Amazon SageMaker Demonstration

Xuchong Du, Preethika Pradeep, Daniil Filienko

Introducing the Technology

What is AWS SageMaker:
AWS SageMaker is a fully managed machine learning service provided by Amazon Web Services (AWS) that simplifies the process of building, training, and deploying machine learning models at scale.

1. **AWS SageMaker is a Software-as-a-Service to use machine learning**
2. **Amazon launched SageMaker in 2017**
3. **SageMaker supports multiple programming languages**
4. **You can integrate SageMaker with other AWS services to build ETL pipelines.**

History of the Technology: Who

Who invented the technology:
Amazon Company (AWS)
Initial release on 29 November 2017

Competing platform:
Vertex AI (Google)
Azure Machine Learning (Microsoft)

Cloud Platforms	AWS Sagemaker	Azure	Vertex AI
LLM	Llama	GPT	Palm
Company	Meta	OpenAI	Google

AWS Sagemaker | Azure Machine Learning | GCP Vertex.ai

---

History of the Technology: Why

Personal Motivation:
ML learner's need to simplify the process of ML deployment.

Company Motivation:
Industry demand for comprehensive accessible ML tools and machine learning service at large scale.

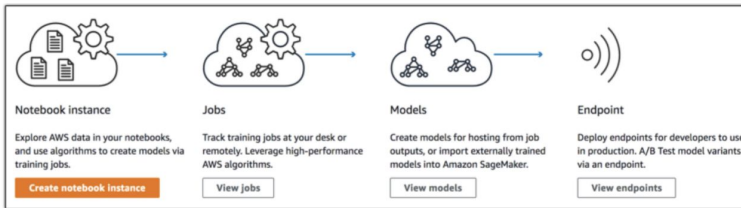**Machine Learning algorithm portion is barely 10% of the entire process!**

https://proceedings.neurips.cc/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf
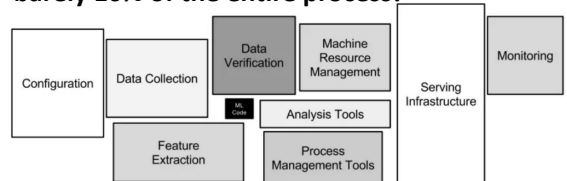
History of the Technology: How

1. Prepare Data   **2. Build Models**   **3. Train Models**
4. Deploy Models   **5. Monitor and Maintain**

SageMaker Data Wrangler :
Data preparation tool with a visual interface.

SageMaker Studio:
Machine learning environment for building, training, deploying, and analyzing models via a single web-based UI.

SageMaker Autopilot:
Automated ML service for deploying and monitoring models without coding.


Technology Now: Features Summary

- **Fully managed service**
- **Flexible notebooks**
- **Integration with AWS services**
- **Auto scaling capabilities**
- **Managed spot training**
- **Automated hyperparameter tuning**
- **Deployment flexibility**
- **A/B testing**
- **Managed data wrangling**
- **Integration with GitHub**

Technology Advantages

- **Reduces operational overhead**
- **Enables faster model development**
- **Lowers costs**
- **Simplifies machine learning**
- **Enhances productivity**
- **Improves model performance**
- **Supports better decision making**
- **Provides flexibility**
- **Facilitates collaboration**
- **Enables hybrid cloud portability**
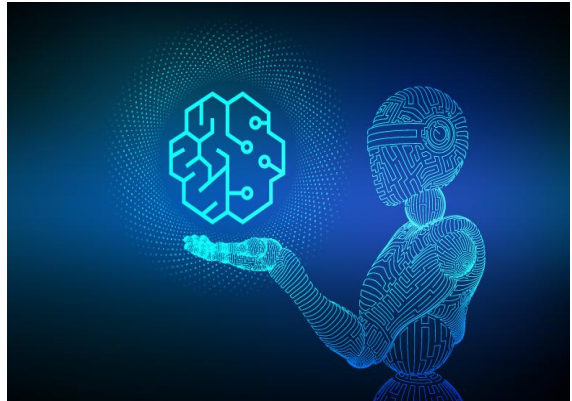
Technology Disadvantages

- **Vendor lock-in**
- **Limited customization**
- **No no-code interface**
- **Steep learning curve**
- **Additional charges possible**
- **Limited visualization options**
- **Spot training risks**
- **Cloud dependencies**
- **Long term commitment**
- **Security management overhead**

Cost Discussion

- **Pay only for what you use**
- **Notebooks are free up to a reasonable usage**
- **Managed spot training can lower costs**
- **Automated model tuning adds charges**
- **Monitoring overall costs needs diligence**
- **Requires understanding cost levers**
- **Benefits include reduced operational expenses**
- **Provides cost visibility**
- **Enables budget safeguards**

aws

Amazon SageMaker

End-to-End Managed Machine Learning Platforms

XENONSTACK

Build Model

Train and Test Model

Register and Manage Model
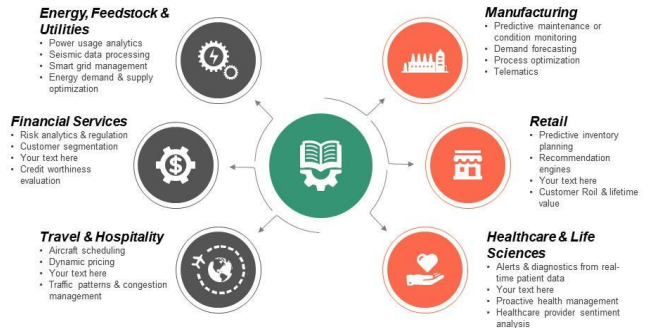
Deploy Service

Build Image

Cost Example

- **Notebook instance - A ml.t3.medium notebook instance used 40 hrs/month @ $0.08 per hour would cost $3.20**
- **Training - A ml.m5.4xlarge training job that runs for 2 hours would cost approx $1 based on spot instance rates**
- **Hosting - ml.m5.large production variant hosted for 30 days @ $0.096 per hour would cost approx $68**
- **Data storage - 100GB of SSD storage for model artifacts/intermediate data @ $0.023 per GB-month costs $2.3**
- **Inference - 1 million ml.m5.large inference requests @ $0.0001per request would incur a charge of $100**
- **Tuning - An automated tuning job using ml.m5.4xlarge & launching 10 iterations, each running for 2 hrs would cost approx $20**
- **Data labeling - Manual labeling of 2000 images with Ground Truth @ $0.012 per image would cost $24**
- **Data processing - Script run on ml.m5.xlarge instance for 0.2 hrs @ $0.192 would cost approx $0.04**
- **CloudWatch - Storing/accessing SageMaker log data in CloudWatch monitored at scale of 1KB per hour @ $0.50 per GB/month**

Example Use Cases

- **Predictive maintenance**
- **Customer churn prediction**
- **Fraud detection**
- **Demand forecasting**
- **Personalized recommendations**
- **Image classification**
- **Natural language processing**
- **Predictive analytics**
- **Anomaly detection**
- **Customer sentiment analysis**

Energy, Feedstock & Utilities
- Power usage analytics
- Seismic data processing
- Smart grid management
- Energy demand & supply optimization

Manufacturing
- Predictive maintenance or condition monitoring
- Demand forecasting
- Process optimization
- Telematics

Financial Services
- Risk analytics & regulation
- Customer segmentation
- Your text here
- Credit worthiness evaluation

Retail
- Predictive inventory planning
- Recommendation engines
- Your text here
- Customer Roil & lifetime value

Travel & Hospitality
- Aircraft scheduling
- Dynamic pricing
- Your text here
- Traffic patterns & congestion management

Healthcare & Life Sciences
- Alerts & diagnostics from real-time patient data
- Your text here
- Proactive health management
- Healthcare provider sentiment analysis

Usability

Moderately easy to learn

- Abundance of tutorials
  - Jumpstart
- Well integrated with other AWS services

Hard to master

- Has a vast amount of potential ML-related use cases
  - 18 services in SageMaker alone
- Development and deployment procedures are interwoven
  - Inherent in that type of service

Demonstration

- Live Presentation of SageMaker Canvas

13

Conclusion

Do we recommend to use it?

- Yes for technical professionals
  - Fully managed service pipeline [end-to-end]
    - Lower Operational Overhead
  - Many integrated services
- Not for amateurs
  - Expensive
    - Complex pricing policy
  - People without any ML experience
    - Services presume basic understanding of ML concepts
    - Can be wasteful

14

Questions?

Thank you!