

# SageMaker Jumpstart

Use MyApps

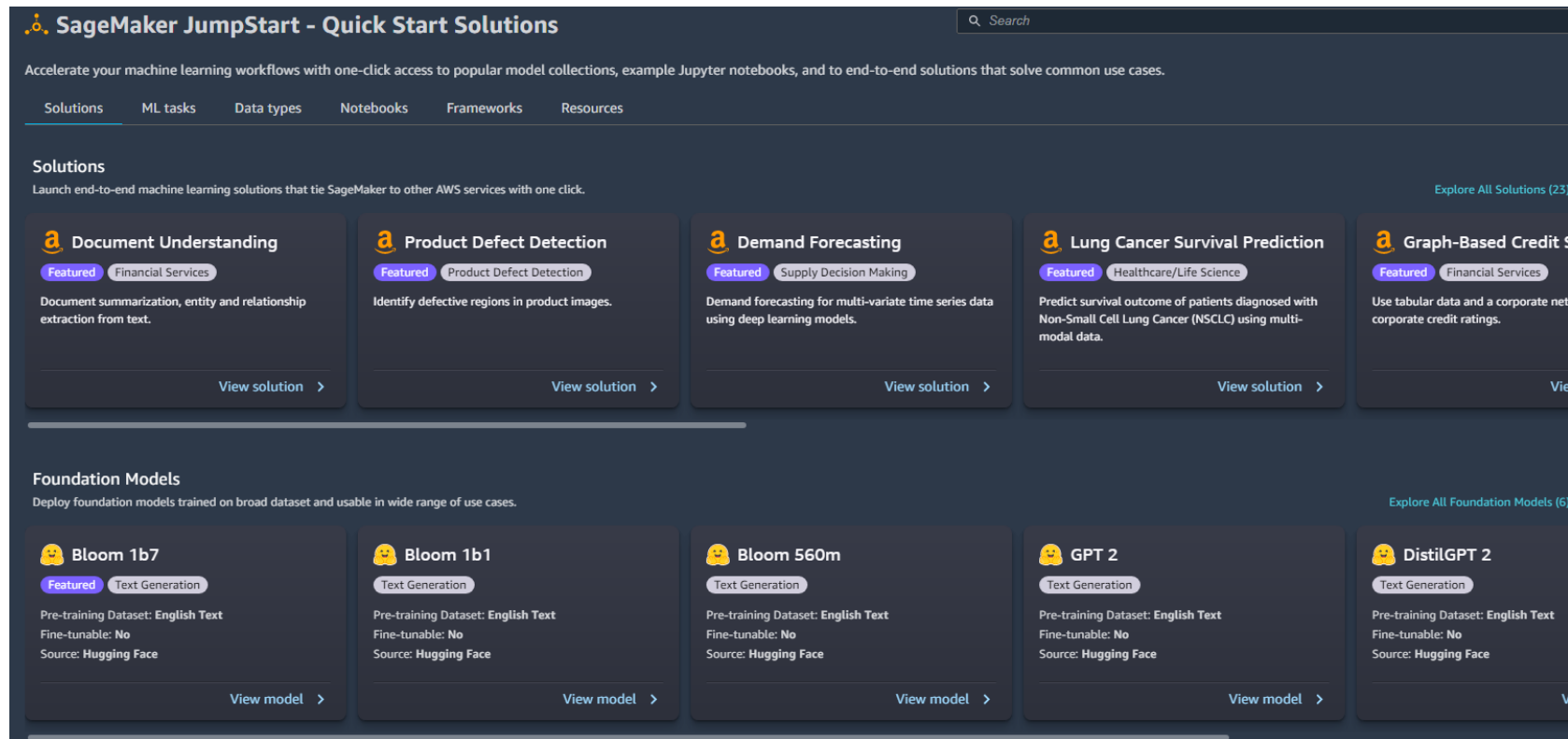
In ca-central

# What is SageMaker Jumpstart

- It is a service that has a set of best practices to create an ML solution
- There are more than a dozen of solutions available in Jumpstart
- The use cases are based on multiple industries like manufacturing, retail and finance

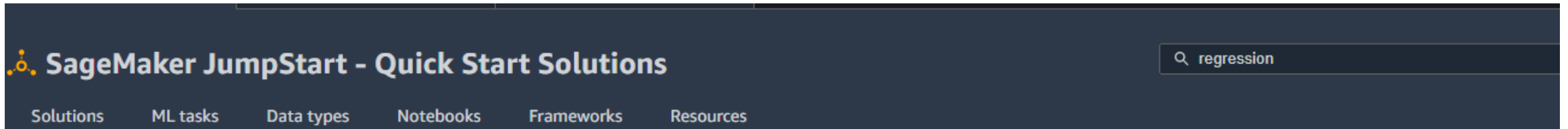
# Interface to work with JumpStart

- Solutions based on ML tasks, Data types, notebooks, frameworks



# We want to try a regression model

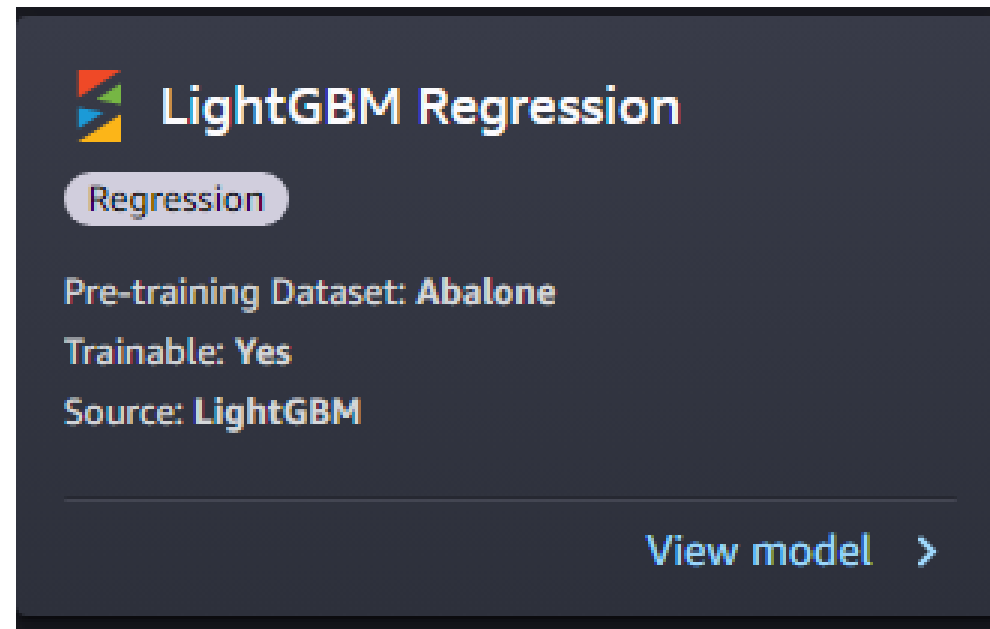
- Search **regression** in the SageMaker Jumpstart



- Select **Regression with XGBoost** → That is the just to see what's inside that notebook. The notebook is about estimating Abalone age by the data from UCI data repository
- But we do not use this data and instead we use a data set to predict the **likelihood of acceptance for university**

# LightGBM Regression

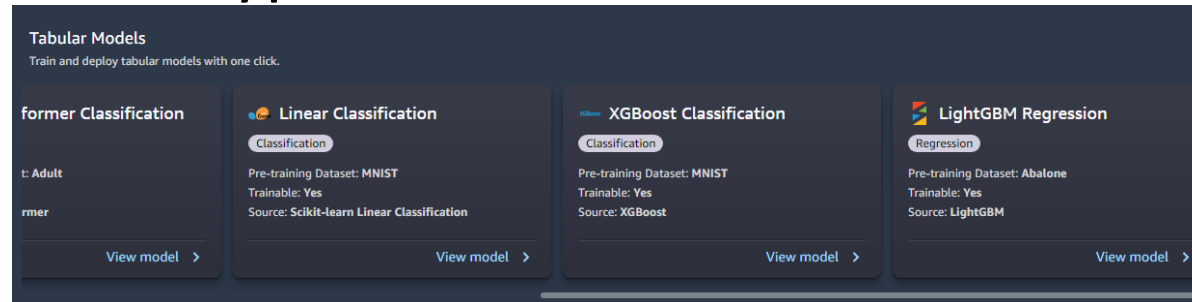
- On the same page you also see a model that is pre-trained on the Abalone data set



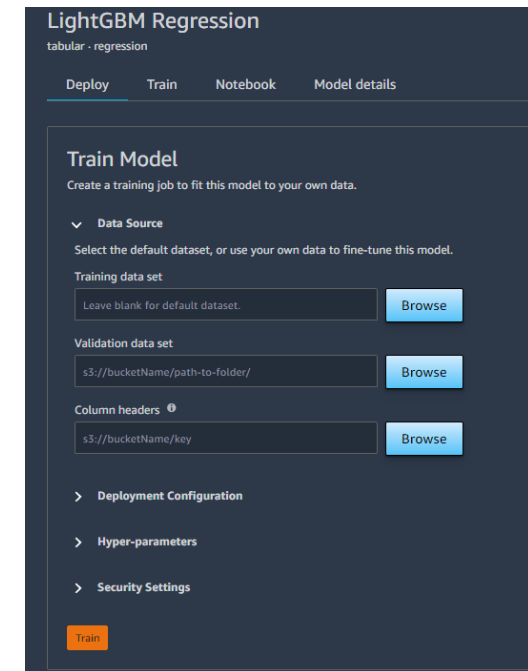
- Click on the LightGBM regression tile

# LightGBM Regression

- Alternatively: In StudioJS → data types → Tabular and select **LightGBM Regression**:



- We can set the parameters in the JS and just click on Train



# Introduction to the data set

- We want to use a data set to create a model to estimate the chance of admission in university based on GRE, TOEFL and other criterial

	A	B	C	D	E	F	G	H	I
1	GRE_Score	TOEFL_Scc	University_SOP	LOR	CGPA	Research	Chance_of_Admission		
2	337	118	4	4.5	4.5	9.65	1	0.92	
3	324	107	4	4	4.5	8.87	1	0.76	
4	316	104	3	3	3.5	8	1	0.72	

▼ Data Source

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

Training data set

Leave blank for default dataset.

Browse

Validation data set

s3://bucketName/path-to-folder/

Browse

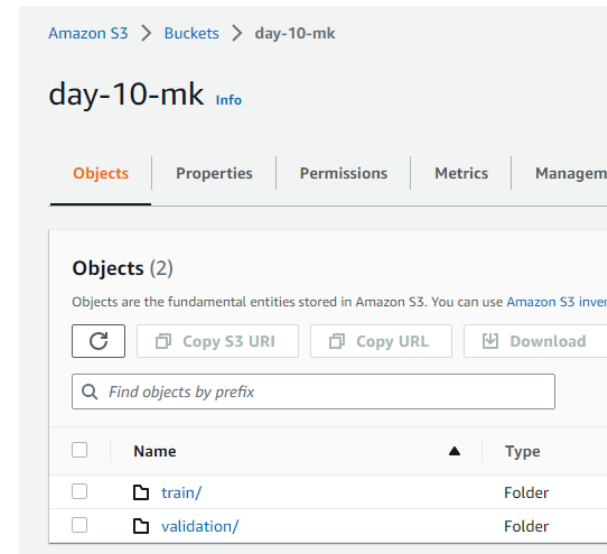
Column headers ⓘ

s3://bucketName/key

Browse

# Upload Dataset

- I have uploaded 3 data files for university admission (training, validation and test data sets)
- Upload the training and validation files in **two folders** folders in a bucket (you need to create a bucket ). The name of folders must be **train** and **validation**.





# Rename the training dataset

- In the notebook it is assumed that the training and validation dataset names are **data.csv**

## Train the Model on a New Dataset

Below are the instructions for how the training data should be formatted for input to the model.

- **Input:** A directory containing two sub-directories 'train/' and 'validation/' (optional), and a json-format file named 'categorical\_index.json' (optional). Each sub-directory contains a 'data.csv' file.
  - The 'data.csv' files under sub-directory 'train/' and 'validation/' are for training and validation, respectively. The validation data is used to compute a validation score at the end of each boosting iteration. An early stopping is applied when the validation score stops improving. If the validation data is not provided, a 20% of training data is randomly sampled to serve as the validation data.
  - The first column of the 'data.csv' should have the corresponding target variable. The rest of other columns should have the corresponding predictor variables (features).

- Rename the datasets to data.csv in both folders

## Rename object [Info](#)

- ⚠ This action creates a copy of the object with a
- Objects copied with customer-provided encryption objects encrypted with SSE-C, use the AW
- If this bucket uses the bucket owner enforced :  
[Learn more](#)

## Rename object "university\_train.csv"

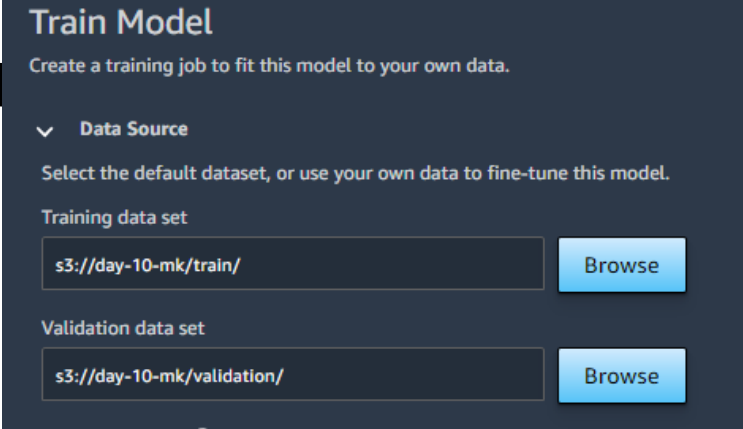
New object name

data.csv

Object names can't contain "/". See rules for naming

# Configure the jumpstart

- Copy the train and validation folder in the respective field
- **NOTE:** You need to put **bucket name/foldername**.



The screenshot shows a 'Train Model' configuration window. It has a title bar 'Train Model' and a subtitle 'Create a training job to fit this model to your own data.' Below this is a section 'Data Source' with a dropdown arrow. Under 'Data Source' is the instruction 'Select the default dataset, or use your own data to fine-tune this model.' There are two input fields: 'Training data set' and 'Validation data set'. The 'Training data set' field contains the text 's3://day-10-mk/train/' and has a 'Browse' button to its right. The 'Validation data set' field contains the text 's3://day-10-mk/validation/' and also has a 'Browse' button to its right.


- In **Deployment Configuration:** Change the instance type to smaller instance like m5.xlarge
- Review hyperparameters
- In Security Settings:
  - Change the IAM role by selecting **Input IAM Role**
  - Change it to something like:  
**arn:aws:iam::239630988601:role/fast-ai-academic-1-Student-Azure**
  - Change the red items according to your account info

# Start training

- Click on **Train**
- After a few minutes the Model is ready

smjs-c-lgb-regression-model-20221122-034557

### Training Status

In Progress  less than 20 seconds ago

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

Base model	LightGBM Regression
Model task	regression
Training job name	smjs-c-lgb-regression-model-20221122-034557
Training job arn	arn:aws:sagemaker:us-east-1:619770596115:training-job/smjs-c-lgb-regression-model-20221122-034557
Training time	~18 seconds
Output path	not available

> Instance Settings

> Hyper Parameters

> Security Settings

Stop Training

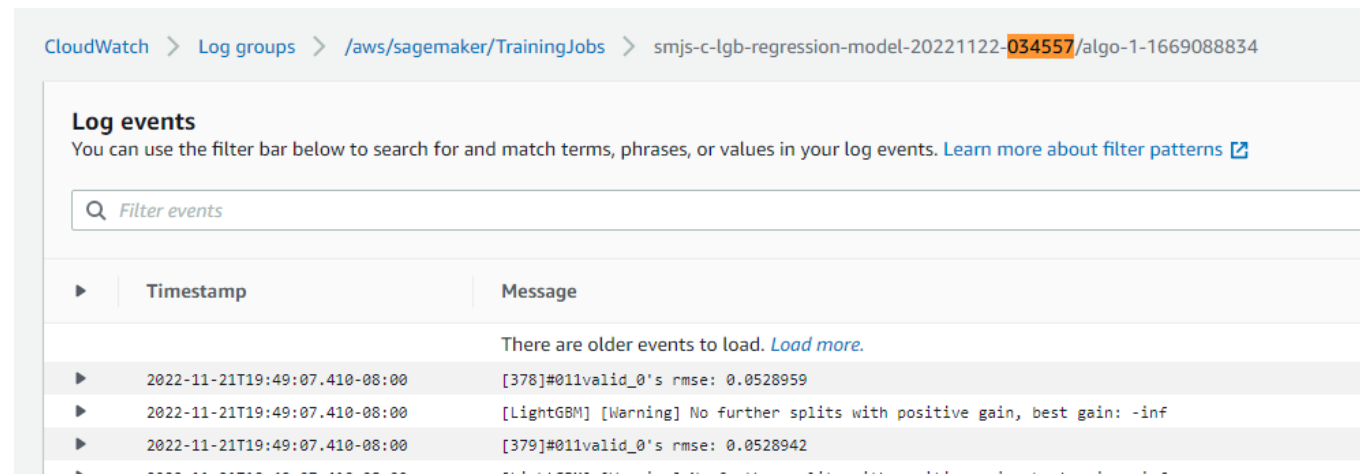
You can also go to SM console and see that the training has started:

# Evaluate the metrics of the model

- Grab the training job name:

smjs-c-lgb-regression-model-20221122-034557

- Search in CloudWatch, the log group and in the **TrainingJob** folder search for the training job that ends with the number of your job name:



# Compare training and validation RMSEs

CloudWatch > Log groups > /aws/sagemaker/TrainingJobs > smjs-c-lgb-regression-model-20221122-034557/algo-1-1669088834

## Log events

You can use the filter bar below to search for and match terms, phrases, or values in your log events. [Learn more about filter patterns](#)

	Timestamp	Message
		There are older events to load. <a href="#">Load more.</a>
▶	2022-11-21T19:49:07.410-08:00	[378]#011valid_0's rmse: 0.0528959
▶	2022-11-21T19:49:07.410-08:00	[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
▶	2022-11-21T19:49:07.410-08:00	[379]#011valid_0's rmse: 0.0528942
▶	2022-11-21T19:49:07.410-08:00	[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
▶	2022-11-21T19:49:07.410-08:00	[380]#011valid_0's rmse: 0.0528843
▶	2022-11-21T19:49:07.410-08:00	[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
▶	2022-11-21T19:49:07.410-08:00	[381]#011valid_0's rmse: 0.0528825
▶	2022-11-21T19:49:07.410-08:00	[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
▶	2022-11-21T19:49:07.410-08:00	[382]#011valid_0's rmse: 0.0528804
▶	2022-11-21T19:49:07.410-08:00	[LightGBM] [Warning] No further splits with positive gain, best gain: -inf

# After completion...

- You also should know where the model is: in the path in front of output path

**Complete** ↻ 2 minutes ago

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.

Base model	LightGBM Regression
Model task	regression
Training job name	smjs-c-lgb-regression-model-20221124-000359
Training job arn	arn:aws:sagemaker:us-east-1:619770596115:training-job/smjs-c-lgb-regression-model-20221124-000359
Training time	~3 minutes
Output path	s3://sagemaker-us-east-1-619770596115/smjs-c-lgb-regression-model-20221124-000359/output/model.tar.gz

# Where is the model

- The model itself is in the S3 bucket:

Amazon S3 > Buckets > sagemaker-us-east-1-619770596115 > smjs-c-lgb-regression-model-20221122-034557/ > output/









## output/


**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.

  Copy S3 URI  Copy URL  Download  Open  Delete **Actions** ▼  Create folder  Upload

 Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size
<input type="checkbox"/>	 <a href="#">model.tar.gz</a>	gz	November 21, 2022, 19:49:21 (UTC-08:00)	

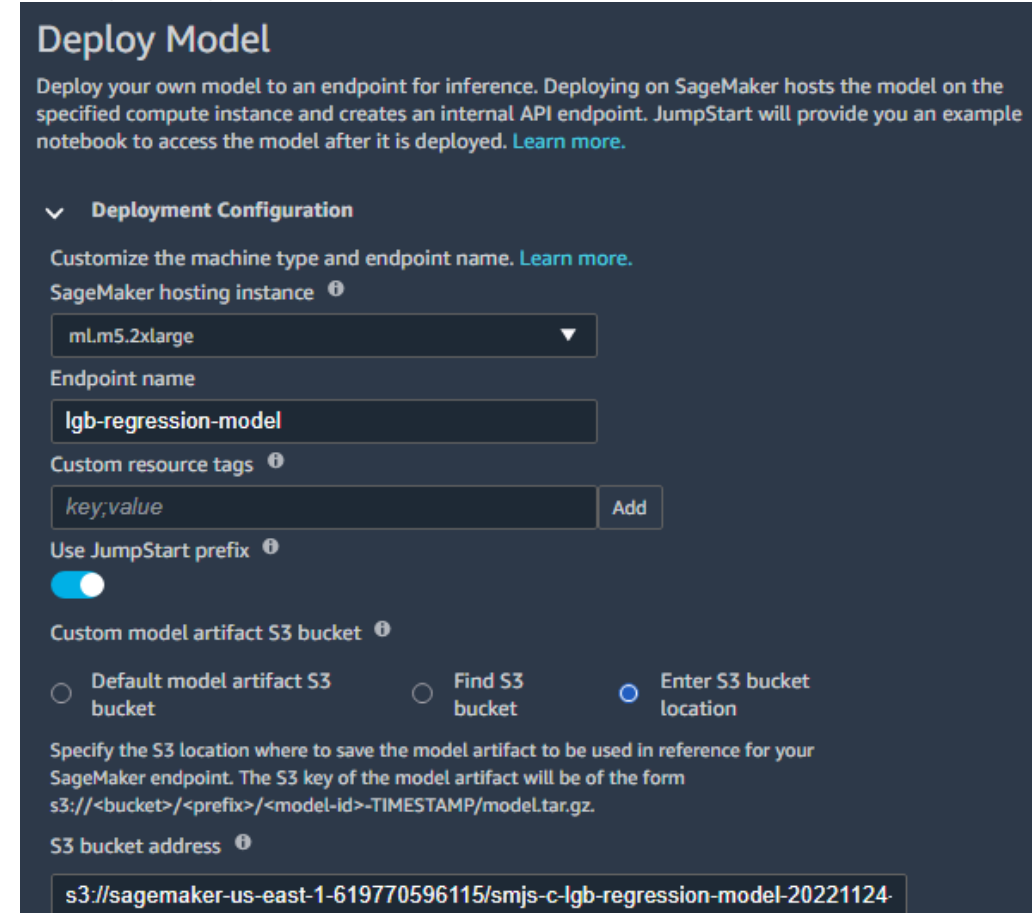
# Deploy the model

- Now that we have the model, we use JS to deploy the model

- Set the instance size and S3 bucket in

## Deployment Configuration

- Set the S3 path by selecting  
“Enter S3 Bucket location”  
and paste the **model.tar.gz** name  
(object **S3 URI**)



**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 ⓘ

ml.m5.2xlarge ▼

Endpoint name

lgb-regression-model

Custom resource tags ⓘ

key;value Add

Use JumpStart prefix ⓘ

☒ ☐ ☐

Custom model artifact S3 bucket ⓘ

☐ Default model artifact S3 bucket ☐ Find S3 bucket ☒ Enter S3 bucket location

Specify the S3 location where to save the model artifact to be used in reference for your SageMaker endpoint. The S3 key of the model artifact will be of the form s3://<bucket>/<prefix>/<model-id>-TIMESTAMP/model.tar.gz.

S3 bucket address ⓘ

s3://sagemaker-us-east-1-619770596115/smjs-c-lgb-regression-model-20221124-



# Security settings

- Set the security setting as shown below:

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 the IAM role you type below.

**Execution role arn** ⓘ

`arn:aws:iam::239630988601:role/fast-ai-academic-`


# Deploying the model

- Click on Deploy and wait until it deploys the endpoint

MODEL ENDPOINT

jumpstart-ffc-lgb-regression-model [Browse JumpStart](#)

### Endpoint Status


Creating  less than 10 seconds ago  
Your endpoint is being created. This may take five to ten minutes.

Base model	LightGBM Regression
Model task	regression
Endpoint source	not available
Endpoint arn	arn:aws:sagemaker:ca-central-1-239630988601:endpoint/jumpstart-ffc-lgb-regression-model
Endpoint name	jumpstart-ffc-lgb-regression-model
Instance	mLm5.2xlarge
Number of instances	1
Model data location	s3://sagemaker-ca-central-1-239630988601/smjs-c-lgb-regression-model-20230312-202548/output/model.tar.gz/lightgbm-regression-model-20230312-204358/model.tar.gz

> Security settings

jumpstart-ffc-lgb-regression-model [Browse JumpStart](#)

### Endpoint Status

In Service  2 minutes ago  
This endpoint is operational and ready to respond to inference requests.

Base model	LightGBM Regression
Model task	regression
Endpoint source	not available
Endpoint arn	arn:aws:sagemaker:ca-central-1-239630988601:endpoint/jumpstart-ffc-lgb-regression-model
Endpoint name	jumpstart-ffc-lgb-regression-model
Instance	mLm5.2xlarge
Number of instances	1
Model data location	s3://sagemaker-ca-central-1-239630988601/smjs-c-lgb-regression-model-20230312-202548/output/model.tar.gz/lightgbm-regression-model-20230312-204358/model.tar.gz

> Security settings

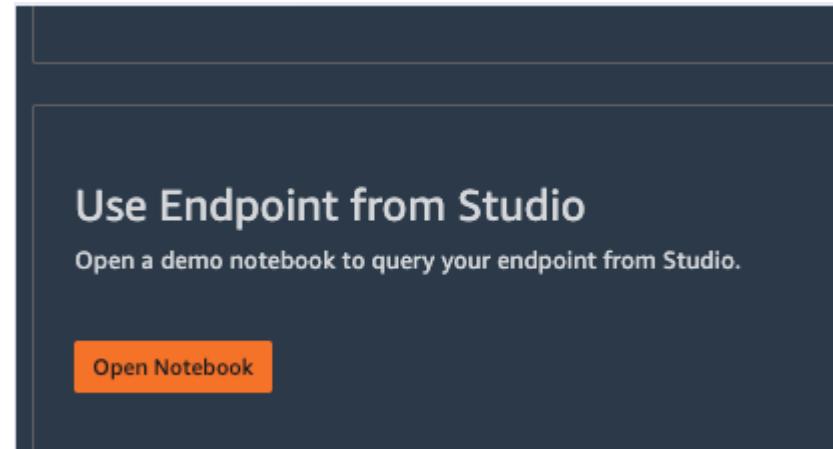
### Use Endpoint from Studio

Open a demo notebook to query your endpoint from Studio.

[Open Notebook](#)

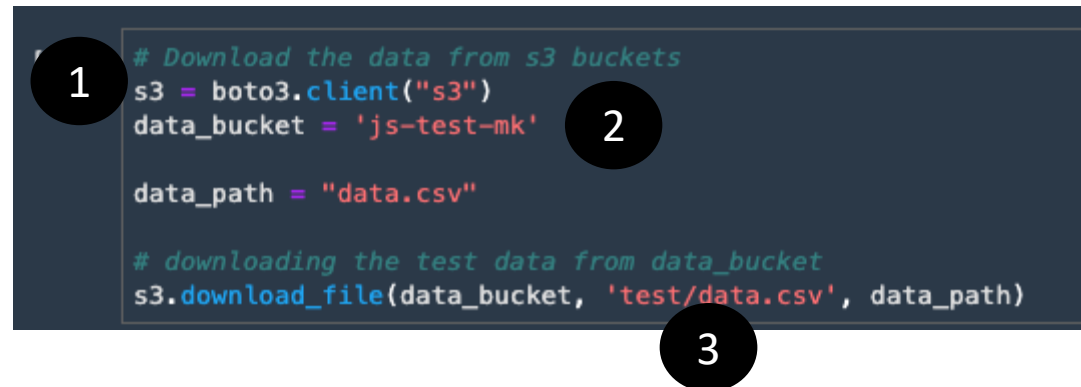
# Prepare for inferencing

- Upload the test data to the same bucket but a new folder called **test**
- Make sure the object name is **data.csv**
- Click on **Open Notebook**, in the **Use Endpoint from Studio**



# After deploying, test the endpoint

- In the notebook, change the bucket to where the test data is located on
- Make sure the object name is **data.csv** in the data bucket, and apply the changes as shown in the picture
- You have to change 3 spots in the cell
- Make sure you do not put extra / or anything else that is not needed



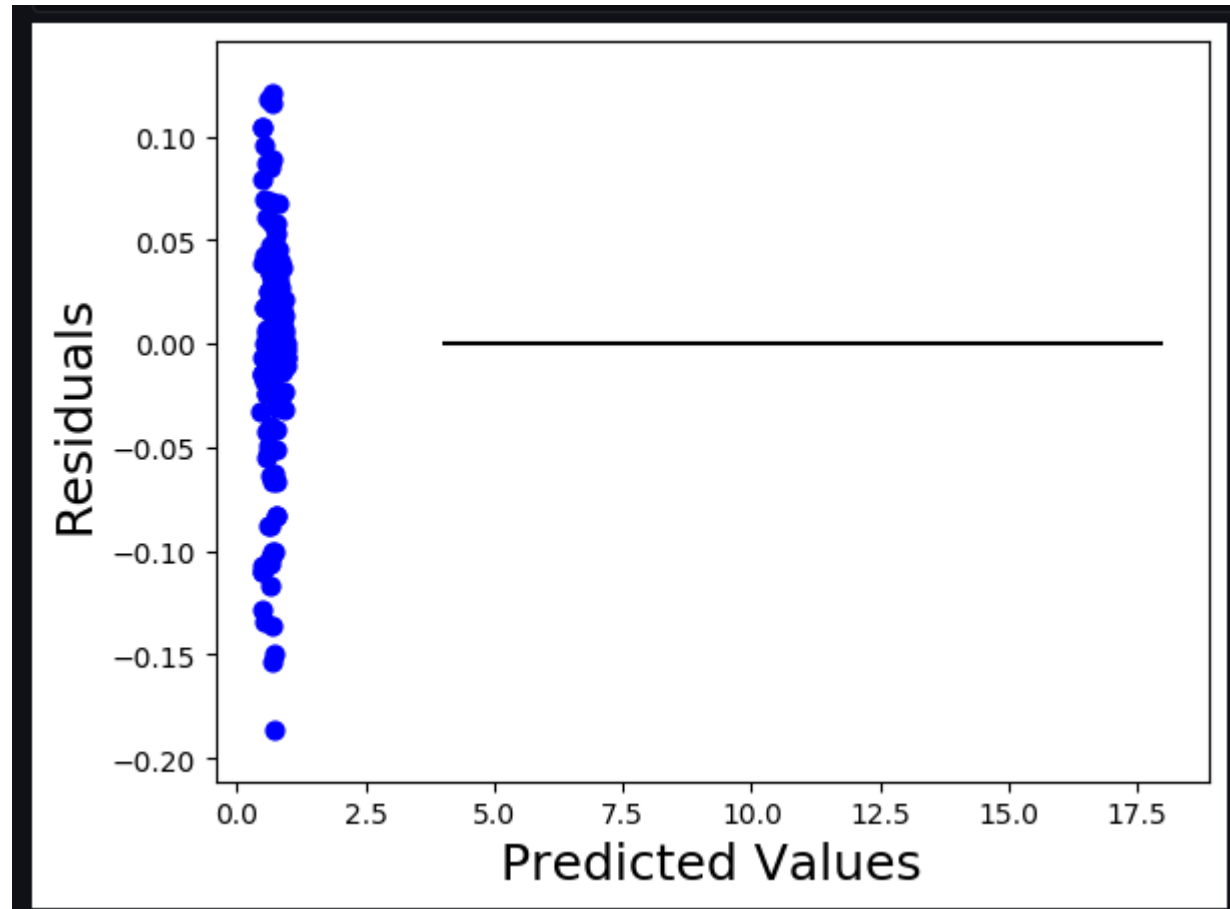
```
# Download the data from s3 buckets
s3 = boto3.client("s3")
data_bucket = 'js-test-mk'

data_path = "data.csv"

# downloading the test data from data_bucket
s3.download_file(data_bucket, 'test/data.csv', data_path)
```

# You see now the metric against the test data

- As you see the residual of the resulting model was pretty good



# Delete the endpoint

- Before deleting the endpoint, read the assignment page in the next slide, capture everything you need to include in the report and then delete the endpoint
- After finishing your assignment, make sure you delete the endpoint
- Save the completed notebook for the assignment into your local computer since you need to explain it in the presentation date

# Assignment

- The training job you used , we started based on a pre-trained model with Abalone data
- Did that really make a difference? What if you use the data I gave you and start a new training from scratch and use LighGBM algorithm to train a new mode?  
(<https://docs.aws.amazon.com/sagemaker/latest/dg/lightgbm.html>)
- You need to write code (not console) in the notebook to do this assignment
- Use the same hyperparameters you used during the class in JS to train from scratch and compare at least the following items:
  - Training time
  - Quality of model against validation data
  - Residual values
- Write a report to explain the steps you took and what you learned out of this activity
- Submit the completed notebook and report to blackboard. The notebook should have all the cells run already and I want to see the result of each cell
- You just open the file in the presentation day and explain what happens in each cell. You do not run those cells in the presentation day. We just talk about results and reports.