



Machine Learning In a Day Workshop

Technology Solution Professional, Data & AI
Ahmed Sherif



Agenda

- Quick Introduction
- Techniques
- Process
- Tools
- Demo

Quick Introduction

ALLOW MYSELF TO INTRODUCE...MYSELF



What is Machine Learning?

Machine Learning is the study of the building blocks of systems that can learn from data

What is Machine Learning?

“Learning is any process in which a system improves performance by experience.”

“Machine Learning is concerned with computer programs that automatically improve their performance with experience”

--Herbert Simon

Turing Award 1975

Nobel Prize in Economics 1978



Why Machine Learning now?

Flood of Data

Increasing computational power

Growing progress in available algorithms from researchers

Growth of open source

Support from Industries

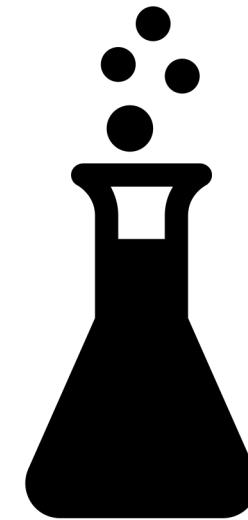
How are all 3 related: Data Science, AI, and ML

AI

- Focus on core human abilities such as vision, speech, language, decision making, and other complex tasks
- Designing machines and software to emulate these processes.

Machine
Learning

- Create models that predict, prescribe, group, or identify



Data Science

Data Science, AI, and Machine Learning

- Data science produces insights
- Machine learning produces predictions
- Artificial intelligence produces actions

How are all 3 used together

Suppose we were building a self-driving car, and were working on the specific problem of stopping at stop signs.

We would need skills drawn from all three of these fields.



Machine Learning

The car has to recognize a stop sign using its cameras. We construct a dataset of millions of photos of street side objects, and train an algorithm to **predict** which have stop signs in them.



Artificial Intelligence

Once our car can recognize stop signs, it needs to decide when to take the **action** of applying the brakes. It's dangerous to apply them too early or too late, and we need it to handle varying road conditions (for example, to recognize on a slippery road that it's not slowing down quickly enough)



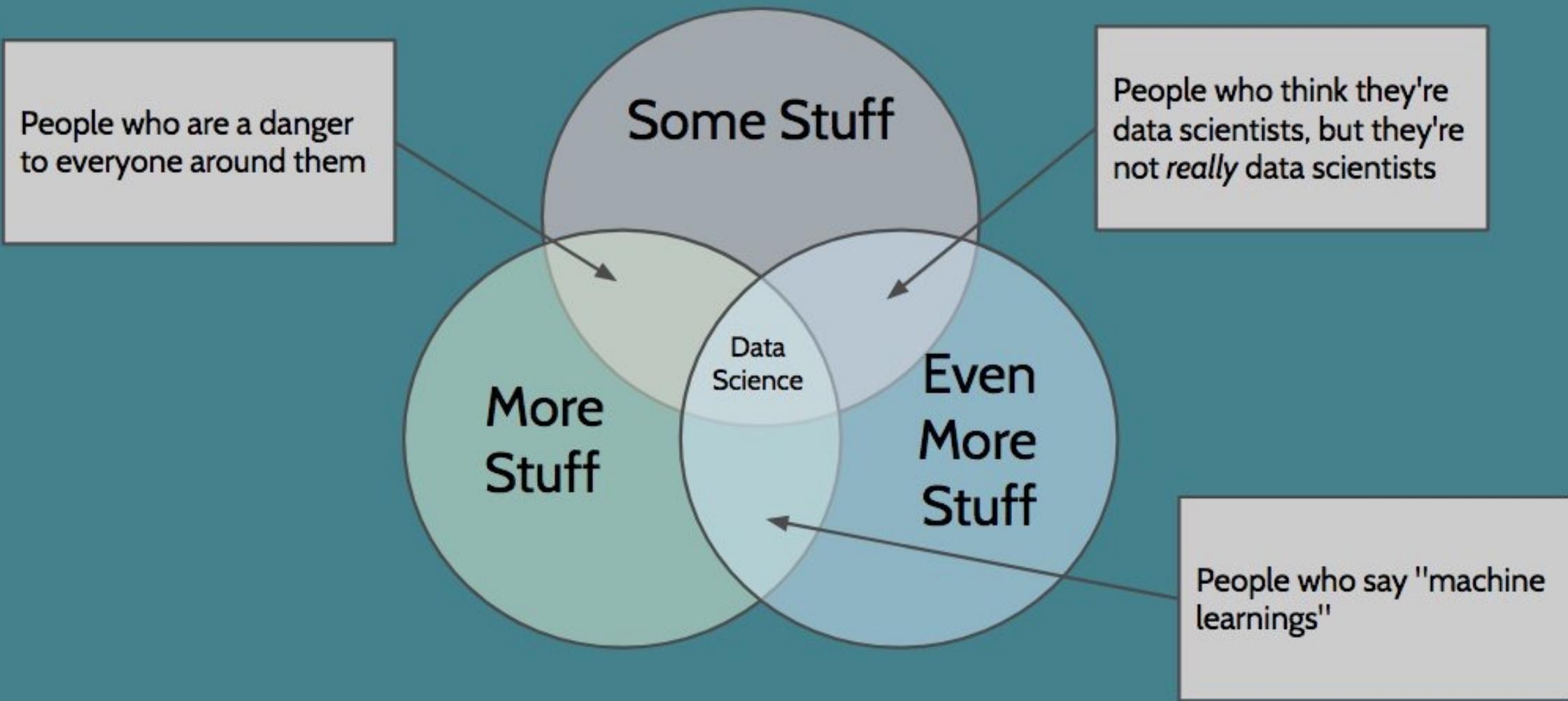
Data Science

In street tests we find that the car's performance isn't good enough, with some false negatives in which it drives right by a stop sign.

After analyzing the street test data, we gain the **insight** that the rate of false negatives depends on the time of day: it's more likely to miss a stop sign before sunrise or after sunset.

We realize that most of our training data included only objects in full daylight, so we construct a better dataset including nighttime images and go back to the machine learning step.

Data Science Is A Broad Field



Is Machine Learning a new concept?

The term ***Machine Learning*** may be relatively new

The algorithms behind them are anything but new

Techniques

Classification



Identifying to which of a set of **categories** (sub-populations) a new **observation belongs**, on the basis of a training set of data containing observations (or instances) whose category membership is known.

Is the movie coming out Fresh or Rotten?



Regression



A set of statistical processes for **estimating the relationships among variables**. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a **dependent variable** and one or more **independent variables** (or 'predictors').

How much money did the movie make?



Incredibles 2

Domestic Total as of Jun. 27, 2018:
\$384,163,124

Distributor: Buena Vista	Release Date: June 15, 2018
Genre: Animation	Runtime: 1 hrs. 58 min.
MPAA Rating: PG	Production Budget: N/A

Clustering



The task of **grouping a set of objects** in such a way that objects in the same group (called a **cluster**) are **more similar** (in some sense) to each other than to those in other groups (clusters).



Which customers are the most similar?

Can we create groups from this data?

I don't know anything about these entities where do I start?

Anomaly Detection



The identification of **items, events or observations** which **do not conform** to an **expected pattern** or other items in a dataset. Typically the anomalous items will translate to some kind of problem such as bank fraud, a structural defect, medical problems or errors in a text.

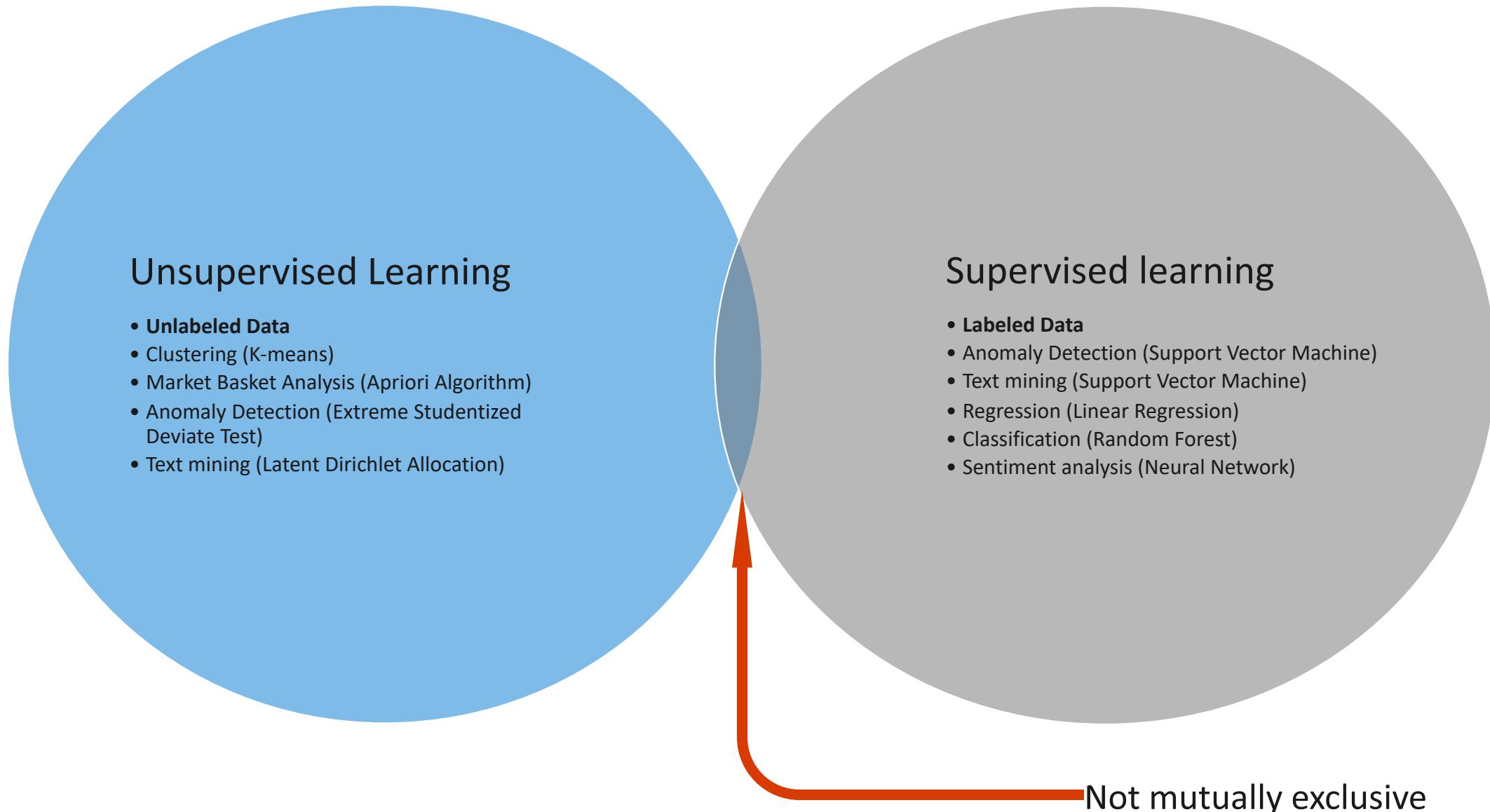


How can we spot possible fraud?

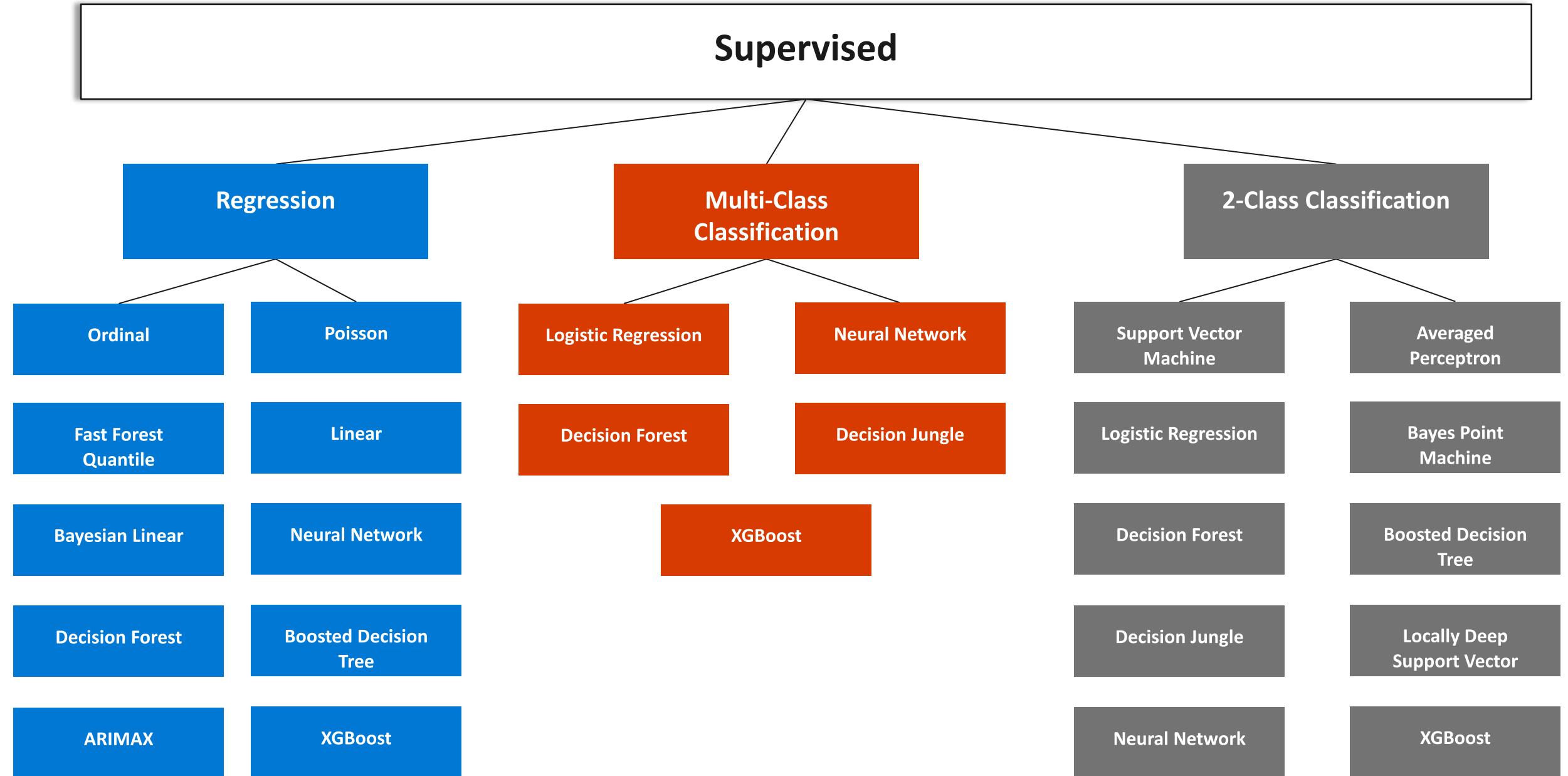
Can I automate identifying outliers?

Process

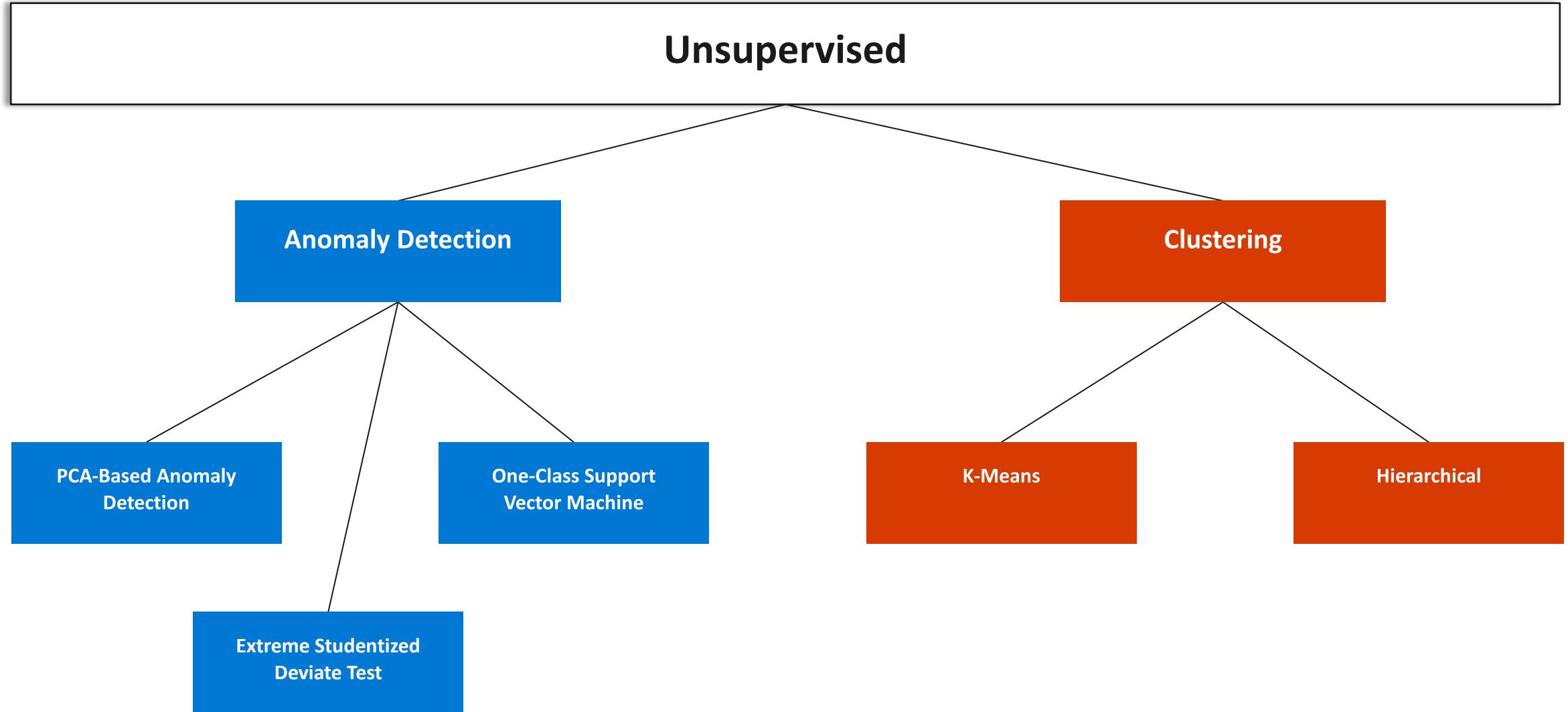
Supervised vs Unsupervised Learning



Algorithm selection (Supervised Learning)

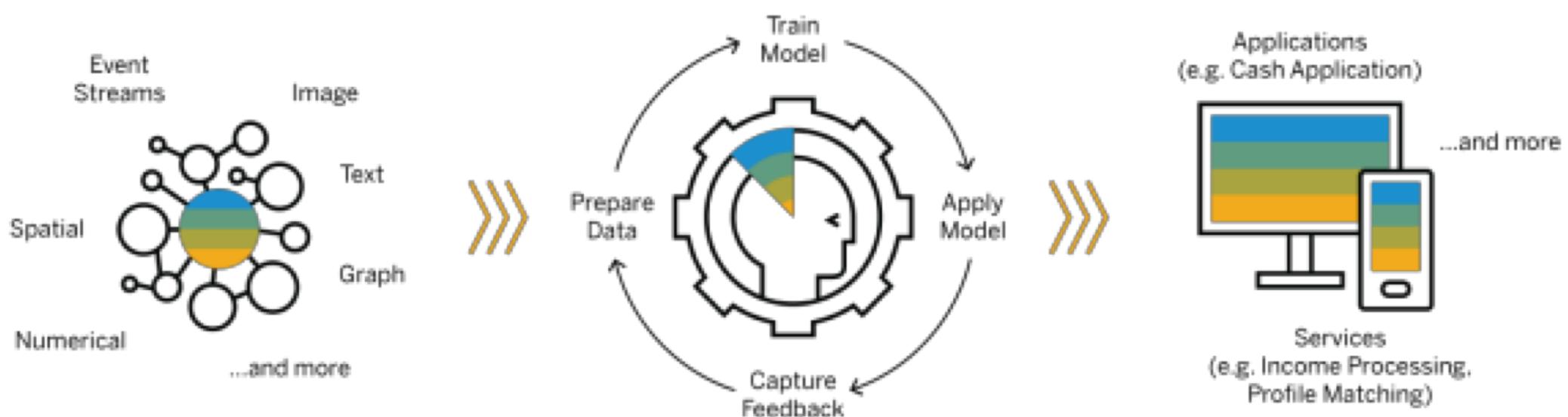


Algorithm selection (Unsupervised Learning)



All models are wrong...but some are useful!

End to End Solution for Building Machine Learning Algorithms



Tools

Machine Learning Options on Azure

Pre-Built



BOT SERVICE



COGNITIVE SERVICES

Custom



AZURE ML Studio

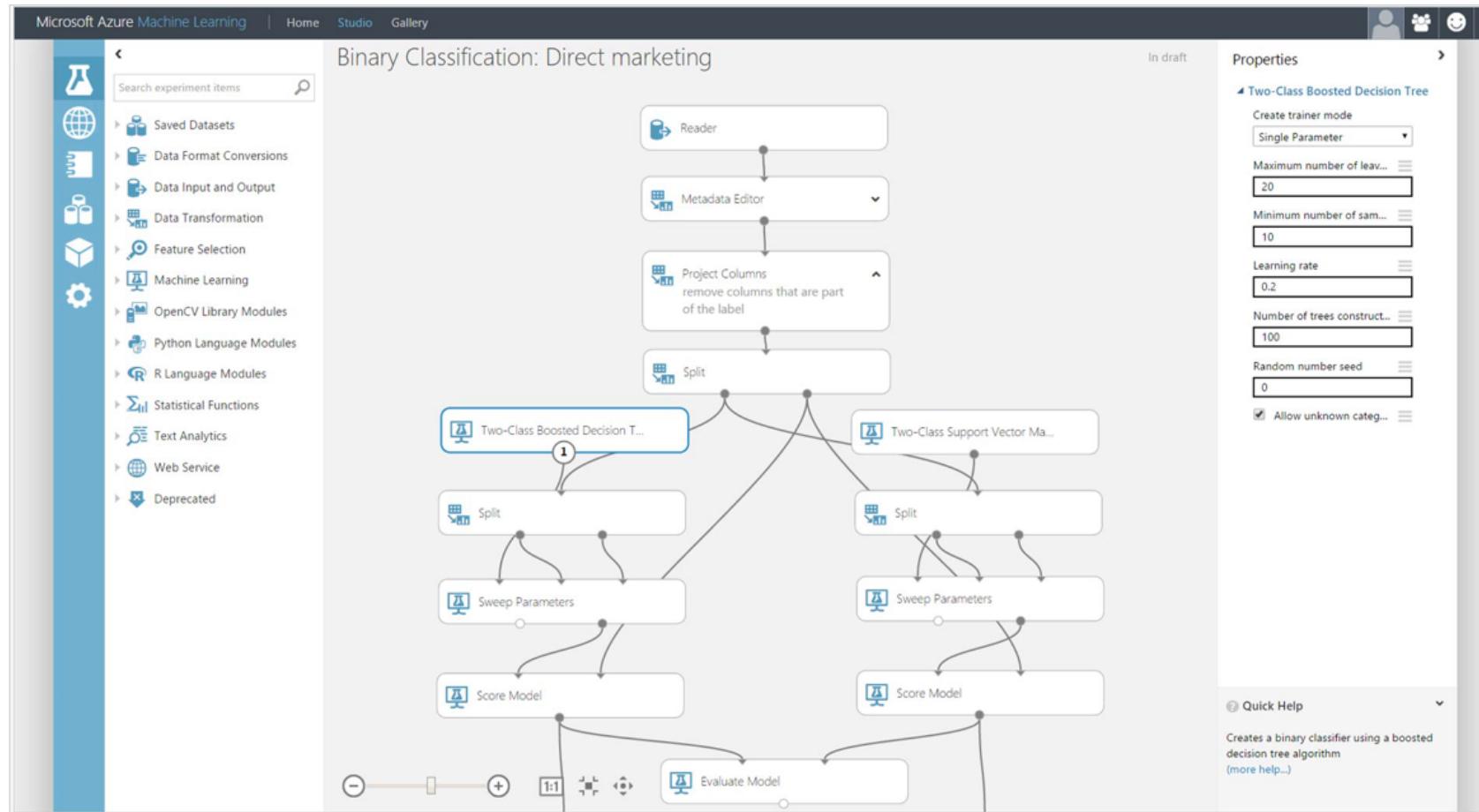


ML SERVER



AZURE
DATABRICKS

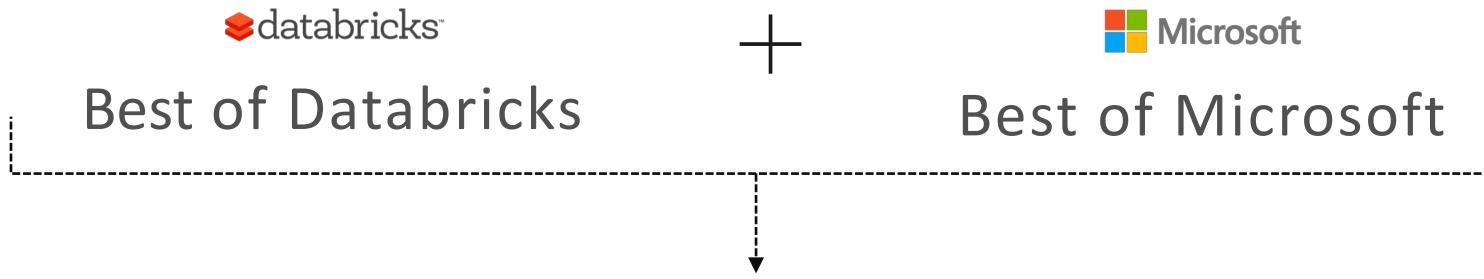
Azure Machine Learning Studio



VISUAL DRAG-AND-DROP

What is Azure Databricks?

A fast, easy and collaborative Apache® Spark™ based analytics platform optimized for Azure

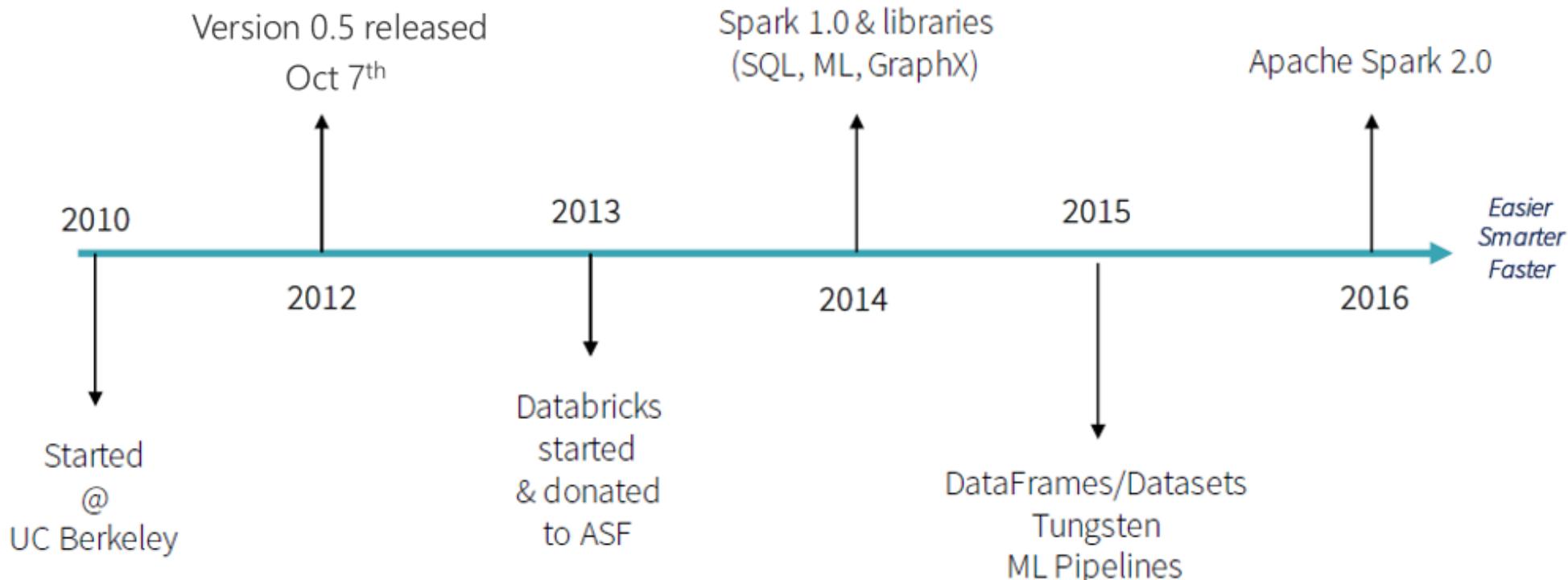


-  Designed in collaboration with the founders of Apache Spark
-  One-click set up; streamlined workflows
-  Interactive workspace that enables collaboration between data scientists, data engineers, and business analysts.
-  Native integration with Azure services (Power BI, SQL DW, Cosmos DB, Blob Storage)
-  Enterprise-grade Azure security (Active Directory integration, compliance, enterprise-grade SLAs)

SPARK IN ONE SENTENCE

A group of engineers started the Spark project with the singular goal to “democratize” the “superpower” of big data, by offering high-level APIs and a unified engine to do machine learning, ETL, streaming and interactive SQL.

SPARK: A BRIEF HISTORY

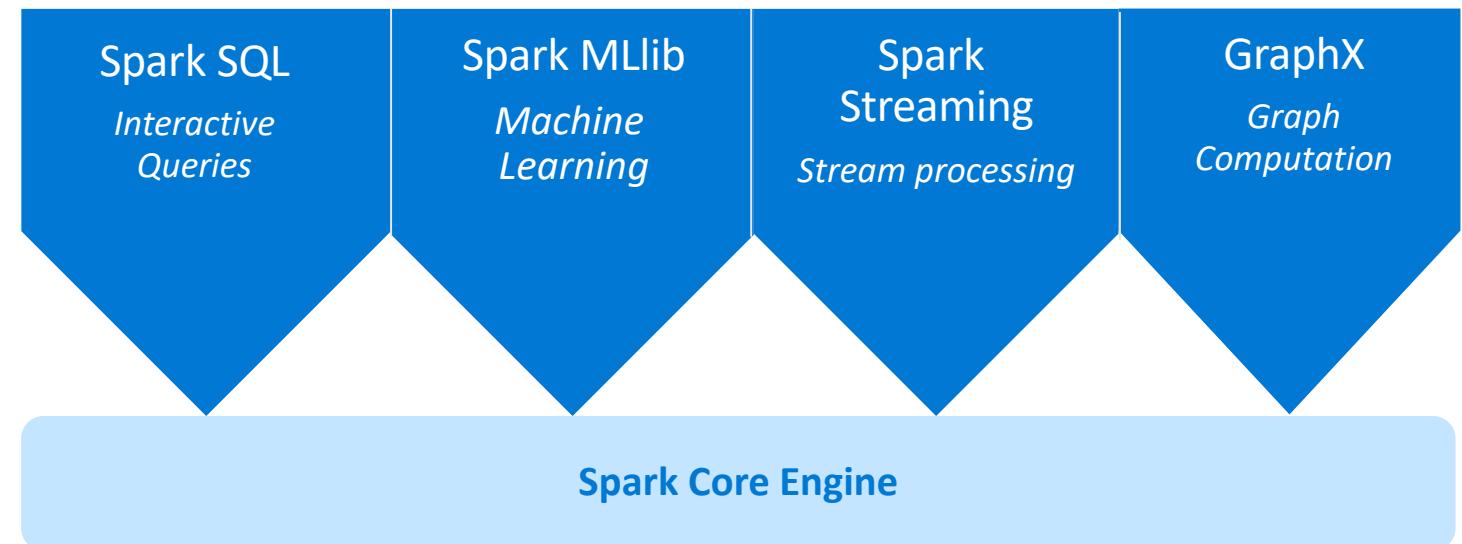


A P A C H E S P A R K

An unified, open source, parallel, data processing framework for Big Data Analytics

Spark Unifies:

- Batch Processing
- Interactive SQL
- Real-time processing
- Machine Learning
- Deep Learning
- Graph Processing



Azure Databricks key audiences & benefits



Data scientist

Integrated workspace

Easy data exploration

Collaborative experience

Interactive dashboards

Faster insights

- Best Spark & serverless

- Databricks managed Spark



Data engineer

Improved ETL performance

- Zero management clusters, serverless

Easy to schedule jobs

Automated workflows

Enhanced monitoring & troubleshooting

- Automated alerts & easy access to logs

Zero Management Spark

Cluster democratization (serverless)



CDO, VP of analytics

Fast, collaborative analytics platform accelerating time to market

No dev-ops required

Enterprise grade security

- Encryption
- End-to-end auditing
- Role-based control
- Compliance

Unified analytics platform

Workshop