Amazon SageMaker

AutoML Demo

Nicholaus Lawson
Solution Architect

The AWS ML stack

Broadest and most complete set of machine learning capabilities

AI SERVICES

Vision
Rekognition

Chatbots
Lex

Business tools
Personalize
Forecast
Fraud Detector
Lookout for Metrics

Search
Kendra

Healthcare
HealthLake
Comprehend Medical
Transcribe Medical

Speech
Polly
Transcribe

Text
Comprehend
Translate
Textract

Contact centers
Contact Lens
Connect
Voice ID

Code + DevOps
CodeGuru
DevOps Guru

Industrial
Panorama Appliance and SDK, Monitron, Lookout for Equipment, Lookout for Vision

AMAZON SAGEMAKER

Label data

SAGEMAKER STUDIO IDE

Data collection prep | Store features | Detect bias and explain predictions | Visualize in notebooks | Pick algorithm | Train models faster | Tune parameters | Deploy in production | Manage & monitor

CI/CD

Manage edge devices

ML FRAMEWORKS & INFRASTRUCTURE

TensorFlow, PyTorch, Apache MXNet | Deep learning AMIs & containers | GPUs | Inferentia | Elastic inference | FPGA

aws

Quick intro to Amazon SageMaker

aws machine learning

Integrated Workbench

IDE designed specifically for ML, data preparation, experiment management, and workflows

Managed Infrastructure

Designed for ultra low latency and high throughput, automatic scaling, and distributed training

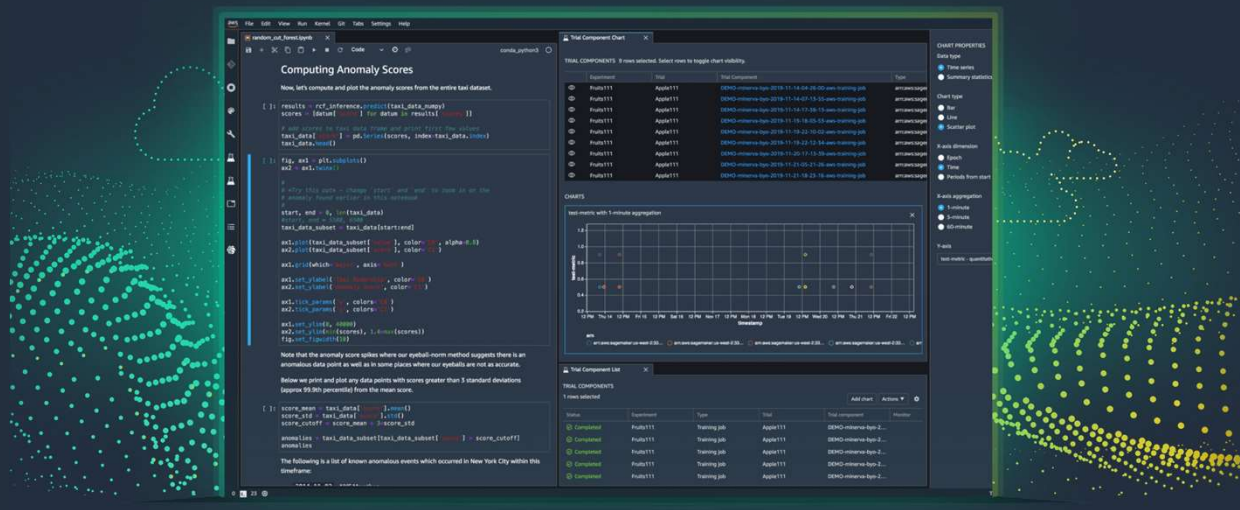
Managed Tooling

Purpose-built from the ground up to work together including auto ML, collaboration, debugger, profiler, bias analyzer, and explainability

https://aws.amazon.com/sagemaker

Amazon SageMaker

Most complete, end-to-end ML service



aws machine learning

AI/ML life cycle

Amazon SageMaker

PREPARE

SageMaker Ground Truth
Label training data for machine learning

SageMaker Data Wrangler
Aggregate and prepare data for machine learning

SageMaker Processing
Built-in Python, BYO R/Spark

SageMaker Feature Store
Store, update, retrieve, and share features

SageMaker Clarify
Detect bias and understand model predictions

BUILD

SageMaker Studio Notebooks
Jupyter notebooks with elastic compute and sharing

Built-in and Bring your-own Algorithms
Dozens of optimized algorithms or bring your own

Local Mode
Test and prototype on your local machine

SageMaker Autopilot
Automatically create machine learning models with full visibility

SageMaker JumpStart
Pre-built solutions for common use cases

TRAIN & TUNE

Managed Training
Distributed infrastructure management

SageMaker Experiments
Capture, organize, and compare every step

Automatic Model Tuning
Hyperparameter optimization

Distributed Training
Training for large datasets and models

SageMaker Debugger
Debug and profile training runs

Managed Spot Training
Reduce training cost by 90%

DEPLOY & MANAGE

Managed Deployment
Fully managed, ultra low latency, high throughput

Kubernetes & Kubeflow Integration
Simplify Kubernetes-based machine learning

Multi-Model Endpoints
Reduce cost by hosting multiple models per instance

SageMaker Model Monitor
Maintain accuracy of deployed models

SageMaker Edge Manager
Manage and monitor models on edge devices

SageMaker Pipelines
Workflow orchestration and automation

SageMaker Studio
Integrated development environment (IDE) for ML

aws machine learning

Quick intro to Amazon SageMaker Studio

aws machine learning

Amazon SageMaker Studio

Fully integrated development environment (IDE) for Machine Learning

Collaboration at scale
Without tracking code dependencies

Easy experiment management
Organize, track, and compare thousands of experiments

Automatic model generation
Full visibility and control without writing code

Higher quality ML models
Automatically debug errors, monitor models, and maintain high quality

Increased productivity
Code, build, train, deploy, and monitor in a unified visual interface

aws machine learning

Each algorithm solves a type of prediction problem

Classification
- Linear Learner *
- XGBoost *
- KNN
- Factorization Machines

Working with Text
- Blazing Text
  - Supervised
  - Unsupervised *

Sequence Translation
- Seq2Seq *

Computer Vision
- Image Classification <>
- Object Detection <>
- Semantic Segmentation

Recommendation
- Factorization Machines * (+ KNN)

Anomaly Detection
- Random Cut Forests *
- IP Insights *

Regression
- Linear Learner

Topic Modeling
- LDA
- NTM

Forecasting
- DeepAR *

Clustering
- Kmeans *
- KNN

Feature Reduction
- PCA
- Object2Vec

* = distributed training <> = incremental training

aws machine learning

AutoML and SageMaker

aws machine learning

Amazon SageMaker Autopilot

Provide data

Data in tabular form

Specify column to predict

Support for regression and classification

Create model

Feature generation, algorithm selection, and parameter tuning

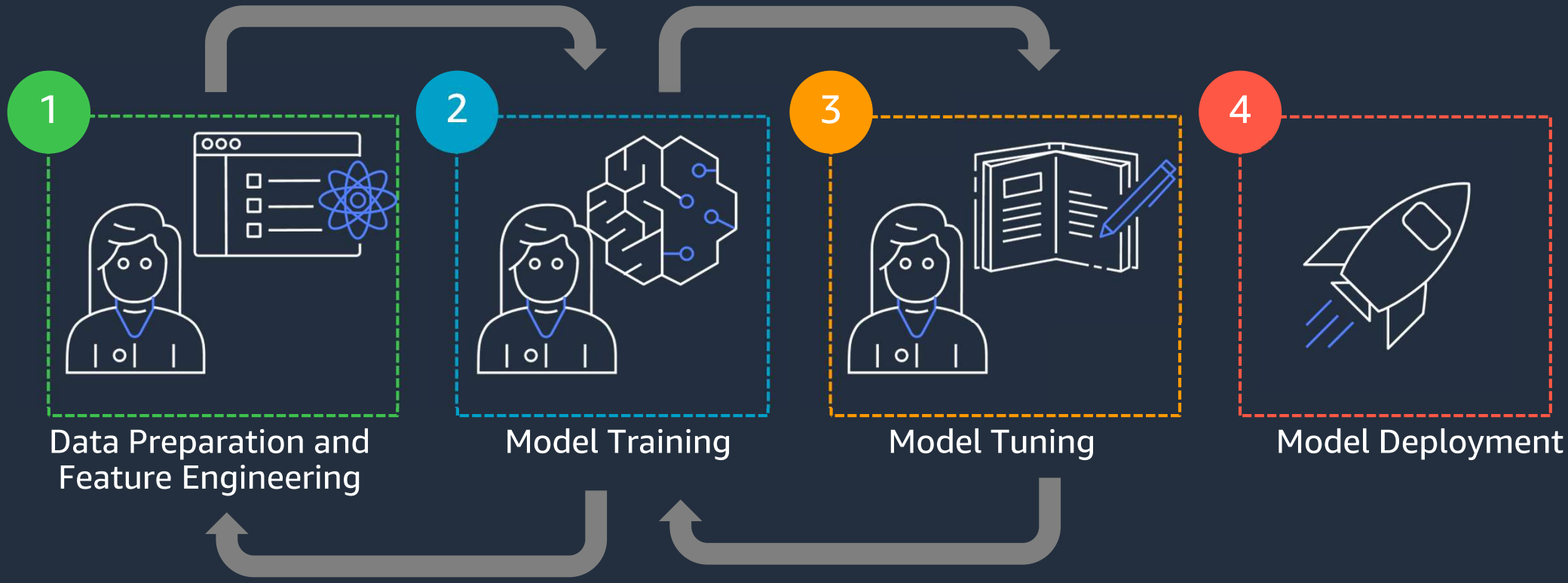
Track experiment

Automatically tracked as an experiment

Transparent

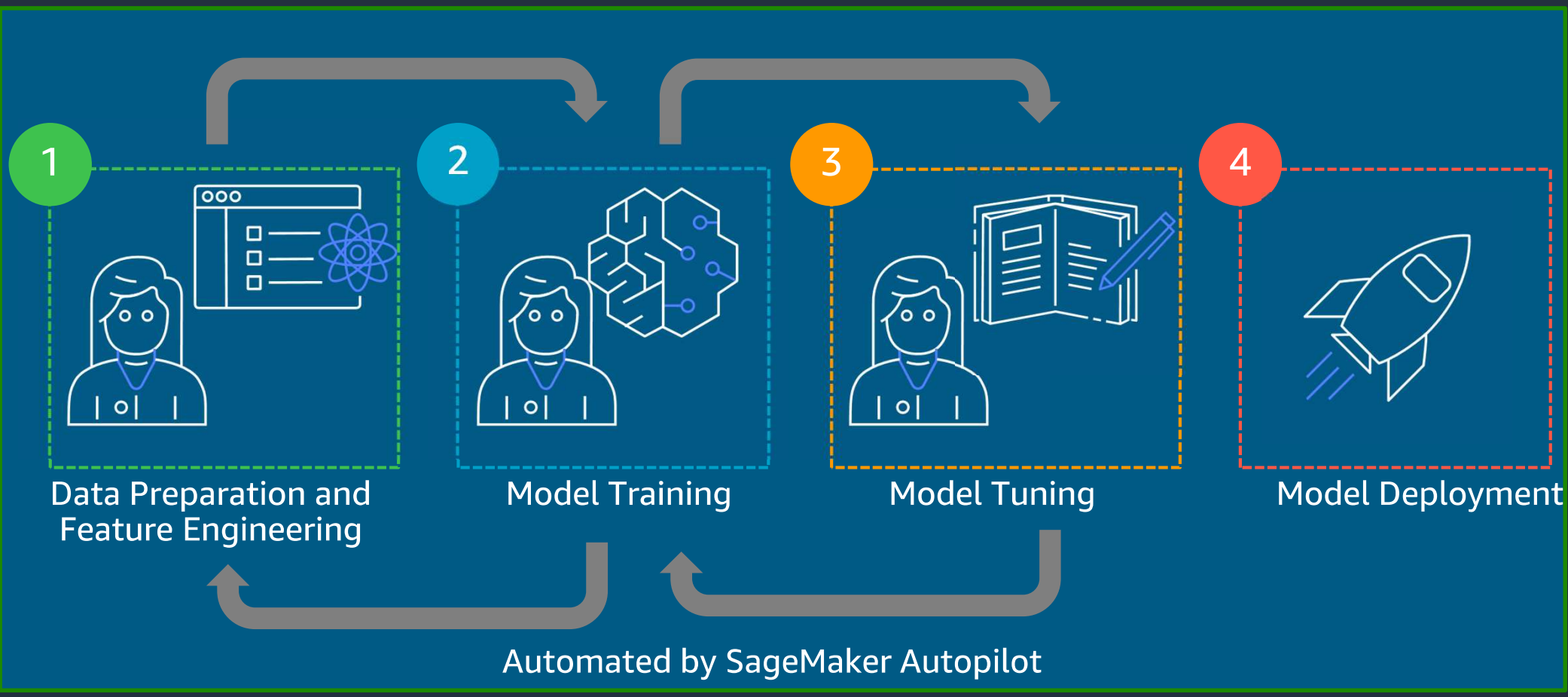
Get notebook with source code

Why building ML models is hard...



aws machine learning

Why building ML models is hard…



aws machine learning

AutoML with Amazon SageMaker Autopilot

SageMaker Autopilot covers all steps

- *Problem identification*: looking at the data set, what class of problem are we trying to solve?
- *Algorithm selection*: which algorithm is best suited to solve the problem?
- *Data preprocessing*: how should data be prepared for best results?
- *Hyperparameter tuning*: what is the optimal set of training parameters?

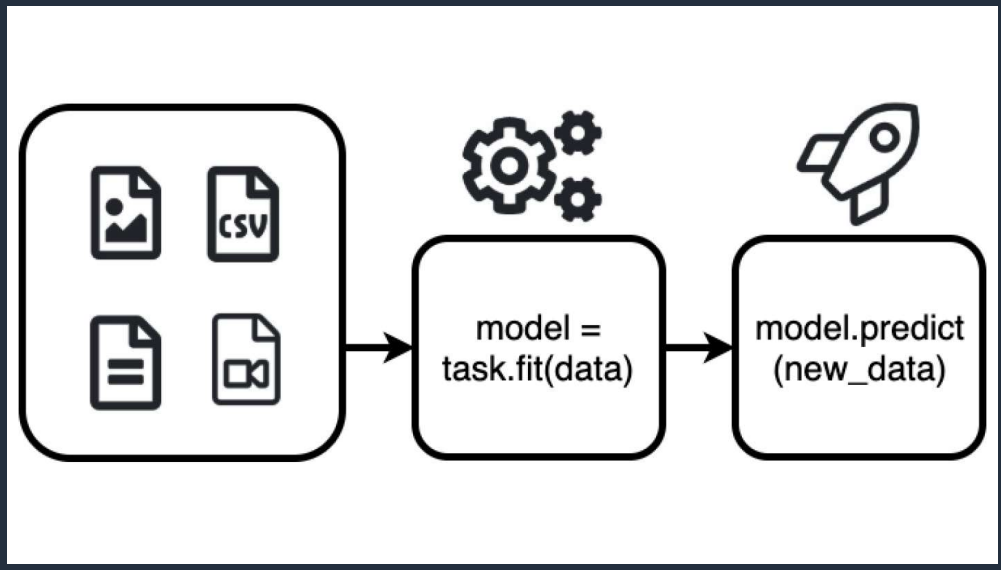
Supported algorithms at launch:

Linear Learner, XGBoost

AutoGluon

AutoGluon enables easy-to-use and easy-to-extend AutoML with a focus on automated stack ensembling, deep learning, and real-world applications spanning image, text, and tabular data.

- Quickly prototype deep learning and classical ML solutions for your raw data with a few lines of code.
- Automatically utilize state-of-the-art techniques (where appropriate) without expert knowledge.
- Leverage automatic hyperparameter tuning, model selection/ensembling, architecture search, and data processing.
- Easily improve/tune your bespoke models and data pipelines, or customize *AutoGluon* for your use-case.



aws machine learning

AutoGluon

Tabular Prediction

How to predict variables in tabular datasets.

Image Prediction

How to classify images into various categories.

Object Detection

How to detect objects and their location in images.

Text Prediction

How to solve NLP problems via supervised learning from raw text.

Multimodal Prediction

How to solve problems that contain Image, Text, and Tabular features at the same time.

Time Series Forecasting

How to train time series models for forecasting.

Getting started with SageMaker

aws machine learning

SageMaker JumpStart

Easily and quickly bring machine learning applications to market

15+ pre-built solutions for common ML use cases

Solutions can be used out-of-the-box or can be customized for a specific business problem

Accelerate time to deploy over 150 open source models

Use one-click deployable ML models and algorithms from popular model zoos

Get started with just a few clicks

Easily bring ML applications to market using pre-built solutions, ML models, and algorithms from popular model zoos, and getting started content

aws machine learning

SageMaker JumpStart

Use cases

Autonomous driving

Predictive maintenance

Churn prediction

Computer vision

Personalized recommendations

Fraud detection

Extract data from documents

Demand forecasting

Credit risk prediction

aws machine learning

Amazon SageMaker JumpStart pre-built solutions

Predictive Maintenance

Predictive maintenancefor manufacturing >

Predictive maintenance for vehicle fleets >

Demand Forecasting

Demand forecasting with deep learning >

Fraud Detection

Detect malicious users and transactions >

Fraud detection in financial transactions using deep graph library >

Credit Risk Prediction

Explain credit decisions >

Extract & Analyze Data from Documents

Document summarization, entity, and relationship extraction >

Handwriting recognition >

Filling in missing values in tabular records >

Differential privacy for sentiment classification >

Computer Vision

Product defect detection in images >

Autonomous Driving

Visual perception with active learning >

Personalized Recommendations

Entity resolution in identity graphs >

Purchase modeling >

Churn Prediction

Churn prediction with text >

Learn more about solutions:
https://aws.amazon.com/sagemaker/getting-started/

aws machine learning

DEMO!

Bear with me:
-This is live, we have no animatronics here
-Depends on the internet to connect to notebook
-Depends on guest wifi

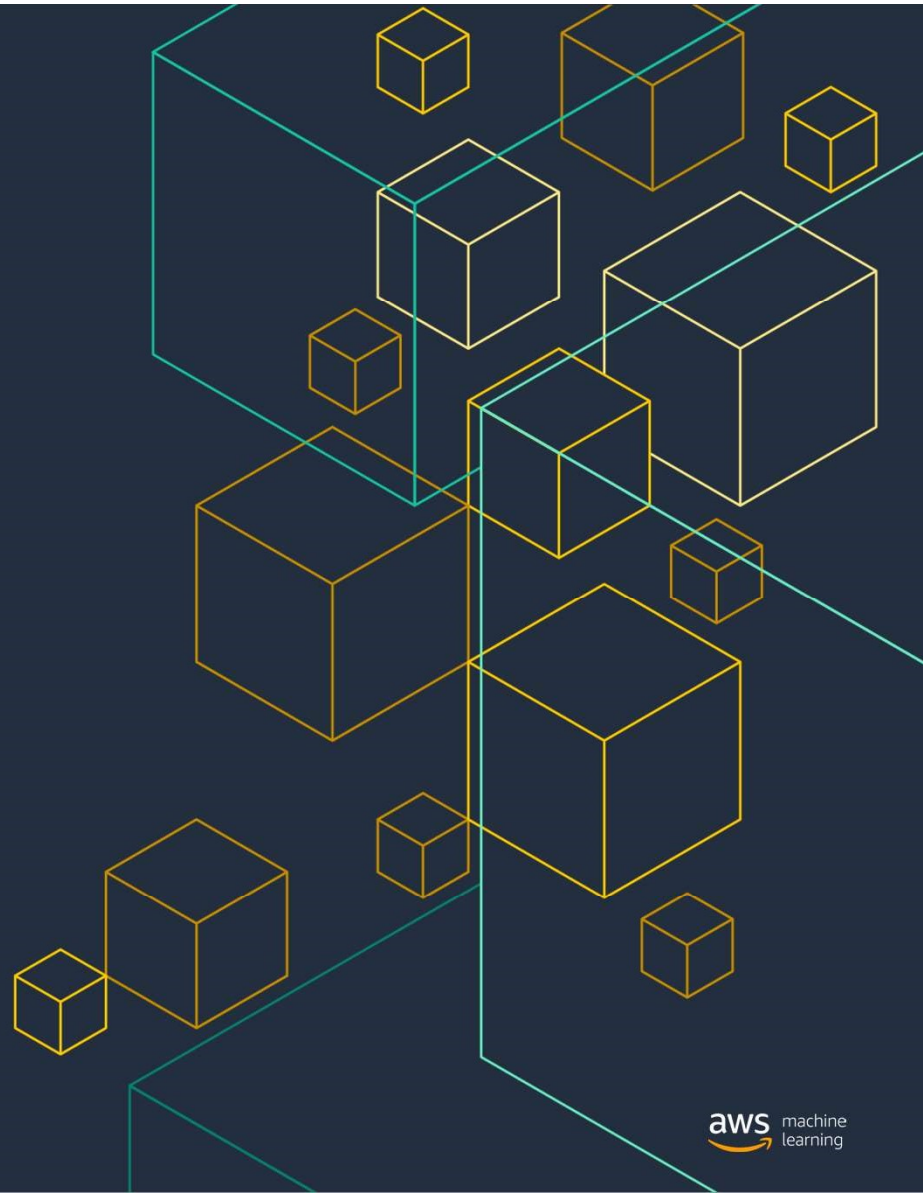
What we are going to cover
1\ Data Wrangler
2\ Hi scoring Kaggle notebook
3\ My significantly lower scoring notebook
4\ AutoGluon notebook
5\ SageMaker AutoPilot result

Code Repo

https://github.com/ccrngd1/AutoAIDemo

aws machine learning

Thank you



aws machine learning

References

https://aws.amazon.com/free

https://aws.amazon.com/sagemaker

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

https://aws.amazon.com/sagemaker/autopilot/

https://auto.gluon.ai/stable/index.html#

https://github.com/awslabs/autogluon/

https://github.com/ccrngd1/AutoAIDemo