Build, train, and deploy ML models with Amazon SageMaker

Jonathan Dion
Senior Technical Evangelist
Amazon Web Services
@jotdion
linkedin.com/in/jotdion

Agenda

- The AWS ML Stack
- ML Services overview
- Labs

- What we'll cover today:
  - Preparing our data set
  - Training and deploying with built-in algorithms
  - Finding optimal hyperparameters with automatic model tuning
  - Deploying multiple models for A/B testing

aws

Our mission at AWS

Put machine learning in the hands of every developer

aws

MACHINE LEARNING IS HAPPENING IN COMPANIES OF EVERY SIZE AND INDUSTRY

Tens of thousands customers have chosen AWS for their ML workloads | More than twice as many customers using ML than any other cloud provider

aws

The AWS ML Stack

Broadest and most complete set of Machine Learning capabilities

AI SERVICES

VISION
- Amazon Rekognition

SPEECH
- Amazon Polly
- Amazon Transcribe *+Medical*

TEXT
- Amazon Comprehend *+Medical*
- Amazon Translate
- Amazon Textract

SEARCH
- Amazon Kendra

CHATBOTS
- Amazon Lex

PERSONALIZATION
- Amazon Personalize

FORECASTING
- Amazon Forecast

FRAUD
- Amazon Fraud Detector

DEVELOPMENT
- Amazon CodeGuru

CONTACT CENTERS
- Contact Lens *For Amazon Connect*

ML SERVICES

Amazon SageMaker

Ground Truth

ML Marketplace

SageMaker Studio IDE

Built-in algorithms | Notebooks | Experiments | Processing & Model Evaluation | Model training & tuning | Debugger | Autopilot | Model hosting | Model Monitor

Neo

Augmented AI

ML FRAMEWORKS & INFRASTRUCTURE

TensorFlow | mxnet | PYTORCH | GLUON | Keras | learn | HOROVOD | DeepGraphLibrary

Deep Learning AMIs & Containers | GPUs & CPUs | Elastic Inference | Inferentia | FPGA

aws

The AWS ML Stack

Broadest and most complete set of Machine Learning capabilities

AI SERVICES

VISION
- Amazon Rekognition

SPEECH
- Amazon Polly
- Amazon Transcribe *+Medical*

TEXT
- Amazon Comprehend *+Medical*
- Amazon Translate
- Amazon Textract

SEARCH
- Amazon Kendra

CHATBOTS
- Amazon Lex

PERSONALIZATION
- Amazon Personalize

FORECASTING
- Amazon Forecast

FRAUD
- Amazon Fraud Detector

DEVELOPMENT
- Amazon CodeGuru

CONTACT CENTERS
- Contact Lens *For Amazon Connect*

ML SERVICES

Amazon SageMaker

Ground Truth

ML Marketplace

SageMaker Studio IDE

Built-in algorithms | Notebooks | Experiments | Processing & Model Evaluation | Model training & tuning | Debugger | Autopilot | Model hosting | Model Monitor

Neo

Augmented AI

ML FRAMEWORKS & INFRASTRUCTURE

TensorFlow | mxnet | GLUON | Keras | PYTORCH | DeepGraphLibrary

Deep Learning AMIs & Containers | GPUs & CPUs | Elastic Inference | Inferentia | FPGA

aws

The AWS ML Stack

Broadest and most complete set of Machine Learning capabilities

AI SERVICES

VISION	SPEECH		TEXT			SEARCH	CHATBOTS	PERSONALIZATION	FORECASTING	FRAUD	DEVELOPMENT	CONTACT CENTERS
Amazon Rekognition	Amazon Polly	Amazon Transcribe *+Medical*	Amazon Comprehend *+Medical*	Amazon Translate	Amazon Textract	Amazon Kendra	Amazon Lex	Amazon Personalize	Amazon Forecast	Amazon Fraud Detector	Amazon CodeGuru	Contact Lens *For Amazon Connect*

ML SERVICES

Amazon SageMaker	Ground Truth	ML Marketplace	SageMaker Studio IDE								Neo	Augmented AI
			Built-in algorithms	Notebooks	Experiments	Processing & Model Evaluation	Model training & tuning	Debugger	Autopilot	Model hosting	Model Monitor	

ML FRAMEWORKS & INFRASTRUCTURE

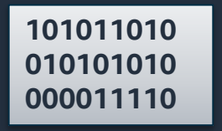
TensorFlow · mxnet · GLUON · Keras · PYTORCH · DeepGraphLibrary

Deep Learning AMIs & Containers | GPUs & CPUs | Elastic Inference | Inferentia | FPGA

aws

The machine learning workflow is iterative and complex

Prepare



Collect and prepare training data

Build

Choose or build an ML algorithm

Train & Tune

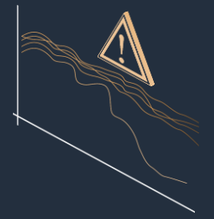
Set up and manage environments for training

Train, debug, and tune models

Manage training runs

Deploy & Manage

Deploy model in production



Monitor models

Validate predictions

Scale and manage the production environment

aws

Amazon SageMaker helps you build, train, and deploy models

Prepare

Build

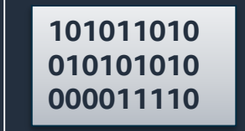
Train & Tune

Deploy & Manage

Web-based IDE for machine learning

Automatically build and train models

Fully managed data processing jobs and data labeling workflows



Collect and prepare training data

One-click collaborative notebooks and built-in, high performance algorithms and models

Choose or build an ML algorithm

One-click training

Set up and manage environments for training

Debugging and optimization

Train, debug, and tune models

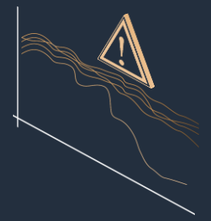
Visually track and compare experiments

Manage training runs

One-click deployment and autoscaling

Deploy model in production

Automatically spot concept drift



Monitor models

Add human review of predictions

Validate predictions

Fully managed with auto-scaling for 75% less

Scale and manage the production environment

aws

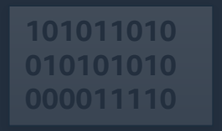
Amazon SageMaker helps you build, train, and deploy models

Prepare | **Build** | **Train & Tune** | **Deploy & Manage**

Web-based IDE for machine learning

Automatically build and train models

Fully managed data processing jobs and data labeling workflows



Collect and prepare training data

One-click collaborative notebooks and built-in, high performance algorithms and models

Choose or build an ML algorithm

One-click training

Set up and manage environments for training

Debugging and optimization

Train, debug, and tune models

Visually track and compare experiments

Manage training runs

One-click deployment and autoscaling

Deploy model in production

Automatically spot concept drift

Monitor models

Add human review of predictions

Validate predictions

Fully managed with auto-scaling for 75% less

Scale and manage the production environment

aws

Amazon SageMaker Studio

Fully integrated development environment (IDE) for machine learning

Collaboration at scale

Share notebooks without tracking code dependencies

Easy experiment management

Organize, track, and compare thousands of experiments

Automatic model generation

Get accurate models with full visibility & control without writing code

Higher quality ML models

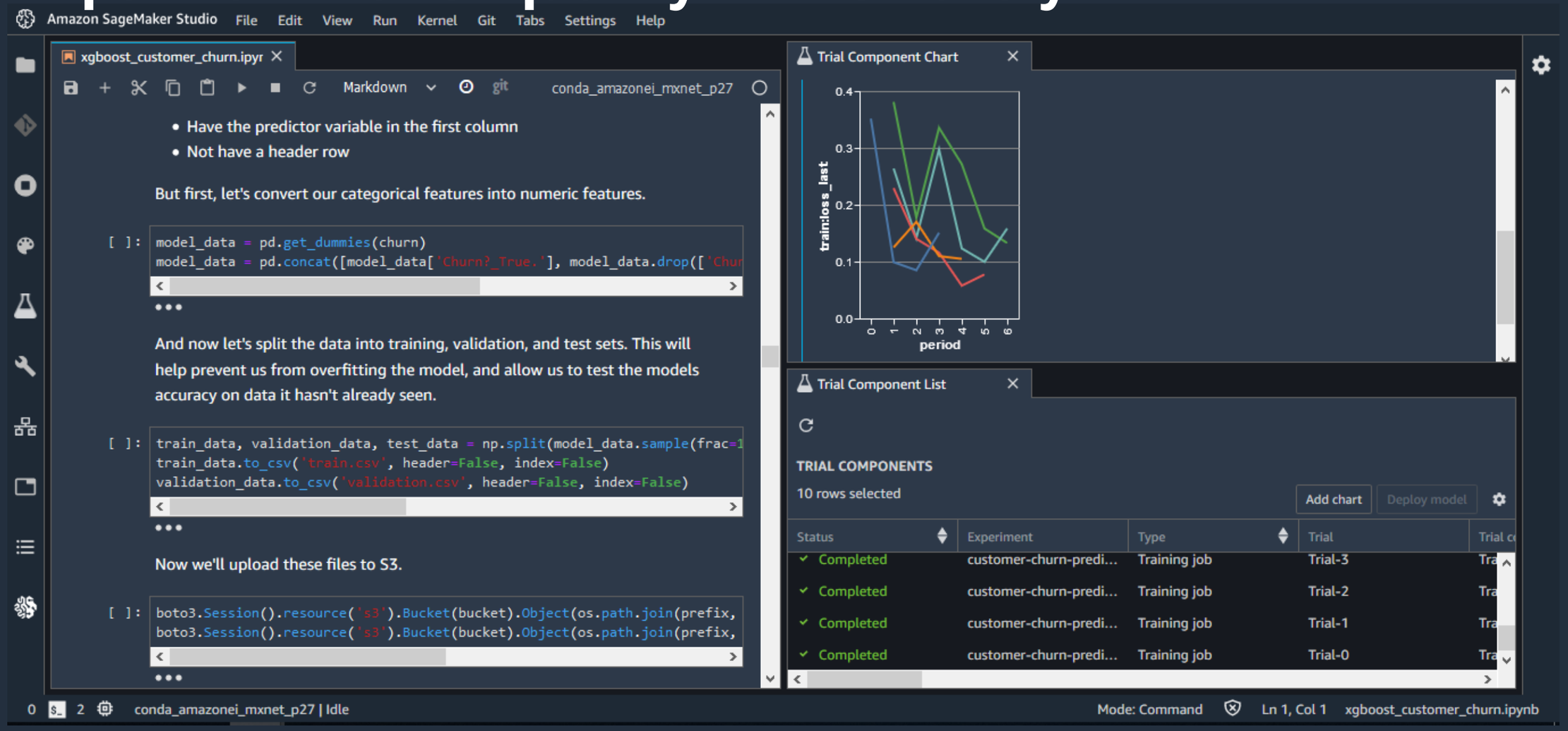
Automatically debug errors, monitor models, & maintain high quality

Increased productivity

Code, build, train, deploy, & monitor in a unified visual interface

aws

Use Amazon SageMaker Studio to update models and see impact on model quality immediately



Amazon SageMaker Autopilot

Automatic model creation with full visibility & control

Quick to start

Provide your data in a tabular form & specify target prediction

Automatic model creation

Get ML models with feature engineering & model tuning automatically done

Visibility & control

Get notebooks for your models with source code

Recommendations & Optimization

Get a leaderboard & continue to improve your model

aws

Amazon SageMaker helps you build, train, and deploy models

Prepare

Build

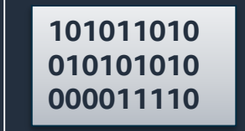
Train & Tune

Deploy & Manage

Web-based IDE for machine learning

Automatically build and train models

Fully managed data processing jobs and data labeling workflows



Collect and prepare training data

One-click collaborative notebooks and built-in, high performance algorithms and models

Choose or build an ML algorithm

One-click training

Set up and manage environments for training

Debugging and optimization

Train, debug, and tune models

Visually track and compare experiments

Manage training runs

One-click deployment and autoscaling

Deploy model in production

Automatically spot concept drift

Monitor models

Add human review of predictions

Validate predictions

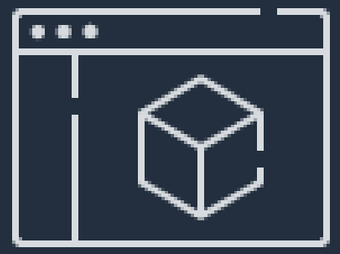
Fully managed with auto-scaling for 75% less

Scale and manage the production environment

aws

Amazon SageMaker Ground Truth

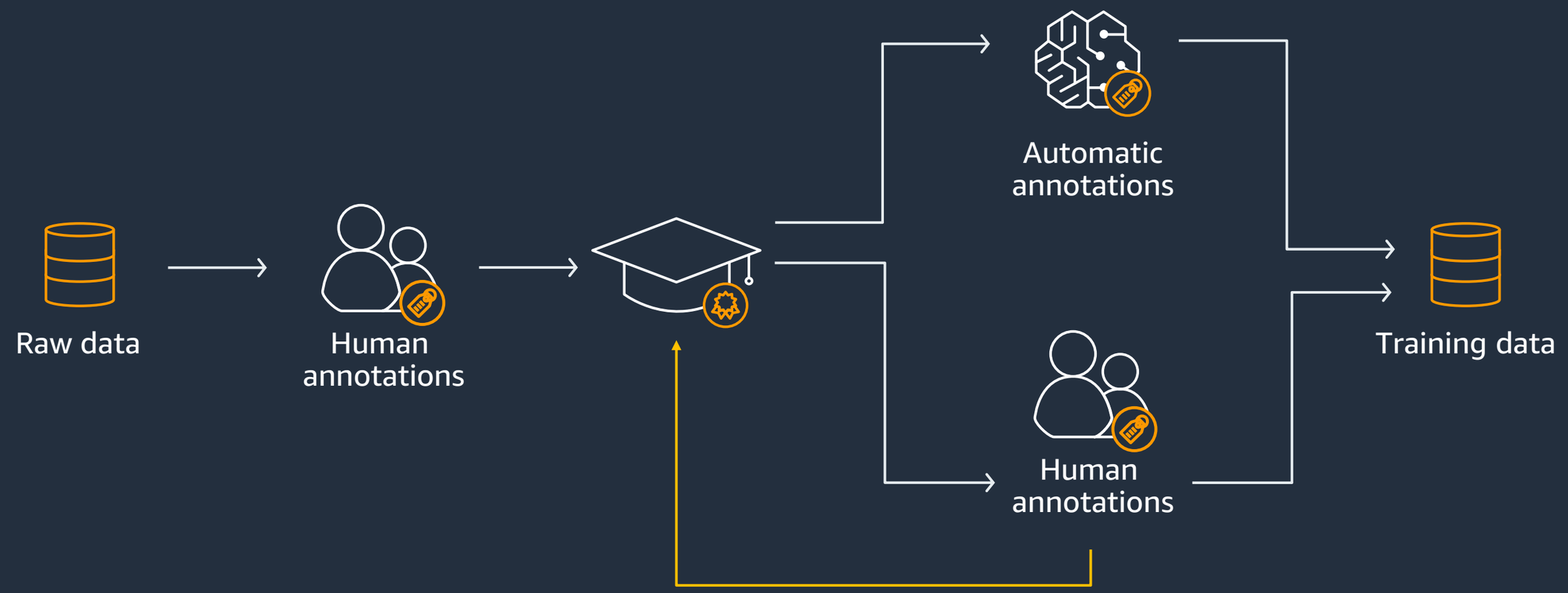
Build highly accurate training datasets using machine learning



- Reduce data labeling costs by up to 70%
- Access labelers through Amazon Mechanical Turk, Amazon approved vendors, or use private human labelers
- Achieve accurate results quickly

aws

How Amazon SageMaker Ground Truth Works



aws

Amazon SageMaker Processing

Analytics jobs for data processing and model evaluation

Fully managed

Achieve distributed processing for clusters

Custom processing

Bring your own script for feature engineering

Container support

Use SageMaker's built-in containers or bring your own

Security and compliance

Leverage SageMaker's security & compliance features

Automatic creation & termination

Your resources are created, configured, & terminated automatically

aws

Amazon SageMaker helps you build, train, and deploy models

Prepare

Build

Train & Tune

Deploy & Manage

Web-based IDE for machine learning

Automatically build and train models

Fully managed data processing jobs and data labeling workflows

101011010
010101010
000011110

Collect and prepare training data

One-click collaborative notebooks and built-in, high performance algorithms and models

Choose or build an ML algorithm

One-click training

Set up and manage environments for training

Debugging and optimization

Train, debug, and tune models

Visually track and compare experiments

Manage training runs

One-click deployment and autoscaling

Deploy model in production

Automatically spot concept drift

Monitor models

Add human review of predictions

Validate predictions

Fully managed with auto-scaling for 75% less

Scale and manage the production environment

aws

Amazon SageMaker Notebooks

Fast-start sharable notebooks (in preview)

Easy access with Single Sign-On (SSO)

Access your notebooks in seconds

Fully managed and secure

Administrators manage access and permissions

Fast setup

Start your notebooks without spinning up compute resources

Easy collaboration

Share notebooks with a single click

Flexible

Dial up or down compute resources (coming soon)

aws

Amazon SageMaker has built-in algorithms or bring your own

Classification
- Linear Learner
- XGBoost
- KNN

Working with Text
- BlazingText
- Supervised
- Unsupervised

Sequence Translation
- Seq2Seq

Computer Vision
- Image Classification
- Object Detection
- Semantic Segmentation

Recommendation
- Factorization Machines

Anomaly Detection
- Random Cut Forests
- IP Insights

Regression
- Linear Learner
- XGBoost
- KNN

Topic Modeling
- LDA
- NTM

Forecasting
- DeepAR

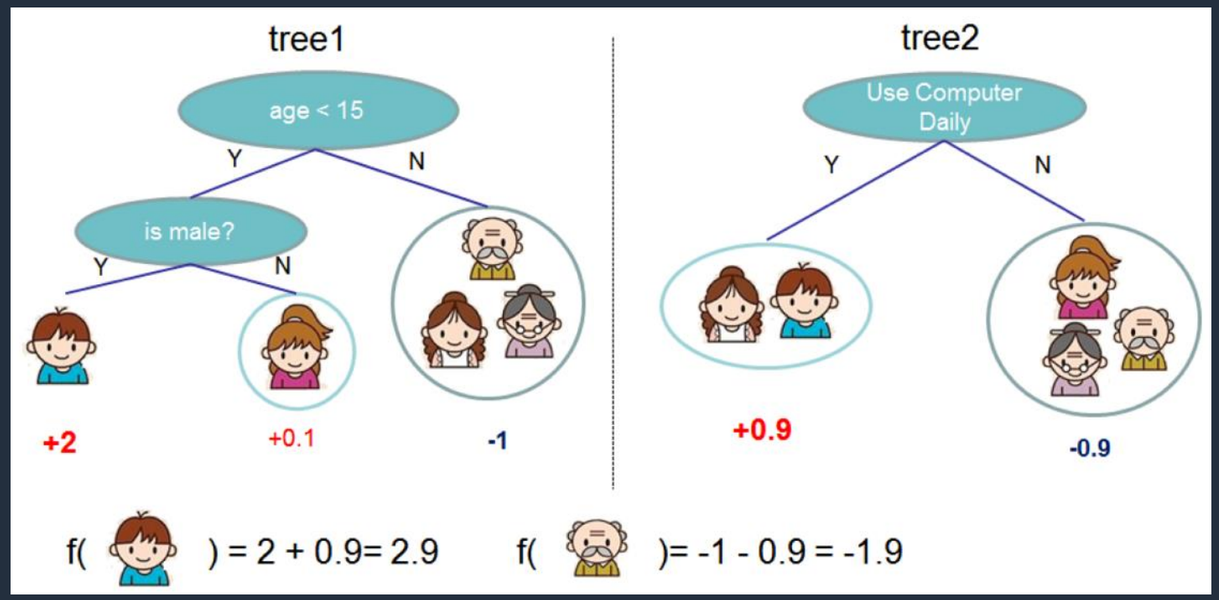
Clustering
- KMeans

Feature Reduction
- PCA
- Object2Vec

aws

XGBoost

- Open-source project
- Popular tree-based algorithm for regression, classification, and ranking
- Handles missing values and sparse data
- Supports distributed training
- Can work with datasets larger than RAM



https://github.com/dmlc/xgboost
https://xgboost.readthedocs.io/en/latest/
https://arxiv.org/abs/1603.02754

AWS Marketplace

You can shop for algorithms, models, and data in AWS Marketplace

Browse or search
AWS Marketplace

Subscribe in a
single click

Available in
Amazon SageMaker

aws

Amazon SageMaker helps you build, train, and deploy models

Prepare

Build

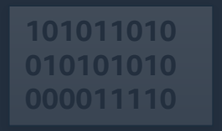
Train & Tune

Deploy & Manage

Web-based IDE for machine learning

Automatically build and train models

Fully managed data processing jobs and data labeling workflows



Collect and prepare training data

One-click collaborative notebooks and built-in, high performance algorithms and models

Choose or build an ML algorithm

One-click training

Set up and manage environments for training

Debugging and optimization

Train, debug, and tune models

Visually track and compare experiments

Manage training runs

One-click deployment and autoscaling

Deploy model in production

Automatically spot concept drift

Monitor models

Add human review of predictions

Validate predictions

Fully managed with auto-scaling for 75% less

Scale and manage the production environment

aws

Train your model with one click using Amazon SageMaker

Train with your own algorithms

Distributed by default

Train on a data stream

Single pass training

Not memory bound

Checkpoint for re-training

aws

Amazon SageMaker Automatic Model Tuning

Automatically tune hyperparameters across algorithms

Tuning at scale

Adjust thousands of different combinations of algorithm parameters

Automated

Uses ML to find the best parameters

Faster

Eliminate days or weeks of tedious manual work

Examples

Decision Trees	Neural Networks
Tree depth	Number of layers
Max leaf nodes	Hidden layer width
Gamma	Learning rate
Eta	Embedding dimensions
Lambda	
Alpha	Dropout

aws

Amazon SageMaker Experiments

Organize, track, and compare training experiments

Tracking at scale

Track parameters & metrics across experiments & users

Custom organization

Organize experiments by teams, goals, & hypotheses

Visualization

Easily visualize experiments and compare

Metrics and logging

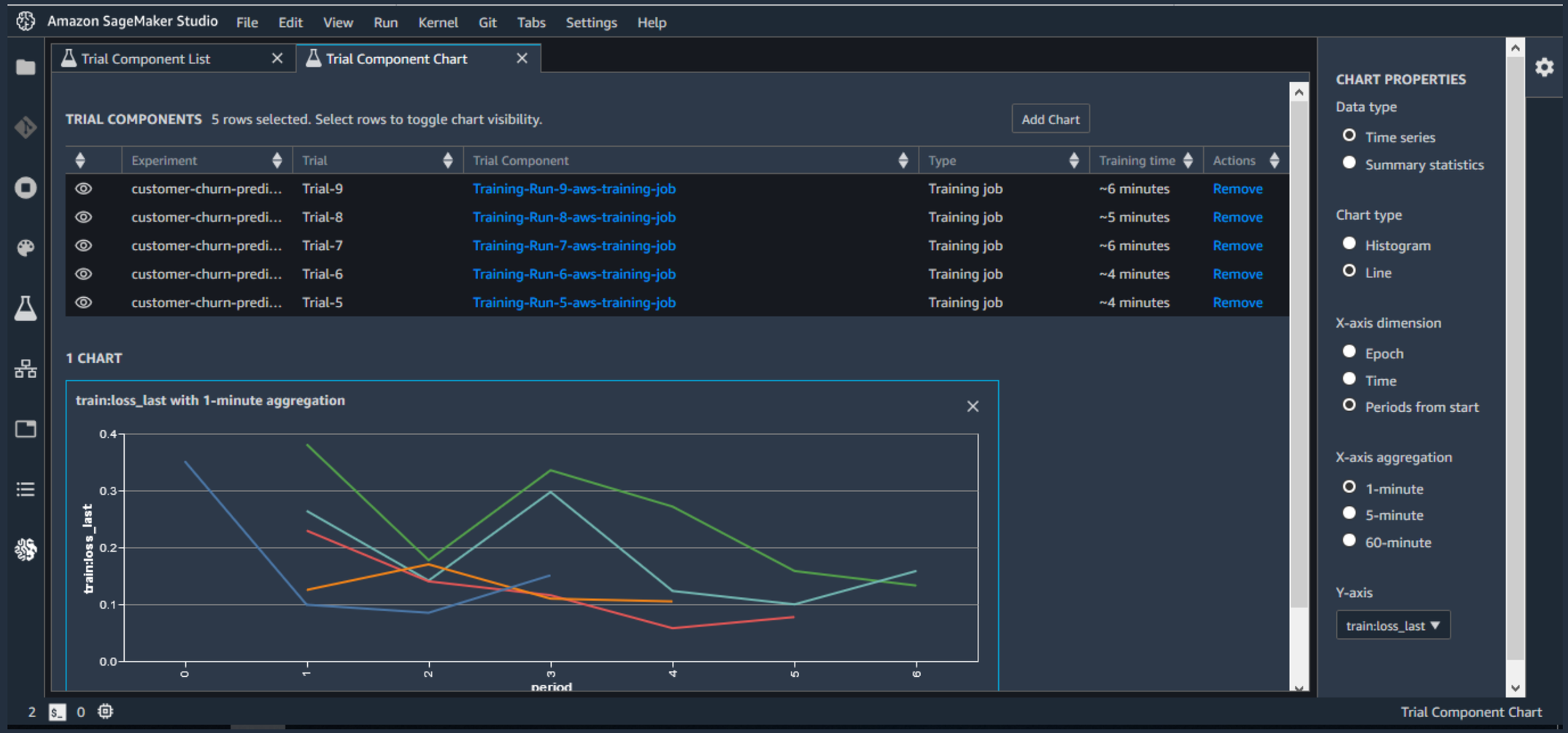
Log custom metrics using the Python SDK & APIs

Fast Iteration

Quickly go back & forth & maintain high-quality

aws

Use Amazon SageMaker Experiments to track and manage thousands of experiments

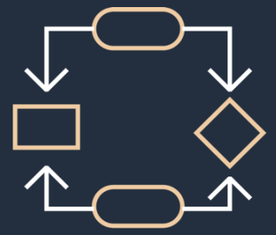


Amazon SageMaker Debugger

Analysis and debugging, explainability, and alert generation

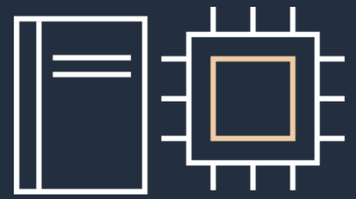
Relevant data capture

Data is automatically captured for analysis



Data analysis & debugging

Analyze & debug data with no code changes



Automatic error detection

Errors are automatically detected based on rules

Improved productivity with alerts

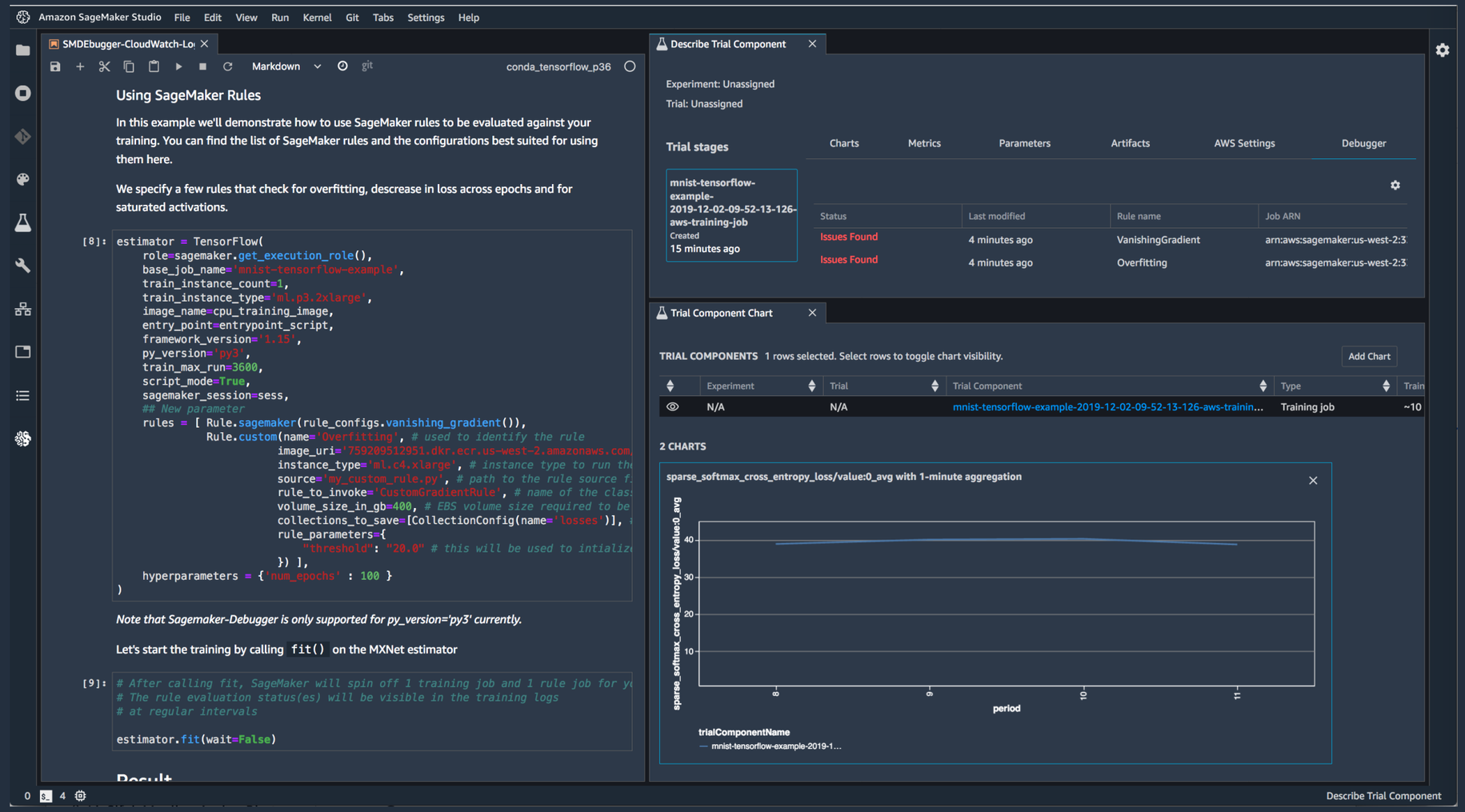
Take corrective action based on alerts

Visual analysis and debugging

Visually analyze & debug from SageMaker Studio

aws

Use Amazon SageMaker Debugger to identify issues such as vanishing gradients



Amazon SageMaker helps you build, train, and deploy models

Prepare

Build

Train & Tune

Deploy & Manage

Web-based IDE for machine learning

Automatically build and train models

Fully managed data processing jobs and data labeling workflows

101011010
010101010
000011110

Collect and prepare training data

One-click collaborative notebooks and built-in, high performance algorithms and models

Choose or build an ML algorithm

One-click training

Set up and manage environments for training

Debugging and optimization

Train, debug, and tune models

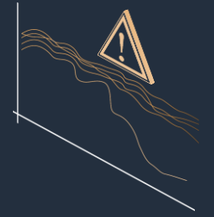
Visually track and compare experiments

Manage training runs

One-click deployment and autoscaling

Deploy model in production

Automatically spot concept drift



Monitor models

Add human review of predictions

Validate predictions

Fully managed with auto-scaling for 75% less

Scale and manage the production environment

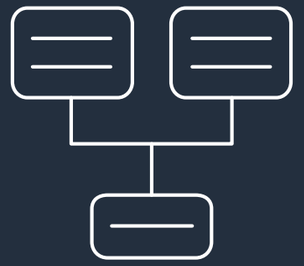
aws

Amazon SageMaker is fully managed

One click model deployment

Auto-scaling

Low latency and high throughput



Bring your own model

Python SDK

Deploy multiple models on an endpoint

aws

Amazon SageMaker Model Monitor

Continuous monitoring of models in production

Automatic data collection

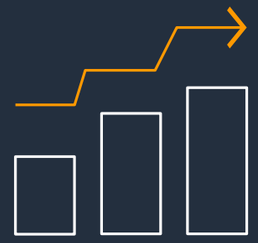
Data is automatically collected from your endpoints

Continuous Monitoring

Define a monitoring schedule and detect changes in quality against a pre-defined baseline

Flexibility with rules

Use built-in rules to detect data drift or write your own rules for custom analysis



Visual data analysis

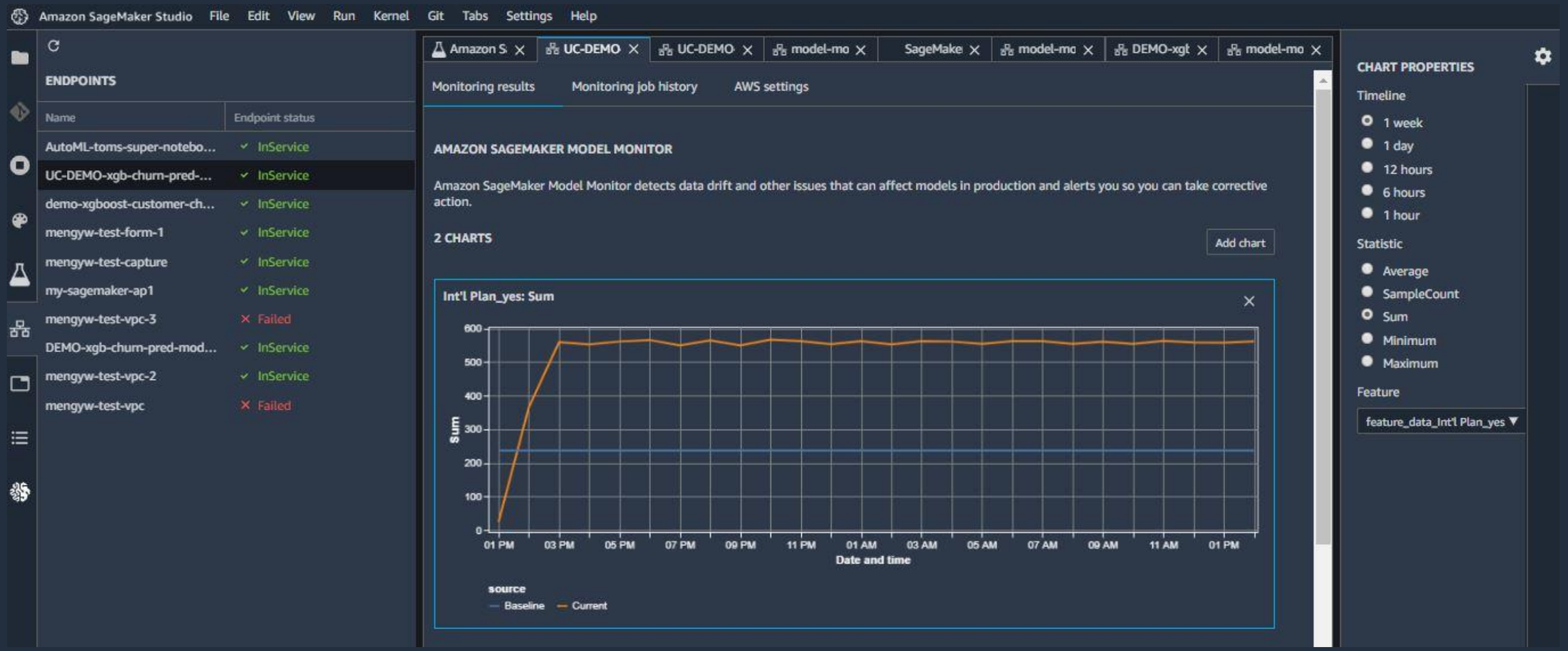
See monitoring results, data statistics, and violation reports in SageMaker Studio

CloudWatch Integration

Automate corrective actions based on Amazon CloudWatch alerts

aws

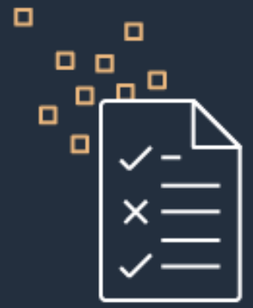
Use Amazon SageMaker Model Monitor to identify model drift and take action



aws

Amazon Augmented AI

Easily build workflows required for human review of predictions



Easily implement human review workflows

Reduce time to market with pre-built workflows and UIs

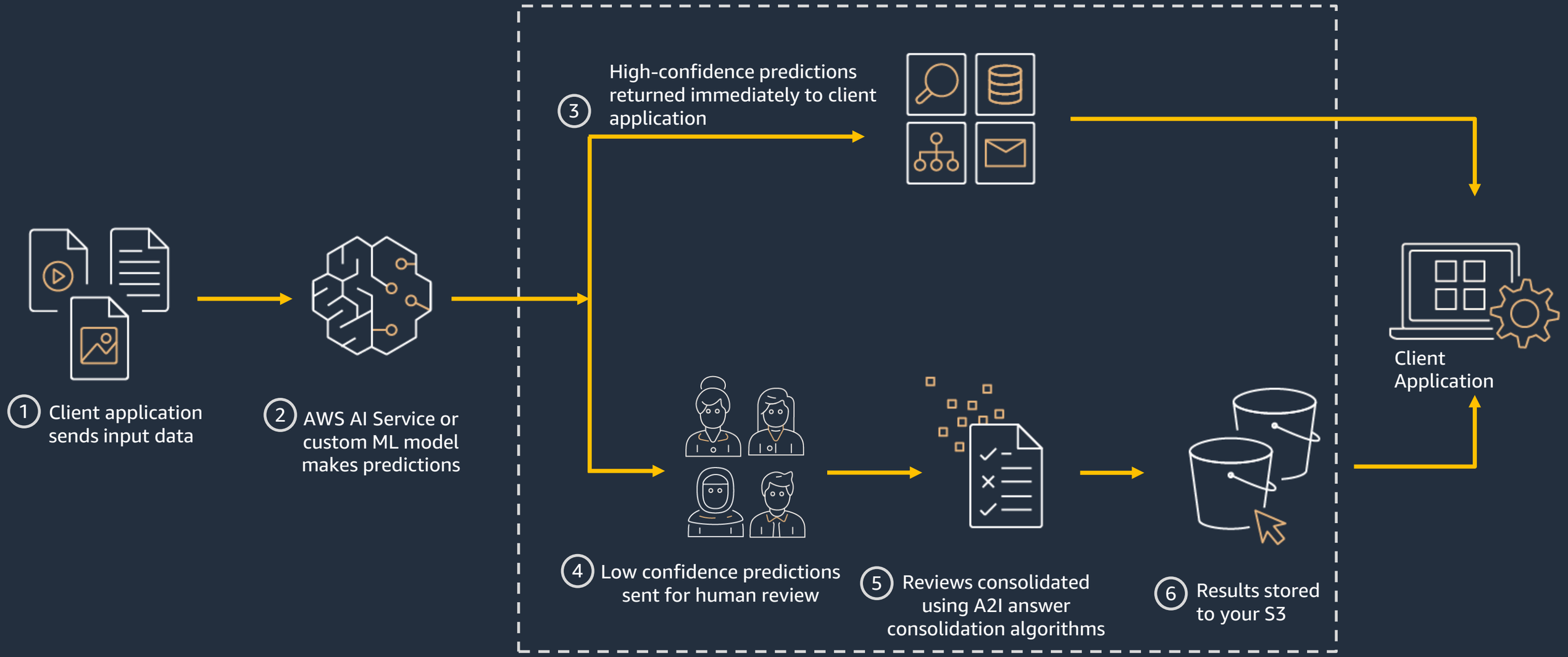
Multiple workforce options

Integrate with your custom ML models

Pre-built algorithms to increase accuracy

aws

How Amazon Augmented AI works

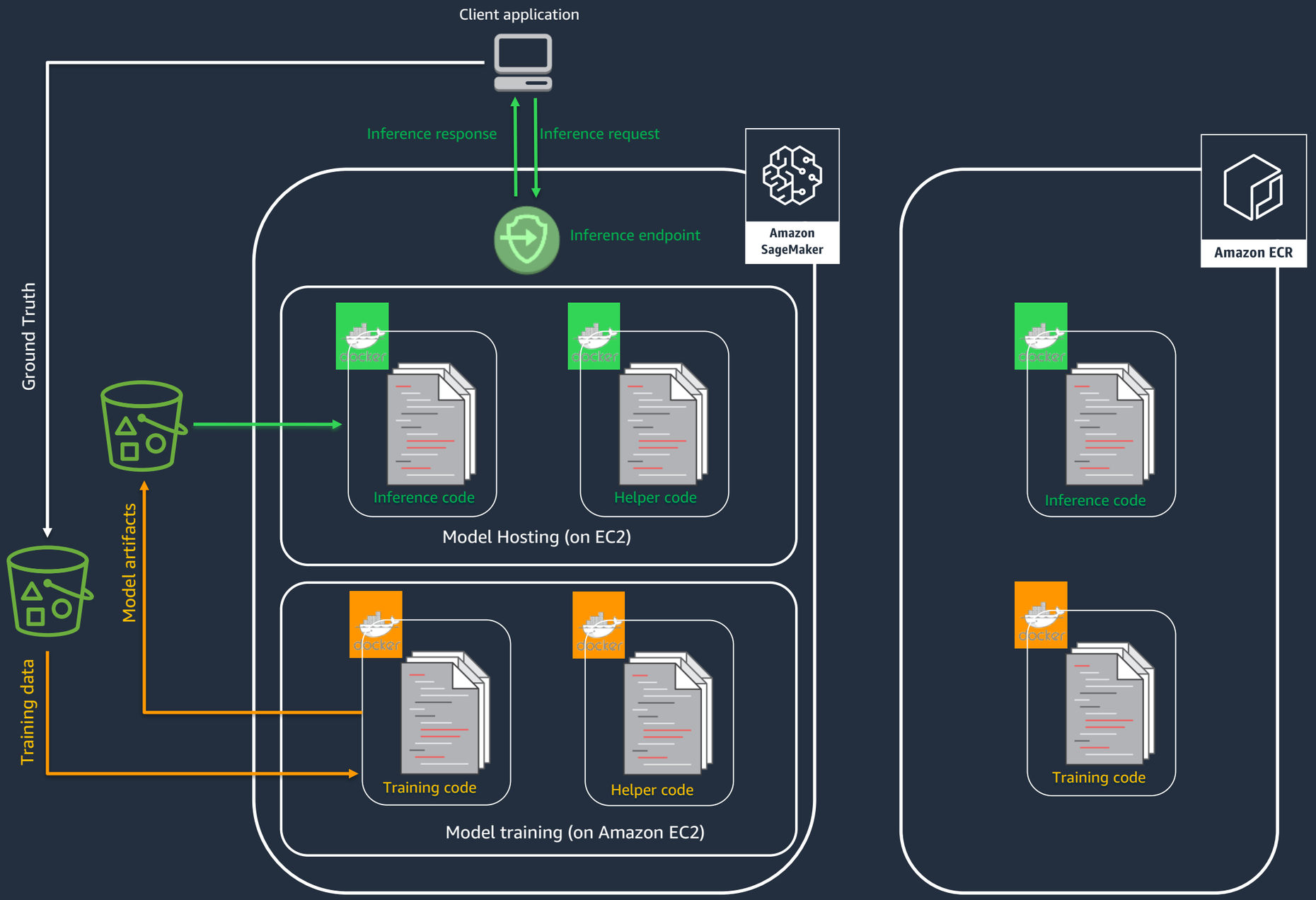


Get started with Amazon SageMaker

Prepare	Build	Train & Tune	Deploy & Manage
Amazon SageMaker Studio *Integrated Development environment(IDE) for Machine Learning*			
Amazon SageMaker Autopilot *Automatically build and train models*			**One Click Deployment** *Supports real-time, batch & multi-model*
Amazon SageMaker GroundTruth *Build and manage training dataset*	**Amazon SageMaker Notebooks** *One-click notebooks with elastic compute*	**One Click Training** *Supports supervised, unsupervised & RL*	**Amazon SageMaker Model Monitor** *Automatically detect concept drift*
Processing Job *Supports Python or Spark*	**AWS Marketplace** *Pre-built algorithms, models, and data*	**Automatic Model Tuning** *One-click hyperparameter optimization*	**Amazon SageMaker Neo** *Train once, deploy anywhere*
		Amazon SageMaker Experiments *Capture, organize, and compare every step*	**Amazon Elastic Inference** *Auto scaling for 75% less*
		Amazon SageMaker Debugger *Debug and profile training runs*	**Amazon Augmented AI** *Add human review of model predictions*

aws

Useful information for the labs



Model options

Training code

Factorization machines
Linear learner
Principal component analysis
K-means
XGBoost
And more

Built-in algorithms

mxnet Chainer TensorFlow PyTorch

Bring your own script

Bring your own container

aws

Amazon SageMaker SDK

- AWS SDK for Python orchestrating all Amazon SageMaker activity
  - Algorithm selection, training, deploying, hyperparameter optimization, and so on
  - There's also a Spark SDK (Python and Scala), which we won't cover today

- High-level objects for:
  - Some built-in algos: K-means, PCA, and the like
  - Deep-learning libraries: TensorFlow, MXNet, PyTorch, Chainer
  - Sagemaker.estimator.estimator for everything else

https://github.com/aws/sagemaker-python-sdk
https://sagemaker.readthedocs.io/en/latest/

aws

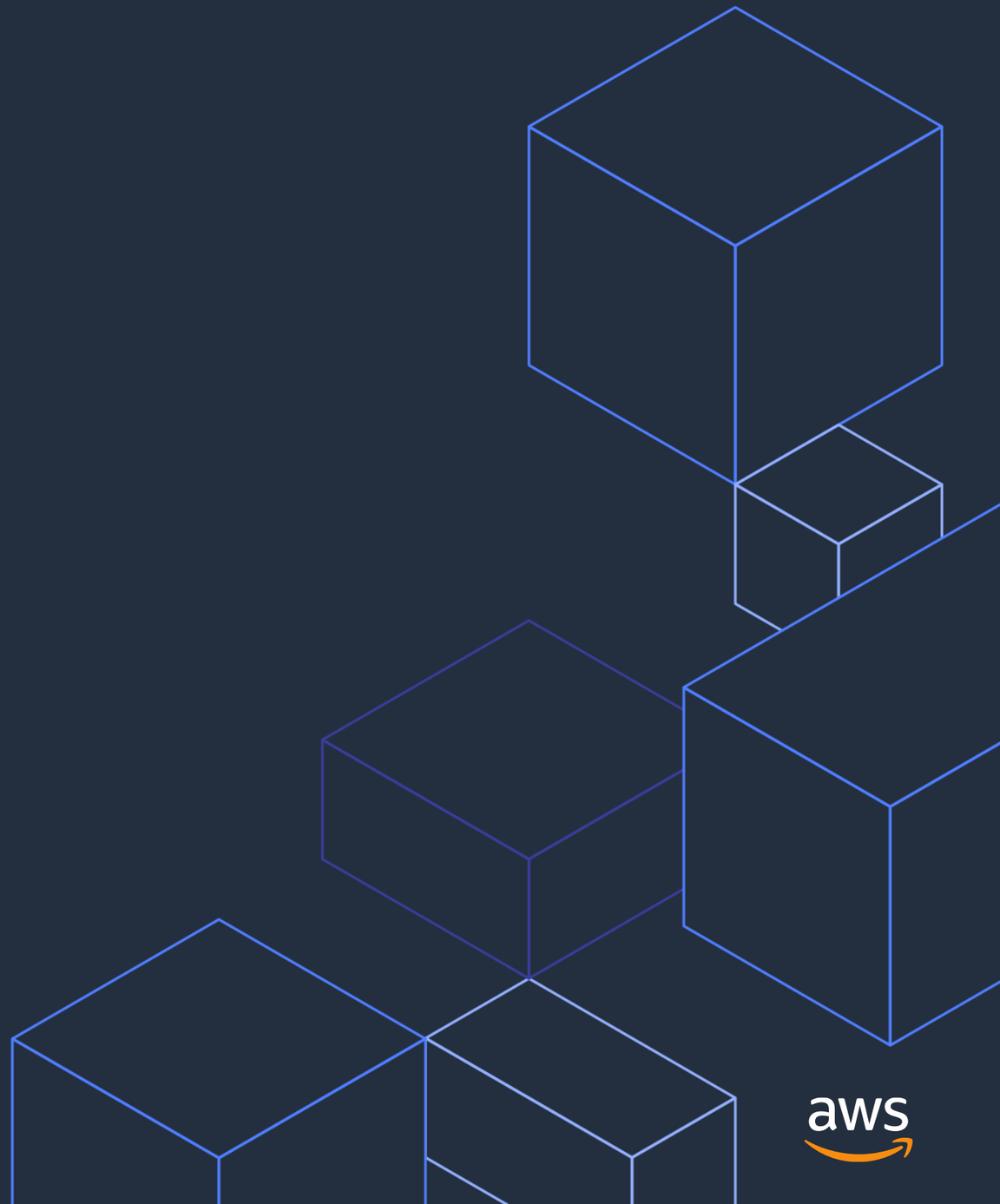
Confusion matrix

Actual \ Predict	0	1
0	True negative	False positive
1	False negative	True positive

Actual \ Predict	0	1
0	3567	71
1	355	126

aws

Labs



Problem statement

Direct marketing is a common tactic to acquire customers. Because resources and a customer's attention are limited, the goal is to target only the subset of prospects who are likely to engage with a specific offer.

Predicting those potential customers based on readily available information like demographics, past interactions, and environmental factors is a common machine-learning problem.

We will train a model using XGBoost on a bank marketing dataset provided by UCI's ML Repository to predict if a customer will enroll for a term deposit at a bank after one or more phone calls.

aws

Walkthrough: Notebook instance setup

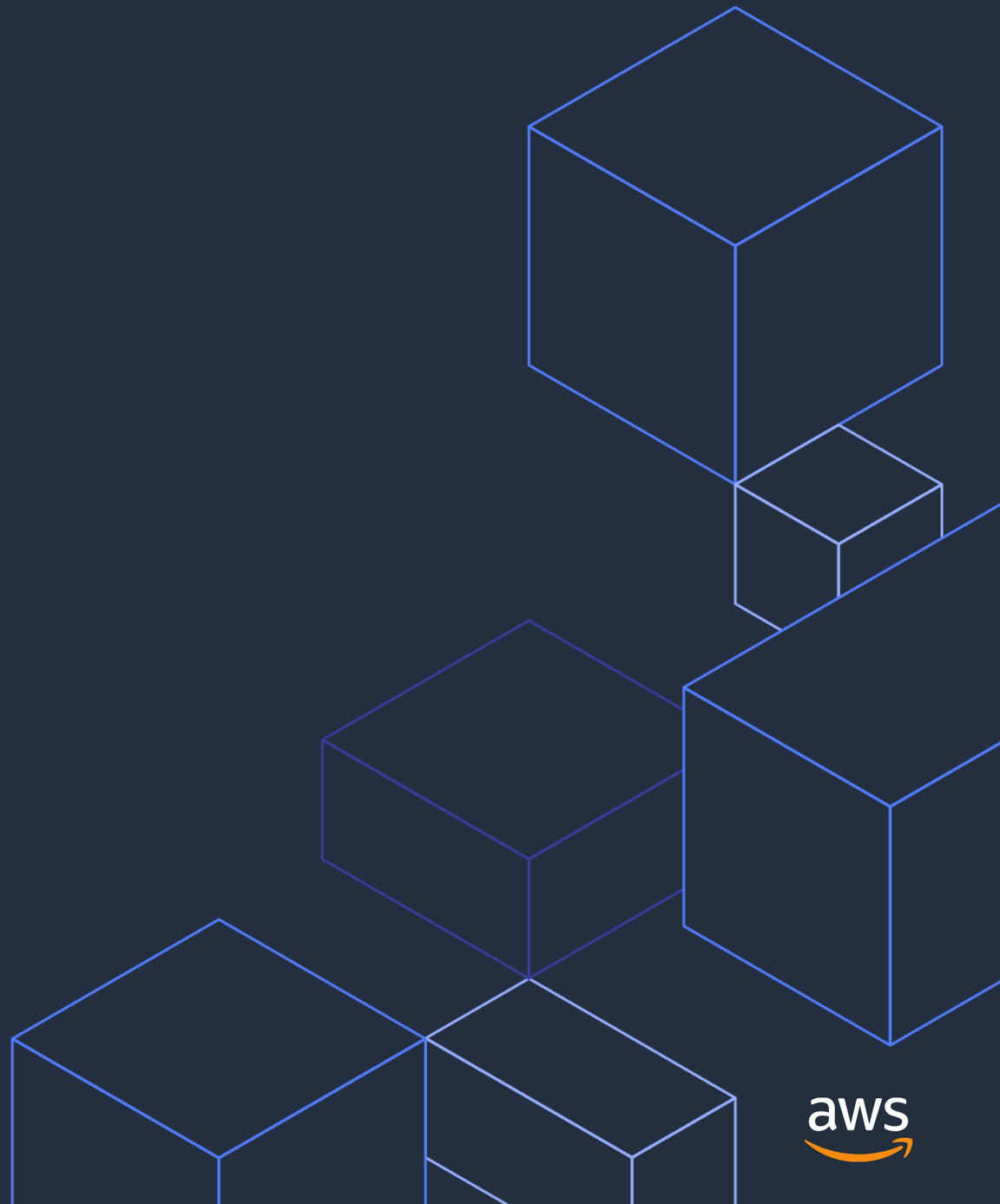
Labs

1. Preparing the data
2. Training our first model with XGBoost
3. Deploying our model
4. Predicting with our model
5. Manually tuning our model
6. Finding optimal hyperparameters with automatic model tuning
7. Deploying our best 2 models
8. Predicting with our best 2 models

bit.ly/2OEdvFM

aws

Resources



Resources

https://ml.aws

https://aws.amazon.com/sagemaker
https://github.com/awslabs/amazon-sagemaker-examples
https://github.com/aws/sagemaker-python-sdk

https://github.com/awslabs/amazon-sagemaker-workshop

aws

Thank you!

Jonathan Dion
Senior Technical Evangelist
Amazon Web Services
@jotdion
linkedin.com/in/jotdion