
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

AIM314

Accelerate your ML journey with Amazon SageMaker low-code tools

Claire O'Brien Rajkumar
Sr. Product Manager
Amazon Web Services

Vadim Omeltchenko
Sr. AI/ML Solutions Architect
Amazon Web Services

Ori Goshen
Co-Founder, Co-CEO
AI21 Labs

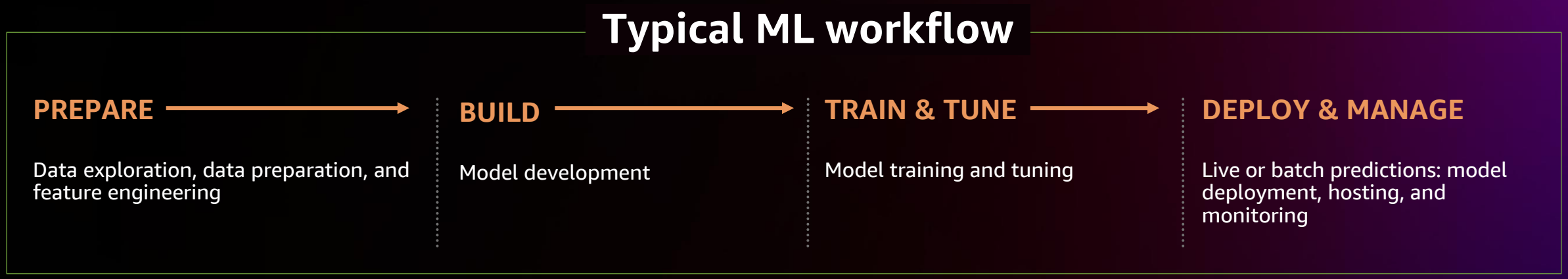
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What this session will cover

1. Explanation of low-code ML
2. Overview of Amazon SageMaker low-code tools
3. Demos of Amazon SageMaker low-code tools

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The machine learning workflow



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Common challenges with machine learning

1. Requires deep expertise to prepare the data and build the models
2. Experimentation is time-consuming and resource-intensive
3. Data scientists are oversubscribed, and needs are only increasing

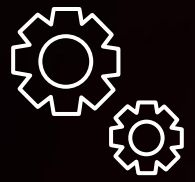
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Low-code machine learning



Faster ML workflows



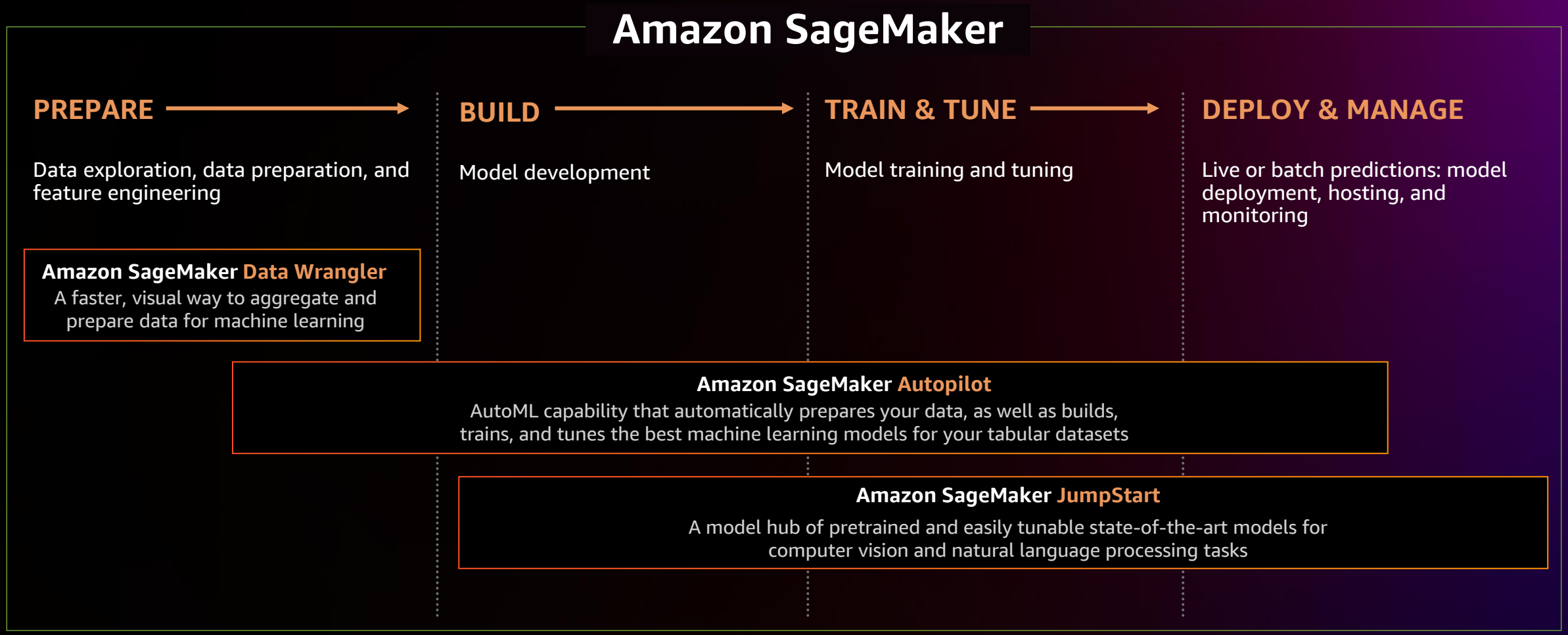
Open and customizable solutions



Appropriate for both new and advanced ML practitioners

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Low-code machine learning from AWS



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker Data Wrangler

Prepare

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

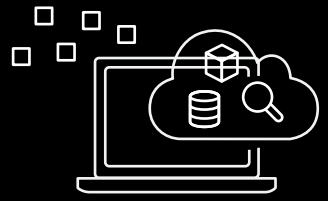
How SageMaker Data Wrangler works

LOW-CODE TOOL FOR PREPARING DATA FOR ML



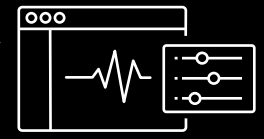
Amazon SageMaker Data Wrangler

A faster, visual way to aggregate and prepare data for machine learning



Connect & import data

Import data from a wide variety of data sources in various supported formats



Get insights on data quality

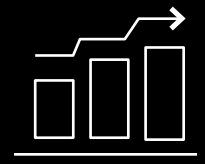
Detect data quality issues & bias with automatic analyses & visualizations



Cleanse & enrich

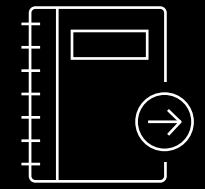
Perform feature engineering with 300+ built-in data transforms

Also bring your own custom transforms



Analyze and visualize

Leverage built-in analyses like target leakage, quick model, and more to understand your data



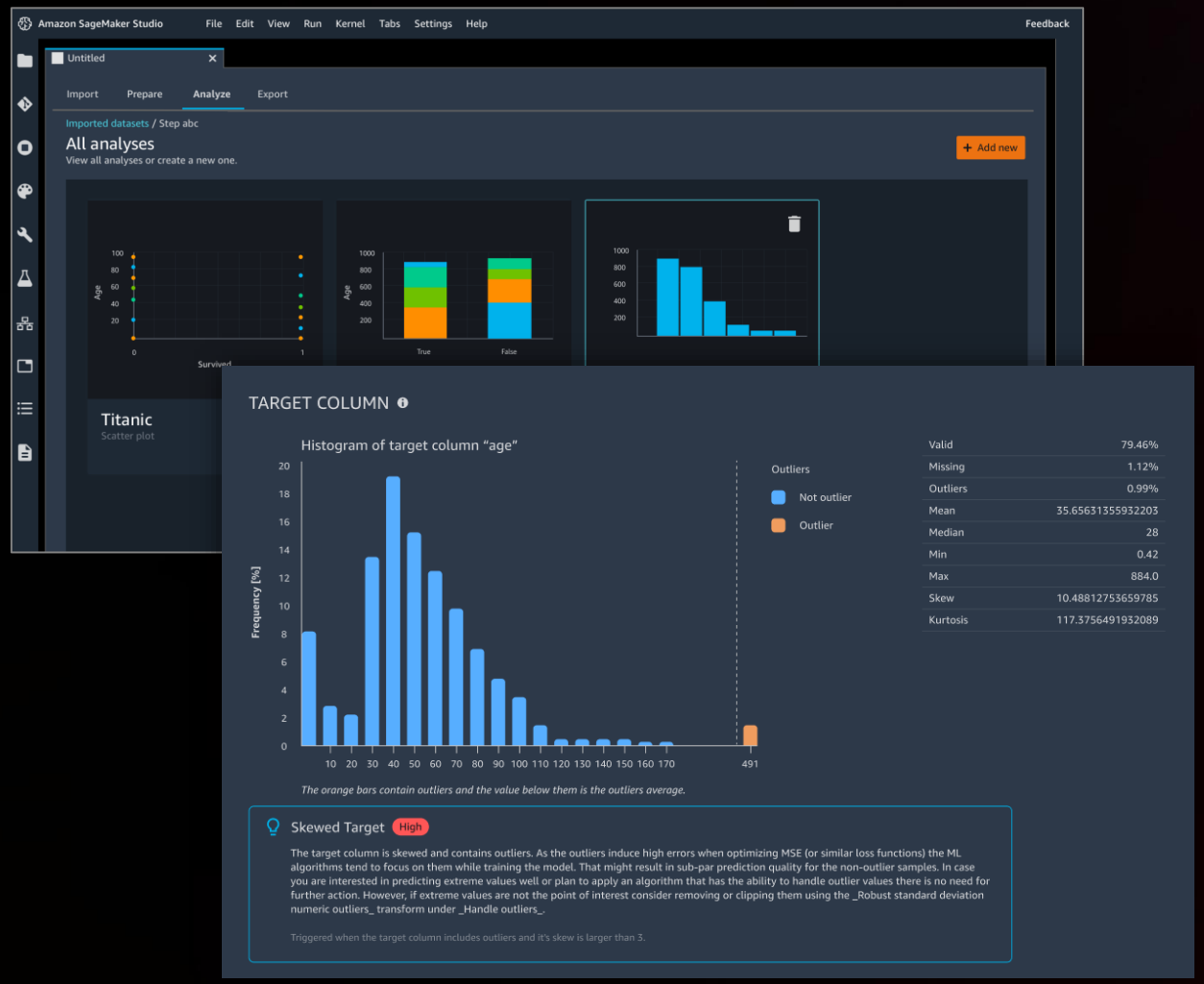
Export prepared data to train and deploy models

Export prepared data to train & deploy ML models via SageMaker Autopilot, notebooks, or code scripts

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

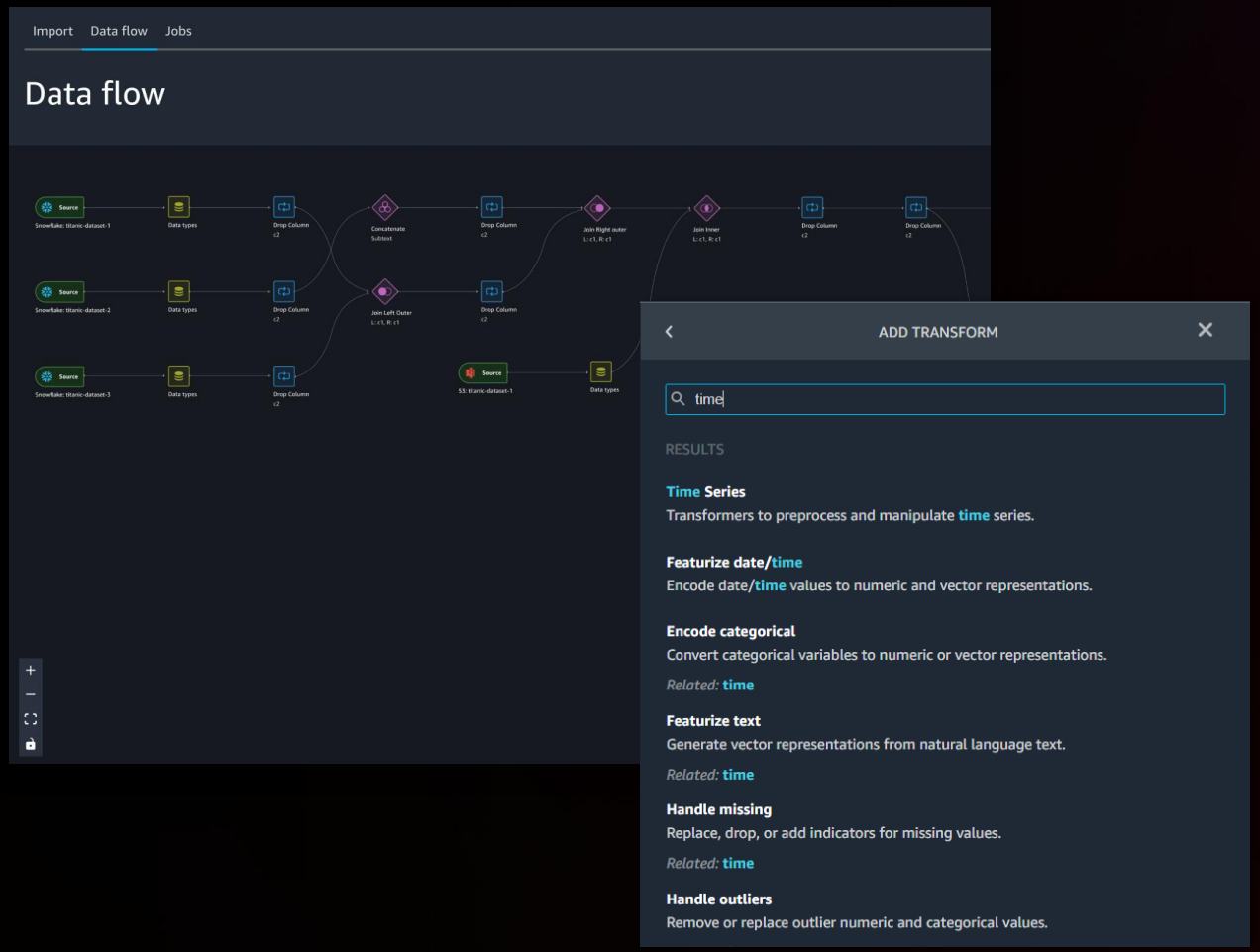
Quickly understand and iterate



- **ML-powered analyses** to diagnose data quality issues
- **SageMaker Clarify** integration to detect and mitigate bias
- **Quick Model** to quickly estimate quality of a model you train on your data

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

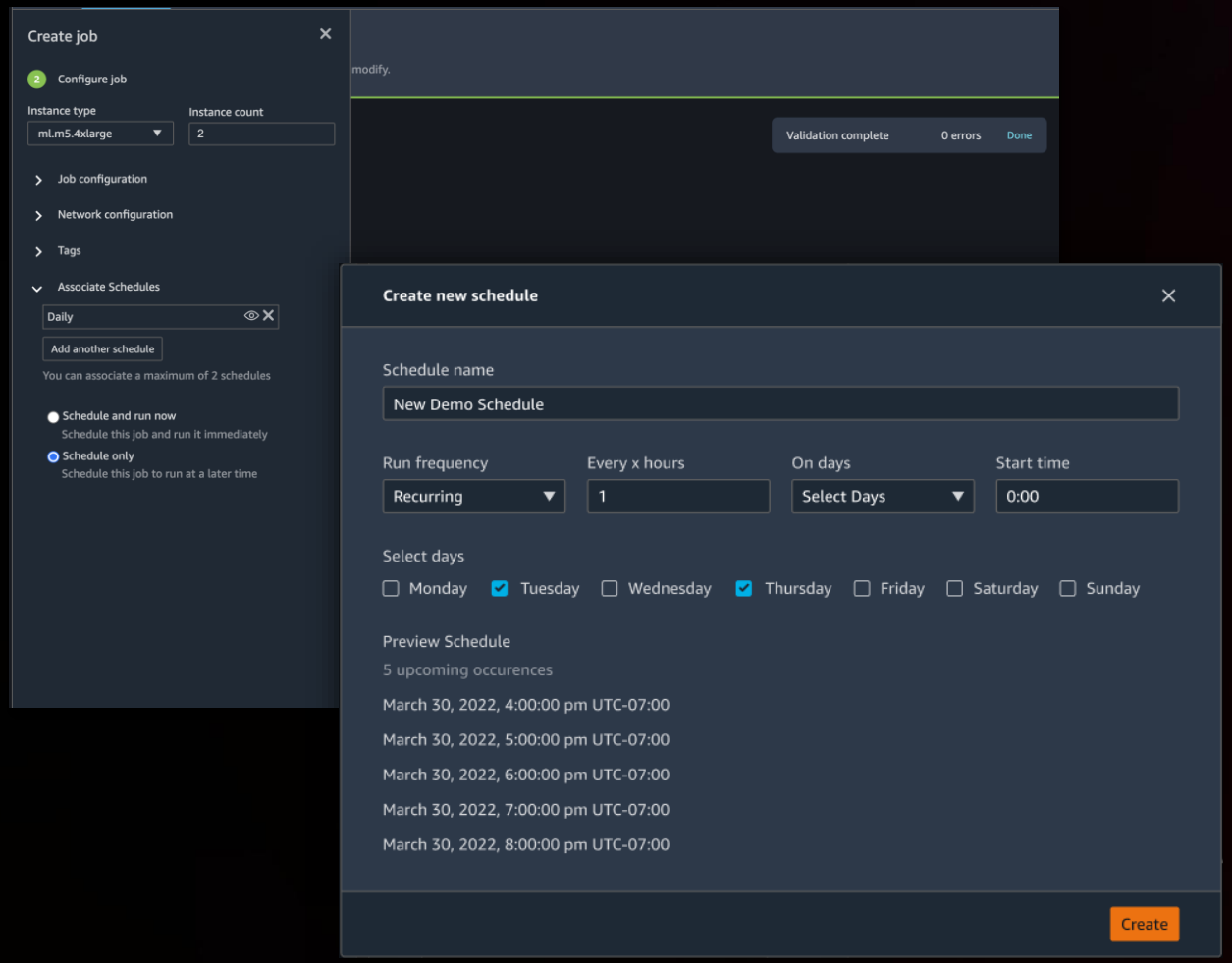
Transform with point and click, code snippets, and code



- **Point and click** to implement PySpark-based transformations
- **Edit code snippets** authored in PySpark
- **Write** your own PySpark, Pandas, or SQL-based custom transformations

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

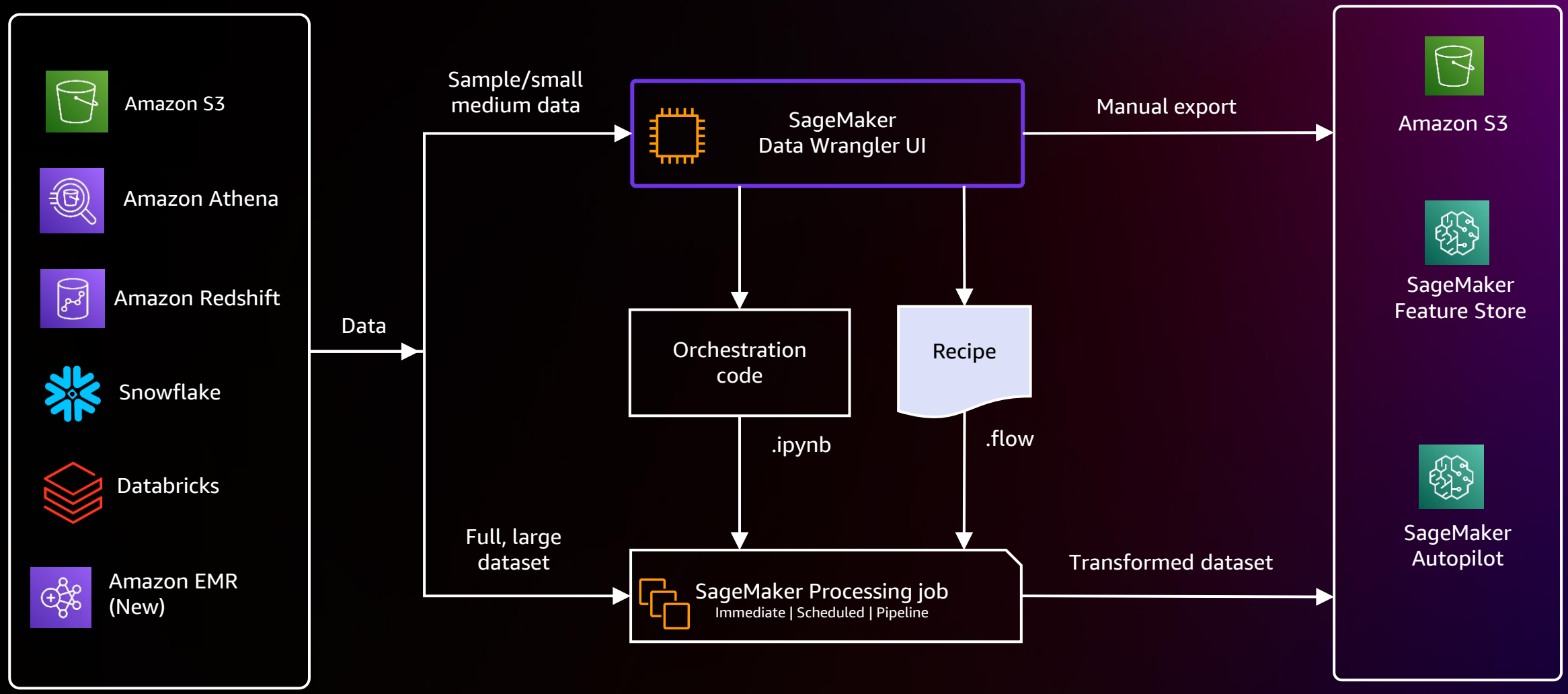
Process data at scale from UI or notebooks



- Schedule **serverless processing job**
- Export orchestration code to **notebook or Python file** to use in pipeline
- **SageMaker Feature Store** and Amazon S3 supported destinations

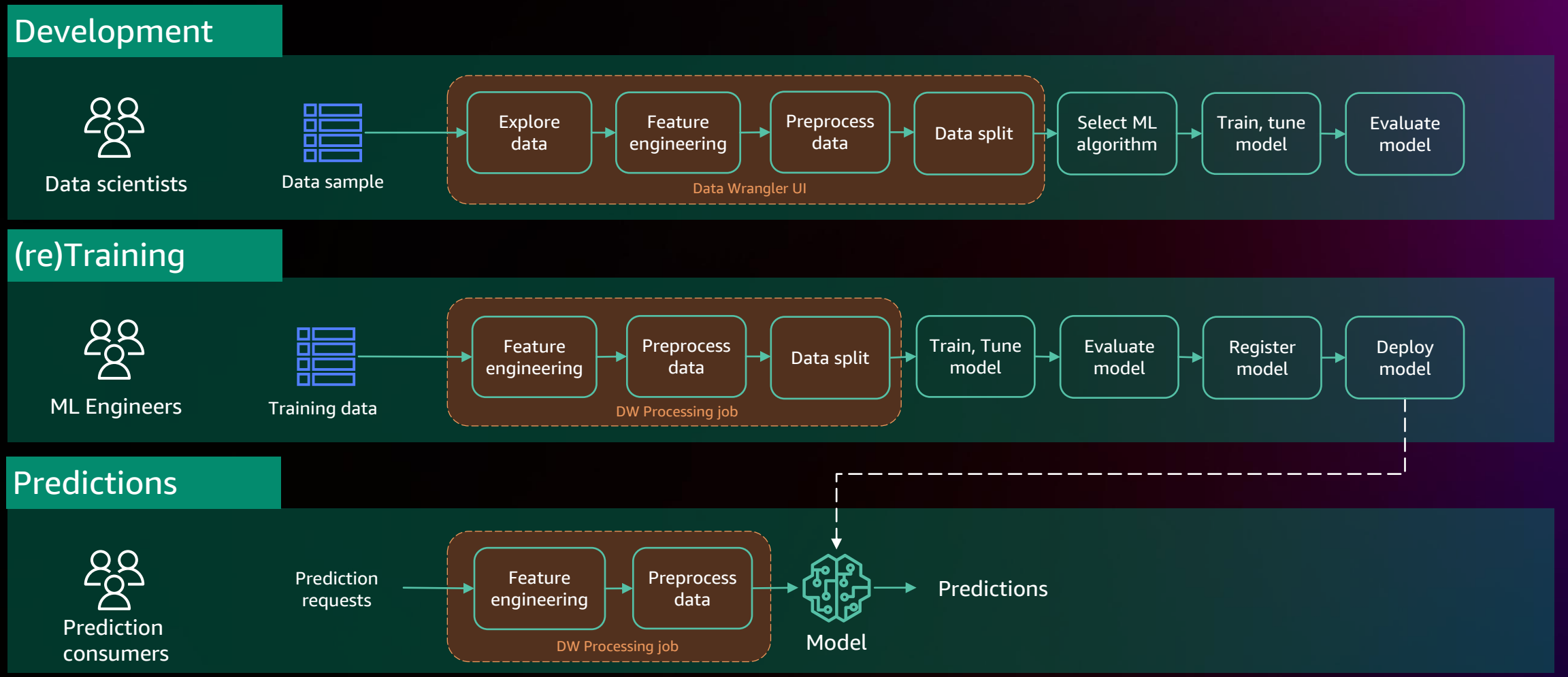
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Feature engineering with SageMaker Data Wrangler



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SageMaker Data Wrangler in your organization



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker Autopilot

Build > Train & tune > Deploy & manage

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

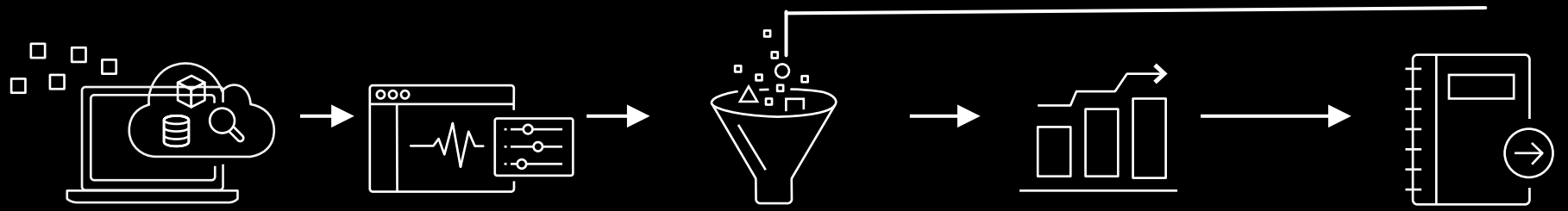
How SageMaker Autopilot works

LOW-CODE TOOL FOR BUILDING, TRAINING, AND TUNING MACHINE LEARNING MODELS



Amazon SageMaker Autopilot

A faster, visual way to train, tune, and deploy ML



Upload data and select target

Load tabular data from S3 and select target column for prediction

Automatic model creation

Automatic feature selection, feature engineering, and selection of algorithms and training modes

Model leaderboard

Review model leaderboard to deploy the best model endpoint for inferencing

Full visibility and controls

Data exploration, model insights, and explainability reports

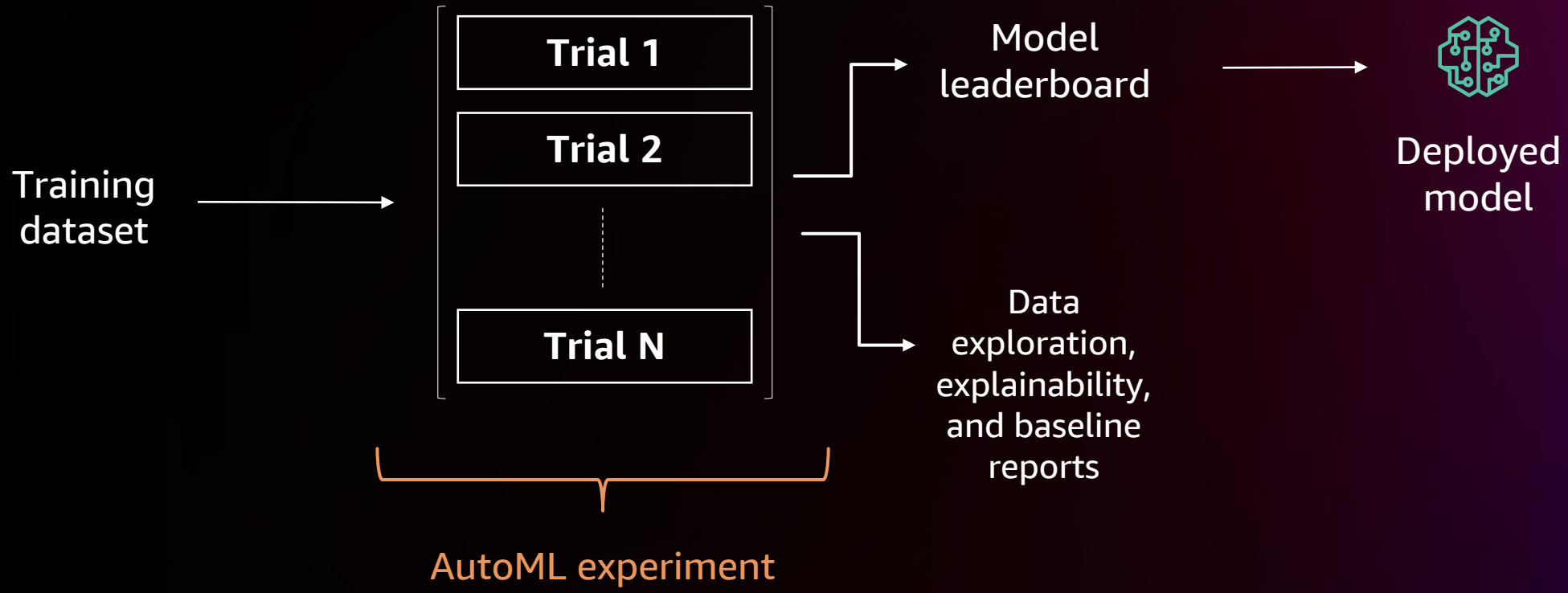
Evaluate, deploy, and register models

Auto-deploy or one-click deploy Autopilot model; easily build training, evaluation, and deployment ML workflows

Experimentation — Customize & optimize — Move to production

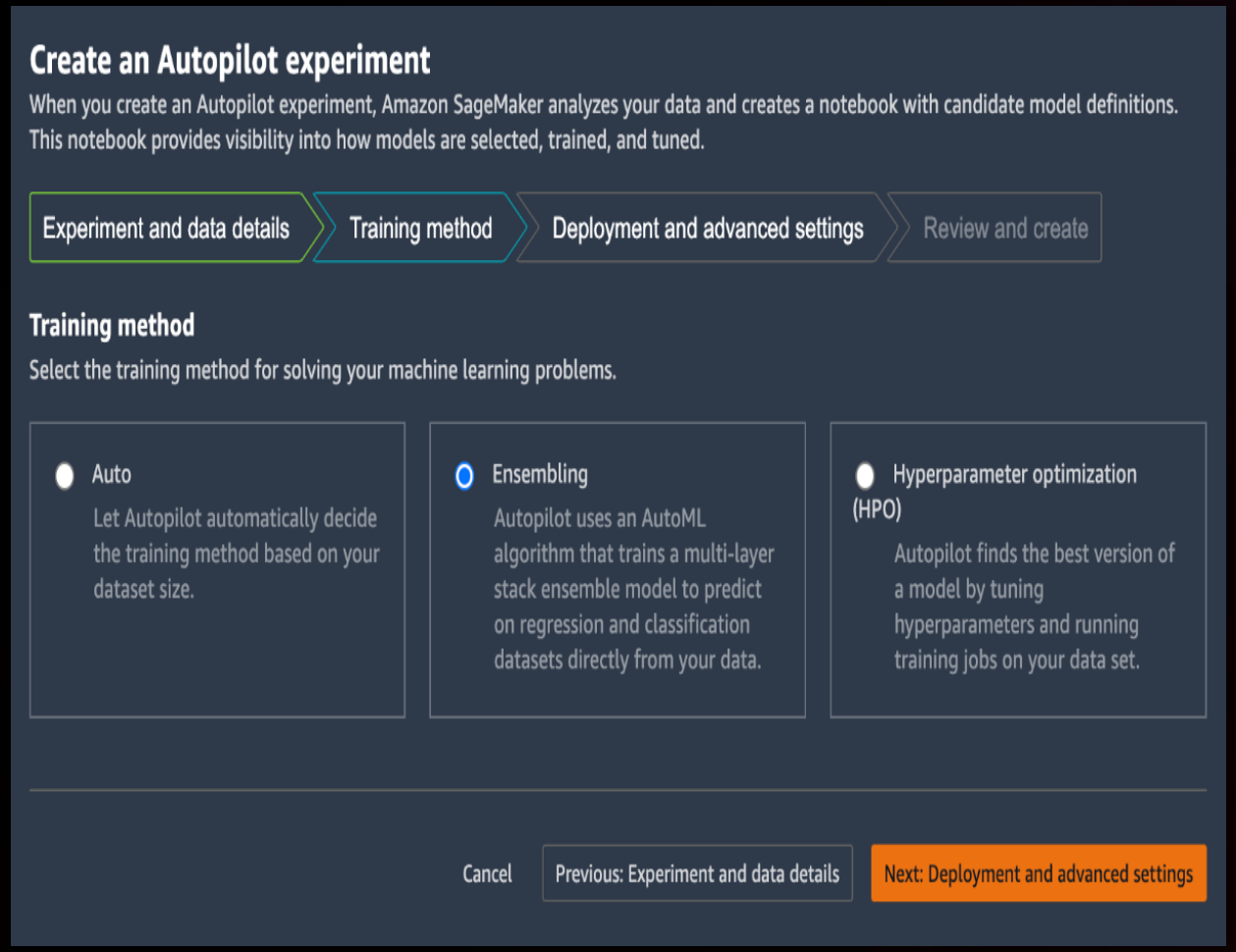
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Accelerating model development



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

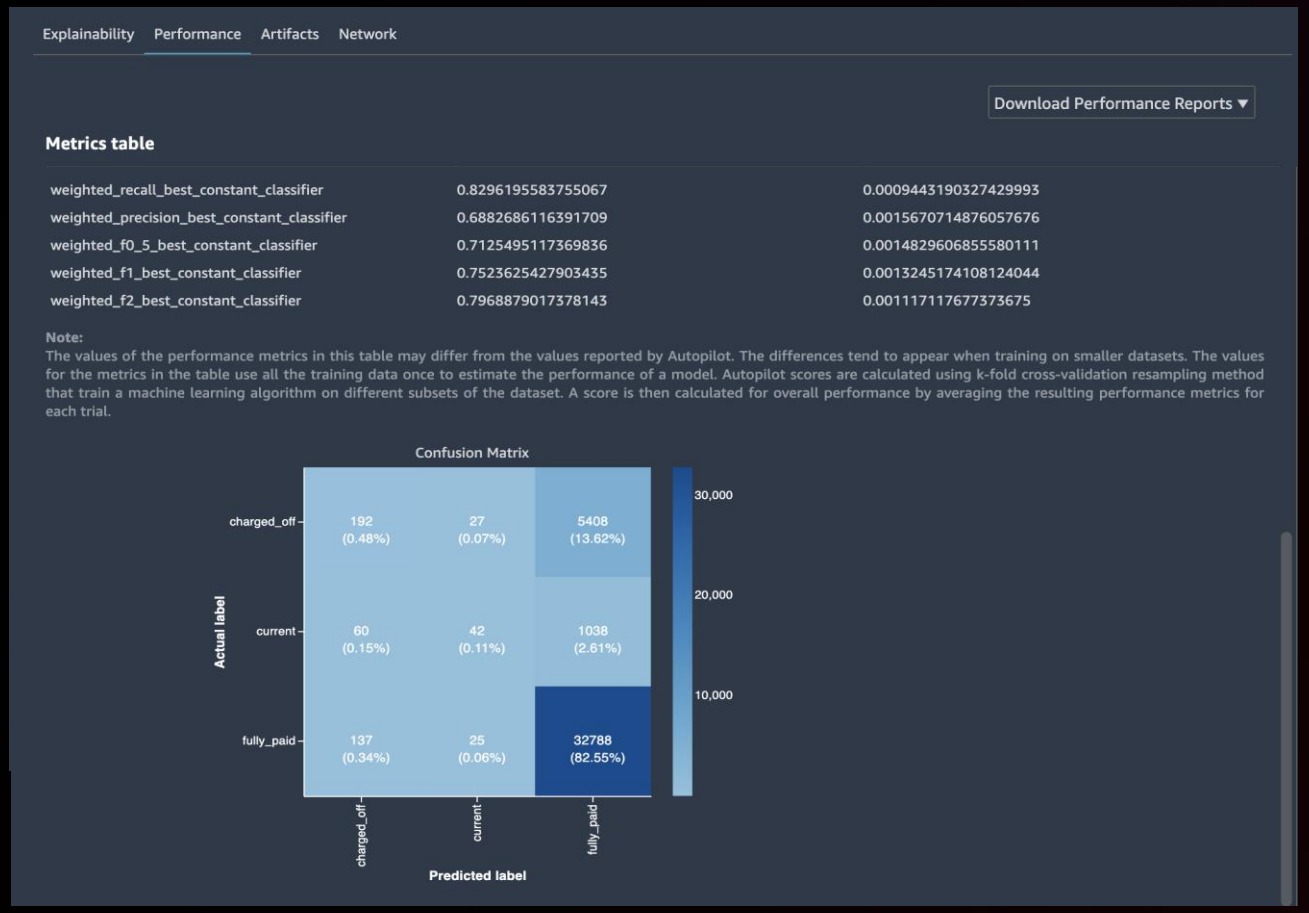
Control: Customizable AutoML journey



- Auto **data preprocessing** or bring your own transforms from **SageMaker Data Wrangler**
- Auto **feature selection** or select features and change auto-inferred data types
- Auto **data split,** specify data split ratio, or upload validation dataset
- **Training modes**
  - Ensemble powered by AutoGluon
  - Hyperparameter optimization (HPO)
  - Auto

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

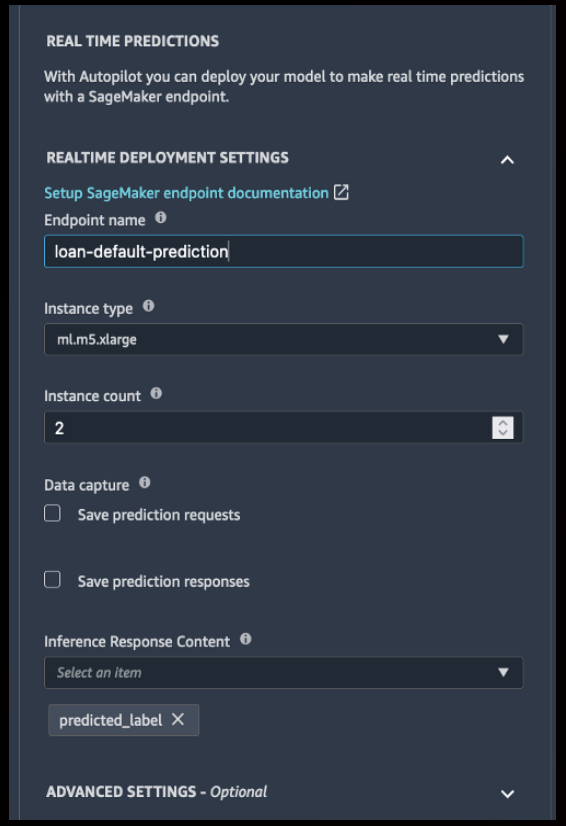
Visibility: Understanding your SageMaker Autopilot model

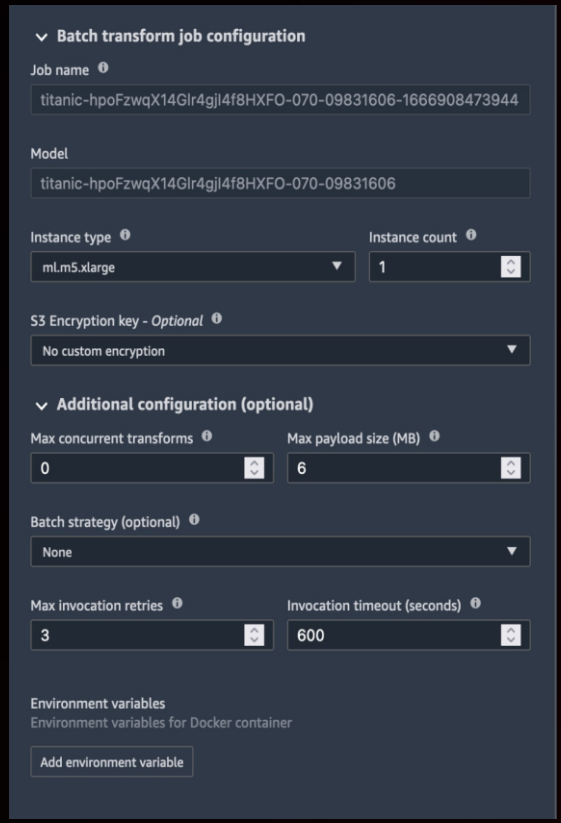


- Model leaderboard and **model assets,** including training and validation data, feature engineering code, deployable model
- Review performance metrics, model insights charts such as confusion matrix, ROC curve, etc., with **model insights report**
- **Explainability** report to understand summary statistics and feature importance
- **Data exploration report** for insights on data and data quality

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Deploy to SageMaker endpoint

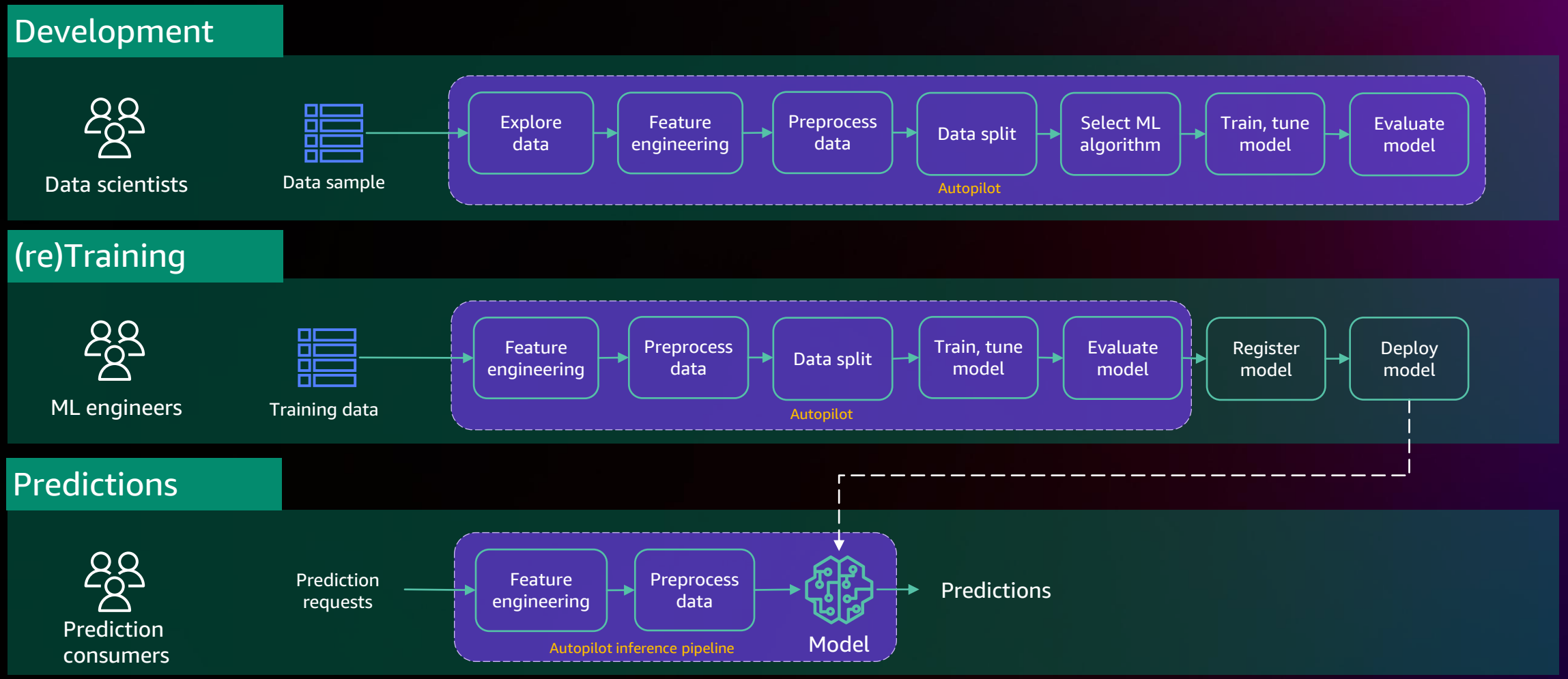




- Deploy your model to make **real-time predictions** with a SageMaker endpoint via UI or invoke with API
- Get **batch predictions** on the provided dataset
- Experiments launched from **SageMaker Data Wrangler** will include SageMaker Data Wrangler transforms when deploying models for inference

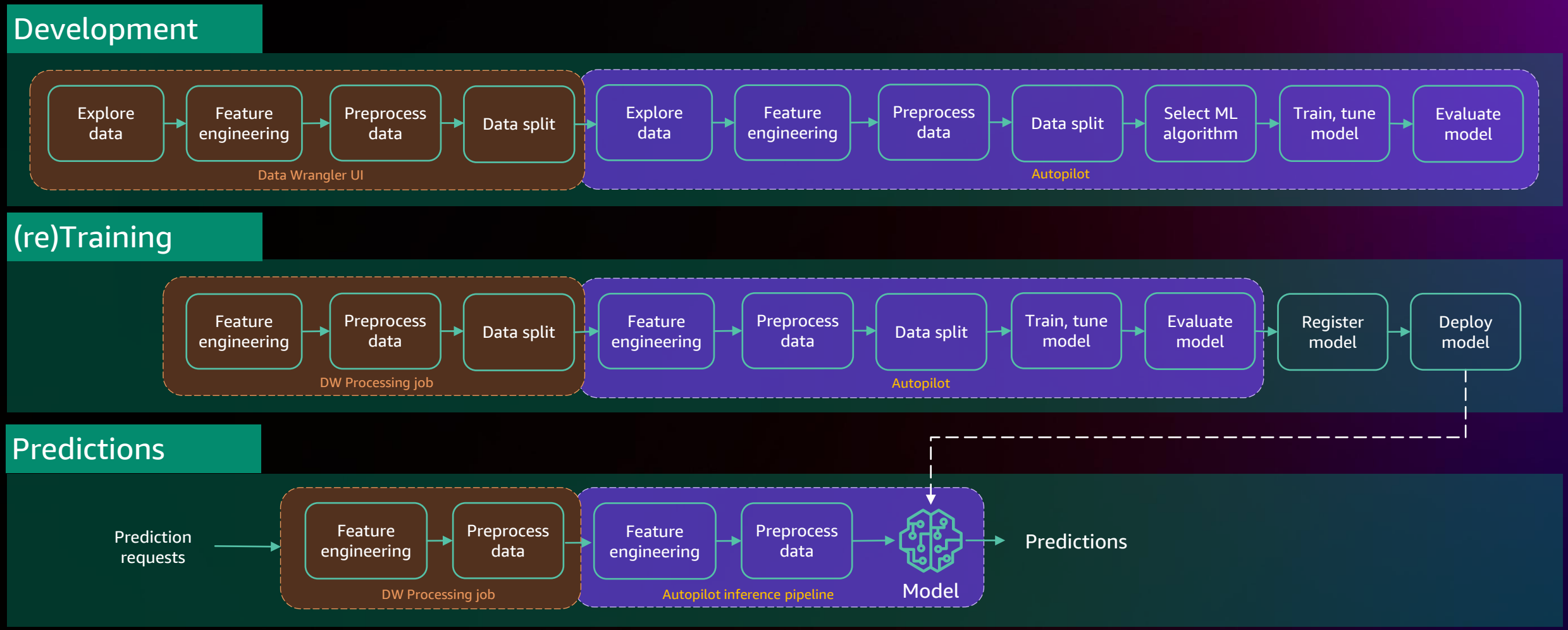
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SageMaker Autopilot in ML workflow



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SageMaker Data Wrangler and SageMaker Autopilot in ML workflow



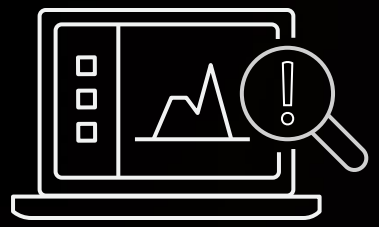
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker JumpStart

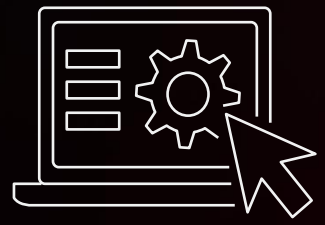
Build > Train & tune > Deploy & manage

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

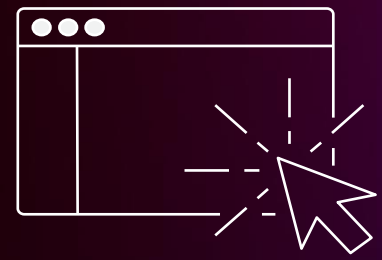
SageMaker Jumpstart: ML hub for SageMaker customers



SageMaker Studio



SageMaker Console



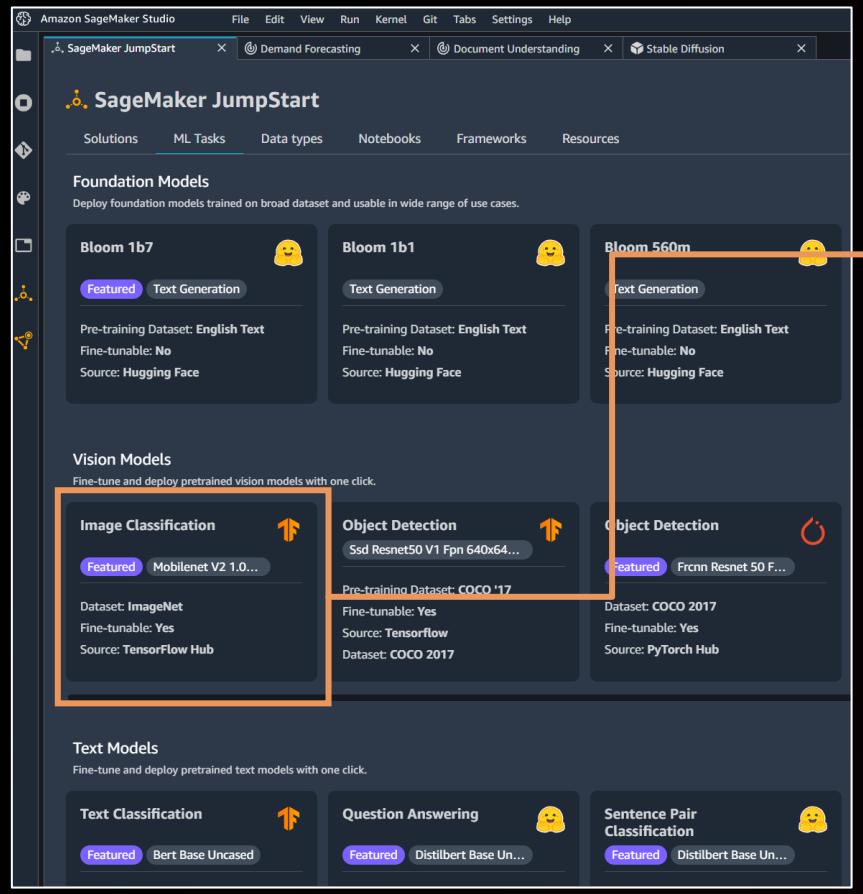
SageMaker SDK

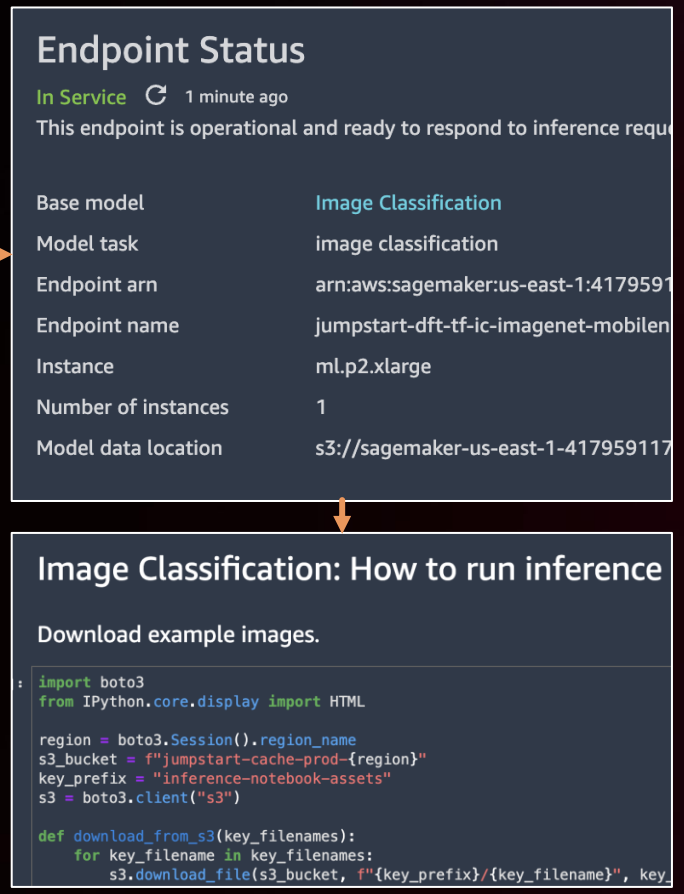
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SageMaker JumpStart – Pretrained models

ACCELERATE TIME TO DEPLOY OVER 350 PUBLICLY AVAILABLE MODELS





Pretrained model example: Image classification

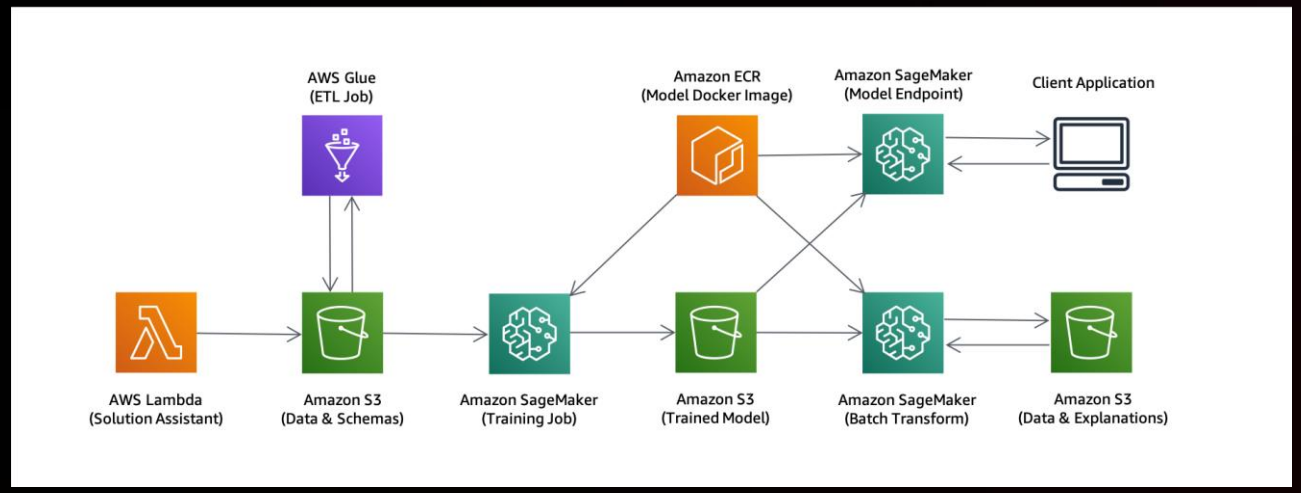
Model categories

- Foundation/generative
- Tabular
- Computer vision
  - Image classification
  - Object detection
  - Semantic segmentation
  - Instance segmentation
  - Image embedding
- Natural language processing
  - Text classification
  - Sentence pair classification
  - Question answering
  - Text summarization
  - Named entity recognition
  - Machine translation
  - Fill-mask
  - Text embedding
  - Text generation
- Audio embedding

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SageMaker JumpStart – Solutions templates

23 PRE-BUILT SOLUTIONS FOR COMMON ML USE CASES



Solution example: Explain credit decisions

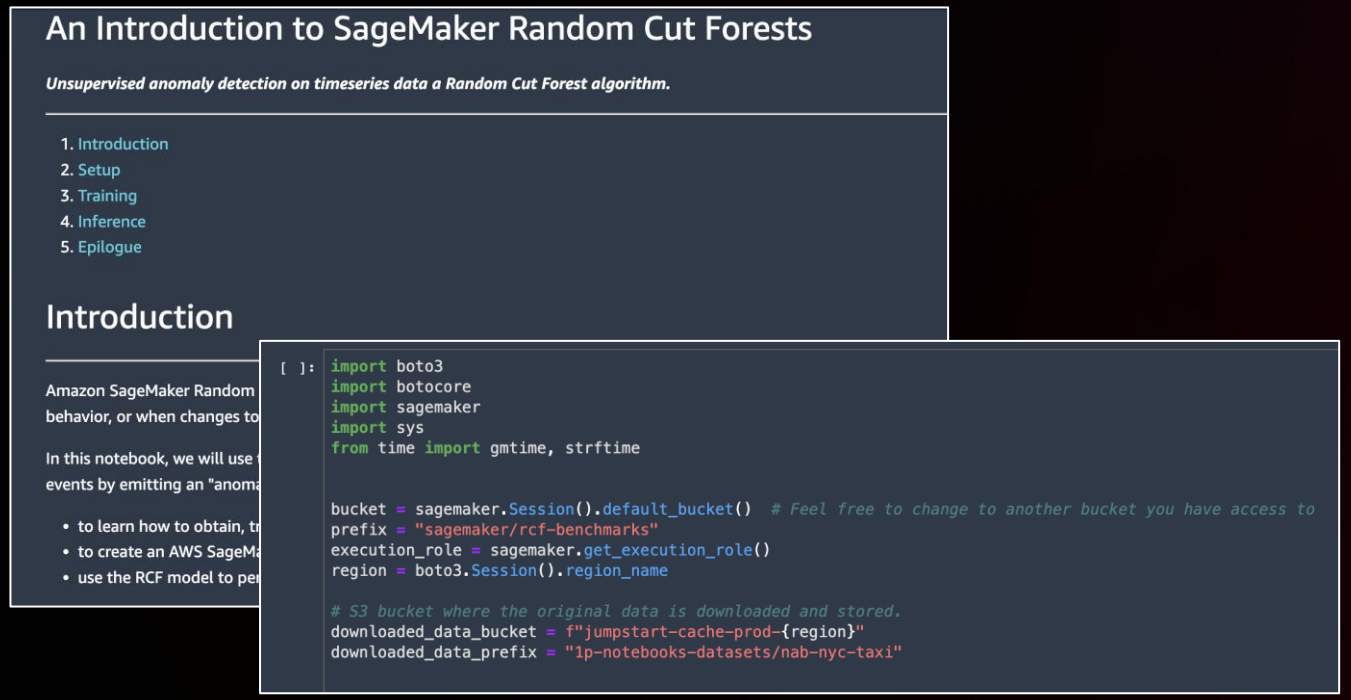
Template categories

- Demand forecasting
- Credit rating prediction
- Fraud detection
- Computer vision
- Extract and analyze documents
- Predictive maintenance
- Churn prediction
- Personalized recommendations
- Reinforcement learning
- Healthcare and life sciences
- Financial pricing

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SageMaker JumpStart – Example notebooks

SAMPLE NOTEBOOKS WITH BUILT-IN ALGORITHMS, PUBLICLY AVAILABLE MODELS, AND POPULAR ML SOLUTIONS



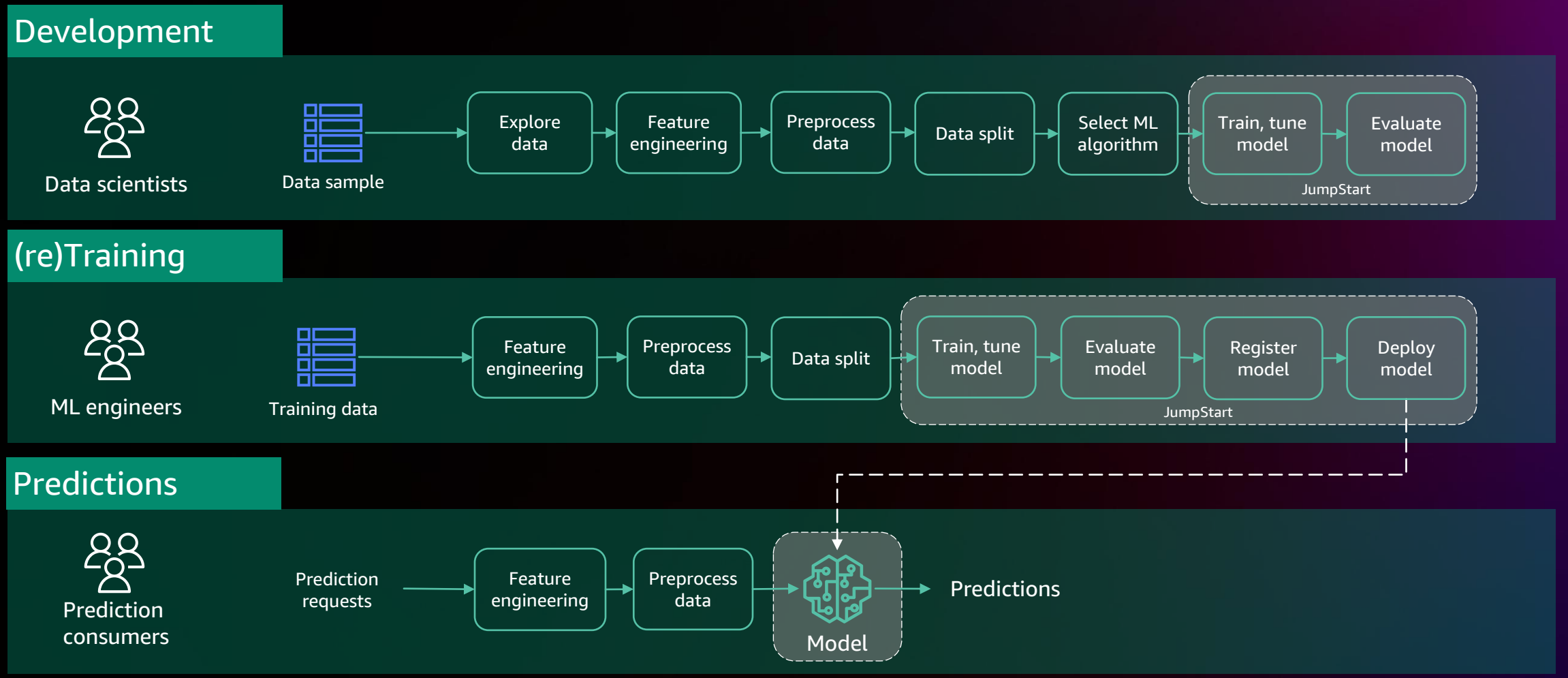
Notebook example: Introduction to SageMaker built-in algorithms

Model categories

- Foundation/generative
- Tabular
- Computer vision
- Natural language processing
- Audio embedding

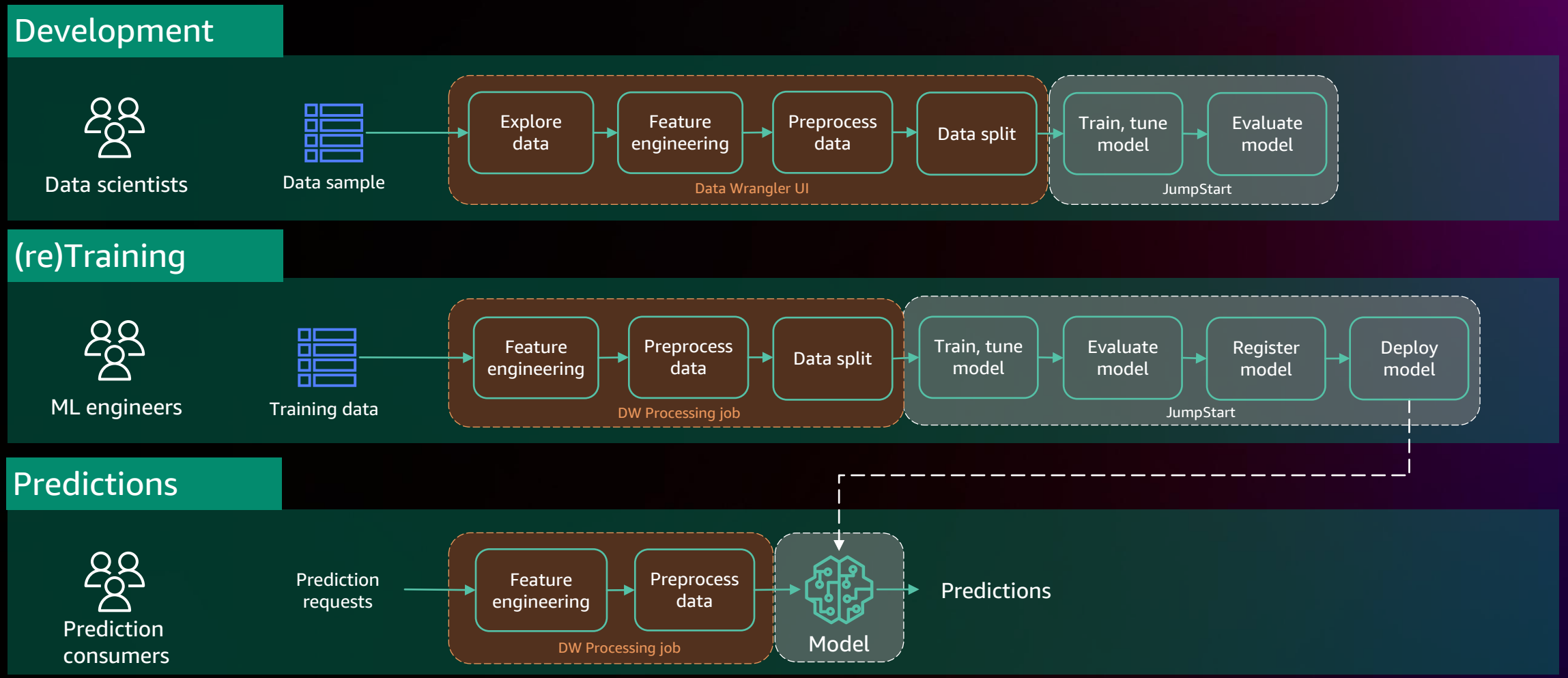
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SageMaker JumpStart in ML workflow



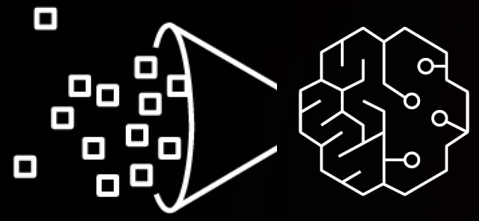
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SageMaker Data Wrangler and SageMaker JumpStart in your ML workflow



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Foundation models now available on SageMaker Jumpstart



Pretrained on large sets of data with billions of parameters



Applicable to many use cases like text summarization, question answering, digital art creation, etc.



Expensive to train from scratch

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Partner and publicly available models



- ✓ Data stays in your account – model inputs, model outputs
- ✓ Fully integrated with the suite of SageMaker features

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AI21 Labs

Transform the way people read and write with Generative AI

Applications



> 5M users

Developer platform

AI21 studio

> 20K developers

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AI21labs

AI21 Studio and Jurassic-1 language models

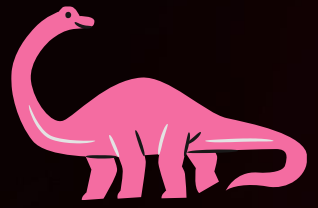
AI21 Studio features:

- Developer-friendly API
- No-code playground environment
- 3-click custom model training
- Pay-as-you-go pricing



Jurassic-1 Large

(7.5B parameters)



Jurassic-1 Grande

(17B parameters)



Jurassic-1 Jumbo

(178B parameters)

aws

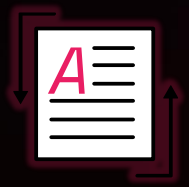
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AI21labs

New valuable applications are unlocked every day


Content generation


Summarization


Paraphrasing


Code generation


Repurposing


Chat

citibank

PUBLICIS GROUPE

UBISOFT

Capgemini

glassdoor

Quora

latitude

Tweet Hunter

jenni

Queue

Center for Research on Foundation Models

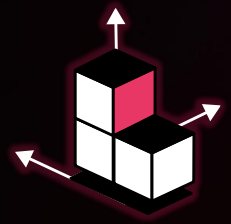
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AI21labs

What's so exciting about Jurassic-1 models?



Striking the right price/ performance balance



Optimal width/depth ratio transformer network architecture (Levine et al., 2020)



Designed for safety and robustness (Liang et al., 2022)



Trained with 250K items dictionary



Dedicated task-specific endpoints

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

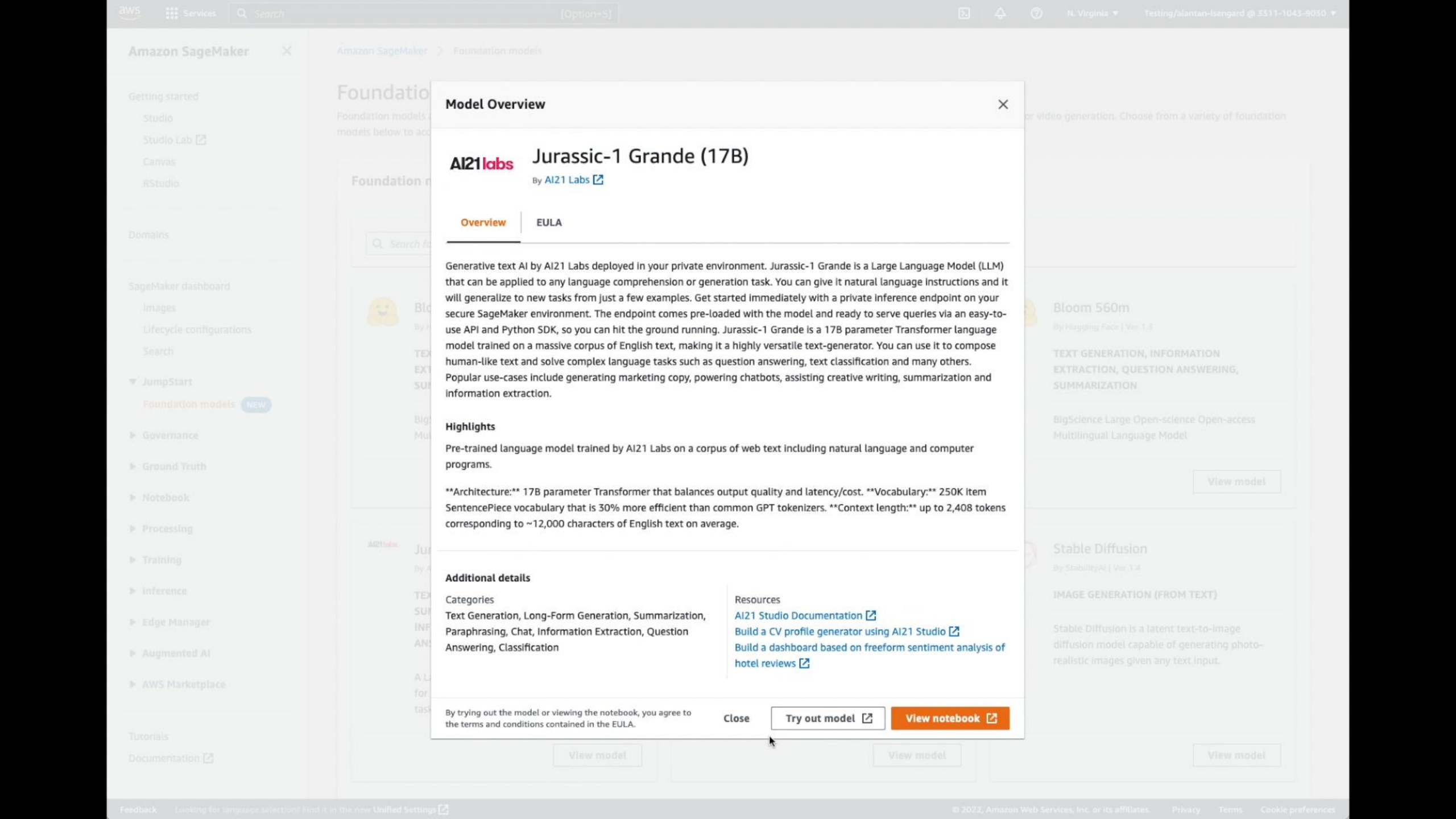
AI21labs

SageMaker + Jurassic-1 Grande

aws

AI21labs

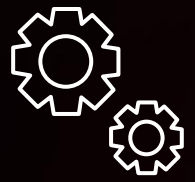
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.


Model Overview
Jurassic-1 Grande (17B)
By AI21 Labs
Overview
EULA
Generative text AI by AI21 Labs deployed in your private environment. Jurassic-1 Grande is a Large Language Model (LLM) that can be applied to any language comprehension or generation task. You can give it natural language instructions and it will generalize to new tasks from just a few examples. Get started immediately with a private inference endpoint on your secure SageMaker environment. The endpoint comes pre-loaded with the model and ready to serve queries via an easy-to-use API and Python SDK, so you can hit the ground running. Jurassic-1 Grande is a 17B parameter Transformer language model trained on a massive corpus of English text, making it a highly versatile text-generator. You can use it to compose human-like text and solve complex language tasks such as question answering, text classification and many others. Popular use-cases include generating marketing copy, powering chatbots, assisting creative writing, summarization and information extraction.
Highlights
Pre-trained language model trained by AI21 Labs on a corpus of web text including natural language and computer programs.
Architecture: 17B parameter Transformer that balances output quality and latency/cost. **Vocabulary:** 250K item SentencePiece vocabulary that is 30% more efficient than common GPT tokenizers. **Context length:** up to 2,408 tokens corresponding to ~12,000 characters of English text on average.
Additional details
Categories
Text Generation, Long-Form Generation, Summarization, Paraphrasing, Chat, Information Extraction, Question Answering, Classification
Resources
AI21 Studio Documentation
Build a CV profile generator using AI21 Studio
Build a dashboard based on freeform sentiment analysis of hotel reviews
By trying out the model or viewing the notebook, you agree to the terms and conditions contained in the EULA.
Close
Try out model
View notebook

Low-code machine learning



Faster ML workflows



Open and customizable solutions



Appropriate for both new and advanced ML practitioners

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

No-code | low-code re:Invent sessions

- AIM337 Build better ML models for business decisions using Amazon SageMaker Canvas
- TNC107 Use Amazon SageMaker Canvas to make your first ML model
- FSI305 Use AWS no-code services to build a credit score model
- TNC107 Use Amazon SageMaker Canvas to make your first ML model
- AIM201 Accelerate your ML lifecycle with Amazon SageMaker low-code tools
- AIM314 Accelerate your ML journey with Amazon SageMaker low-code tools
- AIM332 Accelerate data preparation with Amazon SageMaker Data Wrangler
- AIM326 Prepare data and model features for ML with ease, speed, and accuracy

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Claire O'Brien Rajkumar
claireob@amazon.com

Vadim Omeltchenko
vadimo@amazon.com

Ori Goshen



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.