INNOVATE

DATA EDITION

23 August, 2022

Simplify customer purchase intent predictions with analytics and ML

Kamal Manchanda
Solutions Architect, AISPL

Sureshkumar K V
Prototyping Engineer, AISPL

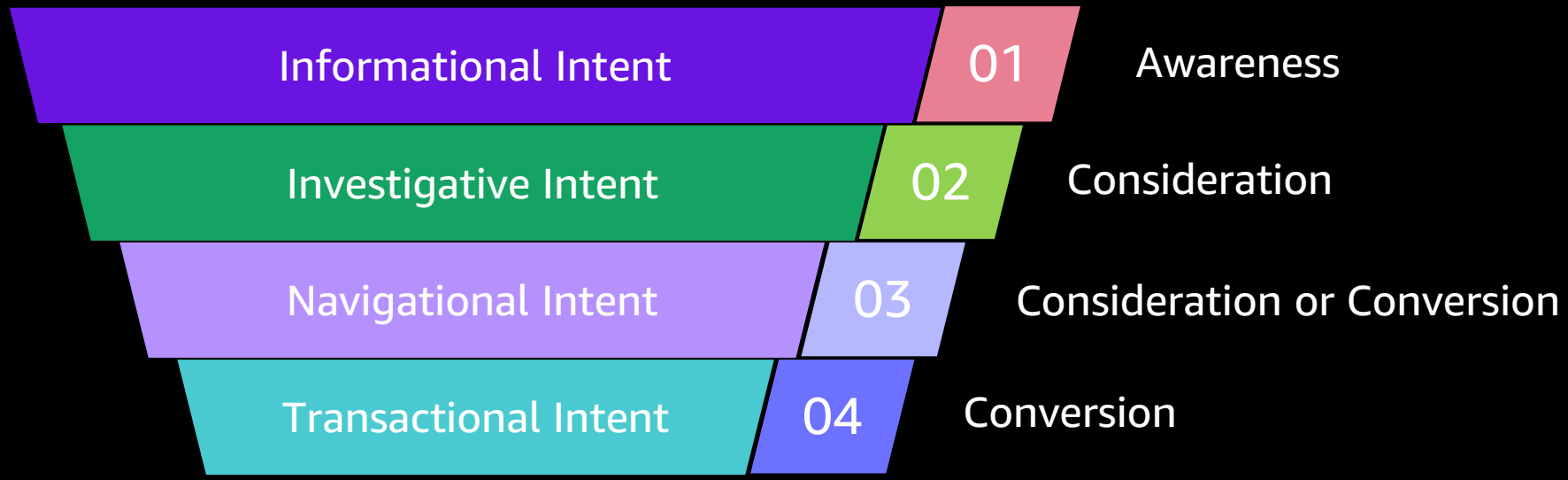
Agenda

- Purchase intent prediction – Use case overview
- How to predict purchase intent
- Overview of key services
- Demo 1 - Real time ingestion and inference
- Use Auto ML to create high-quality models with just a few clicks
- Demo 2 – Model training and deployment
- Key takeaways
- How to get started

Purchase intent prediction – Use case overview

What is purchase intent

Purchase intent is the probability that a consumer will buy a product or service



Why predict purchase intent

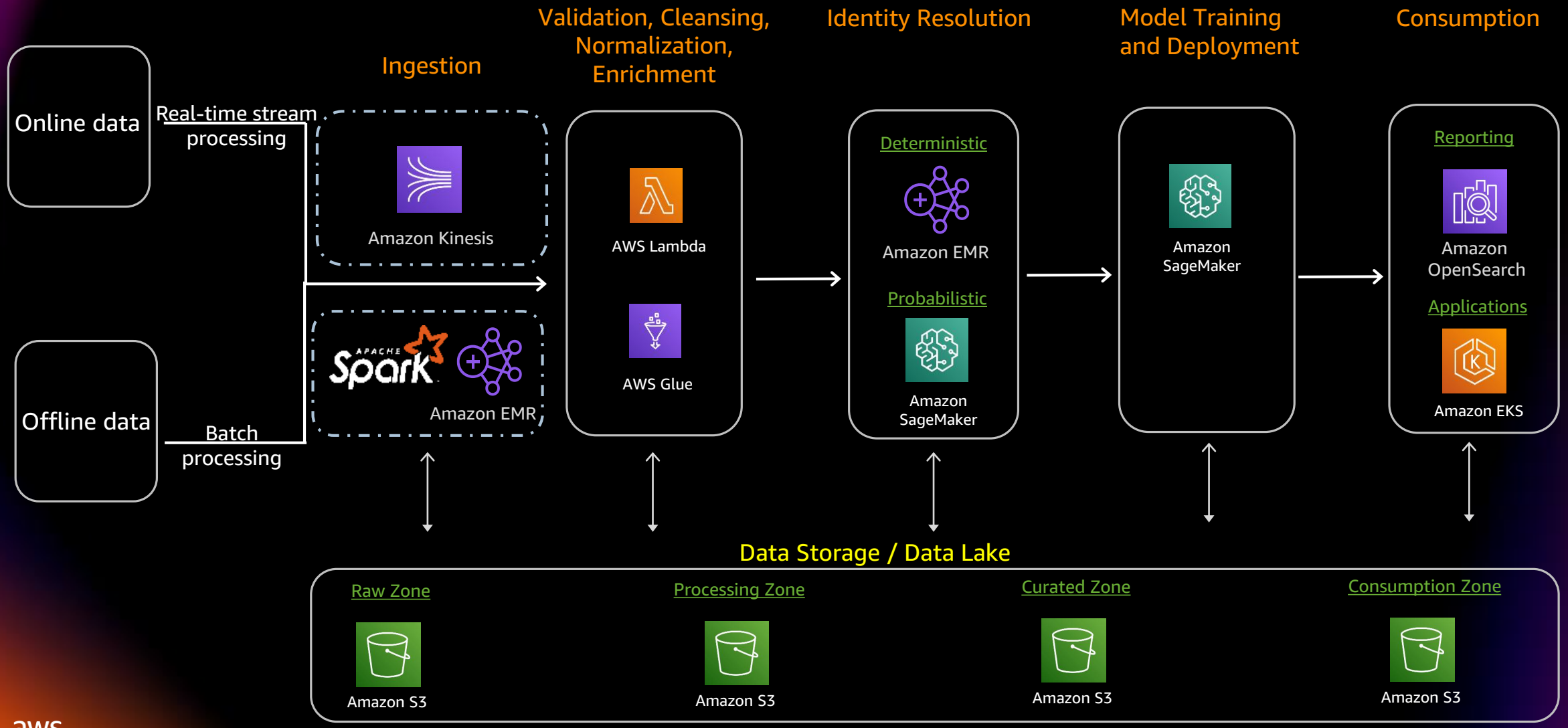
- Improve conversion metrics
- Increase margins
- Protect brand equity and perception
- Improve profitability and marketing channel ROI
- Personalize the customer experience
- Drive customer lifetime value and retention

Key Factors for Predicting Purchase Intent

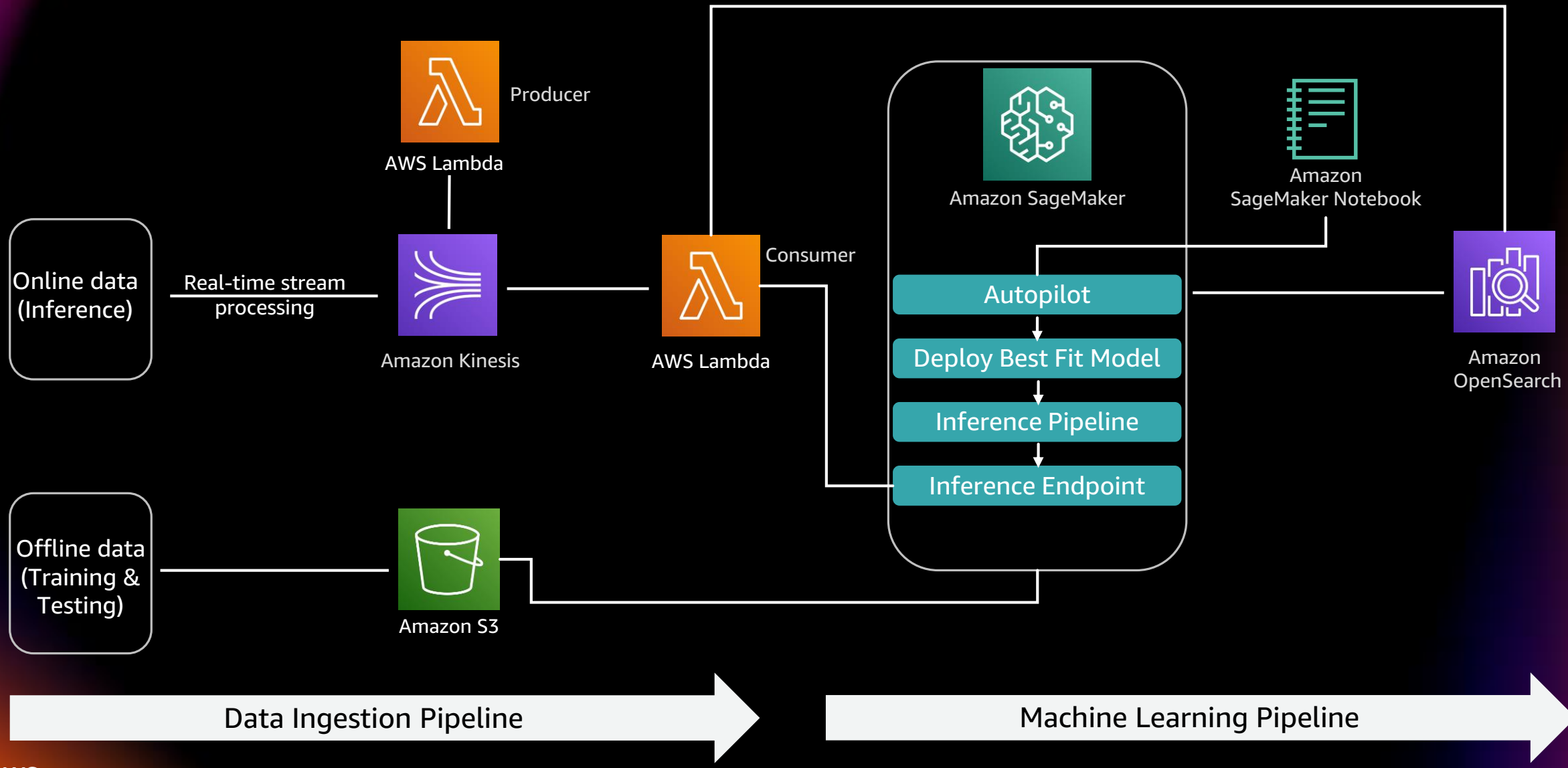
- Demographies
- Website engagement
- Past purchases
- Interaction with marketing messages
- Product reviews on website or social media

How to predict purchase intent

Reference Architecture



Demo Architecture



Overview of AWS key services

Amazon Kinesis

Easily collect, process, and analyze data streams in real time

Kinesis Data Streams

Collect and store data streams for analytics

Kinesis Data Firehose

Load data streams into AWS data stores

Kinesis Data Analytics

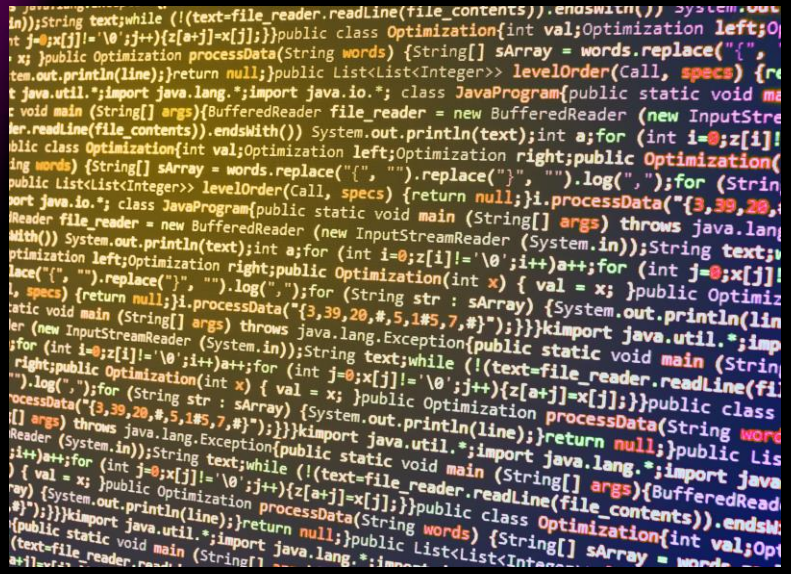
Analyze data streams with KDA Studio or Apache Flink

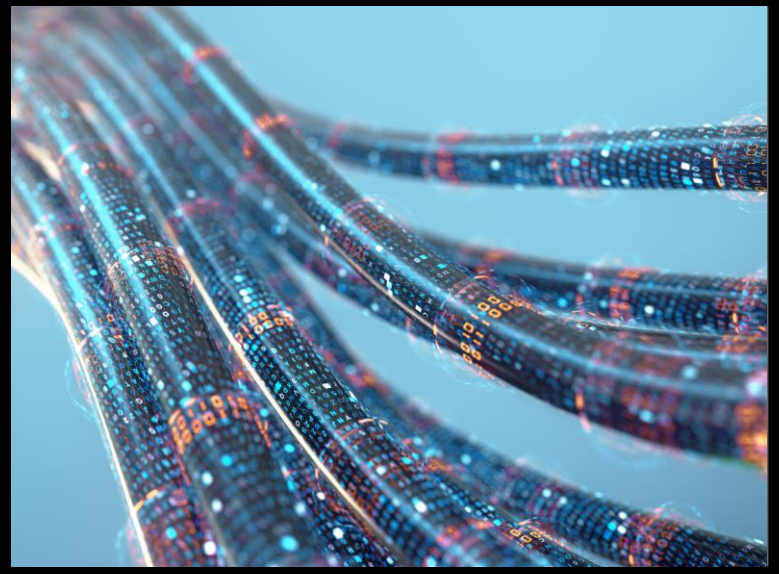
Kinesis Video Streams

Collect and store video streams for analytics

Amazon OpenSearch

Amazon OpenSearch is a powerful analytics engine





Text search

Natural language

Boolean queries

Relevance

Streaming data

High-volume ingest

Near real time

Distributed storage

Analysis

Time-based visualizations

Nestable statistics

Time series tools

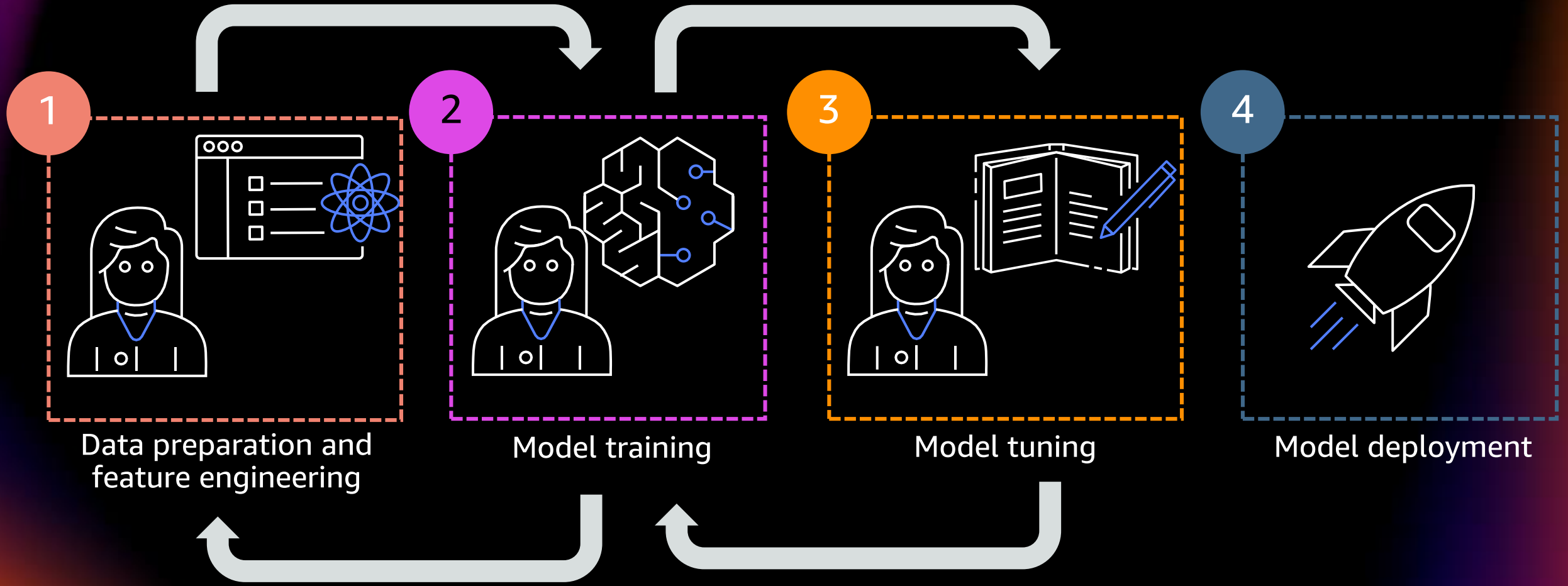
Amazon SageMaker

Amazon SageMaker helps organizations harness ML

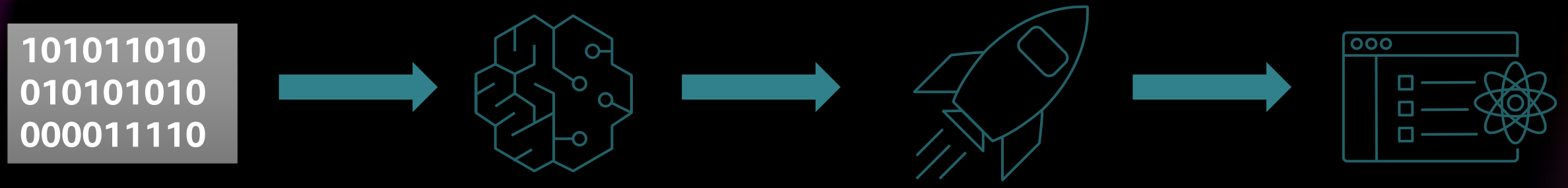
Demo 1 – Real time ingestion and inference

Use Auto ML to create high-quality models with just a few clicks

Why building ML models is time consuming



Auto-ML with Amazon SageMaker Autopilot

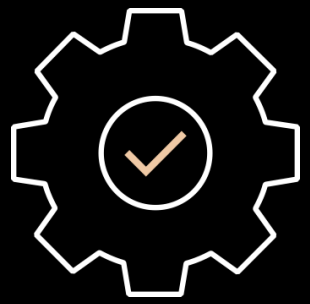


Amazon SageMaker Autopilot automatically builds, trains, and tunes the best machine learning models based on your data, while allowing you to maintain full control and visibility.

Amazon SageMaker Autopilot core features

Quick to start

Provide your data in a tabular form and specify target prediction



Automatic model creation

Get ML models with feature engineering and model tuning automatically done

Visibility and control

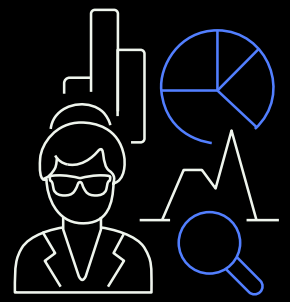
Get notebooks for your models with source code

Recommendations and optimization

Get a leaderboard and continue to improve your model

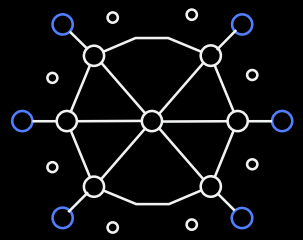
Automatic model creation for tabular data with full visibility and control

Build your models securely



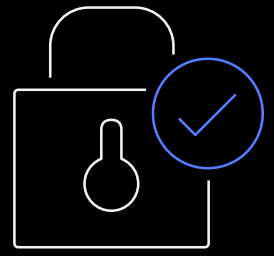
Access control

Use IAM policies to control access to data, models, and endpoints



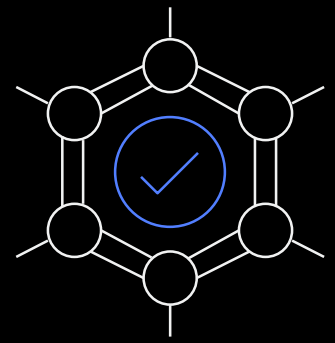
End-to-end encryption

Support for PrivateLink VPC endpoints, encryption at rest and in-transit



Secure auditing

Amazon CloudWatch and AWS CloudTrail integration



Support for compliance efforts for multiple standards

SOC, PCI, FedRAMP, HIPAA, and more

Complete integration with Amazon SageMaker security and compliance features

Tabular data as a common data type

- Customer churn prediction
- Customer lifetime value
- Credit risk prediction
- Sale prediction
- Price predictions
- Price elasticity for a customer
- Risk assessment
- Loan risk classification

ID_codesort	var_0	var_1	var_2	var_3	var_4	target
test_0	11.0656	7.7798	12.9536	9.4292	11.4327	0
test_1	8.5304	1.2543	11.3047	5.1858	9.1974	0
test_2	5.4827	-10.358	10.1407	7.0479	10.2628	0
test_3	8.5374	-1.3222	12.022	6.5749	8.8458	0
test_4	11.7058	-0.1327	14.1295	7.7506	9.1035	1
test_5	5.9862	-2.2913	8.6058	7.0685	14.2465	1
test_6	8.4624	-6.1065	7.3603	8.2627	12.0104	0

Data exploration notebook

- Dataset statistics – row-wise and column-wise
- Suggested remedies for common dataset problems

Dataset Sample

The following table is a random sample of **10** rows from the training dataset. For ease of presentation, we are only showing **20 of the 21** columns of the dataset.

Suggested Action Items

- Verify the input headers correctly align with the columns of the dataset sample. If they are incorrect, update the header names of your input dataset in Amazon Simple Storage Service (Amazon S3).

	State	Account Length	Area Code	Phone	Int'l Plan	VMail Plan	VMail Message	Day Mins	Day Calls	Day Charge	...	Eve Calls	Eve Charge	Night Mins	Night Calls	Night Charge	Intl Mins	Intl Calls
0	CO	76	408	412-4185	no	yes	26	214.6	110	36.48	...	87	17.44	134.6	140	6.06	8.1	2
1	NY	104	415	391-1793	no	yes	26	189.1	112	32.15	...	97	15.15	199.3	104	8.97	11.1	4
2	KY	122	408	392-1616	no	yes	27	253.7	84	43.13	...	109	19.48	190.5	123	8.57	9.2	5
3	NH	67	415	355-1113	no	yes	40	104.9	65	17.83	...	93	18.39	217.4	128	9.78	9.6	9
4	WI	153	510	349-3112	no	no	0	159.5	103	27.12	...	90	23.42	176.7	126	7.95	10.1	2
5	NH	146	510	345-2319	no	no	0	115.6	77	19.65	...	100	18.16	218.4	72	9.83	10.7	6
6	WV	63	510	328-9797	no	no	0	261.8	69	44.51	...	135	20.83	202.1	94	9.09	14.7	4
7	NH	90	408	393-7322	no	no	0	140.2	97	23.83	...	102	18.18	120.0	126	5.4	7.1	2

Candidate generation notebook

Override points

- Algorithms considered
- Evaluation metric
- Hyperparameter ranges
- Model search strategy
- Instances used

The SageMaker Autopilot Job has analyzed the dataset and has generated **10** machine learning pipeline(s) that use **2** algorithm(s). Each pipeline contains a set of feature transformers and an algorithm.

Available Knobs

1. The resource configuration: instance type & count
2. Select candidate pipeline definitions by cells
3. The linked data transformation script can be reviewed and updated. Please refer to the README.md for detailed customization instructions.

dpp0-xgboost: This data transformation strategy first transforms 'numeric' features using RobustImputer (converts missing values to nan), 'categorical' features using ThresholdOneHotEncoder, 'text' features using MultiColumnTfidfVectorizer. It merges all the generated features and applies RobustStandardScaler. The transformed data will be used to tune a *xgboost* model. Here is the definition:

[ ]:
```
automl_interactive_runner.select_candidate({
    "data_transformer": {
        "name": "dpp0",
        "training_resource_config": {
            "instance_type": "ml.m5.4xlarge",
            "instance_count": 1,
            "volume_size_in_gb":  50
        },
        "transform_resource_config": {
            "instance_type": "ml.m5.4xlarge",
            "instance_count": 1,
        },
        "transforms_label": True,
        "transformed_data_format": "application/x-recordio-protobuf",
        "sparse_encoding": True
    }
```

Demo 2 – Model training and deployment

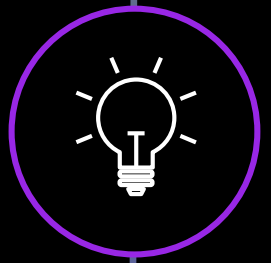
Key Takeaways

- Purchase intent prediction is critical in terms of conversion and improving your business metrics.
- AWS offers real time analytical services which we can use to collect the relevant data and use the same for training and inference.
- With Amazon SageMaker Autopilot, the model training and deployment becomes effortless.

Get started with your analytics and ML journey

Get started

Get started with Amazon Kinesis, Amazon OpenSearch and Amazon SageMaker directly from the AWS console



Learn more

https://aws.amazon.com/kinesis/

https://aws.amazon.com/opensearch-service/

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

Hands on Workshop

https://catalog.us-east-1.prod.workshops.aws/workshops/63069e26-921c-4ce1-9cc7-dd882ff62575/en-US/lab4/automl#overview

aws

Visit the AWS Data resource hub

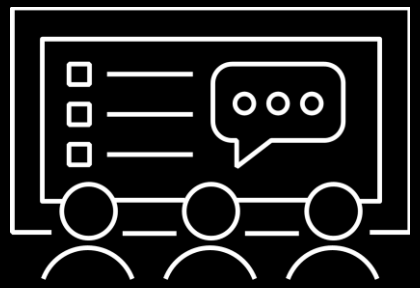
A modern data strategy can help you manage, act on, and react to your data so you can make better decisions, respond faster, and uncover new opportunities. Dive deeper with these resources today.

- Harness data to reinvent your organization
- In unpredictable times, a data strategy is key
- Make data a strategic asset
- Rewiring your culture to be data-driven
- Put your data to work with a modern analytics approach
- … and more!

https://tinyurl.com/data-hub-aws

Visit resource hub

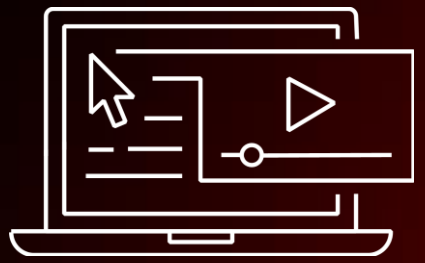
AWS Training and Certification for Data and Analytics



AWS Data & Analytics FREE Training Resources

Discover how to harness data, one of the world's most valuable resources, and innovate at scale.

https://bit.ly/3Ntlhy7



AWS Data Analytics Learning Plan

This learning plan expose you to the fastest way to get answers from all your data to all your users. It can also help prepare you for the AWS Certified Data Analytics - Specialty certification exam.

https://bit.ly/3wBVjD1

AWS Certified Data Analytics - Specialty

Earning AWS Certified Data Analytics – Specialty validates expertise in using AWS data lakes and analytics services.

https://go.aws/3lwF0RR

Thank you for attending AWS Innovate – Data Edition

We hope you found it interesting! A kind reminder to **complete the survey.**
Let us know what you thought of today's event and how we can improve the event experience for you in the future.


aws-apj-marketing@amazon.com


twitter.com/AWSCloud


facebook.com/AmazonWebServices


youtube.com/user/AmazonWebServices


slideshare.net/AmazonWebServices


twitch.tv/aws

Thank you!