



# Sagemaker

## Session-5



# Table of Contents

- ▶ Cloud Computing Market Size
- ▶ What is SageMaker?
- ▶ SageMaker Process
- ▶ Algorithms
- ▶ Notebook Instance
- ▶ S3
- ▶ Billing

# Table of Contents

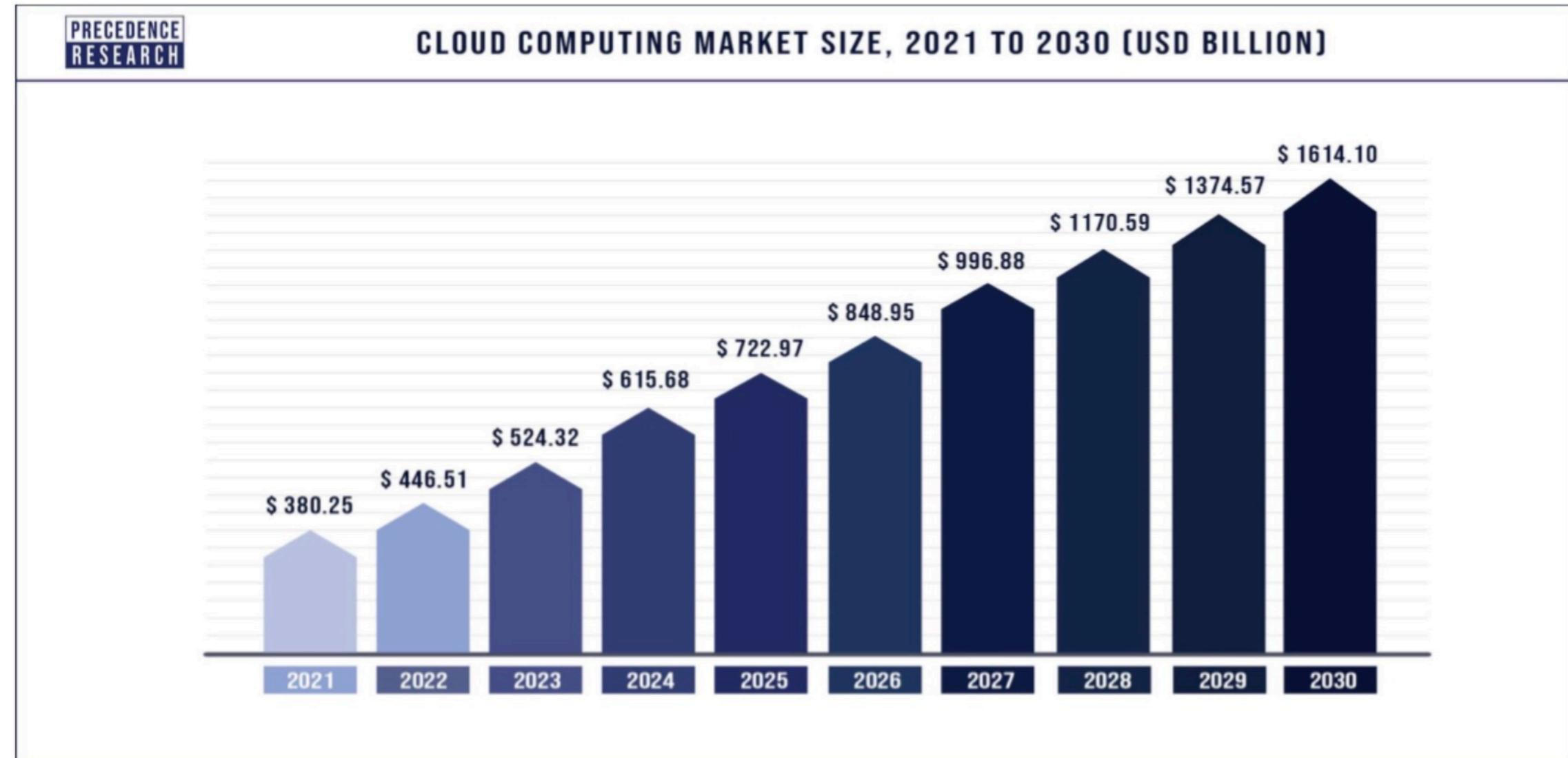
- ▶ Cloud Computing Market Size
- ▶ What is SageMaker?
- ▶ SageMaker Process
- ▶ Algorithms
- ▶ Notebook Instance
- ▶ S3
- ▶ Billing



# SageMaker

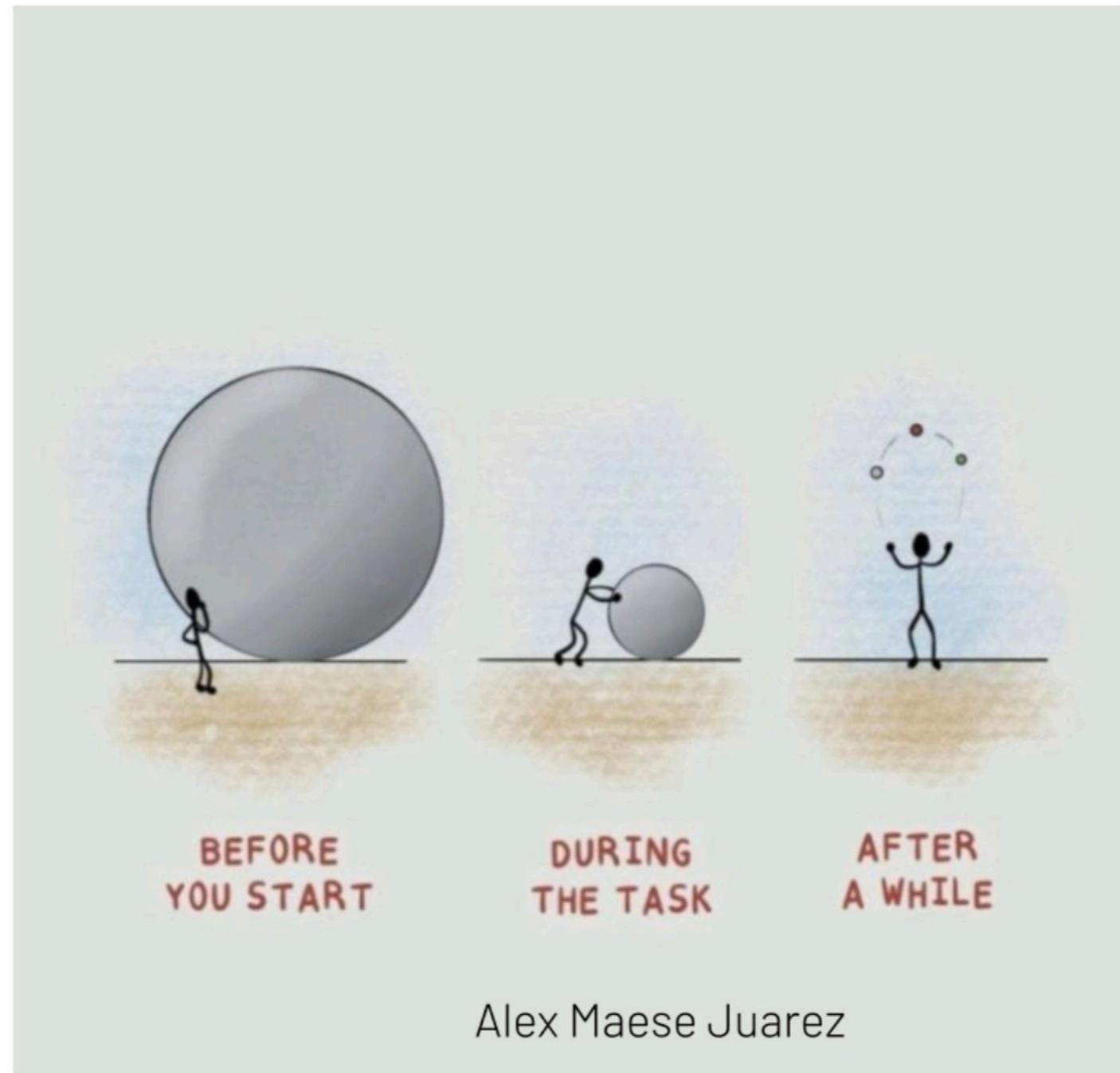


# Cloud Computing Market Size



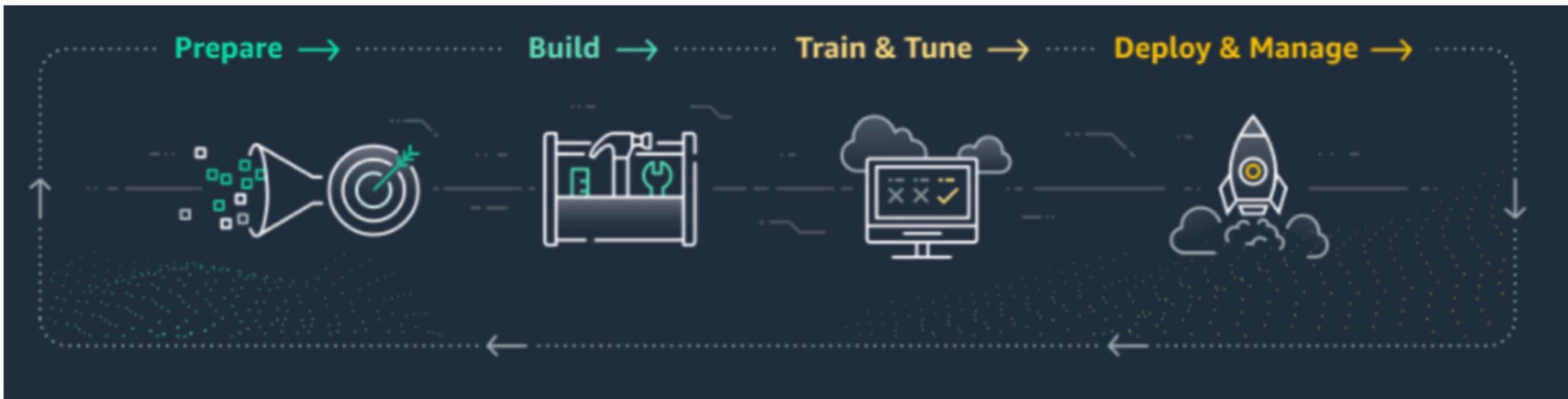
- ▶ The rising popularity of the latest novel technologies like **artificial intelligence** and **machine learning** and its rapid adoption in the cloud computing is empowering the growth of the global cloud computing market.

# SageMaker



# What is SageMaker

Amazon **SageMaker** is a **platform** service that simplifies the process of **building, training, and deploying ML models** by providing everything organizations need to connect to their training data, select and optimize the **best algorithm and framework**, and **deploy their model** on auto-scaling clusters of Amazon **EC2**.



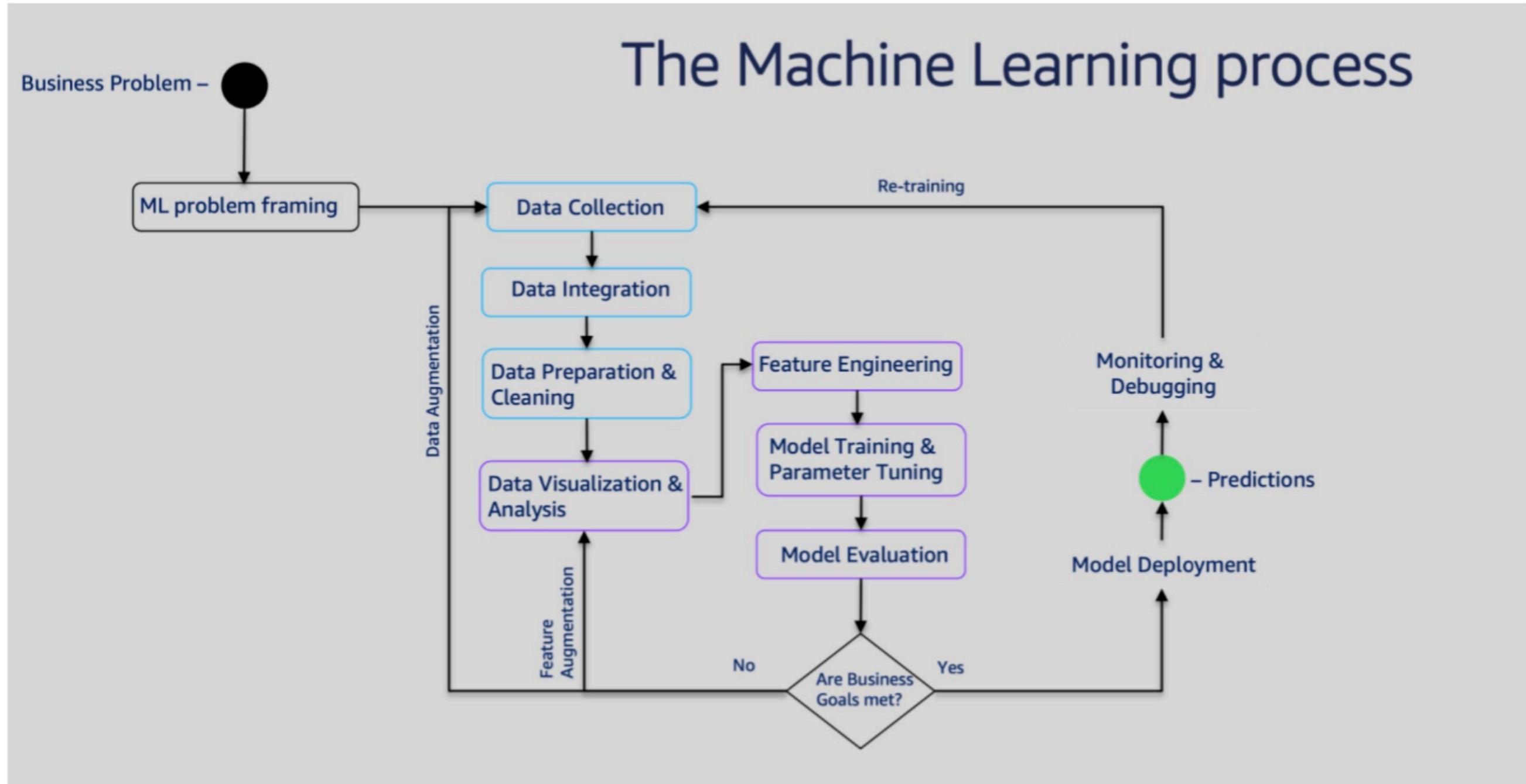
# WHY Amazon SageMaker?

- ▶ Accelerating machine learning innovation through security
- ▶ Security features from Amazon SageMaker and the AWS Cloud can help organizations go from idea to production faster
- ▶ <https://amer.resources.awscloud.com/ai-ml/accelerating-machine-learning-innovation-through-security>

# ML with AWS, by the numbers

- ▶ Reduce training time by 50%
- ▶ Provide 90% scaling efficiency
- ▶ Deliver 3x faster network throughput
- ▶ Improve price and performance by 25%
- ▶ 91% of cloud-based PyTorch runs on AWS
- ▶ 92% of cloud-based TensorFlow runs on AWS

# ML process



# SageMaker Free Tier (2 months)

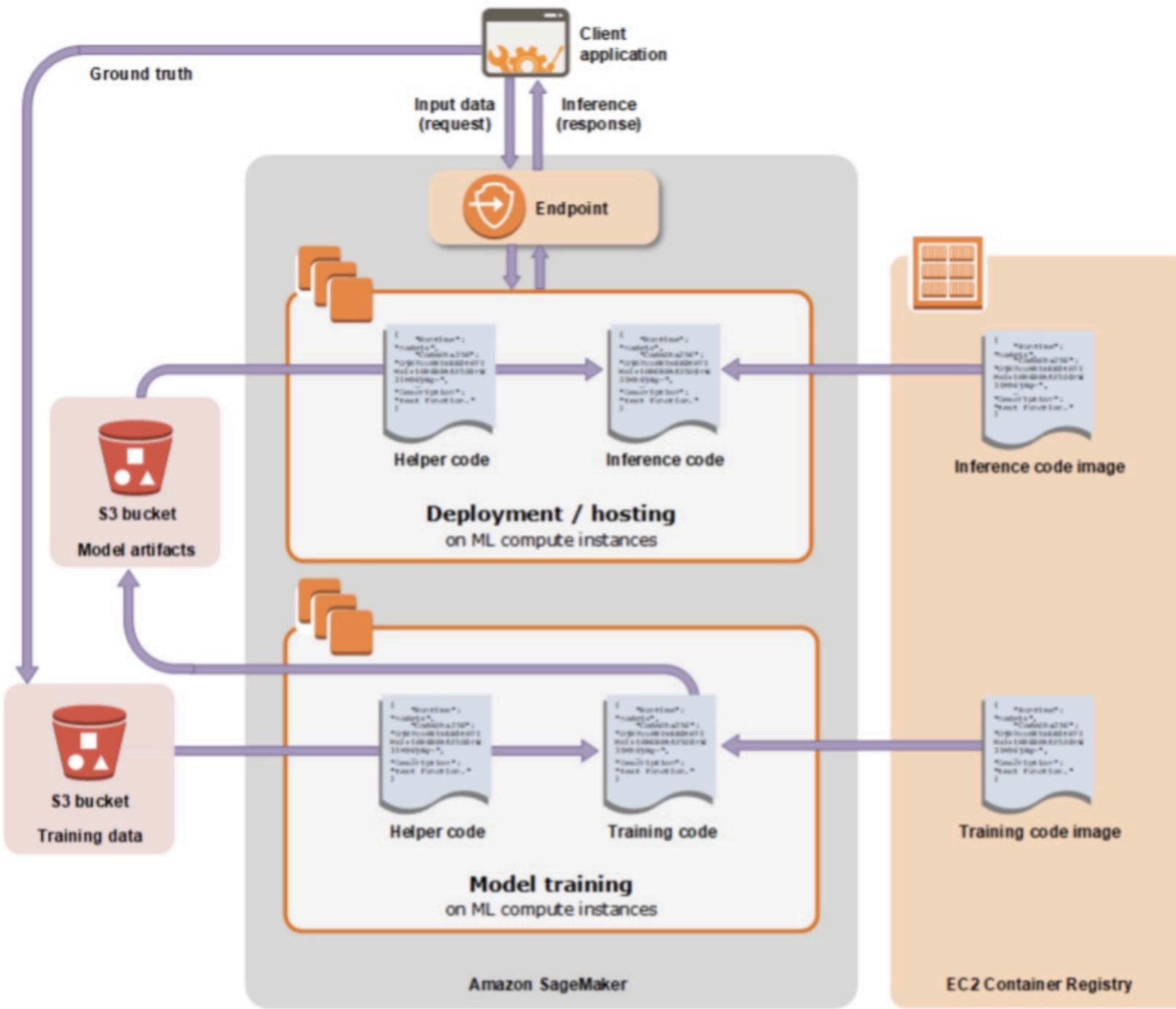
Two months free tier – starts from the first month you create a SageMaker resource

Development – 250 Hours/Month t2.medium or t3.medium

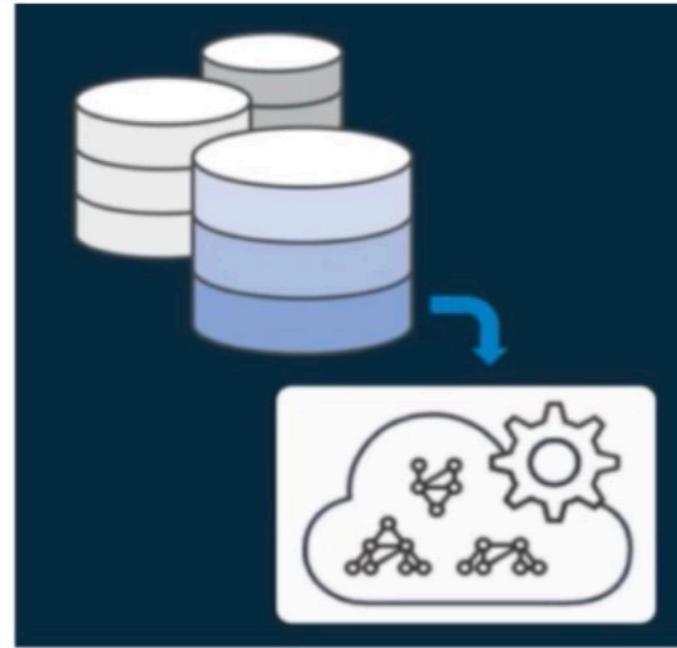
Train – 50 Hours/Month m4.xlarge or m5.xlarge

Deploy – 125 Hours/Month m4.xlarge or m5.xlarge

# Model Deployment Architecture



# SageMaker Process

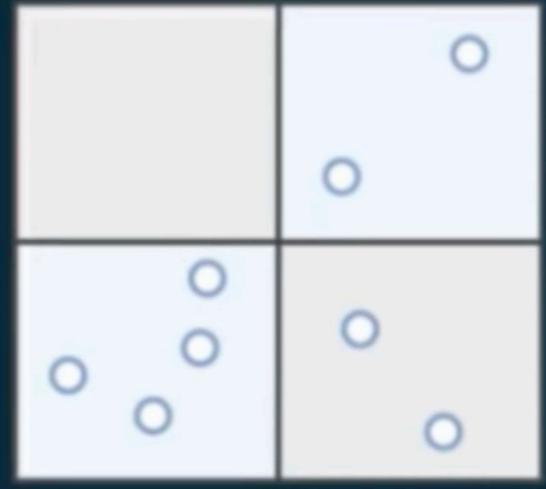


- Get your data
- Explore and refine models in a single Notebook Instance
- Train on the full dataset in a cluster of GPU instances..
- Deploy to production

# Built-in algorithms



XGBoost, FM,  
Linear, k-NN, and  
Forecasting for  
supervised  
learning



k-Means, PCA, and  
Random Cut  
Forest for  
unsupervised  
learning

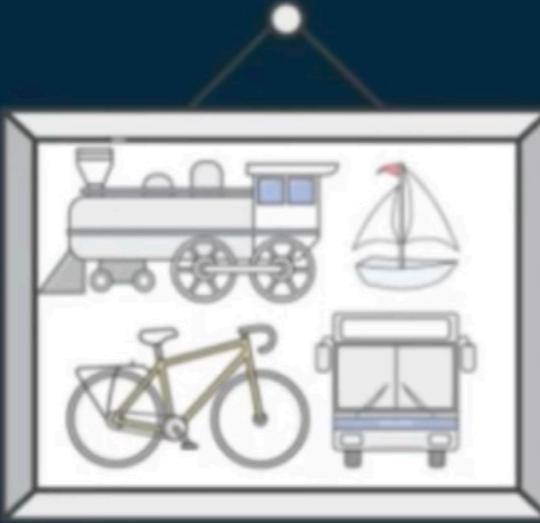


Image  
classification and  
object detection  
for computer  
vision



LDA, Neural Topic  
Model, Seq2seq,  
and Word2Vec for  
text and NLP

# Built-in algorithms-Supervised

- ▶ Linear Learner: regression, classification
- ▶ K-Nearest Neighbors: non-parametric regression and classification
- ▶ XGBoost: regression, classification, ranking
- ▶ Factorization Machines: regression, classification, recommendation
- ▶ Semantic Segmentation: Deep Learning
- ▶ Image Classification: Deep Learning (ResNet)
- ▶ Object Detection (SSD): Deep Learning
- ▶ (VGG or ResNet)
- ▶ Sequence to Sequence: machine translation, speech to text and more
- ▶ DeepAR: time-series forecasting (RNN)

# Built-in algorithms- Unsupervised

- ▶ K-Means: clustering
- ▶ Principal Component Analysis: dimensionality reduction
- ▶ Random Cut Forest: anomaly detection
- ▶ Object2Vec: general-purpose embedding
- ▶ Neural Topic Model: topic modeling
- ▶ Latent Dirichlet Allocation: topic modeling (mostly)
- ▶ Blazing Text: GPU-based Word2 Vec, and text classification
- ▶ IP Insights: usage patterns for IP addresses

# Bring your own algorithm



Pick your  
preferred  
framework...



... add algorithm  
code to a Docker  
container...



... publish to ECR

# S3 BUCKET and OBJECTS



- ▶ iden-ml-sagemaker/
  - bikerental/
    - train/
      - train.csv
    - validation/
      - validation.csv
    - test/
      - test.csv
    - model/
      - xgboost-bikerental-v1-2022-04-21-10-23-10-964/ (Job\_name+datetime)
        - output
          - model.tar.gz

Amazon S3>Buckets>iden-ml-sagemaker>bikerental/output/xgboost-bikerental-v1-2022-04-21-10-23-10-964/output/model.tar.gz

# Notebook Instance

An *Amazon SageMaker notebook instance* is a machine learning (ML) compute instance running the Jupyter Notebook App. SageMaker manages creating the instance and related resources.

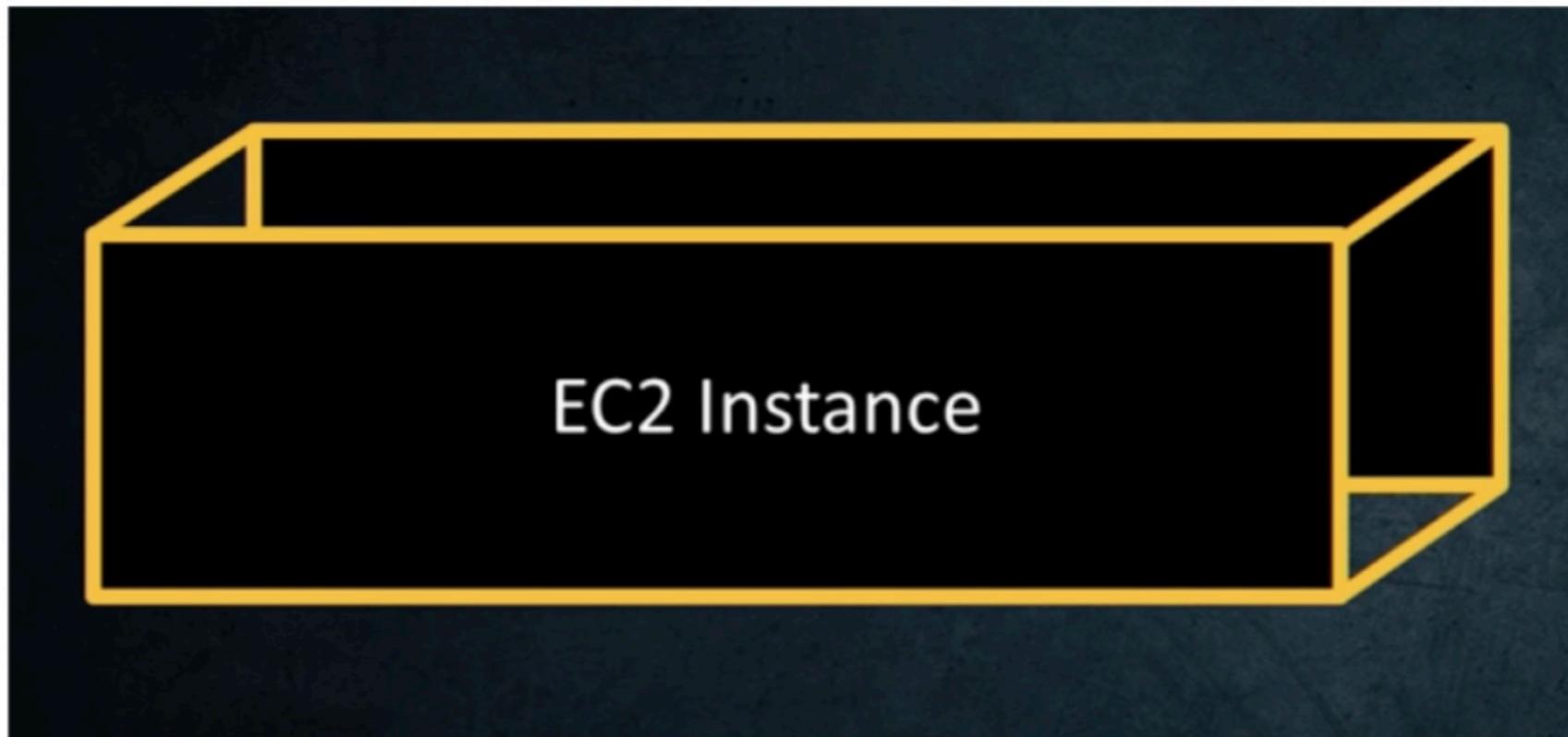


# Notebook Instance

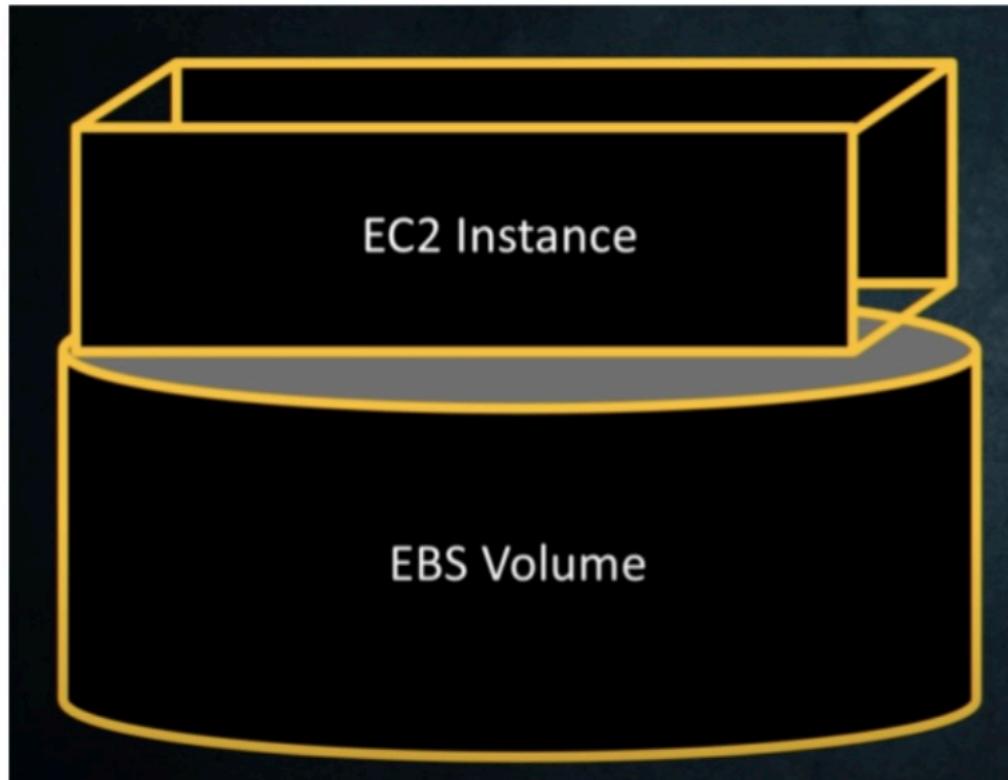


Use Jupyter notebooks in your notebook instance

- To prepare and process data,
- Write code to train models,
- Deploy models to SageMaker hosting,
- Test or validate your models.



# Notebook Instance



- Anaconda Packages,
- Tensorflow and Apache MXnet,
- Storage volume,
- Sample notebooks that contain complete code walkthroughs



You are using CSV formatted files to train on SageMaker's built-in XGBoost algorithm. SageMaker expects your training and validation to follow this convention:

- ▶ CSV must have column headers and target variable must be the last column
- ▶ CSV must have column headers with the target variable in the first column
- ▶ CSV must not have a column header record. Target variable must be the last column
- ▶ CSV must not have a column header record. Target variable must be the first column

# How does SageMaker built-in know the target variable?

- ▶ For CSV training, the algorithm assumes that the target variable is in the first column and that the CSV does not have a header record.(Train,Validation)
- ▶ For CSV inference, the algorithm assumes that CSV input does not have the label column.

<https://docs.aws.amazon.com/sagemaker/latest/dg/xgboost.html>

# SageMaker Billing Dashboard



Cost Management	▶ EC2 Container Registry (ECR)	\$0.00
Cost Explorer	▶ Elastic Compute Cloud	\$0.00
Budgets	▶ Key Management Service	\$0.00
Budgets Reports	▶ SageMaker	\$0.30
Savings Plans 	▼ US East (N. Virginia)	\$0.30
Preferences	Amazon SageMaker CreateVolume-Gp2	\$0.00
Billing preferences	\$0.00 for SageMaker Debugger Built-in Rule Volume \$0.14 per GB-Mo of Endpoint ML storage \$0.14 per GB-Mo of Notebook Instance ML storage \$0.14 per GB-Mo of Training Job ML storage	0.026 GB-Mo 0.002 GB-Mo 0.034 GB-Mo 0.000278 GB-Mo
Payment methods		
Consolidated billing 	Amazon SageMaker Invoke-Endpoint	\$0.00
Tax settings	\$0.016 per GB for Endpoint Data IN \$0.016 per GB for Endpoint Data OUT	0.001 GB 0.000230 GB
	Amazon SageMaker RunInstance	\$0.30
	\$0.0 for SageMaker Debugger Built-in Rule Instance \$0.00 for Host:ml.m4.xlarge per hour under monthly free tier \$0.00 for Notebk:ml.t2.medium per hour under monthly free tier \$0.00 for Notebk:ml.t3.medium per hour under monthly free tier \$0.23 per Notebook ml.m5.xlarge hour in US East (N. Virginia) \$0.478 per Training ml.c4.2xlarge hour in US East (N. Virginia) Managed Spot Training Job ml.m5.xlarge hour in US East (N. Virginia)	0.005 Hrs 0.169 Hrs 2.695 Hrs 1.153 Hrs 1.258 Hrs 0.016 Hrs 0.003 Hrs
	▶ US East (Ohio)	\$0.00
	▶ Secrets Manager	\$0.00
	▶ Simple Notification Service	\$0.00
	▶ Simple Storage Service	\$0.00

# SageMaker Use Case Example

## ZAPPOS

*"We are...using analytics and machine learning solutions to personalize sizing and search results for individual users. AWS services (including Amazon SageMaker) allow (our) engineers to focus on improving performance and results rather than DevOps overhead."*

# SageMaker Studio



The screenshot shows the Amazon SageMaker Studio interface with the title bar "model deployment < > d-1nmnhjbogdti.studio.us-east-1.sagemaker.aws". The top navigation bar includes File, Edit, View, Run, Kernel, Git, Tabs, Settings, and Help. The main area is titled "Launcher" and displays several sections:

- Get started**:
  - JumpStart models, algorithms, and solutions**: Includes "SageMaker JumpStart" and "Solution: Detect malicious users and transactions" (with a link), "Solution: Demand forecasting" (with a link), and a "Go to SageMaker JumpStart" button.
  - Build models automatically**: Includes "SageMaker Autopilot" and links to "Video: Get started with Autopilot" and "Blog: Getting started with Autopilot". It also features a "New autopilot experiment" button.
  - Instantly prepare data for ML**: Includes "SageMaker Data Wrangler" and links to "Blog: Getting started with Data Wrangler" and "Blog: Predicting credit risk". It features a "Start now" button and a small icon of a brain with data lines.
- ML tasks and components**:
  - "New compilation job": Create a new compilation job, View compilation jobs.
  - "New project": Organize ML components and automate MLOps with built-in or custom project templates, View projects.
  - "New feature group": Create a new feature group in the feature store to logically group and manage features, View feature store.
  - "New Autopilot experiment": Create prediction models from your data and start making predictions in a few clicks, View current experiments.
  - "New data flow": Prepare and visualize your data with SageMaker Data Wrangler, View data flows.
- Notebooks and compute resources**:
  - Select a SageMaker image: Data Science. Options include "Notebook" (Python 3) with a "+" button, "Console" (Python 3) with a "+" button, and "Image Terminal" (Image Terminal) with a "+" button.
  - Select a start-up script: No Script. Options include "Show Contextual Help" with a "+" button, "System terminal" with a "+" button, and "Text File" with a "+" button.
- Utilities and files**:
  - Options include "Markdown File" with a "+" button.

# Hyperparameters

The screenshot shows the 'Hyperparameters' configuration page in the Amazon SageMaker console. At the top, there's a navigation bar with links to S3, Cloud9, IAM, Lambda, API Gateway, EC2, Elastic Container Registry, and Support. Below the navigation bar, the title 'Hyperparameters' is displayed. A descriptive text block states: 'You can use hyperparameters to finely control training. We've set default hyperparameters for the algorithm you've chosen. Learn more' with a blue 'Learn more' link. Below this, a table lists seven hyperparameters with their current values:

Key	Value
eta	0.1
eval_metric	auc
max_depth	3
num_round	20
objective	reg:logistic
scale_pos_weight	2.0
subsample	0.5

At the bottom of the page, there's a section titled 'Input data configuration'.

Feedback Looking for language selection? Find it in the new Unified Settings ↗ Privacy Terms