Accelerate innovation using machine learning

Easily build, train, and deploy machine learning (ML) models

Yan Marim

Tens of thousands of customers use Amazon SageMaker

Amazon SageMaker helps organizations harness ML

Overcoming the barriers to ML

Not enough ML builders →

No-code ML tools

Make ML predictions regardless of ML experience

Access, process, and label massive volumes of data for ML →

Purpose-built data preparation tools

Access, process, and label data for ML

Disparate data science tools →

Integrated ML tools in a single interface

Build, train, and deploy models using IDEs

Tedious, manual ML operations →

Built-in MLOps

Automate and standardize MLOps practices

Challenging to govern ML projects efficiently →

Out-of-box ML governance tools

Simplify access control and enhance transparency across ML lifecycle

No-code ML tools

ML requires a lot more builders

2x

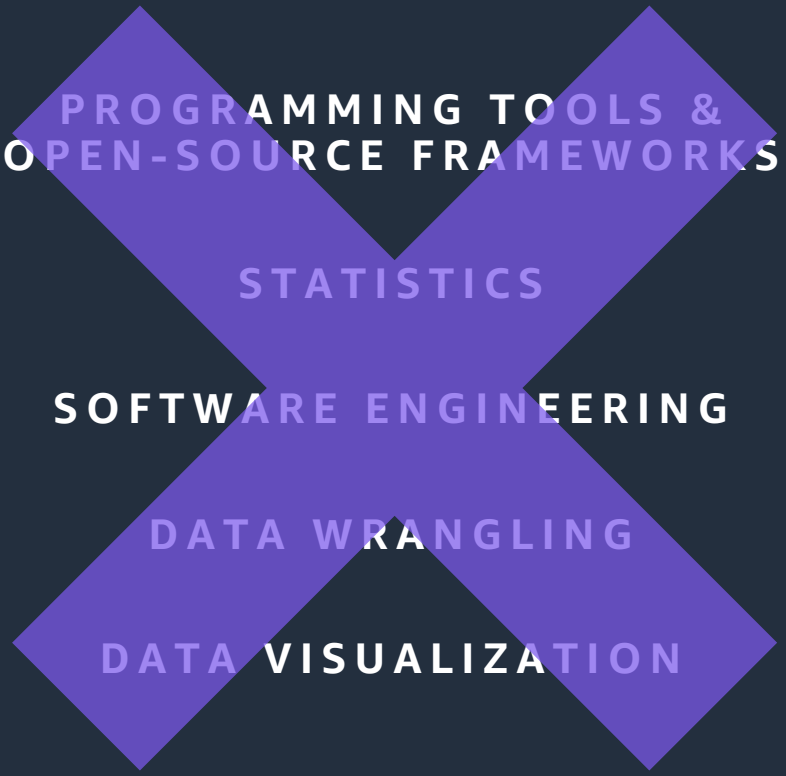
the growth of any other emerging job role

+74%

annual growth in the past 4 years

Source: LinkedIn Emerging Jobs Report 2020

Business analysts require a different set of tools to use ML



Amazon SageMaker Canvas

BUILD ML MODELS AND GENERATE ACCURATE PREDICTIONS— NO CODE REQUIRED

Quickly access and prepare data for Machine Learning

Built-in AutoML to build models and generate accurate predictions

Share models and datasets with data scientists so they can validate and further refine ML models

Usage-based pricing to avoid licensing fees and reduce TCO

Purpose-built data preparation tools

Structured

Unstructured

Amazon SageMaker Data Wrangler

EXPLORE, PREPARE, AND PROCESS DATA WITH LITTLE TO NO CODE

Import data from multiple sources

Get insights on data and data quality

Visually explore, analyze, and prepare data

Quickly perform feature engineering

Automate ML data preparation workflows

Amazon SageMaker Feature Store

SECURELY STORE, DISCOVER, AND SHARE FEATURES FOR ML

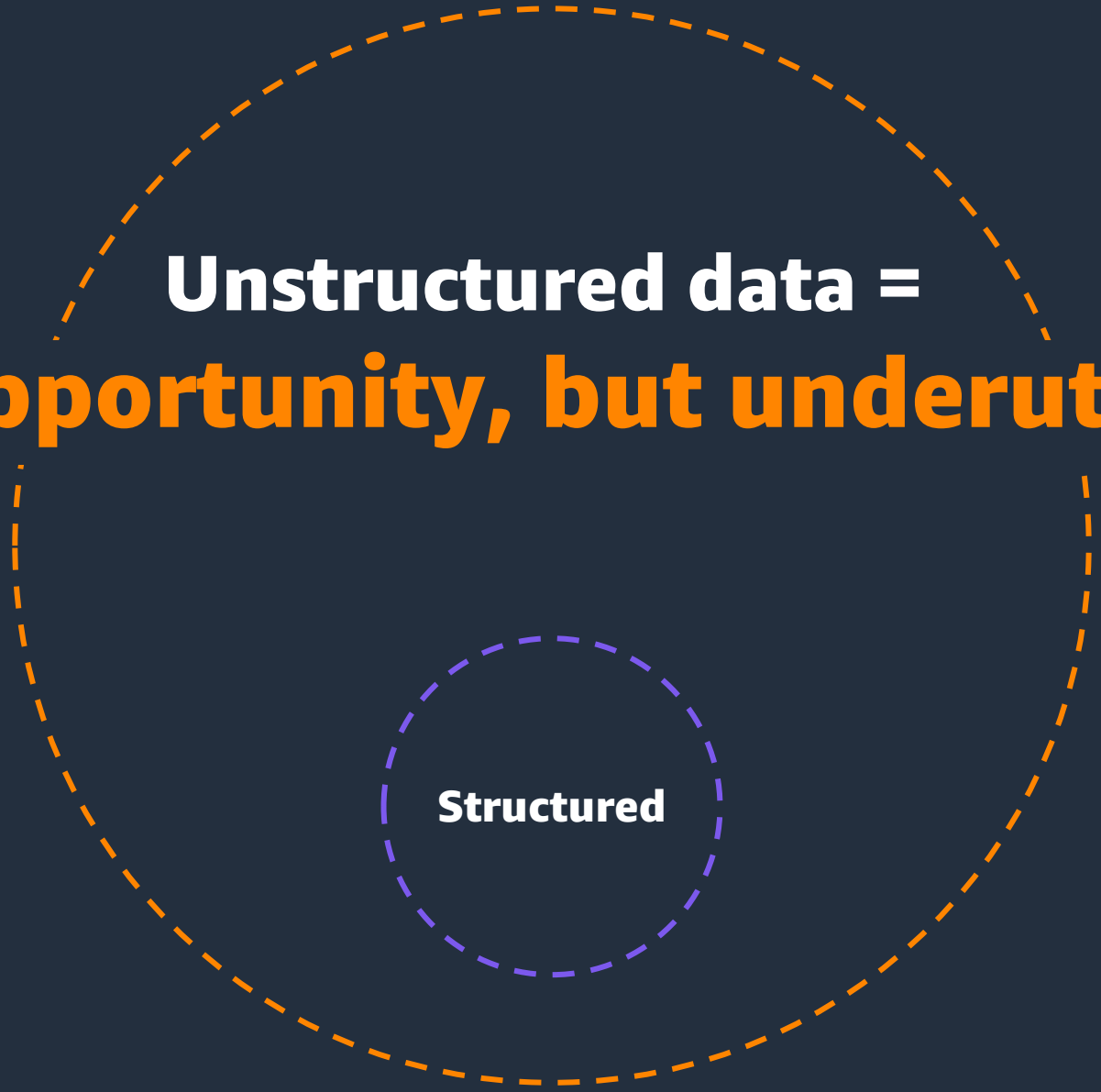
Online and off-line

Millisecond latency

Consistent features

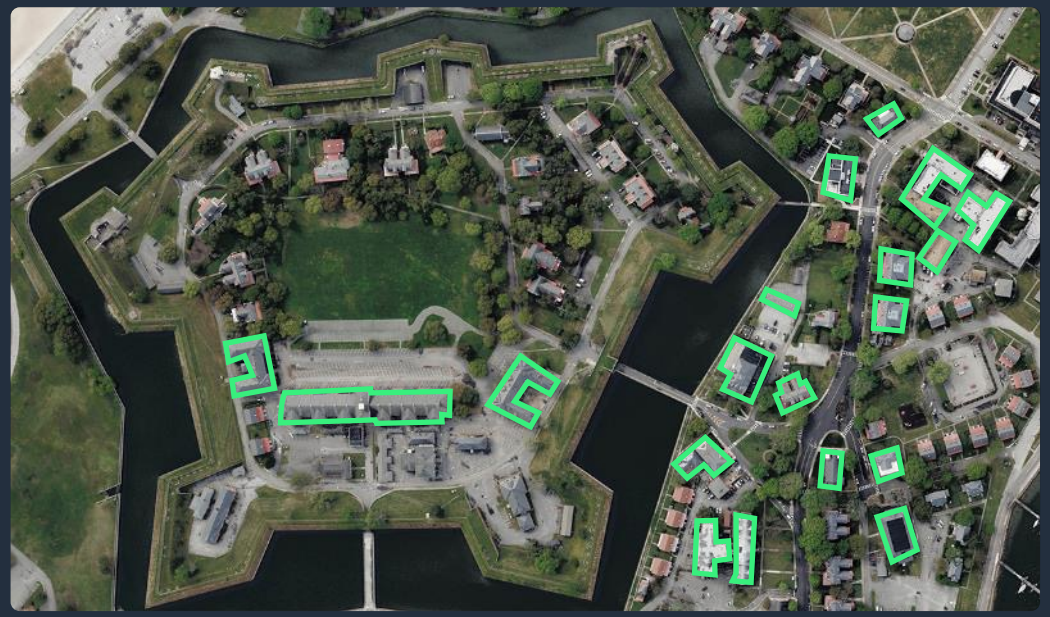
Visual search

Sharing and collaboration



What is Geospatial Data?

Aerial and satellite imagery

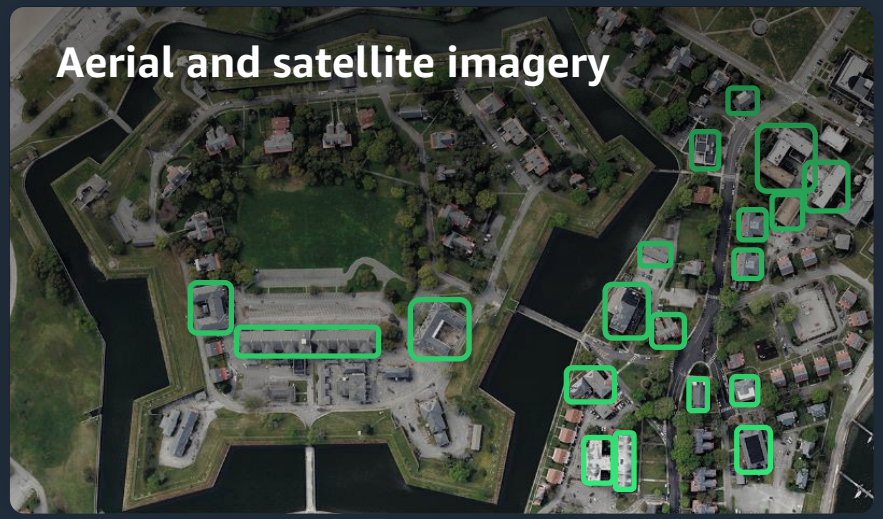


Mapping data

Road mask
(color as speed)

Geospatial ML on Amazon SageMaker

BUILD, TRAIN, AND DEPLOY ML MODELS USING GEOSPATIAL DATA



- Access readily available geospatial data source
- Efficiently process or enrich large-scale geospatial datasets
- Accelerate model building by using built-in pre-trained ML models
- Analyze geospatial data and explore model predictions on an interactive map

Amazon SageMaker Ground Truth Plus

DELIVER HIGH-QUALITY TRAINING DATASETS FAST AND REDUCE DATA LABELING COSTS

Increase data quality through ML-powered data labeling

Access expert data labelers

Reduce data labeling costs with assistive labeling features

Improve operational efficiency by reviewing project metrics

Make data labeling accessible to data operations and program managers

Integrated ML tools in a single interface

Amazon SageMaker brings tools for every step of the ML lifecycle under one unified visual user interface

Build ML models

FULLY MANAGED SHAREABLE NOTEBOOKS ON AMAZON EC2

Fully managed, sharable Jupyter Notebooks
Run notebooks on elastic compute resources

Built-in algorithms
15 built-in algorithms available in prebuilt container images

Prebuilt solutions and open-source models
Over 300 popular open-source models

AutoML
Automatically create ML models with full visibility

Support for major frameworks and toolkits
Optimized for popular deep learning (DL) frameworks such as TensorFlow, PyTorch, Apache MXNet, and Hugging Face

Train & tune ML models

FAST AND COST-EFFECTIVE ML MODEL TRAINING

Many ways to train models
Local mode, script, BYOC, deep learning containers

Distributed training libraries
With data and model parallel libraries complete distributed training 40% faster

Debug and profile training runs
Use real-time metrics to correct performance problems

Automatic model tuning
Find the best version of a model with automatic hyperparameter optimization

Training compiler
Accelerate training times by up to 50% through more efficient use of GPUs

Experiment management
Track ML model iterations easily by automatically capturing input parameters, configurations, and results

Managed spot training
Reduce costs by up to 90% by automatically running training jobs when compute becomes available

Automate training workflows
Create a repeatable process to orchestrate the steps for rapid experimentation and model retraining

Deploy ML models

FULLY MANAGED DEPLOYMENT FOR INFERENCE AT SCALE

Wide selection of infrastructure
70+ instance types with varying levels of compute and memory to meet the needs of every use case

Deploy models in production for inference for any use case
From low latency and high throughput to long-running inference for use cases

Cost-effective deployment
Multi-model/multi-container endpoints, serverless inference, and elastic scaling

Shadow testing
Validate the performance of new ML models against production models to prevent costly outages

Automatic deployment recommendations
Optimal instance type/count and container parameters, and fully managed load testing

Built-in integration for MLOps
ML workflows, CI/CD, lineage tracking, and catalog

SageMaker JumpStart

ML HUB WITH BUILT-IN ALGORITHMS, FOUNDATION MODELS, AND PREBUILT ML SOLUTIONS

20+ prebuilt solutions for common ML use cases
Leverage solutions out-of-the-box or customize for a specific business problem

Accelerate time to deploy over 300 open-source models
Use one-click deployable ML models and algorithms from popular model hubs

Get started with just a few clicks
Bring ML applications to market using built-in algorithms with pretrained models from model hubs, pretrained foundation models, and prebuilt solutions to solve common use cases

Share ML artifacts across your organization
Share models and notebooks with others within your organization, and allow them to train with their own data or deploy as-is for inference

Support the responsible use of ML throughout the model lifecycle

Onboard

Setup ML users with custom permissions

Build

Perform bias analysis during exploratory data analysis

Document model information such as intended use and risk ratings

Train

Conduct bias and explainability analysis after training

Capture model training and evaluation observations

Deploy

Explain individual inferences from models in production

Monitor

Validate bias and relative feature importance over time

Audit performance and lineage of all your models, in one place

Governance

Purpose-built governance tools to help implement ML responsibly

Amazon SageMaker Clarify

DETECT BIAS IN ML MODELS AND UNDERSTAND MODEL PREDICTIONS

Identify imbalances in data
Detect bias during data preparation

Check your trained model for bias
Evaluate the degree to which various types of bias are present in your model

Explain overall model behavior
Understand the relative importance of each feature to your model's behavior

Explain individual predictions
Understand the relative importance of each feature for individual inferences

Detect drift in bias and model behavior over time
Provide alerts and detect drift over time due to changing real-world conditions

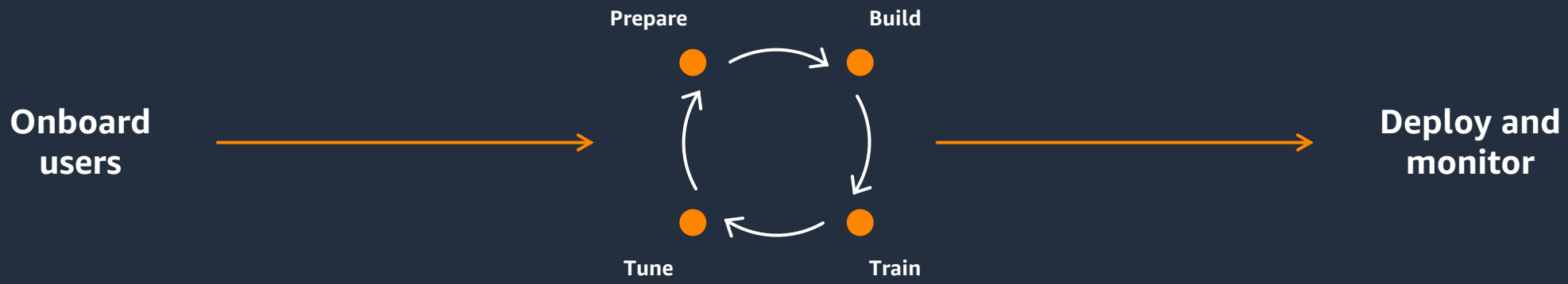
Generated automated reports
Produce reports on bias and explanations to support internal presentations

Key challenges for ML governance

Hand-crafting custom policies is time consuming

Using disparate, manual tools to capture and share model information can be error-prone

Custom instrumentation to get visibility into model performance is expensive



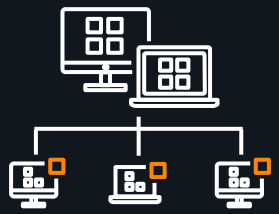
ML governance tools

Amazon SageMaker Role Manager

Define custom permissions for SageMaker users in minutes

Amazon SageMaker Model Cards

Create a single source of truth for model information



Amazon SageMaker Model Dashboard

Audit all your models, lineage and performance in one place

Built-in MLOps capabilities

Operational challenges with managing the ML lifecycle

Manual iterative processes slow down ML innovation

Difficult to scale the number of models in production

CI/CD for ML requires writing custom code

Compliance requirements are difficult to meet

Amazon SageMaker MLOps

STREAMLINE THE ML LIFECYCLE

Automate ML workflows to scale model development

Build CI/CD pipelines for ML to accelerate model deployment

Catalog model versions, metadata, metrics, and approvals for traceability and reusability

Track lineage for troubleshooting and compliance

Maintain accuracy of predictions after models are deployed

Amazon SageMaker key benefits

MOST COMPLETE
END-TO-END ML SERVICE

Democratize ML innovation
Empower more groups of people, including business analysts

Accelerate the ML lifecycle
Reduce training time from hours to minutes

Prepare data at scale
Access, process, and label structured and unstructured data

Streamline ML processes
Automate and standardize MLOps practices

Improve ML governance
Simplify access control and enhance transparency

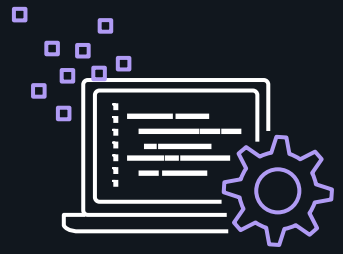
Amazon SageMaker

GETTING STARTED

Business analysts

Get started with Amazon SageMaker Canvas

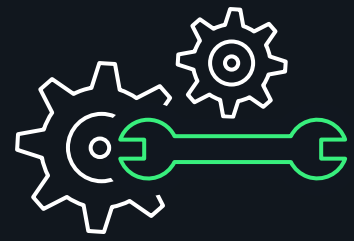
No-code ML



Data scientists

Onboard through Amazon SageMaker Studio

Single IDE for full-code ML



ML engineers

Implement MLOps with Amazon SageMaker Pipelines

Workflow automation and CI/CD

Amazon SageMaker feature tour

PREPARE DATA AND BUILD, TRAIN, AND DEPLOY ML MODEL FOR ANY USE CASE

PREPARE

Geospatial
Visualize geospatial data

Ground Truth
Create high quality datasets for ML

Data Wrangler
Aggregate and prepare data for ML

Processing
Built-in Python, BYO R/Spark

Feature Store
Store, catalog, search, and reuse features

Clarify
Detect bias and understand model predictions

BUILD

Studio Notebooks & Notebook Instances
Fully managed Jupyter notebooks with elastic compute

Studio Lab
Free ML development environment

Built-in Algorithms
Integrated tabular, NLP, and vision algorithms

JumpStart
UI based discovery, training, and deployment of models, solutions, and examples

Autopilot
Automatically create ML models with full visibility

Bring Your Own
Bring your own container and algorithms

Local Mode
Test and prototype on your local machine

TRAIN & TUNE

Fully Managed Training
Broad hardware options, easy to setup and scale

Distributed Training Libraries
High performance training for large datasets and models

Training Compiler
Faster deep learning model training

Automatic Model Tuning
Hyperparameter optimization

Managed Spot Training
Reduce training cost by up to 90%

Debugger and Profiler
Debug and profile training runs

Experiments
Track, visualize, and share model artifacts across teams

Customization Support
Integrate with popular open source frameworks and libraries

DEPLOY & MANAGE

Fully Managed Deployment
Ultra low latency, high throughput inference

Real-Time Inference
For steady traffic patterns

Serverless Inference
For intermittent traffic patterns

Asynchronous Inference
For large payloads or long processing times

Batch Transform
For offline inference on batches of large datasets

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

Multi-Container Endpoints
Reduce cost by hosting multiple containers per instance

Shadow Testing
Validate model performance in production

Inference Recommender
Automatically select compute instance and configuration

Model Monitor
Maintain accuracy of deployed models

Kubernetes Operators & Components
Manage and monitor models on edge devices

Edge Manager
Manage and monitor models on edge devices

MLOps: Pipelines | Projects | Model Registry
Workflow automation, CI/CD for ML, central model catalog

Canvas
Generate accurate machine learning predictions—no code required

Studio | RStudio
Integrated development environment (IDE) for ML

Governance
Model Cards | Dashboard | Permissions

Thank you!

Yan Marim
yamarim@amazon.com