aws machine learning

AI/ML trends

Julien Simon
Global Evangelist, AI & Machine Learning, AWS
@julsimon

AI/ML today

- « Who? »: Everyone. Literally everyone.
- « Why? »: is this still a thing in 2020? I don't think so, can we please move on?
- « How? » is what's on our customers' minds
- Plenty of options for ML software and hardware: open source, vendors, SaaS, clouds.
- Unfortunately, confusion and inefficiency are often the fruits of plentiness.
- Our role is to help customers find their best path to success.
- But are we sure we know who our customers are?

AI/ML customers: one size does not fit all

We (Developer Advocates) see four technical customer personas with distinct challenges (« my problem is… ») and expectations (« I want to solve it with… »)

1. Application Developer / Pragmatic Data Scientist
2. Researcher / Data Scientist
3. Data Engineer
4. DevOps / MLOps

1. Application Developer / Pragmatic Data Scientist

« I want to apply AI/ML to well-defined business use cases in the easiest way possible.

Not all problems require custom algorithms and models, and I don't have enough time or skills anyway.

I'm happy to work with high-level services or pre-trained models that solve the ML problem for me. »

→ Polly, Rekognition, Textract, etc. : call an API, get the job done

→ Rekognition Custom Labels, Transcribe Language Models, Forecast, Personalize, Fraud Detector: train on your data without ML skills

→ AWS Marketplace, SM AutoPilot, SM Penny: build/use « low code » ML models

2. Researcher / Data Scientist

« I focus on exploring data sets, and on building ML models based either on existing algorithms or on my own. I just want to build models of the highest quality. »

Complex persona, coming in different flavors:

- Confused: « I work with local ML infrastructure, and I'd like to move to the cloud. Where do I start? »
- Rational: « I already use a collection of ML tools in the cloud, and I'm looking to standardize my ML platform. »
- Open Source: « I want to build and deploy ML models using only open source tools. »

2.1 Confused Researcher / Data Scientist

« I work with local ML infrastructure, and I'd like to move to the cloud. Where do I start?

I want a productive environment to build, deploy, and evaluate models, without having to manage any local setup on my machine. I'm totally new to AWS and extremely confused about services, instance types, etc. »

→ Understand the starting point: desktop? On premise?
→ Demystify AWS in general and core services in particular (IAM, EC2, S3, VPC)
→ Introduce SageMaker Studio first, then the SDK
→ SM Processing, SM Experiments, batch transform should help a lot
→ Production-related capabilities are usually less interesting to them

2.2 Rational Researcher / Data Scientist

« I already use a collection of ML tools in the cloud, and I'm looking to standardize my ML platform.

I want to focus on ML, not on building tools. I want all the cloud benefits (elasticity, scalability, etc.) and no infrastructure management. »

→ Understand the starting point: EC2? Containers? SageMaker?

→ Show how SageMaker Studio helps with standardization and collaboration

→ Go deep on relevant advanced capabilities (distributed training, debugging, and all the re:Invent launches)

2.3 Open Source Researcher / Data Scientist

« I want to build and deploy ML models using only open source tools.
I want all the cloud benefits, but our company and IT has standardized on select open source offerings (such as K8s and Kubeflow), and only use these services. »

- → Understand the starting point: on premise? EC2? Containers?
- → Understand if this is about « can only » or « only want to »
  - → The former means changing company policy. This is a C-level discussion.
  - → The latter is either ideology or lack of awareness. Both can be answered.
- → EKS, Kubeflow, Deep Learning Containers, custom containers for SageMaker, etc.
- → Ask about advanced capabilities: « How do you debug models? », etc.
- → Show how to use them with the SageMaker SDK

3. Data Engineer

« I'm a data engineer working alongside data scientists. I focus on turning raw data into clean and expressive datasets that can be used for model training.

I have to work with lots of diverse data sources. Pandas is just not scaling. I use Spark a lot, but I don't want to manage clusters and infrastructure.

I just want to quickly and easily build high-quality datasets. »

→ Understand the starting point: EMR? Glue? Notebooks?
→ Data exploration and transformation: Mohave, Elixir, Yavapai
→ Data processing: SM Processing (now with PySpark),
→ Data quality: Deequ, Thundera

4. MLOps

« I'm a DevOps engineer working alongside data scientists. I focus on their infrastructure needs, and on models that are deployed in prod. I just want the best quality of service.

I do too much manual work. I want to automate complex multi-step workflows.

Robustness, scalability, and security are important too. »

→ Automation, automation, automation
  → They're probably already familiar with CloudFormation
  → Go deep on Yosemite
→ Cost optimization is certainly important: spot, Elastic Inference, Inferentia
→ Same for security: VPC training, encryption, etc.

Closing thoughts

- We always do what's best for our customers.
- Not for us, not for marketing, not for service teams.
- Helping customers starts with understanding who they are, where they are in their ML journey, why they're there, and what their #1 pain point is.
- Yes, SageMaker is our best effort at solving ML pain points, based on 20+ years of experience.
- There are many paths that can lead customers there, and they can be confusing.
- Some customers choose to use other AWS services. That's fine with us.
- We never force anything down their throat.
- We only obsess about helping customers find their path to ML success.