated (Column names are taken from the first row of the CSV file)
Data types 10 Numeric Attributes
10 Categorical Attributes
1 Binary Attribute
Target
Edit
Target y (Binary classification)
Row identifier (optional)
Edit
Record ID None
Bitir ve işleme ver (Finish)
Cancel
Previous
Finish
Feedback
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use

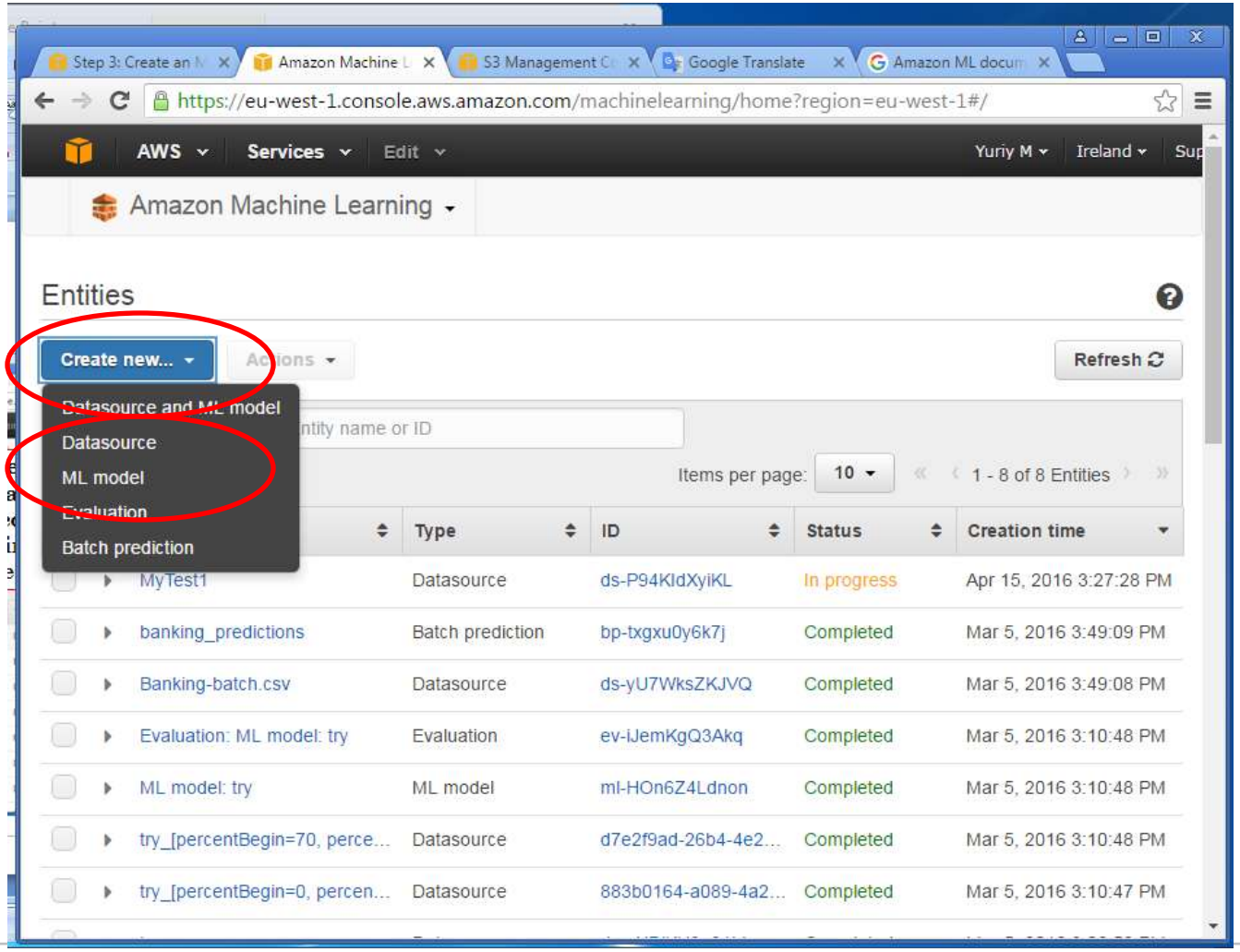
Amazon ML tutorial

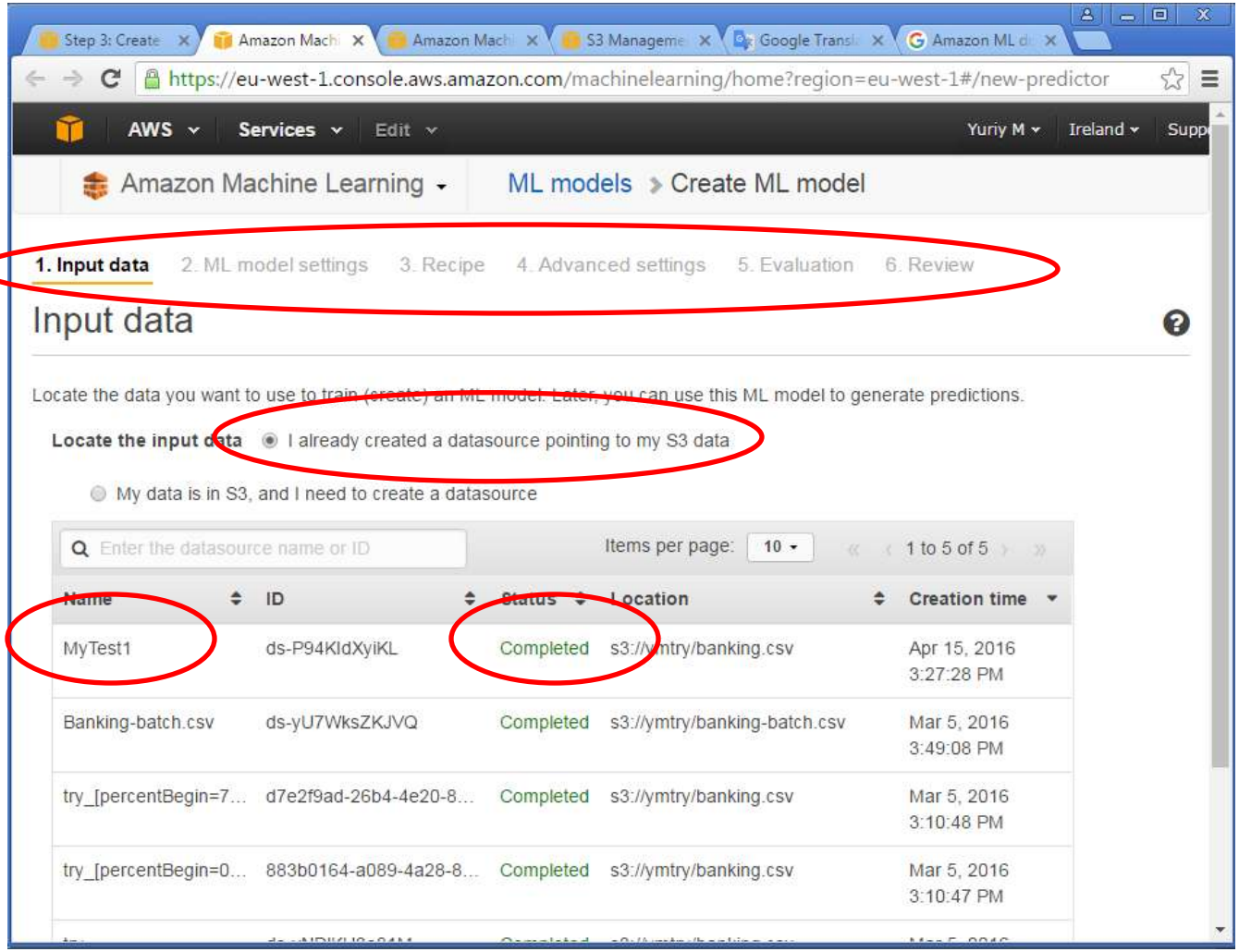
- Oluşturduğunuz veri kaynağı kullanılarak, Amazon ML'e model eğitimi talimatı verilebilir; bunun için

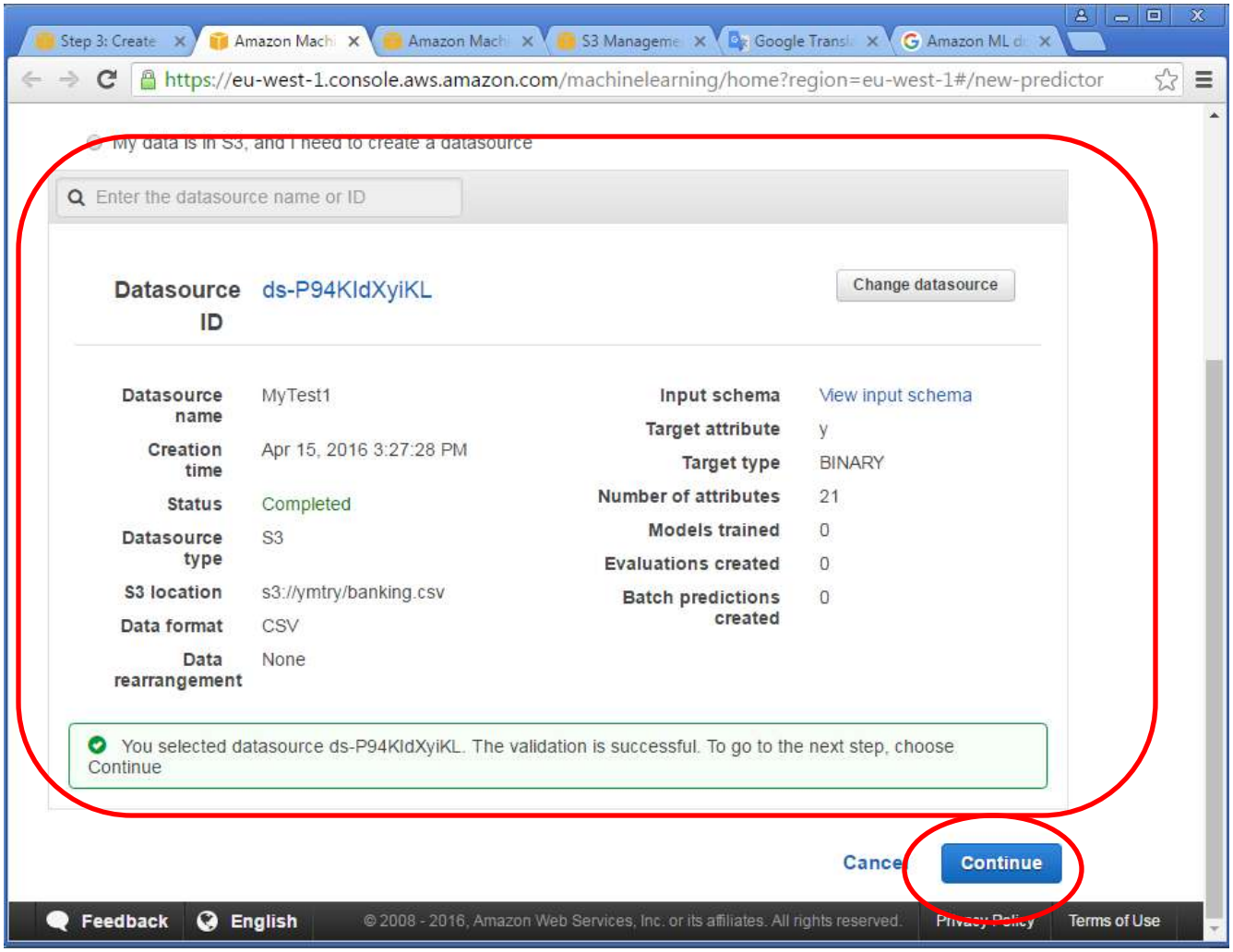

Step 3: Create an
Amazon Machine
S3 Management
Google Translate
Amazon ML docum
https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/insights/ds-P94KI
AWS
Services
Edit
Yuriy M
Ireland
Amazon Machine Learning
Datasources
ds-P94KIdXyiKL
Dashboard
Datasources
ML models
Evaluations
Batch Predictions
Datasource creation is Pending
tasource information
Delete this Datasource
ID ds-P94KIdXyiKL
Name MyTest1
Creation time 04/15/16 15:27:28
Status Pending
Message Not available
Input schema View input schema
Attributes
Binary
Categorical
Numeric
Text
Target information
Target name y
Target type Binary
Input data
S3 location: s3://ymtry/banking.csv
Data format: CSV

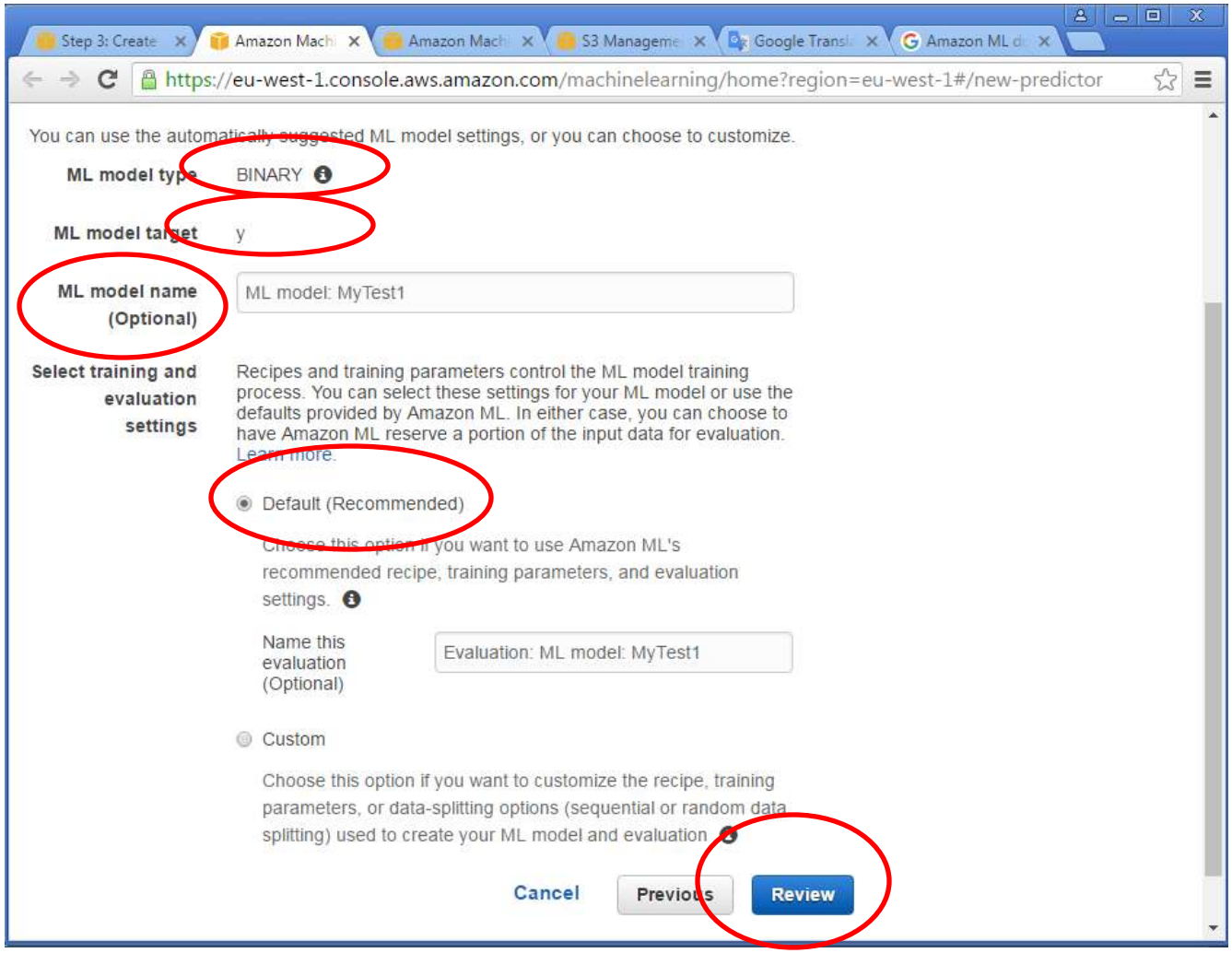


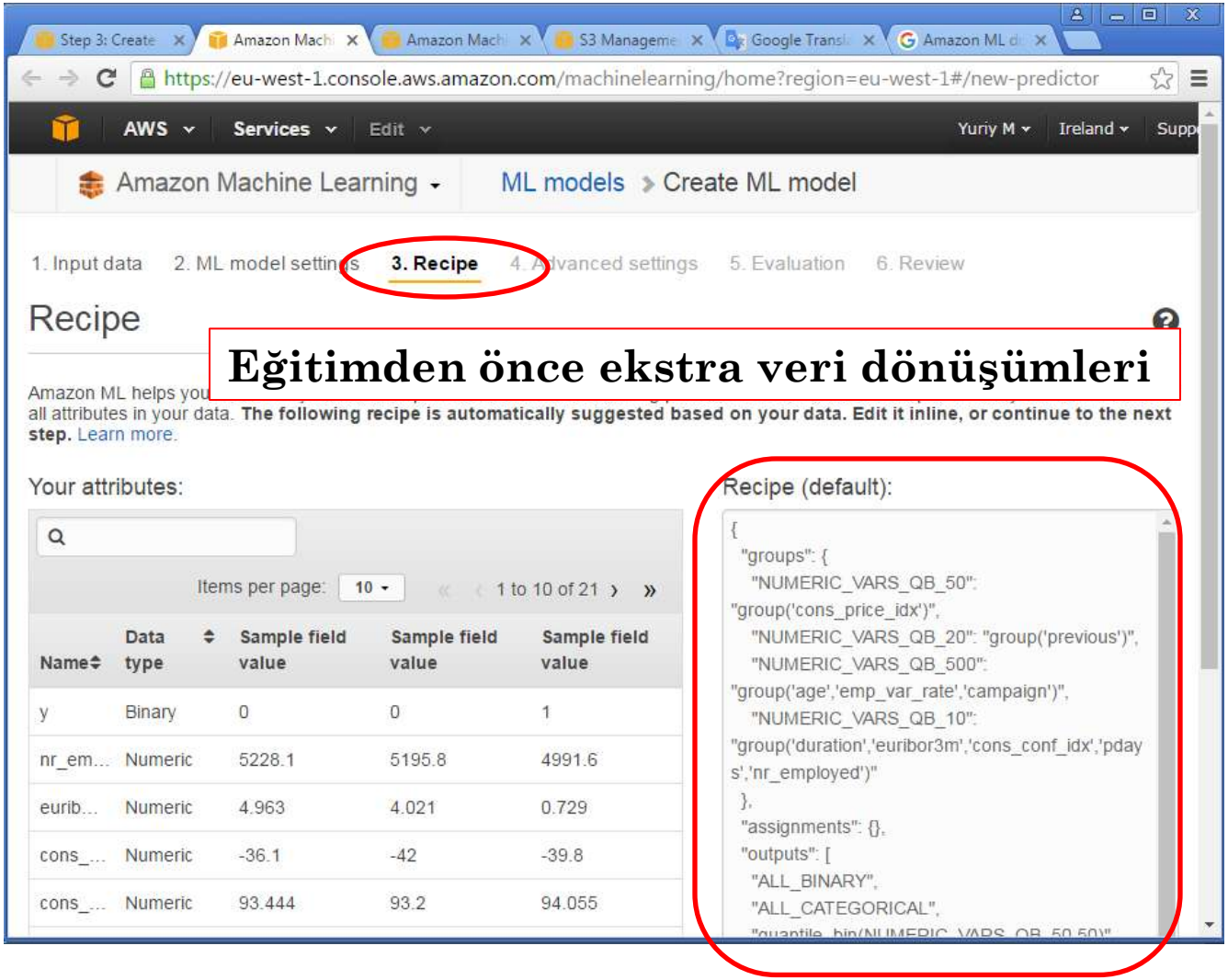
Veri kaynağın hazır olması lazım (completed) ! – bu biraz zaman gerektirebilir


Step 3: Create an
Amazon Machine
S3 Management
Google Translate
Amazon ML docu
https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/
AWS
Services
Edit
Yuriy M
Ireland
Amazon Machine Learning
Entities
Create new...
Refresh
Datasource and ML model
Datasource
ML model
Evaluation
Batch prediction
Items per page:
10
1 - 8 of 8 Entities
Type
ID
Status
Creation time
MyTest1
Datasource
ds-P94KIdXyiKL
In progress
Apr 15, 2016 3:27:28 PM
banking_predictions
Batch prediction
bp-txgxu0y6k7j
Completed
Mar 5, 2016 3:49:09 PM
Banking-batch.csv
Datasource
ds-yU7WksZKJVQ
Completed
Mar 5, 2016 3:49:08 PM
Evaluation: ML model: try
Evaluation
ev-iJemKgQ3Akq
Completed
Mar 5, 2016 3:10:48 PM
ML model: try
ML model
ml-HOn6Z4Ldnon
Completed
Mar 5, 2016 3:10:48 PM
try_[percentBegin=70, perce...
Datasource
d7e2f9ad-26b4-4e2...
Completed
Mar 5, 2016 3:10:48 PM
try_[percentBegin=0, percen...
Datasource
883b0164-a089-4a2...
Completed
Mar 5, 2016 3:10:47 PM

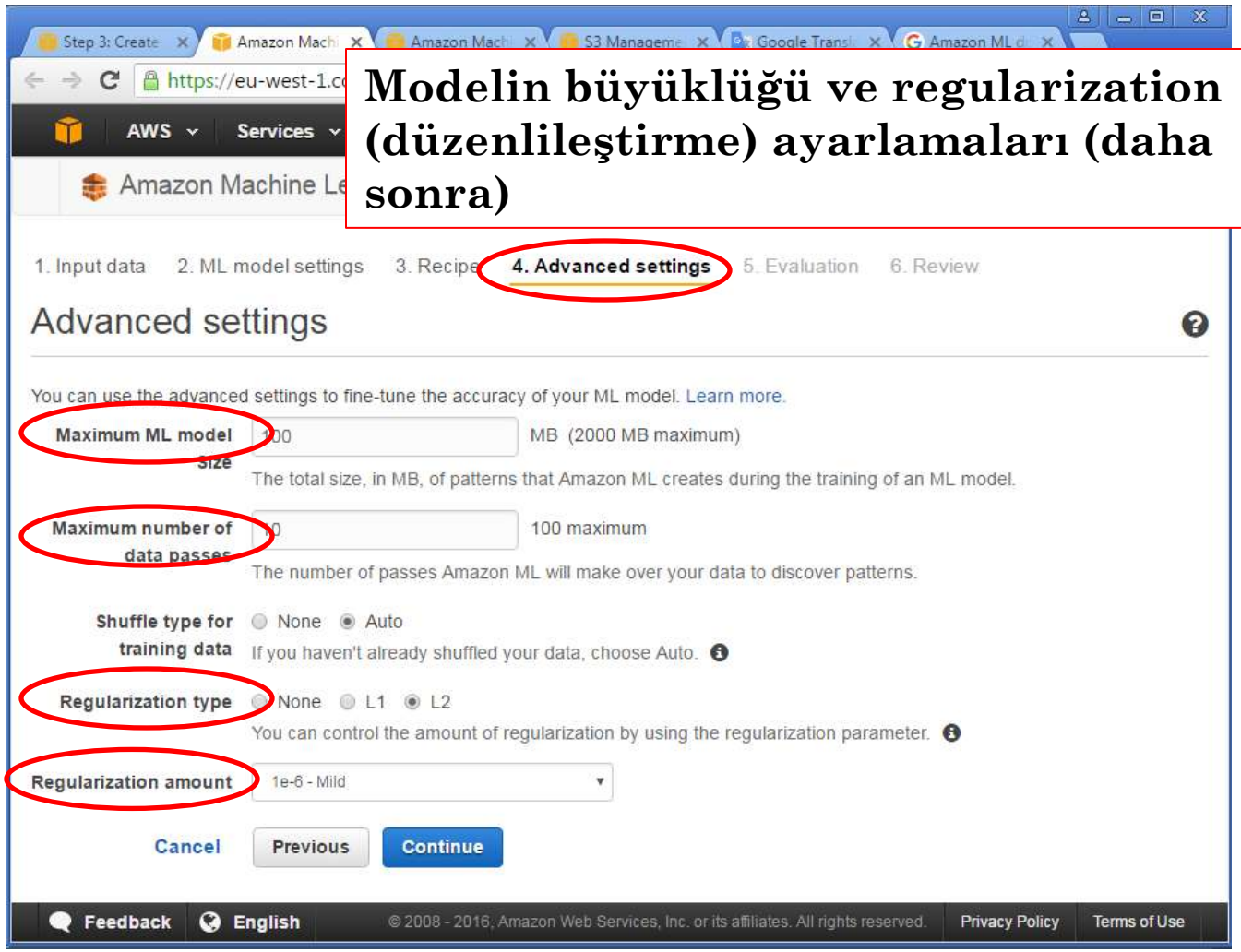

https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/new-predictor
AWS
Services
Edit
Yuriy M
Ireland
Amazon Machine Learning
ML models
Create ML model
1. Input data
2. ML model settings
3. Recipe
4. Advanced settings
5. Evaluation
6. Review
Input data
Locate the data you want 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
Enter the datasource name or ID
Items per page: 10
1 to 5 of 5
Name
ID
Status
Location
Creation time
MyTest1
ds-P94KIdXyiKL
Completed
s3://ymtry/banking.csv
Apr 15, 2016 3:27:28 PM
Banking-batch.csv
ds-yU7WksZKJVQ
Completed
s3://ymtry/banking-batch.csv
Mar 5, 2016 3:49:08 PM
try_[percentBegin=7...
d7e2f9ad-26b4-4e20-8...
Completed
s3://ymtry/banking.csv
Mar 5, 2016 3:10:48 PM
try_[percentBegin=0...
883b0164-a089-4a28-8...
Completed
s3://ymtry/banking.csv
Mar 5, 2016 3:10:47 PM


Step 3: Create
Amazon Mach
Amazon Mach
S3 Manageme
Google Transl
Amazon ML d
https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/new-predictor
My data is in S3, and I need to create a datasource
Enter the datasource name or ID
Datasource ID
ds-P94KIdXyiKL
Change datasource
Datasource name
MyTest1
Creation time
Apr 15, 2016 3:27:28 PM
Status
Completed
Datasource type
S3
S3 location
s3://ymtry/banking.csv
Data format
CSV
Data rearrangement
None
Input schema
View input schema
Target attribute
y
Target type
BINARY
Number of attributes
21
Models trained
0
Evaluations created
0
Batch predictions created
0
You selected datasource ds-P94KIdXyiKL. The validation is successful. To go to the next step, choose Continue
Cancel
Continue
Feedback
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Terms of Use

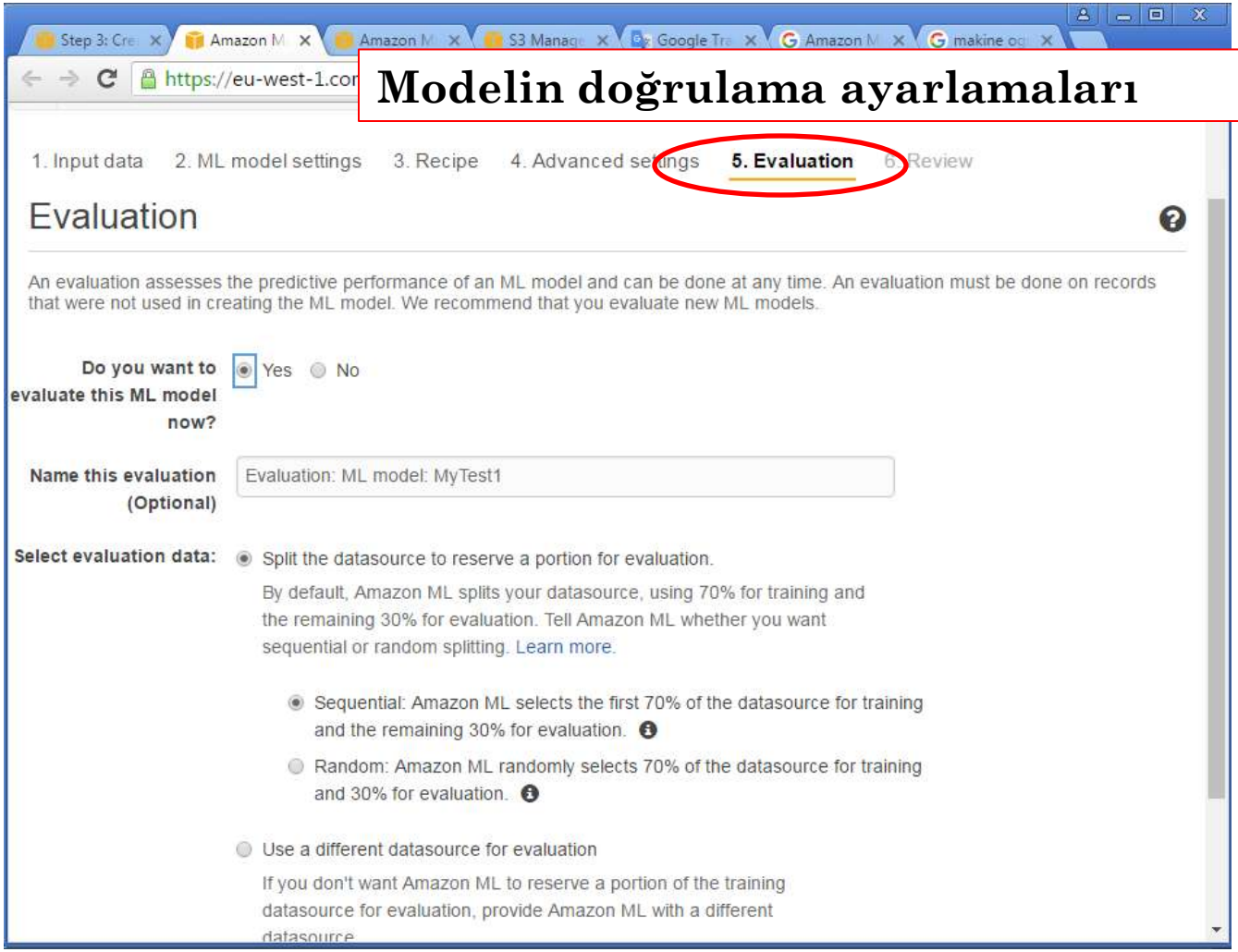

Step 3: Create
Amazon Mach
Amazon Mach
S3 Manageme
Google Transl
Amazon ML d
https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/new-predictor
You can use the automatically suggested ML model settings, or you can choose to customize.
ML model type
BINARY
ML model target
y
ML model name (Optional)
ML model: MyTest1
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)
Choose this option if you want to use Amazon ML's recommended recipe, training parameters, and evaluation settings.
Name this evaluation (Optional)
Evaluation: ML model: MyTest1
Custom
Choose this option if you want to customize the recipe, training parameters, or data-splitting options (sequential or random data splitting) used to create your ML model and evaluation
Cancel
Previous
Review


Step 3: Create
Amazon Machi
Amazon Machi
S3 Manageme
Google Transl
Amazon ML d
https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/new-predictor
AWS
Services
Edit
Yuriy M
Ireland
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
Eğitimden önce ekstra veri dönüşümleri
Amazon ML helps you
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:
Items per page: 10
1 to 10 of 21
Name
Data type
Sample field value
Sample field value
Sample field value
y Binary 0 0 1
nr_em... Numeric 5228.1 5195.8 4991.6
eurib... Numeric 4.963 4.021 0.729
cons_... Numeric -36.1 -42 -39.8
cons_... Numeric 93.444 93.2 94.055
Recipe (default):
{
"groups": {
"NUMERIC_VARS_QB_50":
"group('cons_price_idx')",
"NUMERIC_VARS_QB_20": "group('previous')",
"NUMERIC_VARS_QB_500":
"group('age','emp_var_rate','campaign')",
"NUMERIC_VARS_QB_10":
"group('duration','euribor3m','cons_conf_idx','pdays','nr_employed')"
},
"assignments": {},
"outputs": [
"ALL_BINARY",
"ALL_CATEGORICAL",

Modelin büyüklüğü ve regularization (düzenlileştirme) ayarlamaları (daha sonra)



Modelin doğrulama ayarlamaları



Eğitim ve test (doğrulama) veri kümeleri

- Eğitim ve doğrulama veri kümesi, makine öğrenmesinde önemli bir konu ve kavram
- Bir durumu açıklayan modeli oluşturmak için kullanılan ve anlaşılan durumun örneklerine etiketlenmiş veri (labeled data) denir
- Bu şekilde ML model, etiketlenmiş verileri (bir ölçeğe göre) optimal yani en iyi şekilde açıklayabilecek modeli seçecek

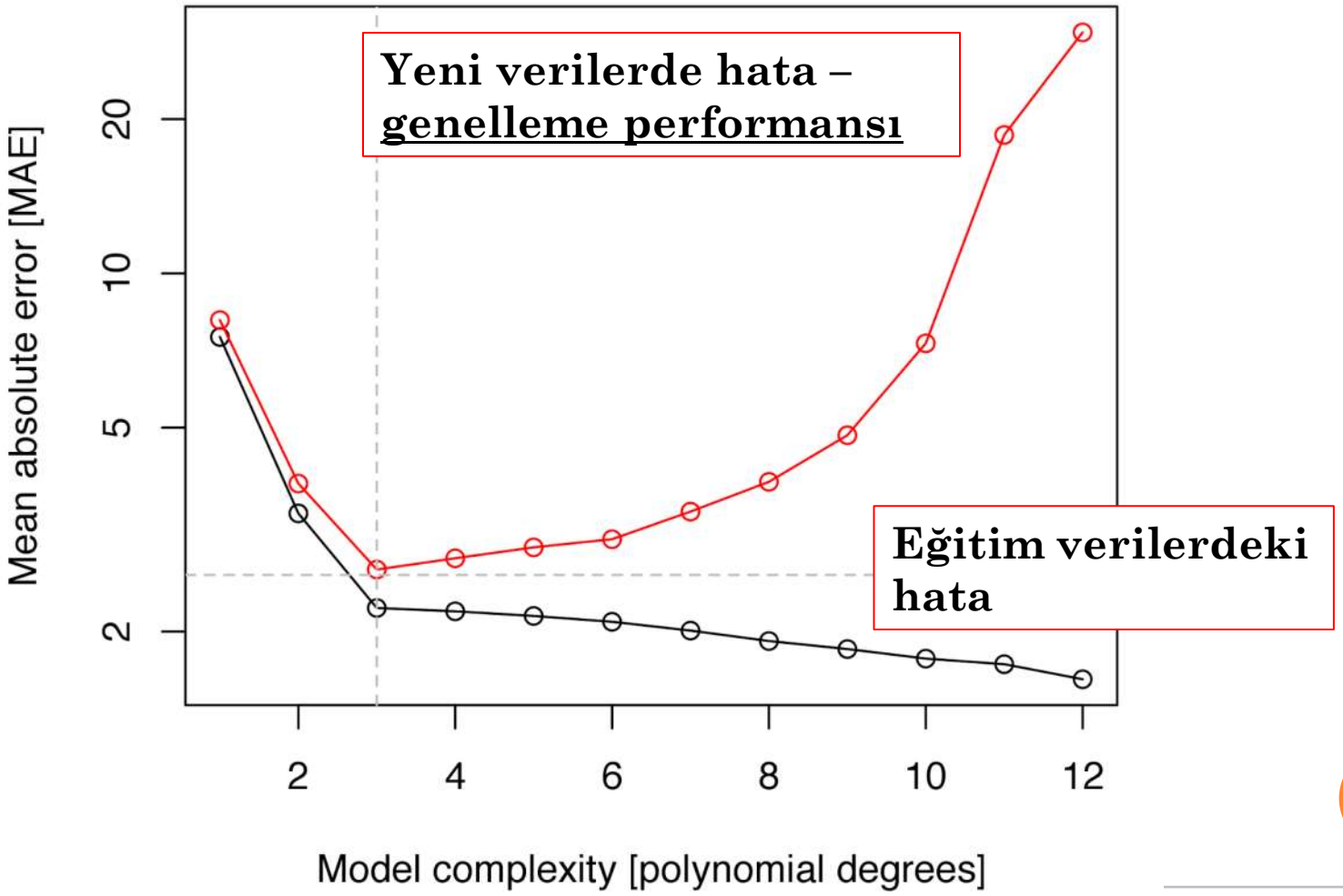
EĞITIM VE TEST (DOĞRULAMA) VERI KÜMELERI

- ML modellerinin, daha önce görülmemiş ve dolayısıyla etiketlenmiş verilerde olmayan durumlarda iyi çalışması istenmekte
- Bu duruma, modellerin gördüğü örnekleri **genelleyebilmesi** denir, yani örnek verisinde yer alan desenlerin daha önce görülmediği verilere uygulanabilmesi

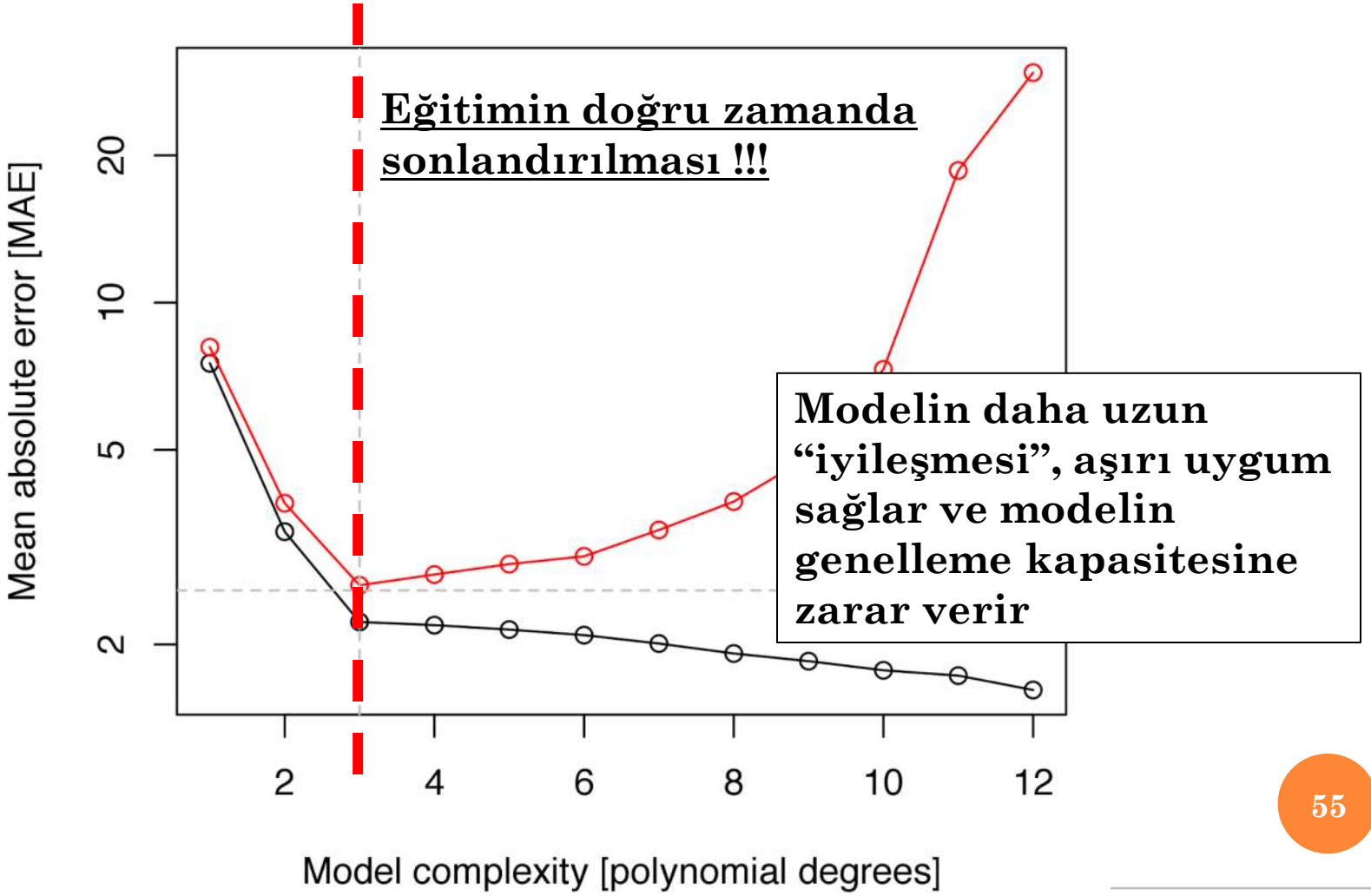
EĞITIM VE TEST (DOĞRULAMA) VERI KÜMELERI

- Makine öğrenmesinde çok görünen durum şöyledir; etiketlenmiş verilerde son derece iyi olan model etiketlenmemiş verilerde, yani daha önce görülmemiş verilerde, son derece kötüdür
- Bu fenomene "**overfitting**" veya "**aşırı uyum**" denir, yani model eğitimdeki örnekleri son derece uygum sağlamak için bu örnekleri esberleyince aynı mantıktan ama daha önce görülmediği örneklerle hiç birşey yapamıyor (bu gerçek hayatın da bir durumudur)

EĞITIM VE TEST (DOĞRULAMA) VERI KÜMELERI



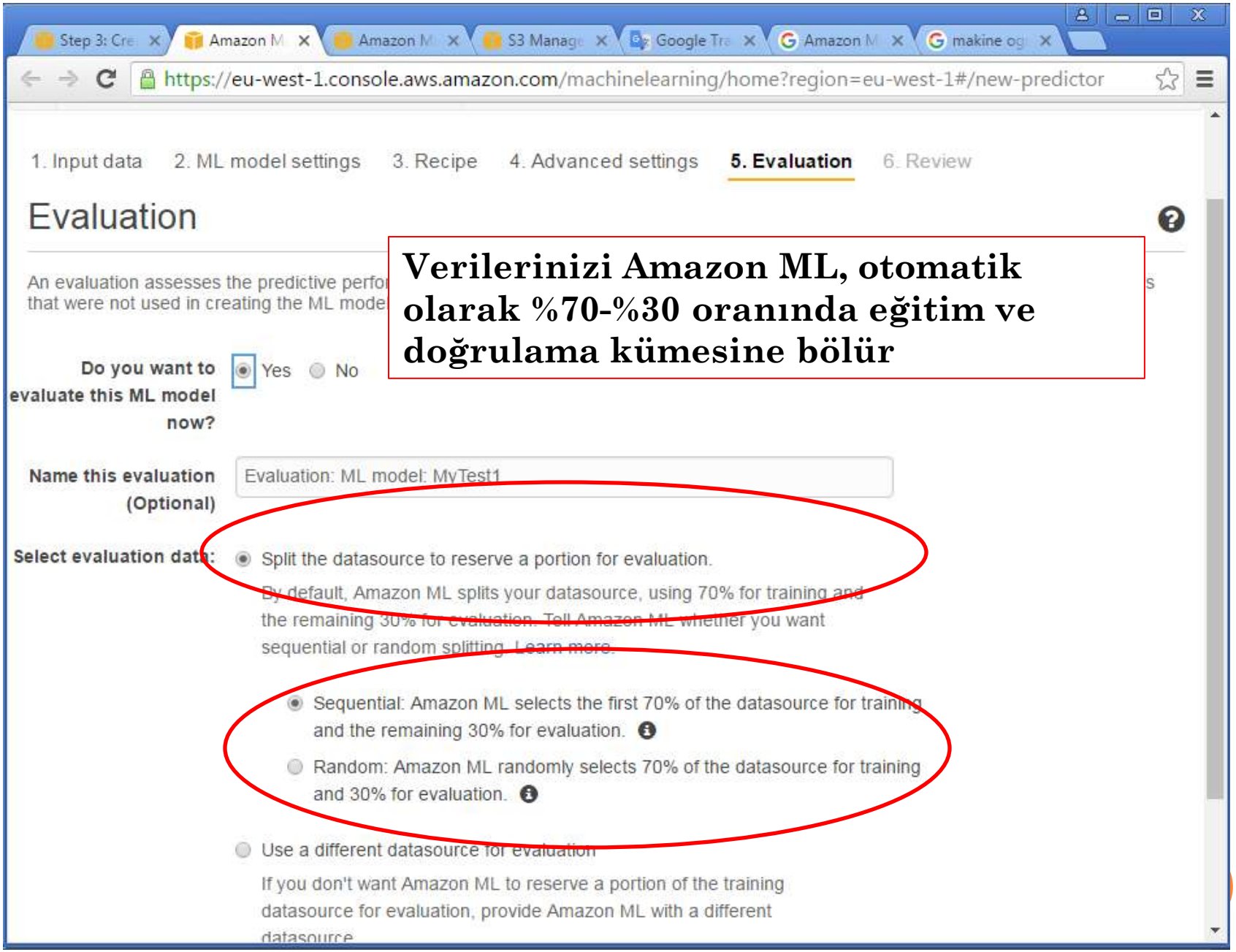
EĞITIM VE TEST (DOĞRULAMA) VERI KÜMELERI

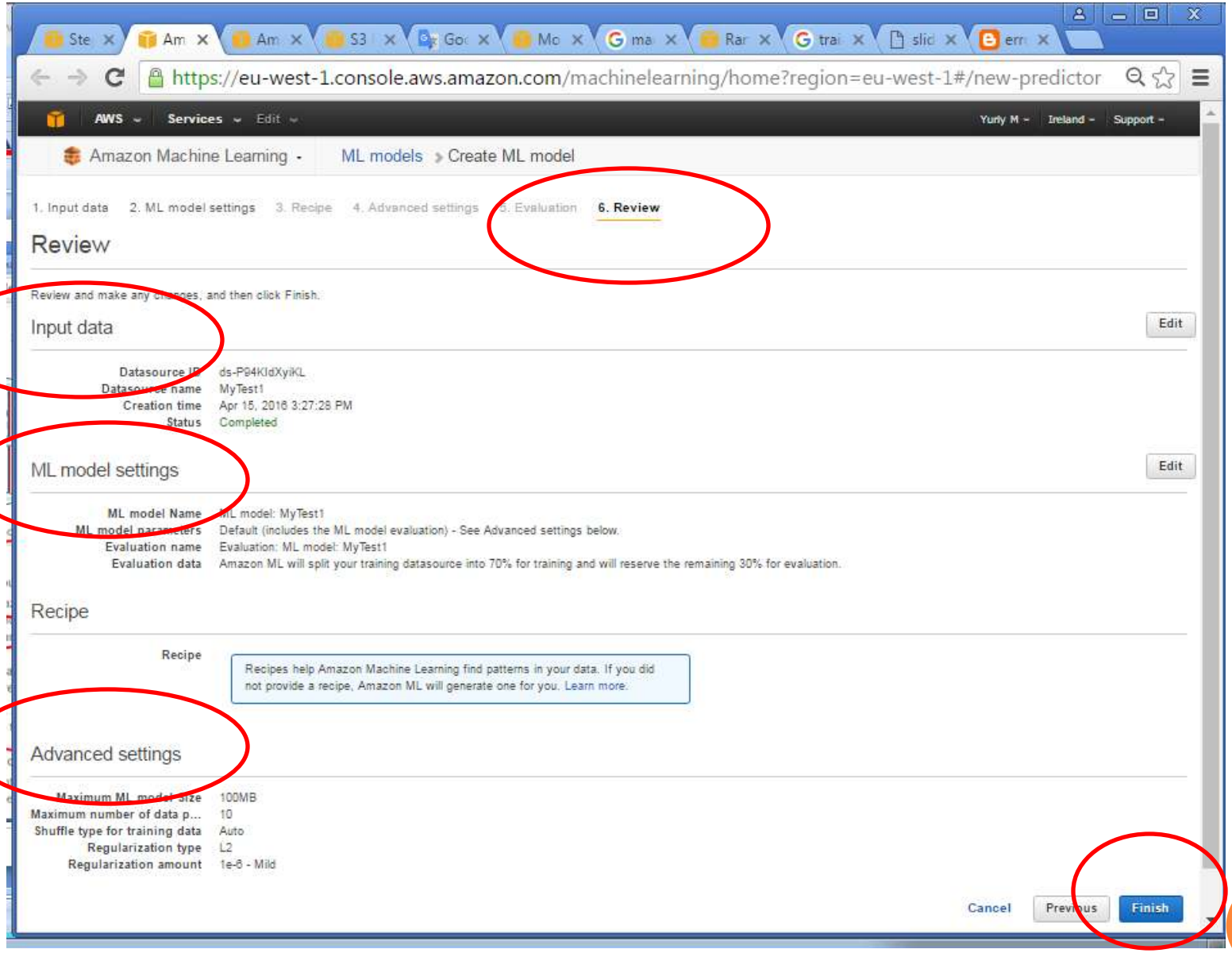


EĞITIM VE TEST (DOĞRULAMA) VERI KÜMELERI

- Modelin genelleme kapasitesini kontrol etmek için kullanılan (etiketlenmiş!) verilere "**test**", "**validation**" veya "**doğrulama**" veri denir
- Modeli iyileştirmek için (yani maliyet fonksiyonu hesaplayarak onun değerini azaltmak için) kullanılan (etiketlenmiş!) verilere "**training**" veya "**eğitim**" veri kümesi denir
- İkisi tabi elimizde var olan etiketlenmiş örneklerden oluşturulur, bunlara **training-test** veya **training-validation split** ("**eğitim-doğrulama bölümü**") denir

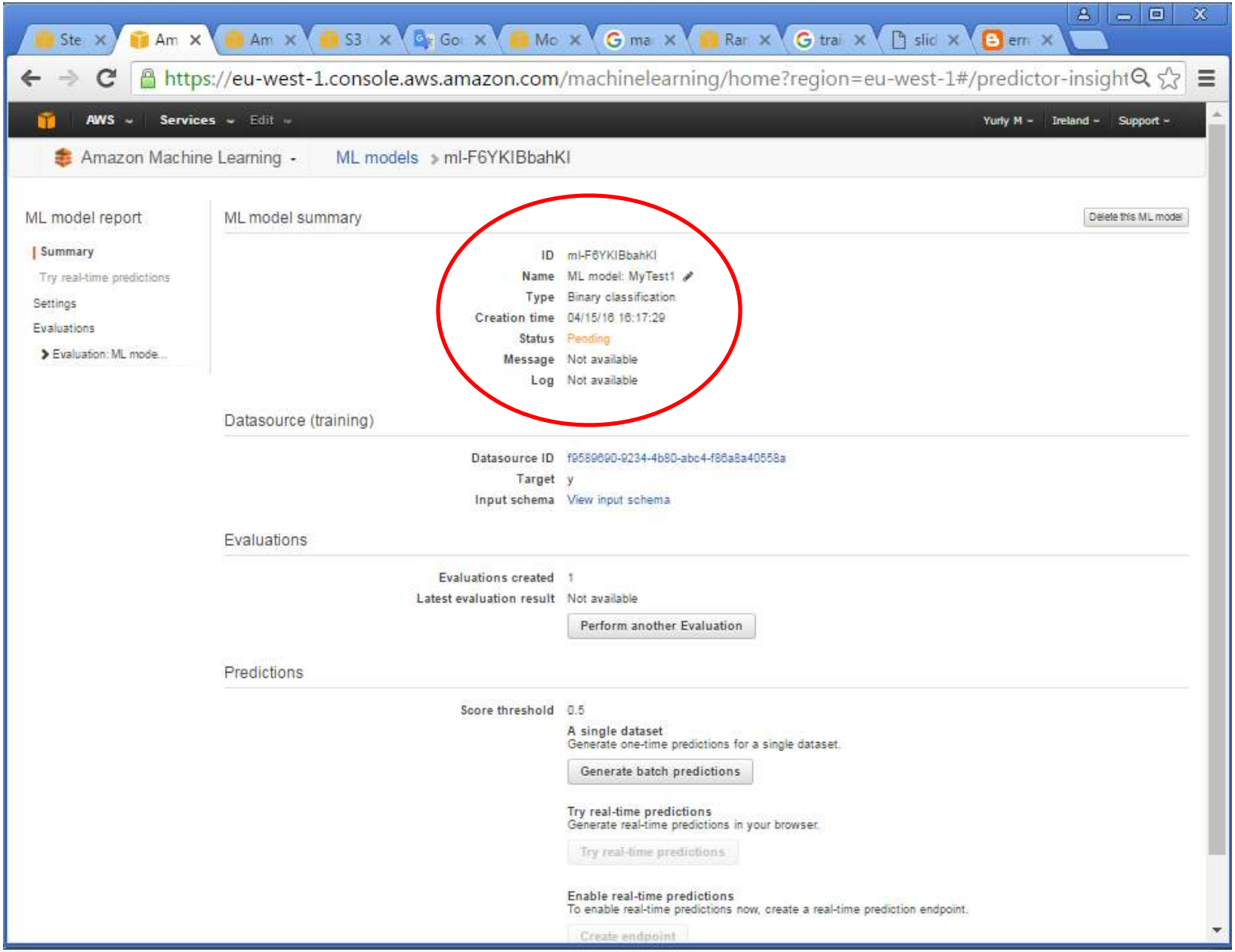
EĞITIM VE TEST (DOĞRULAMA) VERI KÜMELERI


https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/new-predictor
1. Input data
2. ML model settings
3. Recipe
4. Advanced settings
5. Evaluation
6. Review
Evaluation
Verilerinizi Amazon ML, otomatik olarak %70-%30 oranında eğitim ve doğrulama kümesine bölür
Do you want to evaluate this ML model now?
Yes
No
Name this evaluation (Optional)
Evaluation: ML model: MyTest1
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


https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/new-predictor
AWS
Services
Edit
Yuriy M
Ireland
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
Review
Review and make any changes, and then click Finish.
Input data
Edit
Datasource ID ds-P94KIdXyiKL
Datasource name MyTest1
Creation time Apr 15, 2016 3:27:28 PM
Status Completed
ML model settings
Edit
ML model Name ML model: MyTest1
ML model parameters Default (includes the ML model evaluation) - See Advanced settings below.
Evaluation name Evaluation: ML model: MyTest1
Evaluation data Amazon ML will split your training datasource into 70% for training and will reserve the remaining 30% for evaluation.
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.
Advanced settings
Maximum ML model size 100MB
Maximum number of data p... 10
Shuffle type for training data Auto
Regularization type L2
Regularization amount 1e-6 - Mild
Cancel
Previous
Finish

Amazon ML tutorial

- Model tamamladığınızdan sonra, Amazon ML modeli oluşturulacak ve eğitime otomatik olarak alınacak (DİKKAT: bu sonuç Amazon ML tarafından kullanılacak hesaplama kaynaklarının kullanımı için belirli bir masrafa neden olacaktır!)

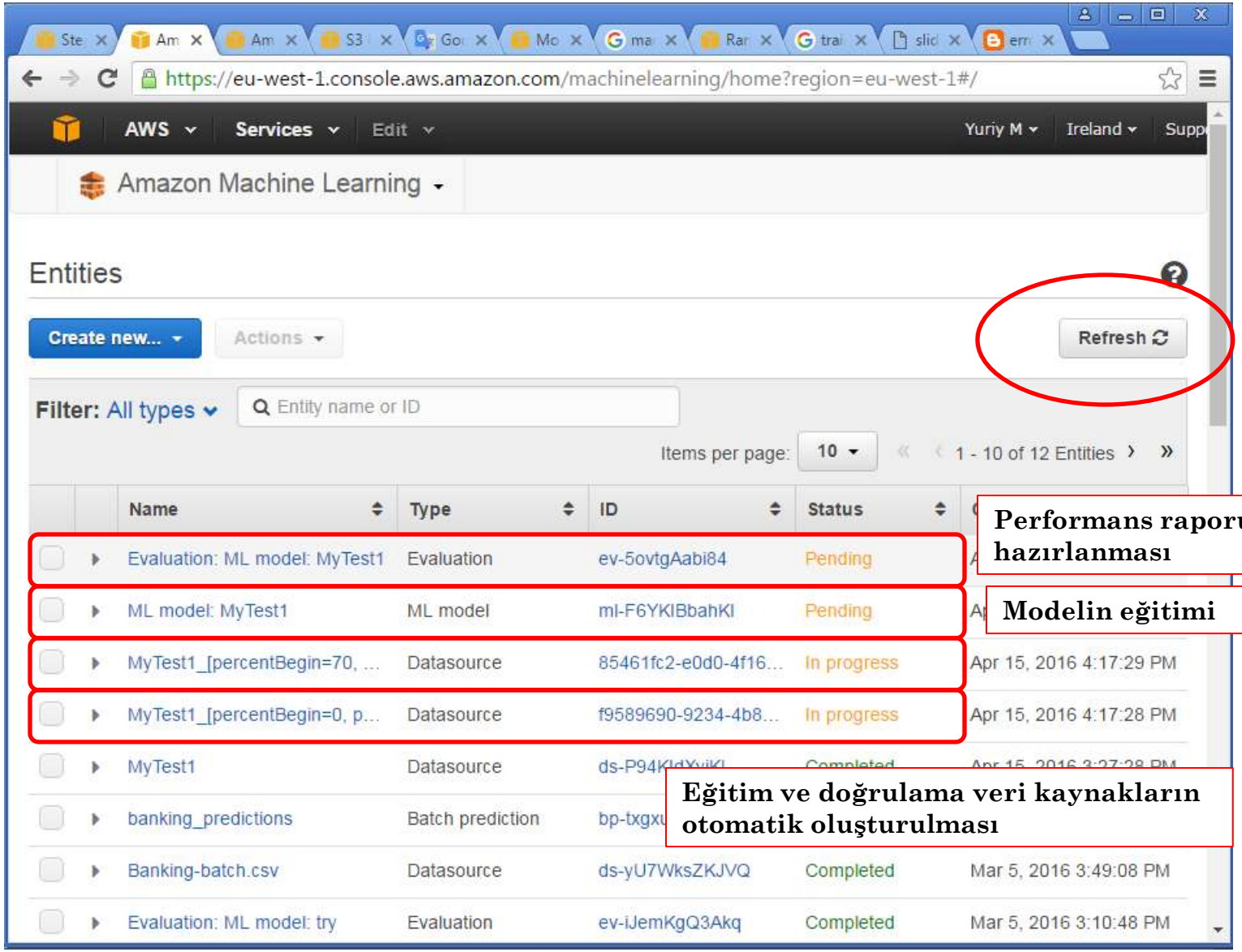

https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/predictor-insight
AWS
Services
Edit
Yuriy M
Ireland
Support
Amazon Machine Learning
ML models
ml-F6YKIBbahKI
ML model report
Summary
Try real-time predictions
Settings
Evaluations
Evaluation: ML mode...
ML model summary
Delete this ML model
ID ml-F6YKIBbahKI
Name ML model: MyTest1
Type Binary classification
Creation time 04/15/16 16:17:29
Status Pending
Message Not available
Log Not available
Datasource (training)
Datasource ID f9589690-9234-4b80-abc4-f86a8a40558a
Target y
Input schema View input schema
Evaluations
Evaluations created 1
Latest evaluation result Not available
Perform another Evaluation
Predictions
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.
Create endpoint

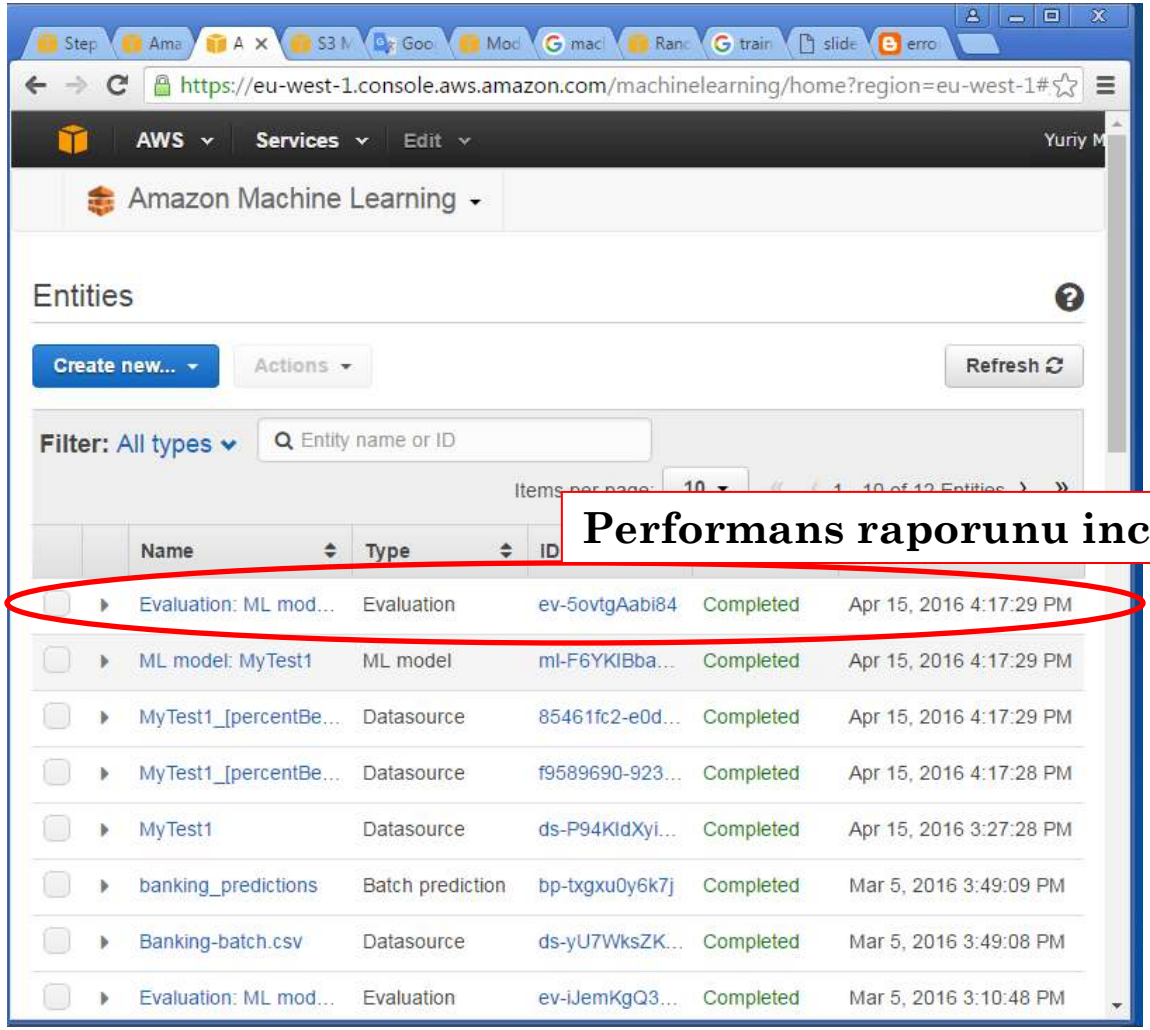
Amazon ML tutorial

- Modelin eğitimi tamamlandıktan sonra (bu biraz vakit gerektirebilir), modelin performans değerlendirme raporu hazırlanacak ve, modelinizi tahmin etme için hazırlamanız için, incelemeniz gerekecek

Amazon ML tutorial

- Modelin durumu, Amazon ML kontrol panelinden (yukarıdaki Amazon Machine Learning menüsündeki Dashboard seçenek) takip edilebilir

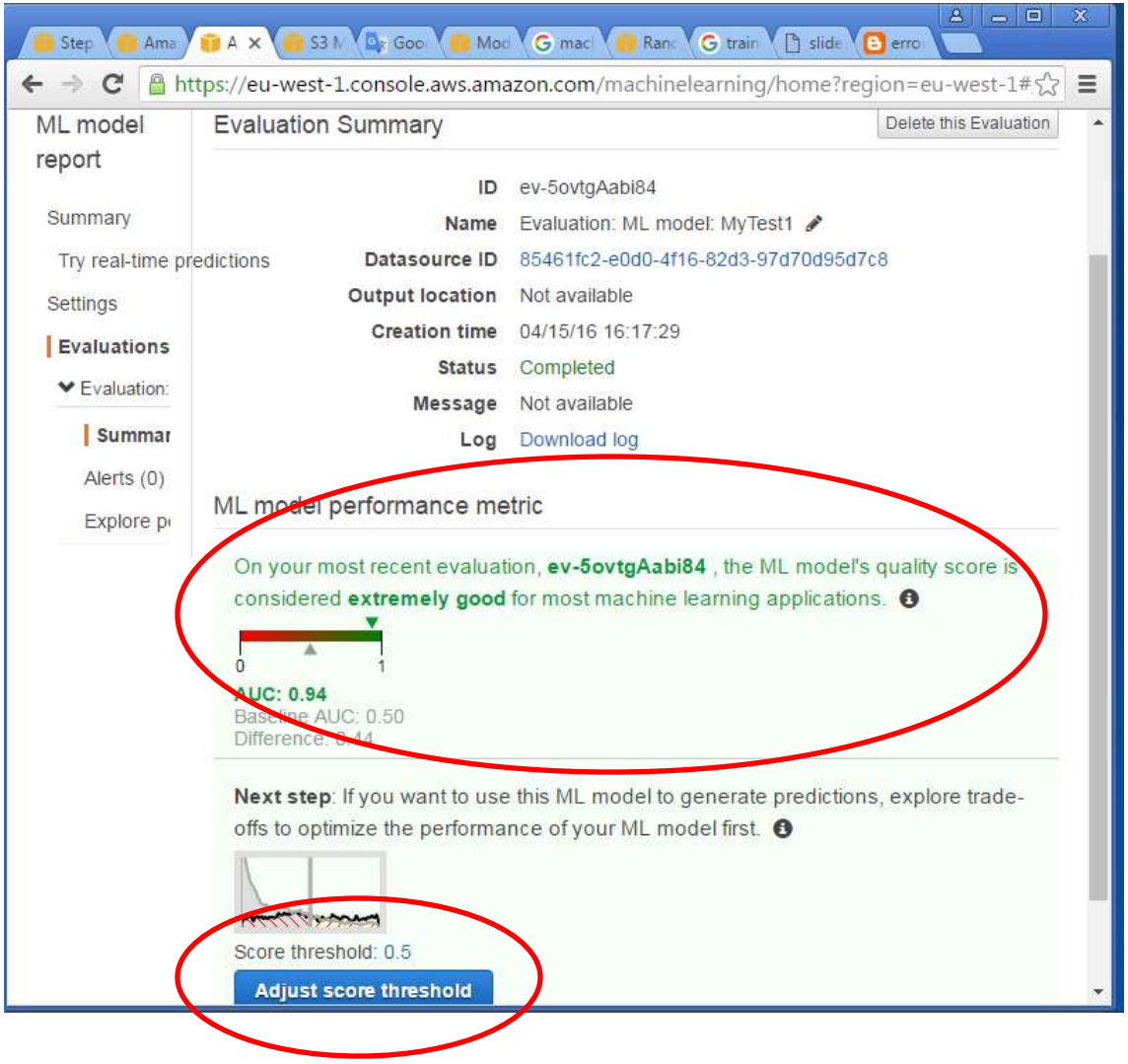

https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/
AWS
Services
Edit
Yuriy M
Ireland
Amazon Machine Learning
Entities
Create new...
Actions
Refresh
Filter: All types
Entity name or ID
Items per page: 10
1 - 10 of 12 Entities
Name
Type
ID
Status
Evaluation: ML model: MyTest1 Evaluation ev-5ovtgAabi84 Pending
ML model: MyTest1 ML model ml-F6YKIBbahKI Pending
MyTest1_[percentBegin=70, ... Datasource 85461fc2-e0d0-4f16... In progress Apr 15, 2016 4:17:29 PM
MyTest1_[percentBegin=0, p... Datasource f9589690-9234-4b8... In progress Apr 15, 2016 4:17:28 PM
MyTest1 Datasource
banking_predictions Batch prediction
Banking-batch.csv Datasource ds-yU7WksZKJVQ Completed Mar 5, 2016 3:49:08 PM
Evaluation: ML model: try Evaluation ev-iJemKgQ3Akq Completed Mar 5, 2016 3:10:48 PM
Performans raporu hazırlanması
Modelin eğitimi
Eğitim ve doğrulama veri kaynakların otomatik oluşturulması



Performans raporunu inceleyin

Amazon ML tutorial

- Performans değerlendirme raporunda, modelin tahmin etme kalite ve sınıflandırma problemleri için eşikin seçilme işlem yapılır

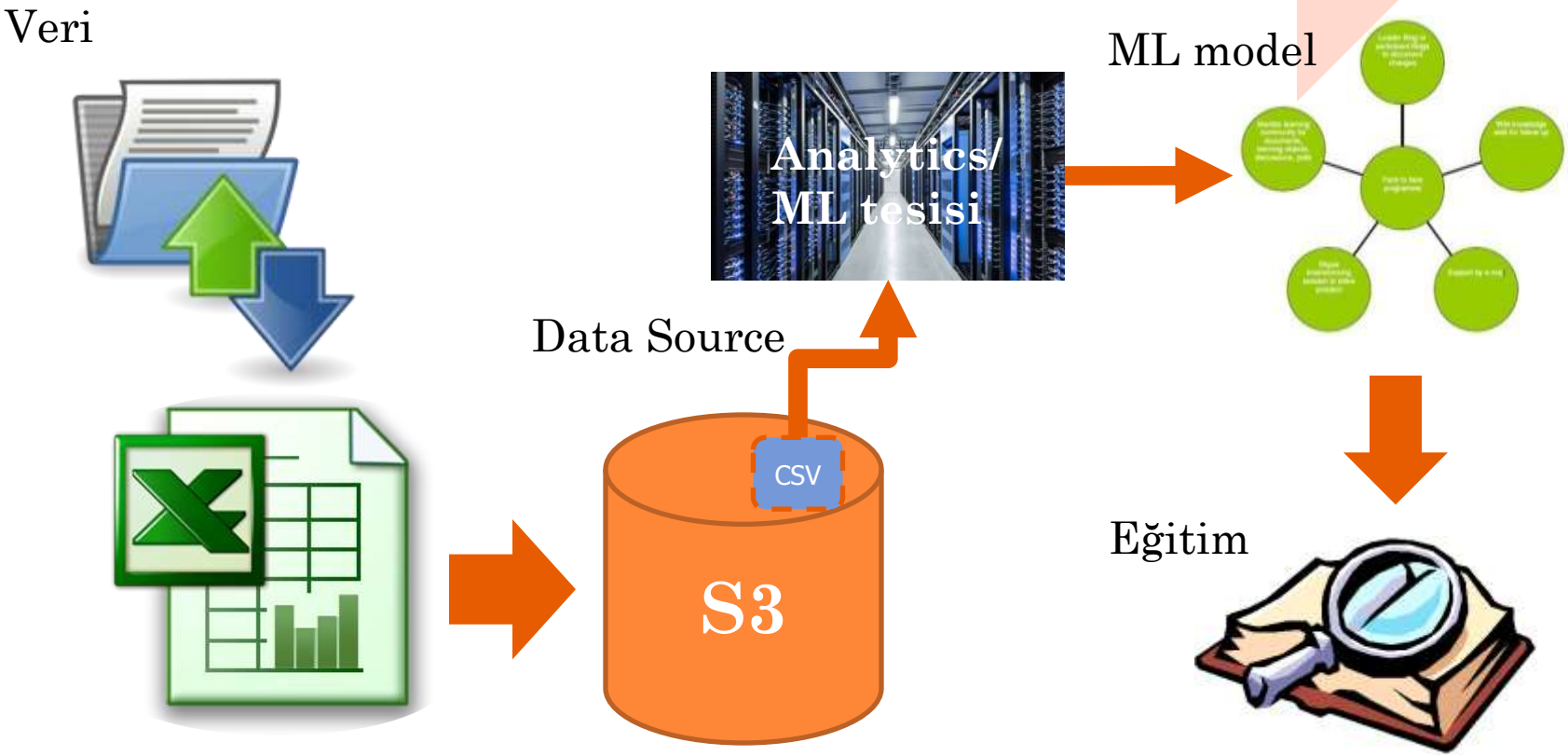

https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#
ML model report
Summary
Try real-time predictions
Settings
Evaluations
Evaluation:
Summar
Alerts (0)
Explore p
Evaluation Summary
Delete this Evaluation
ID ev-5ovtgAabi84
Name Evaluation: ML model: MyTest1
Datasource ID 85461fc2-e0d0-4f16-82d3-97d70d95d7c8
Output location Not available
Creation time 04/15/16 16:17:29
Status Completed
Message Not available
Log Download log
ML model performance metric
On your most recent evaluation, ev-5ovtgAabi84 , the ML model's quality score is considered extremely good for most machine learning applications.
0
1
AUC: 0.94
Baseline AUC: 0.50
Difference: 0.44
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

Veri ler, CVS tablosu olarak hazırlanır

Veri CVS dosyası, Amazon S3'e yüklenir

Amazon S3'deki veri, Analytics/ML tesisine veri kaynağı üzerinde bağlanır

ML tesisinde ML modeli oluştulur ve eğitime verilir



CSV verileri: sütün – bir öznitelik,
satır – bir örnek

ML MODELLERININ KALITE ÖLÇEKLERI

- Bir durumu açıklayan ML modelin kalitesini belirlemek için, makine öğrenmede iki yaklaşım var dır
  - Öğrenimde kullanılan *maliyet fonksiyonu* direk kullanan modelin kalite ölçekleri
  - Modelin performansı ayrı bir şekilde karakterize eden ölçekleri

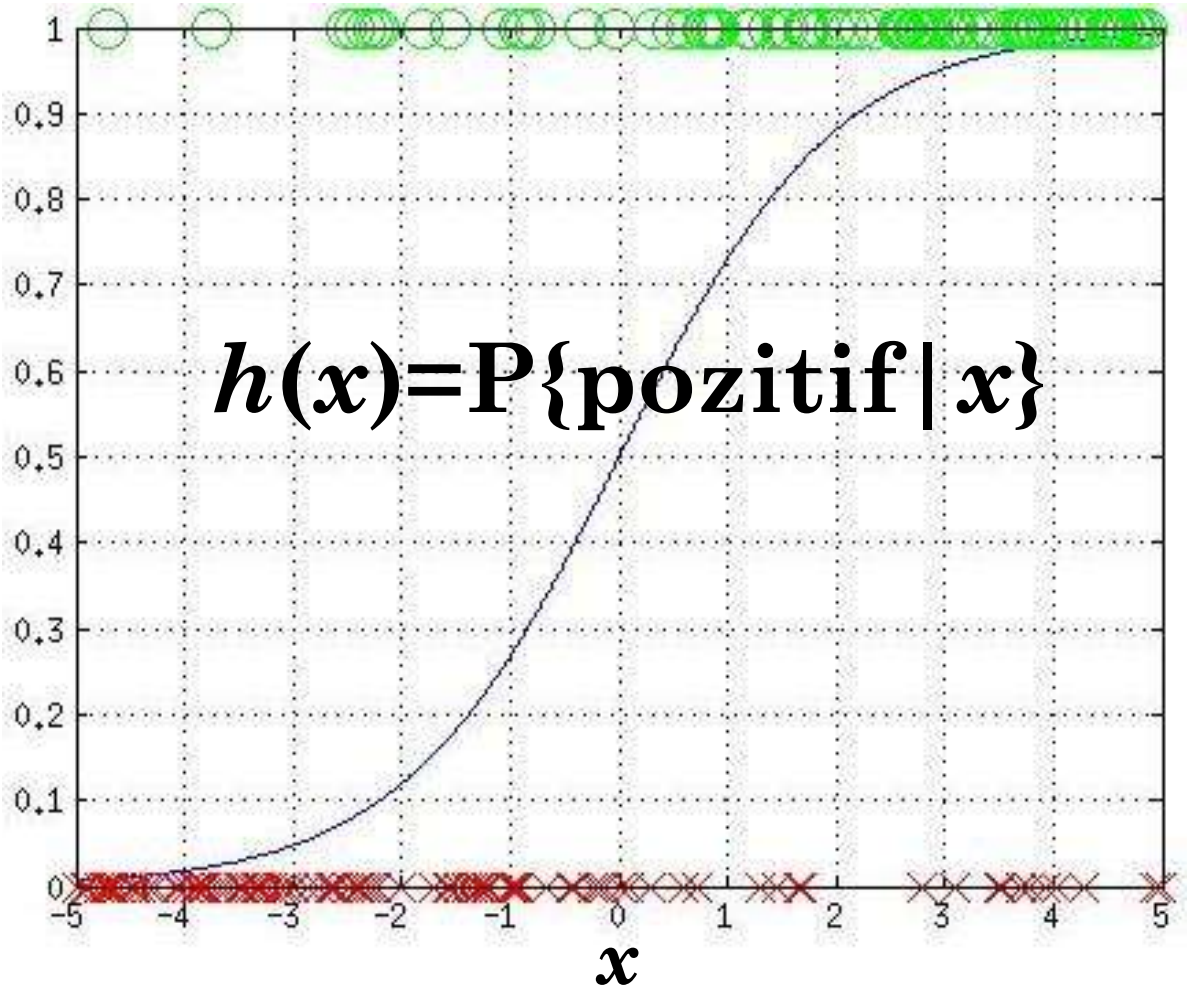
Modellerin performans değerlendirilmesi - sınıflandırma

- Sınıflandırma problemlerinde biz öncelikle ikili sınıflandırma problemi düşünüyoruz (pozitif veya negatif, 0 veya 1, evet veya hair, olur veya olmaz)
- Çok sınıflı sınıflandırma problemleri tipik olarak birkaç ikili sınıflandırma problemine dönüştürmek mümkündür

Modellerin performans değerlendirilmesi - sınıflandırma

- İkili sınıflandırmada çok tipik bir durum öyle ki, h-modeli sürekli ve farklı durumların bir yada diğer sınıfa ait olma olasılığı olarak düşünülür

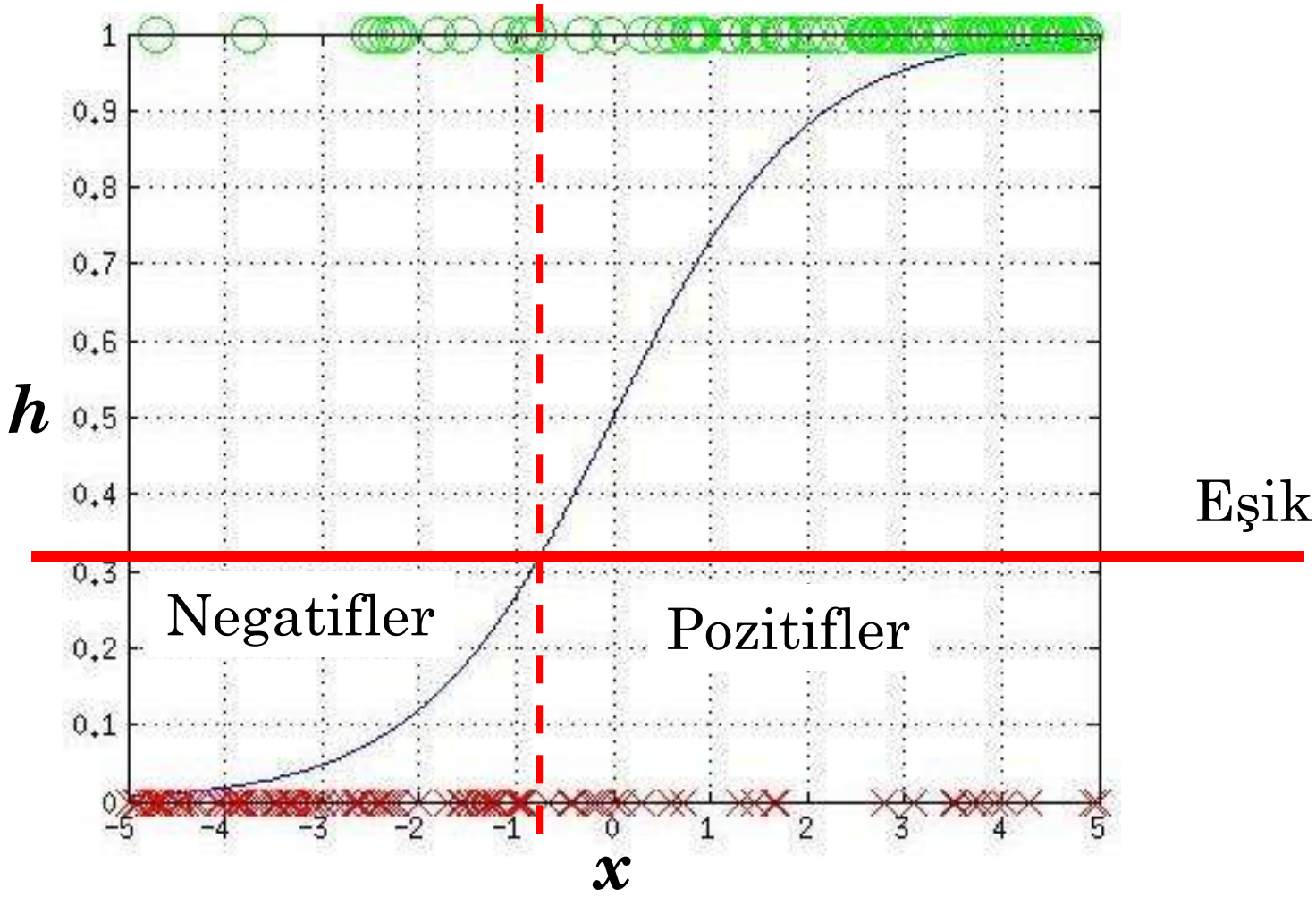
Modellerin performans değerlendirilmesi - sınıflandırma



Modellerin performans değerlendirilmesi - sınıflandırma

- Sınıflandırma için, bir x'in $h(x)$ belirli eşik (threshold) ile karşılaştırılır ve buna göre x'in sınıfı belirtilir

Modellerin performans değerlendirilmesi - sınıflandırma



Modellerin performans değerlendirilmesi - sınıflandırma

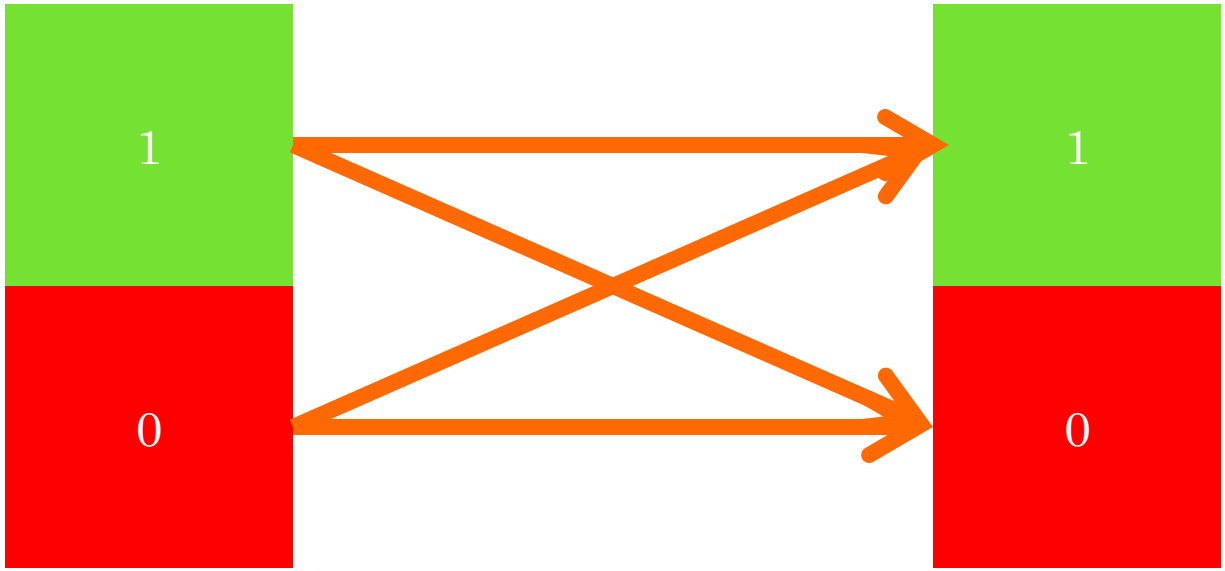
- Bu şekilde, *h*-modeli kendisi aslında örnek durumların bir sınıfa ait olma olasılığı modelidir
- Böyle "olasılık" modeli seçmek için kullanılan *maliyet fonksiyonu* direk olarak sınıflandırma performansını ifade etmiyor, ayrıca sınıflandırmanın performansı belli ki eşike bağlı dır

Modellerin performans değerlendirilmesi - sınıflandırma

- Sınıflandırma problemlerinde elde edilen modellerin farklı performans ölçekleri kullanılır, bunlar – *True Positive, True Negative, False Positive, False Negative skorları, Accuracy, Precision, Recall, F1 measure, Response-Operation curves, Precision-Recall curves, AUC measure, ve Confusion Matrix* tir
- Bunlar var olan makine öğrenme paketler tarafından raporlanıyor ve sınıflandırmanın performansını farklı taraftan karakterize ediyor

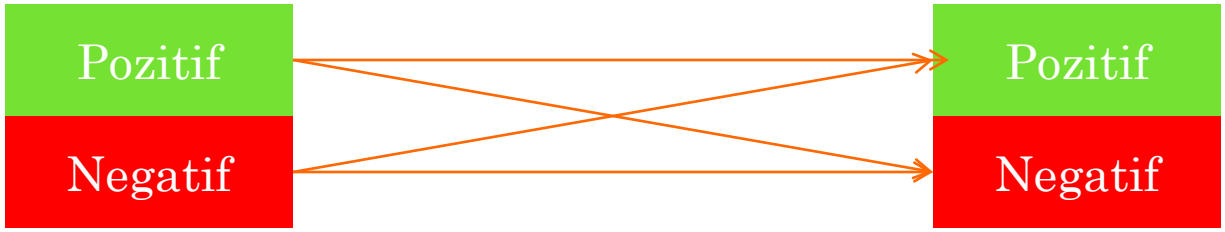
Modellerin performans değerlendirilmesi - sınıflandırma

- İkili sınıflandırmada durumlar iki sınıfa ayırdırılıyor ve kendileri iki sınıftan gelebilmektedir



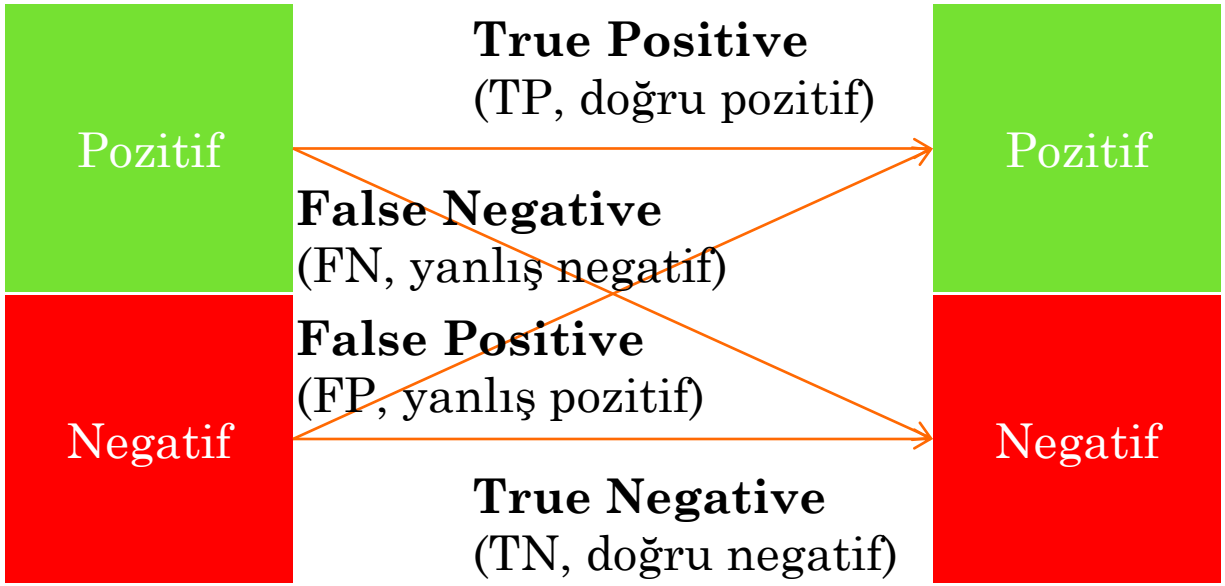
Modellerin performans değerlendirilmesi - sınıflandırma

- Bu iki sınıfa gelellikle “pozitif” ve “negatif” denir
- Buna göre sonuçlarda 4 durum gözlenebilir
  - Pozitif örnekler “pozitif” olarak sınıflandırılmış
  - Pozitif örnekler “negatif” olarak sınıflandırılmış
  - Negatif örnekler “pozitif” olarak sınıflandırılmış
  - Negatif örnekler “negatif” olarak sınıflandırılmış



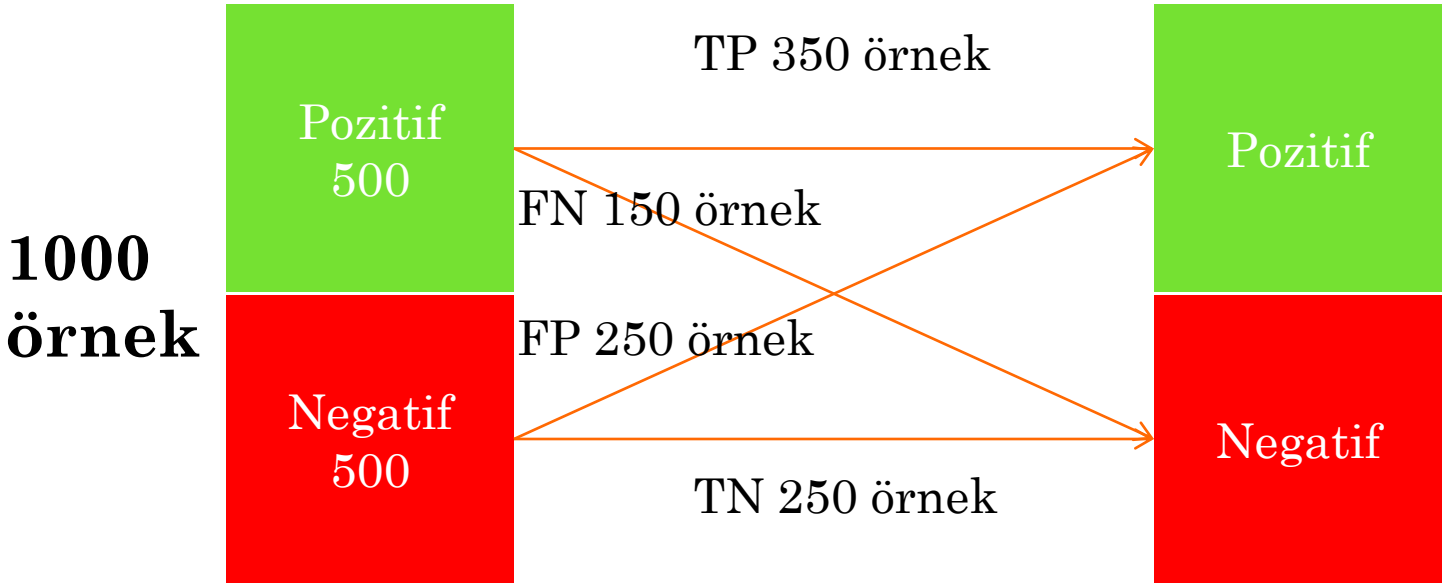
Modellerin performans değerlendirilmesi - sınıflandırma

- Bu 4 durum ilgili şekilde etiketlenir



Modellerin performans değerlendirilmesi - Sınıflandırma

- Bu 4 durum ilgili şekilde etiketlenir



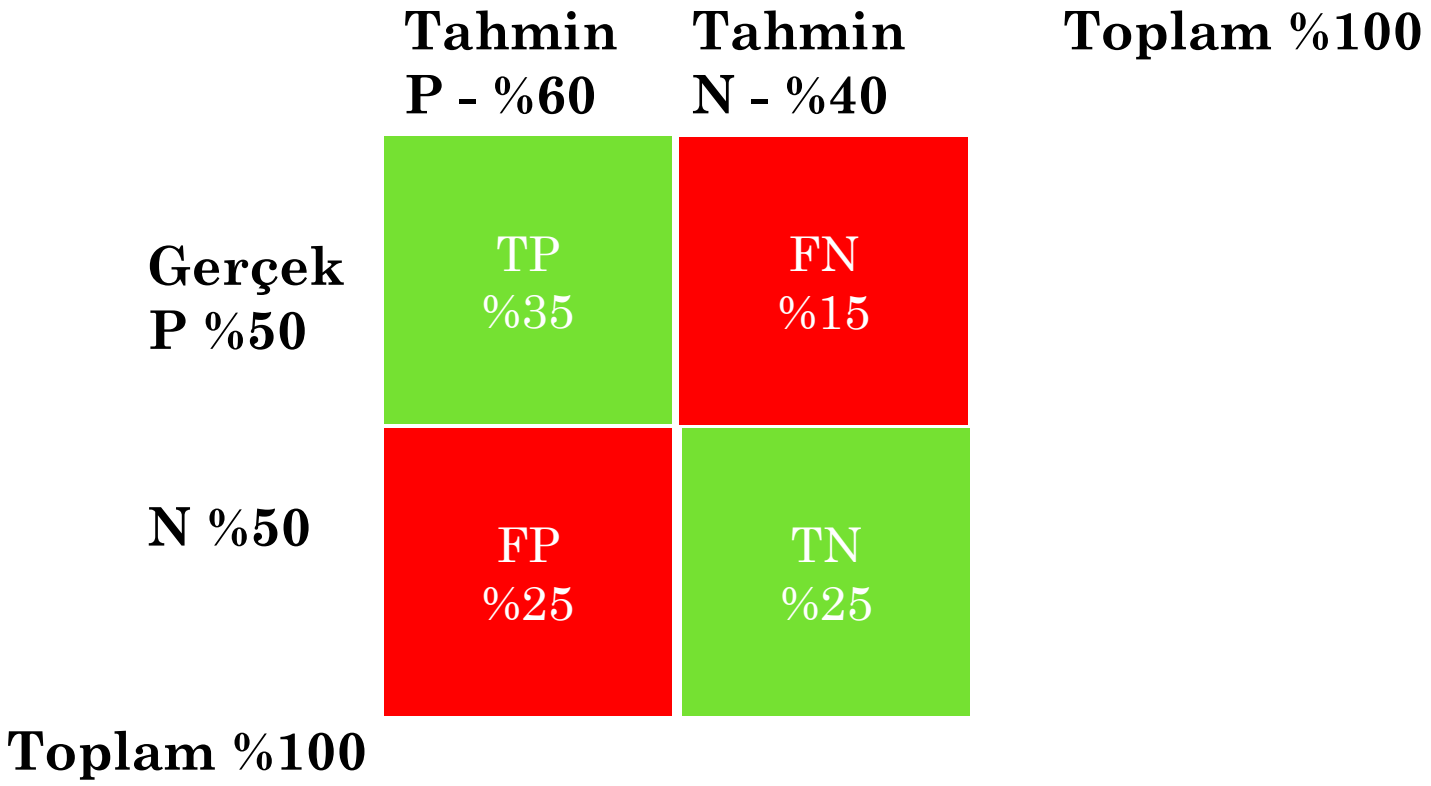
Modellerin performans değerlendirilmesi - sınıflandırma

- İşte **karıştırma matrisi (confusion matrix)**:

	Tahmin P - 600	**Tahmin N - 400**	**Toplam 1000 tahmin**
Gerçek P - 500	TP 350	FN 150	
Gerçek N - 500	FP 250	TN 250	
Toplam 1000 örnek			

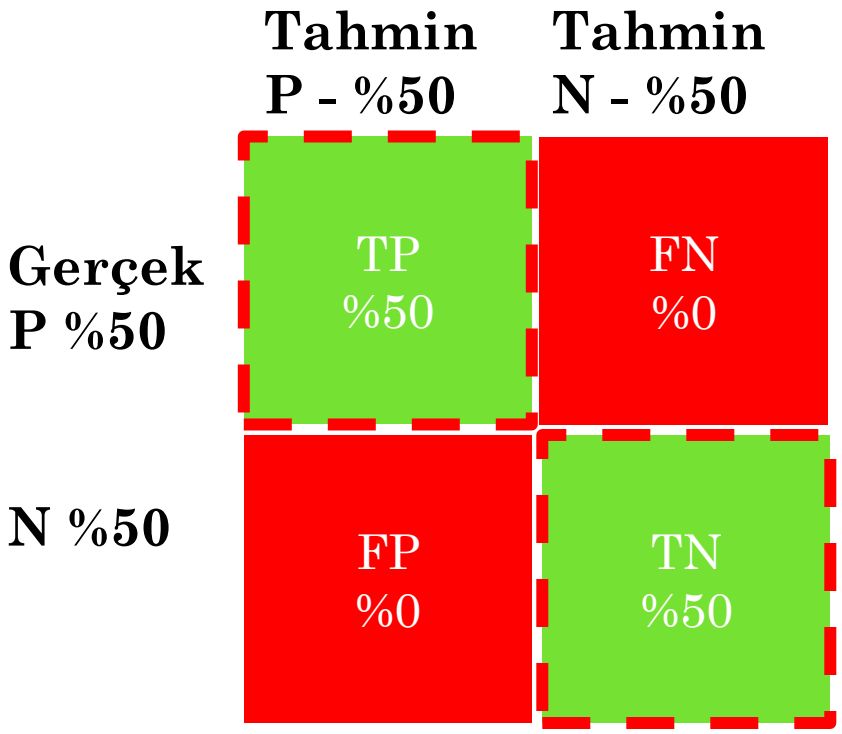
Modellerin performans değerlendirilmesi - sınıflandırma

- Tabi ilgili durumlar yüzdeleri olarak genelde kodlanır



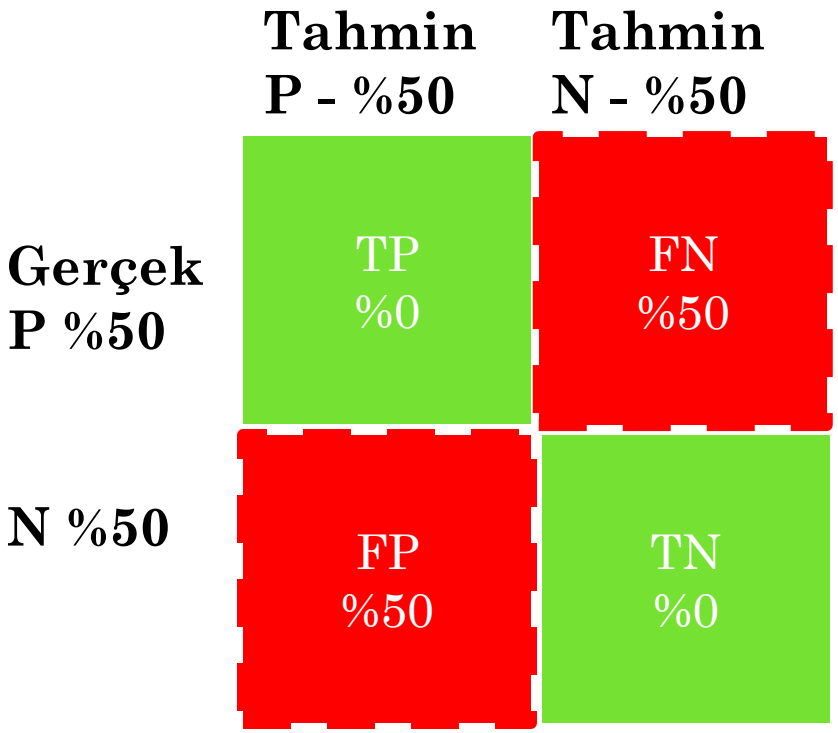
Modellerin performans değerlendirilmesi - Sınıflandırma

- İdeal durum: TP=P ve TN=N



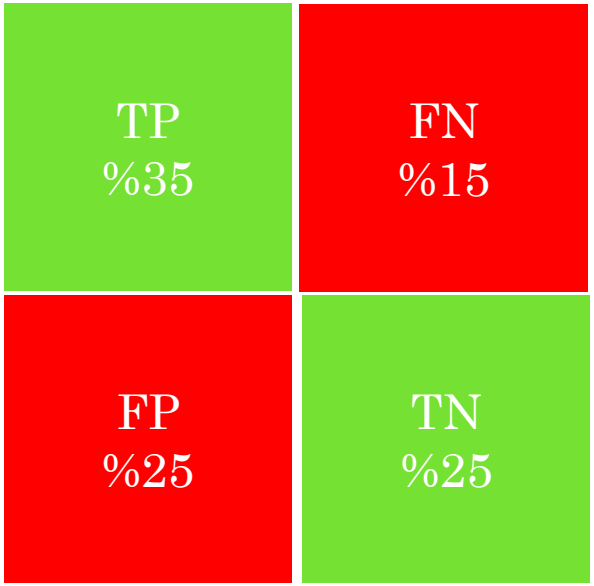
Modellerin performans değerlendirilmesi - Sınıflandırma

- FN ve FP iki tür hatayı ifade eder
  (*bu arada fark edelim ki TP+FN+FN+TN=%100*)



Modellerin performans değerlendirilmesi - Sınıflandırma

- **TP, FN, FP ve TN** yüzdeleri sınıflandırma modelin performansını tamamen belirtirler



Modellerin performans değerlendirilmesi - Sınıflandırma

- Özel ölçekleri:
  Accuracy (Doğruluk) = TP+TN

Accuracy = %60

Modellerin performans değerlendirilmesi - sınıflandırma

- **Doğruluk=TP+TN**, modellerin "genel" (yani P ve N ikisi üzerinde) doğruluğu belirtir
- Doğruluk, modelin genel olarak doğru olup olmadığını karakterize eder; dolayısıyla doğruluğu artırmak genel olarak iyi dir

Modellerin performans değerlendirilmesi - sınıflandırma

- Özel ölçekleri:
  Recall (Anma) = TP/P

Recall = %70

Modellerin performans değerlendirilmesi - sınıflandırma

- **Anma=TP/P**, modellerin pozitif durumları algılayabilmesi belirtir
- Baze durumlarda “pozitif” durumları yakalamak son derecede önemli halbuki yanlışlıkla pozitif olarak tespit edildiği negatif durumlar okadar önemli değil – örneğin, hastalar sırasında kanser durumları tespit etmek
- Bu durumlarda anmaya odaklanmak faydalı dır

Modellerin performans değerlendirilmesi - Sınıflandırma

- Özel ölçekleri:
  Precision (Kesinlik) = TP/(TP+FP)

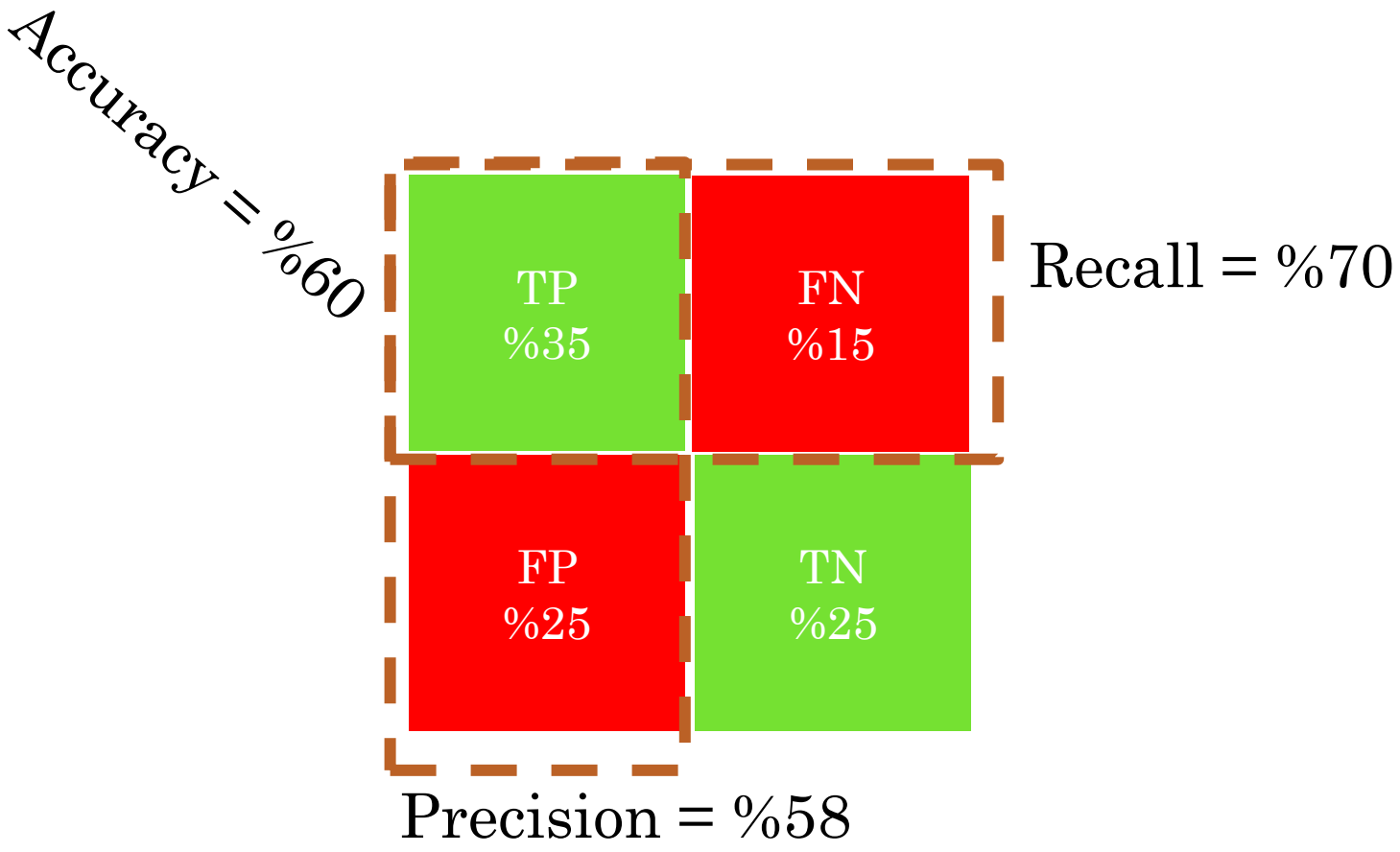
Precision = %58

Modellerin performans değerlendirilmesi - sınıflandırma

- **Kesinlik=TP/(TP+FP)**, modellerin pozitif tahminleri sırasında gerçekten pozitif durumların oranı belirtir
- Bazen "pozitif" olarak tahmin edilen durumlara gerçekten emin olmamız gerekiyor halbuki pozitif durumların yakalanmaması büyük sorun değil – mesela, telefonla hedeflenmiş reklam yaparken yanlış insanları aramamak oldukça isteyebiliriz
- Bu durumlarda kesinliğe odaklanmak faydalı dır

Modellerin performans değerlendirilmesi - Sınıflandırma

- Tüm sınıflandırma performans ölçekleri



Modellerin performans değerlendirilmesi - Sınıflandırma

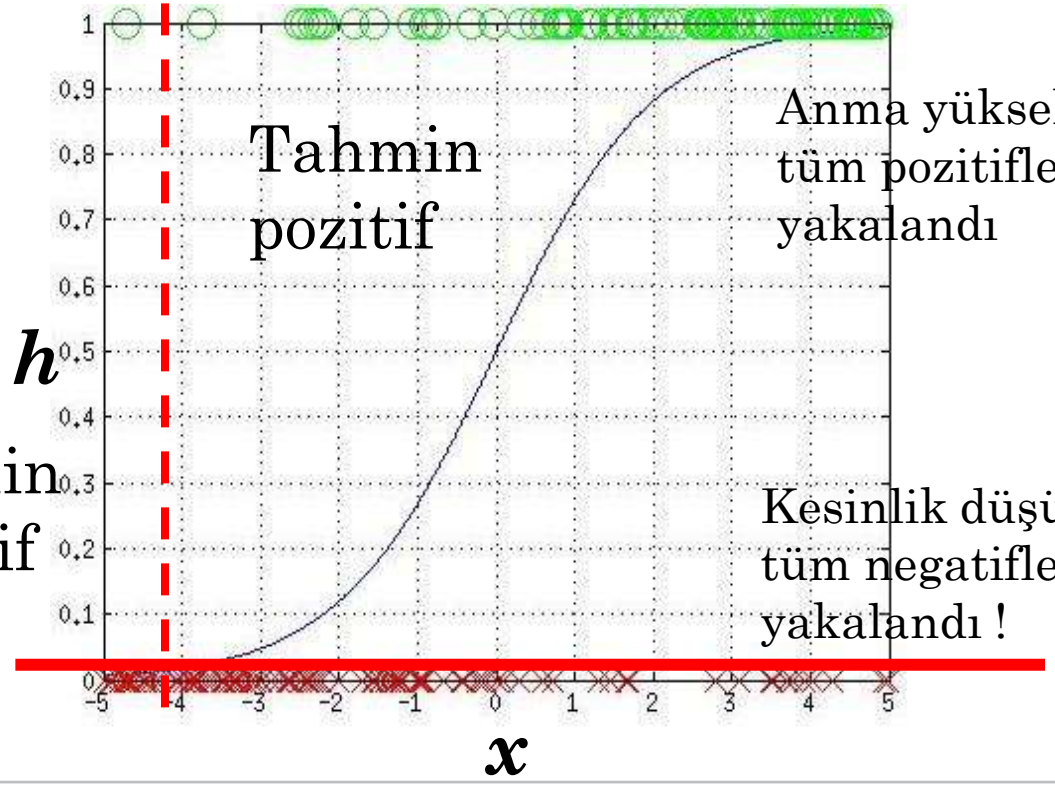
- Sınıflandırma çözümü için eşik seçilmesi ile ilgili baze ayrı ölçekleri kullanılır
- Bunlar ROC (response-operation curve) ve PRC (precision-recall curve) eğrileri

Modellerin performans değerlendirilmesi - sınıflandırma

- Sınıflandırma çözümlerinin eşiği değişken genelde TP, FN, FP, TN ve anma ve kesinlik gibi ölçekler de değişir

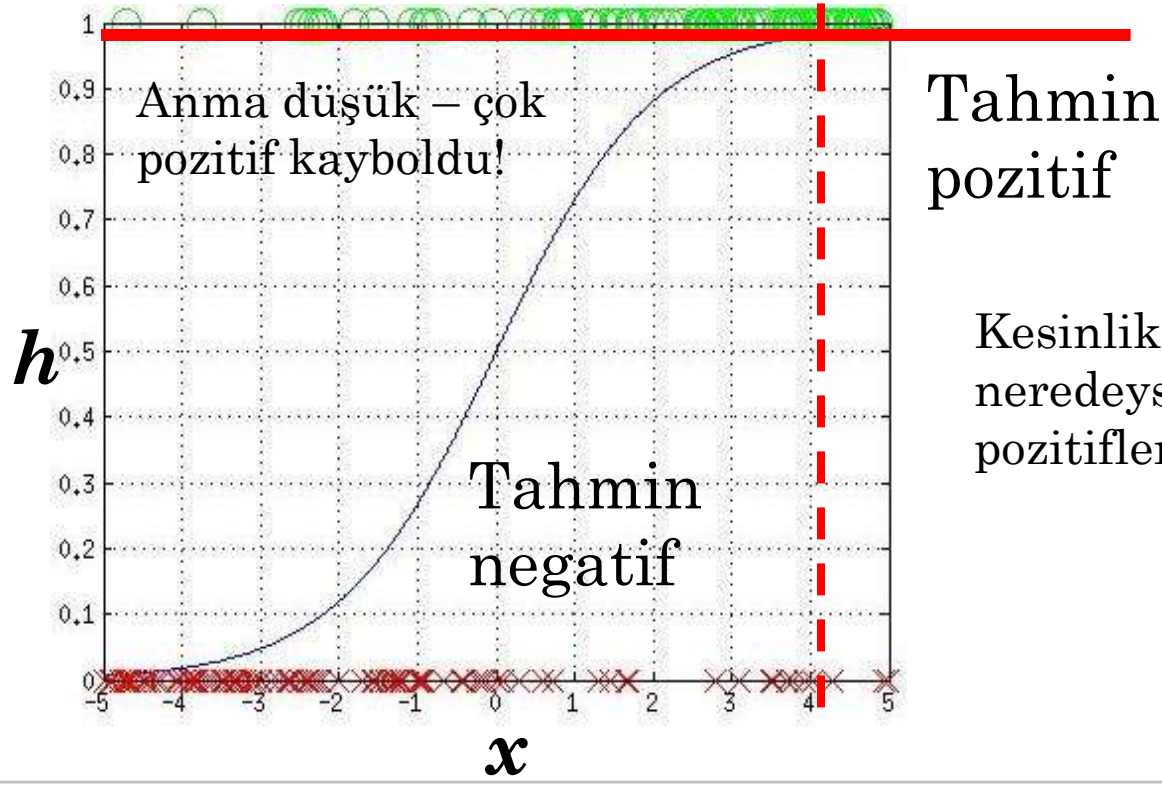
Modellerin performans değerlendirilmesi - sınıflandırma

- 0'a yakın eşik demek *anması* yüksek (%100) fakat *kesinliği* düşük (%P)



Modellerin performans değerlendirilmesi - sınıflandırma

- Aynı zamanda, 1'e yakın eşik demek *kesinlik* yüksek (%100) fakat *anma* düşük (%P)



Modellerin performans değerlendirilmesi - sınıflandırma

- Aynı zamanda, eğer h-modeli 1'e yakın eşikle kullanırsak – yani $h(x)>0.999...$ iken durumlar pozitif olarak tahmin edersek – böyle sınıflandırma çözümün kesinliği yüksek (%100) fakat anması düşük (%0) olacak

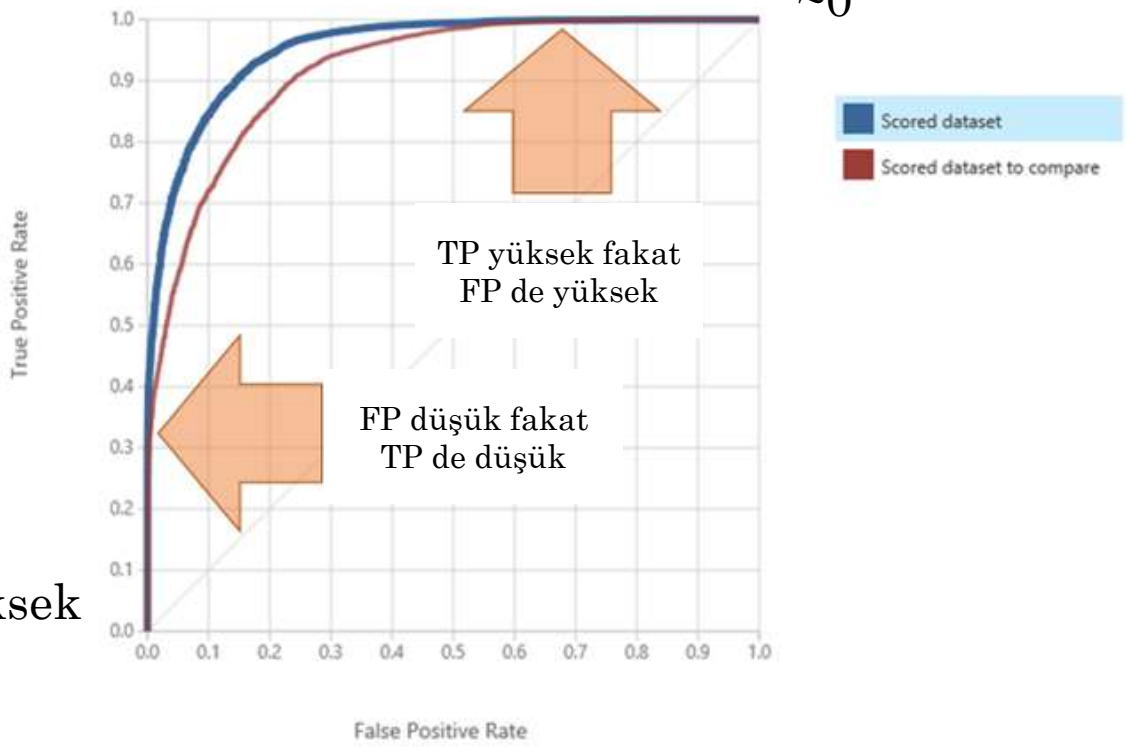
Modellerin performans değerlendirilmesi - sınıflandırma

- Buna göre *h*-modeli ile beraber kullanılan eşik 0'dan 1'e değişiyor olup, buna göre TP, TN, FP, FN gibi ölçekler bir taraftan diğer tarafa tamamen değişebilir

Modellerin performans değerlendirilmesi - sınıflandırma

- **ROC eğrisi**, eşik değişimine göre TP ve FP değiştiğini gösteriyor

Eşik düşük
~0



Eşik yüksek
~1

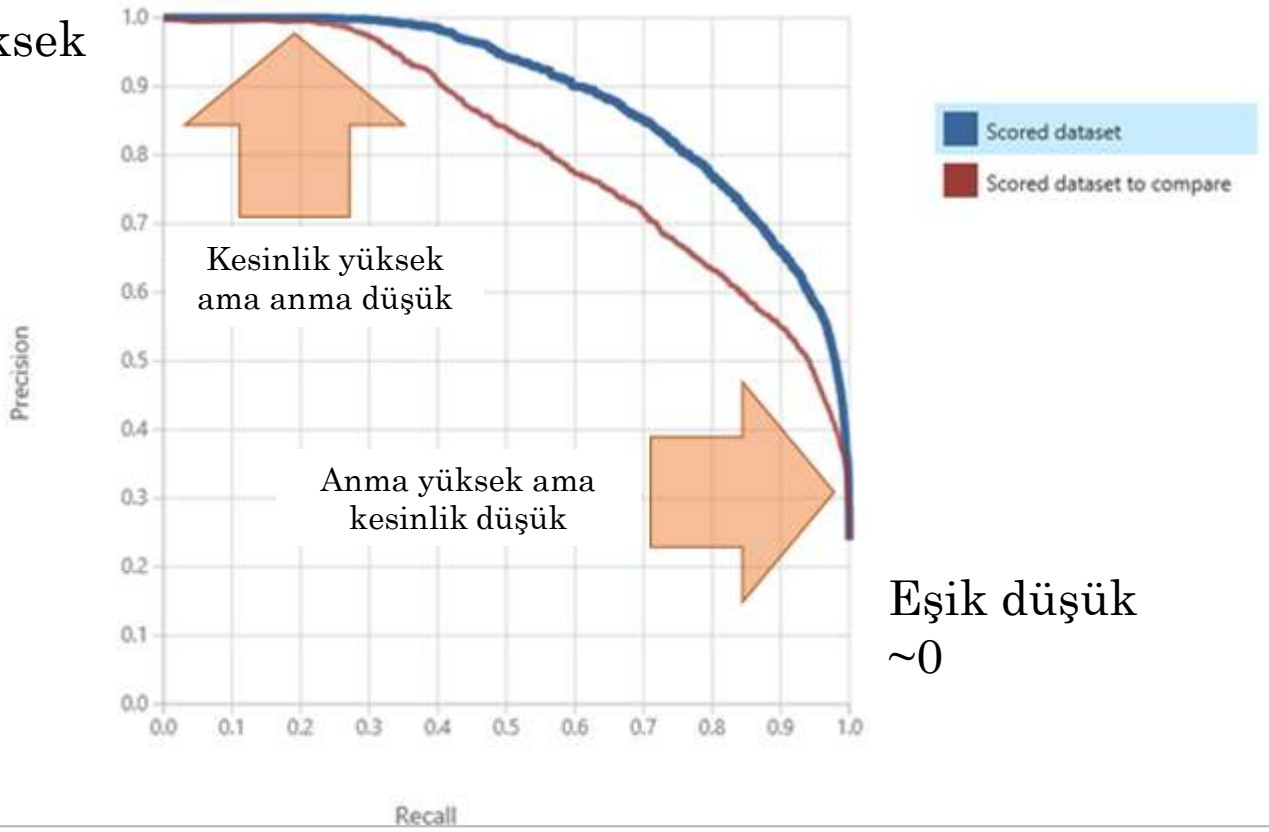
Modellerin performans değerlendirilmesi - sınıflandırma

- ROC eğrisi monoton olarak yukarı gider
- Sol alt koşe yüksek eşike karşılıklı gelir (h~1)
- Sağ üst koşe düşük eşike karşılıklı gelir (h~0)
- Yüksek eşik demek ki, biz sadece gerçekten doğru durumlar pozitif olarak yakalıyoruz (geçme eşiği yüksek) fakat fazla pozitif durum kaybediyoruz (geçme eşik yüksek)
- Düşük eşik demek ki, biz neredeyse tüm durumlar pozitif olarak etiketliyoruz, dolayısıyla anma yüksek (tüm pozitif durumlar yakaladık tabi) fakat fazla negatif durum yanlışlıkla yakalıyoruz (geçme eşik düşük)

Modellerin performans değerlendirilmesi - sınıflandırma

- **PRC eğrisi**, eşik değişimine göre Precision (kesinlik) ve Recall (anma) değiştiğini gösteriyor

Eşik yüksek ~1



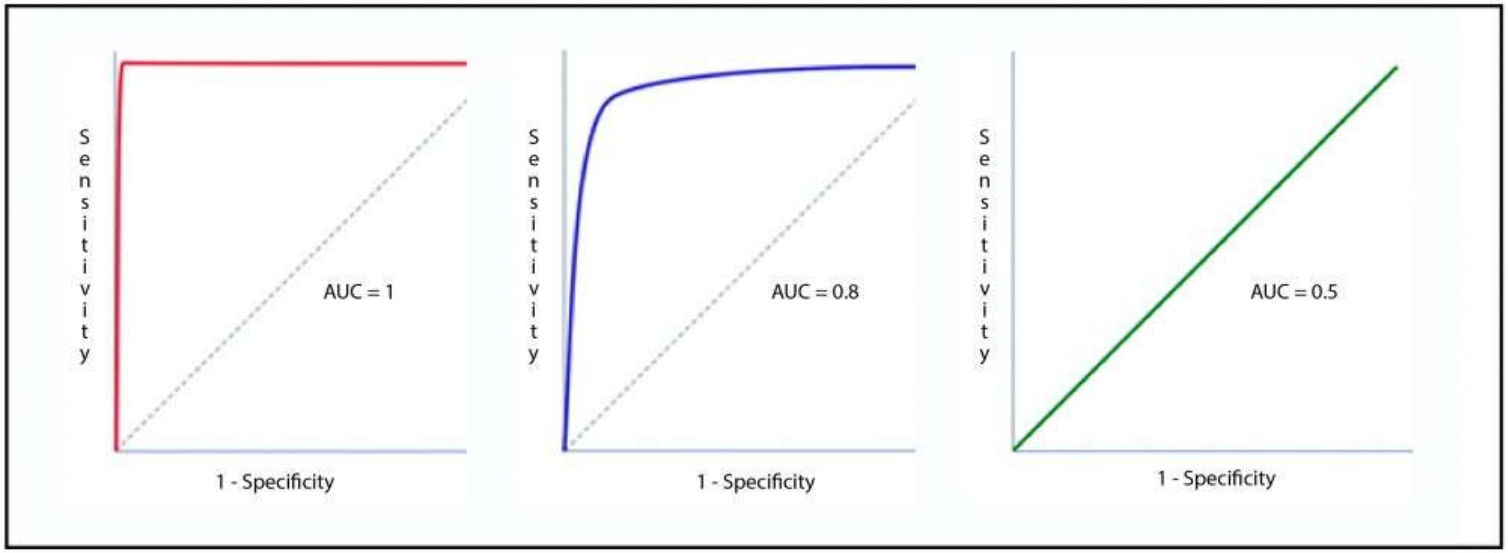
Eşik düşük ~0

Modellerin performans değerlendirilmesi - sınıflandırma

- PRC eğrisi monoton aşağa gider
- Sol üst koşe yüksek eşike karşılıklı gelir (h~1)
- Sağ alt koşe düşük eşike karşılıklı gelir (h~0)
- Yüksek eşik demek biz gerçekten doğru durumlar tek pozitif olarak değerlendiriyoruz (geçme eşik yüksek) fakat fazla pozitif durum yakalayamıyoruz (eşik yüksek!)
- Düşük eşik demek biz neredeyse tüm durumlar pozitif olarak değerlendiriyoruz, dolayısıyla anma yüksek (geçme eşik düşük!) fakat fazla negatif durum yanlışlıkla yakalıyoruz (eşik düşük)

Modellerin performans değerlendirilmesi - sınıflandırma

- Bir sınıflandırma çözümünün genel performansını karakterize etmek için Area Under Curve – Eğri Altı Alanı – veya AUC ölçeği kullanılır



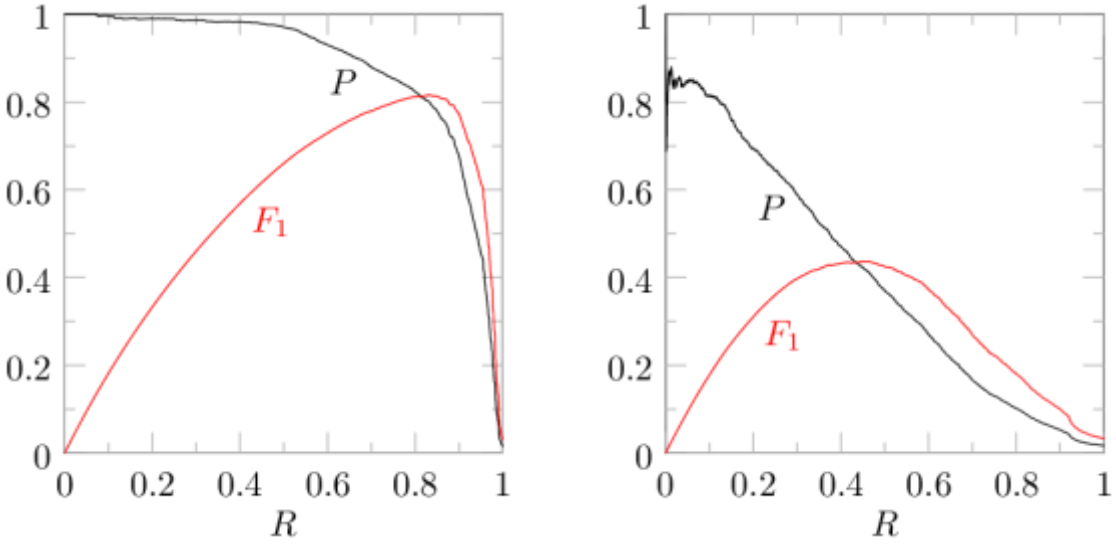
Modellerin performans değerlendirilmesi - Sınıflandırma

- ROC veya PRC eğrilerinden eşiği seçmek için F1 skoru genelde kullanılır

F1=2*Anma*Kesinlik/(Anma + Kesinlik)

Modellerin performans değerlendirilmesi - sınıflandırma

- Eşiğe bağlı F1 skoru, 0..1 arasında değişir ve yüksek F1 skoru genelde aynı zamanda anma ve kesinlik yüksek olduğunu belirtir
- Böylece, eşiği seçmek için F1 skoru maksimuma çıkartan eşik çok sık seçilir



Modellerin performans değerlendirilmesi - regresyon

- Regresyon problemlerinde kalite, maliyet fonksiyonu direk kullanılarak ifade edilir

Modellerin performans değerlendirilmesi - regresyon

- Regresyon problemlerinde modelin tahmini ve gerçek değer arasındaki farka hata denir

$$\varepsilon_i = y_i - h(x_i; \theta)$$

Modellerin performans değerlendirilmesi - regresyon

- Böyle ise, maliyet fonksiyonu modelin tahminlerinin ortalama hatasıdır

$$J(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left| \varepsilon_i \right|^2$$

Modellerin performans değerlendirilmesi - regresyon

- Standard ismi – **ortalama karesel hatası (Mean-Square Error veya MSE)**

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left|h(x_i;\theta) - y_i\right|^2$$

Modellerin performans değerlendirilmesi - regresyon

- MSE en tipik kullanılan regresyon modellerin tahmin edebilme ölçeğidir

Modellerin performans değerlendirilmesi - regresyon

- Sabit modelin MSE'sine **varians** denir *(yani y'leri her zaman aynı olarak tahmin eden model)*

$$VAR = \frac{1}{n}\sum_{i=1}^{n} \left| m - y_i \right|^2$$

Modellerin performans değerlendirilmesi - regresyon

- Makine öğrenmenin modellerinin MSE'leri garantili sabit modelinden düşüktür (genellikle); bu durumda modelin performansı **kalan varyans yüzdesi (fraction of varians unexplained)** olarak da değerlendirilebilir

$$FVU = MSE / VAR$$

Modellerin performans değerlendirilmesi - regresyon

- Böyle ise, **açıklanan varyans (explained varians)** model sayesinde sabit modeline göre tahminlerin hatalarında azalışa denir, kalan tahmin hatalarına **açıklanmayan varyans (unexplained varians)** denir
- Açıklanan varyans yüzdesine de, **belirtme katsayısı** (coeffisient of determination) denir ve R^2 sembol ile belirtilir

$$R^2 = 1 - MSE / VAR$$

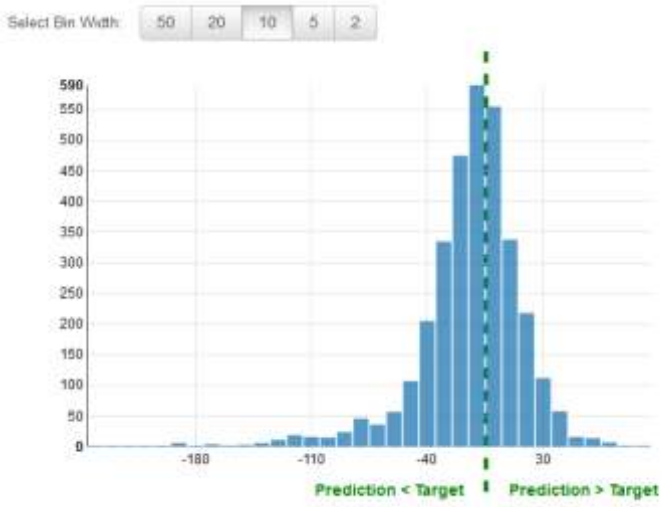
Modellerin performans değerlendirilmesi - regresyon

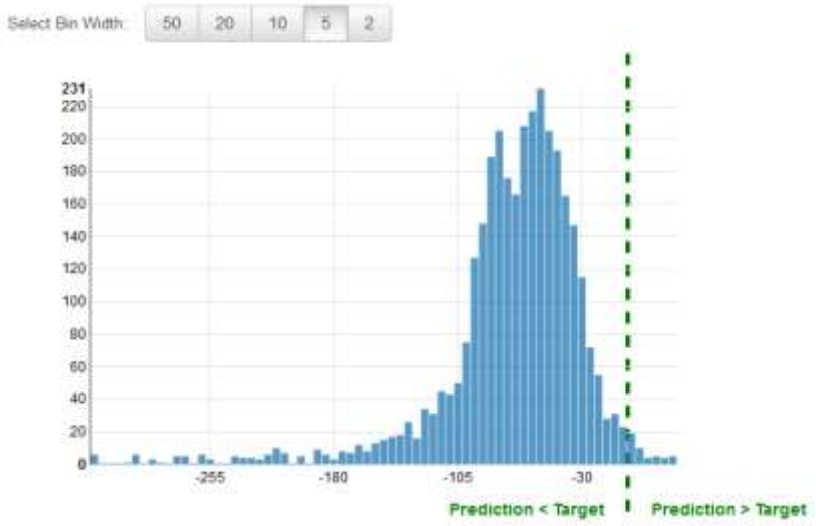
- Modelin hatalarına, **hatalar (errors)**, **kalanlar (residuals)** ve **yenilik (innovation)** de denir

$$\varepsilon_i = y_i - h(x_i; \theta)$$

MODELLERIN PERFORMANS DEĞERLENDIRILMESI - REGRESYON

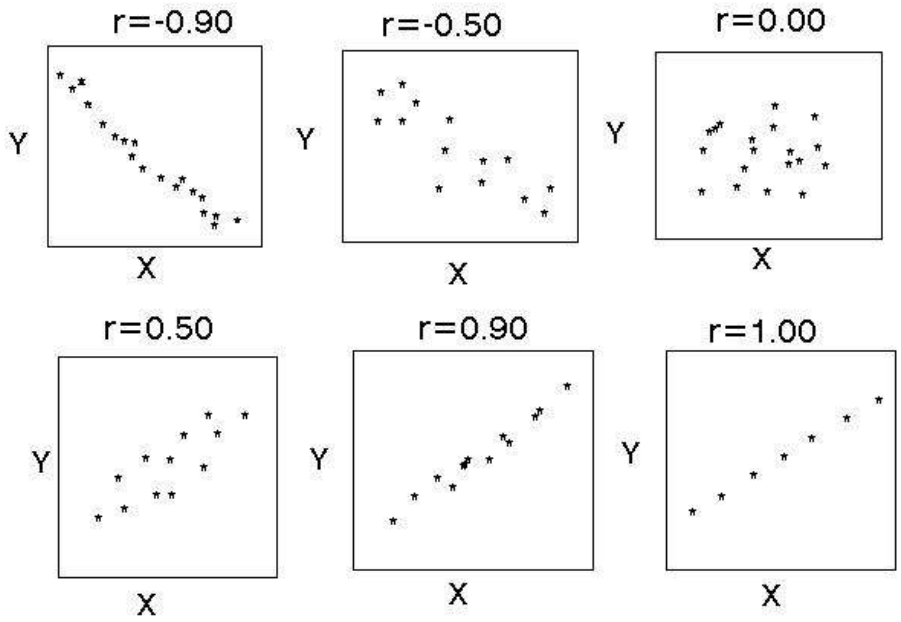
- Hataların daha ayrıntılı ölçeği, hatalar/kalanların dağılımıdır; kalanların dağılımı hataların nekadar büyük ve ne şekilde olması hakkında bilgi sağlar





Modellerin performans değerlendirilmesi - regresyon

- Regresyon modeller için başka da bir ölçek, r^2 sembol ile belirtilen **korelasyon katsayısı (correlation coefficient)**



Modellerin performans değerlendirilmesi - regresyon

- Korelasyon katsayısı, modelin tahminleri ve gerçek niteliğin değerleri doğru orantılı olup olmadığı hakkında bilgi sağlar, fakat eşitliği ima etmez

$$y_i = \frac{1}{2} x_i \Rightarrow corr\ (y, x) = 1$$

Modellerin performans değerlendirilmesi - regresyon

- **MSE**, model tahminlerinin tipik hatasına kare-ortalama anlamında ifade eder ve en çok kullanılan regresyon kalite ölçeği dir
- MSE tabi herzaman pozitif olur (MSE>0); bundan hariç fakat başka her hangi bir kısıtı yok (*yani 1 olabilir, 10 olabilir, 100 de olabilir*) – bu nedenle MSE bir değeri doğru yorumlamak için onun bir tipik değerine ihtiyacı var
- Böyle "tipik" değeri olarak "sabit" modelin varyansı en çok kullanılır ve MSE kendisi onun yüzdesi olarak FVU veya R^2 olarak ifade edilir

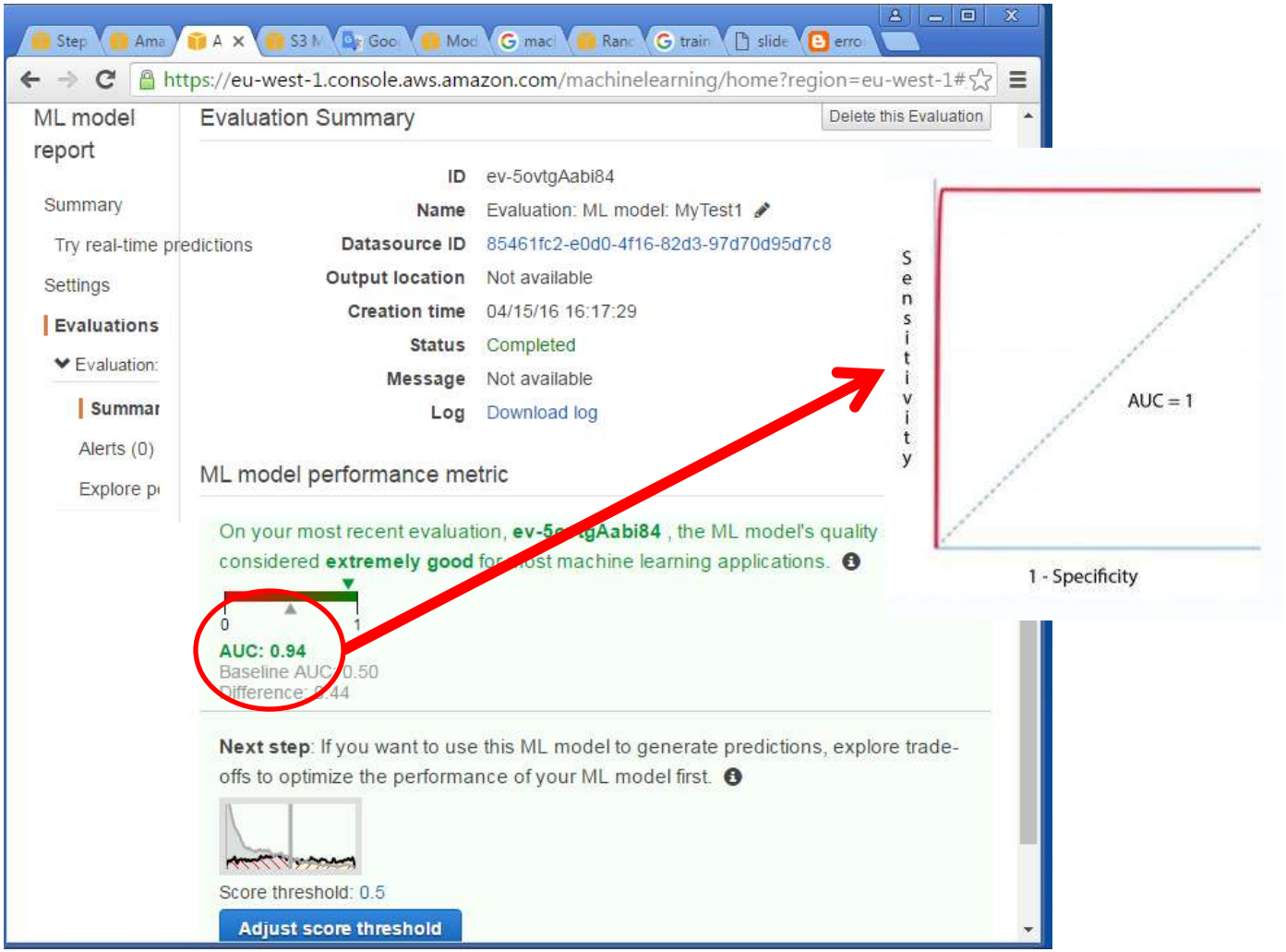
Modellerin performans değerlendirilmesi - regresyon

- **FVU** oranı 0 ve 1 arasında değişiyor olup, sıfıra eşitliği kalan varyans yok ve model eksiksiz demek, bire eşit ise modelin bilgisiz ve sabit tahminlerden daha iyi olmamasını demek tir
- **R^2** tam ters şekilde, sıfıra eşit ise modelin sabit tahminlere göre hata azalışı sağlamaması, bire eşit ise, tahminlerin tam doğru olmasını ve verinin varyansının %100 anlatmasını demek tir

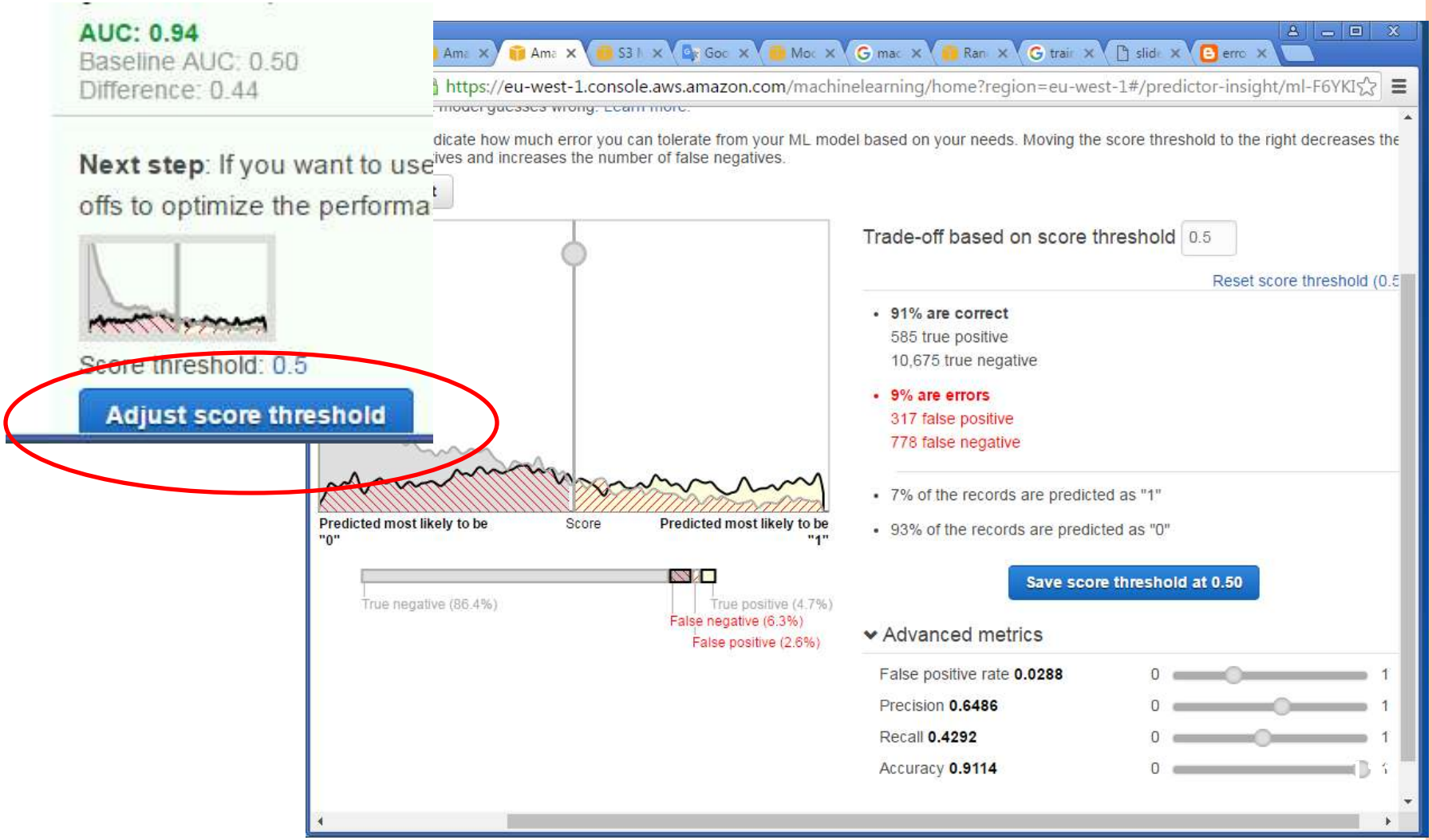
Modellerin performans değerlendirilmesi - regresyon

- **Korelasyon katsayısı r^2**, model tahminlerinin ve gerçek hedef değerlerinin beraber veya doğru orantılı olup olmadığını ifade eder
- r^2, -1...1 arasında değişiyor olup, -1'e eşitse tahminlerin gerçeğin tam ters olması, 0'a eşitse tahminlerin gerçek değerlerle alakasız olması, +1'e eşit ise tahminlerin gerçek değerler tam gibi değiştiği demek tir

Amazon ML tutorial

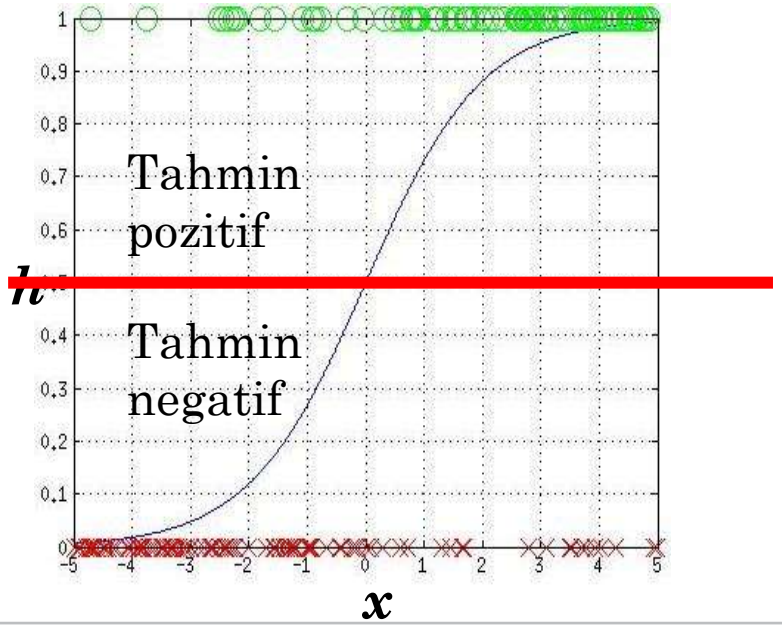


Amazon ML tutorial



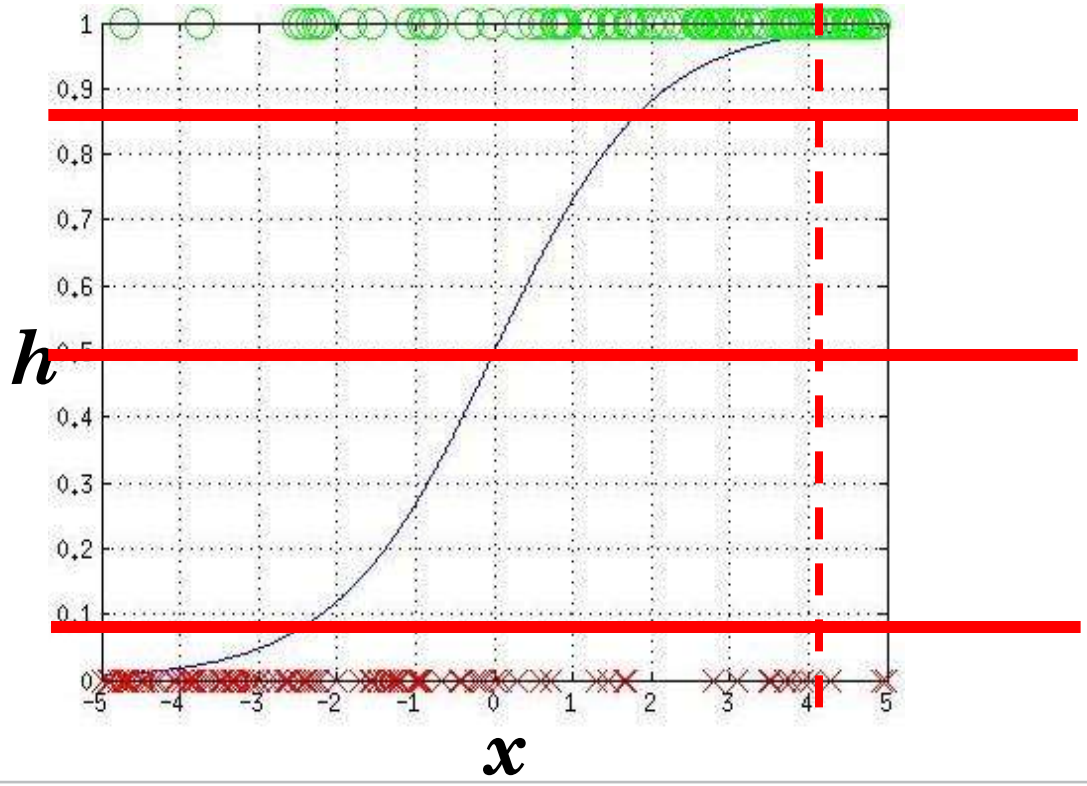
AMAZON ML TUTORIAL

- Sınıflandırma problemlerinde ML model iki durumu ayıran bir "olasılık" değeri veriyor
- Sonuç durumların sınıflara konulması aslında bu olasılıkla beraber bir eşike göre yapılıyor



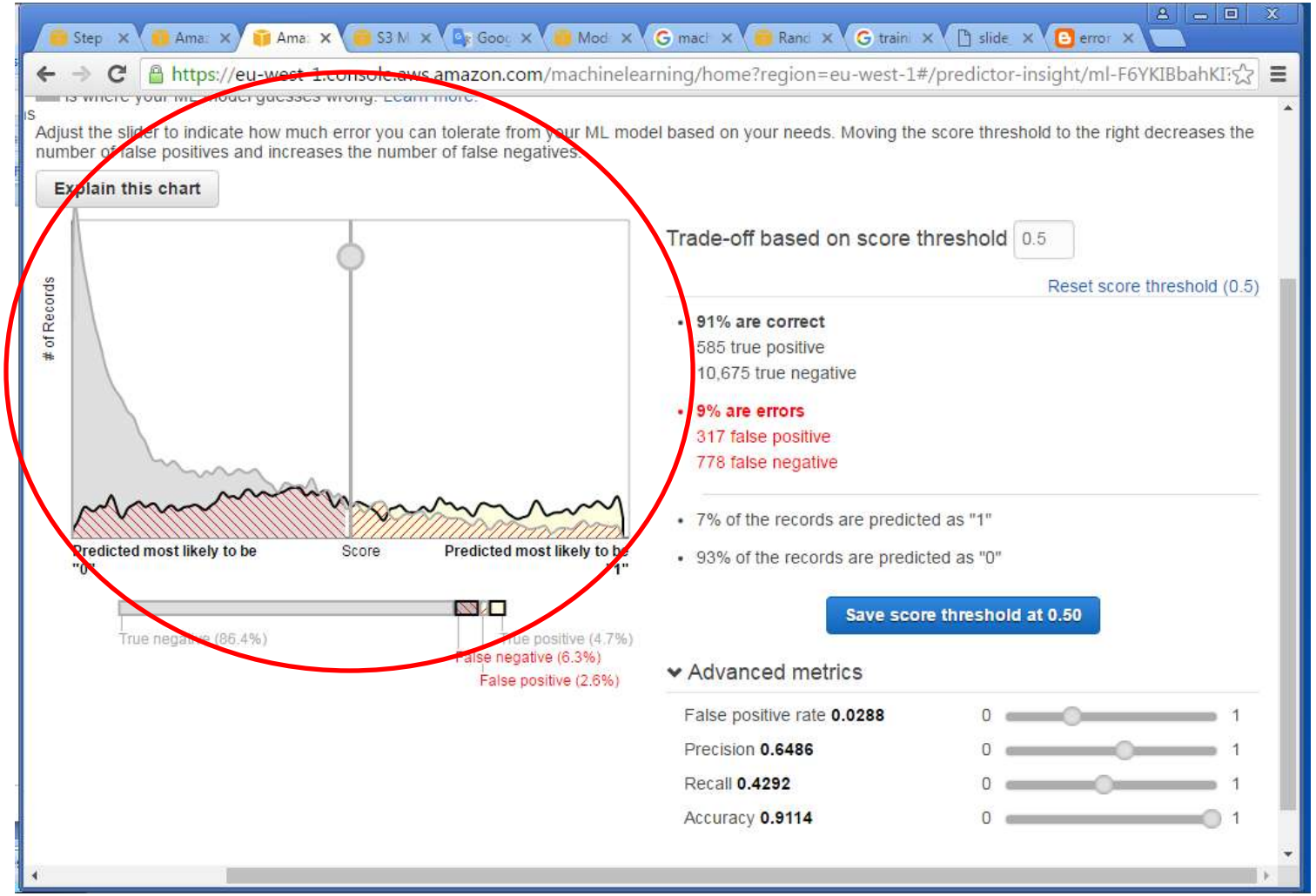
AMAZON ML TUTORIAL

- Eşike göre bir veya diğer sınıfından daha çok veya daha az tahmin üretilebilir



Amazon ML tutorial

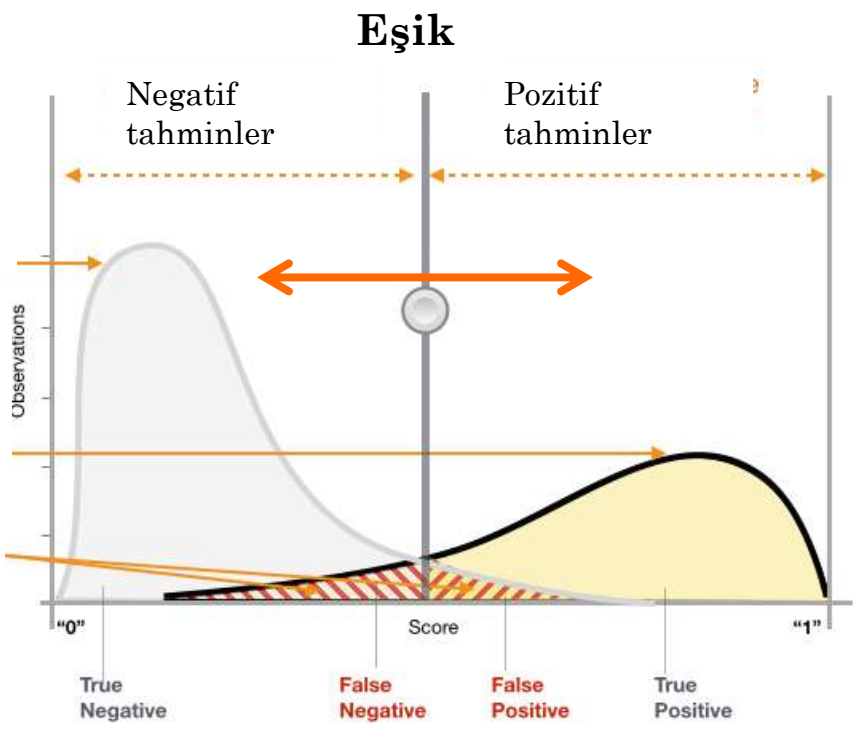
- Eşikin farklı değerleri tahminlerin (a) ortalama doğruluk, (b) tahminlerin kesinlik veya (c) anma ayarlayabilir ve hedeflerinize göre seçilmesi gerekir (örneğin – tahminleri yüksek kesinlikliğini gerektiren uygulamalarda daha yüksek eşik seçilmeli, olayların düşük kaybolma olasılığını gerektiren uygulamalarda daha düşük eşik seçilmeli ve hem daha yüksek tahminlerin kesinliği hemde daha düşük olayların kaybolma olasılığını (yani dorğuluğu) gerektiren uygulamalarda orta eşik değeri seçilebilir)


https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/predictor-insight/ml-F6YKIBbahKI
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
# of Records
Predicted most likely to be "0"
Score
Predicted most likely to be "1"
True negative (86.4%)
True positive (4.7%)
False negative (6.3%)
False positive (2.6%)
Trade-off based on score threshold 0.5
Reset score threshold (0.5)
91% are correct
585 true positive
10,675 true negative
9% are errors
317 false positive
778 false negative
7% of the records are predicted as "1"
93% of the records are predicted as "0"
Save score threshold at 0.50
Advanced metrics
False positive rate 0.0288
Precision 0.6486
Recall 0.4292
Accuracy 0.9114

CSV dosyanızdaki ML modelin çıktılarına göre negatif örneklerin değerleri/dağılımı

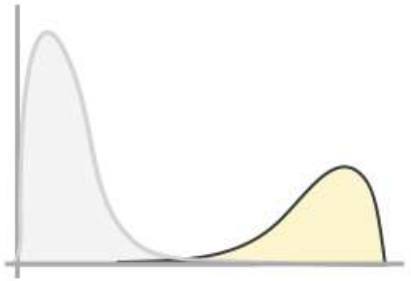
CSV dosyanızdaki ML modelin çıktılarına göre pozitif örneklerin değerleri/dağılımı

Bunlar belirli eşik için hataları gösteriyor

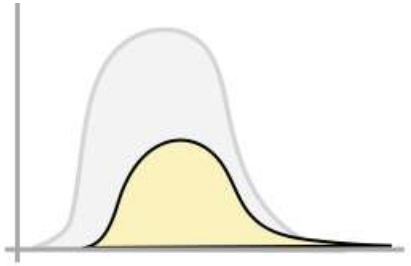


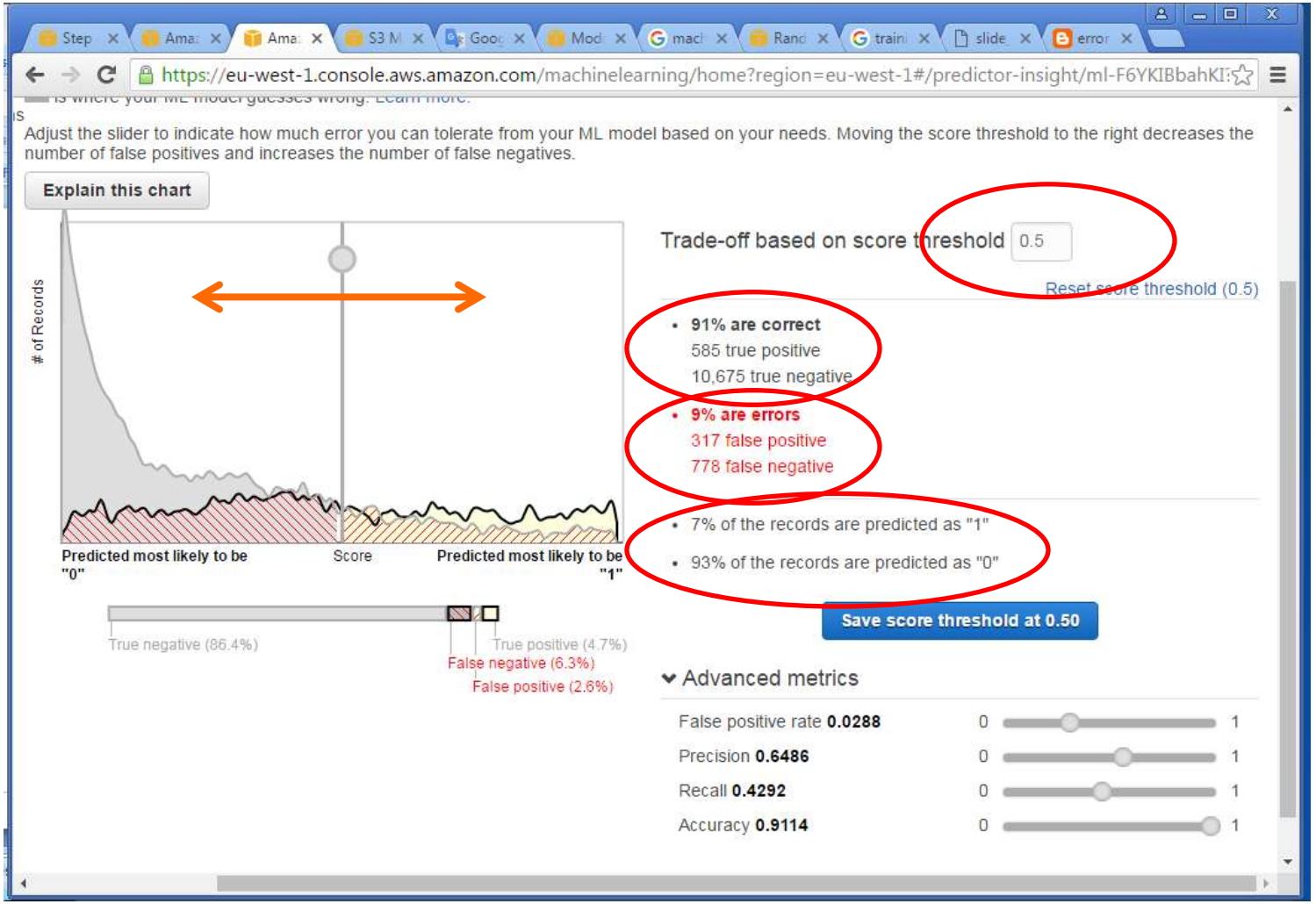
CSV dosyanızda olan örnekler için ML modelin çıktıları (yani modelin "olasılık" değerleri)

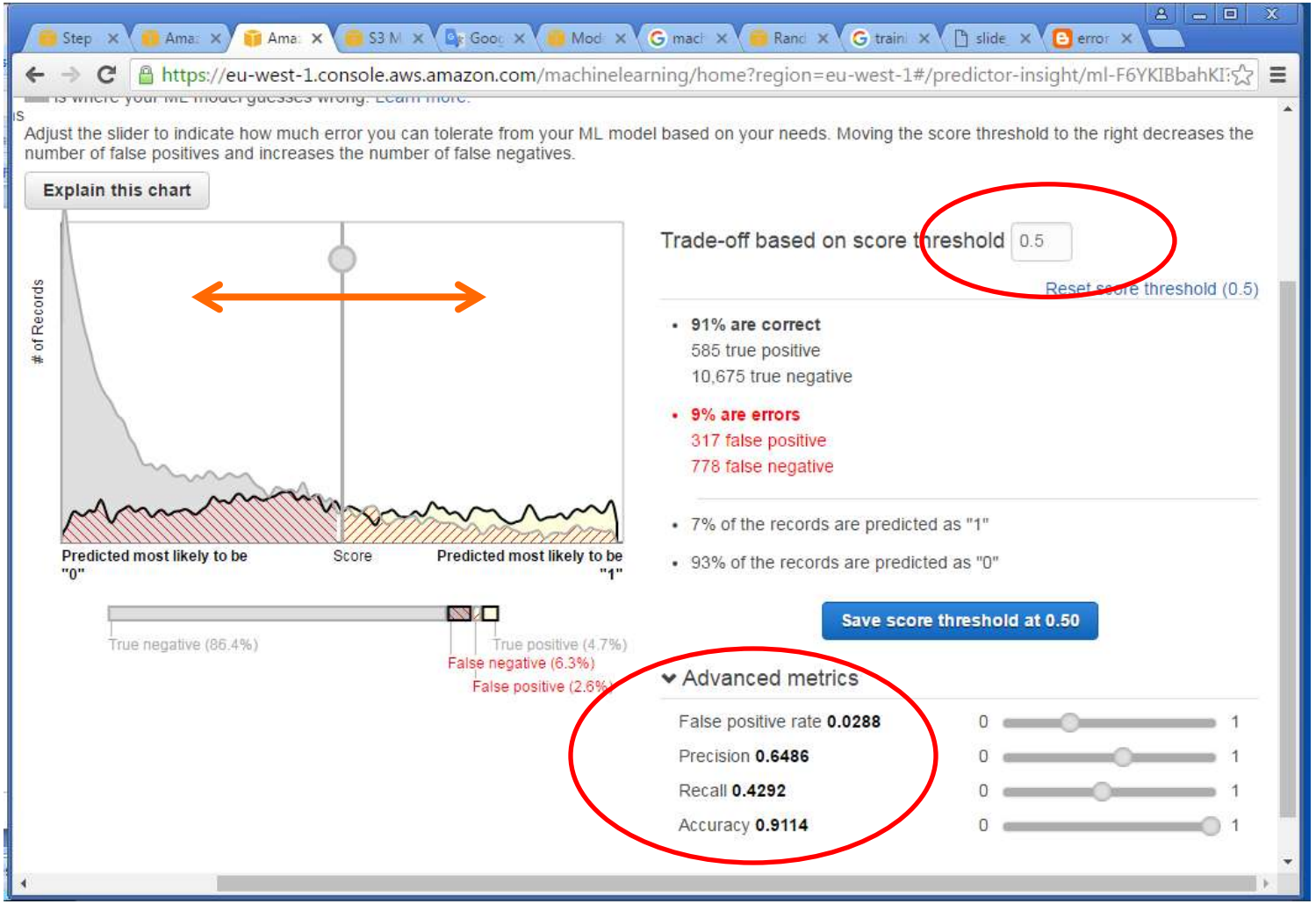
İyi ML modelin sonucu, durumlar ayırt ediliyor

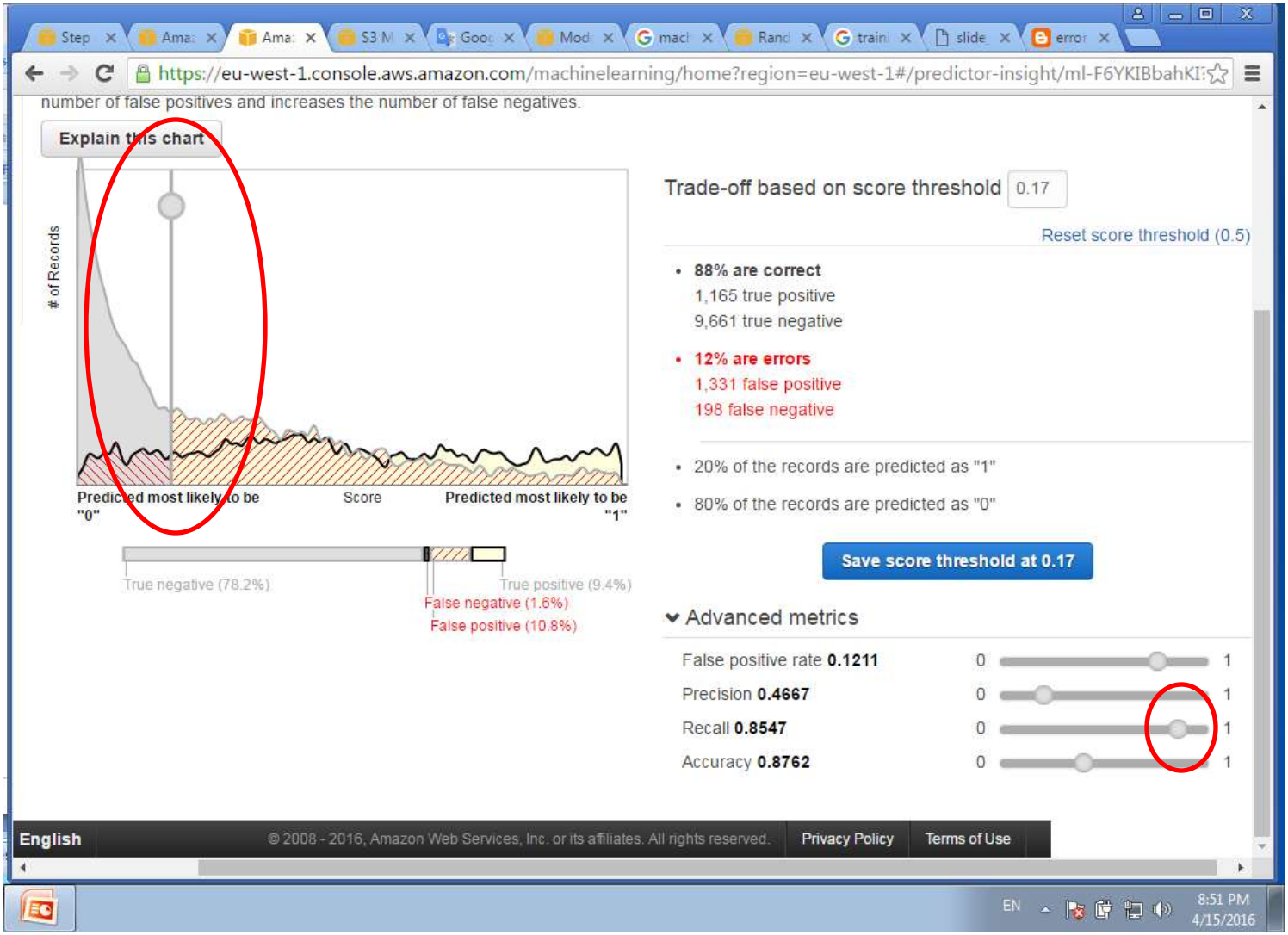


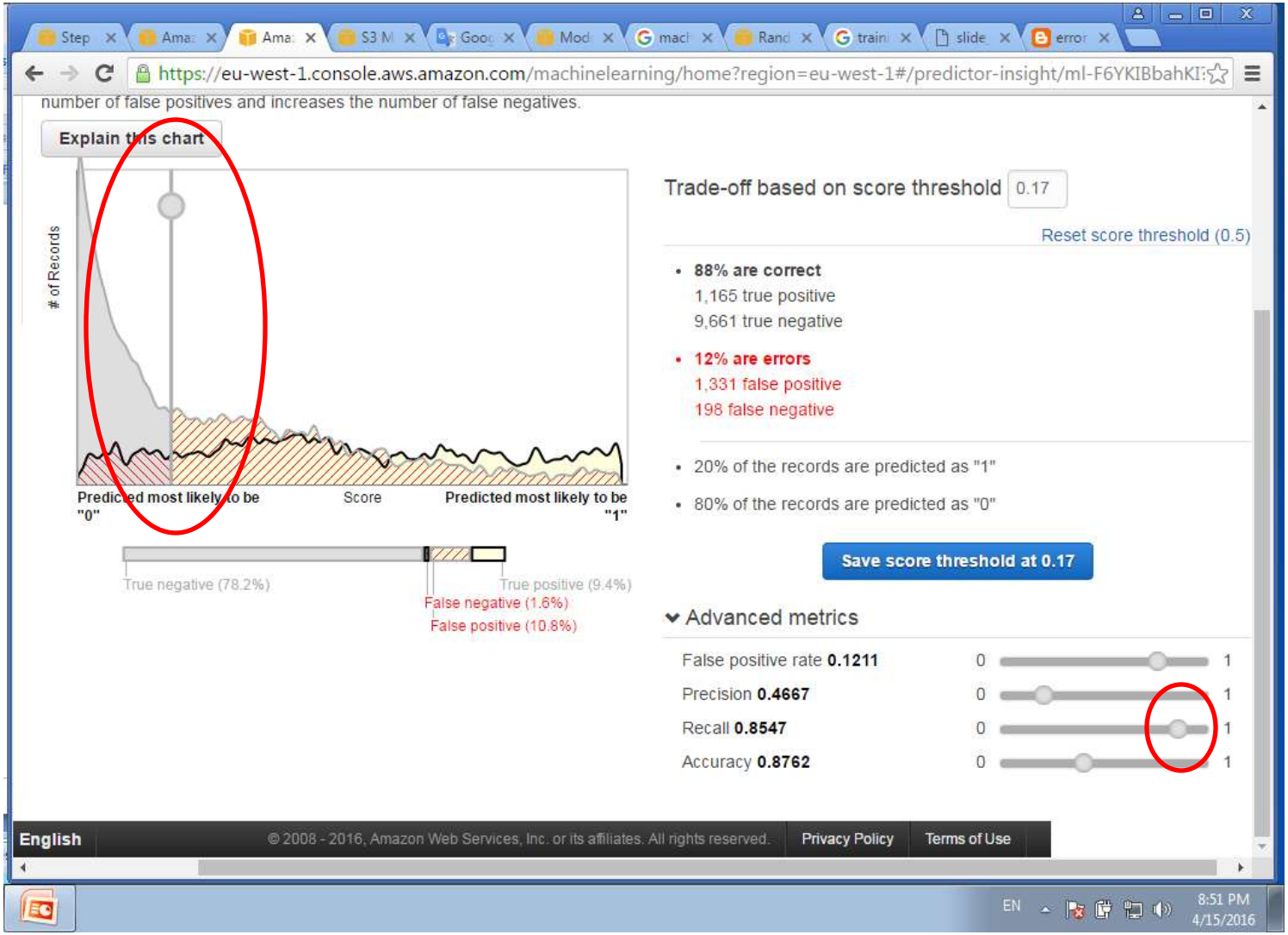
Kötü ML modelin sonucu, durumlar ayırt edilmiyor

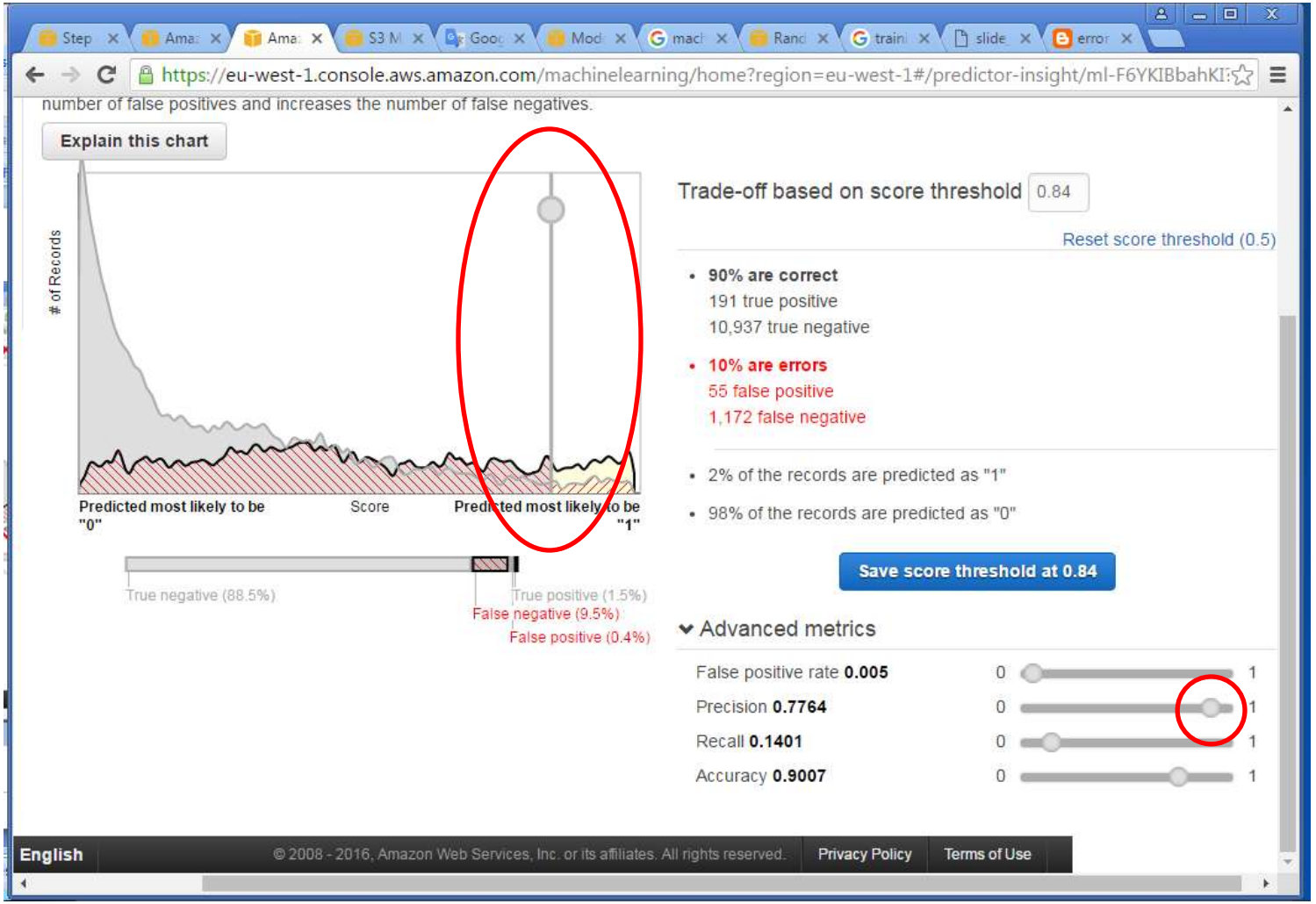


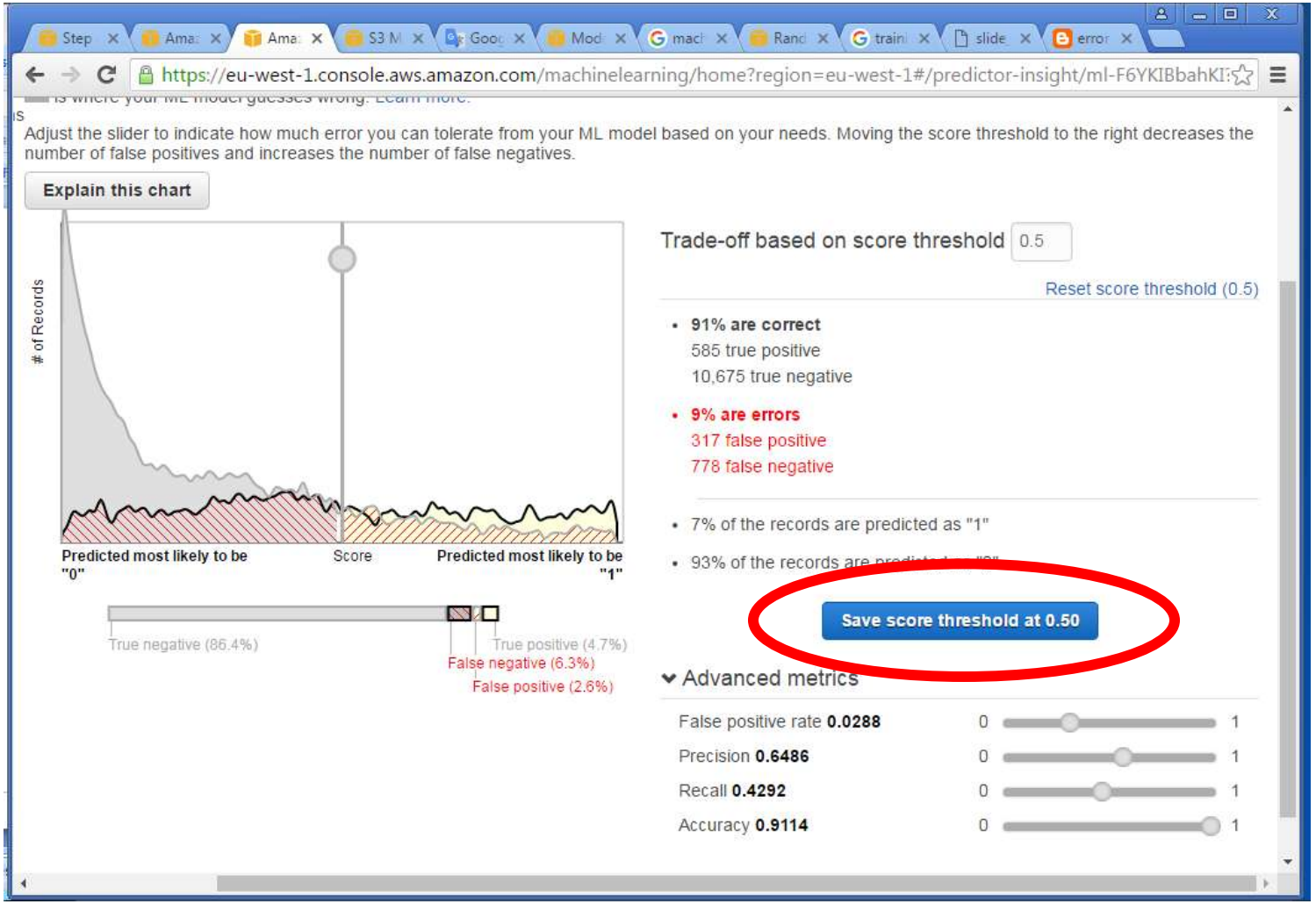

https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/predictor-insight/ml-F6YKIBbahKI
Adjust the slider to indicate how much error you can tolerate from your ML model based on your needs. Moving the score threshold to the right decreases the number of false positives and increases the number of false negatives.
Explain this chart
# of Records
Predicted most likely to be "0"
Score
Predicted most likely to be "1"
True negative (86.4%)
True positive (4.7%)
False negative (6.3%)
False positive (2.6%)
Trade-off based on score threshold 0.5
Reset score threshold (0.5)
91% are correct
585 true positive
10,675 true negative
9% are errors
317 false positive
778 false negative
7% of the records are predicted as "1"
93% of the records are predicted as "0"
Save score threshold at 0.50
Advanced metrics
False positive rate 0.0288
Precision 0.6486
Recall 0.4292
Accuracy 0.9114


https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/predictor-insight/ml-F6YKIBbahKI
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
# of Records
Predicted most likely to be "0"
Score
Predicted most likely to be "1"
True negative (86.4%)
False negative (6.3%)
True positive (4.7%)
False positive (2.6%)
Trade-off based on score threshold 0.5
Reset score threshold (0.5)
91% are correct
585 true positive
10,675 true negative
9% are errors
317 false positive
778 false negative
7% of the records are predicted as "1"
93% of the records are predicted as "0"
Save score threshold at 0.50
Advanced metrics
False positive rate 0.0288
Precision 0.6486
Recall 0.4292
Accuracy 0.9114


https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/predictor-insight/ml-F6YKIBbahKI
number of false positives and increases the number of false negatives.
Explain this chart
# of Records
Predicted most likely to be "0"
Score
Predicted most likely to be "1"
True negative (78.2%)
True positive (9.4%)
False negative (1.6%)
False positive (10.8%)
Trade-off based on score threshold 0.17
Reset score threshold (0.5)
88% are correct
1,165 true positive
9,661 true negative
12% are errors
1,331 false positive
198 false negative
20% of the records are predicted as "1"
80% of the records are predicted as "0"
Save score threshold at 0.17
Advanced metrics
False positive rate 0.1211
Precision 0.4667
Recall 0.8547
Accuracy 0.8762
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
EN
8:51 PM
4/15/2016

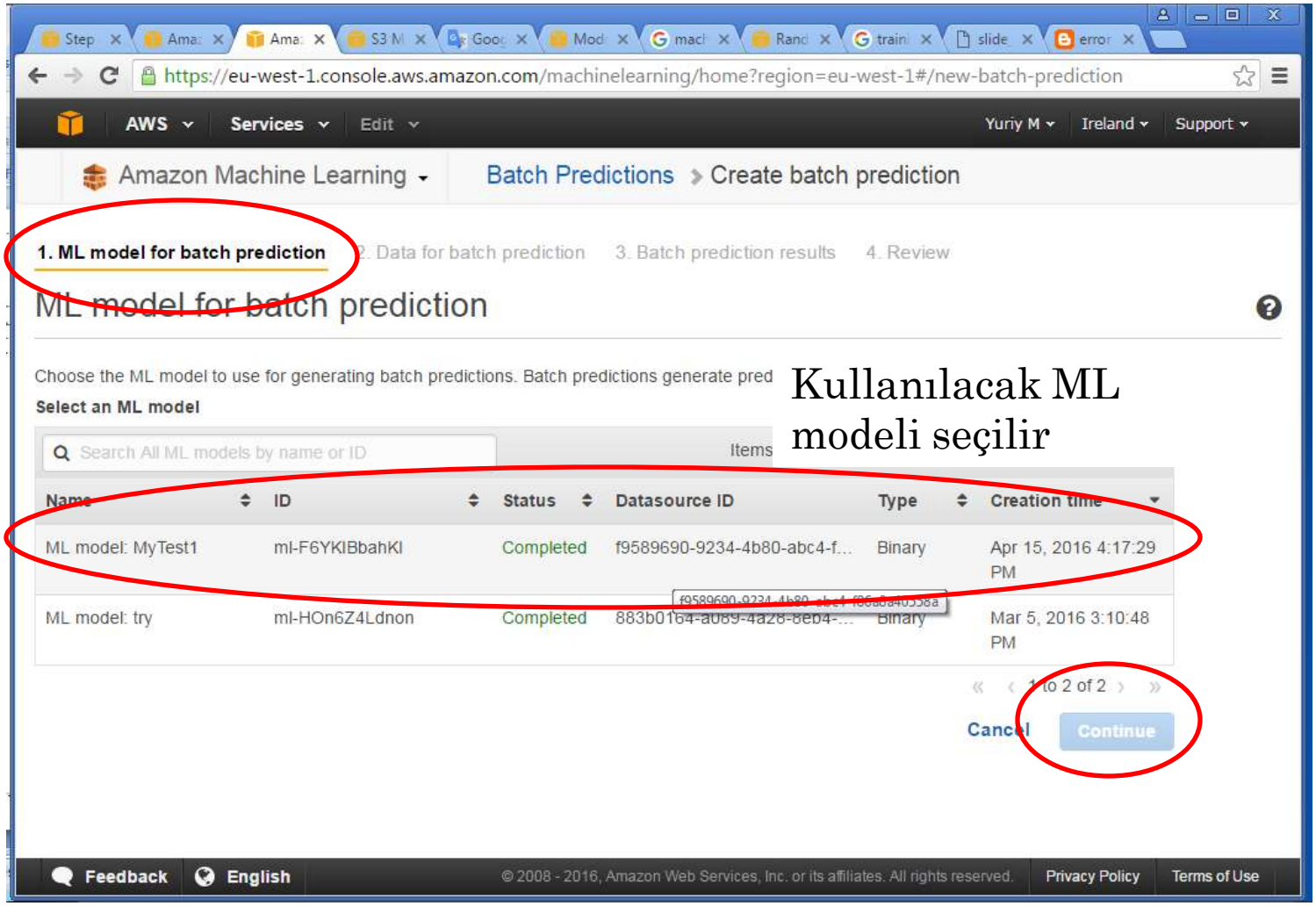

https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/predictor-insight/ml-F6YKIBbahKI
number of false positives and increases the number of false negatives.
Explain this chart
# of Records
Predicted most likely to be "0"
Score
Predicted most likely to be "1"
True negative (78.2%)
True positive (9.4%)
False negative (1.6%)
False positive (10.8%)
Trade-off based on score threshold 0.17
Reset score threshold (0.5)
88% are correct
1,165 true positive
9,661 true negative
12% are errors
1,331 false positive
198 false negative
20% of the records are predicted as "1"
80% of the records are predicted as "0"
Save score threshold at 0.17
Advanced metrics
False positive rate 0.1211
Precision 0.4667
Recall 0.8547
Accuracy 0.8762
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
EN
8:51 PM
4/15/2016


https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/predictor-insight/ml-F6YKIBbahKI
number of false positives and increases the number of false negatives.
Explain this chart
# of Records
Predicted most likely to be "0"
Score
Predicted most likely to be "1"
True negative (88.5%)
True positive (1.5%)
False negative (9.5%)
False positive (0.4%)
Trade-off based on score threshold 0.84
Reset score threshold (0.5)
90% are correct
191 true positive
10,937 true negative
10% are errors
55 false positive
1,172 false negative
2% of the records are predicted as "1"
98% of the records are predicted as "0"
Save score threshold at 0.84
Advanced metrics
False positive rate 0.005
Precision 0.7764
Recall 0.1401
Accuracy 0.9007
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use


https://eu-west-1.console.aws.amazon.com/machinelearning/home?region=eu-west-1#/predictor-insight/ml-F6YKIBbahKI
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
# of Records
Predicted most likely to be "0"
Score
Predicted most likely to be "1"
True negative (86.4%)
True positive (4.7%)
False negative (6.3%)
False positive (2.6%)
Trade-off based on score threshold 0.5
Reset score threshold (0.5)
91% are correct
585 true positive
10,675 true negative
9% are errors
317 false positive
778 false negative
7% of the records are predicted as "1"
93% of the records are
Save score threshold at 0.50
Advanced metrics
False positive rate 0.0288
Precision 0.6486
Recall 0.4292
Accuracy 0.9114

Amazon ML tutorial

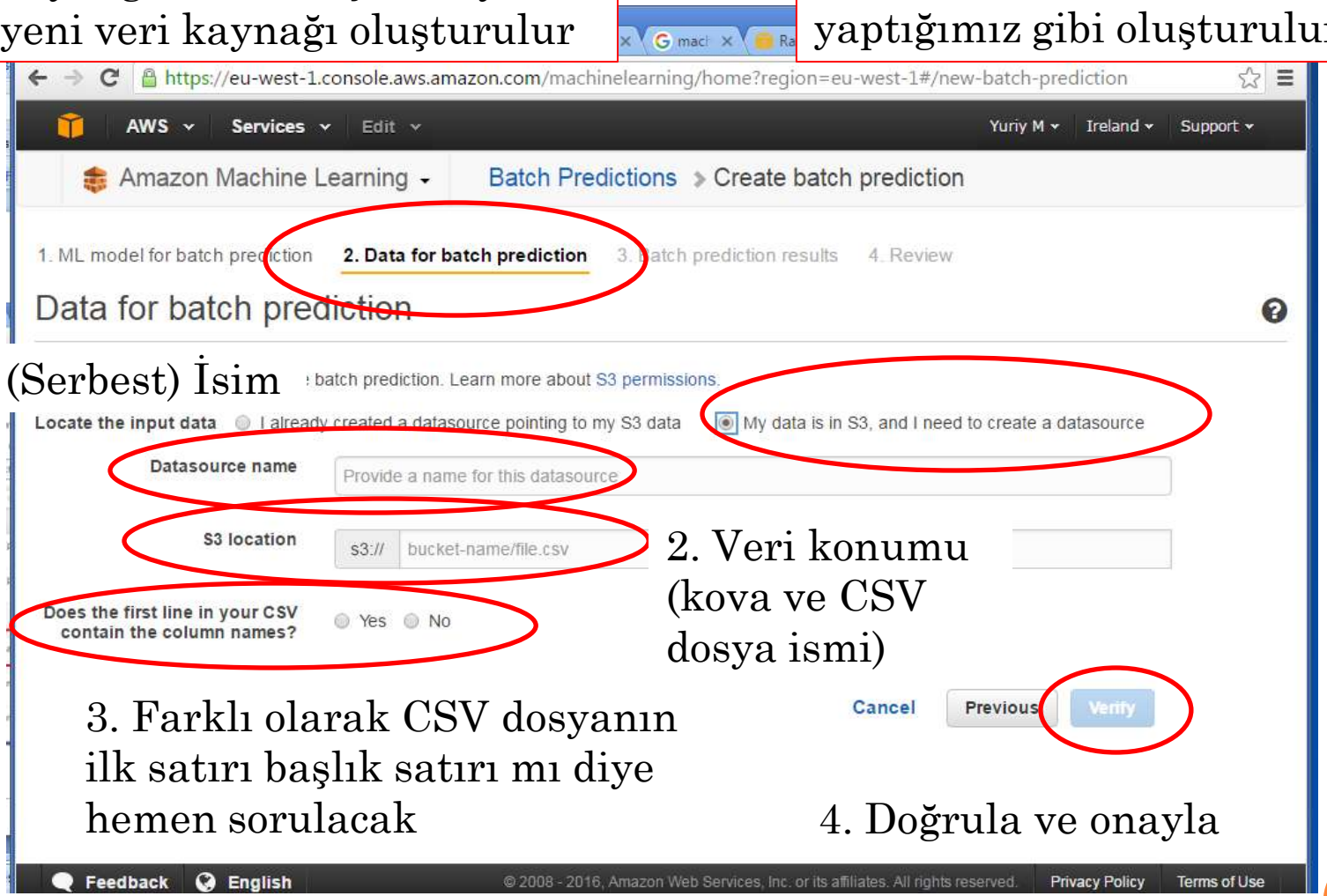
- Elde edilen modeli yeni verilerine uygulamak için, Amazon ML kontrol panelinden “Create New/Batch Predictions” (Batch tahminler) seçilmeli
- “Batch predictions” modunda Amazon ML hesabınızda kayıtlı olan model S3 depolamasındaki CSV dosyasına uygulanabilir ve sonuçlar yukarıdaki şekilde incelenebilir



Kullanılacak ML modeli seçilir

Kullanılacak veri, veri kaynağı olarak seçilir veya yeni veri kaynağı oluşturulur

Yeni veri kaynağı, S3'e yüklenen bir CSV dosya için yaptığımız gibi oluşturulur



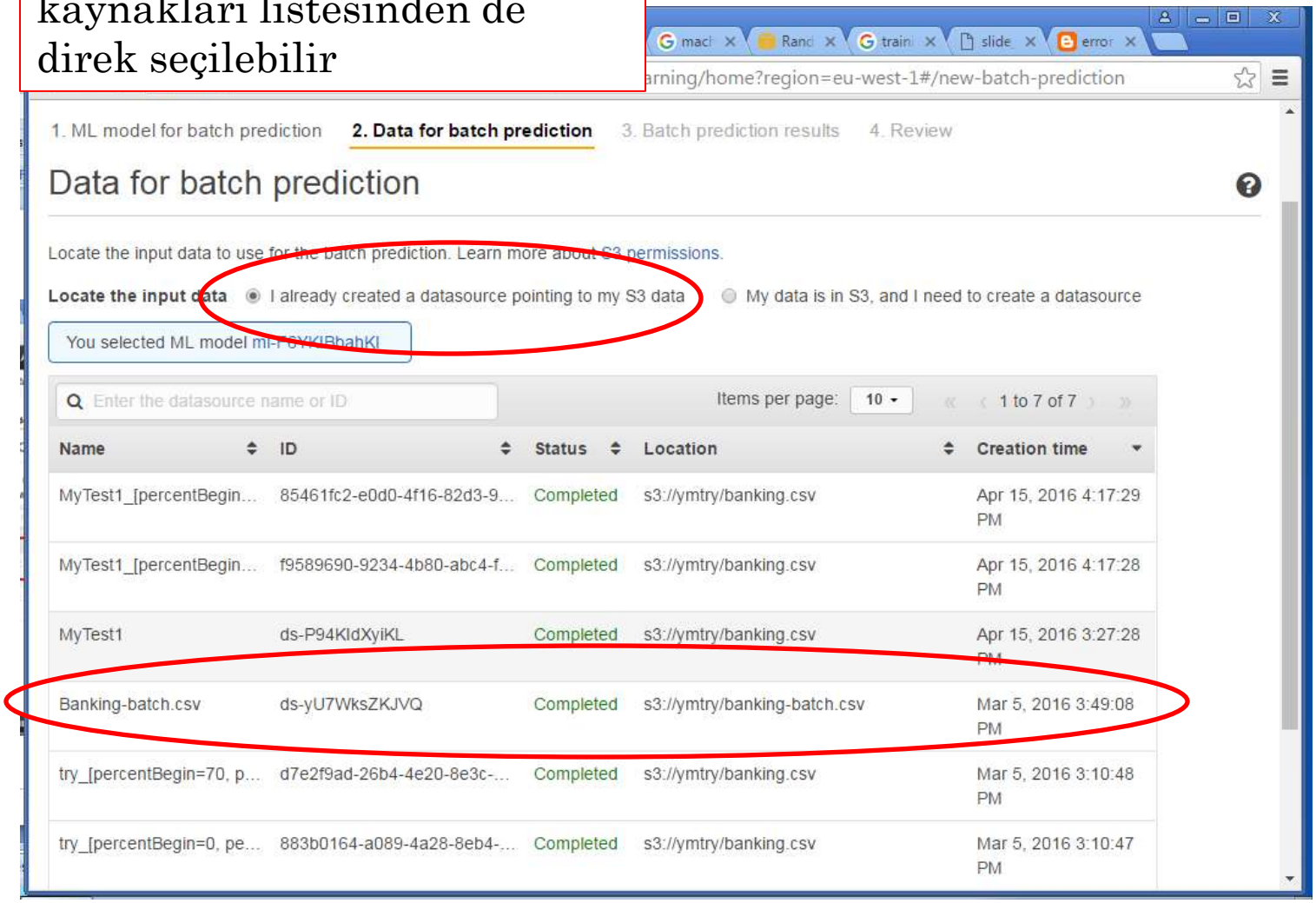
1. (Serbest) İsim

2. Veri konumu (kova ve CSV dosya ismi)

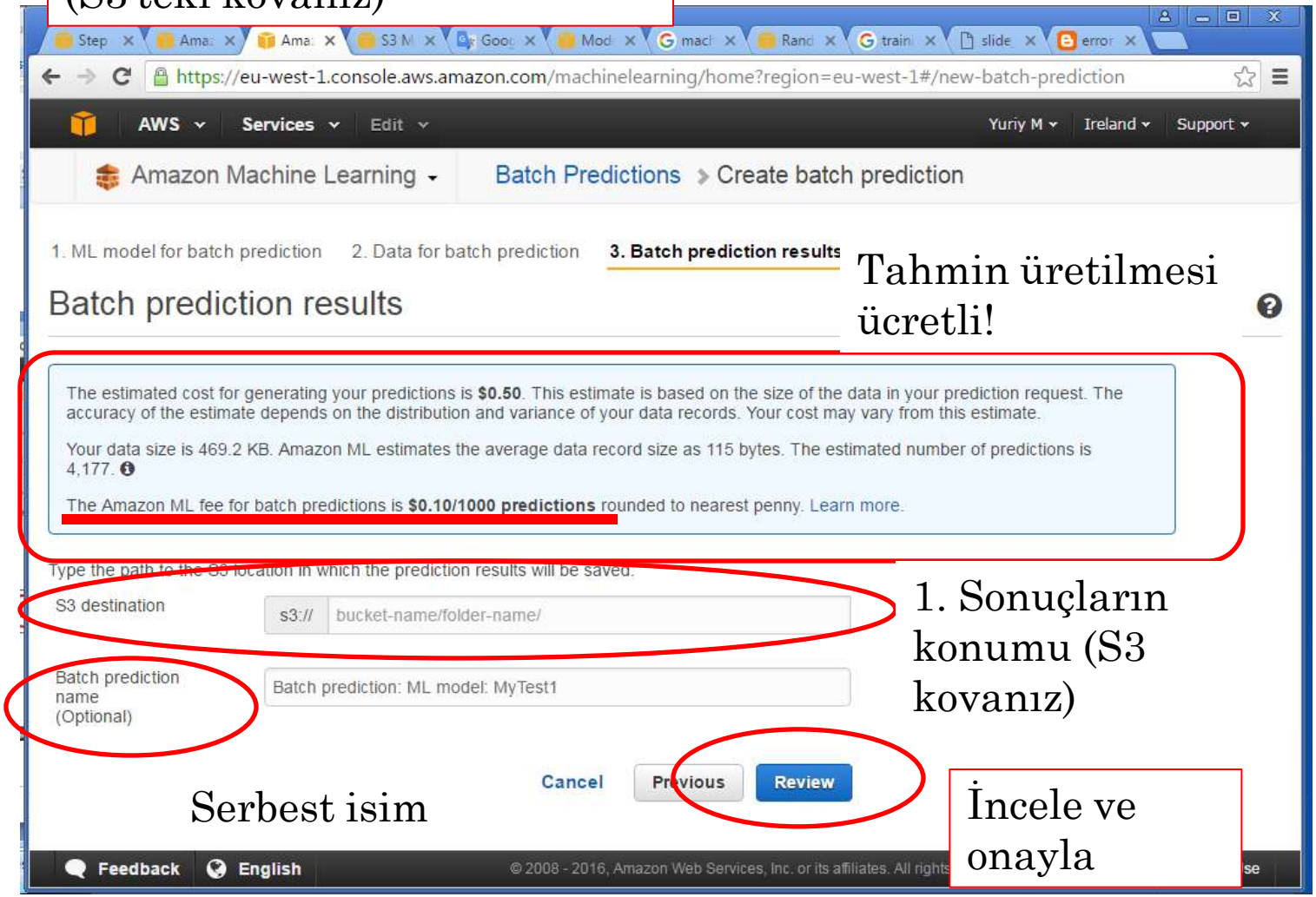
3. Farklı olarak CSV dosyanın ilk satırı başlık satırı mı diye hemen sorulacak

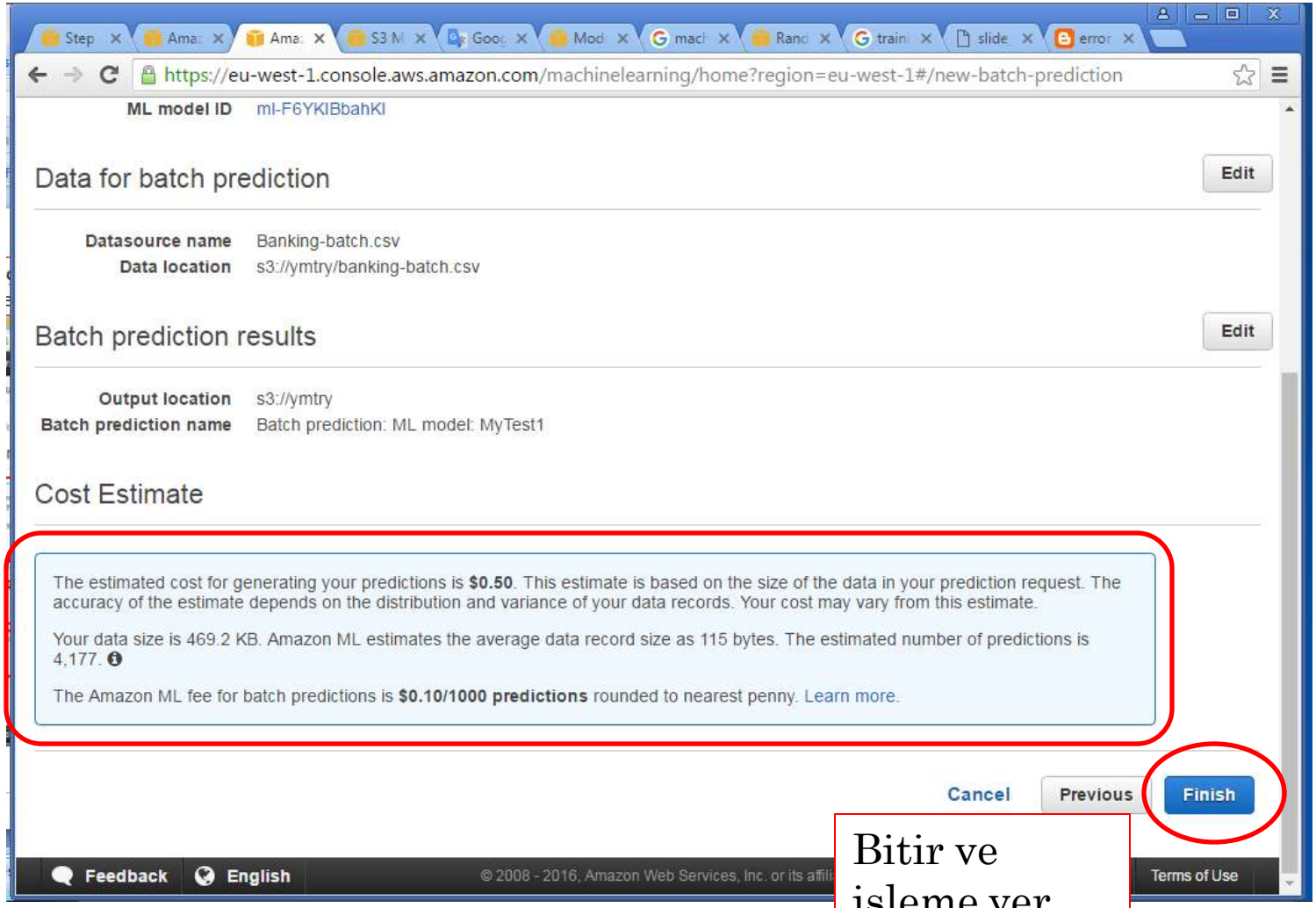
4. Doğrula ve onayla

Veri kaynağı zaten varsa, veri kaynakları listesinden de direk seçilebilir



Sonuçlar kaydedilecek konum (S3'teki kovanız)



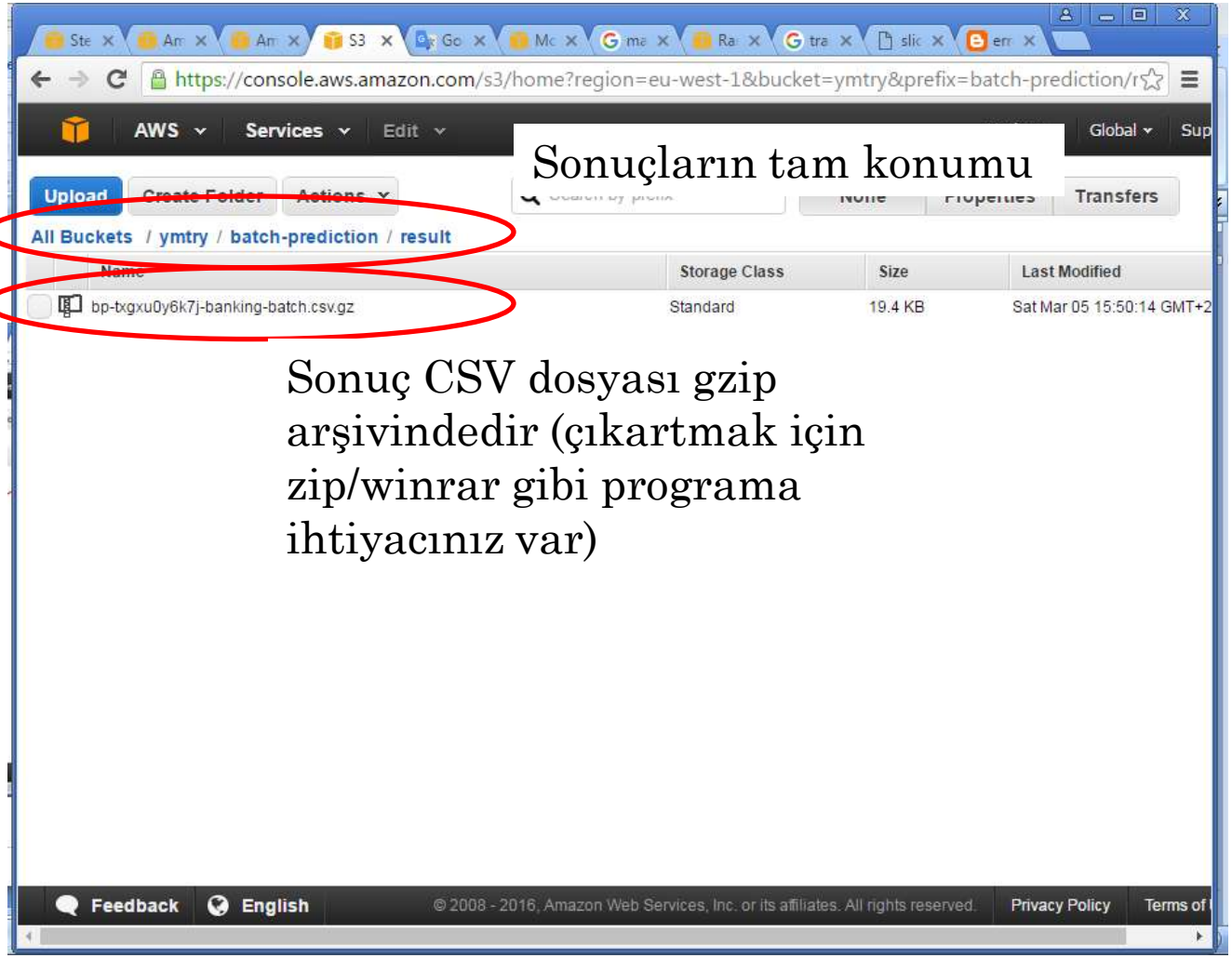


Bitir ve işleme ver

Amazon ML tutorial

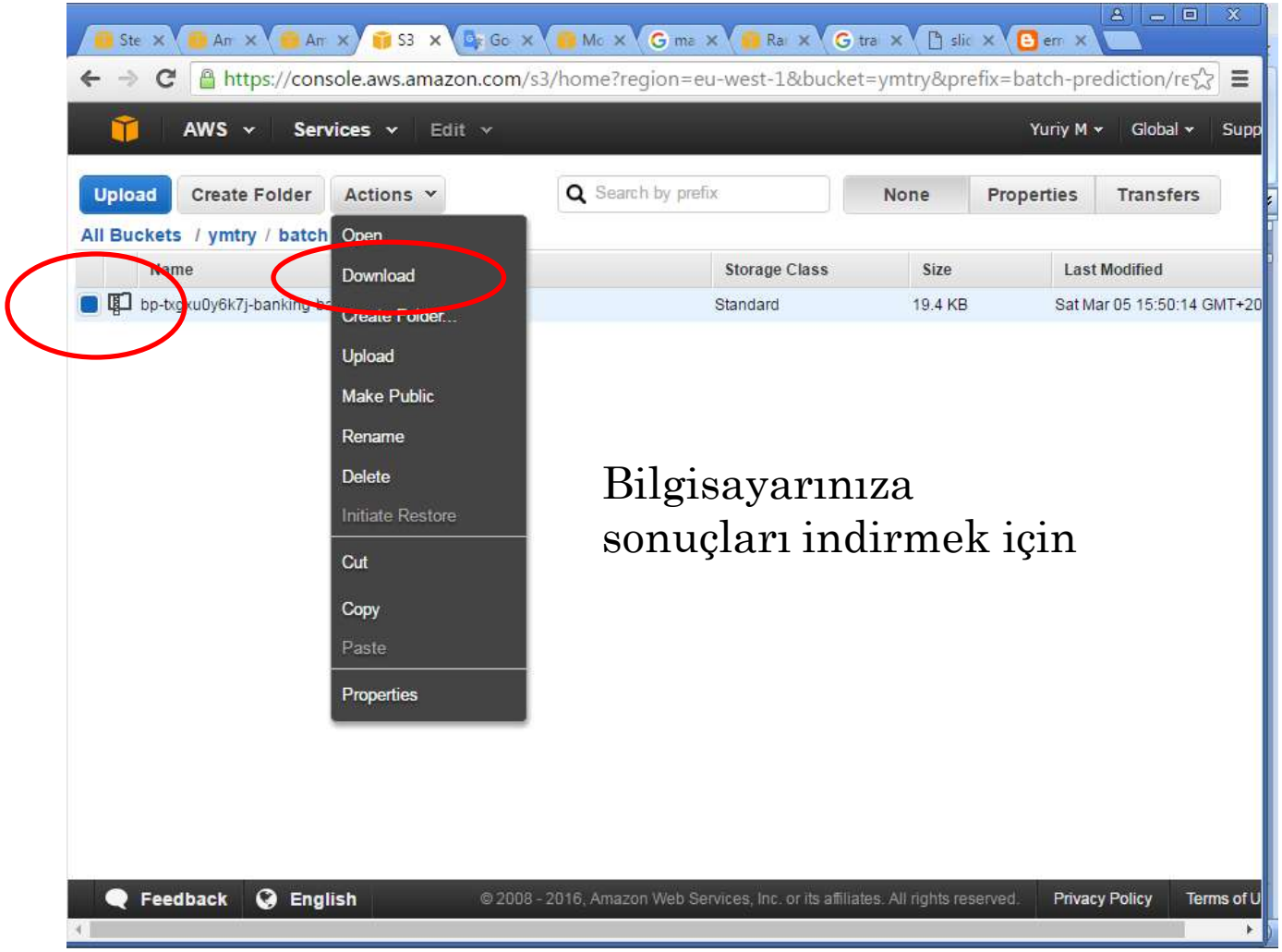
- Sonuçlar CSV dosyası olarak belirttiğiniz S3 kovasınızda oluşturulacak

S3 depolaması

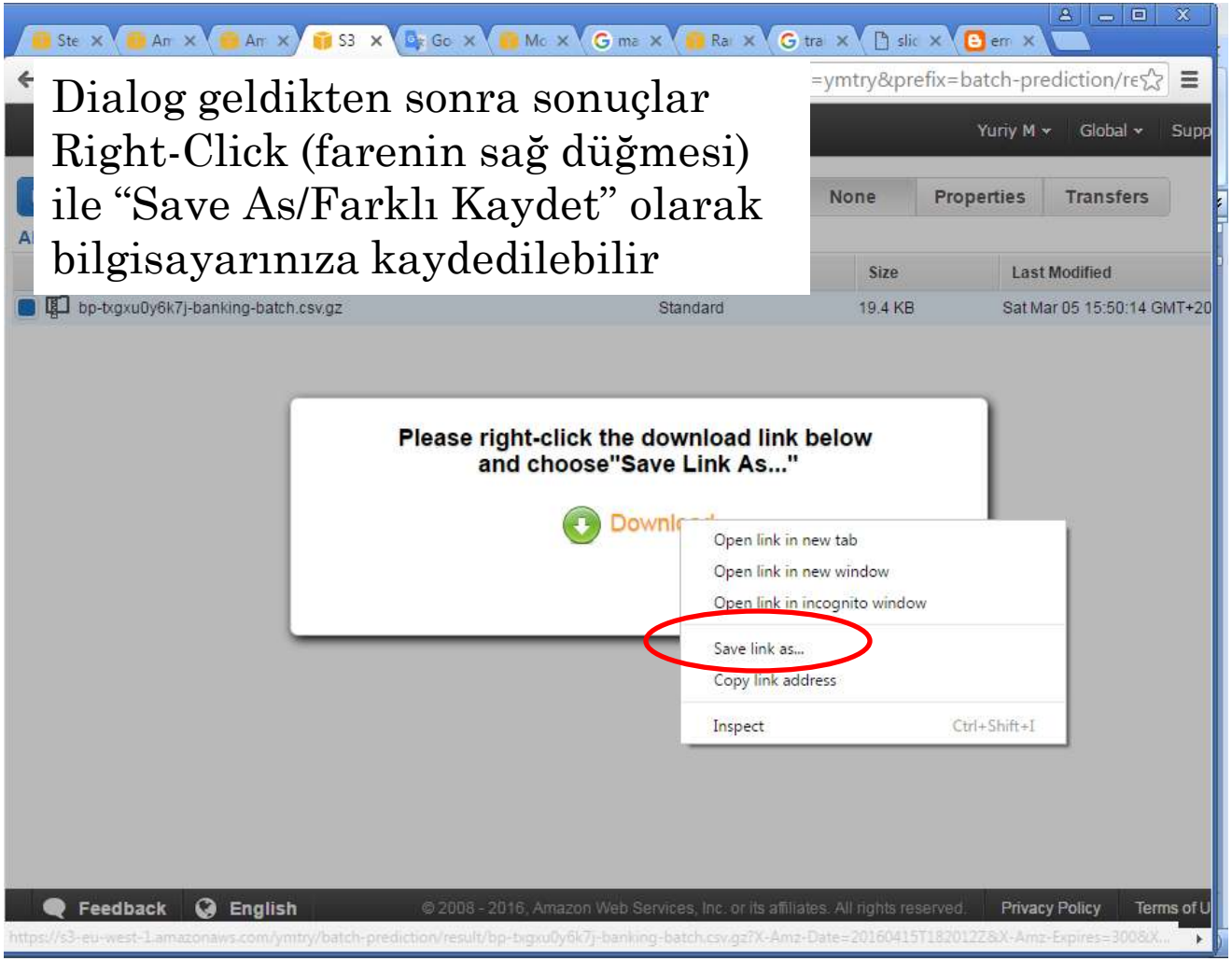


Sonuçların tam konumu

Sonuç CSV dosyası gzip arşivindedir (çıkartmak için zip/winrar gibi programa ihtiyacınız var)



Bilgisayarınıza
sonuçları indirmek için


Dialog geldikten sonra sonuçlar Right-Click (farenin sağ düğmesi) ile “Save As/Farklı Kaydet” olarak bilgisayarınıza kaydedilebilir
Yuriy M
Global
None
Properties
Transfers
Size
Last Modified
bp-txgxu0y6k7j-banking-batch.csv.gz
Standard
19.4 KB
Sat Mar 05 15:50:14 GMT+20
Please right-click the download link below and choose"Save Link As..."
Open link in new tab
Open link in new window
Open link in incognito window
Save link as...
Copy link address
Inspect
Ctrl+Shift+I
Feedback
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy

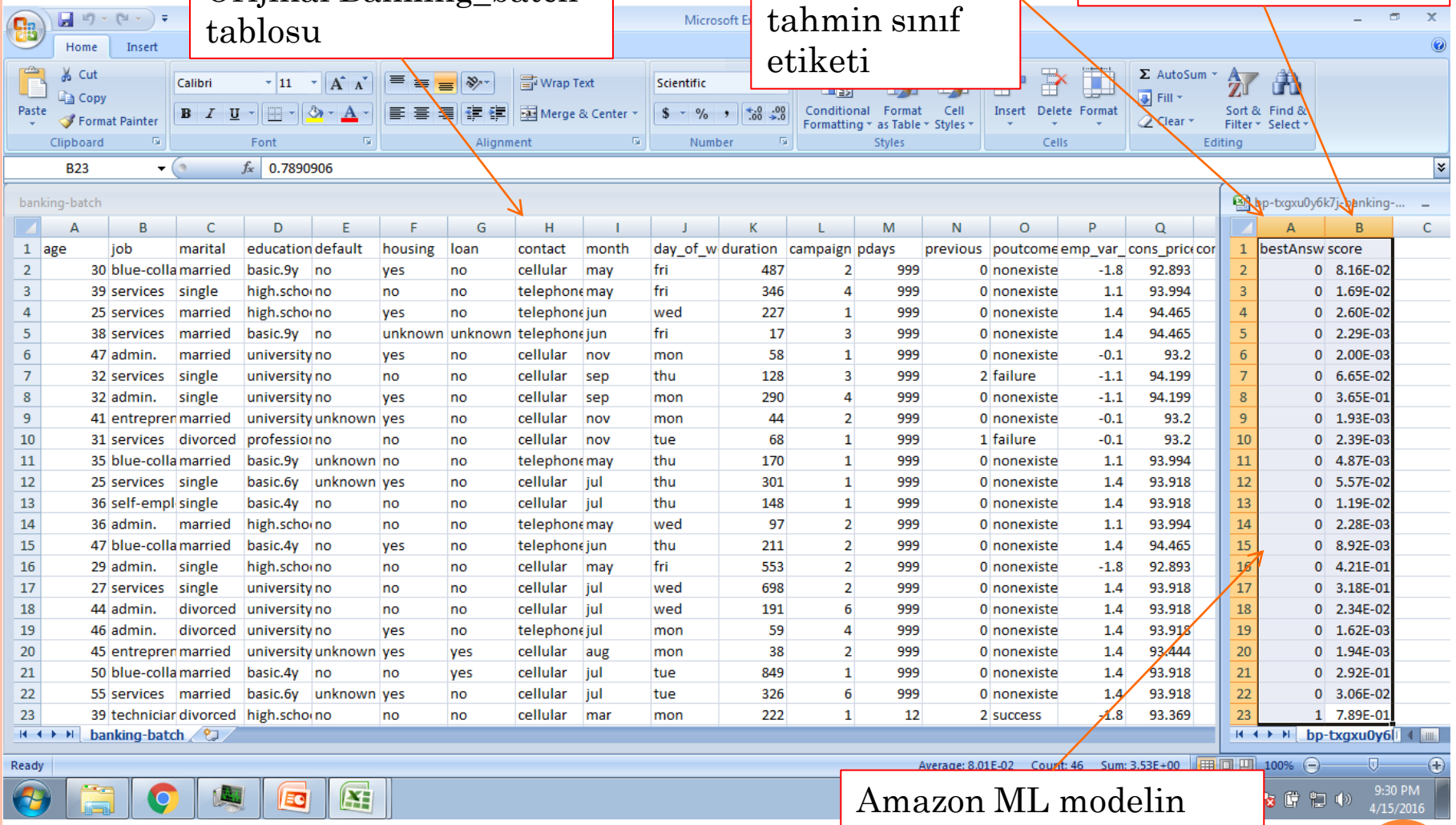
Amazon ML tutorial

- Sonuç CSV dosyasında, girdi CSV dosyanızın satır sırasına göre tahmin etiketleri (yani örnekler için sonuç sınıfları) ve gerçel olasılık değerleri ile beraber gösterilmektedir

Orijinal Banking_batch tablosu

"Best Answer" sütünü – tahmin sınıf etiketi

"Score" sütünü – tahmin olasılıkları



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	age	job	marital	education	default	housing	loan	contact	month	day_of_w	duration	campaign	pdays	previous	poutcome	emp_var_	cons_pric
2	30	blue-colla	married	basic.9y	no	yes	no	cellular	may	fri	487	2	999	0	nonexiste	-1.8	92.893
3	39	services	single	high.scho	no	no	no	telephone	may	fri	346	4	999	0	nonexiste	1.1	93.994
4	25	services	married	high.scho	no	yes	no	telephone	jun	wed	227	1	999	0	nonexiste	1.4	94.465
5	38	services	married	basic.9y	no	unknown	unknown	telephone	jun	fri	17	3	999	0	nonexiste	1.4	94.465
6	47	admin.	married	university	no	yes	no	cellular	nov	mon	58	1	999	0	nonexiste	-0.1	93.2
7	32	services	single	university	no	no	no	cellular	sep	thu	128	3	999	2	failure	-1.1	94.199
8	32	admin.	single	university	no	yes	no	cellular	sep	mon	290	4	999	0	nonexiste	-1.1	94.199
9	41	entrepren	married	university	unknown	yes	no	cellular	nov	mon	44	2	999	0	nonexiste	-0.1	93.2
10	31	services	divorced	professio	no	no	no	cellular	nov	tue	68	1	999	1	failure	-0.1	93.2
11	35	blue-colla	married	basic.9y	unknown	no	no	telephone	may	thu	170	1	999	0	nonexiste	1.1	93.994
12	25	services	single	basic.6y	unknown	yes	no	cellular	jul	thu	301	1	999	0	nonexiste	1.4	93.918
13	36	self-empl	single	basic.4y	no	no	no	cellular	jul	thu	148	1	999	0	nonexiste	1.4	93.918
14	36	admin.	married	high.scho	no	no	no	telephone	may	wed	97	2	999	0	nonexiste	1.1	93.994
15	47	blue-colla	married	basic.4y	no	yes	no	telephone	jun	thu	211	2	999	0	nonexiste	1.4	94.465
16	29	admin.	single	high.scho	no	no	no	cellular	may	fri	553	2	999	0	nonexiste	-1.8	92.893
17	27	services	single	university	no	no	no	cellular	jul	wed	698	2	999	0	nonexiste	1.4	93.918
18	44	admin.	divorced	university	no	no	no	cellular	jul	wed	191	6	999	0	nonexiste	1.4	93.918
19	46	admin.	divorced	university	no	yes	no	telephone	jul	mon	59	4	999	0	nonexiste	1.4	93.918
20	45	entrepren	married	university	unknown	yes	yes	cellular	aug	mon	38	2	999	0	nonexiste	1.4	93.444
21	50	blue-colla	married	basic.4y	no	no	yes	cellular	jul	tue	849	1	999	0	nonexiste	1.4	93.918
22	55	services	married	basic.6y	unknown	yes	no	cellular	jul	tue	326	6	999	0	nonexiste	1.4	93.918
23	39	technician	divorced	high.scho	no	no	no	cellular	mar	mon	222	1	12	2	success	-1.8	93.369

	A	B
1	bestAnsw	score
2	0	8.16E-02
3	0	1.69E-02
4	0	2.60E-02
5	0	2.29E-03
6	0	2.00E-03
7	0	6.65E-02
8	0	3.65E-01
9	0	1.93E-03
10	0	2.39E-03
11	0	4.87E-03
12	0	5.57E-02
13	0	1.19E-02
14	0	2.28E-03
15	0	8.92E-03
16	0	4.21E-01
17	0	3.18E-01
18	0	2.34E-02
19	0	1.62E-03
20	0	1.94E-03
21	0	2.92E-01
22	0	3.06E-02
23	1	7.89E-01

Amazon ML modelin sonuç tablosu

Amazon Web Services'nin Amazon Machine Learning hizmeti