Build Machine Learning Models with Amazon SageMaker

Julien Simon
Global Evangelist, AI & Machine Learning
@julsimon

aws SUMMIT

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The Amazon ML Stack: Broadest & Deepest Set of Capabilities

AI SERVICES

Vision

REKOGNITION IMAGE

REKOGNITION VIDEO

TEXTRACT

Speech

POLLY

TRANSCRIBE

Language

TRANSLATE

COMPREHEND

COMPREHEND MEDICAL

Chatbots

LEX

Forecasting

FORECAST

Recommendations

PERSONALIZE

ML SERVICES

AMAZON SAGEMAKER

BUILD

Pre-built algorithms & notebooks

Data labeling (GROUND TRUTH)

Algorithms & models (AWS MARKETPLACE)

TRAIN

One-click model training & tuning

Optimization (NEO)

Models without training data (REINFORCEMENT LEARNING)

DEPLOY

One-click deployment & hosting

ML FRAMEWORKS & INFRASTRUCTURE

Frameworks

TensorFlow

mxnet

PYTORCH

Interfaces

GLUON

Keras

Infrastructure

EC2 P3 & P3dn

EC2 C5

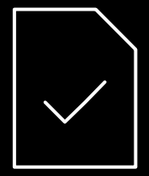
FPGAs

GREENGRASS

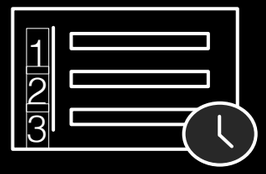
ELASTIC INFERENCE

aws SUMMIT

Amazon SageMaker:
Build, Train, and Deploy ML Models at Scale

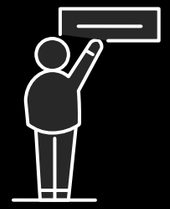


Collect and prepare training data



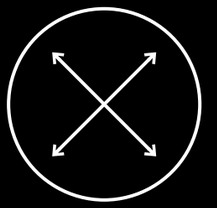
Choose and optimize your ML algorithm

Set up and manage environments for training



Train and Tune ML Models

Deploy models in production



Scale and manage the production environment

intuit

F1

NFL

tinder

KIA

Liberty Mutual.

CONVOY

SIEMENS

DOW JONES

Celgene

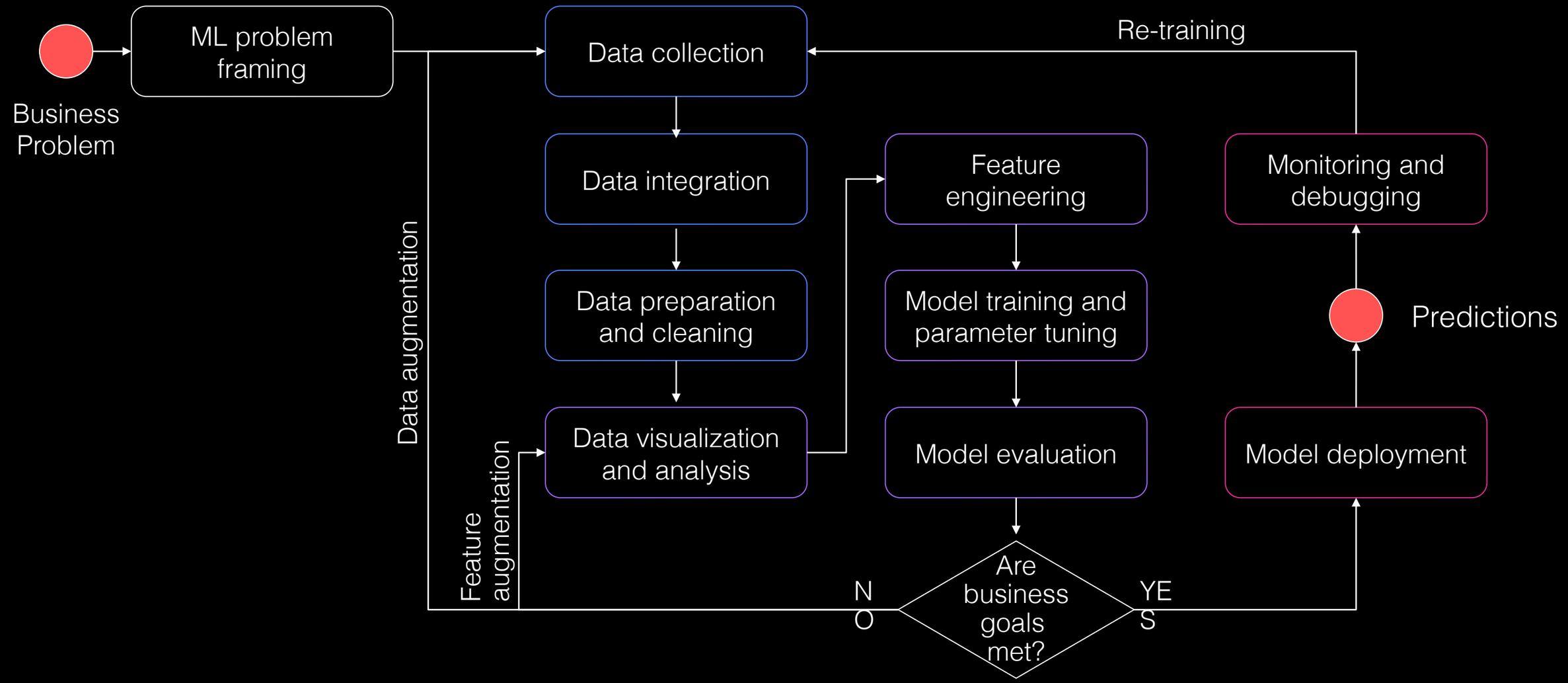
SONY

GE Healthcare

aws SUMMIT

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning cycle

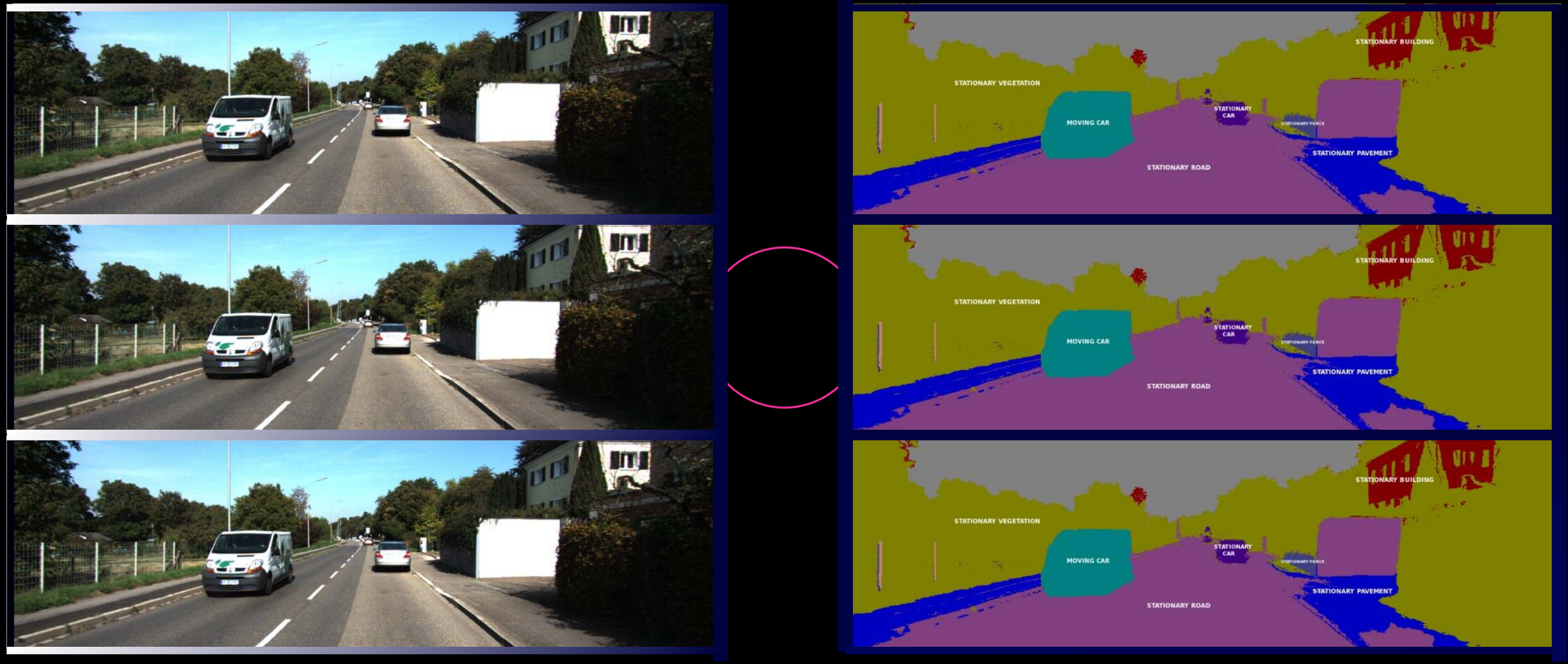


© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Successful models require high-quality data

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

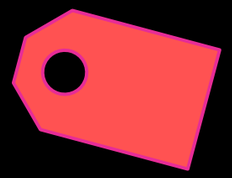
Successful models require high-quality data



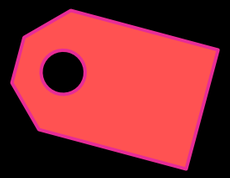
© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker Ground Truth

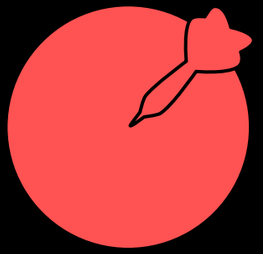
https://aws.amazon.com/blogs/aws/amazon-sagemaker-ground-truth-build-highly-accurate-datasets-and-reduce-labeling-costs-by-up-to-70



Quickly label training data



Easily integrate human labelers



Get accurate results

KEY FEATURES

Automatic labeling via machine learning

Ready-made and custom workflows for image bounding box, segmentation, and text

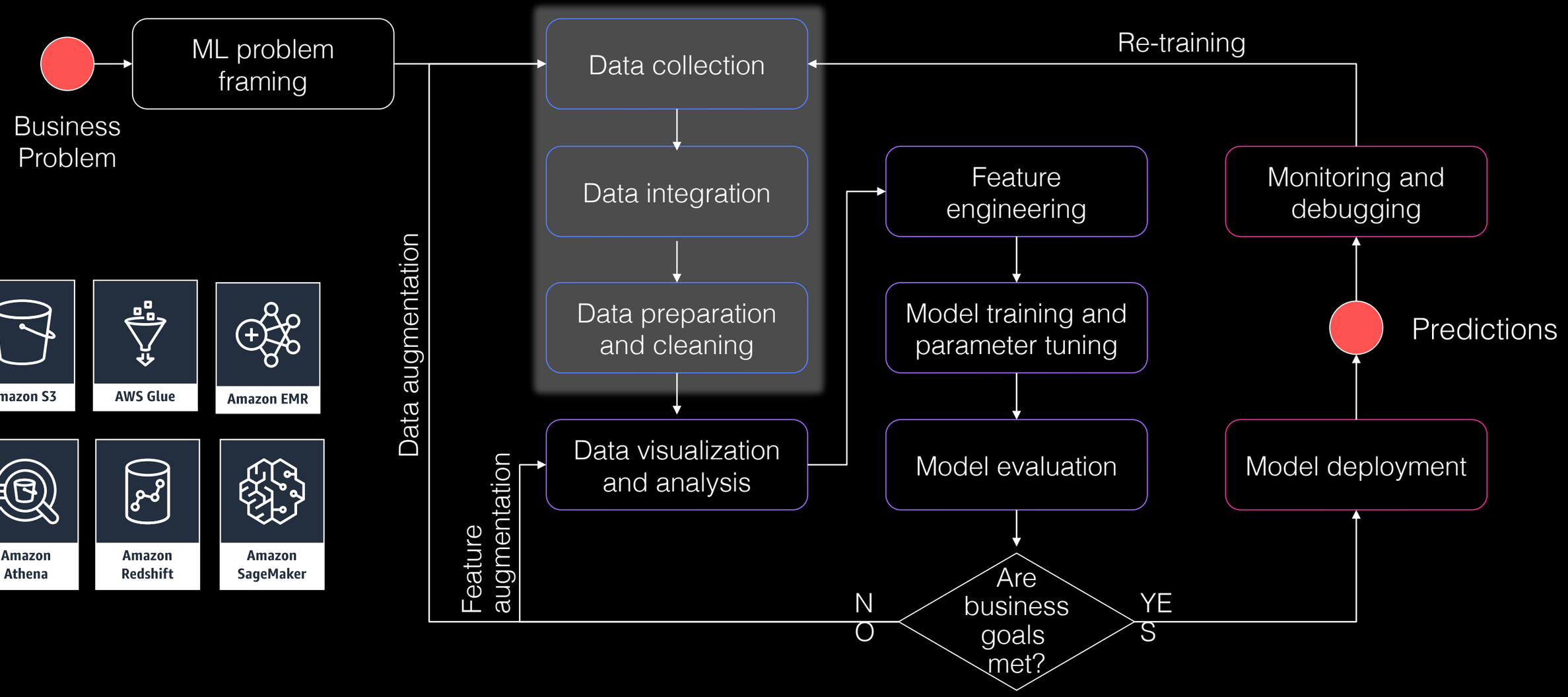
Private and public human workforce

Label management

aws SUMMIT

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

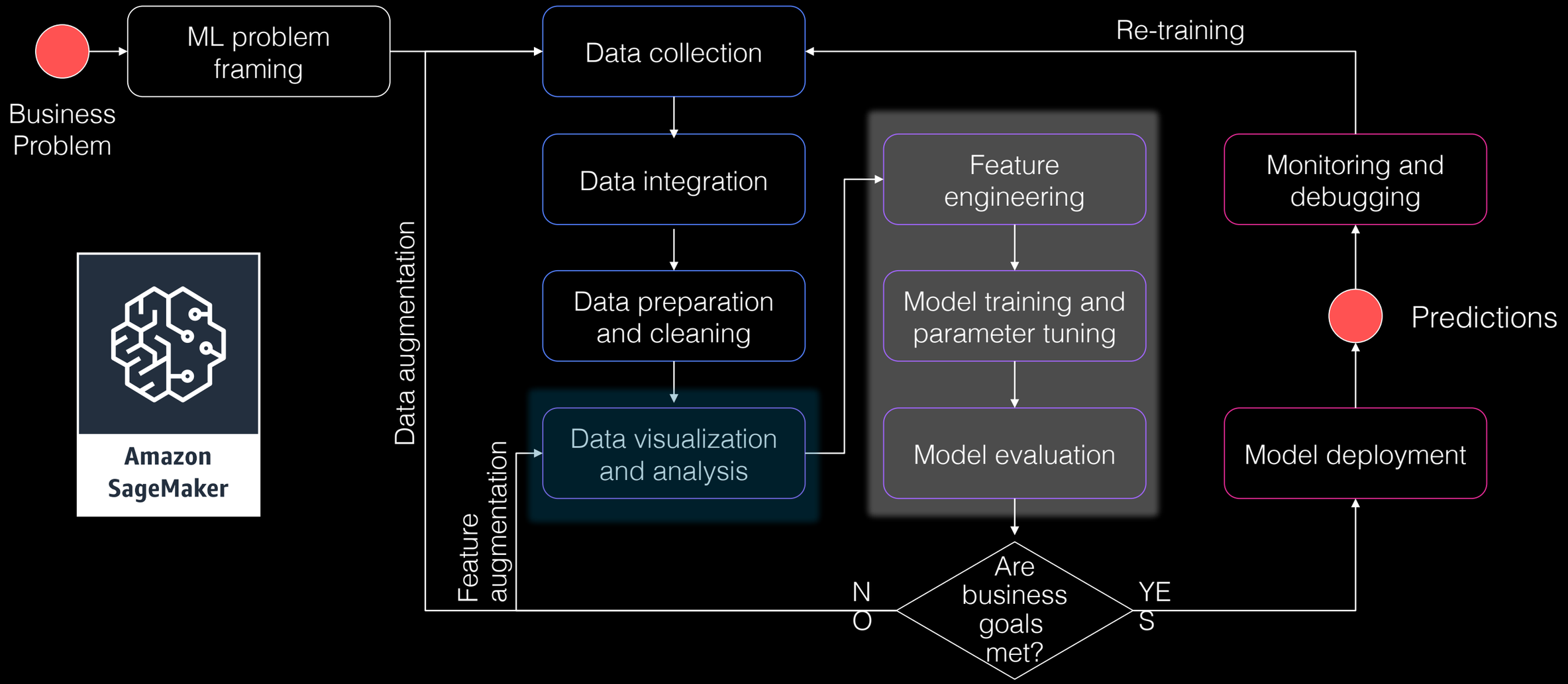
Manage data on AWS



aws SUMMIT

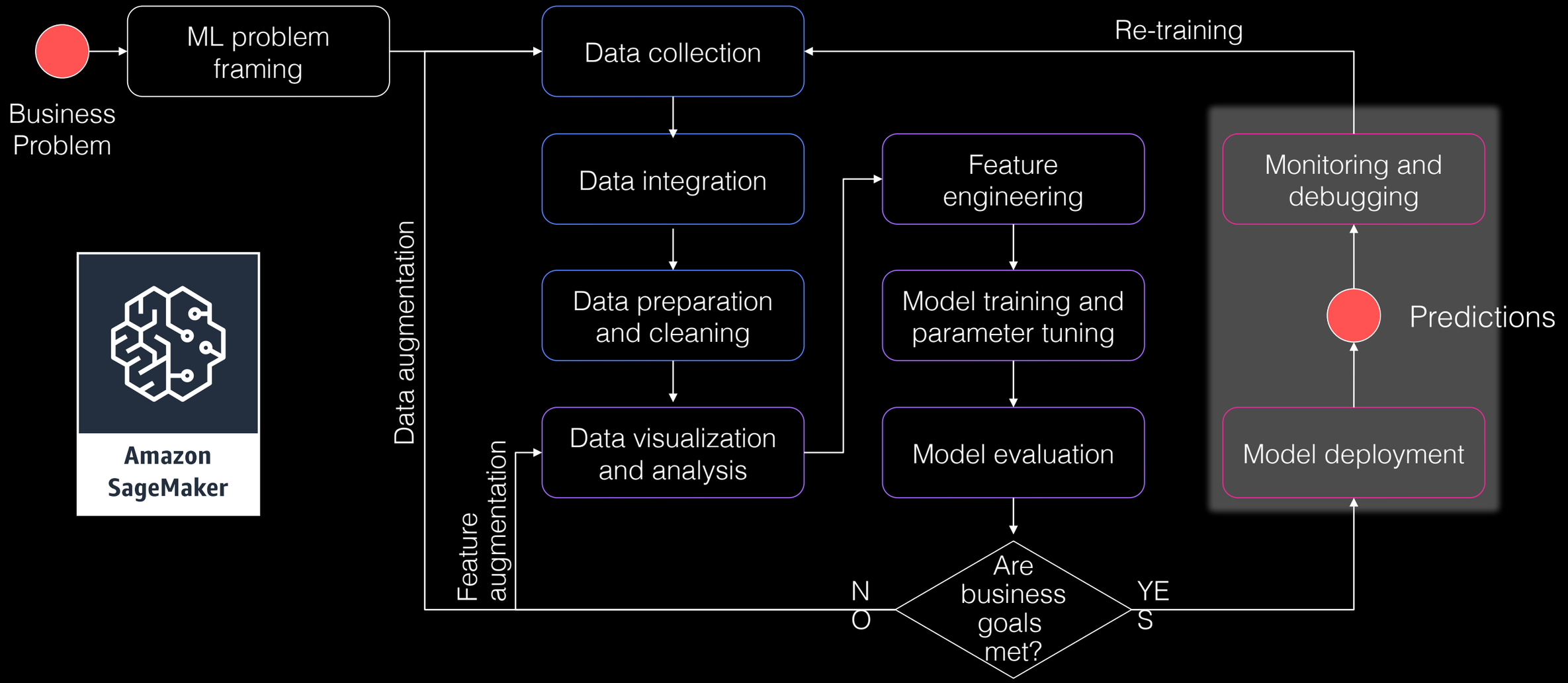
© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Build and train models using SageMaker



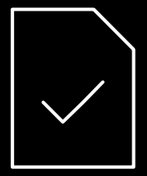
© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Deploy models using SageMaker



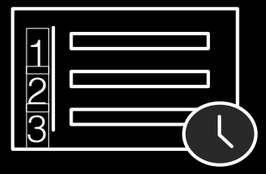
© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker



Collect and prepare training data

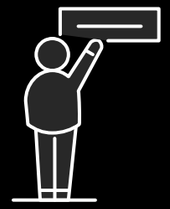
Git integration
Elastic inference



Choose and optimize your ML algorithm

Set up and manage environments for training

New built-in algorithms
scikit-learn environment
Model marketplace
Search

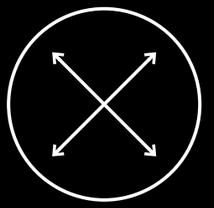


Train and Tune ML Models

P3DN, C5N
TensorFlow on 256 GPUs
Resume HPO tuning job

Deploy models in production

Model compilation
Elastic inference
Inference pipelines

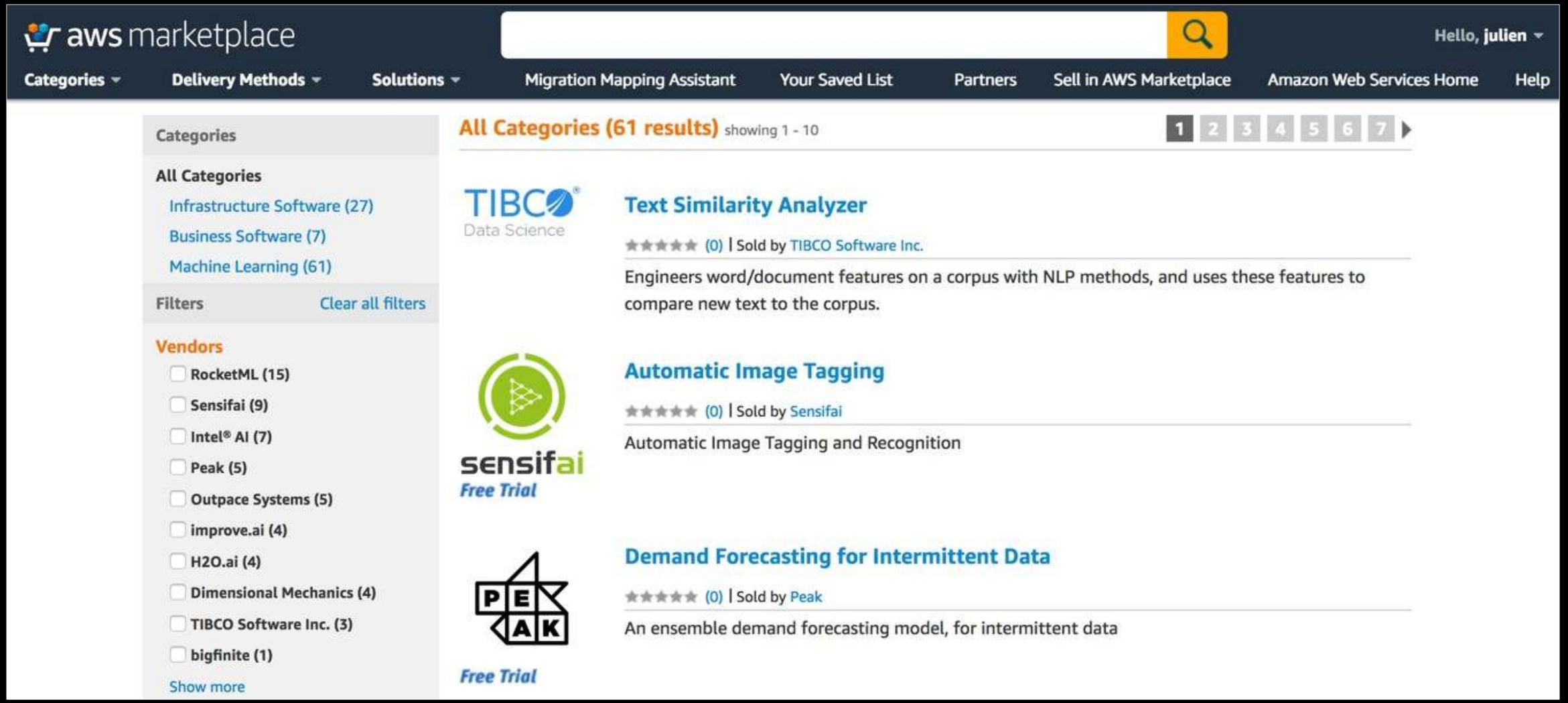


Scale and manage the production environment

aws SUMMIT

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Machine Learning Marketplace



© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Working with Amazon SageMaker

aws SUMMIT

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The Amazon SageMaker API

- Python SDK orchestrating all Amazon SageMaker activity
  - High-level objects for algorithm selection, training, deploying, automatic model tuning, etc.
  - Spark SDK (Python & Scala)

- AWS CLI: '*aws sagemaker*'

- AWS SDK: boto3, etc.

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Model options

Training code

Factorization Machines
Linear Learner
Principal Component Analysis
K-Means Clustering
XGBoost
And more

Built-in Algorithms

mxnet Chainer TensorFlow PyTorch scikit learn

Bring Your Own Script

Bring Your Own Container

aws SUMMIT

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

IMAGE RECOGNITION | *for the good of* | PRODUCT SEARCH

Sébastien BUTREAU
sebastien.butreau@tarkett.com
Group IT Projects & CCOE Manager
Tarkett

aws SUMMIT

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Tarkett

A worldwide leader in flooring & sports surfaces

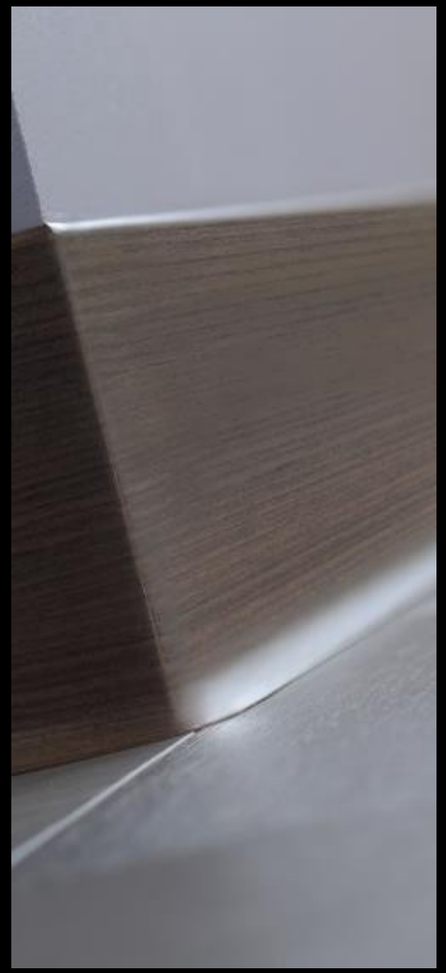
- → **€2.8B** Net sales
  (2018 figures)
- → **13,000** employees
- → Present in more than **100** countries
- → **1.3M square meters** of flooring sold each day

Could we recommend products to each user?

VINYL & LINOLEUM

CARPET

WOOD & LAMINATE



ACCESSORIES & RUBBER

SPORTS SURFACES

aws SUMMIT

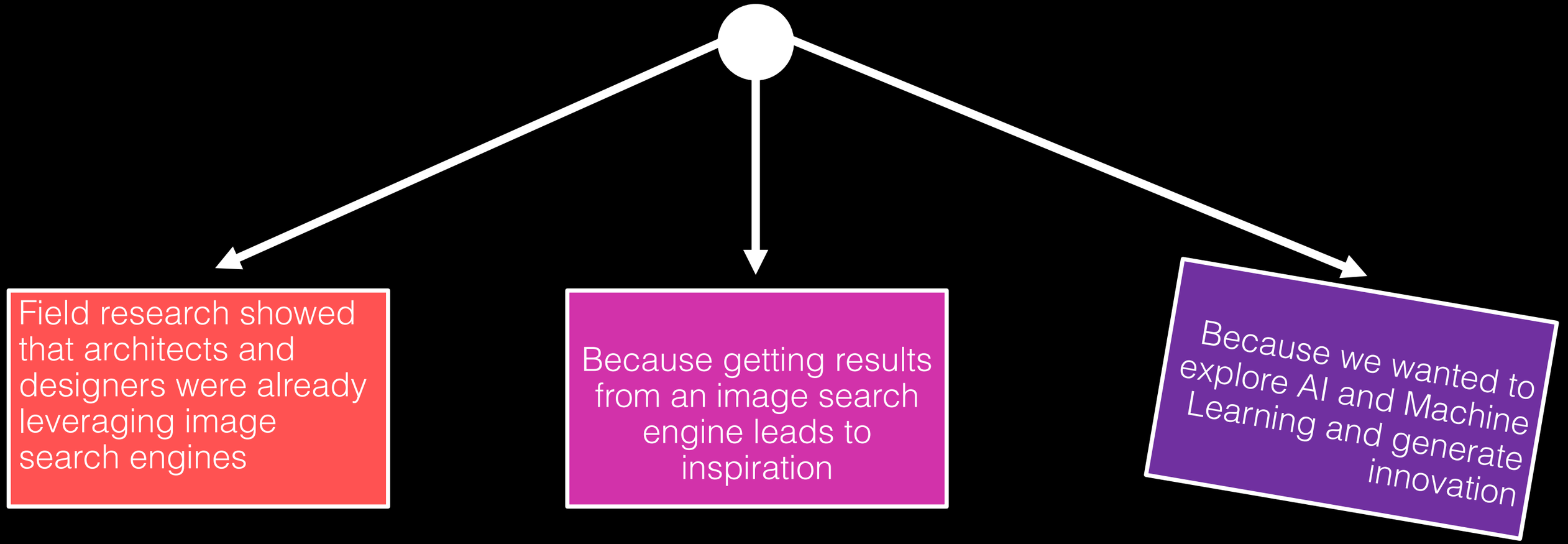
© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

«Building a bot»

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

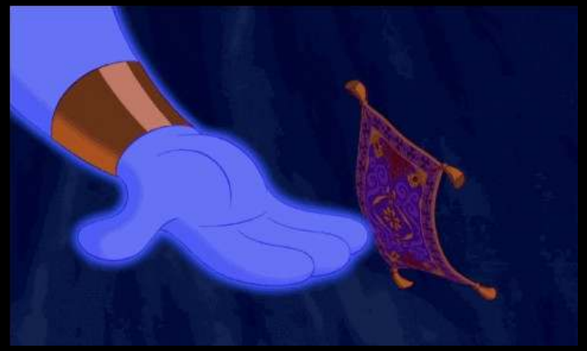
Why **image search?**

Why does it make sense in the context of flooring?



© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

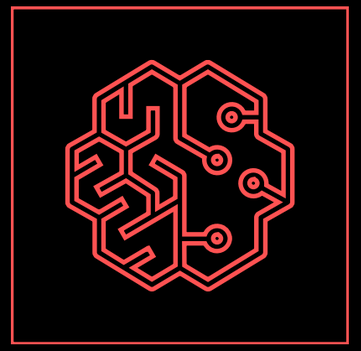
Project Aladdin



- Initial experimentation with computer vision algorithms, then switch to Deep Learning
- Data set construction
  - Thumbnails from the web site
  - Pictures of our products in the showroom
  - Data augmentation
- Train a VGG16 architecture with GPU instances (Amazon EC2 P3)
- Deploy the model and integrate it in the web application

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

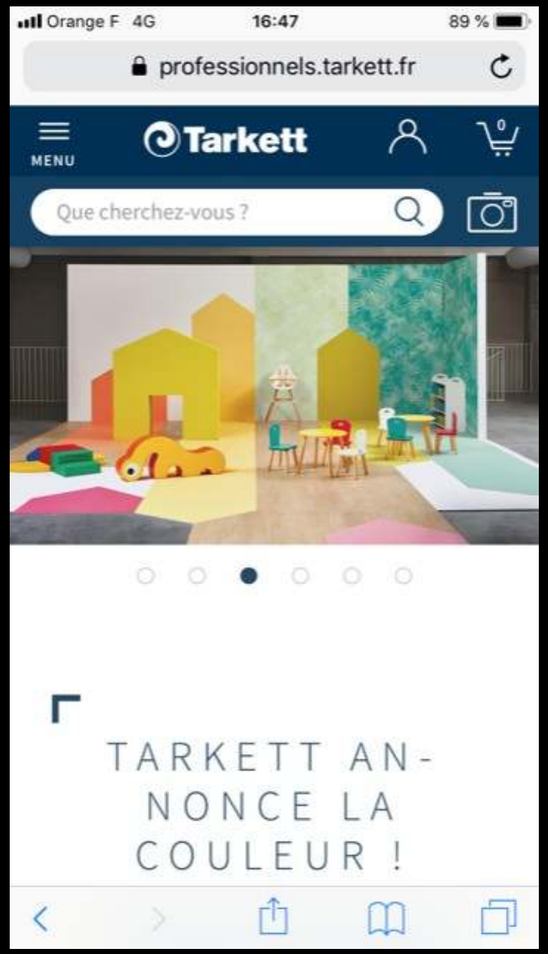
Why we used Amazon SageMaker

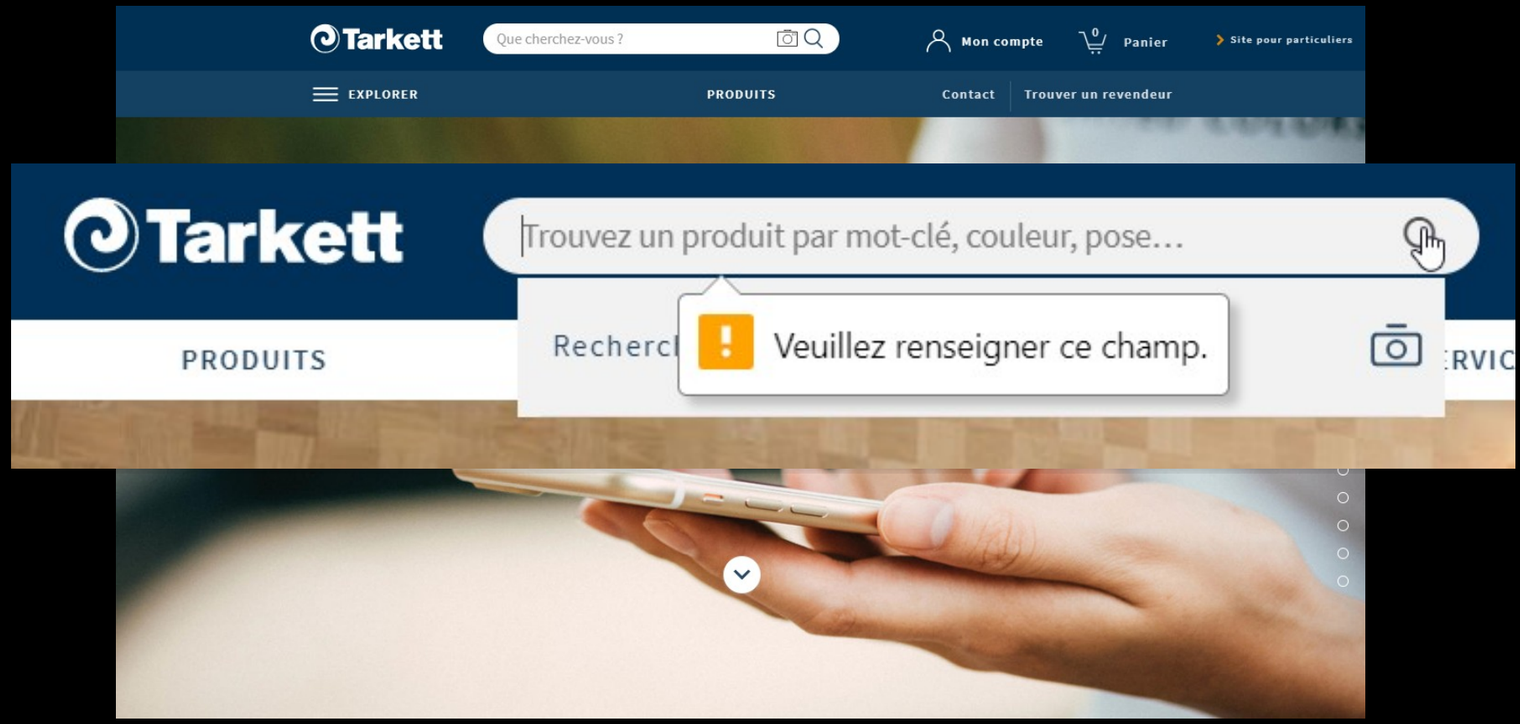


- Go quicker from idea to production
- Distributed training out of the box
- One line of code to deploy models
- Almost the same cost as Amazon EC2 ($300/month)

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Demo: https://professionnels.tarkett.fr





© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Next steps

Explore new possibilities opened by image search

- Find the best substitution when a product is out of stock
- Understand « in-situ » the real user demand
  - Incorporate external factors (seasons, fashion, styles)
  - Position our products against the competition
  - Generate new designs

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Merci!

Sébastien BUTREAU
sebastien.butreau@tarkett.com
Group IT Projects & CCOE Manager
Tarkett

Built-in algorithms

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Built-in algorithms

orange: supervised, yellow: unsupervised

Linear Learner: regression, classification	**Image Classification**: Deep Learning (ResNet)
Factorization Machines: regression, classification, recommendation	**Object Detection (SSD)**: Deep Learning (VGG or ResNet)
K-Nearest Neighbors: non-parametric regression and classification	**Neural Topic Model**: topic modeling
XGBoost: regression, classification, ranking https://github.com/dmlc/xgboost	**Latent Dirichlet Allocation**: topic modeling (mostly)
K-Means: clustering	**Blazing Text**: GPU-based Word2Vec, and text classification
Principal Component Analysis: dimensionality reduction	**Sequence to Sequence**: machine translation, speech to text and more
Random Cut Forest: anomaly detection	**DeepAR**: time-series forecasting (RNN)
Object2Vec: general-purpose embedding	**IP Insights**: usage patterns for IP addresses
Semantic Segmentation: Deep Learning	

, Inc. or its affiliates. All rights reserved.

Demo:
Text Classification with BlazingText

https://github.com/awslabs/amazon-sagemaker-examples/tree/master/introduction_to_amazon_algorithms/blazingtext_text_classification_dbpedia

BlazingText: Scaling and Accelerating Word2Vec using Multiple GPUs

Saurabh Gupta
Amazon Web Services
gsaur@amazon.com

Vineet Khare
Amazon Web Services
vkhare@amazon.com

https://dl.acm.org/citation.cfm?id=3146354

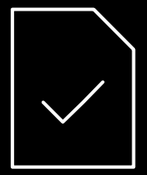
© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Demo:
Image classification with Caltech-256

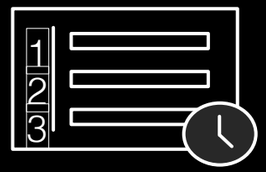
https://gitlab.com/juliensimon/dlnotebooks/sagemaker/

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker

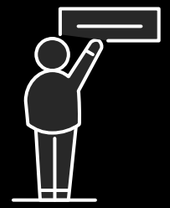


Collect and prepare training data



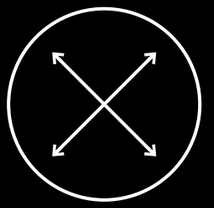
Choose and optimize your ML algorithm

Set up and manage environments for training



Train and Tune ML Models

Deploy models in production



Scale and manage the production environment

Build

Train

Deploy

aws SUMMIT

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Getting started

http://aws.amazon.com/free

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

https://github.com/aws/sagemaker-python-sdk
https://github.com/aws/sagemaker-spark
https://github.com/awslabs/amazon-sagemaker-examples
https://gitlab.com/juliensimon/ent321

https://medium.com/@julsimon
https://gitlab.com/juliensimon/dlnotebooks

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Julien Simon
Global Evangelist, AI and Machine Learning

@julsimon
https://medium.com/julsimon

aws SUMMIT

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Please complete the session survey.

aws SUMMIT

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.