

PROJECT CARD

[SKIP IF FAMILIAR]



INTRODUCTION AND KEY LEARNING OUTCOMES

- We will use an XG-Boost SageMaker Built-in algorithm using SageMaker JumpStart.
- We will learn how to:
 1. Train an XG-boost algorithm in SageMaker to predict university admission using AWS JumpStart
 2. Train an XG-boost algorithm in SageMaker to predict life expectancy (capstone project) using AWS JumpStart
 3. List XG-Boost hyperparameters
 4. Assess trained regression models performance
 5. Plot the residuals
 6. Deploy an endpoint and perform inference

PROJECT CARD

GOAL:

- Build, train, test and deploy a machine learning model to predict chances of university admission into a particular university given student's profile.

TOOL:

- AWS SageMaker Studio – Jump Start

PRACTICAL REAL-WORLD APPLICATION:

- This project can be effectively used by university admission departments to determine top qualifying students.

DATA:

INPUTS (FEATURES):

- GRE Scores (out of 340)
- TOEFL Scores (out of 120)
- University Rating (out of 5)
- Statement of Purpose (SOP)
- Letter of Recommendation (LOR) Strength (out of 5)
- Undergraduate GPA (out of 10)
- Research Experience (either 0 or 1)

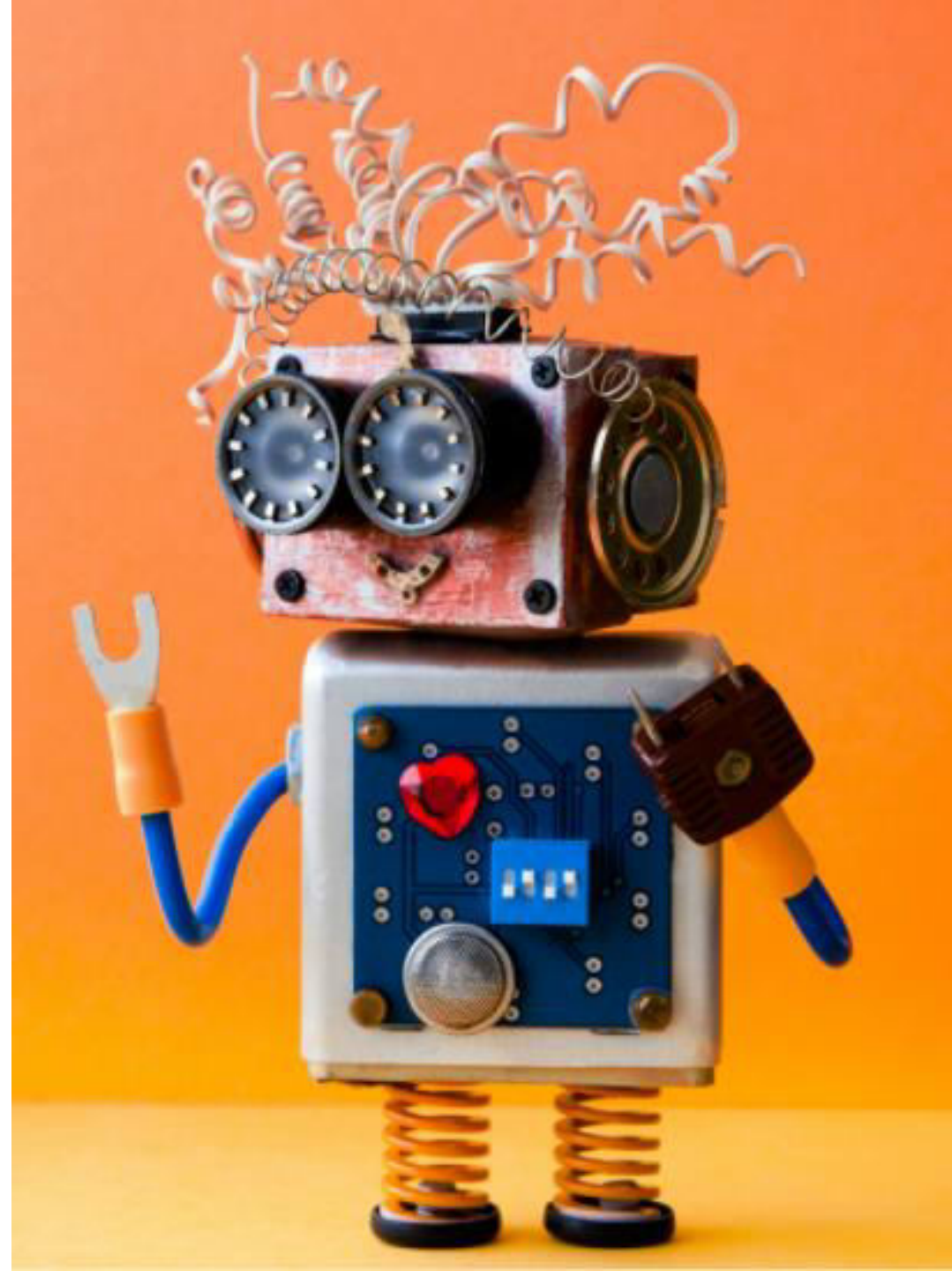
OUTPUTS:

- Chance of admission (ranging from 0 to 1)

- Data Source: <https://www.kaggle.com/robertmiller/graduate-admissions>
- Photo Credit: <https://www.pexels.com/photo/achievement-ceremony-education-graduation-267885/>

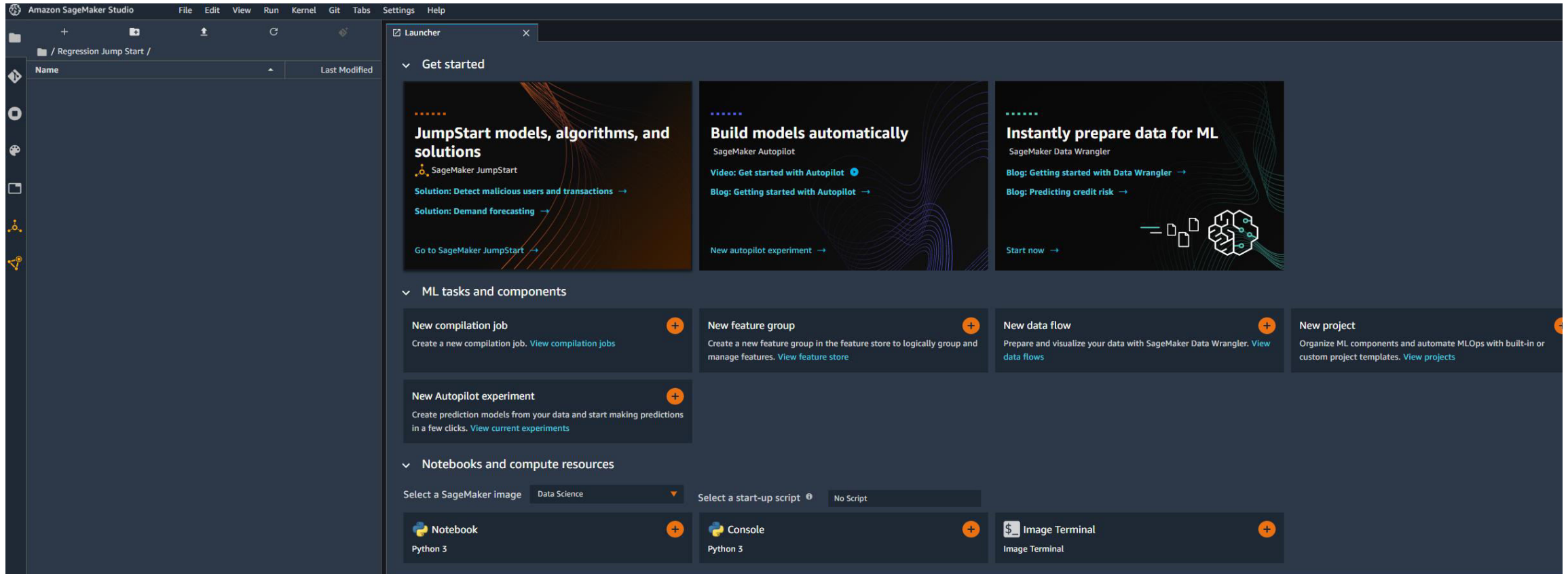


JUMPSTART DEMO: OVERVIEW



JUMP START DEMO

NAVIGATE TO AMAZON SAGEMAKER STUDIO HOMEPAGE AND
CLICK ON SAGEMAKER JUMPSTART

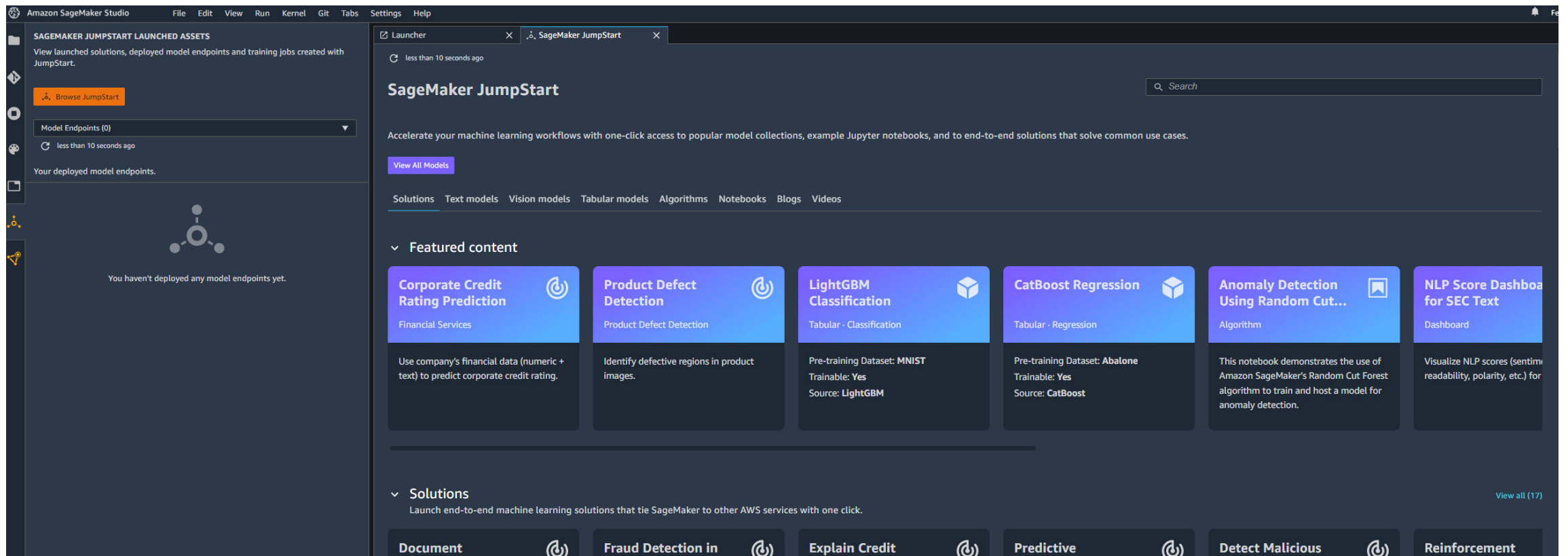


<https://www.flickr.com/photos/pasa/6757993805>

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

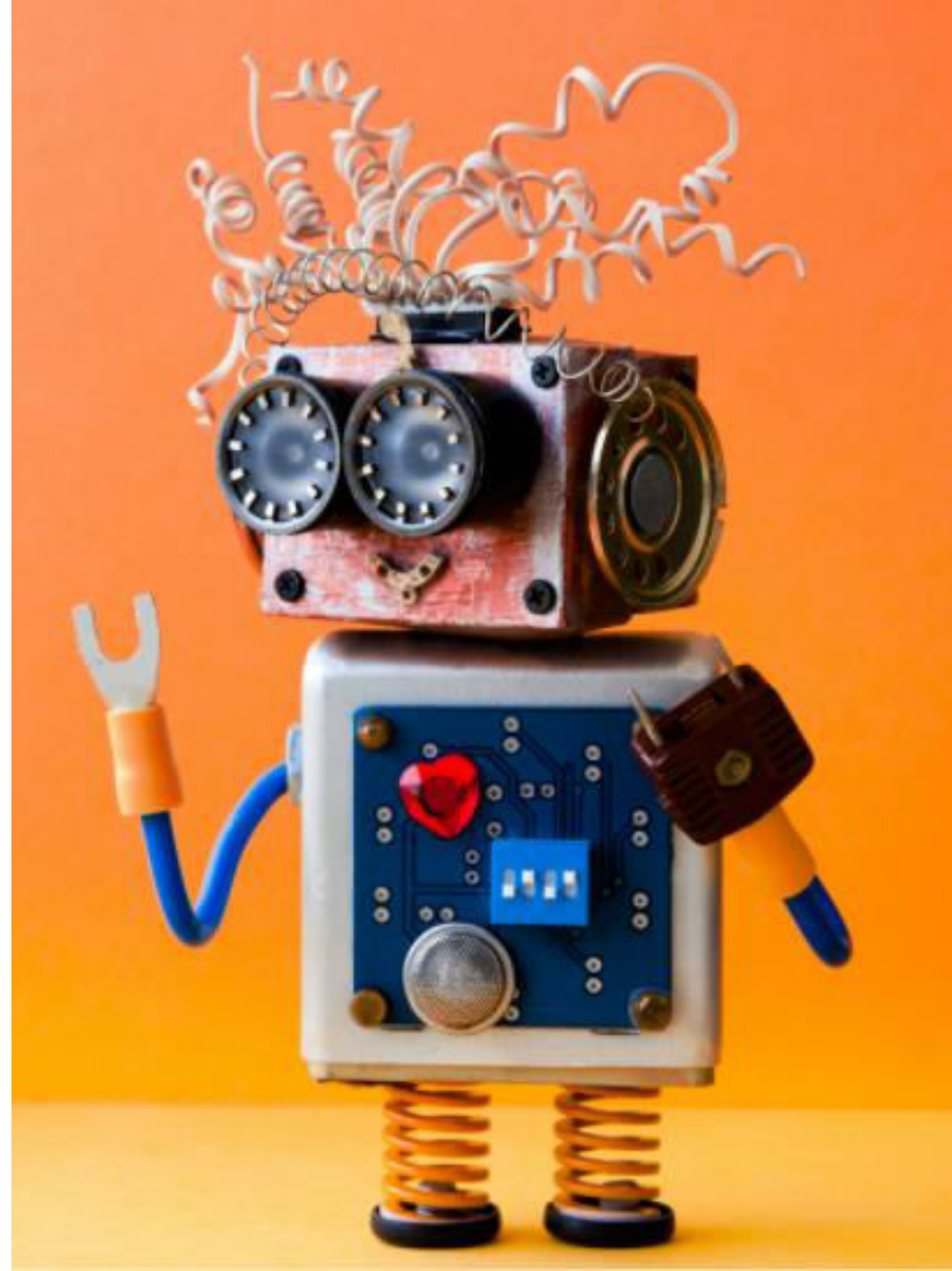
YOU SHOULD FIND SEVERAL ML WORKFLOWS THAT CAN BE DEPLOYED WITH ONE CLICK ACCESS. THERE ARE VISION, TEXT, AND TABULAR MODELS..ETC.



<https://www.flickr.com/photos/pasa/6757993805>

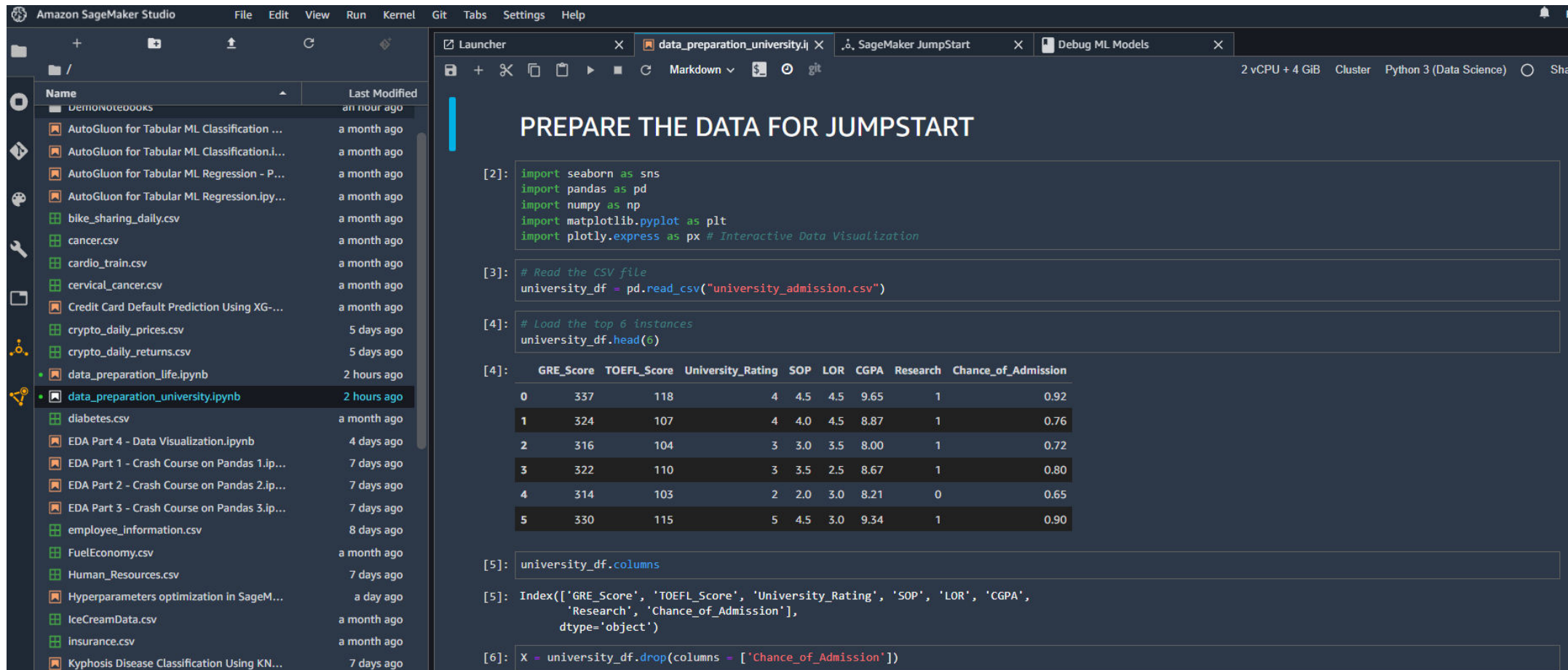
<https://www.kaggle.com/ljanjughazyan/cars1>

JUMPSTART DEMO: DATA PREP PRIOR TO TRAINING



DATA PREP

IN THIS NOTEBOOK, WE WILL SPLIT THE DATA INTO TRAINING, VALIDATION AND TESTING. WE WILL ALSO SET THE FIRST COLUMN IN THE DATA TO BE OUR TARGET (OUTPUT) COLUMN.



PREPARE THE DATA FOR JUMPSTART

```
[2]: import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import plotly.express as px # Interactive Data Visualization
```

```
[3]: # Read the CSV file
university_df = pd.read_csv("university_admission.csv")
```

```
[4]: # Load the top 6 instances
university_df.head(6)
```

	GRE_Score	TOEFL_Score	University_Rating	SOP	LOR	CGPA	Research	Chance_of_Admission
0	337	118	4	4.5	4.5	9.65	1	0.92
1	324	107	4	4.0	4.5	8.87	1	0.76
2	316	104	3	3.0	3.5	8.00	1	0.72
3	322	110	3	3.5	2.5	8.67	1	0.80
4	314	103	2	2.0	3.0	8.21	0	0.65
5	330	115	5	4.5	3.0	9.34	1	0.90

```
[5]: university_df.columns
```

```
[5]: Index(['GRE_Score', 'TOEFL_Score', 'University_Rating', 'SOP', 'LOR', 'CGPA',
        'Research', 'Chance_of_Admission'],
        dtype='object')
```

```
[6]: X = university_df.drop(columns = ['Chance_of_Admission'])
```

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMPSTART DEMO: DATA UPLOAD



JUMP START DEMO

CLICK ON XGBOOST REGRESSION

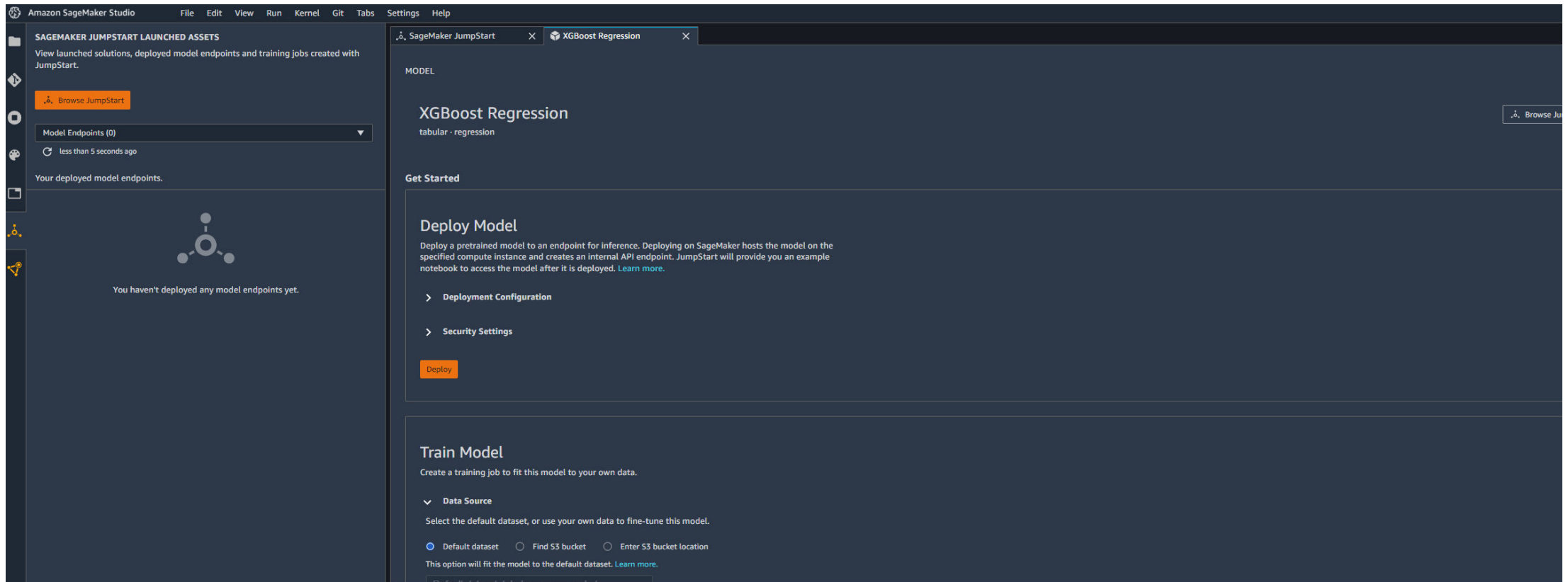
The screenshot displays the Amazon SageMaker Studio interface. On the left sidebar, under 'SAGEMAKER JUMPSTART LAUNCHED ASSETS', there is a 'Browse JumpStart' button and a section for 'Model Endpoints (0)' with a refresh icon and '2 minutes ago'. The main content area has a 'Launcher' tab selected. Below the 'SageMaker JumpStart' header, there are tabs for 'Solutions', 'Text models', 'Vision models', 'Tabular models', 'Algorithms', 'Notebooks', 'Blogs', and 'Videos'. The 'Tabular models' tab is active, showing a grid of model cards. The cards include 'Regression', 'CatBoost Classification', 'XGBoost Regression', 'XGBoost Classification', 'Linear Regression', and 'Linear Classification'. The 'XGBoost Regression' card is highlighted, showing details like 'Pre-training Dataset: Abalone', 'Trainable: Yes', and 'Source: XGBoost'. A 'View all (8)' link is visible at the bottom right of the 'Tabular models' section.

<https://www.flickr.com/photos/pasa/6757993805>

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

YOU CAN DEPLOY A PRETRAINED MODEL (MODEL TRAINED ON THE ABALONE DATASET) OR YOU CAN TRAIN THE MODEL WITH YOUR OWN CUSTOM DATASET



<https://www.flickr.com/photos/pasa/6757993805>

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

YOU CAN READ THE DESCRIPTION BELOW.

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

SAGEMAKER JUMPSTART LAUNCHED ASSETS

View launched solutions, deployed model endpoints and training jobs created with JumpStart.

Browse JumpStart

Model Endpoints (0)

17 minutes ago

Your deployed model endpoints.

You haven't deployed any model endpoints yet.

SageMaker JumpStart

XGBoost Regression

Description

XGBoost Regression

This is the XGBoost algorithm for tabular regression task. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements machine learning algorithms under the Gradient Boosting framework.

The model available for deployment is created by training XGBoost on the ABALONE dataset. The dataset contains examples of eight physical measurements such as length, diameter, and height to predict the age of abalone. Among the eight physical measurements (features), there are one categorical feature and seven numerical features.

Use the Deployed Model for Inference

The deployed model can be used for running inference on the ABALONE test set. Example python code for how to run inference on the deployed model is given in a notebook you can open by clicking 'Open Notebook' on the model endpoint page that is created when deploying the model.

For each test example, the model will output a numerical value to estimate the corresponding target value.

Below is the screenshot of the first 5 examples in the ABALONE test set. All of the test examples with features from Feature_1 to Feature_10 are sent into the deployed model to get model predictions, to estimate the ground truth Target column. Note. The categorical feature with number of classes as three has been one-hot encoded into three columns from Feature_1 to Feature_3.

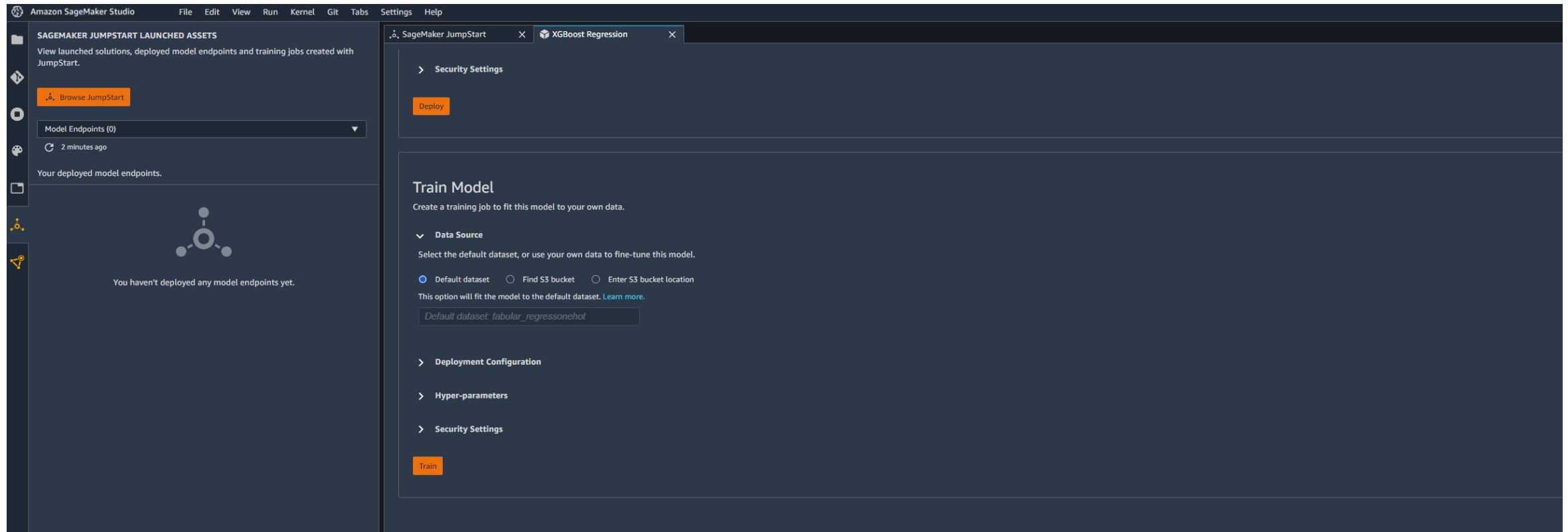
	Target	Feature_1	Feature_2	Feature_3	Feature_4	Feature_5	Feature_6	Feature_7	Feature_8	Feature_9	Feature_10
0	11	1.0	0.0	0.0	0.585	0.455	0.150	0.9870	0.4355	0.2075	0.3100
1	5	0.0	0.0	1.0	0.325	0.245	0.075	0.1495	0.0605	0.0330	0.0450
2	9	0.0	0.0	1.0	0.580	0.420	0.140	0.7010	0.3285	0.1020	0.2255
3	12	0.0	1.0	0.0	0.480	0.380	0.145	0.5900	0.2320	0.1410	0.2300
4	11	0.0	1.0	0.0	0.440	0.355	0.115	0.4150	0.1585	0.0925	0.1310

Below is the residual plot to compare the model predictions and ground truth targets. For each example, the residual value is the subtraction between the prediction and ground truth target. We can see that the points in the residual plot are randomly dispersed around the horizontal axis y = 0, which indicates the fitted regression model is appropriate for the ABALONE data.

<https://www.flickr.com/photos/pasa/6757993805>
<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

- INPUT: A DIRECTORY CONTAINS 2 SUB-DIRECTORY 'TRAIN/' 'VALIDATION/'.
- EACH SUB-DIRECTORY CONTAINS A 'DATA.CSV' FILE. THE 'DATA.CSV' FILES UNDER SUB-DIRECTORY 'TRAIN/' AND 'VALIDATION/' ARE FOR TRAINING/VALIDATION.
- THE FIRST COLUMN OF THE 'DATA.CSV' SHOULD HAVE THE CORRESPONDING TARGET VARIABLE AND THE REST CONTAINS THE FEATURES.
- YOU MUST ENCODE ALL CATEGORICAL FEATURES FIRST.
- OUTPUT: A TRAINED MODEL THAT CAN BE DEPLOYED FOR INFERENCE.



JUMP START DEMO

NAVIGATE TO S3, CREATE A DIRECTORY AND LIST THE TRAIN
AND TEST DATASET

The screenshot shows the AWS Management Console interface for an Amazon S3 bucket named 'sagemaker-us-east-1-359953057718'. The left sidebar contains navigation options for Buckets, Storage Lens, and various settings. The main content area displays the bucket's 'Objects' tab, showing a list of 17 objects. Two folders are visible: 'JumpStart-Regression/' and 'linear_learner/'. The interface includes a search bar, action buttons (Copy S3 URI, Copy URL, Download, Open, Delete, Actions, Create folder, Upload), and a table of objects with columns for Name, Type, Last modified, Size, and Storage class.

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	JumpStart-Regression/	Folder	-	-	-
<input type="checkbox"/>	linear_learner/	Folder	-	-	-

<https://www.flickr.com/photos/pasa/6757993805>

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

CREATE A NEW FOLDER IN THE DEFAULT SAGEMAKER BUCKET

aws Services Search for services, features, blogs, docs, and more [Alt+S]

S3

Amazon S3

Buckets

- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- Access analyzer for S3

Block Public Access settings for this account

▼ **Storage Lens**

- Dashboards
- AWS Organizations settings

Feature spotlight 3

► AWS Marketplace for S3

Create folder [Info](#)

Use folders to group objects in buckets. When you create a folder, S3 creates an object using the name that you specify followed by a slash (/). This object then appears as folder on the console. [Learn more](#)

ⓘ Your bucket policy might block folder creation

If your bucket policy prevents uploading objects without specific tags, metadata, or access control list (ACL) grantees, you will not be able to create a folder using this configuration. Instead, you can use the [upload configuration](#) to upload an empty folder and specify the appropriate settings.

Folder

Folder name

 /

Folder names can't contain "/". [See rules for naming](#)

Server-side encryption

ⓘ The following settings apply only to the new folder object and not to the objects contained within it.

Server-side encryption

☒ Disable

☐ Enable

Cancel Create folder

7993805

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

CREATE A SUB-DIRECTORY AND NAME IT TRAIN

aws Services Search for services, features, blogs, docs, and more [Alt+S]

S3

Amazon S3 X

Buckets

- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- Access analyzer for S3

Block Public Access settings for this account

▼ **Storage Lens**

- Dashboards
- AWS Organizations settings

Feature spotlight 3

► AWS Marketplace for S3

Amazon S3 > sagemaker-us-east-1-359953057718 > Jump-start-demo/ > Create folder

Create folder [Info](#)

Use folders to group objects in buckets. When you create a folder, S3 creates an object using the name that you specify followed by a slash (/). This object then appears as folder on the console. [Learn more](#)

Your bucket policy might block folder creation

If your bucket policy prevents uploading objects without specific tags, metadata, or access control list (ACL) grantees, you will not be able to create a folder using this configuration. Instead, you can use the [upload configuration](#) to upload an empty folder and specify the appropriate settings.

Folder

Folder name

train /

Folder names can't contain "/". [See rules for naming](#)

Server-side encryption

The following settings apply only to the new folder object and not to the objects contained within it.

Server-side encryption

☒ Disable

☐ Enable

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

CREATE A SUB-DIRECTORY AND NAME IT VALIDATION

aws Services Search for services, features, blogs, docs, and more [Alt+S]

S3

Amazon S3 X

Buckets

- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- Access analyzer for S3

Block Public Access settings for this account

▼ **Storage Lens**

- Dashboards
- AWS Organizations settings

Feature spotlight 3

► AWS Marketplace for S3

Amazon S3 > sagemaker-us-east-1-359953057718 > Jump-start-demo/ > Create folder

Create folder [Info](#)

Use folders to group objects in buckets. When you create a folder, S3 creates an object using the name that you specify followed by a slash (/). This object then appears as folder on the console. [Learn more](#)

Your bucket policy might block folder creation

If your bucket policy prevents uploading objects without specific tags, metadata, or access control list (ACL) grantees, you will not be able to create a folder using this configuration. Instead, you can use the [upload configuration](#) to upload an empty folder and specify the appropriate settings.

Folder

Folder name

validation /

Folder names can't contain "/". [See rules for naming](#)

Server-side encryption

The following settings apply only to the new folder object and not to the objects contained within it.

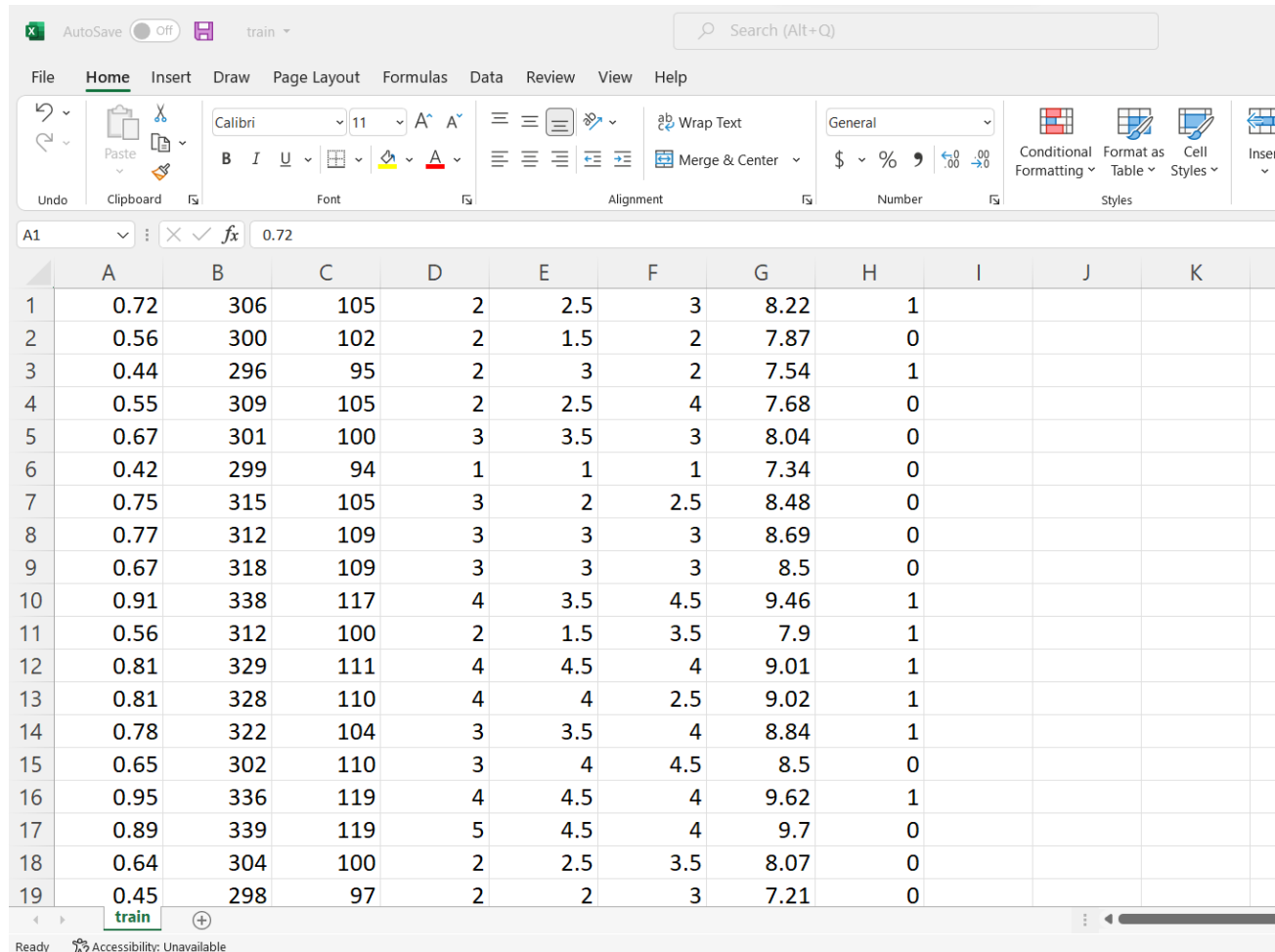
Server-side encryption

☒ Disable

☐ Enable

JUMP START DEMO

UPLOAD THE TRAINING DATASET TO TRAIN DIRECTORY. NOTE THAT THE FOLLOWING: (1) NO HEADERS, (2) TARGET VARIABLE IS LISTED FIRST, (3) YOU WILL NEED TO RENAME THE FILE TO DATA.CSV AFTER IT'S UPLOADED TO S3



The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K
1	0.72	306	105	2	2.5	3	8.22	1			
2	0.56	300	102	2	1.5	2	7.87	0			
3	0.44	296	95	2	3	2	7.54	1			
4	0.55	309	105	2	2.5	4	7.68	0			
5	0.67	301	100	3	3.5	3	8.04	0			
6	0.42	299	94	1	1	1	7.34	0			
7	0.75	315	105	3	2	2.5	8.48	0			
8	0.77	312	109	3	3	3	8.69	0			
9	0.67	318	109	3	3	3	8.5	0			
10	0.91	338	117	4	3.5	4.5	9.46	1			
11	0.56	312	100	2	1.5	3.5	7.9	1			
12	0.81	329	111	4	4.5	4	9.01	1			
13	0.81	328	110	4	4	2.5	9.02	1			
14	0.78	322	104	3	3.5	4	8.84	1			
15	0.65	302	110	3	4	4.5	8.5	0			
16	0.95	336	119	4	4.5	4	9.62	1			
17	0.89	339	119	5	4.5	4	9.7	0			
18	0.64	304	100	2	2.5	3.5	8.07	0			
19	0.45	298	97	2	2	3	7.21	0			

[sa/6757993805](#)
[hazyan/cars1](#)

JUMP START DEMO

CLICK ON RENAME OBJECT

The screenshot shows the Amazon S3 console interface. The left sidebar displays the 'Amazon S3' service with a search bar and a list of buckets. The main content area shows the 'train.csv' object in the 'Jump-start-demo' bucket. The 'Object actions' menu is open, showing options like 'Download as', 'Share with a presigned URL', 'Calculate total size', 'Copy', 'Move', 'Initiate restore', 'Query with S3 Select', and 'Edit actions'. The 'Edit actions' sub-menu is open, highlighting the 'Rename object' option.

Amazon S3 > sagemaker-us-east-1-359953057718 > Jump-start-demo/ > train/ > train.csv

train.csv [Info](#)

[Copy S3 URI](#) [Download](#) [Open](#) [Object actions](#)

Properties | [Permissions](#) | [Versions](#)

Object overview

Owner	drryanahmedaly	S3 URI	s3://sagemaker-us-east-1-359953057718/Jump-start-demo/train/train.csv
AWS Region	US East (N. Virginia) us-east-1	Amazon Resource Name (ARN)	arn:aws:s3::sagemaker-us-east-1-359953057718/Jump-start-demo/train/train.csv
Last modified			

Edit actions

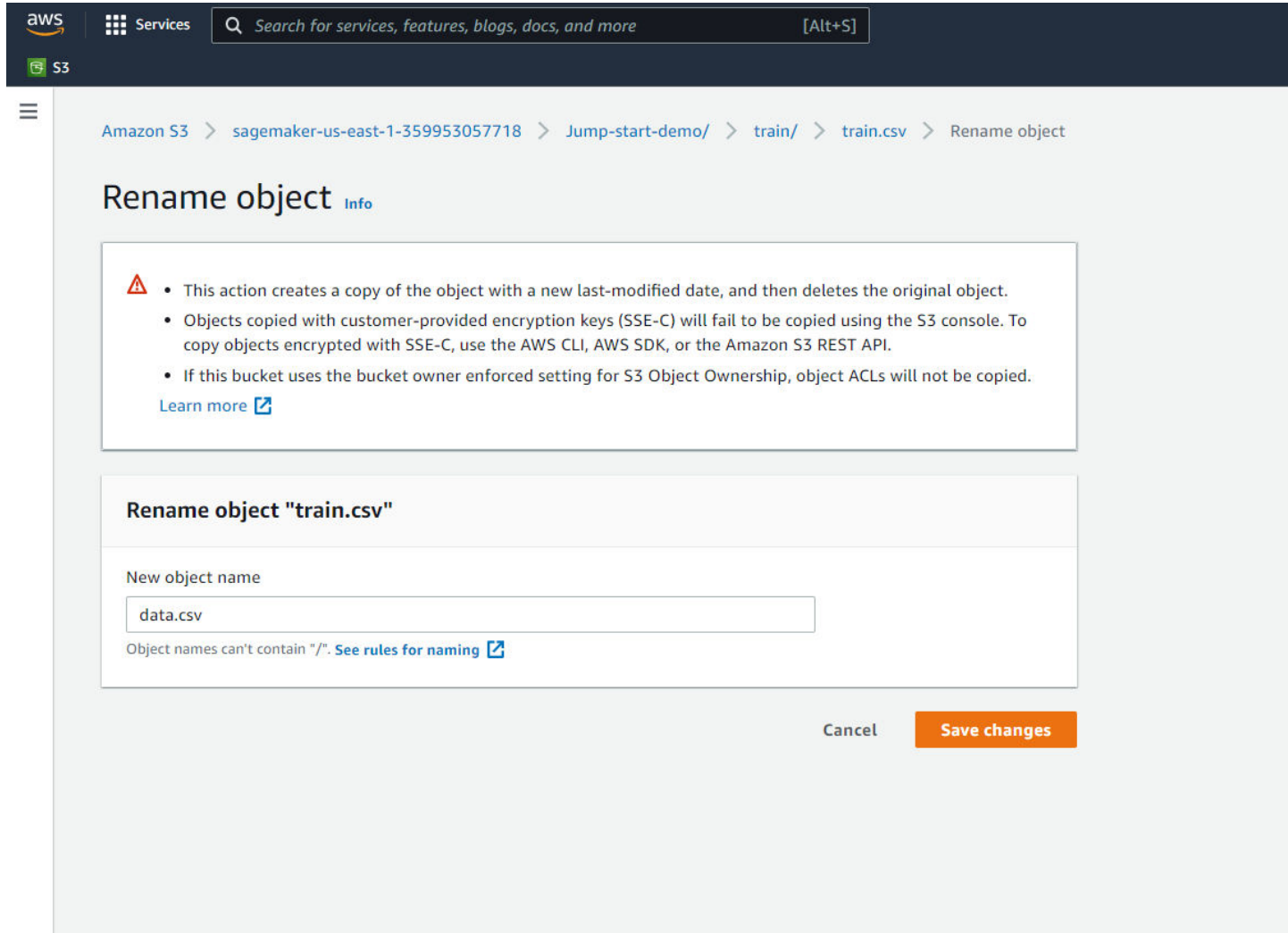
- Rename object
- Edit storage class
- Edit server-side encryption

<https://www.flickr.com/photos/pasa/6757993805>

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

RENAMED IT FROM TRAIN.CSV TO DATA.CSV



The screenshot shows the AWS S3 console interface. At the top, there's a navigation bar with the AWS logo, 'Services' link, a search bar, and an '[Alt+S]' shortcut. Below the navigation bar, the breadcrumb trail reads: 'Amazon S3 > sagemaker-us-east-1-359953057718 > Jump-start-demo/ > train/ > train.csv > Rename object'. The main heading is 'Rename object' with an 'Info' link. A warning box contains three bullet points: 'This action creates a copy of the object with a new last-modified date, and then deletes the original object.', 'Objects copied with customer-provided encryption keys (SSE-C) will fail to be copied using the S3 console. To copy objects encrypted with SSE-C, use the AWS CLI, AWS SDK, or the Amazon S3 REST API.', and 'If this bucket uses the bucket owner enforced setting for S3 Object Ownership, object ACLs will not be copied.' Below the warning box, the title 'Rename object "train.csv"' is followed by a 'New object name' label and a text input field containing 'data.csv'. A note at the bottom of the form states 'Object names can't contain "/"'. At the bottom right of the form are 'Cancel' and 'Save changes' buttons.

aws Services Search for services, features, blogs, docs, and more [Alt+S]

S3

Amazon S3 > sagemaker-us-east-1-359953057718 > Jump-start-demo/ > train/ > train.csv > Rename object

Rename object [Info](#)

- ⚠ This action creates a copy of the object with a new last-modified date, and then deletes the original object.
- Objects copied with customer-provided encryption keys (SSE-C) will fail to be copied using the S3 console. To copy objects encrypted with SSE-C, use the AWS CLI, AWS SDK, or the Amazon S3 REST API.
- If this bucket uses the bucket owner enforced setting for S3 Object Ownership, object ACLs will not be copied.

[Learn more](#)

Rename object "train.csv"

New object name

Object names can't contain "/". [See rules for naming](#)


Cancel Save changes

993805

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

DO THE SAME FOR THE VALIDATION DATASET



Successfully renamed object
"valid.csv" was renamed to "data.csv".

Amazon S3 > sagemaker-us-east-1-359953057718 > Jump-start-demo/ > validation/


validation/


Objects


Properties


Objects (1)


Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)



 Copy S3 URI

 Copy URL


 Download


 Open

Delete

Actions ▼

Create folder

 Upload

<input type="checkbox"/>	Name ▲	Type ▼	Size ▼	Storage class
<input type="checkbox"/>	 data.csv	csv	2.8 KB	Standard

<https://www.flickr.com/photos/pasa/6757993805>
<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

SELECT THE S3 BUCKET AND DATASET DIRECTORY NAME. GO AHEAD, CLICK TRAIN TO START THE TRAINING PROCESS

The screenshot displays the Amazon SageMaker Studio interface. On the left, the 'SAGEMAKER JUMPSTART LAUNCHED ASSETS' panel shows a 'Browse JumpStart' button and a 'Model Endpoints (0)' dropdown. The main workspace area contains a message: 'You haven't deployed any model endpoints yet.' On the right, the 'XGBoost Regression' configuration panel is open, showing the 'Train Model' section. Under 'Data Source', the 'Find S3 bucket' option is selected. The 'S3 bucket name' is set to 'sagemaker-us-east-1-359953057718 us-east-1' and the 'Dataset directory name' is 'Jump-start-demo/'. The 'Train' button is visible at the bottom of the configuration panel.

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

SAGEMAKER JUMPSTART LAUNCHED ASSETS
View launched solutions, deployed model endpoints and training jobs created with JumpStart.

Browse JumpStart

Model Endpoints (0)

41 minutes ago

Your deployed model endpoints.

You haven't deployed any model endpoints yet.

SageMaker JumpStart X XGBoost Regression

> Security Settings

Deploy

Train Model
Create a training job to fit this model to your own data.

> Data Source
Select the default dataset, or use your own data to fine-tune this model.

☐ Default dataset ☒ Find S3 bucket ☐ Enter S3 bucket location

Connect your data. [Learn more.](#)

S3 bucket name ⓘ
sagemaker-us-east-1-359953057718 us-east-1

Dataset directory name ⓘ
Jump-start-demo/

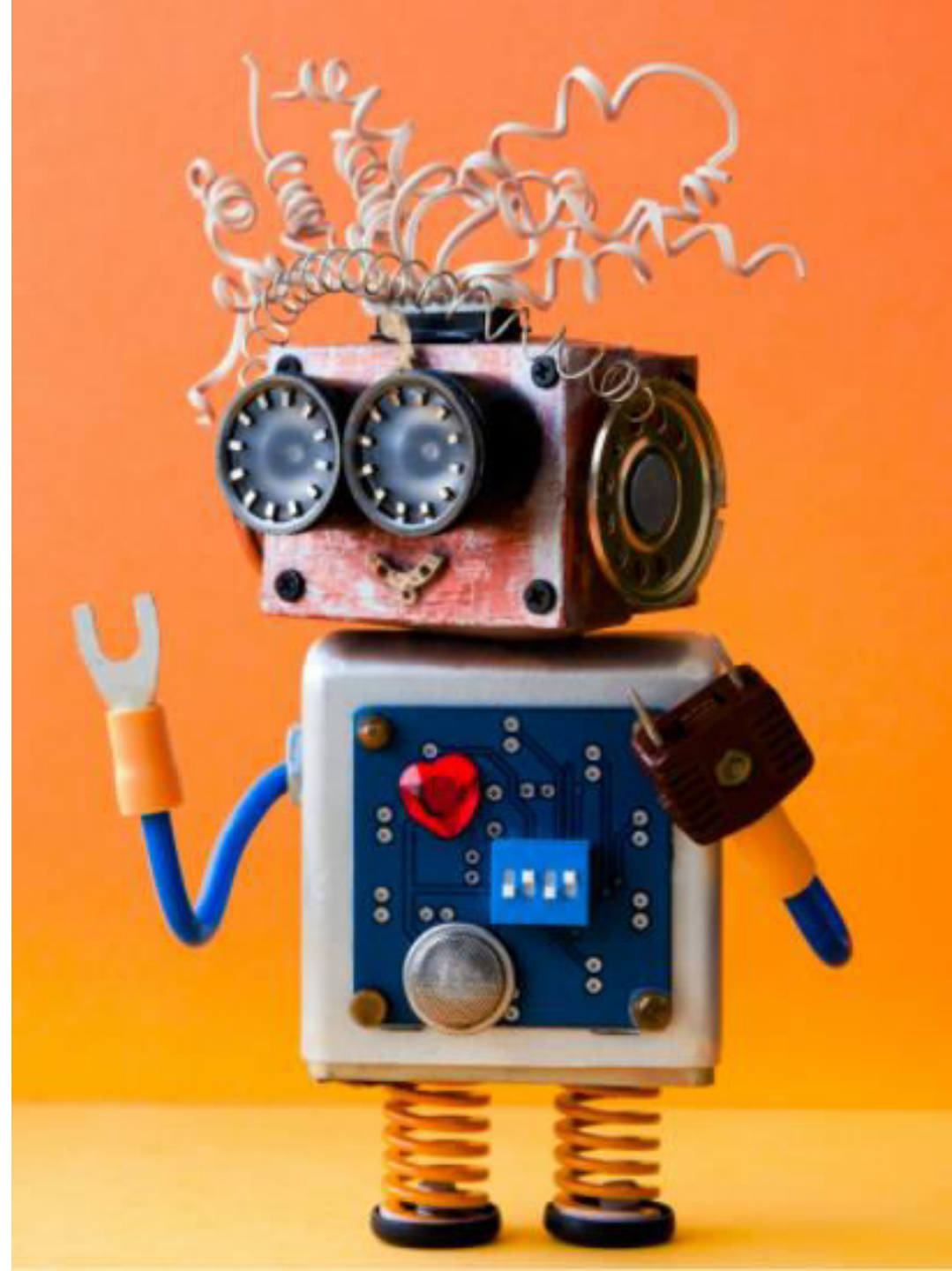
> Deployment Configuration

> Hyper-parameters

> Security Settings

Train

JUMPSTART DEMO: MODEL TRAINING



JUMP START DEMO

TRAINING PROCESS HAS JUST STARTED!

The screenshot displays the Amazon SageMaker Studio interface. On the left sidebar, under 'SAGEMAKER JUMPSTART LAUNCHED ASSETS', there is a 'Browse JumpStart' button and a 'Training Jobs (4)' dropdown menu. Below this, a list of training jobs is shown, including 'smjs-c-xgb-regression-model-20220307-092904' which is currently 'In Progress'. The main panel shows the details for this specific training job, including its name, ARN, and status. The 'Training Status' section indicates it is 'In Progress' and provides a brief description of the task. Below this, there are expandable sections for 'Instance Settings', 'Hyper Parameters', and 'Security Settings'. A 'Stop Training' button is located at the bottom of the details panel.

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

SAGEMAKER JUMPSTART LAUNCHED ASSETS

View launched solutions, deployed model endpoints and training jobs created with JumpStart.

Browse JumpStart

Training Jobs (4)

about 1 hour ago

Your training jobs for fine-tuning.

smjs-c-sklearn-regression-linear-20220226-220310

Complete - 8 days ago

smjs-d-sklearn-regression-linear-20220226-220021

Stopped - 8 days ago

smjs-c-xgb-regression-model-uni-20220226-204620

Complete - 9 days ago

smjs-c-xgb-regression-JumpStart-20220226-201822

Complete - 9 days ago

TRAINING JOB

smjs-c-xgb-regression-model-20220307-092904

Training Status

In Progress less than 10 seconds ago

The training job to fit this model to your data is in progress. This may take some time.

Base model XGBoost Regression

Model task regression

Training job name smjs-c-xgb-regression-model-20220307-092904

Training job arn arn:aws:sagemaker:us-east-1:359953057718:training-job/smjs-c-xgb-regression-model-20220307-092904

Training time ~8 seconds

Output path not available

> Instance Settings

> Hyper Parameters

> Security Settings

Stop Training

<https://www.kick.com/photos/pasa/6737993805>

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

YOU CAN VIEW THE HYPERPARAMETERS AND
INSTANCE SETTINGS

The screenshot displays the Amazon SageMaker Studio interface. On the left, a sidebar titled 'SAGEMAKER JUMPSTART LAUNCHED ASSETS' lists several training jobs. The main panel on the right shows the details for the selected job, 'smjs-c-xgb-regression-model-20220307-092904', including its instance settings, hyperparameters, and security settings.

SAGEMAKER JUMPSTART LAUNCHED ASSETS
View launched solutions, deployed model endpoints and training jobs created with JumpStart.

[Browse JumpStart](#)

Training Jobs (5)

half a minute ago

Your training jobs for fine-tuning.

- smjs-c-xgb-regression-model-20220307-092904
In Progress
- smjs-c-sklearn-regression-linear-20220226-220310
Complete · 8 days ago
- smjs-d-sklearn-regression-linear-20220226-220021
Stopped · 8 days ago
- smjs-c-xgb-regression-model-uni-20220226-204620
Complete · 9 days ago
- smjs-c-xgb-regression-JumpStart-20220226-201822
Complete · 9 days ago

Model task regression

Training job name smjs-c-xgb-regression-model-20220307-092904

Training job arn arn:aws:sagemaker:us-east-1:359953057718:training-job/smjs-c-xgb-regression-model-20220307-092904

Training time ~49 seconds

Output path not available

Instance Settings

Instance type ml.m5.4xlarge

Number of instances 1

Hyper Parameters

Colsample_bytree 1

Early_stopping_rounds 30

Gamma 0

Learning_rate 0.3

Max_depth 6

Min_child_weight 1

Num_boost_round 5000

Reg_alpha 0

Reg_lambda 1

Subsample 1

Security Settings

Network isolation enabled

IAM Role Default SageMaker Role (arn:aws:iam::359953057718:role/service-role/AmazonSageMaker-ExecutionRole-20220222T064696)

Volume encryption No custom KMS key applied.

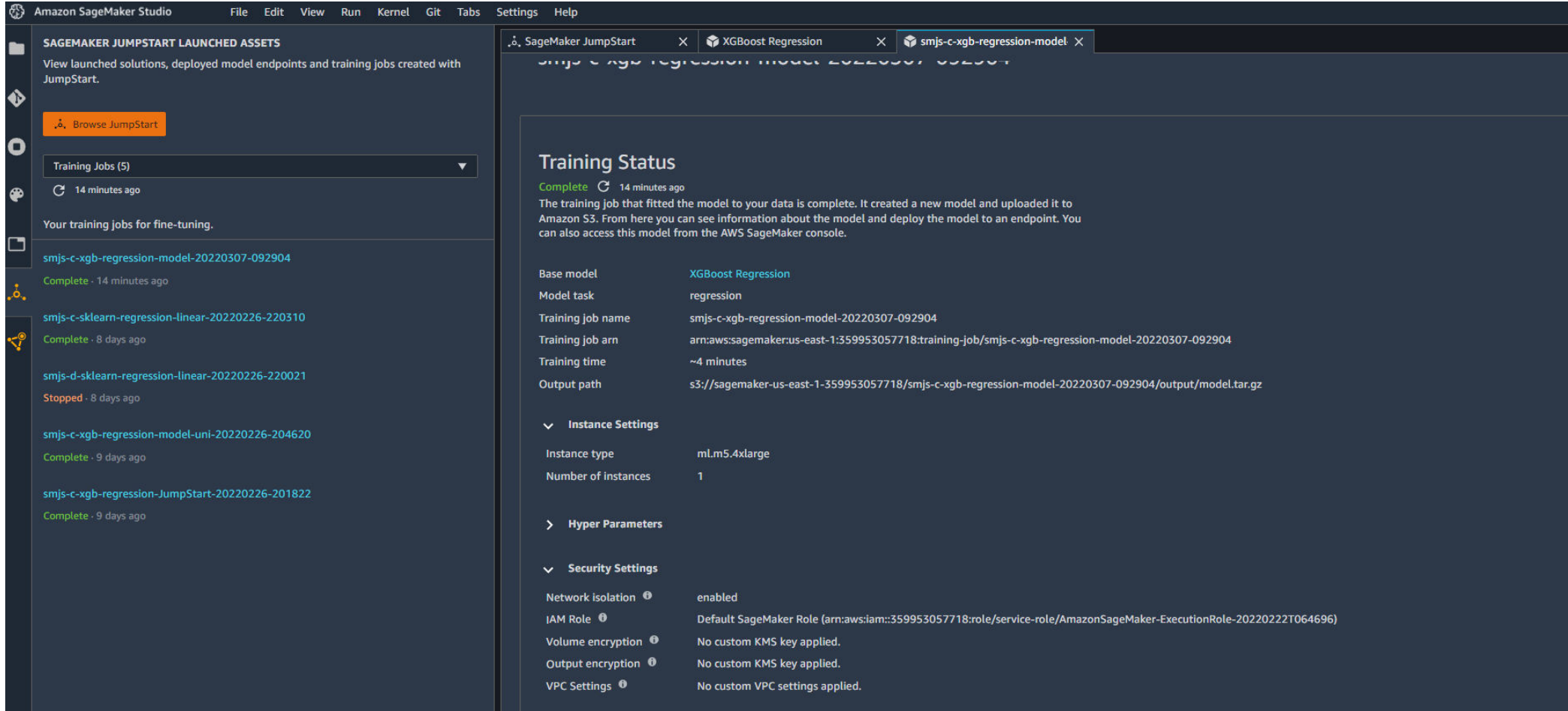
Output encryption No custom KMS key applied.

VPC Settings No custom VPC settings applied.

<https://www.kaggle.com/janjugnazyan/cars1>

JUMP START DEMO

TRAINING IS NOW COMPLETE!



The screenshot displays the Amazon SageMaker Studio interface. The left sidebar, titled "SAGEMAKER JUMPSTART LAUNCHED ASSETS", lists several training jobs. The job "smjs-c-xgb-regression-model-20220307-092904" is highlighted as "Complete" and "14 minutes ago". Below it, other jobs like "smjs-c-sklearn-regression-linear-20220226-220310" and "smjs-d-sklearn-regression-linear-20220226-220021" are also listed with their respective statuses and completion times.

The main panel shows the "Training Status" for the selected job. It indicates that the training job is "Complete" and was finished "14 minutes ago". A message states: "The training job that fitted the model to your data is complete. It created a new model and uploaded it to Amazon S3. From here you can see information about the model and deploy the model to an endpoint. You can also access this model from the AWS SageMaker console."

Below the message, the "Base model" is identified as "XGBoost Regression". The "Model task" is "regression". The "Training Job name" is "smjs-c-xgb-regression-model-20220307-092904". The "Training job arn" is "arn:aws:sagemaker:us-east-1:359953057718:training-job/smjs-c-xgb-regression-model-20220307-092904". The "Training time" is "~4 minutes". The "Output path" is "s3://sagemaker-us-east-1-359953057718/smjs-c-xgb-regression-model-20220307-092904/output/model.tar.gz".

The "Instance Settings" section shows the "Instance type" as "ml.m5.4xlarge" and the "Number of instances" as "1".

The "Hyper Parameters" section is currently collapsed.

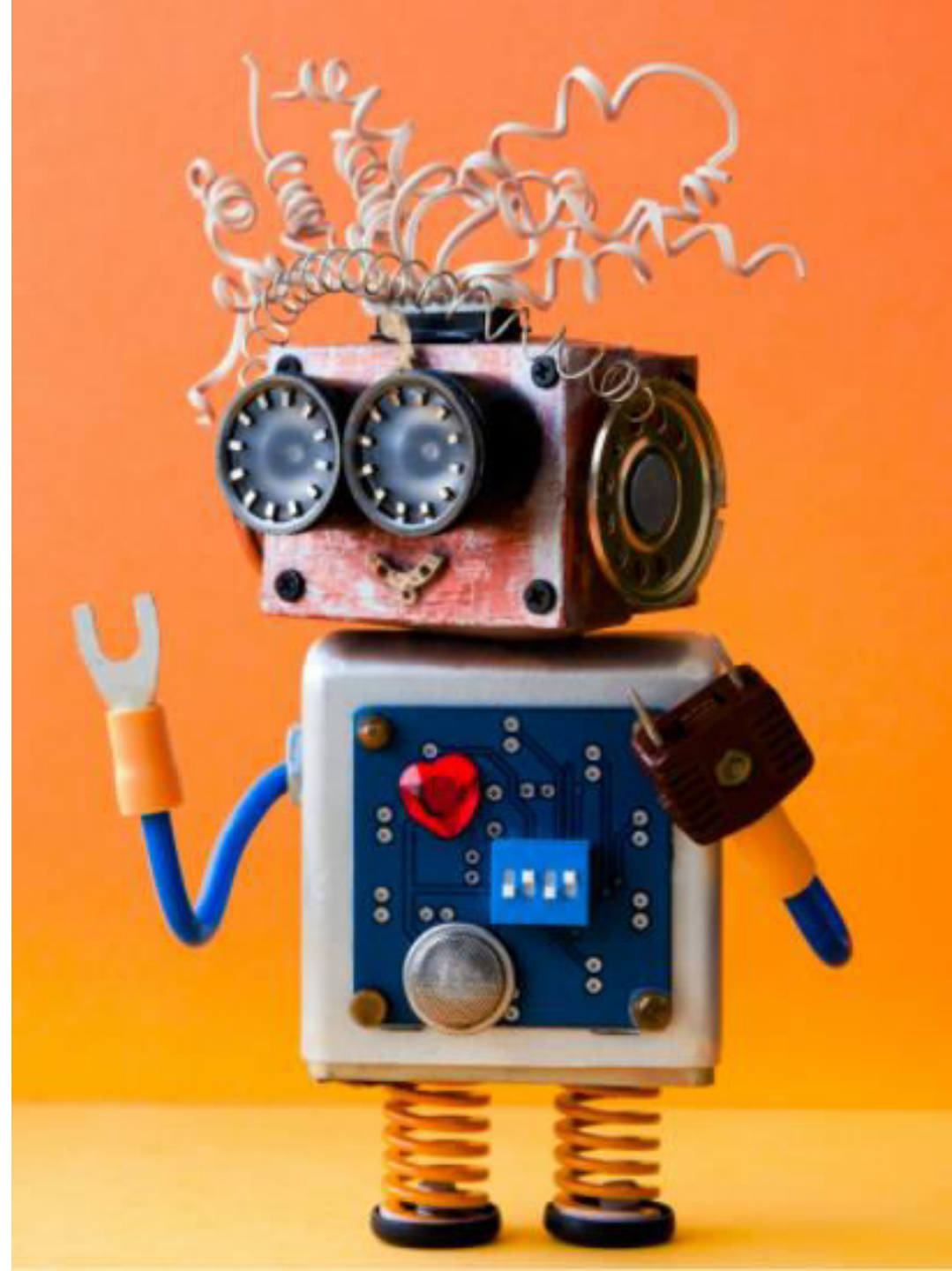
The "Security Settings" section shows the following configurations:

- Network isolation: enabled
- IAM Role: Default SageMaker Role (arn:aws:iam::359953057718:role/service-role/AmazonSageMaker-ExecutionRole-20220222T064696)
- Volume encryption: No custom KMS key applied.
- Output encryption: No custom KMS key applied.
- VPC Settings: No custom VPC settings applied.

<https://www.flickr.com/photos/pasa/6757993805>

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMPSTART DEMO: ENDPOINT DEPLOYMENT



JUMP START DEMO

MODEL DEPLOYMENT

The screenshot displays the Amazon SageMaker Studio interface. On the left sidebar, under 'SAGEMAKER JUMPSTART LAUNCHED ASSETS', there is a list of training jobs. The main panel on the right is titled 'Deploy Model' and contains configuration options for deploying a model.

SAGEMAKER JUMPSTART LAUNCHED ASSETS
View launched solutions, deployed model endpoints and training jobs created with JumpStart.

[Browse JumpStart](#)

Training Jobs (5)
21 minutes ago

Your training jobs for fine-tuning.

- smjs-c-xgb-regression-model-20220307-092904
Complete · 21 minutes ago
- smjs-c-sklearn-regression-linear-20220226-220310
Complete · 8 days ago
- smjs-d-sklearn-regression-linear-20220226-220021
Stopped · 8 days ago
- smjs-c-xgb-regression-model-uni-20220226-204620
Complete · 9 days ago
- smjs-c-xgb-regression-JumpStart-20220226-201822
Complete · 9 days ago

Deploy Model
Deploy your own model to an endpoint for inference. Deploying on SageMaker hosts the model on the specified compute instance and creates an internal API endpoint. JumpStart will provide you an example notebook to access the model after it is deployed. [Learn more.](#)

Deployment Configuration
Customize the machine type and endpoint name. [Learn more.](#)

SageMaker hosting instance
mLm5.xlarge

Endpoint name
xgb-regression-model

[Reset to default](#)

Security Settings
This model runs in network isolation. [Learn more.](#)

Specify the IAM role that Amazon SageMaker should use to deploy your model. [Learn more.](#)

☒ Default IAM role ☐ Find IAM role ☐ Input IAM role

Amazon SageMaker will deploy your model using your Studio execution role.

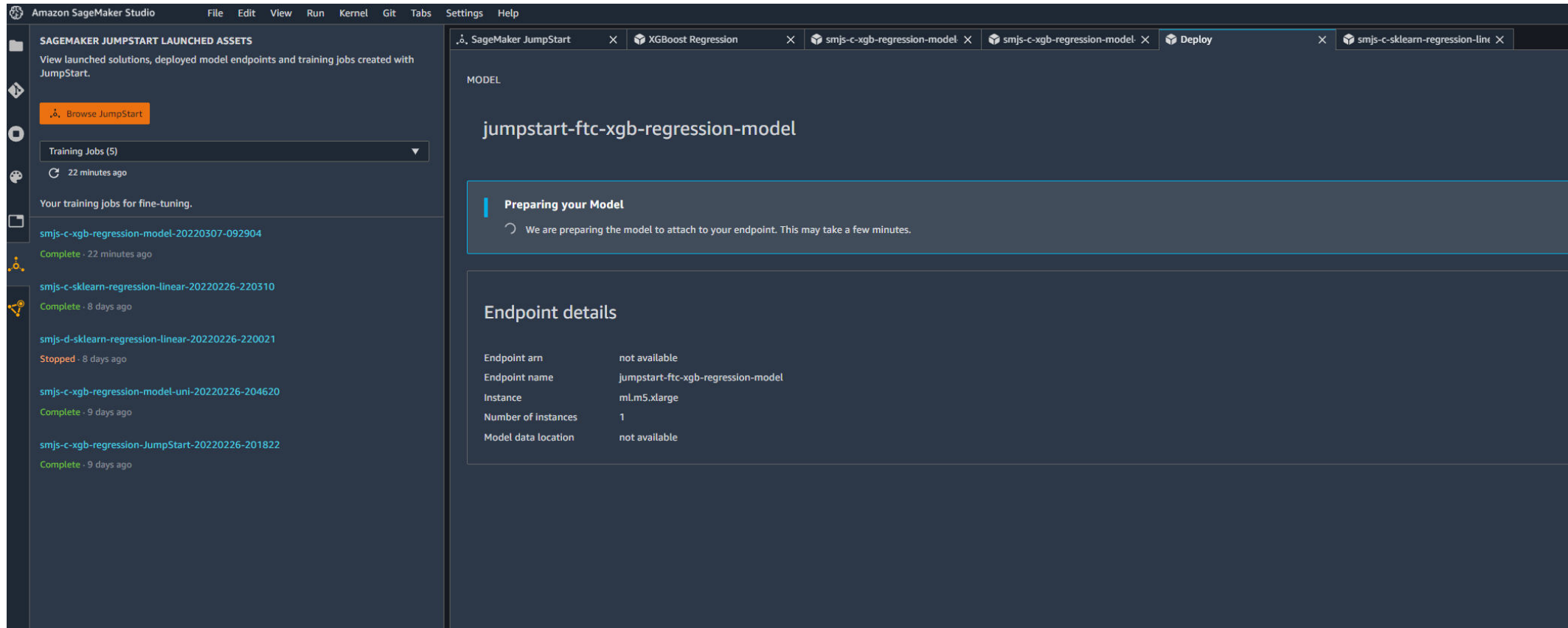
Specify whether your model should connect to a virtual private cloud (VPC). [Learn more.](#)

805

<https://www.kaggle.com/janjugnazyan/cars1>

JUMP START DEMO

MODEL IS BEING DEPLOYED...THIS SHOULD TAKE COUPLE IF MINUTES.



The screenshot displays the Amazon SageMaker Studio interface. The left sidebar, titled 'SAGEMAKER JUMPSTART LAUNCHED ASSETS', lists several training jobs. The main panel shows the 'Deploy' tab for the 'jumpstart-rtc-xgb-regression-model'. A status bar indicates 'Preparing your Model' with a progress indicator and the message: 'We are preparing the model to attach to your endpoint. This may take a few minutes.' Below this, the 'Endpoint details' section provides the following information:

Endpoint details	
Endpoint arn	not available
Endpoint name	jumpstart-rtc-xgb-regression-model
Instance	mLm5.xlarge
Number of instances	1
Model data location	not available

<https://www.flickr.com/photos/pasa/6757993805>

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

NOW THE ENDPOINT IS IN-SERVICE

The screenshot displays the Amazon SageMaker Studio interface. The left sidebar shows the 'SAGEMAKER JUMPSTART LAUNCHED ASSETS' section with a 'Browse JumpStart' button and a list of model endpoints. The main panel shows the 'MODEL ENDPOINT' details for 'jumpstart-ftc-xgb-regression-model'. The 'Endpoint Status' section indicates the endpoint is 'In Service' and operational. Below this, a table lists the endpoint's configuration details.

Property	Value
Base model	XGBoost Regression
Model task	regression
Endpoint arn	arn:aws:sagemaker:us-east-1:359953057718:endpoint/jumpstart-ftc-xgb-regression-model
Endpoint name	jumpstart-ftc-xgb-regression-model
Instance	mL.m5.xlarge
Number of instances	1
Model data location	s3://sagemaker-us-east-1-359953057718/xgboost-regression-model-20220307-095513/model.tar.gz

Below the table, there is a 'Security settings' section with a right-pointing arrow. At the bottom of the main panel, there is a 'Use Endpoint from Studio' section with a description and an 'Open Notebook' button.

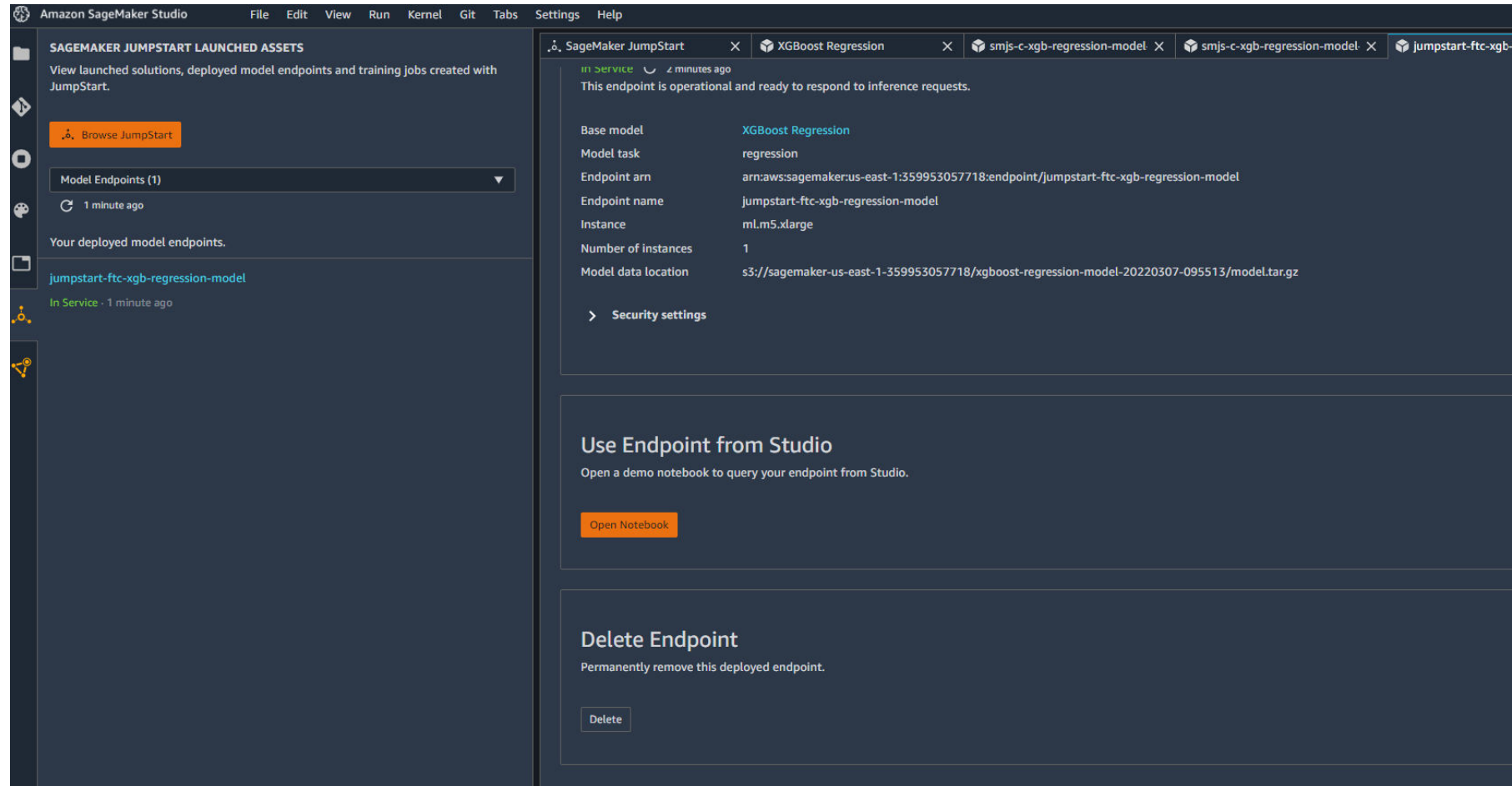
<https://www.kaggle.com/ljanjughazyan/cars1>

JUMPSTART DEMO: INVOKE ENDPOINT



JUMP START DEMO

TO MAKE INFERENCE ON THE ENDPOINT, CLICK ON OPEN
NOTEBOOK TO START MAKING QUERIES

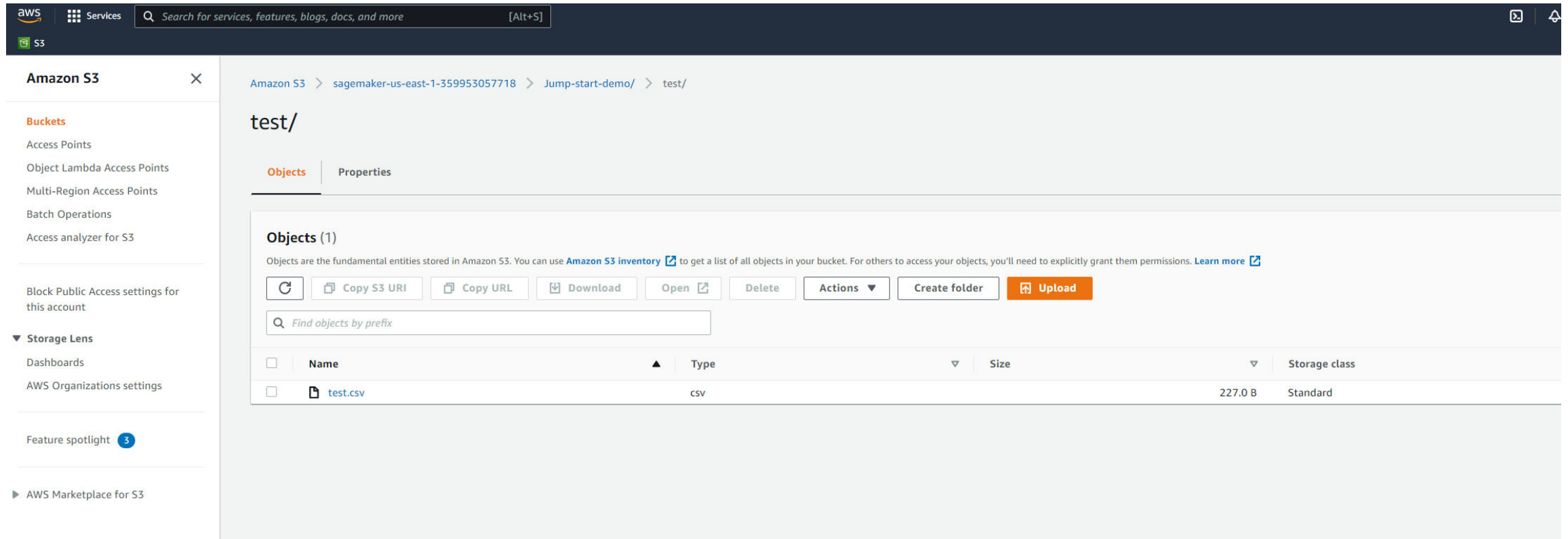


<https://www.flickr.com/photos/pasa/6757993805>

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

UPLOAD THE TESTING DATASETS TO THE S3 BUCKET



The screenshot shows the Amazon S3 console interface. The left sidebar contains navigation options for Amazon S3, including Buckets, Access Points, Object Lambda Access Points, Multi-Region Access Points, Batch Operations, and Access analyzer for S3. The main content area displays the 'test/' bucket. The 'Objects' tab is selected, showing a list of objects. There is one object named 'test.csv' with a size of 227.0 B and a storage class of Standard. The interface includes a search bar, a list of actions (Copy S3 URI, Copy URL, Download, Open, Delete, Actions, Create folder, Upload), and a table of objects.

Amazon S3 > sagemaker-us-east-1-359953057718 > Jump-start-demo/ > test/

test/

Objects Properties

Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Refresh Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

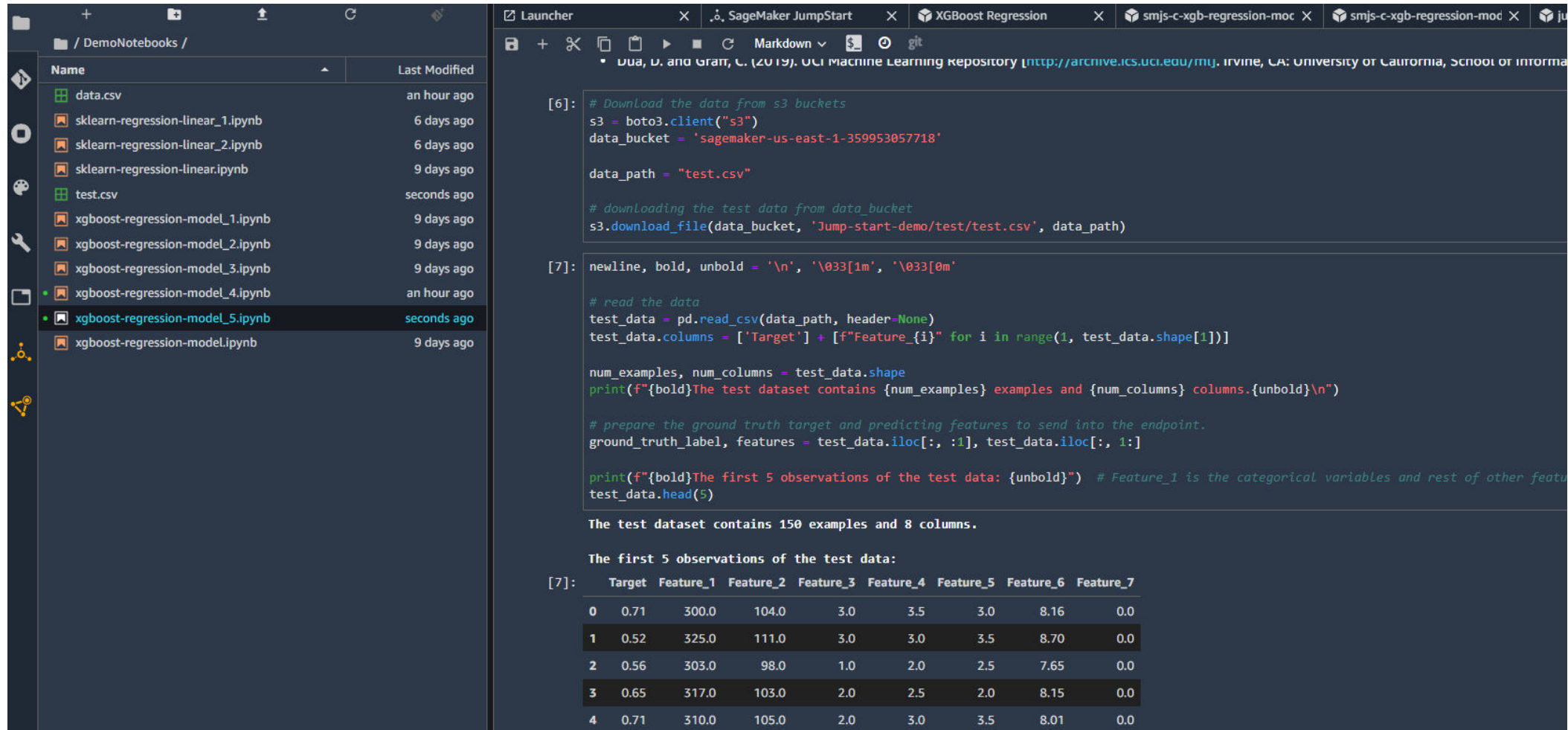
Find objects by prefix

<input type="checkbox"/>	Name	Type	Size	Storage class
<input type="checkbox"/>	test.csv	csv	227.0 B	Standard

<https://www.flickr.com/photos/pasa/6757993805>
<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

TO MAKE INFERENCE ON THE ENDPOINT, CLICK ON OPEN
NOTEBOOK TO START MAKING QUERIES



The screenshot displays the SageMaker JumpStart interface. On the left, a file explorer shows the directory structure under 'DemoNotebooks /'. The files listed include 'data.csv', several 'sklearn-regression-linear.ipynb' files, 'test.csv', and multiple 'xgboost-regression-model_*.ipynb' files. The file 'xgboost-regression-model_5.ipynb' is currently selected and highlighted.

The main area shows the content of the selected notebook, which is a Jupyter notebook. The code in the notebook is as follows:

```
[6]: # Download the data from s3 buckets
s3 = boto3.client("s3")
data_bucket = 'sagemaker-us-east-1-359953057718'

data_path = "test.csv"

# downloading the test data from data_bucket
s3.download_file(data_bucket, 'Jump-start-demo/test/test.csv', data_path)

[7]: newline, bold, unbold = '\n', '\033[1m', '\033[0m'

# read the data
test_data = pd.read_csv(data_path, header=None)
test_data.columns = ['Target'] + [f"Feature_{i}" for i in range(1, test_data.shape[1])]

num_examples, num_columns = test_data.shape
print(f"{bold}The test dataset contains {num_examples} examples and {num_columns} columns.{unbold}\n")

# prepare the ground truth target and predicting features to send into the endpoint.
ground_truth_label, features = test_data.iloc[:, :1], test_data.iloc[:, 1:]

print(f"{bold}The first 5 observations of the test data: {unbold}") # Feature_1 is the categorical variables and rest of other featu
test_data.head(5)

The test dataset contains 150 examples and 8 columns.

The first 5 observations of the test data:
```

	Target	Feature_1	Feature_2	Feature_3	Feature_4	Feature_5	Feature_6	Feature_7
0	0.71	300.0	104.0	3.0	3.5	3.0	8.16	0.0
1	0.52	325.0	111.0	3.0	3.0	3.5	8.70	0.0
2	0.56	303.0	98.0	1.0	2.0	2.5	7.65	0.0
3	0.65	317.0	103.0	2.0	2.5	2.0	8.15	0.0
4	0.71	310.0	105.0	2.0	3.0	3.5	8.01	0.0

<https://www.kaggle.com/ijanjughazyah/cars1>

JUMP START DEMO

ASSESS TRAINED MODEL PERFORMANCE

+

+

↑

↺

⚙

/ DemoNotebooks /

Name	Last Modified
data.csv	an hour ago
sklearn-regression-linear_1.ipynb	6 days ago
sklearn-regression-linear_2.ipynb	6 days ago
sklearn-regression-linear.ipynb	9 days ago
test.csv	3 minutes ago
xgboost-regression-model_1.ipynb	9 days ago
xgboost-regression-model_2.ipynb	9 days ago
xgboost-regression-model_3.ipynb	9 days ago
xgboost-regression-model_4.ipynb	an hour ago
xgboost-regression-model_5.ipynb	seconds ago
xgboost-regression-model.ipynb	9 days ago

Launcher

.SageMaker JumpStart

XGBoost Regression

smjs-c-xgb-regression-model

+

+

✂

📄

▶

■

↺

Code

⌵

\$

⌂

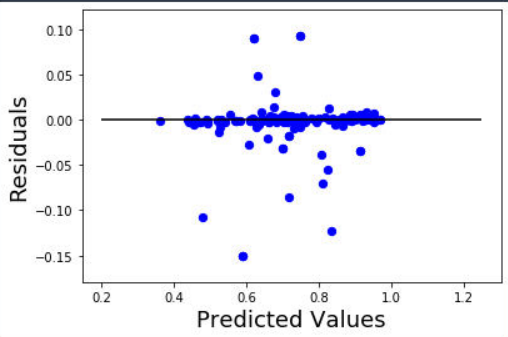
git

```
try:
    model_predictions = parse_response(query_response)
except (TypeError, KeyError) as e:
    raise Exception(
        "Backend scripts have been updated in February '22 to standardize response "
        "format of endpoint response."
        "Response from previous endpoints may not be consistent with this notebook."
        f"To use this notebook, please launch the endpoint again. Error: {e}."
    )
```

Evaluate the predictions results returned from the endpoint

[13]: *# Visualization: a residual plot to compare the model predictions and ground truth targets. For each example, # is the subtraction between the prediction and ground truth target. # We can see that the points in the residual plot are randomly dispersed around the horizontal axis y = 0, # which indicates the fitted regression model is appropriate for the ABALONE data*

```
residuals = ground_truth_label.values[:, 0] - model_predictions
plt.scatter(model_predictions, residuals, color="blue", s=40)
plt.hlines(y=0, xmin=0.2, xmax=1.25)
plt.xlabel('Predicted Values', fontsize=18)
plt.ylabel('Residuals', fontsize=18)
plt.show()
```



<https://www.flickr.com/photos/pasa/6757993805>

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

PRINT OUT REGRESSION MODELS EVALUATION METRICS

```
[10]: # Evaluate the model predictions quantitatively.
eval_r2_score = r2_score(ground_truth_label.values, model_predictions)
eval_mse_score = mean_squared_error(ground_truth_label.values, model_predictions)
eval_mae_score = mean_absolute_error(ground_truth_label.values, model_predictions)
print (
    f"{bold}Evaluation result on test data{unbold}:{newline}"
    f"{bold}{r2_score.__name__}{unbold}: {eval_r2_score}{newline}"
    f"{bold}{mean_squared_error.__name__}{unbold}: {eval_mse_score}{newline}"
    f"{bold}{mean_absolute_error.__name__}{unbold}: {eval_mae_score}{newline}"
)
```

```
Evaluation result on test data:
r2_score: 0.9557161949003675
mean_squared_error: 0.0008903605413045446
mean_absolute_error: 0.011734435749053957
```

```
[ ]:
```

<https://www.flickr.com/photos/pasa/6757993805>

<https://www.kaggle.com/ljanjughazyan/cars1>

JUMP START DEMO

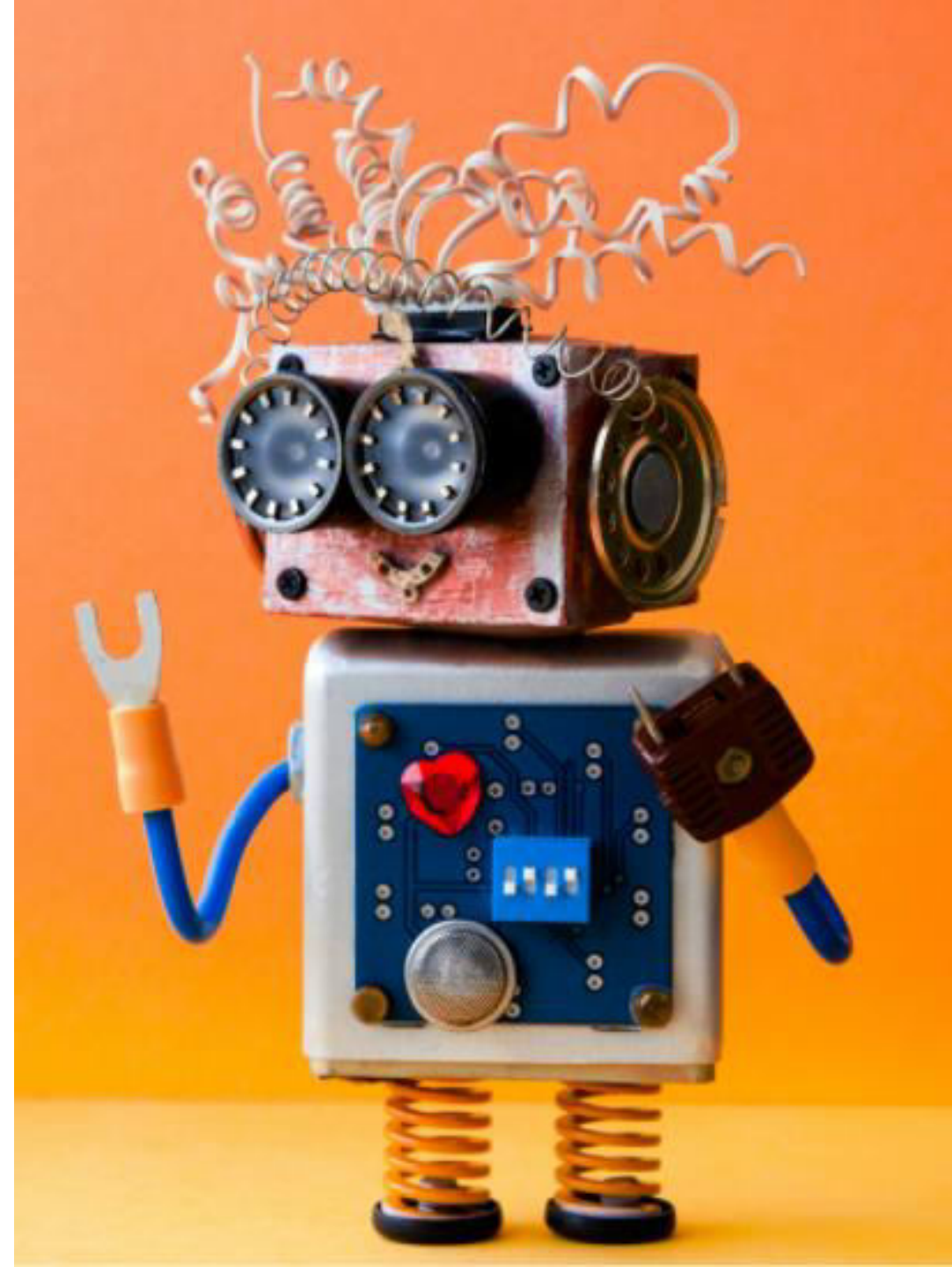
**MAKE SURE TO DELETE THE DEPLOYED ENDPOINT AND
SHUT DOWN ALL RUNNING INSTANCES**

The screenshot shows the Amazon SageMaker console interface. The top navigation bar includes the AWS logo, a search bar, and the user's profile. The left sidebar contains the navigation menu, with 'Endpoints' selected under the 'Inference' section. The main content area displays the 'Endpoints' page, which includes a search bar and a table of endpoints.

Name	ARN	Status
jumpstart-ffc-xgb-regression-model-1	arn:aws:sagemaker:us-east-1:359953057718:endpoint/jumpstart-ffc-xgb-regression-model-1	InService
jumpstart-ffc-xgb-regression-model	arn:aws:sagemaker:us-east-1:359953057718:endpoint/jumpstart-ffc-xgb-regression-model	InService

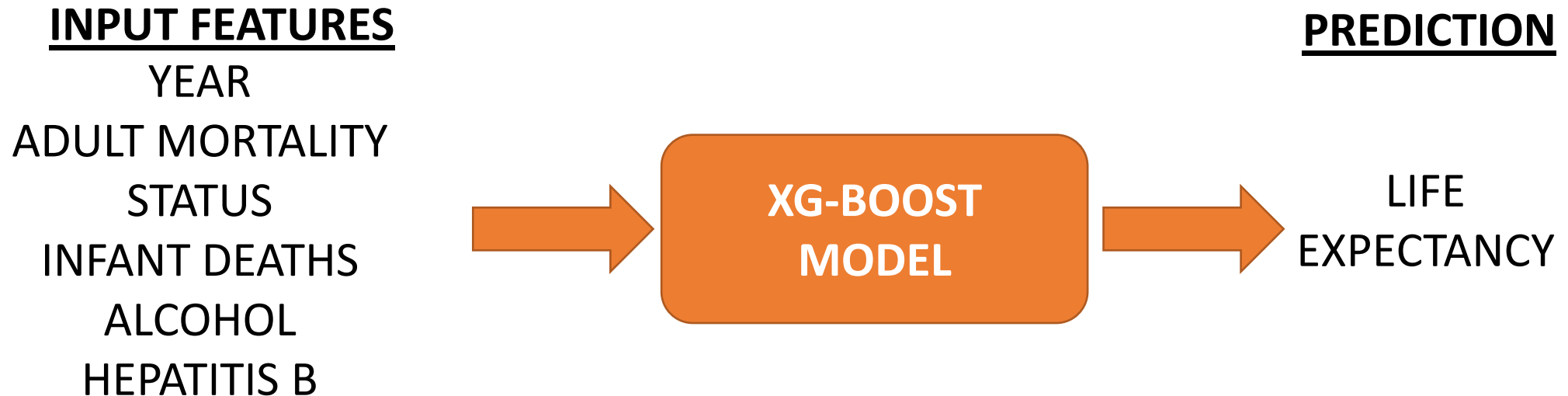
<https://www.flickr.com/photos/pasa/6757993805>
<https://www.kaggle.com/ljanjughazyan/cars1>

FINAL END-OF-DAY CAPSTONE PROJECT

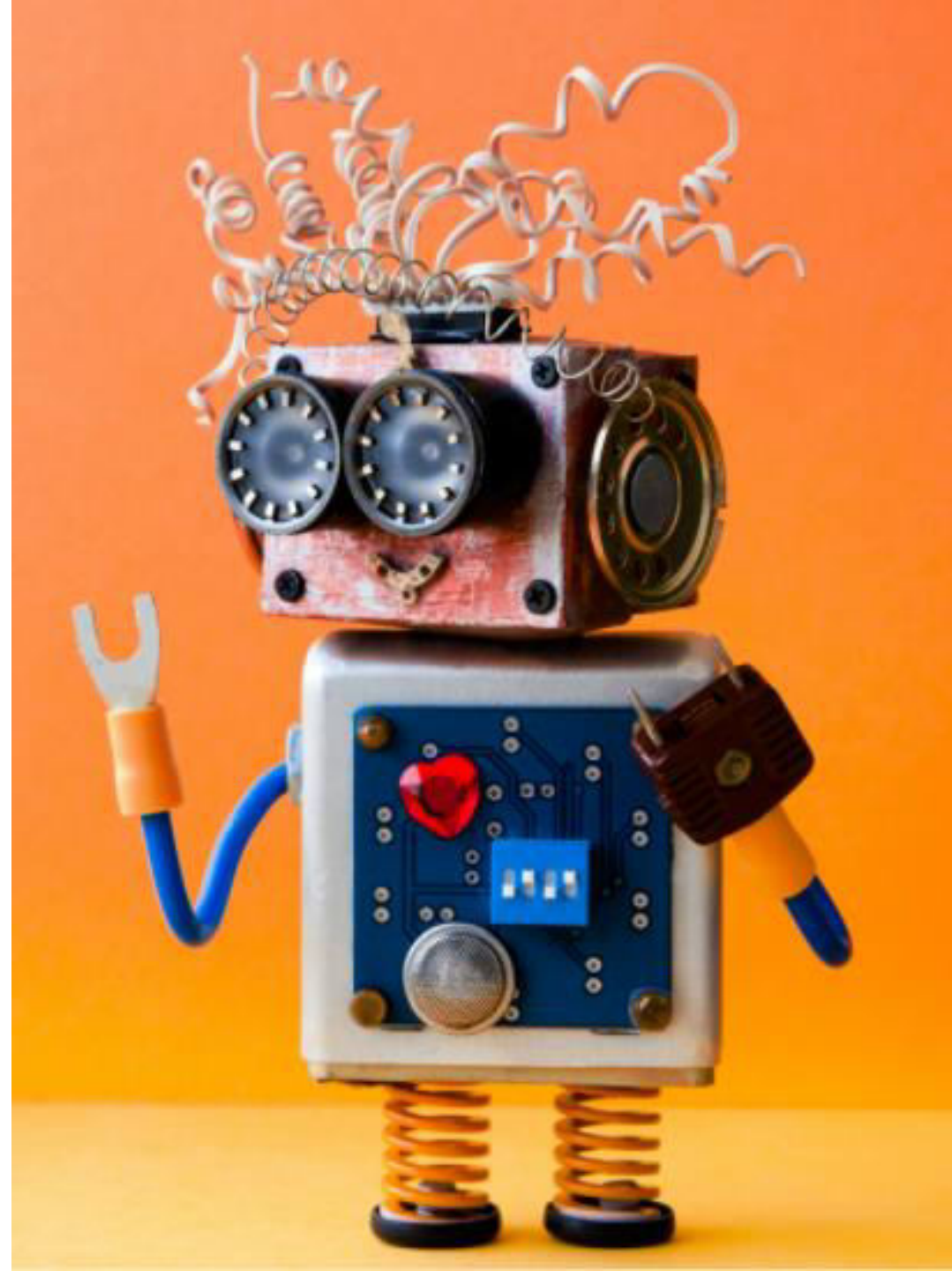


PROJECT OVERVIEW: SAGEMAKER JUMPSTART

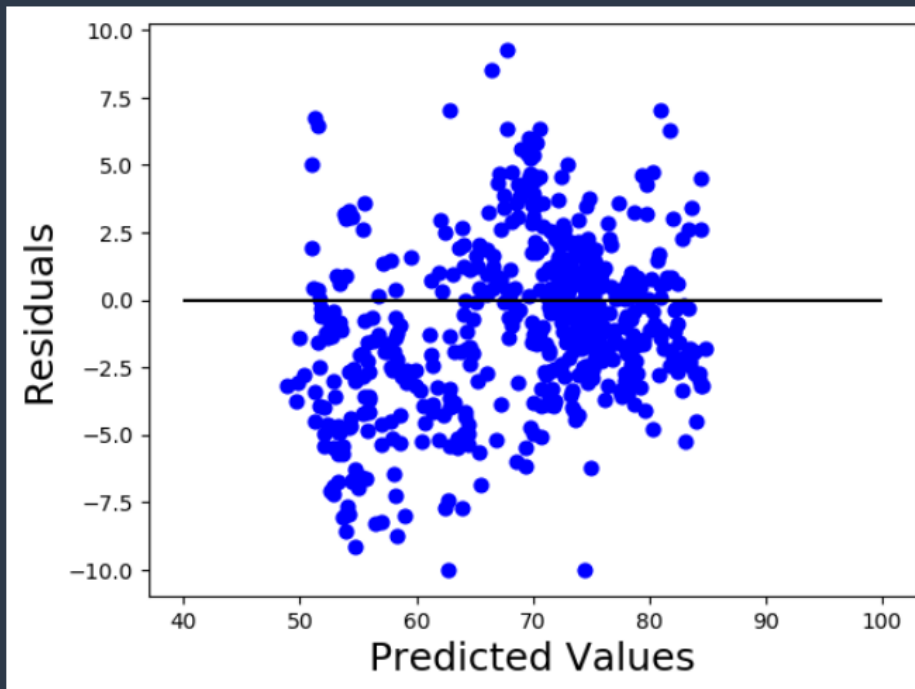
- In this hands-on project, we will train an XG-Boost Regression model to predict life expectancy using SageMaker Jumpstart.
- This data was initially obtained from World Health Organization (WHO) and United Nations Website. Data contains features like year, status, life expectancy, adult mortality, infant deaths, percentage of expenditure, alcohol etc.
- **Tasks:**
 1. Upload the dataset Life_Expectancy_test_NoHeader.csv and Life_Expectancy_train_NoHeader.csv to S3
 2. Using AWS SageMaker JumpStart, train an XG-Boost model to predict life expectancy.
 3. Deploy the model and assess its performance. What's R2?



FINAL END-OF-DAY CAPSTONE PROJECT SOLUTION



FINAL PROJECT SOLUTION



```
[10]: # Evaluate the model predictions quantitatively.
eval_r2_score = r2_score(ground_truth_label.values, model_predictions)
eval_mse_score = mean_squared_error(ground_truth_label.values, model_predictions)
eval_mae_score = mean_absolute_error(ground_truth_label.values, model_predictions)
print (
    f"{bold}Evaluation result on test data{unbold}:{newline}"
    f"{bold}{r2_score.__name__}{unbold}: {eval_r2_score}{newline}"
    f"{bold}{mean_squared_error.__name__}{unbold}: {eval_mse_score}{newline}"
    f"{bold}{mean_absolute_error.__name__}{unbold}: {eval_mae_score}{newline}"
)
```

```
Evaluation result on test data:
r2_score: 0.905099280527262
mean_squared_error: 9.99137901437254
mean_absolute_error: 2.451313835233201
```

<https://www.kaggle.com/janjugnazyan/cars1>

FINAL PROJECT SOLUTION

**MAKE SURE TO DELETE THE DEPLOYED ENDPOINT AND
SHUT DOWN ALL RUNNING INSTANCES**

The screenshot shows the Amazon SageMaker console interface. The top navigation bar includes the AWS logo, a search bar, and the user's profile. The left sidebar contains the navigation menu, with 'Endpoints' selected under the 'Inference' section. The main content area displays the 'Endpoints' page, which includes a search bar and a table of endpoints.

Name	ARN	Status
jumpstart-ffc-xgb-regression-model-1	arn:aws:sagemaker:us-east-1:359953057718:endpoint/jumpstart-ffc-xgb-regression-model-1	InService
jumpstart-ffc-xgb-regression-model	arn:aws:sagemaker:us-east-1:359953057718:endpoint/jumpstart-ffc-xgb-regression-model	InService

<https://www.flickr.com/photos/pasa/6757993805>

<https://www.kaggle.com/ljanjughazyan/cars1>