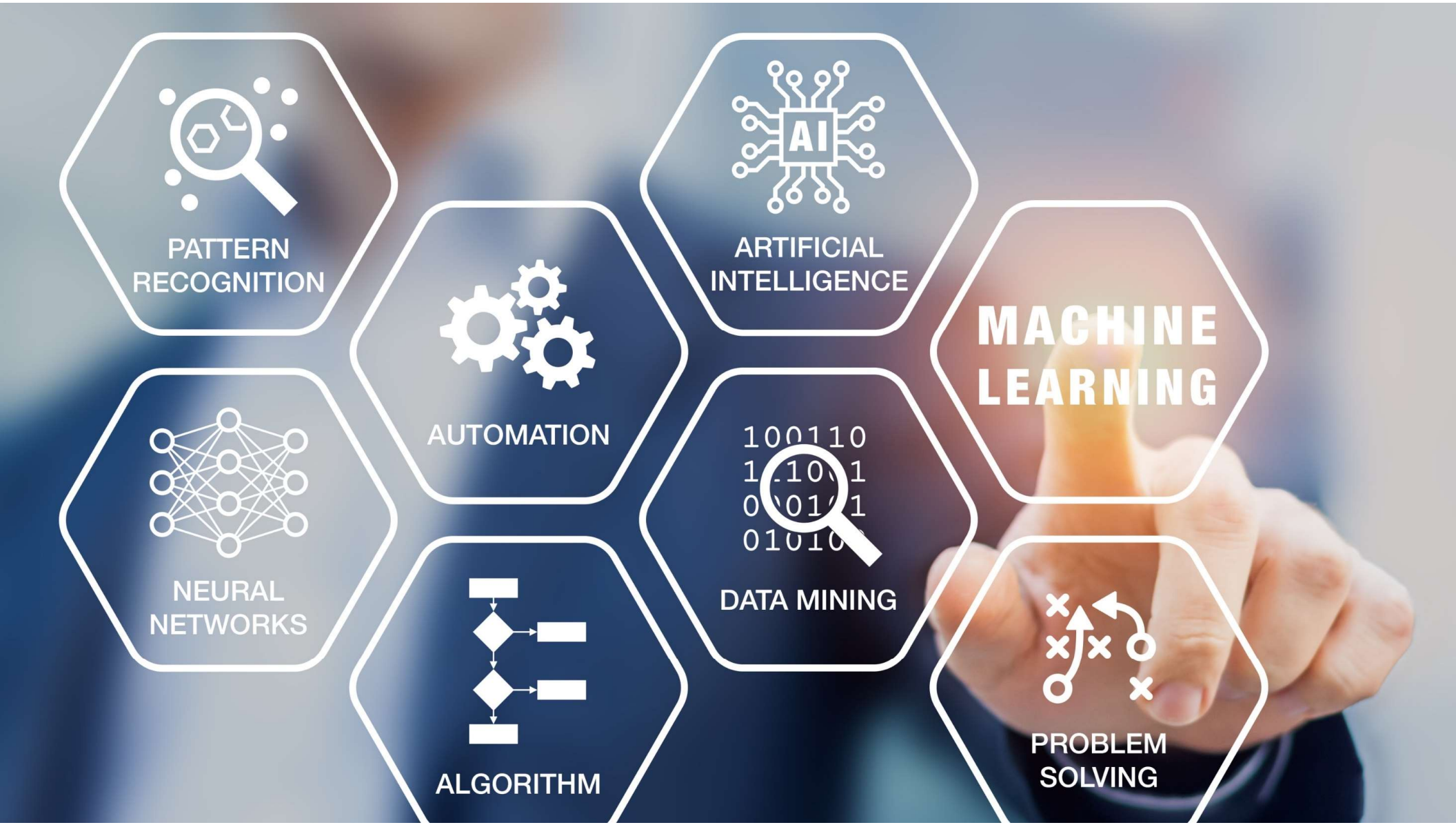
Lecture 2 – Intro to ML Fundamentals

CPE 393 Machine Learning Operations

COMPUTER ENGINEERING

King Mongkut's University of Technology Thonburi

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี


PATTERN RECOGNITION
AI
ARTIFICIAL INTELLIGENCE
MACHINE LEARNING
AUTOMATION
NEURAL NETWORKS
100110
DATA MINING
ALGORITHM
PROBLEM SOLVING

Consumer ML

Face Detection

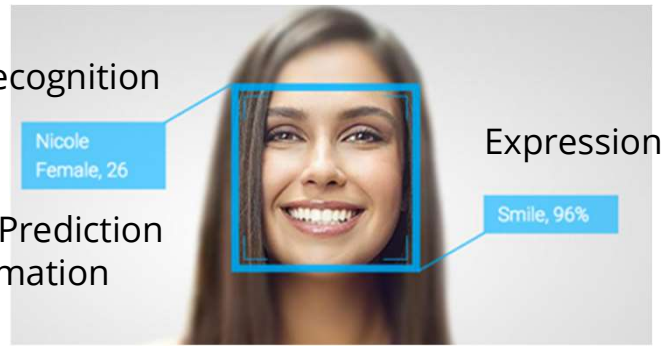
Face Recognition

Nicole
Female, 26

Expression Recognition

Smile, 96%

Gender Prediction
Age Estimation



Smart Home

Voice Interface

What exactly is machine learning?

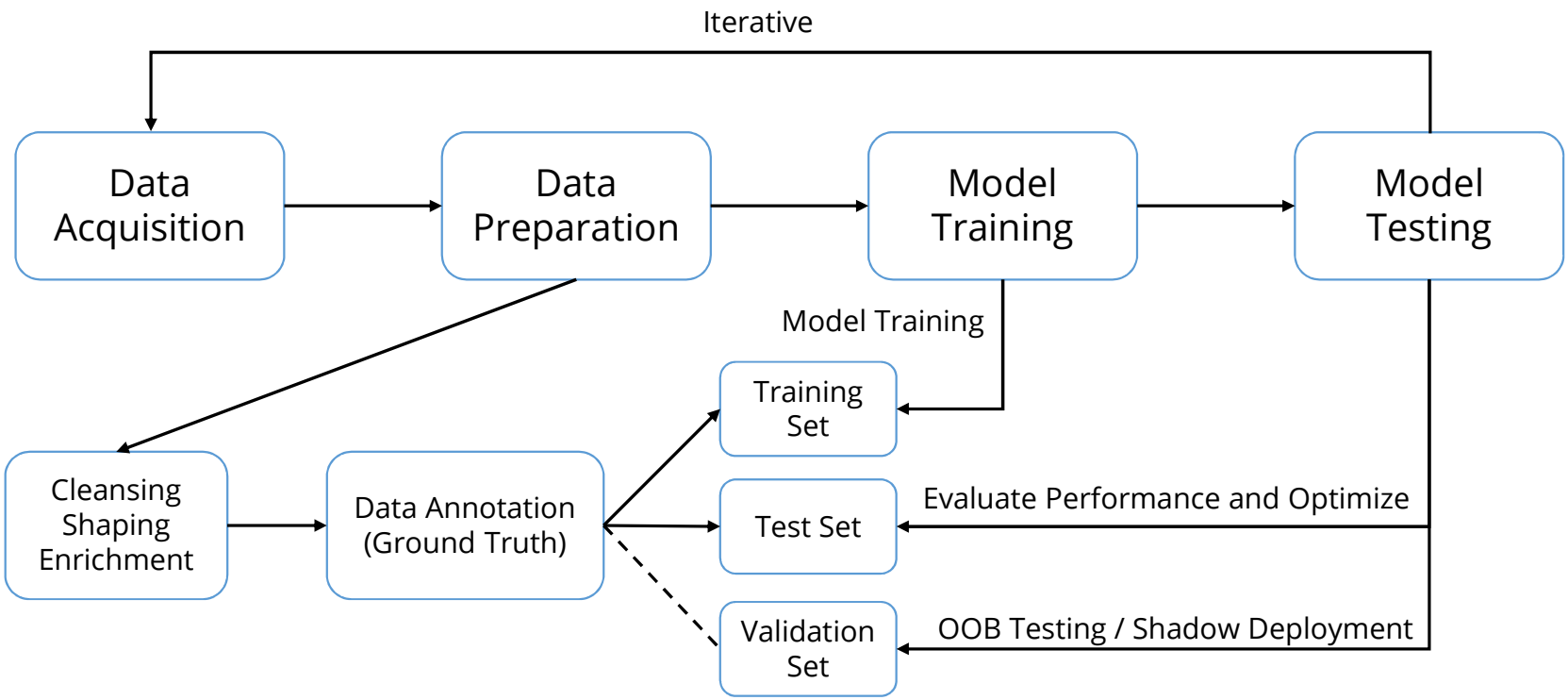
- Machine learning (ML) is a method of data analysis that automates analytical model building.

- Using algorithms that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look.

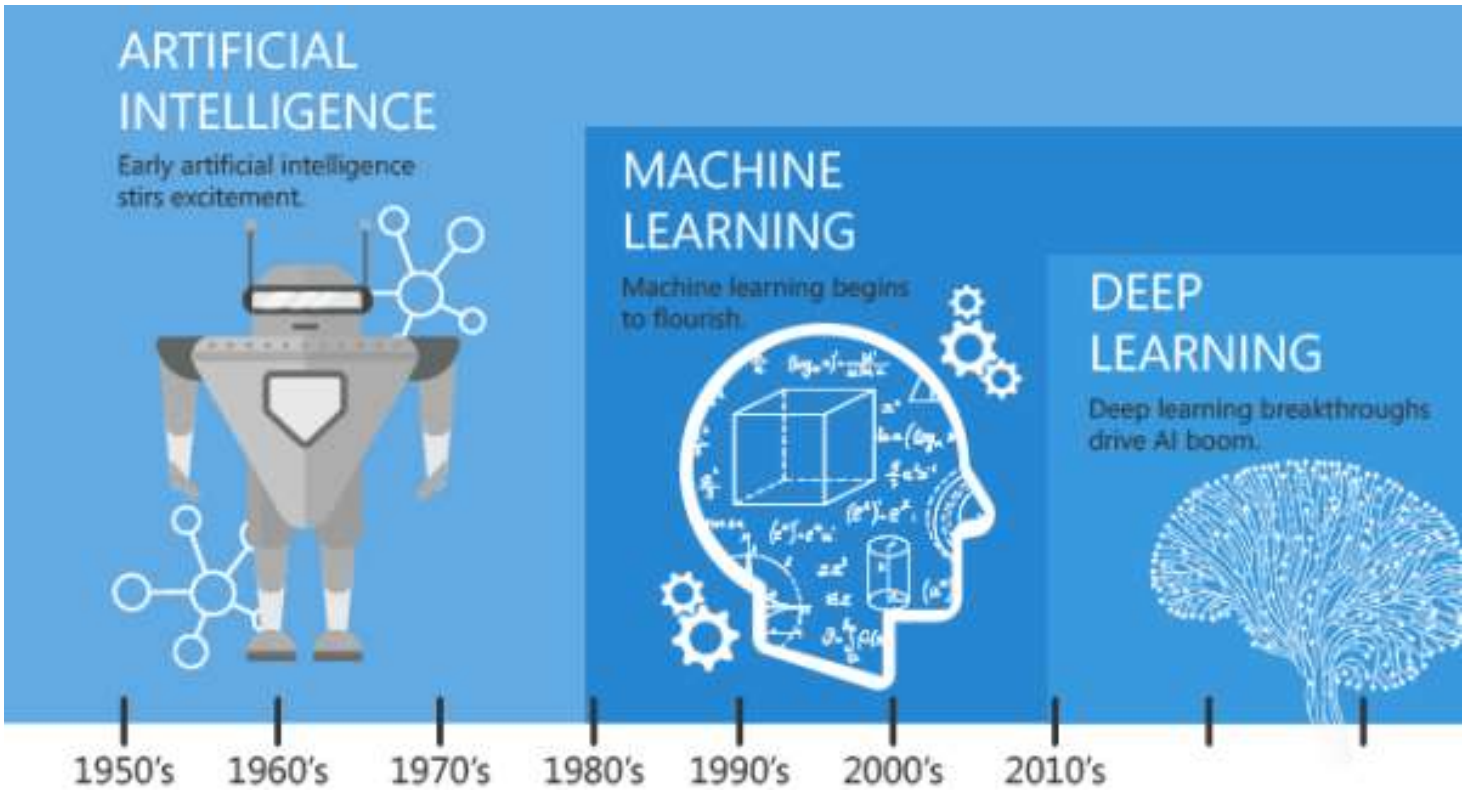
Why machine learning matter

- Automatic: Train it once and it can be run automatically
- Fast: With big data, work faster than human
- Accurate: Can predict groups more accurately than manual methods
- Scale: Able to handle large data

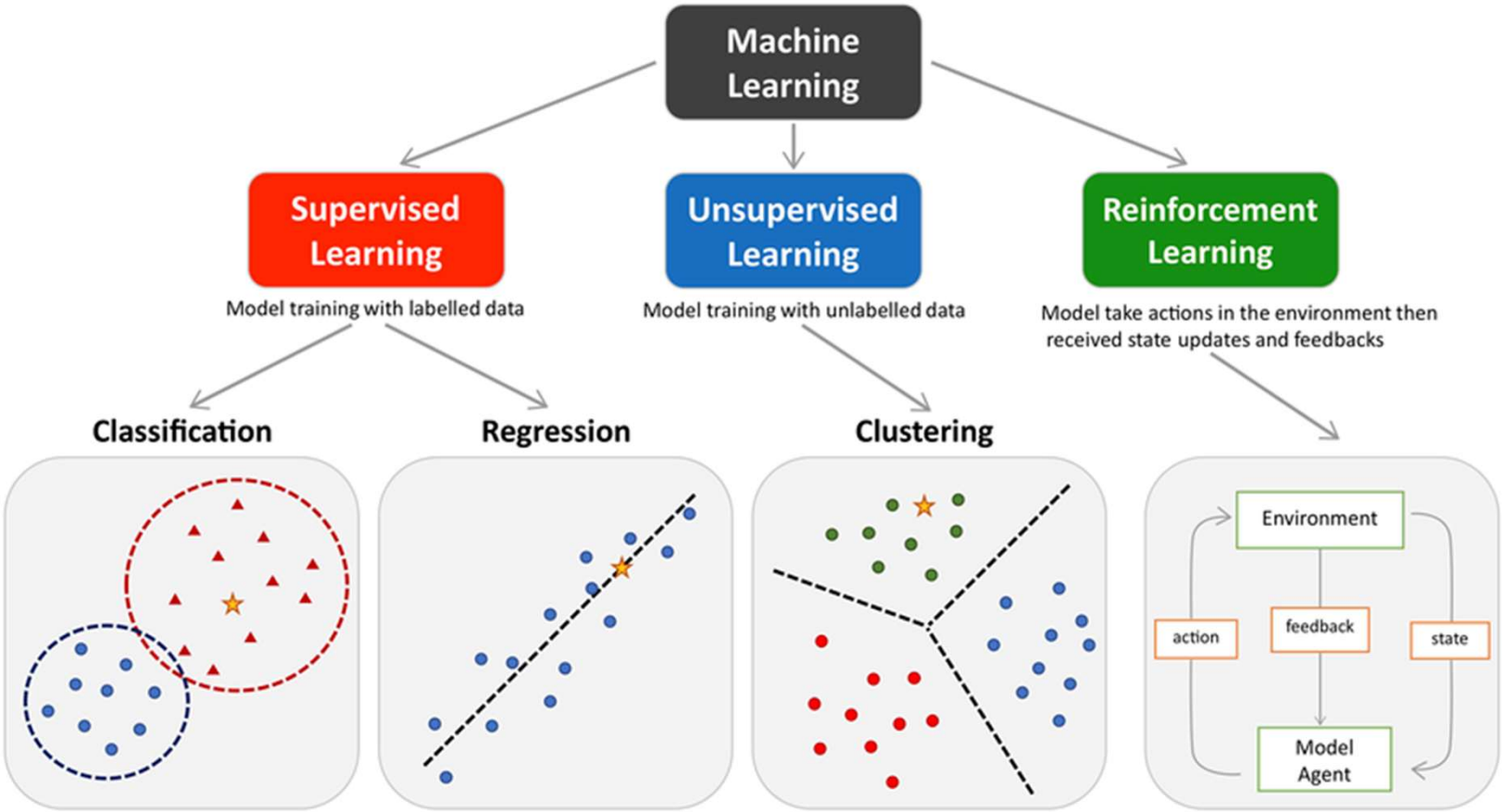
Machine Learning Model Development Workflow



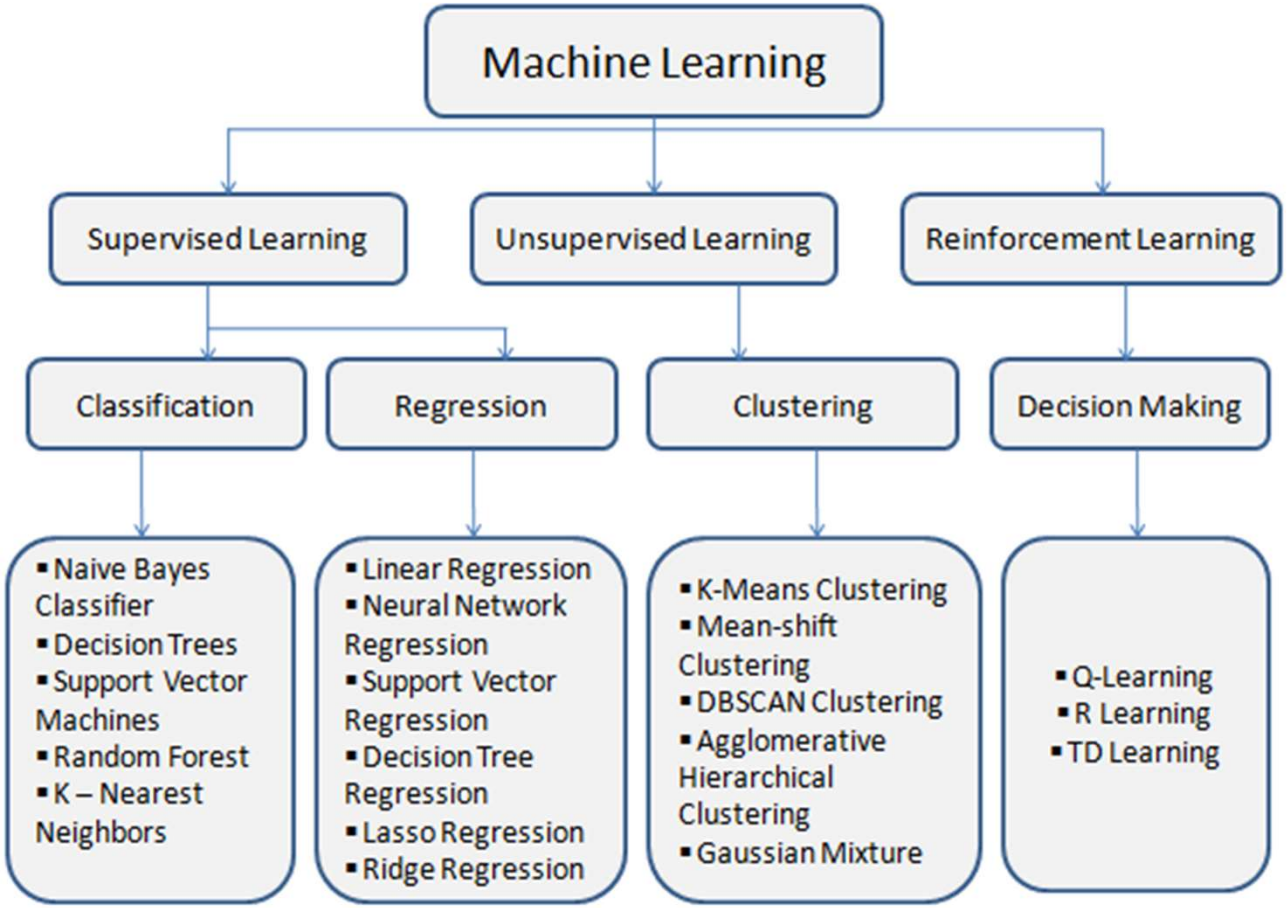
Timeline


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

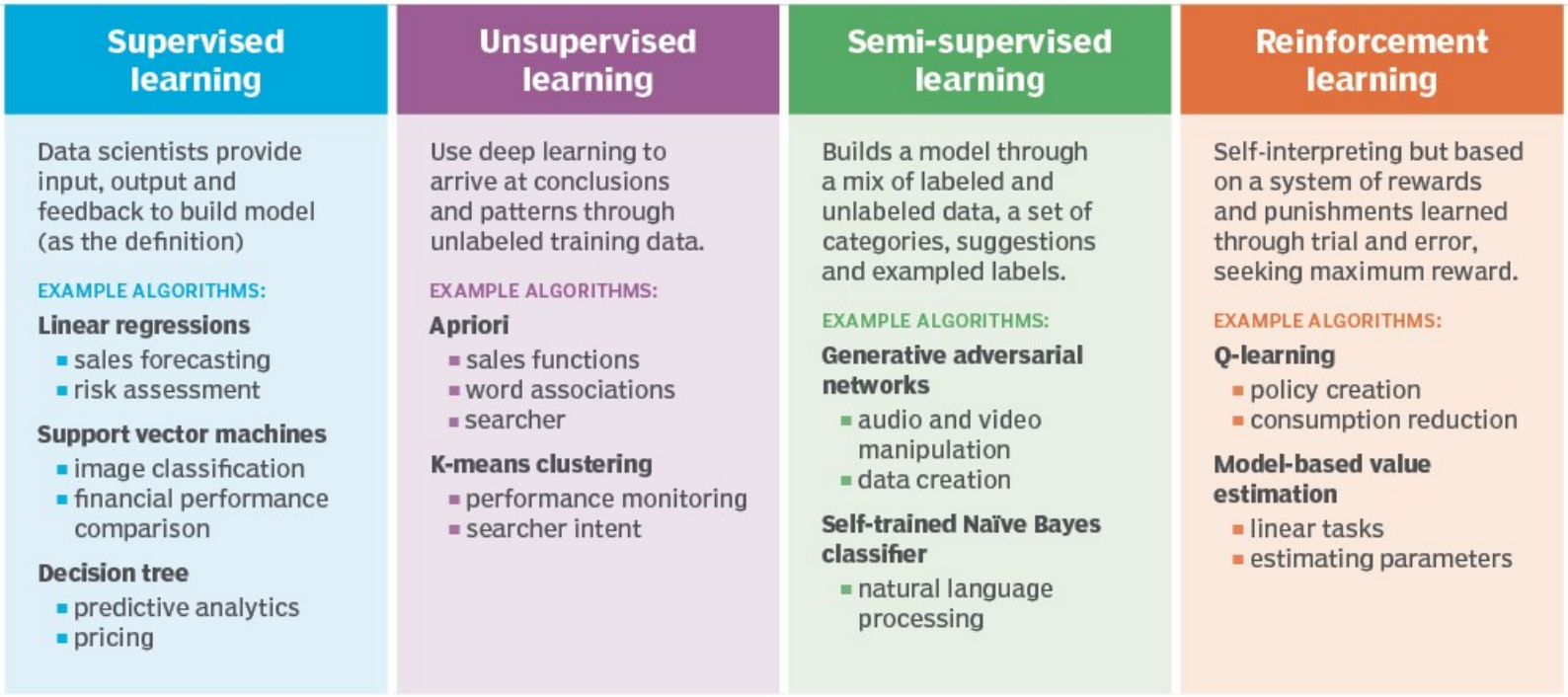
Type of ML


Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Model training with labelled data
Model training with unlabelled data
Model take actions in the environment then received state updates and feedbacks
Classification
Regression
Clustering
Environment
action
feedback
state
Model Agent

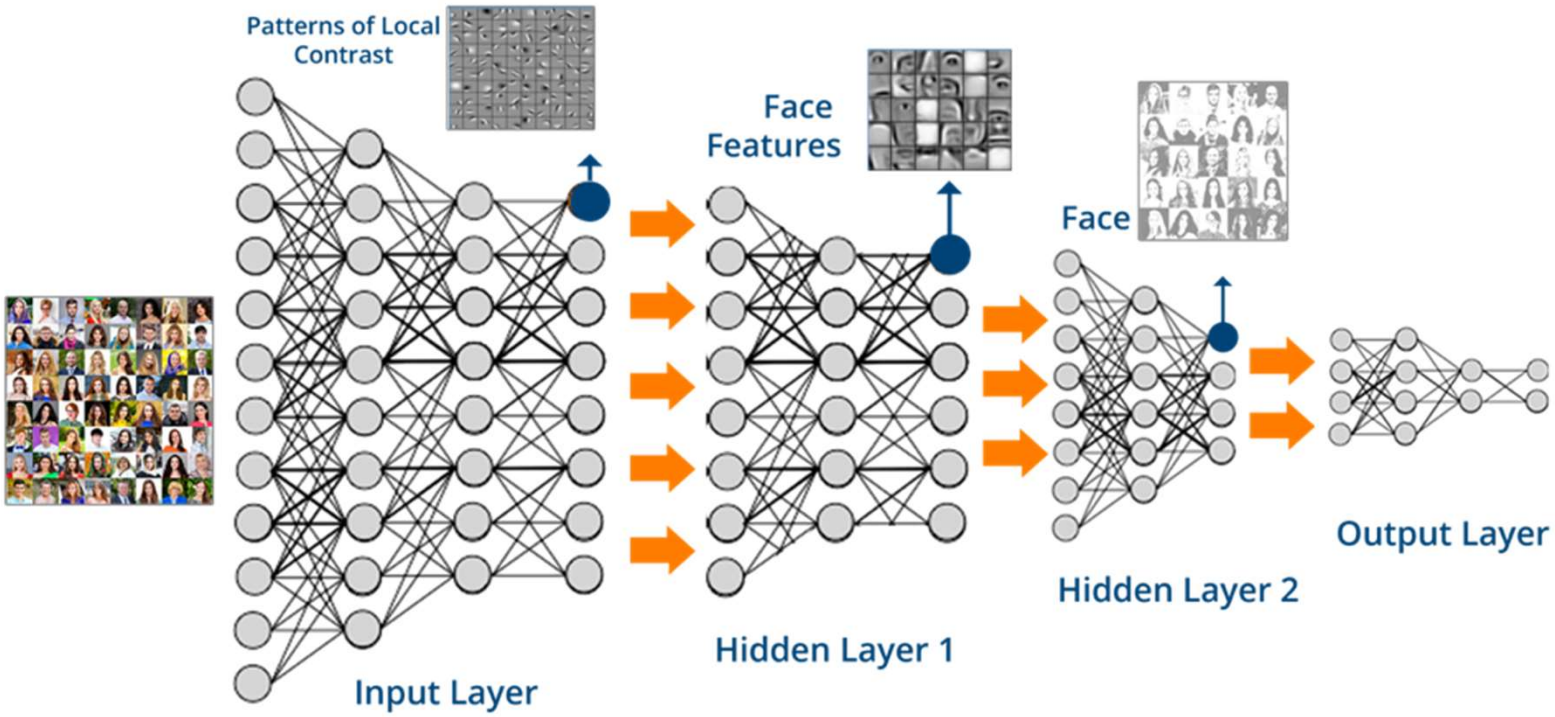
Type of ML


Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Classification
Regression
Clustering
Decision Making
▪ Naive Bayes Classifier
▪ Decision Trees
▪ Support Vector Machines
▪ Random Forest
▪ K – Nearest Neighbors
▪ Linear Regression
▪ Neural Network Regression
▪ Support Vector Regression
▪ Decision Tree Regression
▪ Lasso Regression
▪ Ridge Regression
▪ K-Means Clustering
▪ Mean-shift Clustering
▪ DBSCAN Clustering
▪ Agglomerative Hierarchical Clustering
▪ Gaussian Mixture
▪ Q-Learning
▪ R Learning
▪ TD Learning

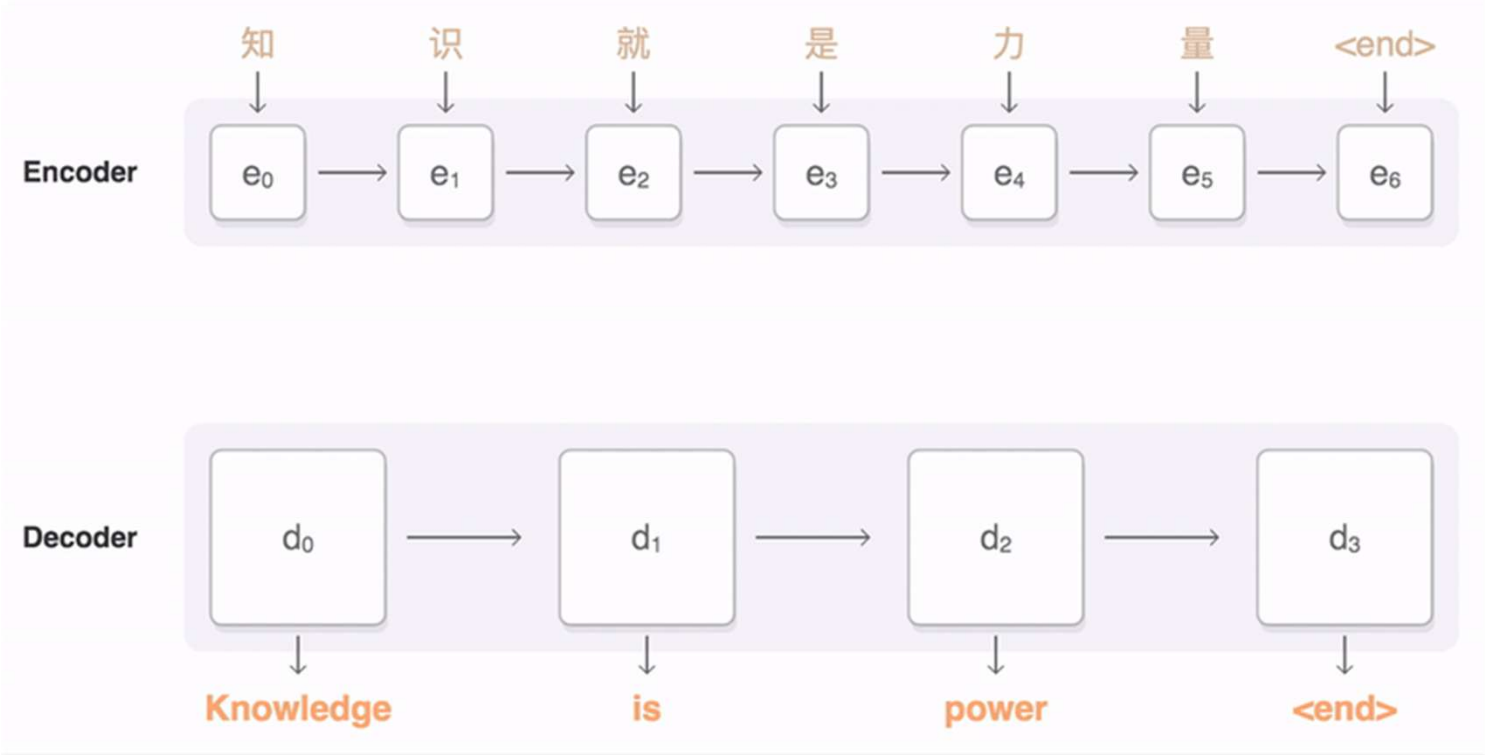
Style of Learning



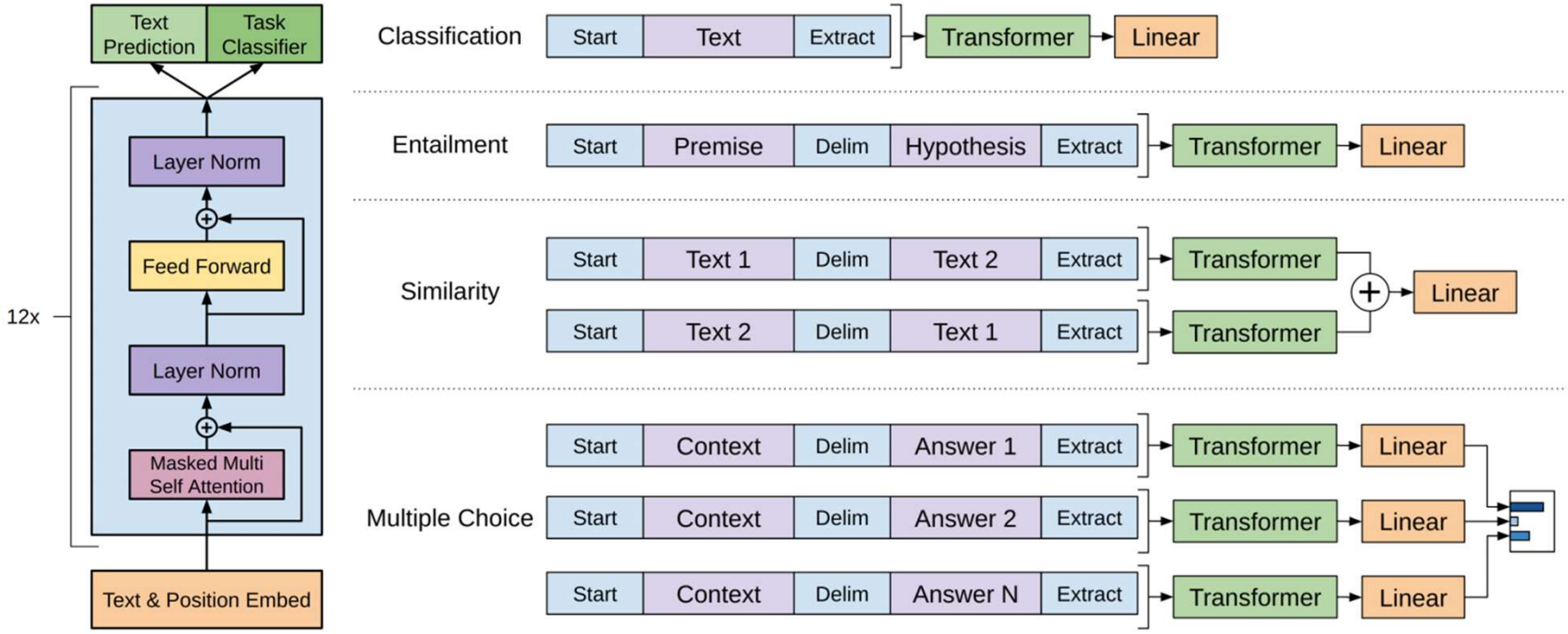
Advanced Machine Learning



Advanced Machine Learning



Advanced Machine Learning

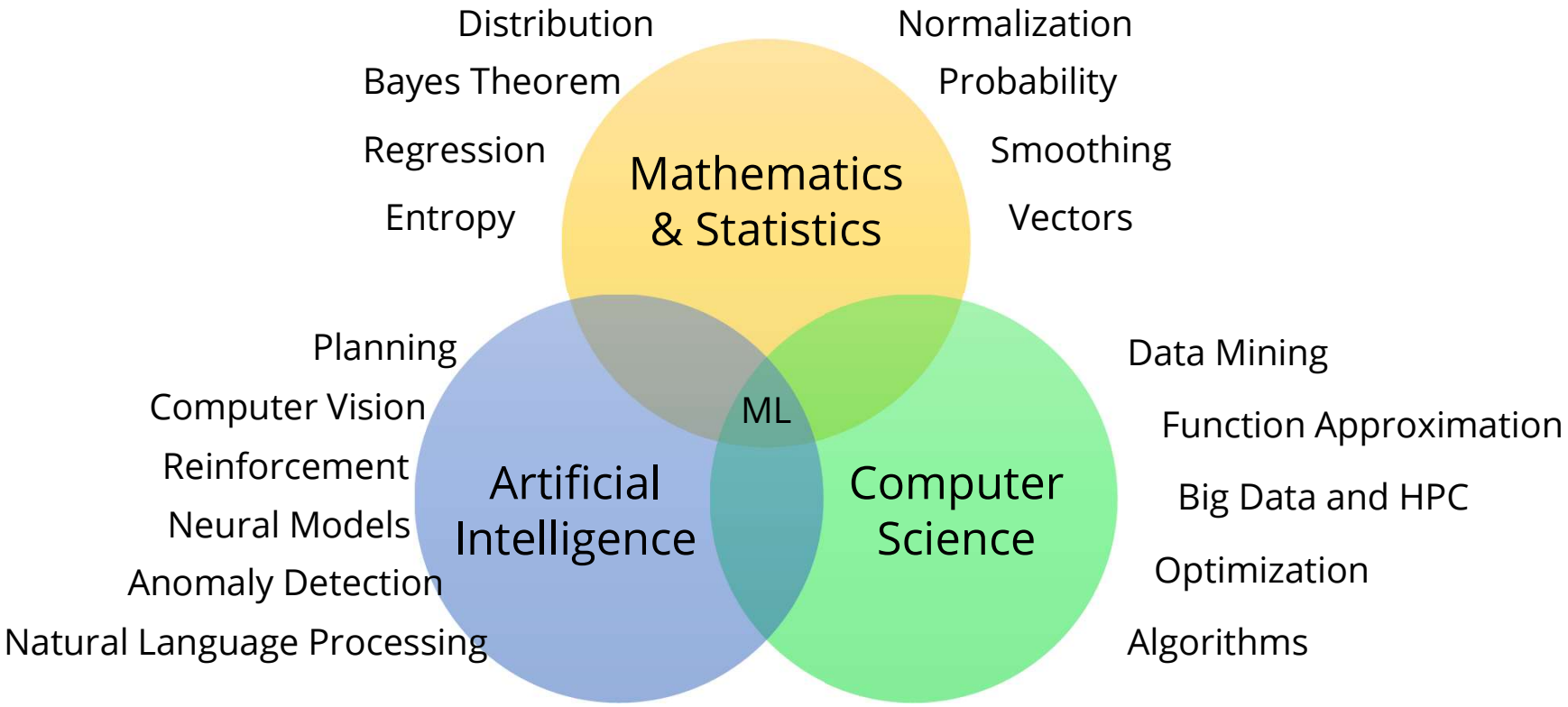


https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

Known issues

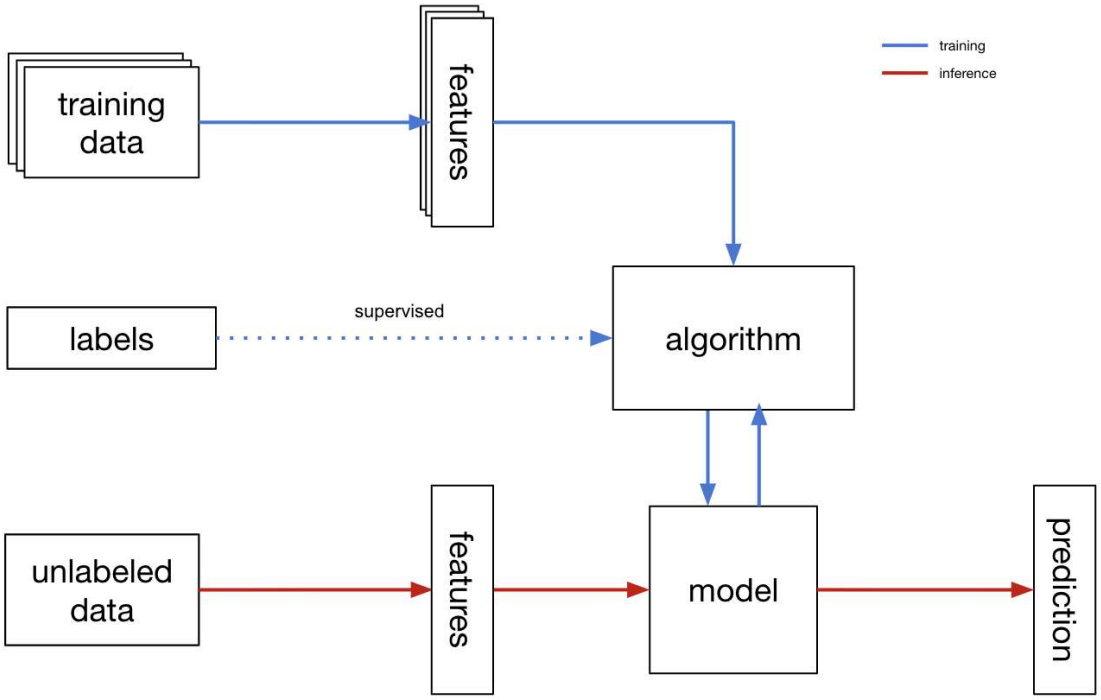
- Overfitting
- Specialized learning vs generalized learning
- Unbalanced data
- Noisy (uncleaned) data
- Optimization

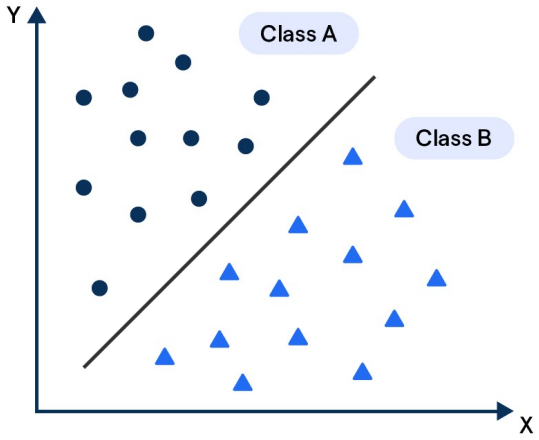
Foundations



Classification Model

Supervised Learning





Classification Model

Supervised Learning

Slide

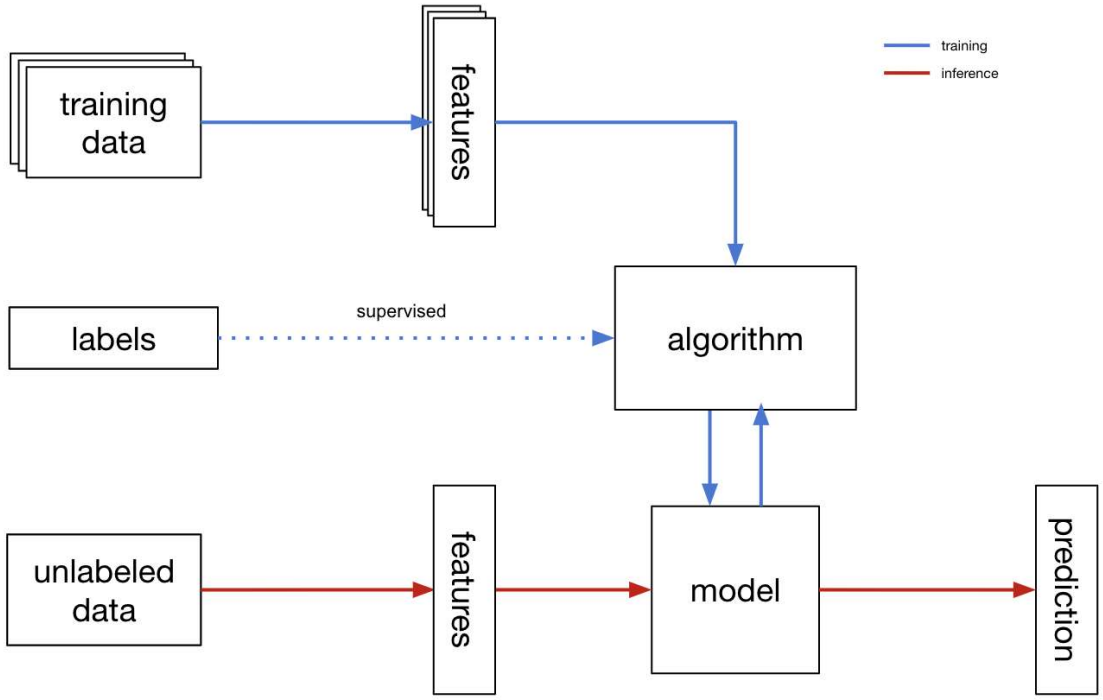
https://github.com/santitham/data-science/blob/main/lectures/8-classification-with-decision-tree.pdf

CoLab

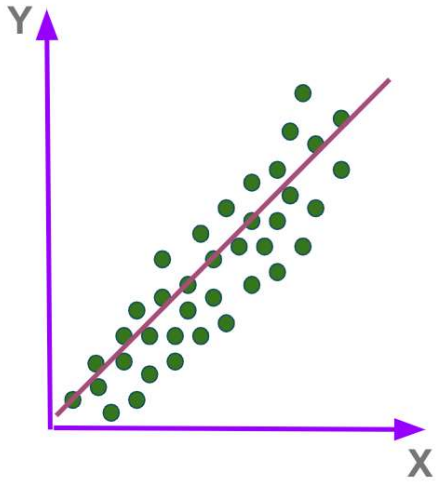
https://colab.research.google.com/drive/1cvP80R1XhTRYG1Jx33wosLCk23UumyOJ#scrollTo=fYenyFzA9cdZ

Regression Model

Supervised Learning



$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_q x_q + \varepsilon$$



Classification Model

Supervised Learning

Slide

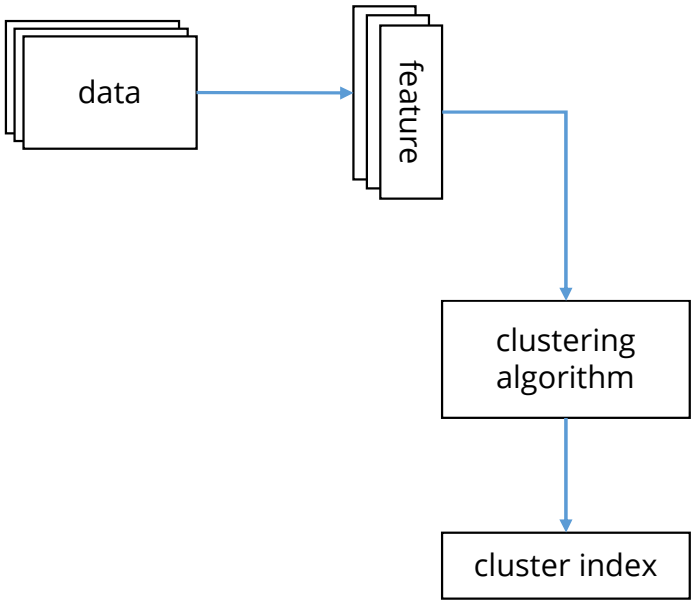
https://github.com/santitham/data-science/blob/main/lectures/7-linear-regression.pdf

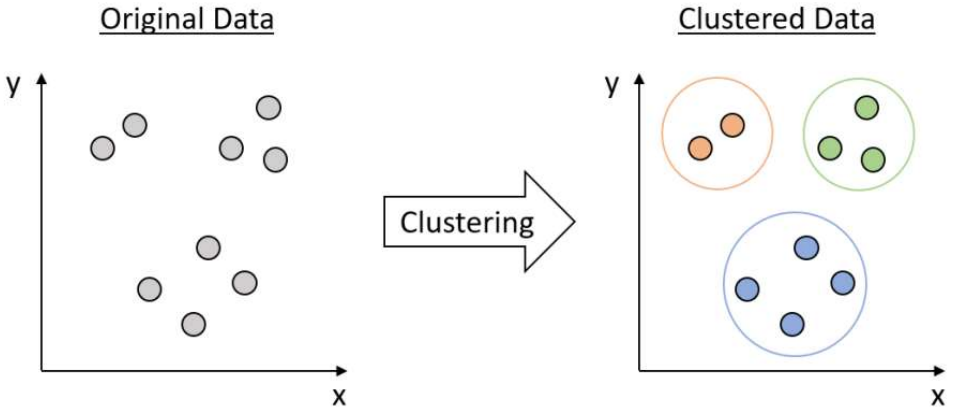
CoLab

https://colab.research.google.com/drive/10ByXWKbpIyoqnWVKBNwjwMz3wUZGsCNy

Clustering

Unsupervised Learning





Clustering
Unsupervised Learning

Slide

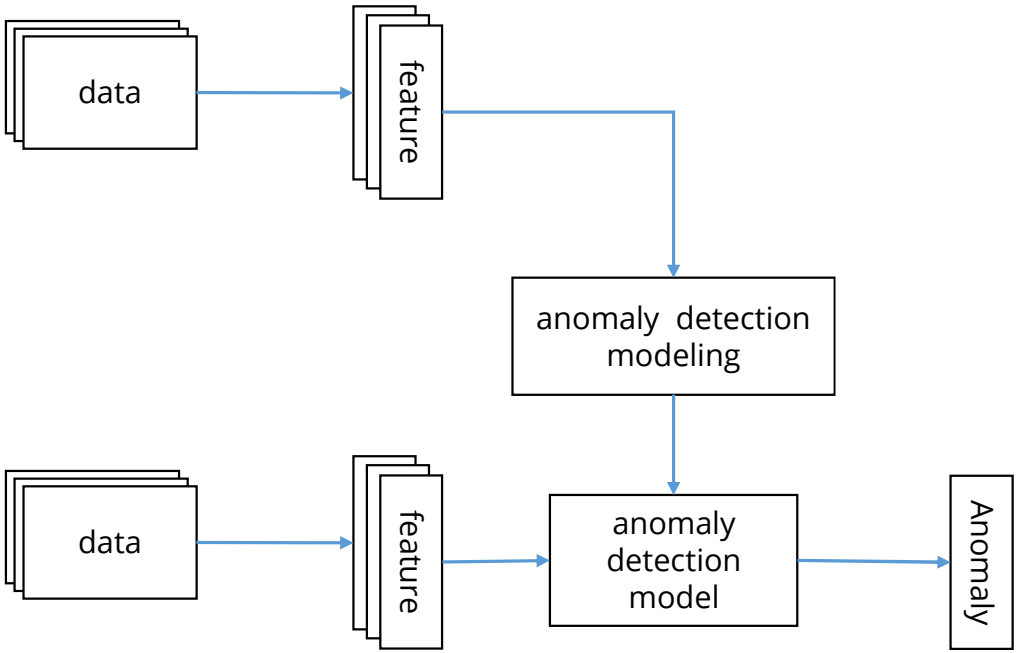
https://github.com/santitham/data-science/blob/main/lectures/10-clustering.pdf

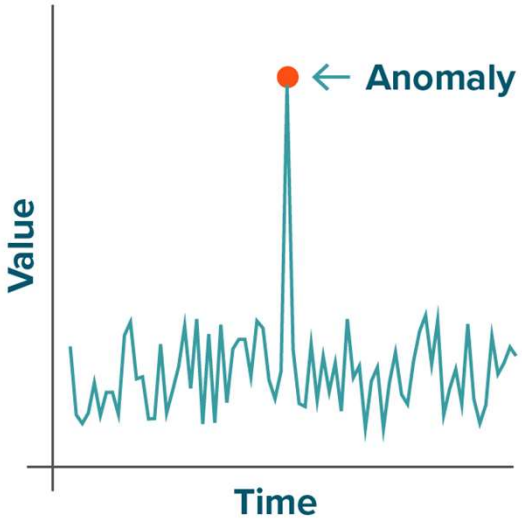
CoLab

https://colab.research.google.com/drive/1Ts03orFSjtPhbCABcsAbBNWjV3Ncf2Zo?usp=sharing

Anomaly detection

Unsupervised Learning





Anomaly Detection

Unsupervised Learning

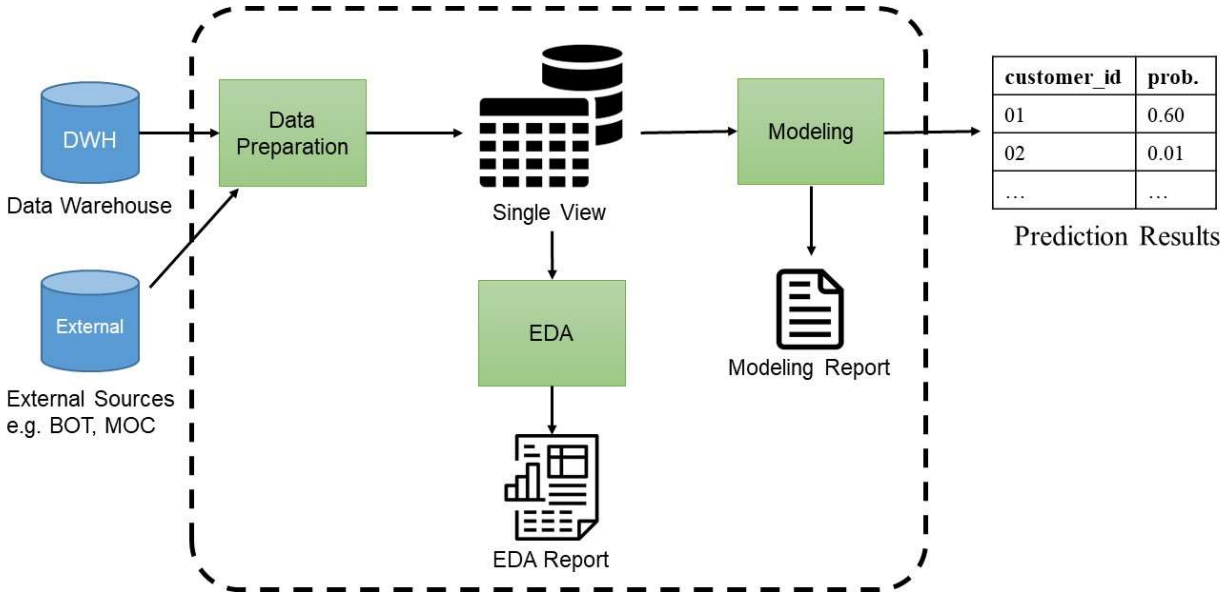
Slide

https://github.com/santitham/data-science/blob/main/lectures/12-anomaly-detection.pdf

CoLab

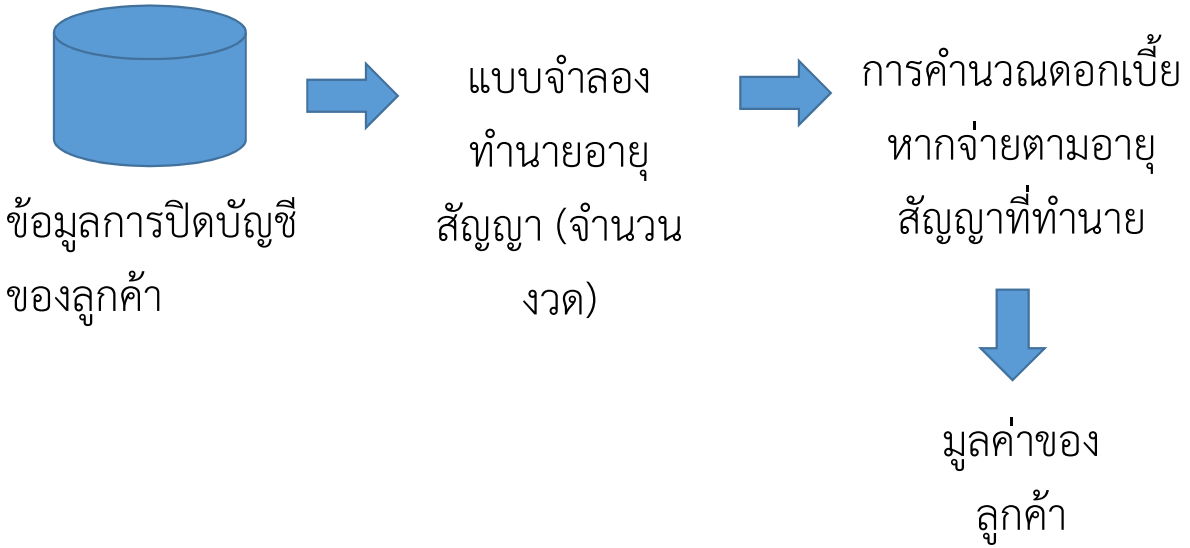
https://colab.research.google.com/drive/1jLNQB5uhxmq7-Auy-79XSDWIcoZLUded

Use Case:
Next Best Product Offering

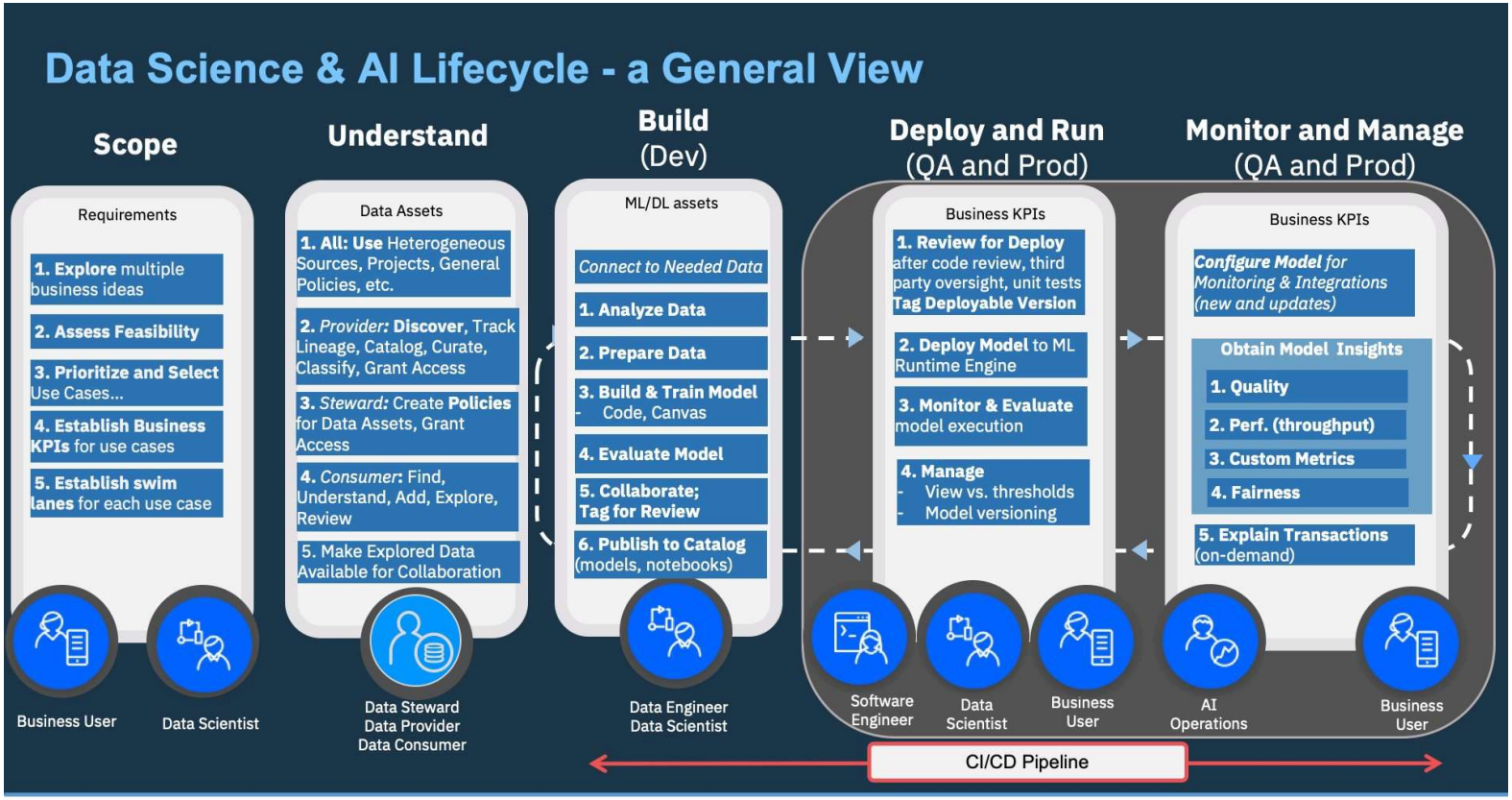


Data: Demographics (all), Previous Contracts (all), Payments (all), Time, etc. | Target: Added Contract

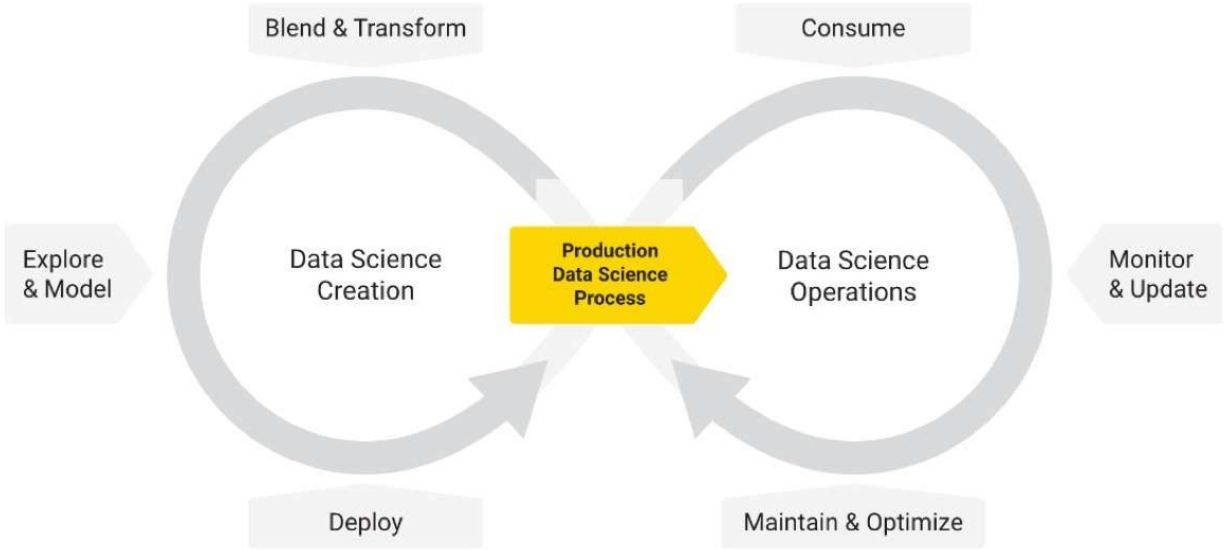
Use Case:
Loan Customer Lifetime Value Prediction



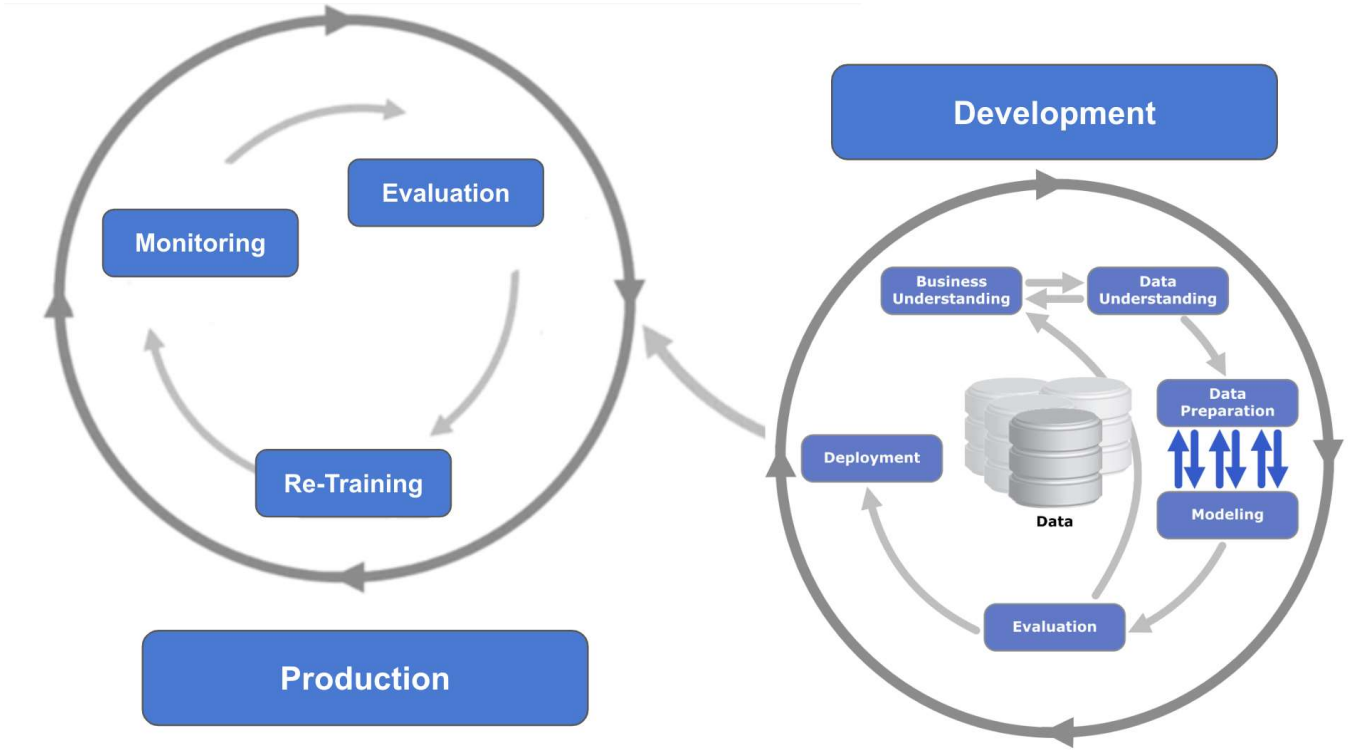
Deployment


Data Science & AI Lifecycle - a General View
Scope
Understand
Build
(Dev)
Deploy and Run
(QA and Prod)
Monitor and Manage
(QA and Prod)
Requirements
1. Explore multiple business ideas
2. Assess Feasibility
3. Prioritize and Select Use Cases...
4. Establish Business KPIs for use cases
5. Establish swim lanes for each use case
Data Assets
1. All: Use Heterogeneous Sources, Projects, General Policies, etc.
2. Provider: Discover, Track Lineage, Catalog, Curate, Classify, Grant Access
3. Steward: Create Policies for Data Assets, Grant Access
4. Consumer: Find, Understand, Add, Explore, Review
5. Make Explored Data Available for Collaboration
ML/DL assets
Connect to Needed Data
1. Analyze Data
2. Prepare Data
3. Build & Train Model
- Code, Canvas
4. Evaluate Model
5. Collaborate; Tag for Review
6. Publish to Catalog (models, notebooks)
Business KPIs
1. Review for Deploy after code review, third party oversight, unit tests Tag Deployable Version
2. Deploy Model to ML Runtime Engine
3. Monitor & Evaluate model execution
4. Manage
- View vs. thresholds
- Model versioning
Business KPIs
Configure Model for Monitoring & Integrations (new and updates)
Obtain Model Insights
1. Quality
2. Perf. (throughput)
3. Custom Metrics
4. Fairness
5. Explain Transactions (on-demand)
Business User
Data Scientist
Data Steward
Data Provider
Data Consumer
Data Engineer
Data Scientist
Software Engineer
Data Scientist
Business User
AI Operations
Business User
CI/CD Pipeline

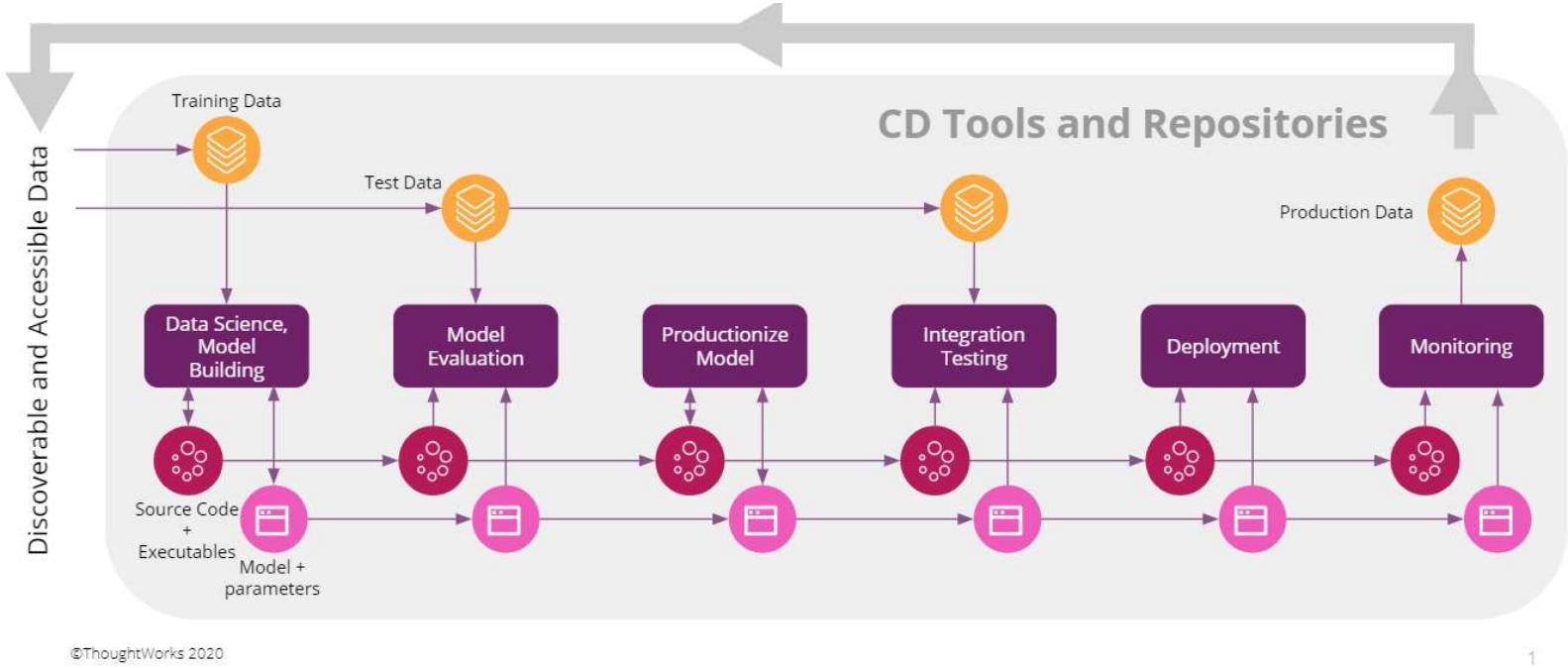
Creation and Operation



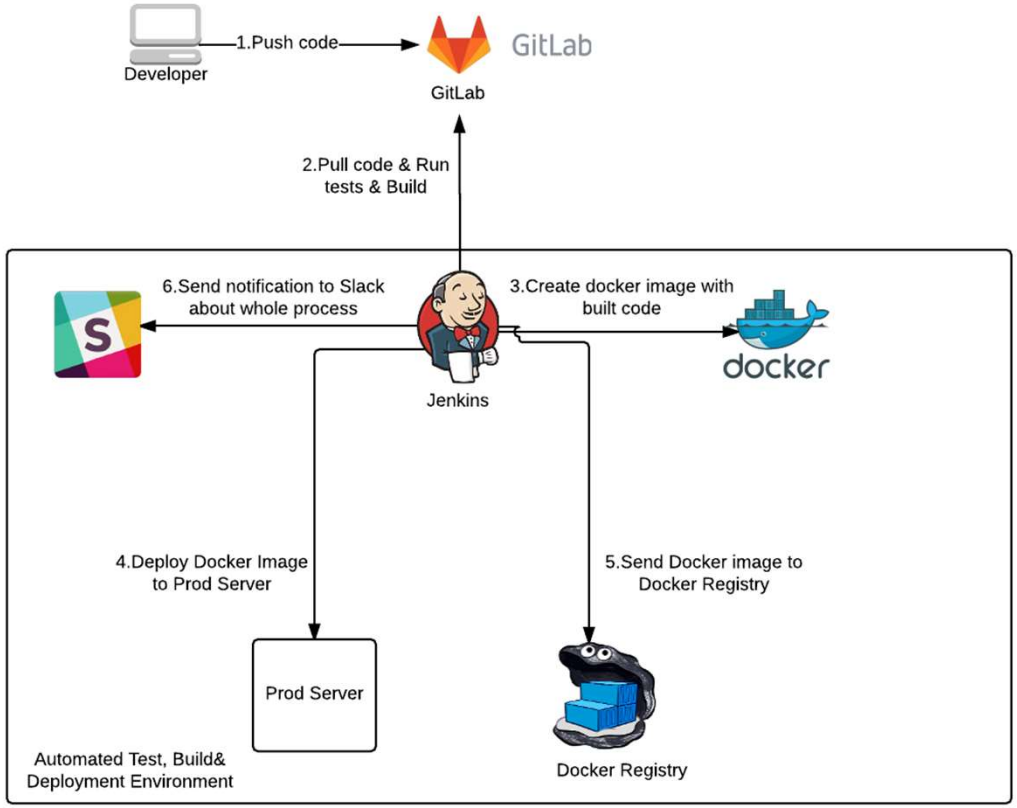
Production environment



Continuous Delivery



Example: Deployment




KMUTT