Analytics Vidhya
Learn everything about analytics

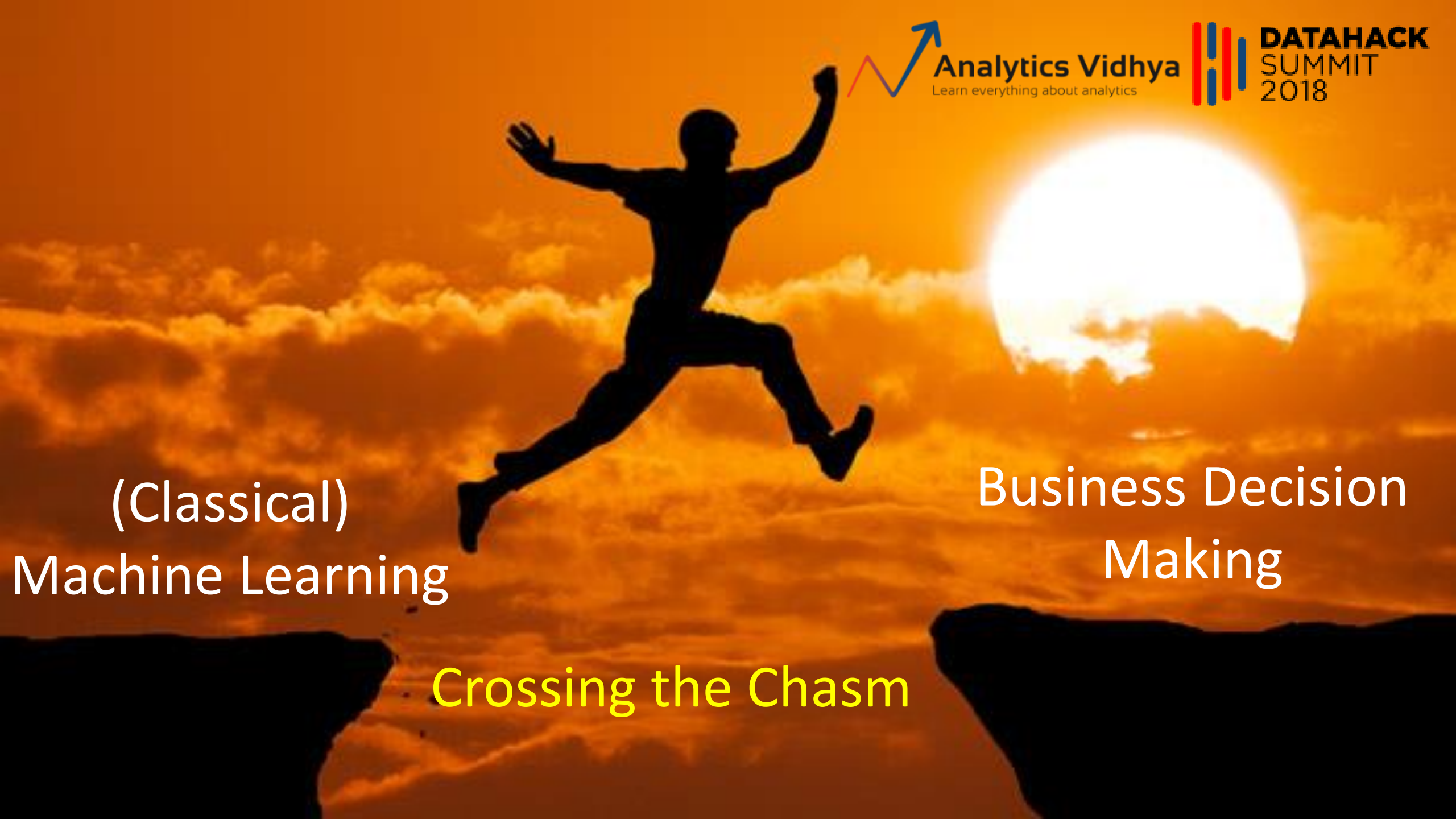
DATAHACK SUMMIT 2018

‘Crossing the Chasm’ – Bridging the Gap Between Machine Learning and Business Decision-Making

Karthikeyan Sankaran

22 – 25 November, 2018

NIMHANS Convention Center, Bengaluru


Analytics Vidhya
Learn everything about analytics
DATAHACK
SUMMIT
2018
(Classical)
Machine Learning
Business Decision
Making
Crossing the Chasm

What is this talk about? (and what it is not!)

What is this talk about:

- ML / DL are fascinating and powerful tools…but
- Significant gap exists in utilizing ML / DL predictions for business decision making
- Introduction to Simulation Modeling
- Explore synergies between Simulation Modeling & Data science techniques

What this talk is not:

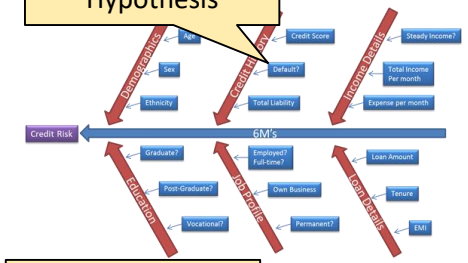
- Does not cover any specific ML / DL / AI techniques
- Does not cover any math / stats concepts
- Does not cover moral & ethical dimensions of data science / AI

GitHub Link: https://github.com/skkeyan-mlai/DHS2018

(Classical) Supervised Machine Learning Pipeline

1 Hypothesis Generation

Formulate Hypothesis

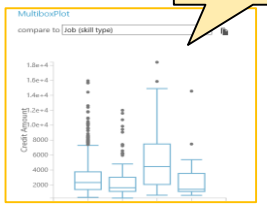


2 Gather Data

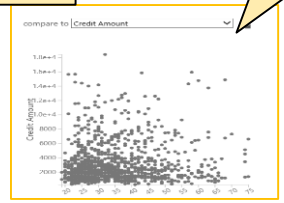
Input Data

3 Exploratory Analysis

Uni-variate Analysis

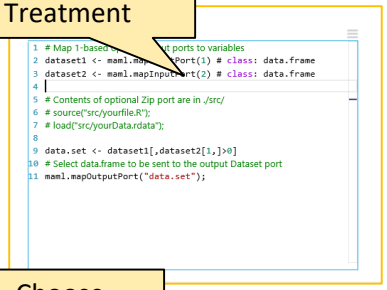


Bi-variate Analysis

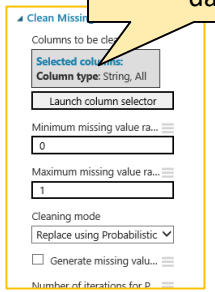


4 Pre-Processing

Outlier Analysis & Treatment

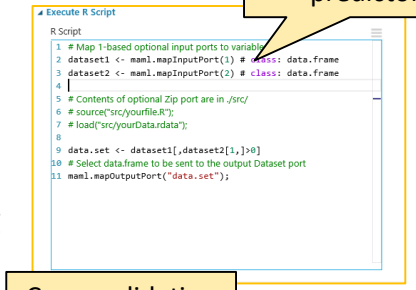


Clean Missing data



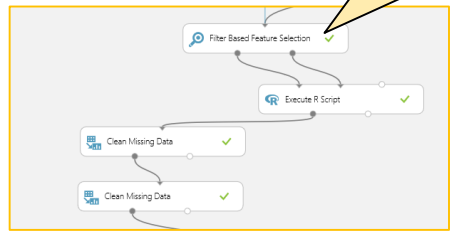
5 Feature Engineering

Generate new predictors



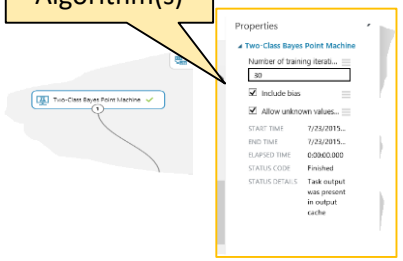
6 Feature Selection

Select final set of predictors



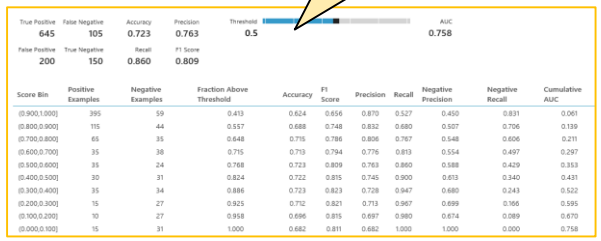
7 Algorithm & Ensembles

Choose Algorithm(s)



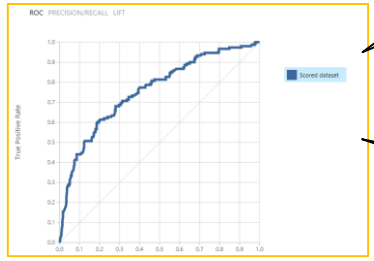
8 Tune Hyper-model Parameters

Cross-validation



9 Output Evaluation

AUC: 0.758



Reduce Bias & Variance

WE ALL LOVE THIS AND IT HAS MANY USE CASES IN THE INDUSTRY …

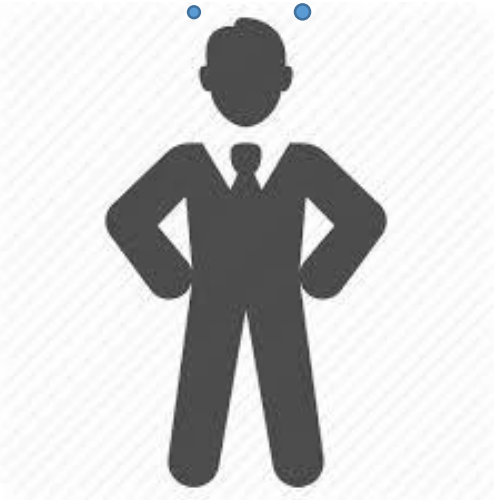
‘Partying’ Data Scientists vs ‘So What’ Business Guy

Stacked Ensemble has an accuracy of 97% without overfitting

DL Network with 25 hidden layers beat the existing benchmarks

Business Value?? Not so sure

Do they understand the business complexity, the context?

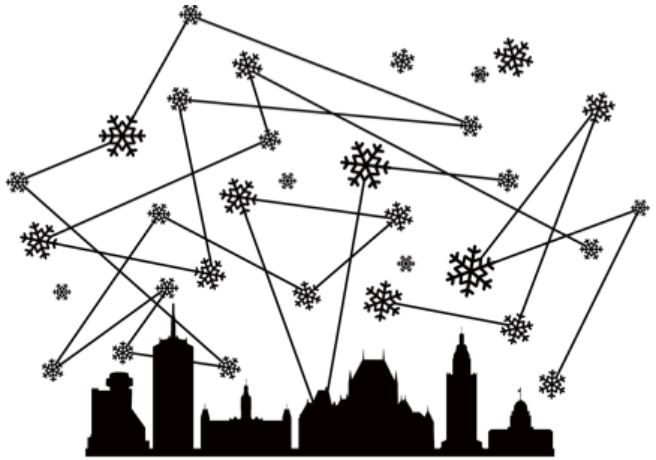


Data Scientists

Business Decision Maker

Decision Making 101

Real World (Business)



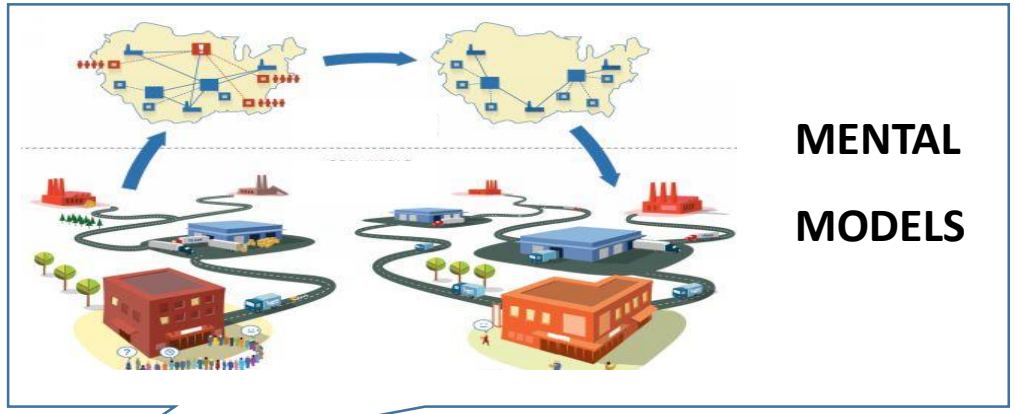
- Multiple Entities
- Complex Non-Linear Relationships
- Delayed Feedback
- ‘Black Swan’ Events

DECISIONS

EXPERTISE

- ✓ Education
- ✓ Training
- ✓ Experience

Business Decision Maker



MENTAL MODELS

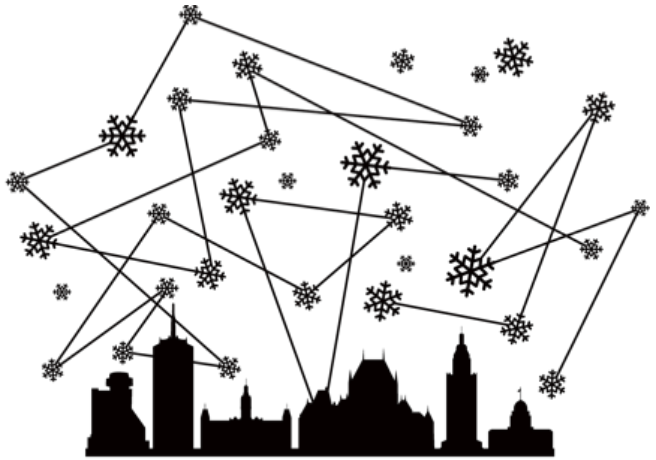
Approximation of the Real World

Characteristics of Mental Models

- Process Centric
- Cause & Effect based
- Estimates risk & uncertainty
- Biased
- Rigid (Relatively difficult to change)

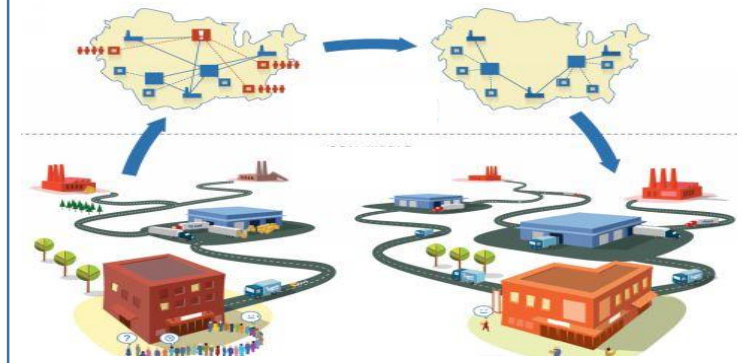
Goal of Data Science (ML / DL)

Real World (Business)



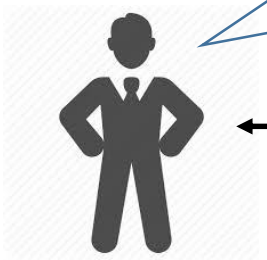
BETTER DECISIONS

EXPERTISE



IMPROVED MENTAL MODELS

Better Approximation of the Real World



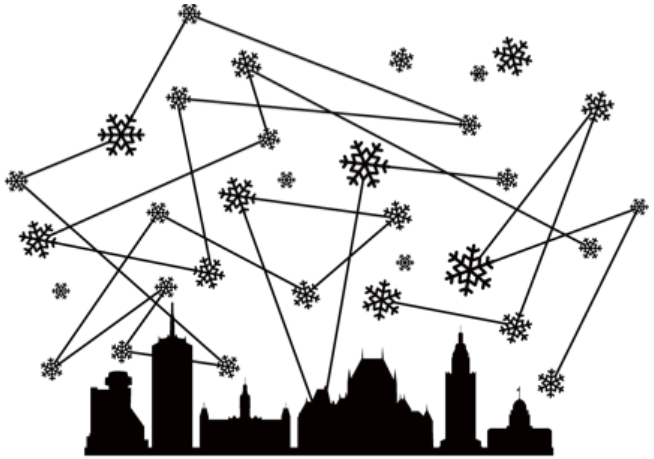
Business Decision Maker

Data Science Pipeline

DATA → ML / DL Algorithms → Inferences / Predictions

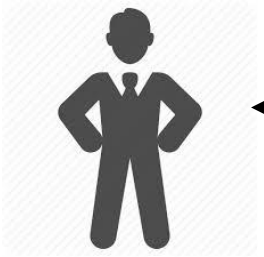
‘Decision Gap’ – What is it?

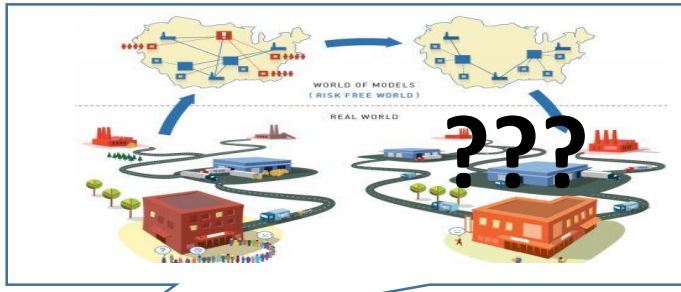
Real World (Business)



???

BETTER DECISIONS

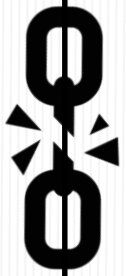




IMPROVED MENTAL MODELS

Better Approximation of the Real World

???



DECISION GAP

Data Science Pipeline

DATA → ML / DL Algorithms → Inferences / Predictions

The Decision Gap – Why it exists?

Predictions from Machine Learning

1	***Black Box Algorithms*** *– Lack of Interpretability of black-box algorithms (GBMs, CNNs etc.) leading to unexplainable predictions at a local level (specific customer or region)*
2	***No Quantification of Uncertainty*** *– ML models are obtained by minimizing a cost function or by maximum likelihood which results in point estimates with no quantification of uncertainty*
3	***Lack of Cause & Effect Relationships*** *- ML models are based on correlations in data and does not provide "*cause & effect*" relationships. Business decisions are about making 'interventions' and that requires an understanding of causation*
4	***No explicit formulation of constraints*** *- Business decision making is always under* constraints *(time, resources etc.). ML models does not typically take these constraints into account*
5	***Does not account for Dynamic, Non-linear Interactions*** *- Hey, what do I do with predicted numbers when the business is dynamic, has multiple entities with non-linear interactions?*

Business Decision Making

Data Science Approach – To Bridge the Decision Gap

Predictions from Machine Learning

1	***Machine Learning Interpretability (MLI)*** *– Techniques (LIME, Shapley values etc.) that deal with explains predictions at a local / individual record level*
2	***Probabilistic Programming*** *– Techniques that leverage Bayesian Inference to provide distribution of parameter values (not point estimates) thus quantifying uncertainty*
3	***Causal Machine Learning*** *– Techniques like Bayes Nets, Structural Equation Models (SEM) etc. that helps determine causal interaction among variables with ability to represent complex interactions in a manageable compact form*
4	***Optimization*** *– Well understood area with techniques like linear programming to help maximize or minimize objective functions along with decision variables and related constraints*
5	***Reinforcement Learning*** *– Techniques that helps in finding optimal policies in an environment through Dynamic Programming, Q-Learning etc. But still very much restricted to "Grid Worlds" and not widely adopted in business setting*

Business Decision Making

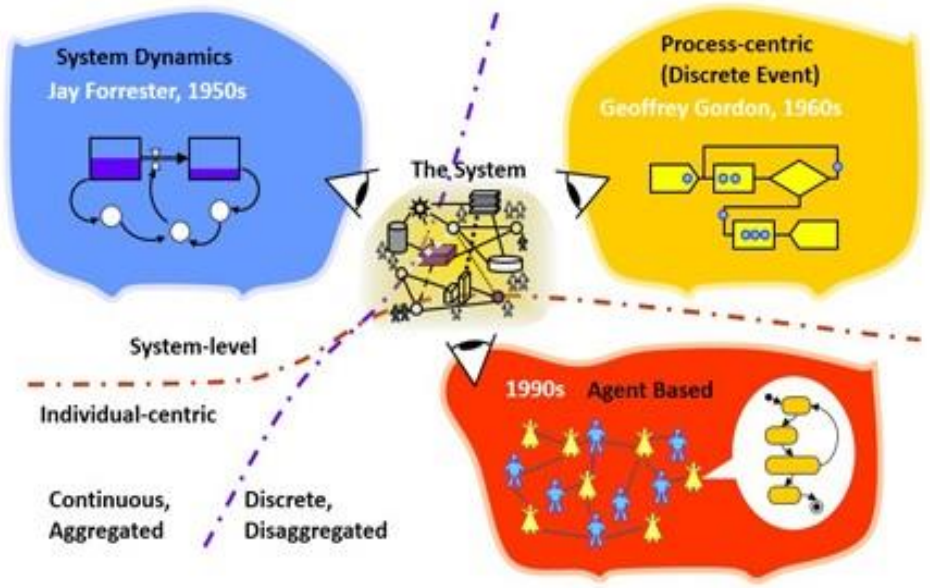
ENTER THE WORLD OF SIMULATION MODELING

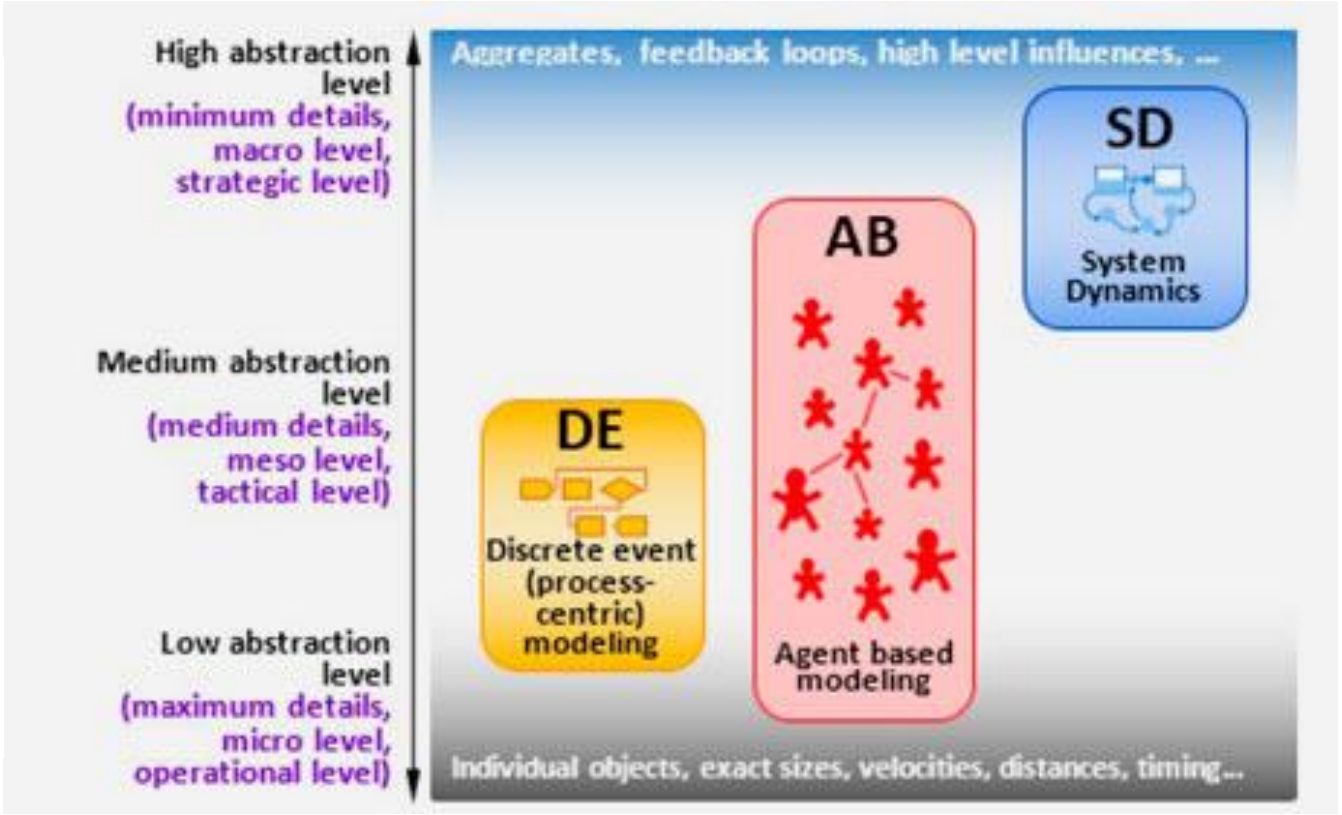
What is Simulation Modeling?

Simulation Modeling (or Business Simulations) helps build realistic approximations of the real world and factors in business complexity

What is AnyLogic?

Techniques for Simulation Modeling



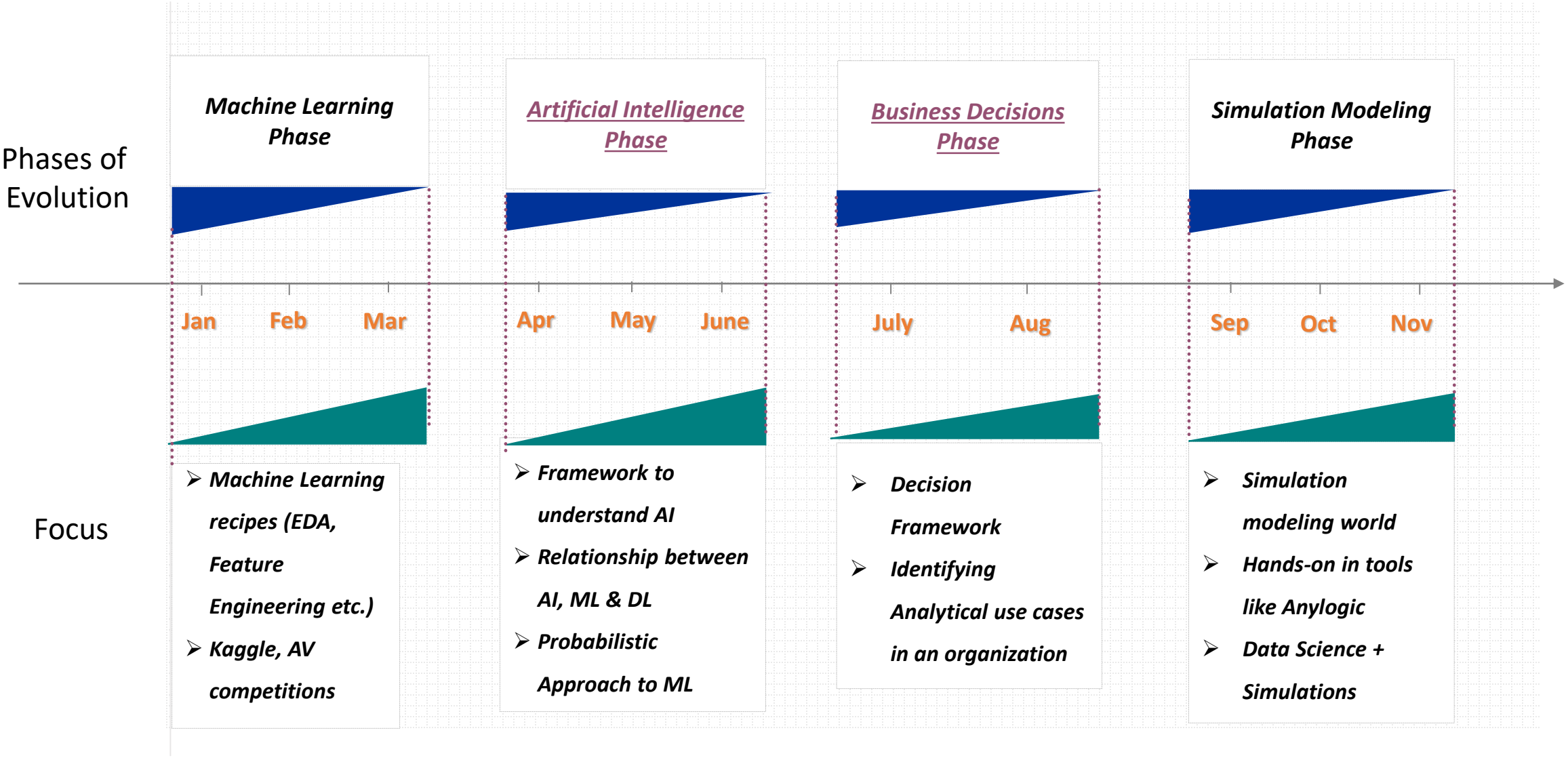


Many Tools: AnyLogic, Vensim, Stella, Simio, Arena, Flexsim, Netlogo etc.

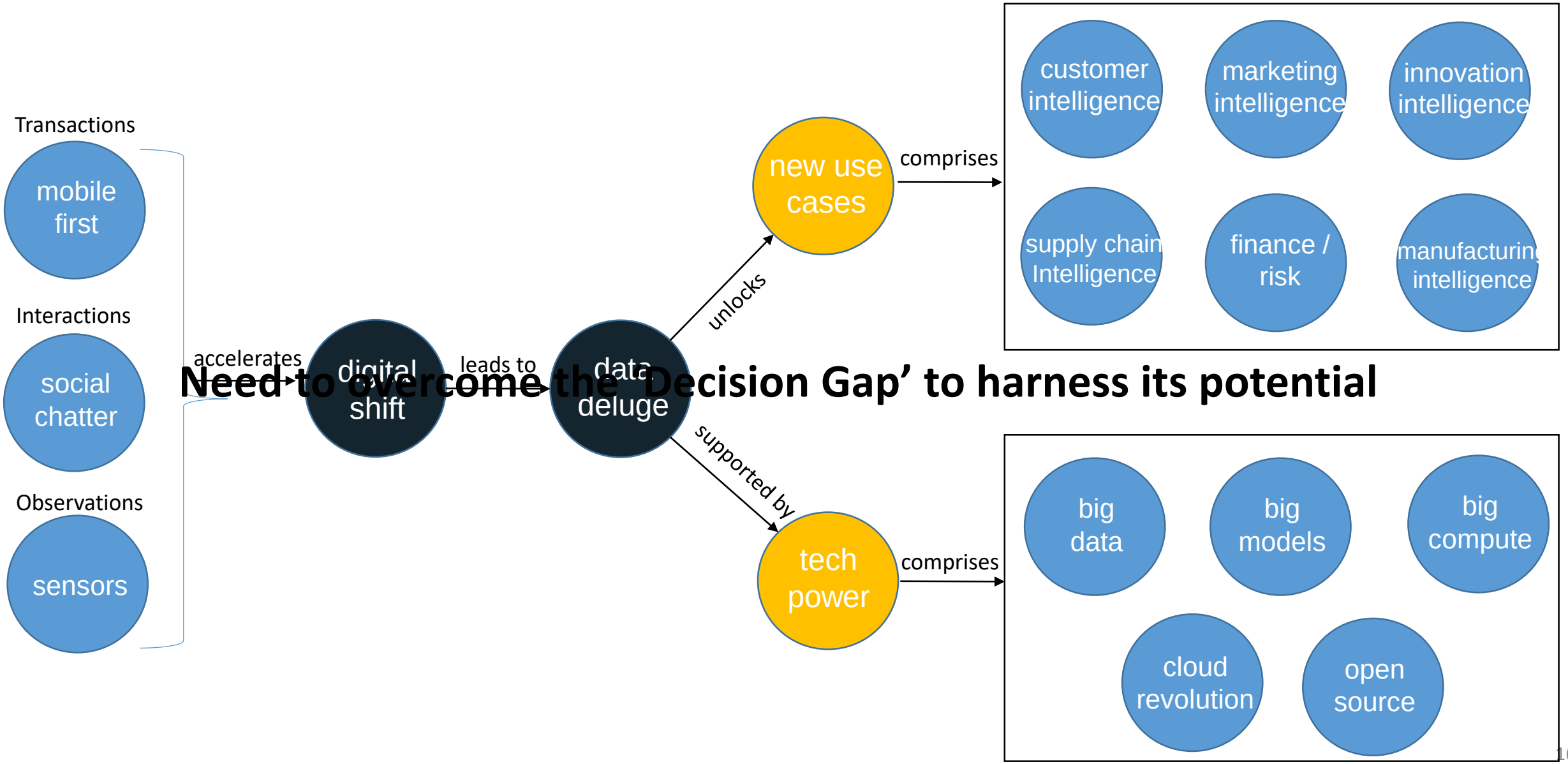
Data Science + Simulation Modeling – A Powerful Combination

	Simulation Modeling		Data Science (ML / DL / AI)	
1	Simulations are created based on the expert view on process relationships between entities – Could be unknown / biased / not realistic	+	Data Science by utilizing real world data and machine learning techniques can help correct the bias, increase precision, provide insights	Python Package PySD
2	Simulations can help generate good quality data from a representative model of the real world under different situations	+	Supervised Machine Learning needs good quality labelled data to generate reliable predictions	
3	Simulation models can provide the environment in which agents can operate across multiple runs to learn the policy	+	In Reinforcement Learning actions have to be performed in an environment multiple times to derive the appropriate policy (Q values)	

My personal ‘Data Science’ journey this year



Data Science & Analytics – 'Clear & Present' Opportunity



Need to overcome the 'Decision Gap' to harness its potential

Thank You!

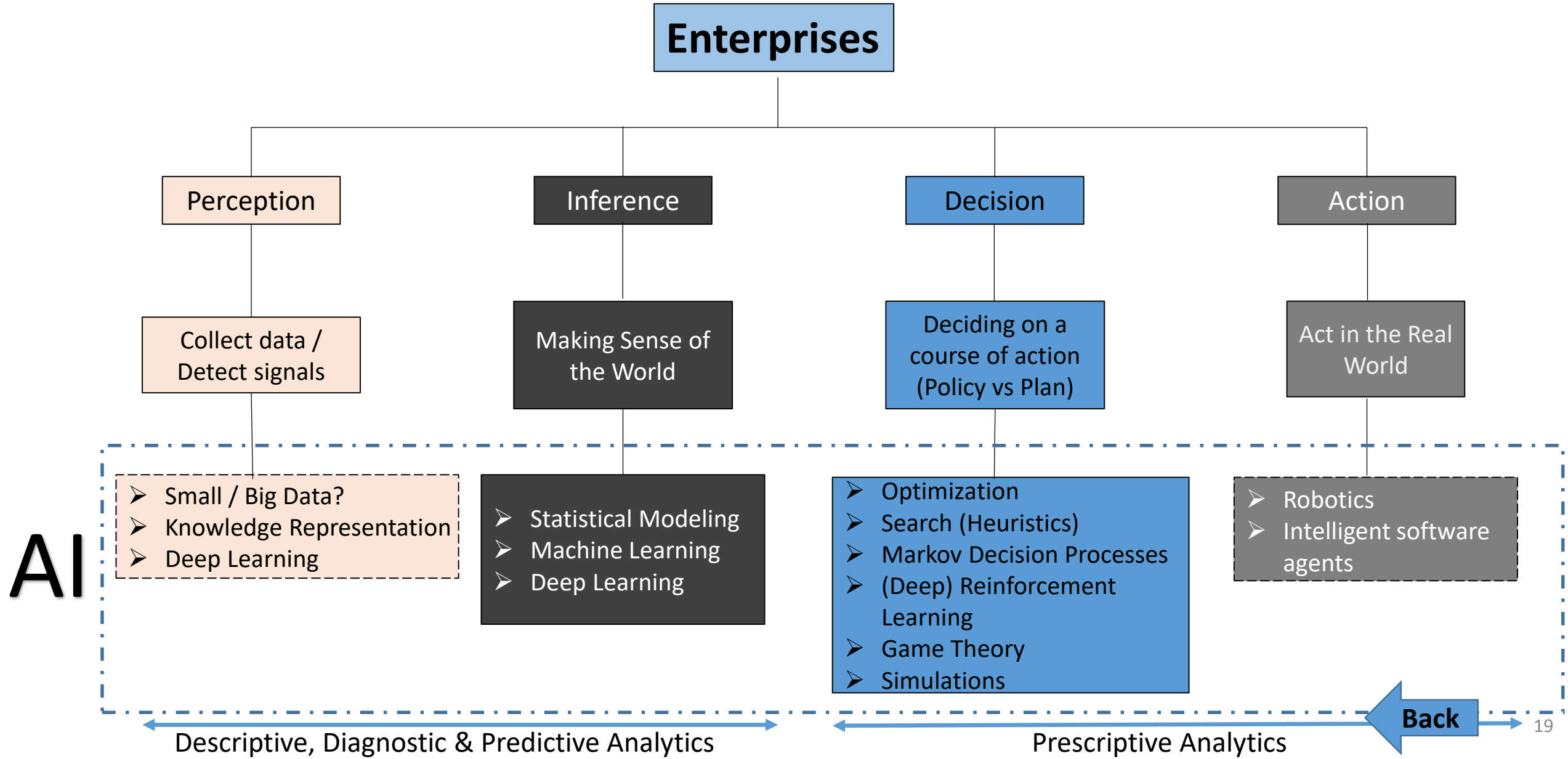
- Karthikeyan Sankaran, Director, LatentView Analytics
- Email ID – Karthikeyan.Sankaran@latentview.com
- Mindmap – bit.ly/31KArT8
- LinkedIn – http://in.linkedin.com/in/karthikeyansankaran
- Tapchief - www.tapchief.com/karthik
- GitHub – https://github.com/skkeyan-mlai/DHS2018

Artificial Intelligence and Human Intelligence

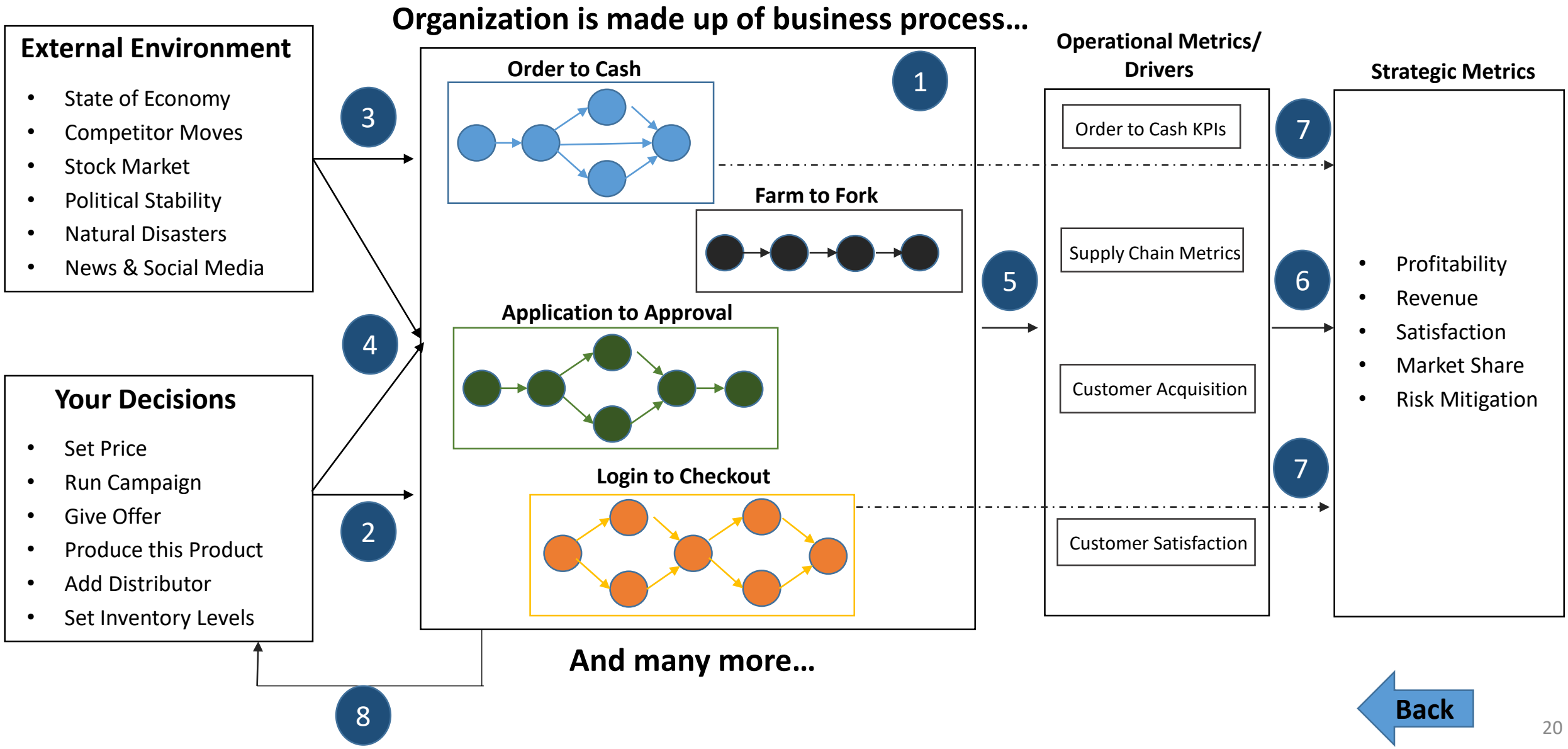
Artificial Intelligence refers to the theory and development of computer systems & machines with the ability to perform tasks normally requiring human intelligence

1. Perceive the world, detect signals and collect data
2. Make sense of the world using data (Insights, Inference, Predictions etc.)
3. Decide on the next course of action
4. Act in the Real World

AI Techniques in Enterprises – Parallels to Human Intelligence



Business Decision Making Framework



Back

My Analytics Mindmap

- Global Trends in Society
- Macro-economy
- Business Fundamentals
- Specific Industry Domain
- Analytical use cases

- Data Management
- Reporting & Self-service
- Quantitative Techniques
- Performance Mgmt
- Insight Delivery

Analytics for Business Value
http://bit.ly/31KArT8

- Scan for New Products
- Evaluate Maturity

- Monitor Ecosystem
- Leverage Resources

Thank You