Manufacturing Data Science

Dr. Chia-Yen Lee (李家岩 博士)

Department of Information Management (資訊管理學系)
National Taiwan University (國立台灣大學)

Education

- Ph.D, 工業與系統工程, Texas A&M University, USA
  (Major: Operations Research 作業研究/運籌學)
- M.S., 工業工程與工程管理, 國立清華大學
- B.S. & B.B.A., 應用數學暨資訊管理, 國立政治大學

Experience

- 教授， 國立台灣大學資訊管理學系
- 教授兼所長，國立成功大學資訊工程學系暨製造資訊與系統研究所
- 副編輯，IEEE Transactions on Automation Science and Engineering (SCI)
- 工業工程學會秘書長、工業局新興技術專家顧問、半導體科技廠智慧製造中心顧問

Award

- IEEE Senior Member (2021)、呂鳳章先生紀念獎(2019)
- 美光教師Micron Teacher Award (2018)
- 李國鼎科技與人文講座研究獎 (2018)
- 科技部吳大猷先生紀念獎 (2017)

Research Interest

- 製造數據科學、智慧型製造系統、生產力與效率分析、多目標決策


智慧工廠
Intelligent Factory is a decision-oriented system which has the computational intelligence and self-learning ability to optimize the manufacturing process.
計算智慧 → Based on Data (資料處理與分析)
自我學習 → Real-time Feedback Control (回饋控制)
INTELLIGENT MANUFACTURING
PRODUCTION PROVEN
PREDICTABLE RESULTS
生產力最佳化實驗室@NCKU
製造資料?科學!?
玉山金控 E.SUN FHC
TREND MICRO 趨勢科技
國泰金控 Cathay Financial Holdings
funP
intel
KKbox
MEDIATEK
QNAP

Introduction

- MDS course supports students learning how to apply artificial intelligence (AI), machine learning, data science (DS) techniques to improve the effectiveness and efficiency of the manufacturing systems. MDS integrates the knowledge domains of the information, engineering, and management. Encourage students to solve the real problem systematically using the design of analytics, from descriptive, diagnostic, predictive, prescriptive to automating, for successfully enhancing decision quality.

Objectives

- Learn the statistical learning and optimization methodologies for intelligent manufacturing systems
- Create a prototype model to solve the problem in real setting related to manufacturing or service systems
- Develop the research skills and prepare a analytic project report

- Learning Requirement:
  - It's better to have prerequisite courses: (1) probability and statistics; (2) operations research
  - Python programming skills
  - Students need to read literature and develop analytical model for term project
- Challenges
  - How to find the problem and identify the right methodology to solve the problem in manufacturing systems?
  - Effectiveness vs. Efficiency
    - From Organization Management Perspective (Drucker, 1977)
      - Do the right thing → Effectiveness
      - Do the thing right → Efficiency
    - From Production System Perspective (Lee and Johnson, 2015)
      - Demand over Output → Effectiveness
      - Output over Input → Efficiency
  - Optimization Theory and Statistics
  - IT Capability & Programming Skills (eg. Python only)

Drucker, P., 1977. An Introductory View of Management. Harper College Press, New York.
Lee, C.-Yen, and A. L. Johnson, 2015. Effective Production: Measuring of the Sales Effect using Data Envelopment Analysis. Annals of Operations Research, 235 (1), 453–486.

Textbook

- Lecture notes

Reference

- Hastie, T., R. Tibshirani, and J. Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer.
- Hillier, F. S., Lieberman, G. J. (2010), Introduction to Operations Research, 9th ed., McGraw-Hill, New York.
- Hopp, W. and M. Spearman (2011), *Factory Physics*, 3rd ed., Waveland Press.
- Montgomery, D. C. (2013), Introduction to Statistical Quality Control, 7 ed.: John Wiley & Sons, Inc.
- Nahmias, S. (2008), *Production and Operations Analysis*, 6th ed., McGraw-Hill/Irwin.
- Pinedo, M. L. (2016), *Scheduling: Theory, Algorithms, and Systems*, 5th edition, Springer-Verlag New York.

Course Contents

- Data Science & Manufacturing Systems
  - Data, Information, Knowledge, and ML/DS Functions
  - Analytics Framework and Data Preprocessing
  - Manufacturing Systems and Factory Dynamics
- Diagnostic and Predictive Analytics
  - Feature Selection and Feature Engineering
  - Regression, Classification, MARS, and Symbolic Regression
  - Tree-based Methods, Random Forest and Boosting
  - SPC, Signal Processing, and PHM
  - Clustering Analysis and Deep Learning
  - Manufacturing Practice
- Prescriptive Analytics
  - Linear Programming and Capacity Planning
  - Metaheuristic Algorithm and Genetic Algorithm
  - Scheduling Optimization and Run-to-Run Control
- Advanced Techniques (if time permits)
  - Concept Drift and Domain Adaptation
  - Transfer Learning, Meta-Learning, Few-shot Learning, Small Samples
- Term-project Presentation

Course Grading Scheme

- 4 Assignments 60%
- Research Project 40% (Analytics with System Demo)

Project Grading Scheme (40%)

- PIP數據不可使用，後果與風險自行承擔。
- Proposal 10%
- Presentation 40%
  - 後一組問前一組問題
  - 同儕互評(其他組別評分後取中位數)
- Final Report 50%

Course Website

- NTU Cool (https://cool.ntu.edu.tw/login/portal)

Teaching Assistant (TA)

- 宋亭遠
  - Expertise
    - Text Mining for Oil Price Forecast
    - Prognostic and Health Management (PHM)
  - Email: r09725045@ntu.edu.tw
- 陳柏儒
  - Expertise
    - Operations Research and Portfolio Optimization
    - Knowledge Graph and Causal Inference
  - Email: r10725008@ntu.edu.tw
- Office Hour: by appointment

注意事項

- 課程需分組
  - 四人一組，盡量不同科系
  - 專案需要跟製造業或服務業有關

- 課程作業基本上沒有標準答案
  - 有答案寫法，但是過沒多久就不新了...甚至效果不好了..(IT進步很快)
  - 建議：會從作業中挑幾份代表性的參考答案，請同學上台分享
    - 專案與作業課堂授權與分享

心理準備

- Prerequisite
  - Probability and Statistics
  - Operations Research
  - Python
- 是研究所課程
  - 有一定的loading…
- 是跨領域
  - 有一定的學習挑戰
  - 與小繁雜…
- 製造業並不是那麼美好與理想…
- 整學期上完大概僅佔15%吧..但掌握了關鍵基礎學理與態度..

Intelligent Manufacturing Systems and Data Science Functions

Dr. Chia-Yen Lee (李家岩 博士)

Department of Information Management (資訊管理學系)
National Taiwan University (國立台灣大學)

Course Contents

- Data Science & Manufacturing Systems
  - Data, Information, Knowledge, and ML/DS Functions
  - Analytics Framework and Data Preprocessing
  - Manufacturing Systems and Factory Dynamics
- Diagnostic and Predictive Analytics
  - Feature Selection and Feature Engineering
  - Regression, Classification, MARS, and Symbolic Regression
  - Tree-based Methods, Random Forest and Boosting
  - SPC, Signal Processing, and PHM
  - Clustering Analysis and Deep Learning
  - Manufacturing Practice
- Prescriptive Analytics
  - Linear Programming and Capacity Planning
  - Metaheuristic Algorithm and Genetic Algorithm
  - Scheduling Optimization and Run-to-Run Control
- Advanced Techniques (if time permits)
  - Concept Drift and Domain Adaptation
  - Transfer Learning, Meta-Learning, Few-shot Learning, Small Samples
- Term-project Presentation

Outline

- Intelligent Manufacturing Systems (IMS)
- Manufacturing Systems and Automation
- Data Science Functions and Applications
  - ML/DS Functions
  - Data Visualization and Interpretation
  - Practical Applications
- Concluding Remarks
  - Data Scientist and Venn Diagram
  - Complexity and Issues
  - Open Data Sources

Overview

- "Intelligent Manufacturing Systems"
  - What is Intelligent?
  - What is Manufacturing?
  - What is Systems?

- Methodologies
  - Computational Intelligence
    - AI, machine learning, …
  - Meta-Heuristics: for engineering optimization
  - Data Science: for pattern recognition in dataset

- Applications
  - Optimization (scheduling, knapsack, network, layout, digital design, etc.)
  - Classification (market segmentation, pattern recognition, bioinformatics, etc.)
  - Forecast (stock market, demand fluctuation, product bundling, product life cycle, etc.)
  - … etc.

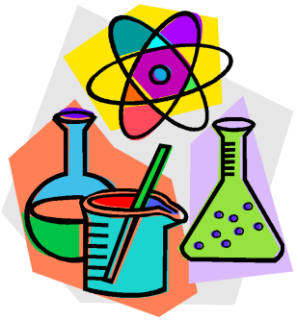
Methodologies …?

Really!?

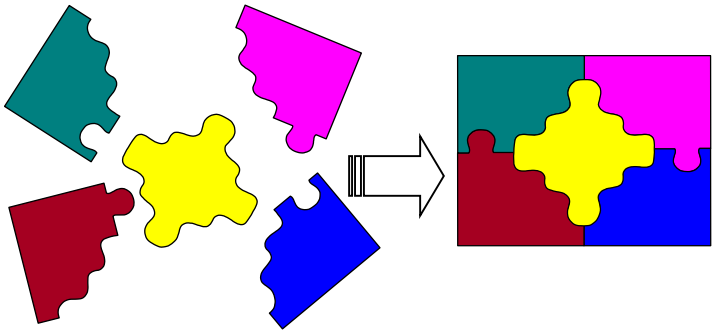
generated by Wordle (2012)

Essentiality of Methodology

- This is the course for "Methodology (方法論)"



  - Science
    - Emphasize on "discovery"
    - impact: long & significant, but 短期可能不見效
      - Newton, 1687, *Philosophiæ Naturalis Principia Mathematica*



  - Methodology
    - Emphasize on "application"
    - Impact: The Science of Better (INFORMS)

  - Technology
    - Emphasize on "realization/ implementation"
    - impact: short, but 短期見效
      - LED light: 1965
      - Cell phone: Apr. 3, 1973
      - MS-DOS 1.0: August 1981

Essentiality of Methodology

- What kind of person you want to be
  - Scientist
    - Explore the "truth" in the world
  - Engineer
    - Solve the "real" engineering problem based on "science"
  - Manager
    - Decision making for "resource" optimization to achieve the "goal"
  - Economist
    - Tell a "story" to illustrate the "trading behavior" among humans
  - Lawyer
    - Judge the "issue" based on the "law"
  - Doctor
    - Suggest the "treatment" according to the "root cause" of "symptom"

What is Manufacturing?

Manufacturing is …

- What is “Manufacturing”?
  - Manus (hand)
  - Factus (make)

Manufacturing is the realization (具現化) of product.

Swann, J., 2003. Manufacturing Systems, Course Lecture Notes, Georgia Institute of Technology.

Manufacturing is …

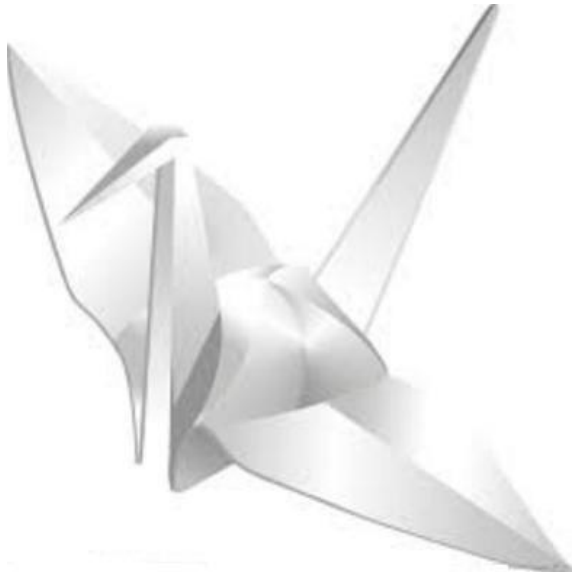
- Black Hand
  - Real field study 實證場域 for implementation (實作) and validation
  - Experience and empirical study
- Resource Optimization
  - 8M1I: 人(Man)、機(Machine)、料(Material)、方法(Method)、測量(Measure)、時間(Minutes)、資金(Money)、環境(Mother nature/environment)、資訊(information)
- Must associated with Productivity and Profitability
  - Efficiency: $\frac{Output}{Input}$
  - Effectiveness: $\frac{Demand}{Output}$

How to throw a “paper” farther?

圖片來源：google 網路截圖

Manufacturing is …

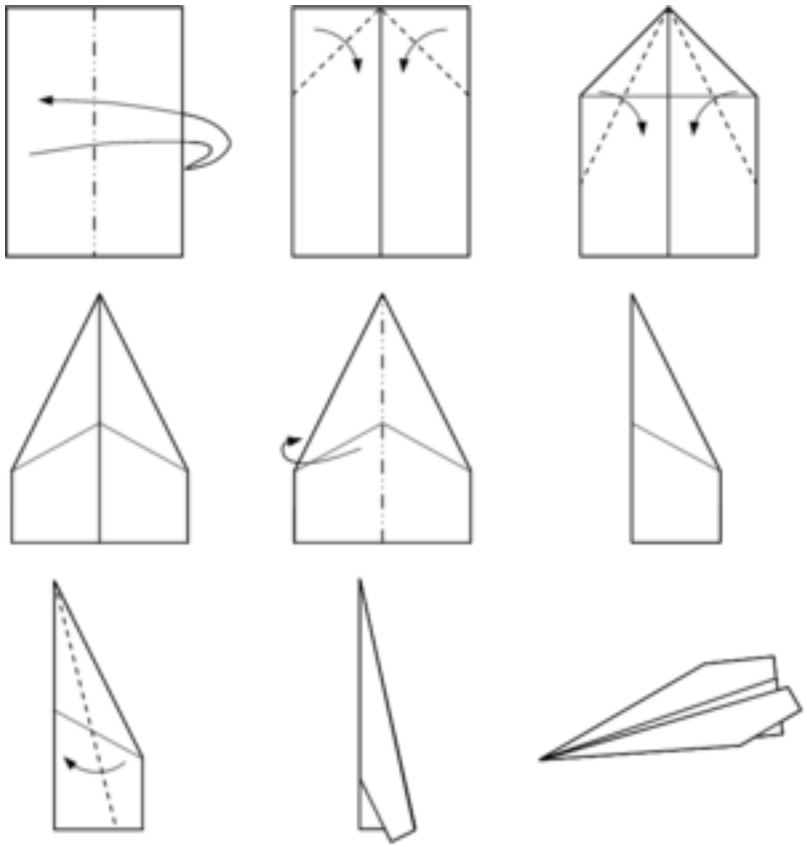
- Idea? (by brain-storming and association…)

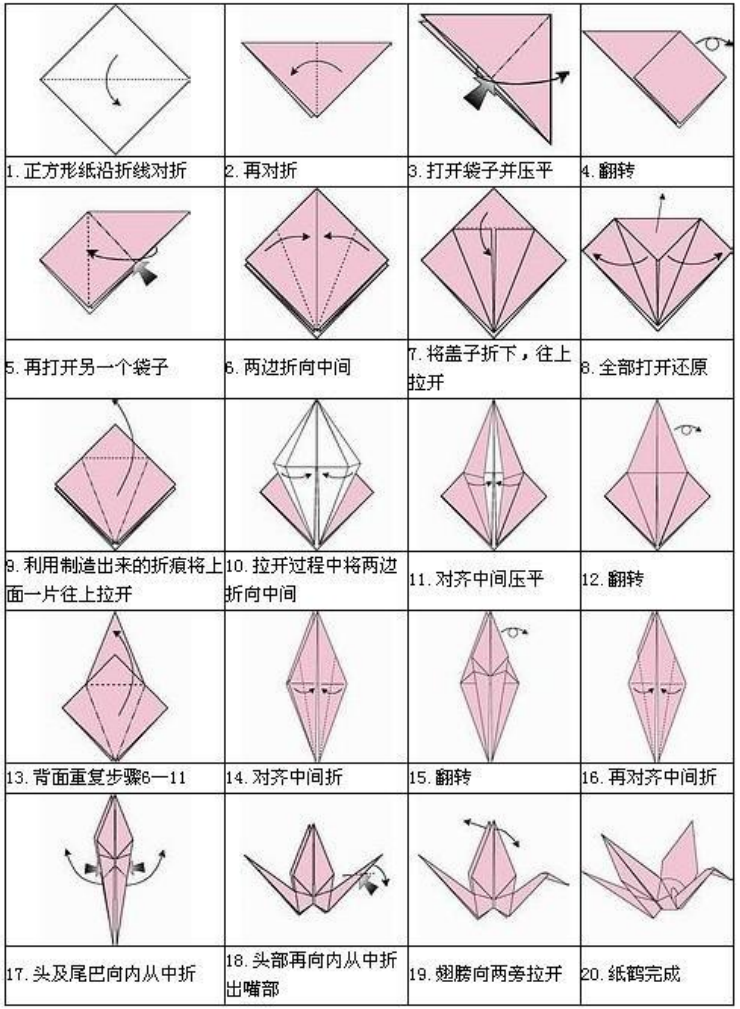


圖片來源：google 網路截圖

Manufacturing is …

- Manufacturing Process… (Handmade)
  - For your needs (function)





XiaYin (2013). 《１６種超酷的紙飛機摺法》教你摺出最會飛的一架紙飛機！http://clickme.net/23163
覃欢 (2011).千紙鶴. 中文百科在線. http://www.zwbk.org/MyLemmaShow.aspx?zh=zh-tw&lid=150882

可是...人性不是這麼容易滿足的...

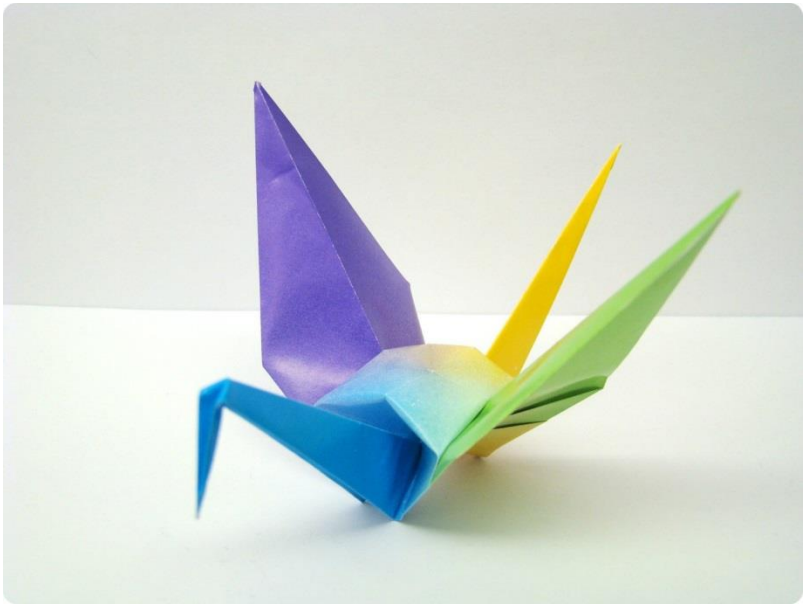
obj1. How to throw a “paper” farther?

obj2. The product should be delightful.

(this is what we call the “customer requirements”)

圖片來源：google 網路截圖

- Valued added?
  - Material change

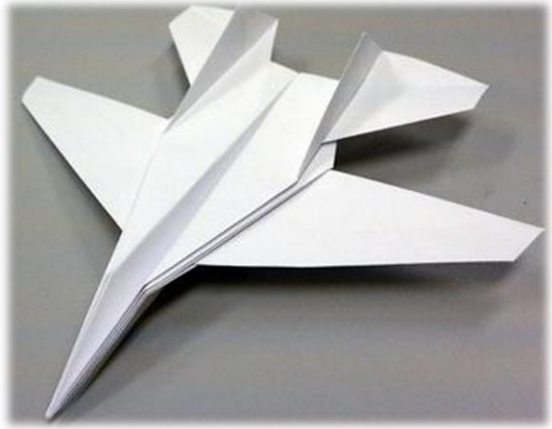


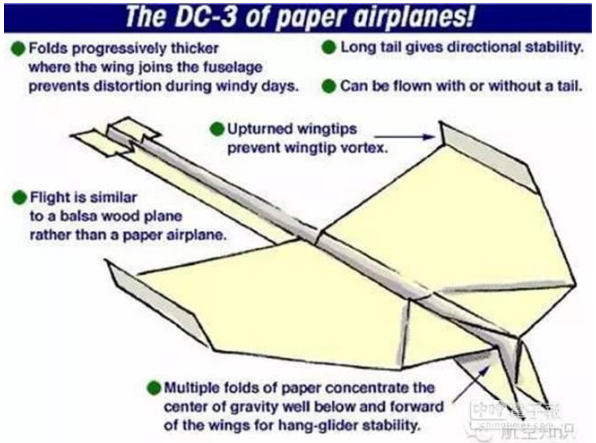
圖片來源：google 網路截圖

- Valued added?
  - Process change

圖片來源：google 網路截圖

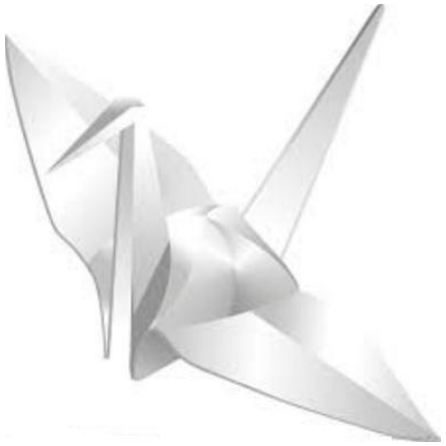
- Valued added?
  - Process change





圖片來源：google 網路截圖

- Valued added?
  - Process change



圖片來源：google 網路截圖

EYNY & dseeyou419 (2013). 紙鶴進化史！https://forum.gamer.com.tw/Co.php?bsn=60084&sn=2179624
阿波羅新聞網(2015). 藝術家的獨特千紙鶴！背後有著超級感人的故事. http://tw.aboluowang.com/2015/1105/639889.html

這樣...顧客就滿足了嗎?

- Customer Requirements
  - Obj1. How to throw a “paper” farther?
  - Obj2. The product should be delightful.
  - Obj3. Low cost (低成本)
  - Obj4. Specification (長、寬、高)
  - Obj5. Firmness (牢固不易壞)
  - …
  - ….
  - ……
  - ……..

製造系統為了達到
顧客需求...
轉而就變成了

製造現場的KPI

Performance Index

Dependability

- The protection within the system
  - Inventory level
  - Ratio of imbedded extras in lead time
- The number of changes in a time period
  - Percentage of production order changes
  - Percentage of sales order changes
- Delivery performance
  - Percentage of quantity over on under original schedule
  - The differences between actual start and completion dates
  - start and completion dates
- Machine availability
  - The ratio of maintenance cost to repair cost
  - Percentage of machine up time
  - Usage of formalized preventative maintenance plans

Time

- Production cycle time

Cost

- Cost per unit

Quality

- Customer response
  - Percentage of repeat sales
  - Unfulfilled customer orders
  - Number of complaints
  - Average time between service calls
- Manufacturing system (conformance to specification)
  - First pass yield
  - Defect ratios
  - Scrap rate

Flexibility

- Lot sizes
- Set-up times
- Percentage of standard, common and unique parts
- Number of different processes
- Cross-training of personnel

Yurdakul, Mustafa (2002),"Measuring a manufacturing system's performance using Saaty's system with feedback approach", Integrated Manufacturing Systems, Vol. 13 Iss 1 pp. 25 - 34

What is Intelligent?

Intelligent…

Shutterstock (www.shutterstock.com)

Intelligent is …

- What is “Data”? (**objective record**)
  - Factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation (Merriam-Webster)
  - Information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer (Cambridge Dictionary)
  - Structured data
    - Transaction: name, location, sex, age, …
    - Weather, temperature, humidity, wind speed, ..
    - Grade: Mandarin, Math, Society, Science, …
    - Machine: log, chamber, parameter, vibration, …
    - Production: type, specification, quantity, routing, recipe, …
  - Unstructured data
    - Video, music, text, image, …

https://www.merriam-webster.com/dictionary/data; https://dictionary.cambridge.org/

Intelligent is …

What is “Data”?

- Metadata (data about data)
  - "data that provides information about other data".
    - Descriptive: used for discovery and identification, such as title, abstract, author, and keywords.
    - Structural: containers of data and indicates how compound objects are put together. It describes the types, versions, relationships and other characteristics of digital materials.
    - Administrative: information to help manage a resource, like resource type, permissions, and when and how it was created.
    - Reference: information about the contents and quality of statistical data.
    - Statistical: also called process data, may describe processes that collect, process, or produce statistical data.

- Data Definition Language (DDL)
  - refers to the set of SQL commands in DBMS, which is a syntax similar to a computer programming language for defining data structures, especially database schemas. DDL statements create and modify database objects such as tables, indexes, and users. Common DDL statements are CREATE, ALTER, and DROP (including indexes, triggers, tables, and views.).

https://en.wikipedia.org/wiki/Metadata; https://en.wikipedia.org/wiki/Data_definition_language
https://study.com/academy/lesson/data-definition-language-ddl-definition-example.html

Intelligent is …

- Information (interpretability, subjective)
  - is processed data (eg. ranking or sorting) which provides meaning and interpretability, so that the reader can receive and understand the "message", which shows some potential value.
  - Information can affect the recipient's thoughts and judgments, and has association and objectives
  - Data + Data + …. + Data ≠ Information
  - From data to information
    - Value-added by operators (eg. calculate, sort, categorize, summarize)
    - Data quality and KPI association are the focuses rather than analysis tool
  - Example
    - Top 10 items with the higher monthly transaction revenue
    - Daily rush hours/peak loadings
    - Preference structure about the music (eg. Rock ⩾ R&B/Soul ⩾ Classical)
    - Feeling to the raining (eg. blue or exciting)

Intelligent is …

- Knowledge (criticize, heterogeneity, insights)
  - the circumstance or condition of knowing something (eg. truth or fact) with familiarity gained through experience, reasoning or association
    - a new framework to integrate experience, value and information
  - From information to knowledge
    - Why? Why? Why? … x N
    - Discussion! Brain-storming! Discussion! Brain-storming! … x N
    - Assumption → Experiments → Validation… xN
    - People or machine/robot participation
    - Reflection (省思)
    - Heterogeneity (異質性) of Information
  - Example
    - Scientific principles (科學的原理原則) (實驗：水的沸點為100度)
    - Discipline: Calculus, Statistics, Algebra, Physics, Chemistry, Economics, Biology, Accounting…

Intelligent is …

DATA → 詮釋的觀點 → Information → 思辨的過程 → Knowledge

科學的原理

修正 (Corrected)
文字化 (Contextualized)
量化 (Quantified)
排序 (Sorting)
計算 (Calculated)
分類 (Categorized)
摘要 (Summarized)
…

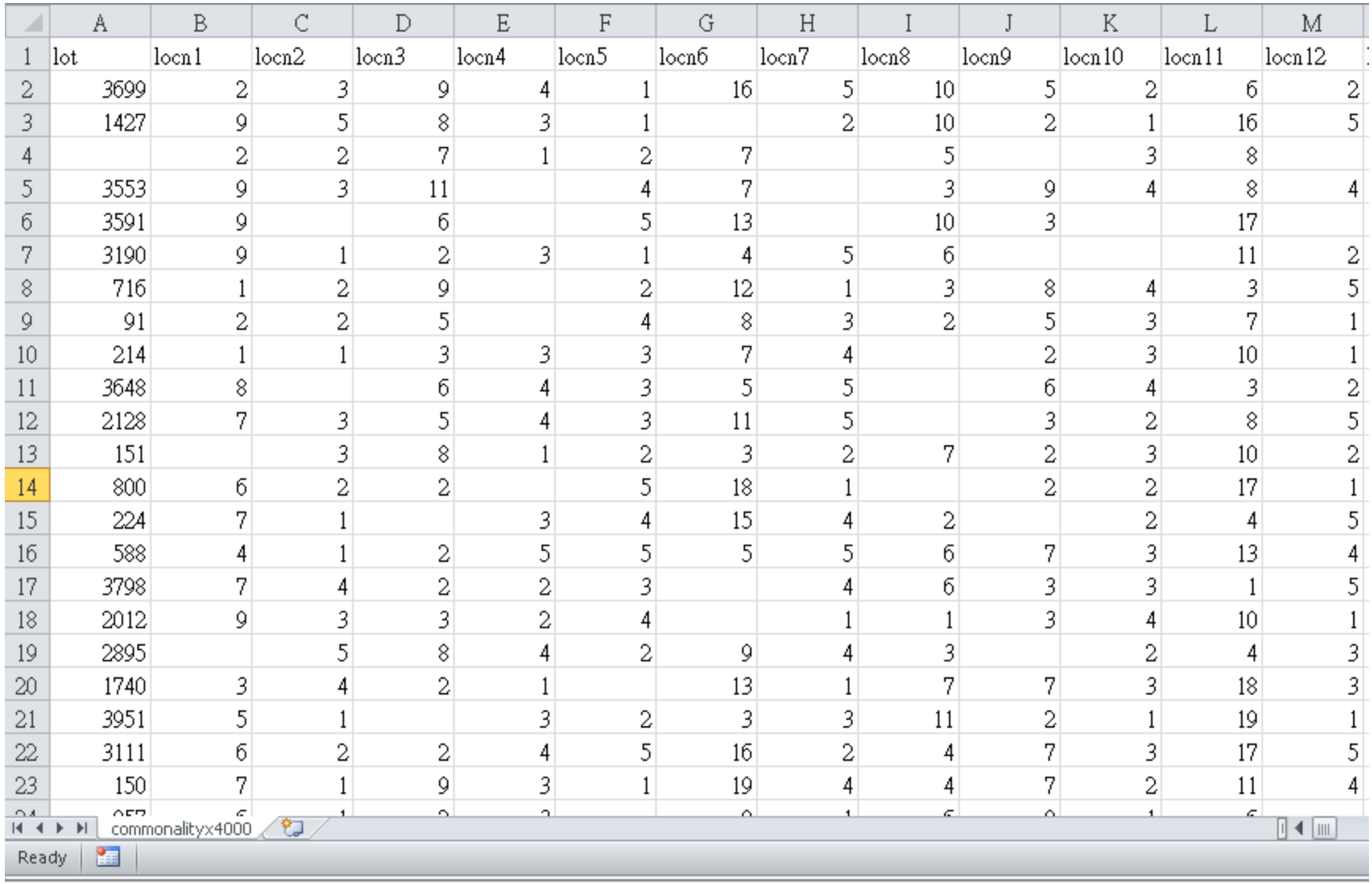
重點在於資料品質與資訊詮釋

附加價值 (Value-added):
- Observation
- Interpretation
- Understanding
- Experience
- Skills

人:
- Conservation
- Comparison
- Association
- Criticize
- Imagination

重點在於人的參與思辨與討論

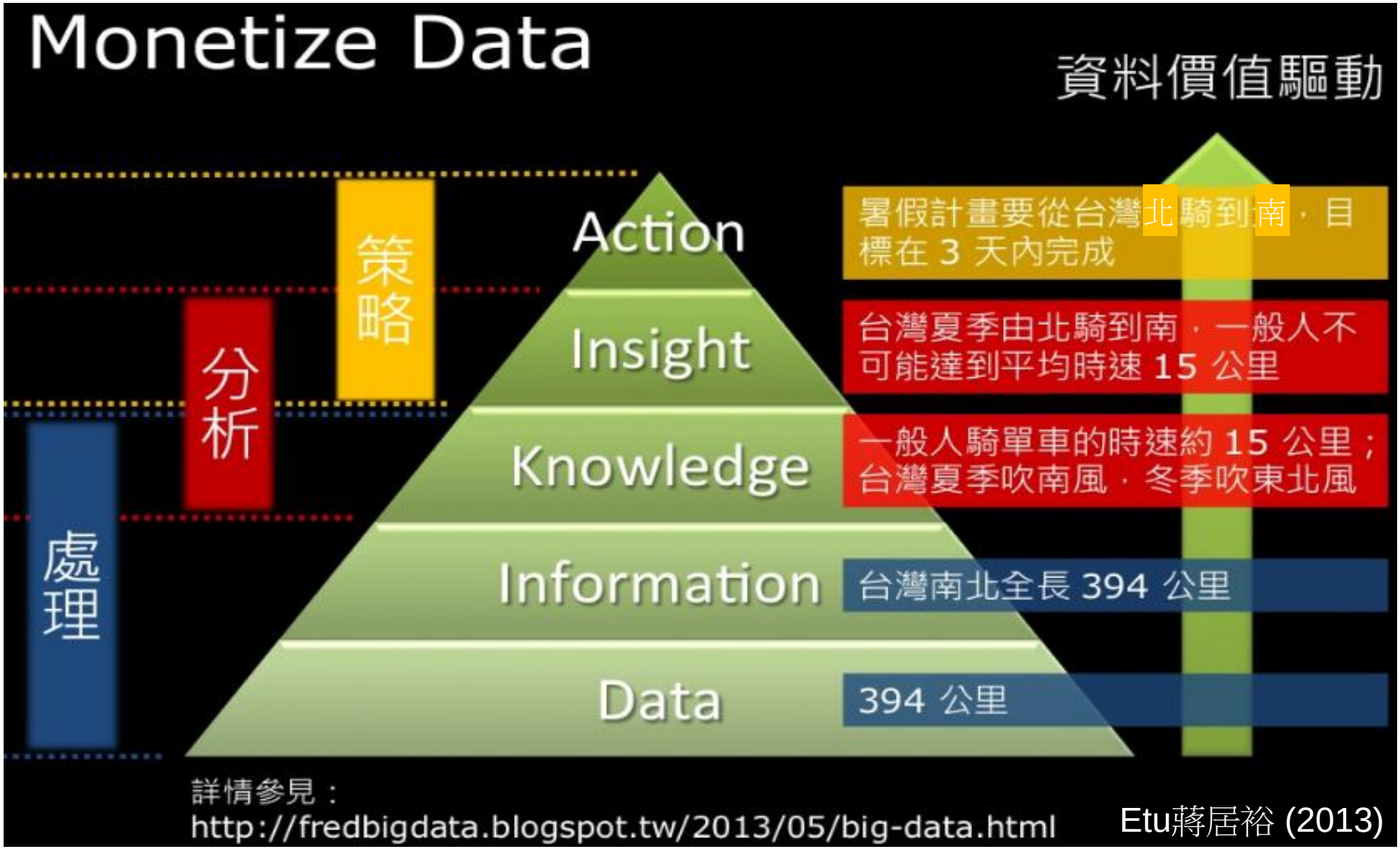
迷思: 資料？資訊?



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	lot	locn1	locn2	locn3	locn4	locn5	locn6	locn7	locn8	locn9	locn10	locn11	locn12
2	3699	2	3	9	4	1	16	5	10	5	2	6	2
3	1427	9	5	8	3	1		2	10	2	1	16	5
4		2	2	7	1	2	7		5		3	8	
5	3553	9	3	11		4	7		3	9	4	8	4
6	3591	9		6		5	13		10	3		17	
7	3190	9	1	2	3	1	4	5	6			11	2
8	716	1	2	9		2	12	1	3	8	4	3	5
9	91	2	2	5		4	8	3	2	5	3	7	1
10	214	1	1	3	3	3	7	4		2	3	10	1
11	3648	8		6	4	3	5	5		6	4	3	2
12	2128	7	3	5	4	3	11	5		3	2	8	5
13	151		3	8	1	2	3	2	7	2	3	10	2
14	800	6	2	2		5	18	1		2	2	17	1
15	224	7	1		3	4	15	4	2		2	4	5
16	588	4	1	2	5	5	5	5	6	7	3	13	4
17	3798	7	4	2	2	3		4	6	3	3	1	5
18	2012	9	3	3	2	4		1	1	3	4	10	1
19	2895		5	8	4	2	9	4	3		2	4	3
20	1740	3	4	2	1		13	1	7	7	3	18	3
21	3951	5	1		3	2	3	3	11	2	1	19	1
22	3111	6	2	2	4	5	16	2	4	7	3	17	5
23	150	7	1	9	3	1	19	4	4	7	2	11	4

Advanced Analytics – Intel: SETFI: Manufacturing data: Semiconductor tool fault isolation. Causality Workbench Repository, http://www.causality.inf.ethz.ch/ repository.php (2008)

Transformation from Data, Information, to Knowledge



Intelligent is …

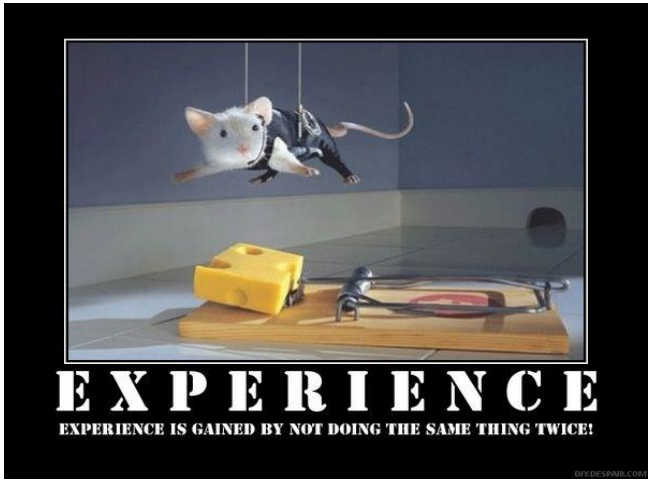
Problem

權衡的觀點

Decision

失敗的藝術

Experience



人生的歷練

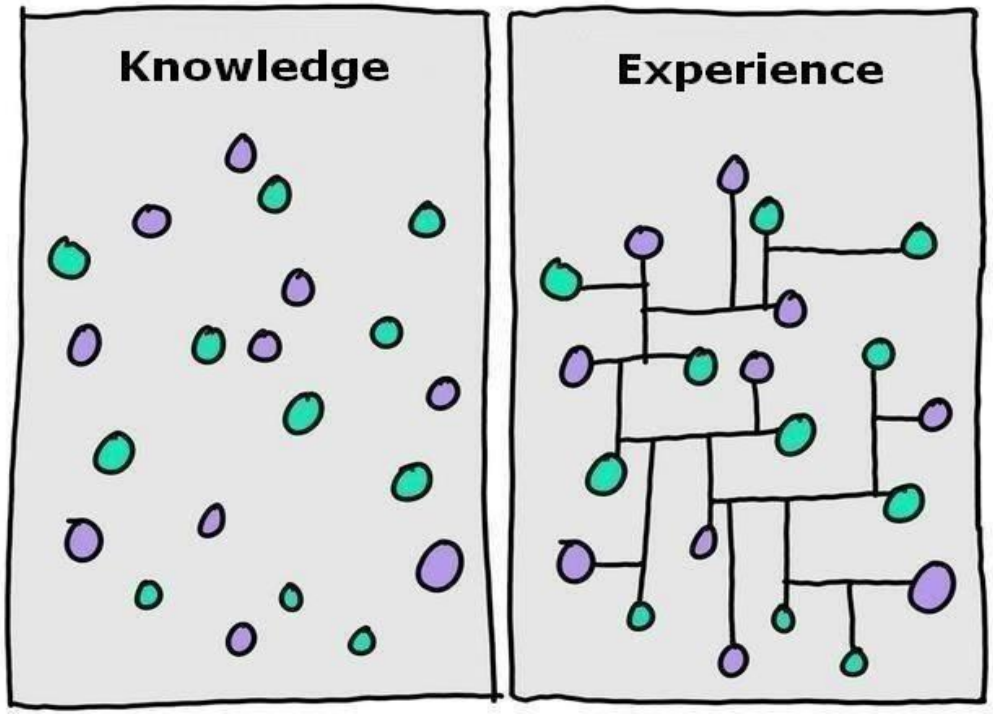
Intelligent is …

知識 與 經驗

之

融會貫通

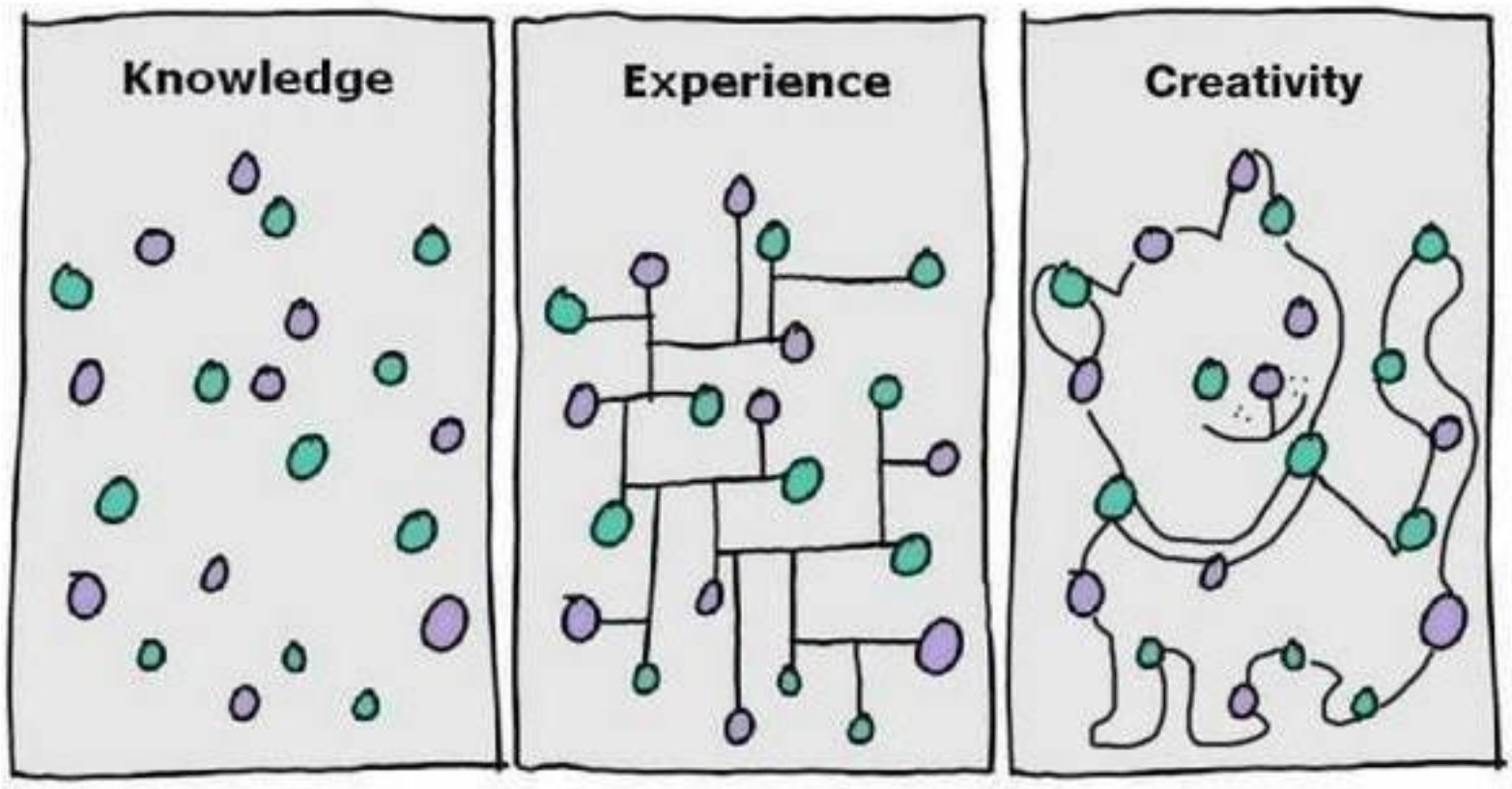
（思辨與討論）



http://www.herogamingjobs.com/2014/01/07/experience-vs-knowledge/

Intelligent is …

- The most important skill in the future will be the ability to "connect the dots" in your own way! (Moioli, 2019)



Moioli, Fabio (2019). The most important skill in the future will be the ability to "connect the dots" in your own way!. https://www.linkedin.com/pulse/most-important-skill-future-ability-connect-dots-your-fabio-moioli

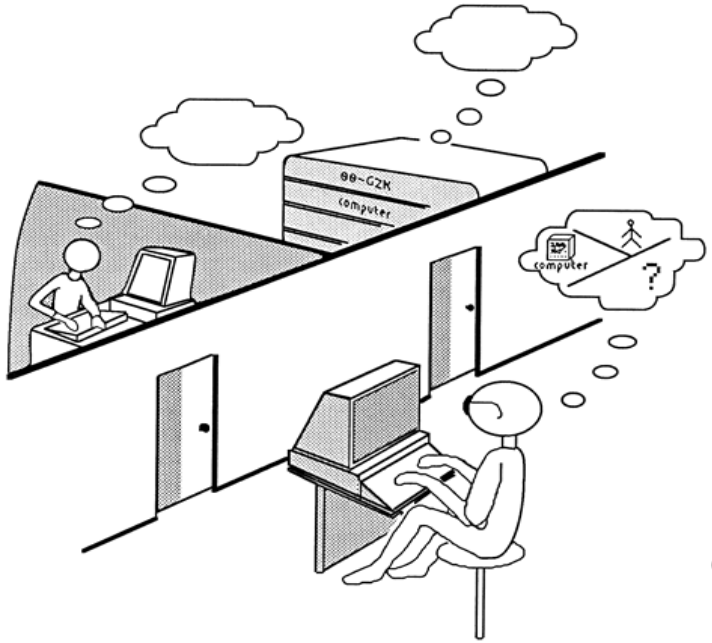
Intelligent is …

- Intelligence (智能)
  - Different and far far from Wisdom (智慧)…
- Human intelligence
  - mental quality that consists of the abilities to learn from experience, adapt to new situations, understand and handle abstract concepts, and use knowledge to manipulate one's environment.
- Artificial intelligence (AI)
  - the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.

Britannica Online Encyclopedia (2012)

Turing Test

- One of the first to believe in AI (Turing, 1950)
  - A machine can mimic the human brain
  - Many people did not want to believe a machine could do the same thing as a human
- Turing Test
  - A person asked questions on a keyboard to a person and a machine, if they could not tell the difference after some time the machine was considered intelligent (BBC News, 1999)



A.L.I.C.E

2000 Turing Test Championship

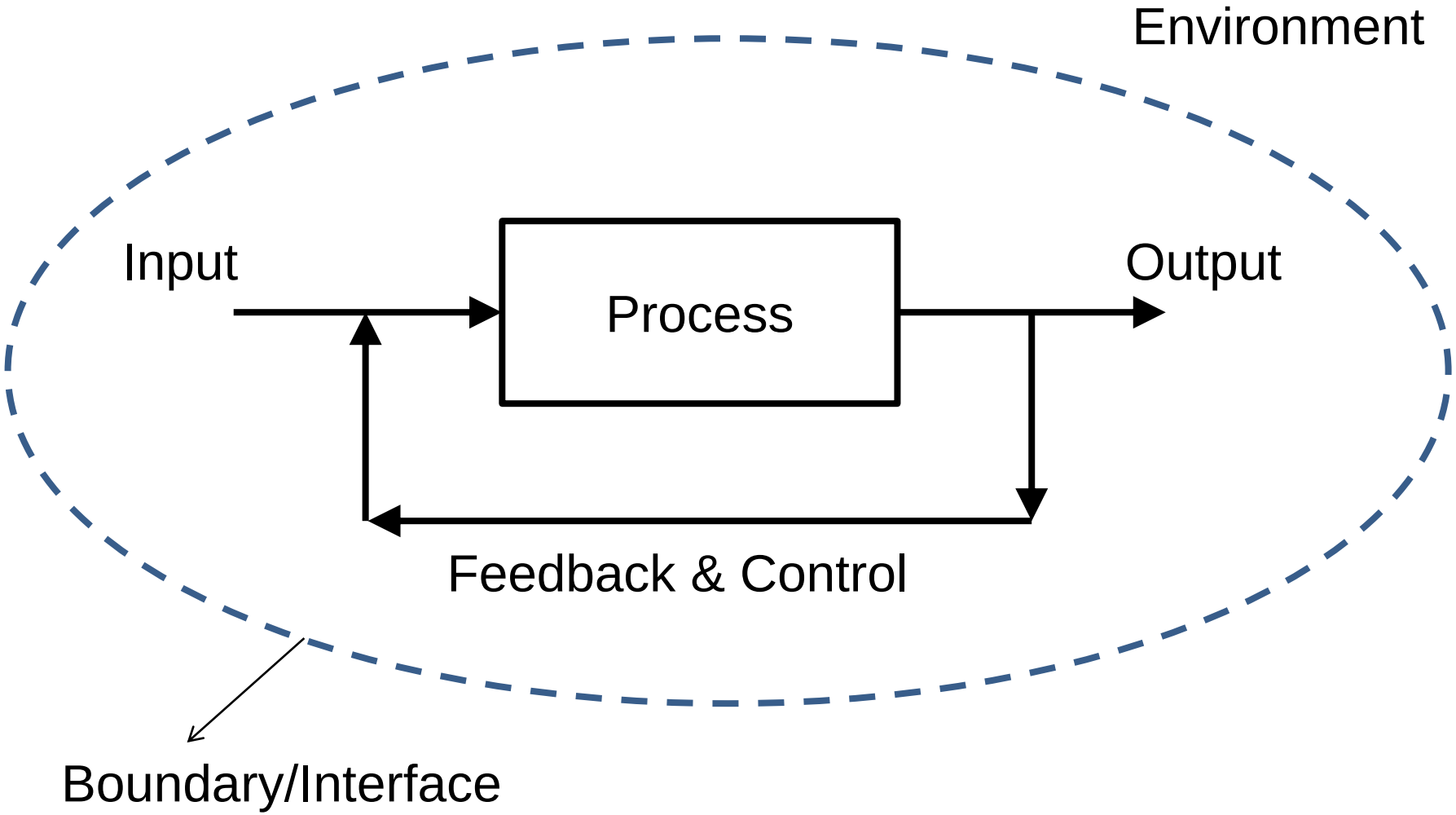
https://www.chatbots.org/chatbot/a.l.i.c.e/

Copeland (2000)

What is Systems?

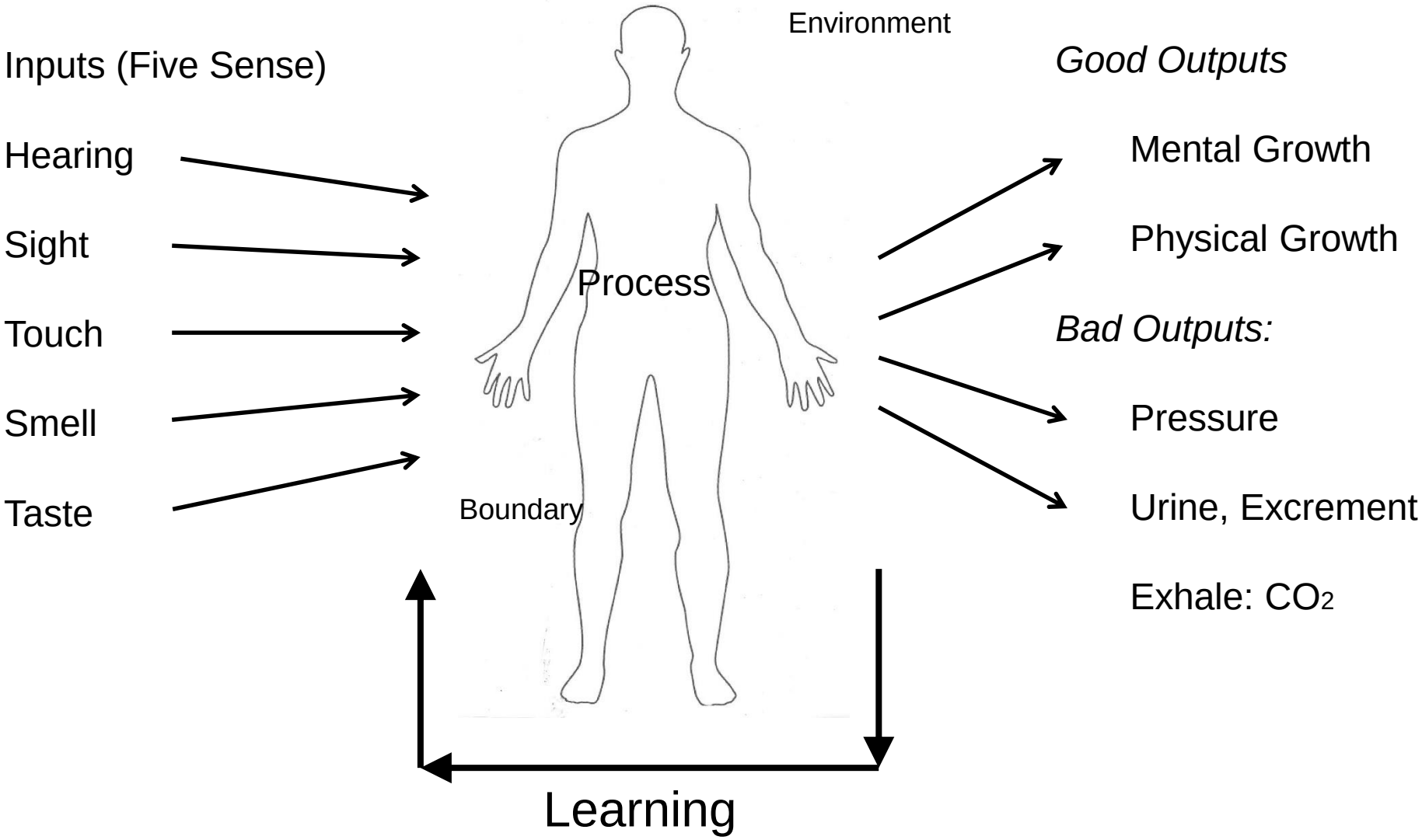
Systems is …

- What is “Systems”?



Human Body

Human Body System



Combine together is …

Intelligent + Manufacturing + Systems = ?

Definition of IMS

Definition

- IMS is a system which improves productivity by systematizing the intellectual aspect involved in manufacturing and flexibility by integrating the entire range of corporate activities so as to foster the optimum in relationship between men and intelligent machines (H. Yoshikawa, University of Tokyo).

- An intelligent manufacturing process has the ability to self-regulate and/or self-control to manufacture the product within its design specifications (Nam Suh, MIT)

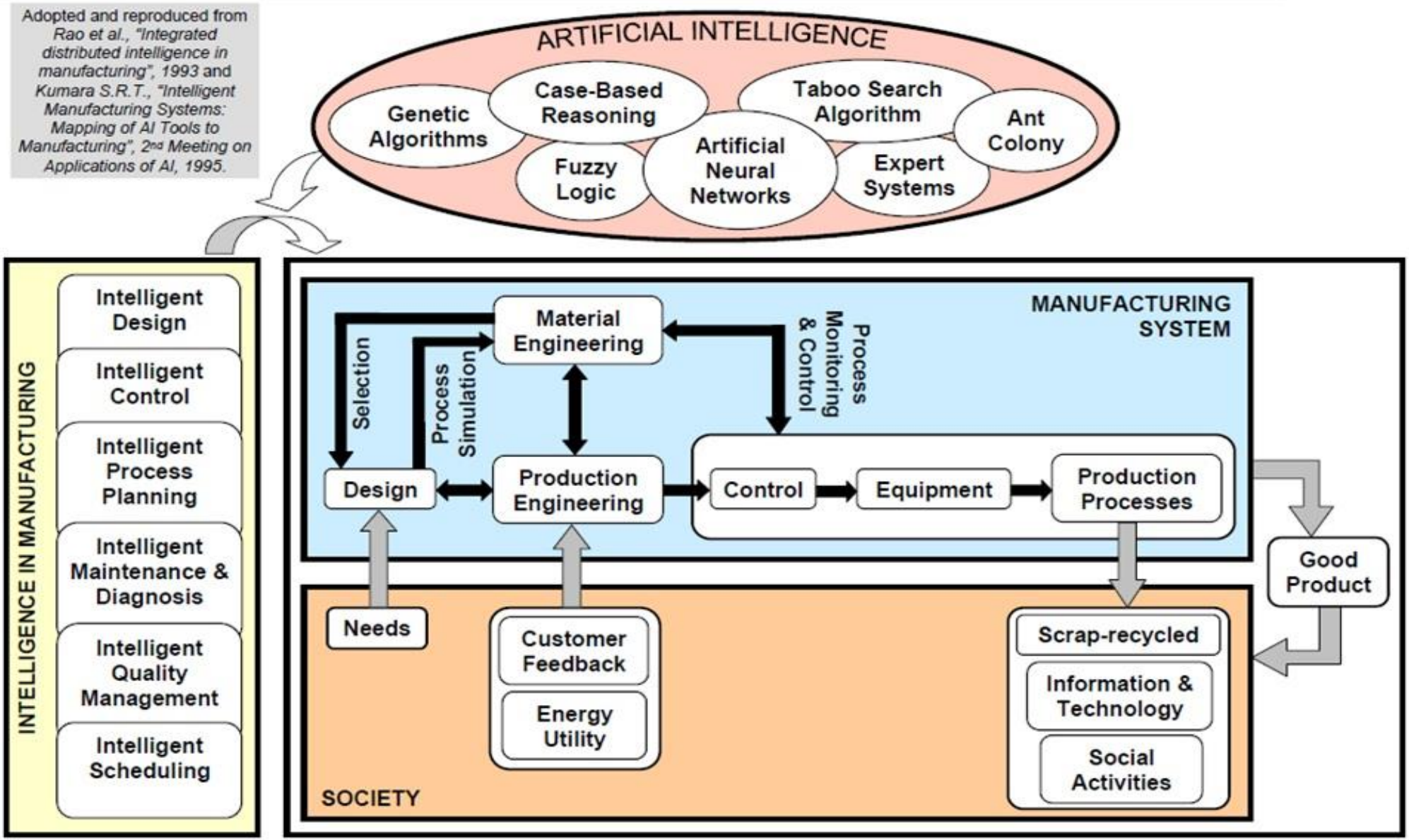
智慧工廠

IMS (smart factory) is a **decision-oriented** system which has the computational intelligence and self-learning ability to optimize the manufacturing process.

計算智慧 → Based on Data (資料處理與分析)

自我學習 → Real-time Feedback Control (回饋控制)

IMS Framework



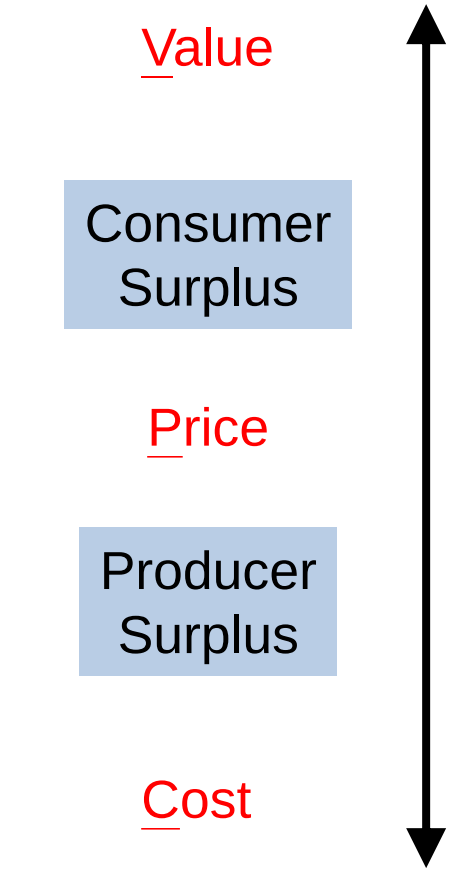
Tolga Bozdana, A., 2012. Intelligent Manufacturing Systems, Course Lecture Notes, Dept. of Mechanical Engineering, University of Gaziantep, Turkey.

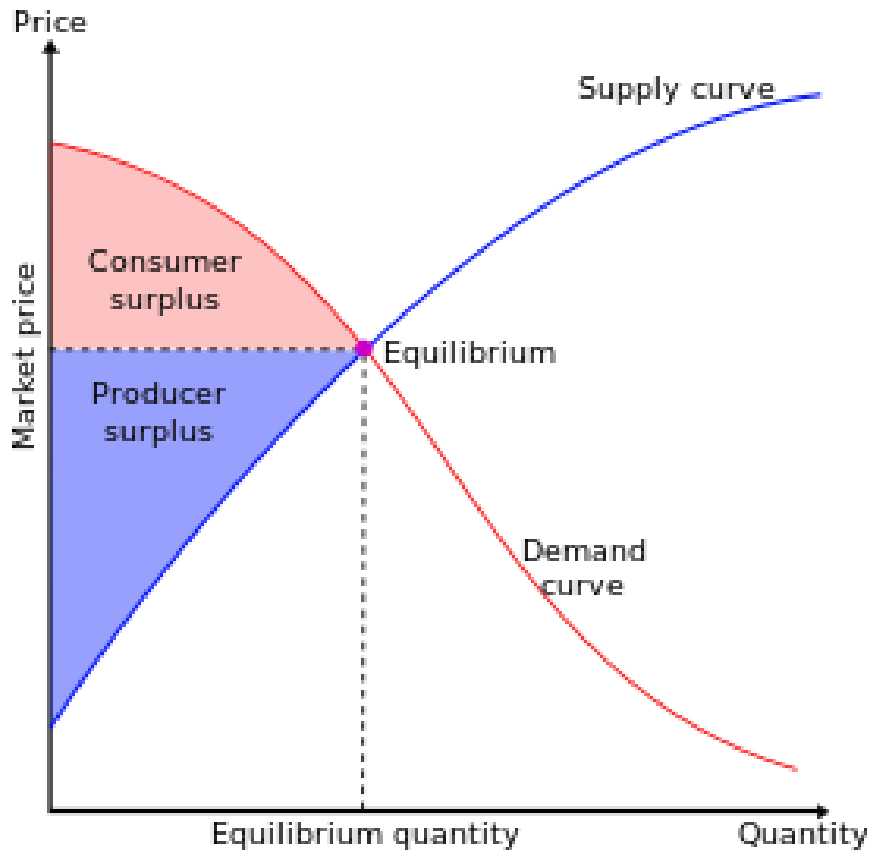
Manufacturing System and Automation

(製造系統與自動化)

Value and Profit

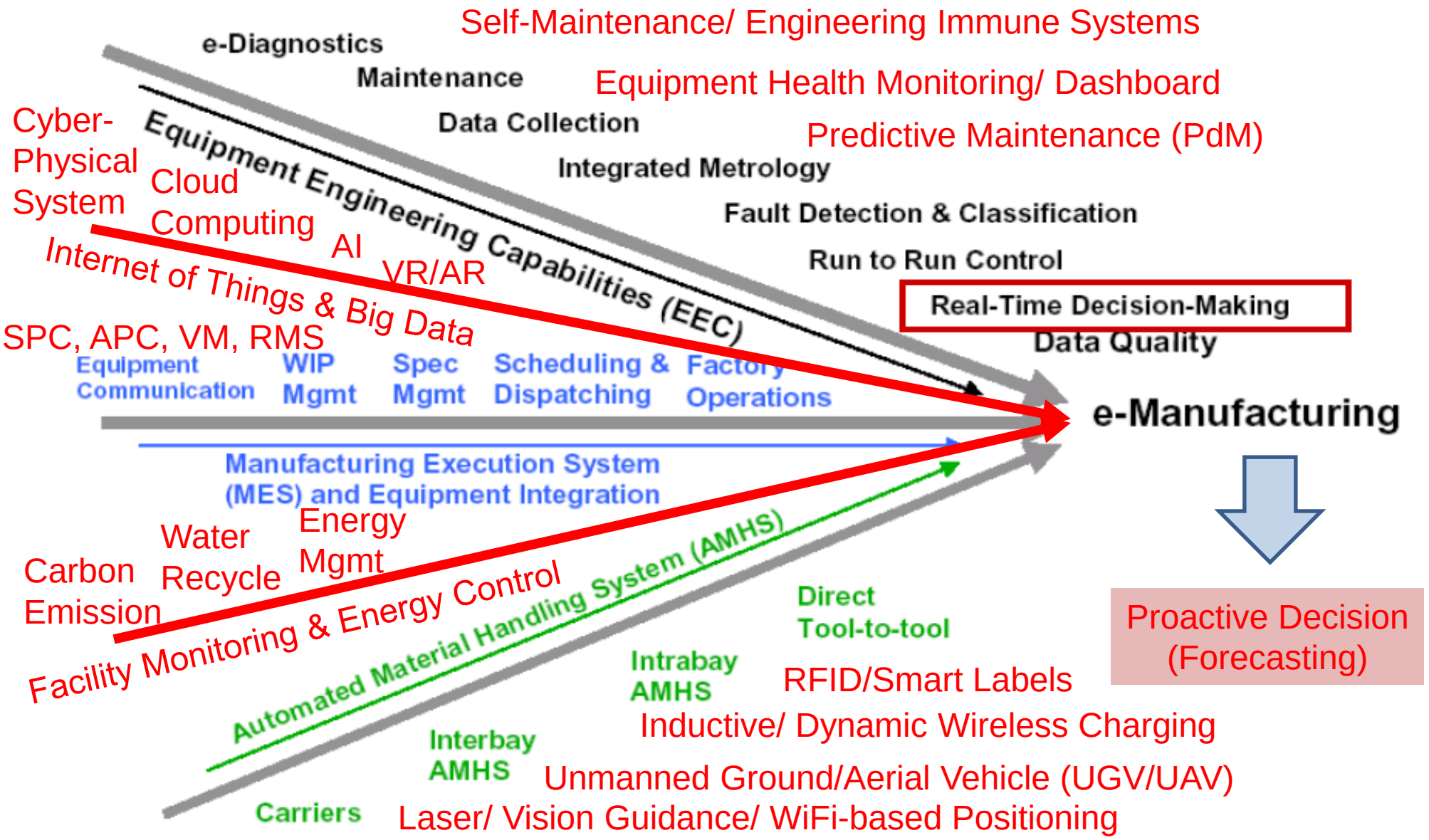
- Value
  - Manufacturing is value-added process for the realization of the product
  - Value – Price – Cost (VPC) framework
  - Customer Surplus vs. Producer Surplus





Wikipedia (2016)

製造智慧：決策



Source: International SEMATECH e-Diagnostics and EEC Guidance 2003

Manufacturing vs. Service

- Difference between Manufacturing and Service

System	Manufacturing	Service
Input	Material	Labor
Process	Capital-intensive Build SOP	Labor-intensive Building SOP is difficult
Output	Physical product	Intangible Non-separable Non-stable Non-storable Time-concerned
feedback	Quality Control Performance criteria	Difficult to measure quality and performance

- Now, manufacturing system moves toward service system.

Manufacturing system vs. Service system

- Manufacturing system vs. Service system
  - Main difference is "Inventory"!

- Inventory
  - Raw materials, Components, Work-in-process (WIP), Finished goods

Capacity + Lead Time + Uncertainty = Inventory

- Inventory Reduction
  - Reduce material and production lead time (includes transport)
  - Reduce information delay times (長鞭效應, Bullwhip effect)
  - Improve quality of information (reduce uncertainty)

半導體製造管理指標

1. 生產效率

- 提高系統產出量[5, 18, 19]
- 系統產能的穩定
- 精準的在製品控制[5, 10, 20]
- 較高效率的生產[16, 20]
- 減少生產流程時間[10,23]
- 降低每層加工的生產流程時間[18]
- 減少機器閒置
- 提高直接員工生產力[18]
- 提高員工總生產力[18]
- 提高瓶頸機台的效能[18]
- 提高設備使用率[19]
- 平衡生產線
- 縮短定貨的前置時間

2. 生產品質

- 良好的生產品質[8, 16, 21]
- 提高生產線良率[18]
- 提高晶粒良率[18]
- 良好的生產控制[5, 8, 19]
- 避免錯誤操作次數
- 機器的預防保養
- 提高機台穩定度
- 人員操作效率[16]

3. 服務

- 生產交期的達成[17, 18]
- 最少化客戶抱怨

4. 成本

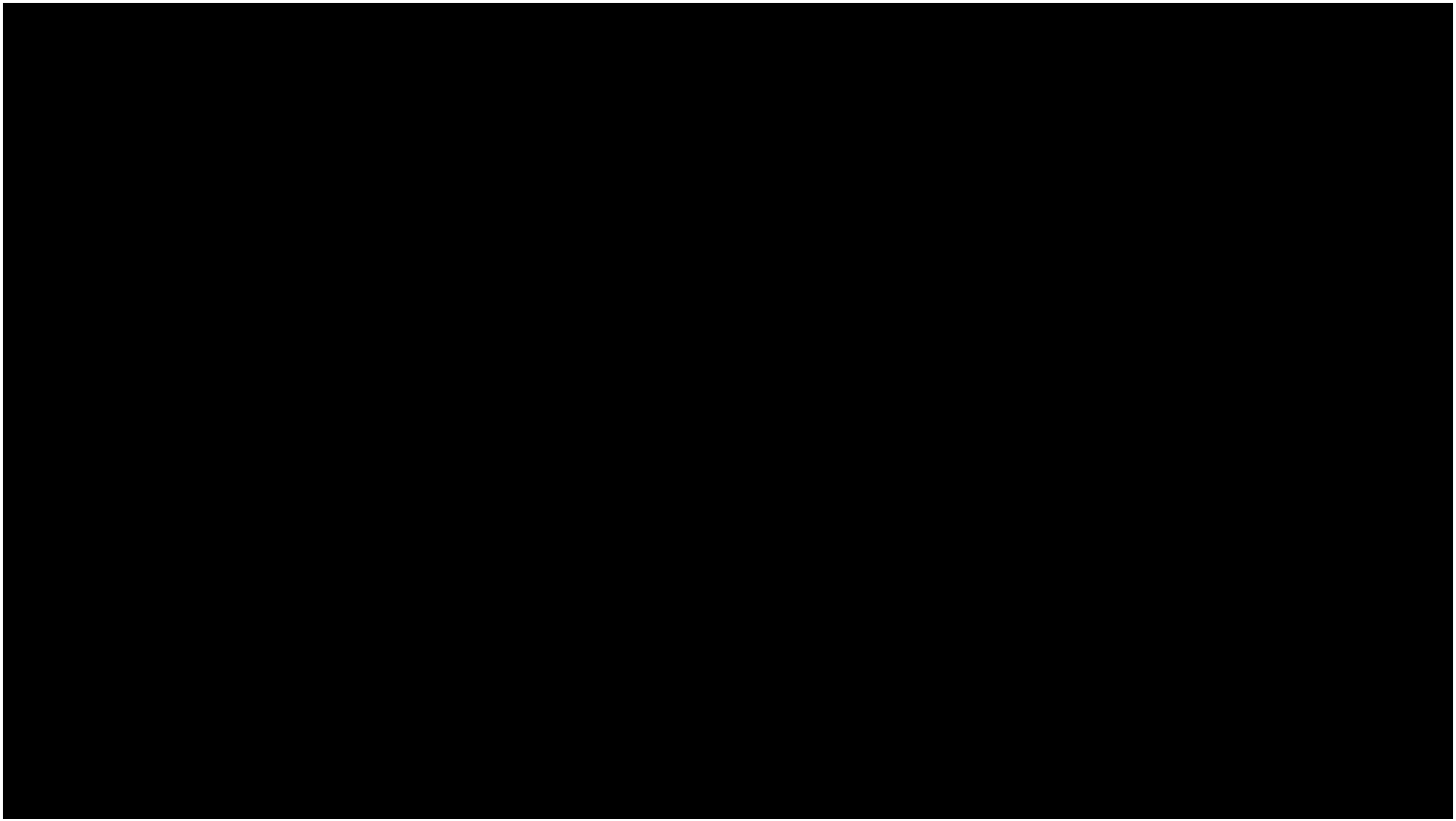
- 較低的生產成本[18]
- 降低每片晶圓的平均成本
- 降低晶圓每層加工的平均成本

Chien et al. (2004)

Let us see what others have done about MANUFACTURING SYSTEMS (Video Time)

Screw Manufacturing Process

- https://www.youtube.com/watch?v=a872eiTjw1w

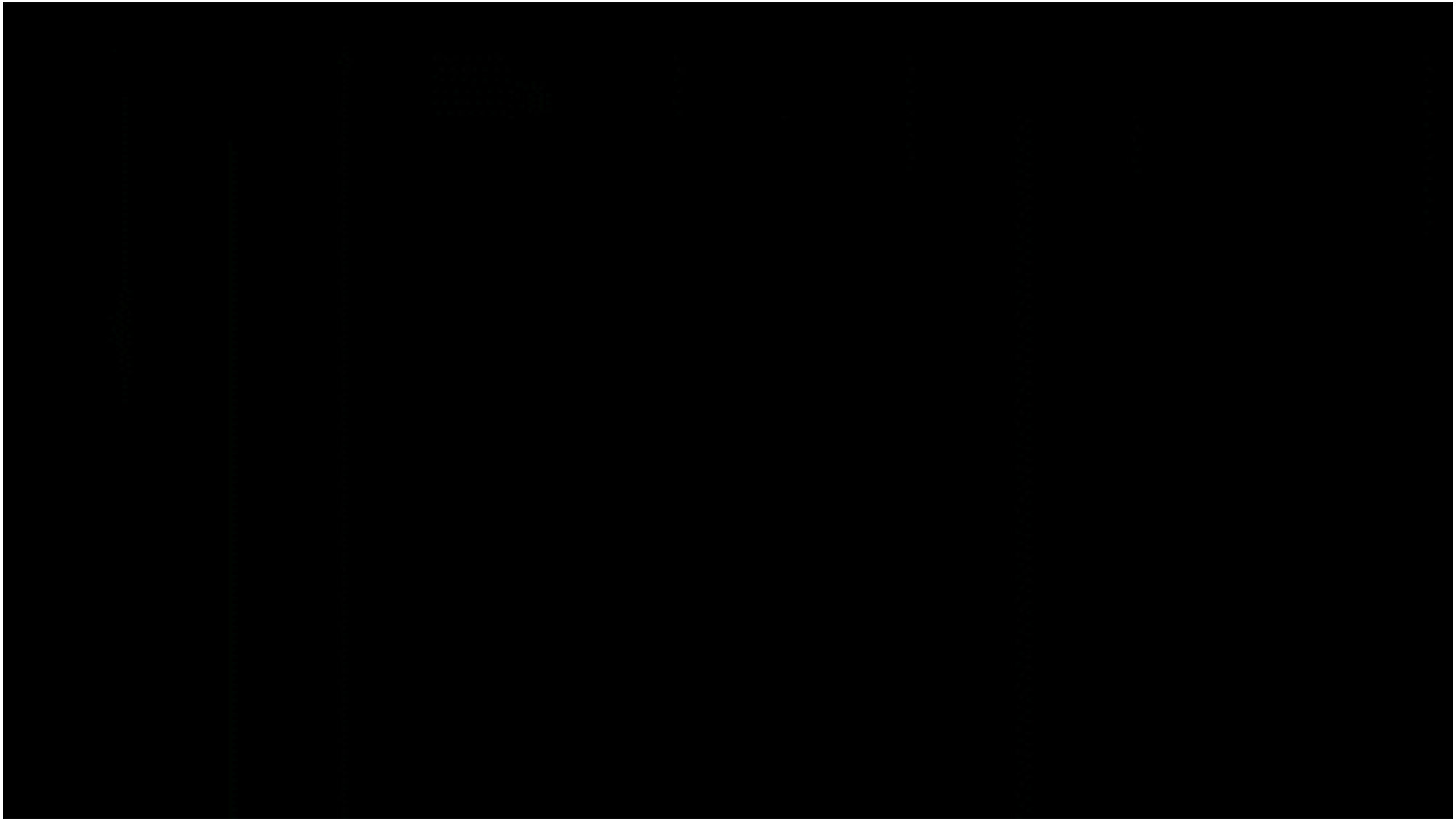


Vehicle Manufacturing System

- https://www.youtube.com/watch?v=xPKRTQbLs0w

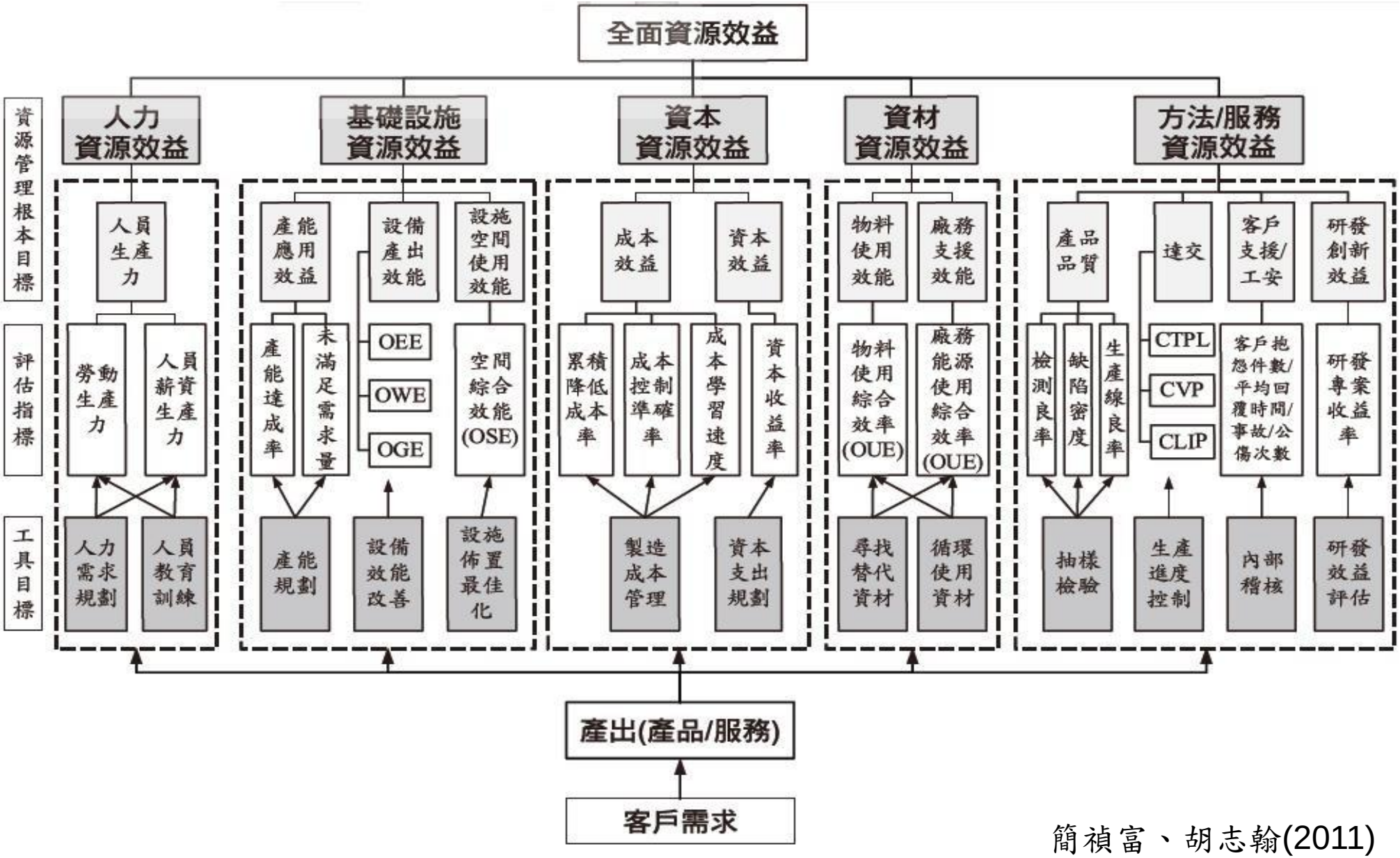
Aircraft Manufacturing System

- https://www.youtube.com/watch?v=WyDy-KdMGEA



Why is it difficult to manage a manufacturing systems?

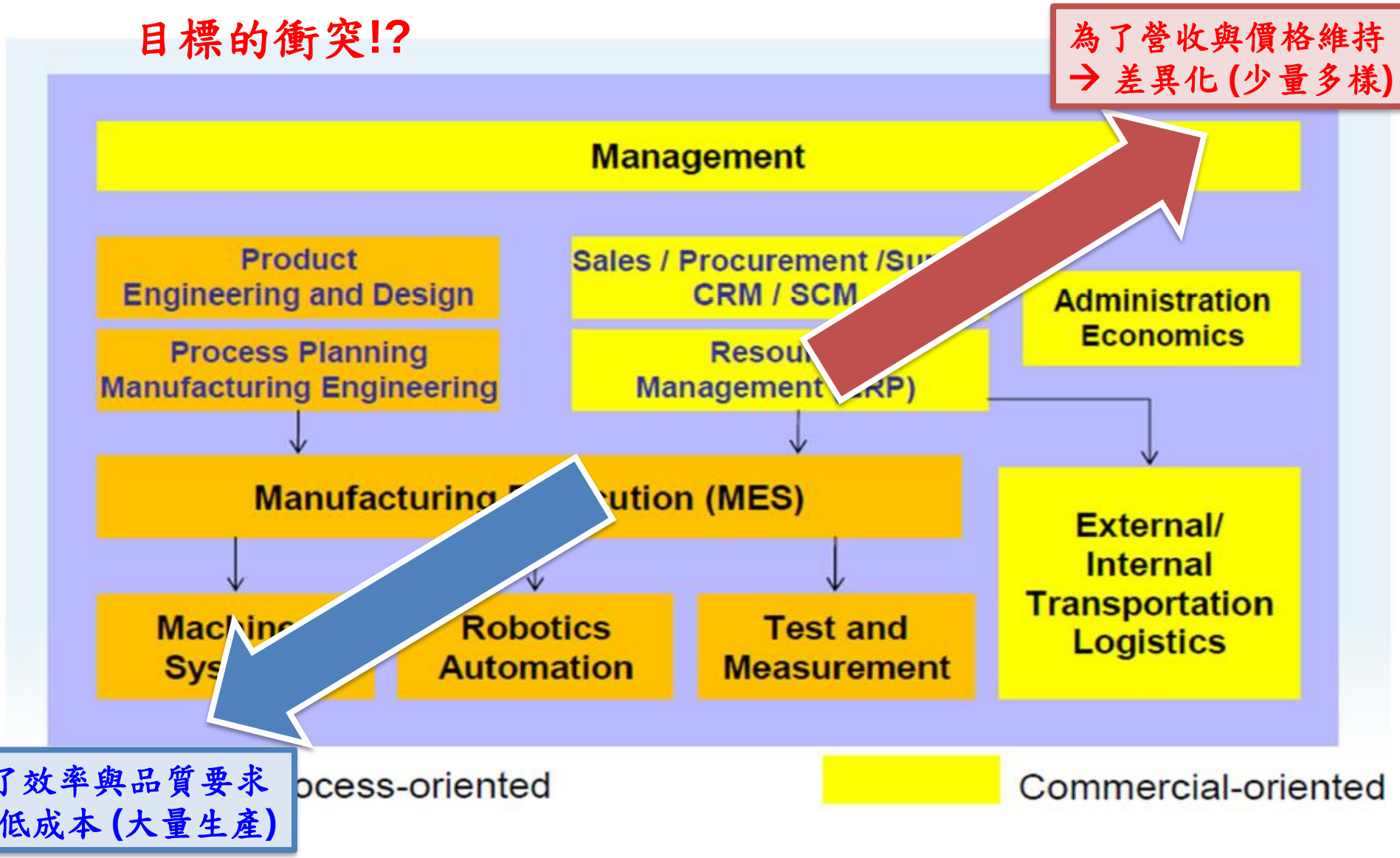
Diverse Resources and KPIs (資源與KPI的多樣性)



簡禎富、胡志翰(2011)

管理階層與製造現場間目標的衝突 (ICT 系統導入的挑戰)

目標的衝突!?

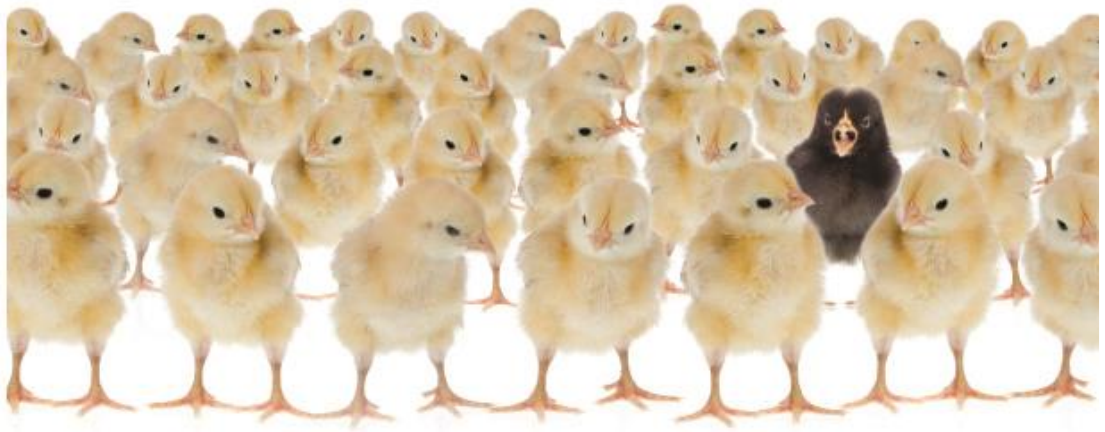


Prof.Dr. Engelbert Westkämper, Fraunhofer IPA Stuttgart, Germany, “Factories of the Future beyond 2013: The role of ICT”
http://cordis.europa.eu/fp7/ict/micro-nanosystems/docs/fof-beyond-2013-workshop/westkaemper-manufuture_en.pdf

Variability (變異)

- Variability is anything that causes the system to depart from regular, predictable behavior.
- Sources of Variability:

setups	workpace variation
machine failures	differential skill levels
materials shortages	engineering change orders
yield loss	customer orders
rework	product differentiation
operator unavailability	material handling



Variability from Resource!

現場「變異(variability)」是不好的!!?

但沒有變異，如何”改善”?
就如同考試成績有好有壞，才能對症下藥~

其實...

變異某種程度上也代表了"資訊量"!?

	English	Math
Student_A	80	76
Student_B	80	91
Student_C	80	83
Student_D	80	62
Student_E	80	88

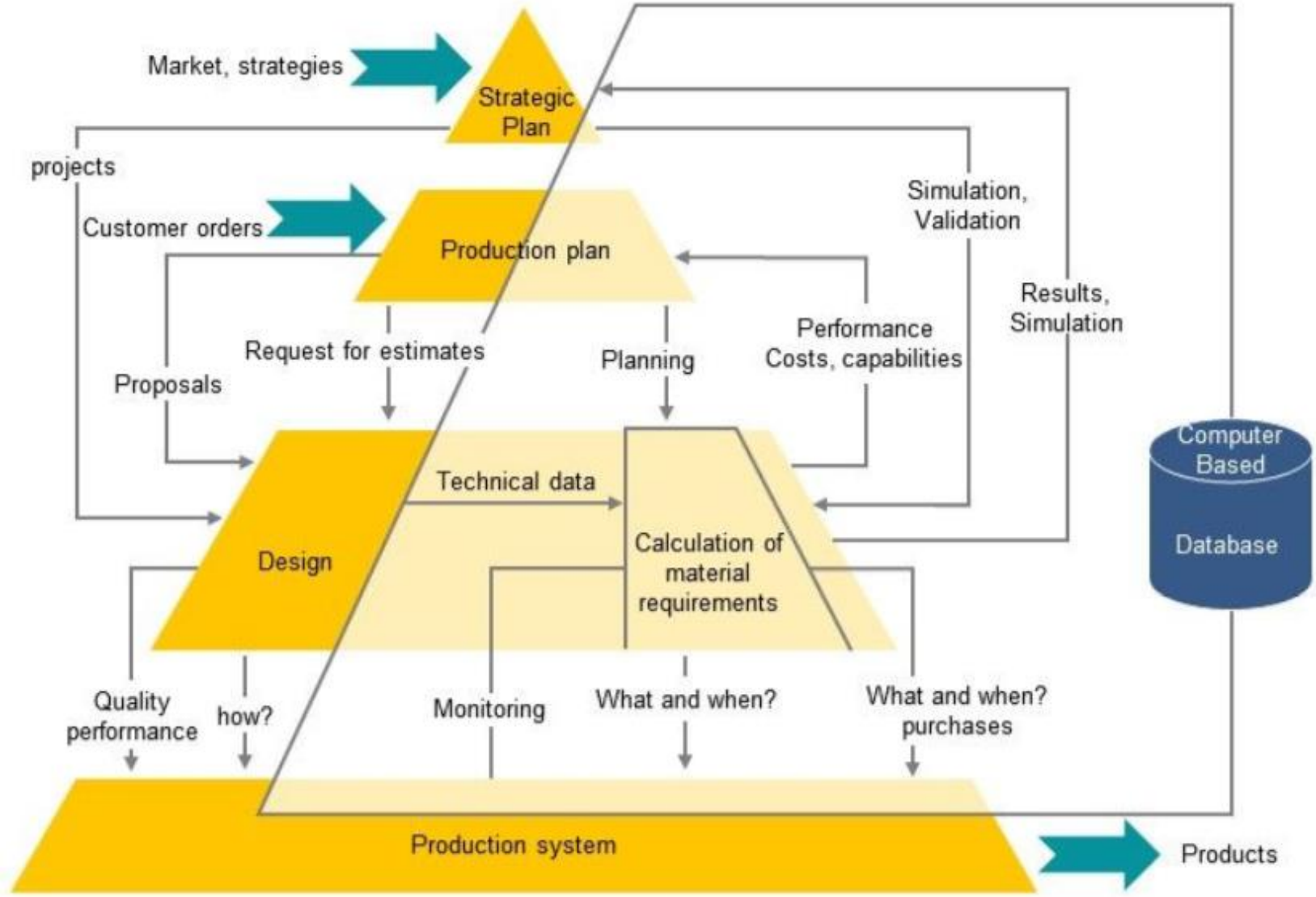
CIM and Automation

- Computer-integrated manufacturing (CIM)
  - manufacturing approach of using computers to control entire production process.
- Distinguish CIM from other manufacturing methodologies:
  - Means for data storage, retrieval, manipulation and presentation;
  - Mechanisms for sensing state and modifying processes;
  - Algorithms for uniting the data processing component with the sensor/modification component.
- CIM is an example of the implementation of information and communication technologies (ICTs) in manufacturing.
- Key challenges
  - Integration of components from different suppliers: When different machines, such as CNC, conveyors and robots, are using different communications protocols (In the case of AGVs, even differing lengths of time for charging the batteries) may cause problems.
  - Data integrity: The higher the degree of automation, the more critical is the integrity of the data used to control the machines. CIM requires extra human labor in ensuring that there are proper safeguards for the data signals that are used to control the machines.
  - Process control: Computers may be used to assist the human operators of the manufacturing facility, but there must always be a competent engineer on hand to handle circumstances which could not be foreseen by the designers of the control software.

Craig, W., and Leonard, A. (2019), Manufacturing engineering and technology, ED-Tech Press.

CIM and Automation

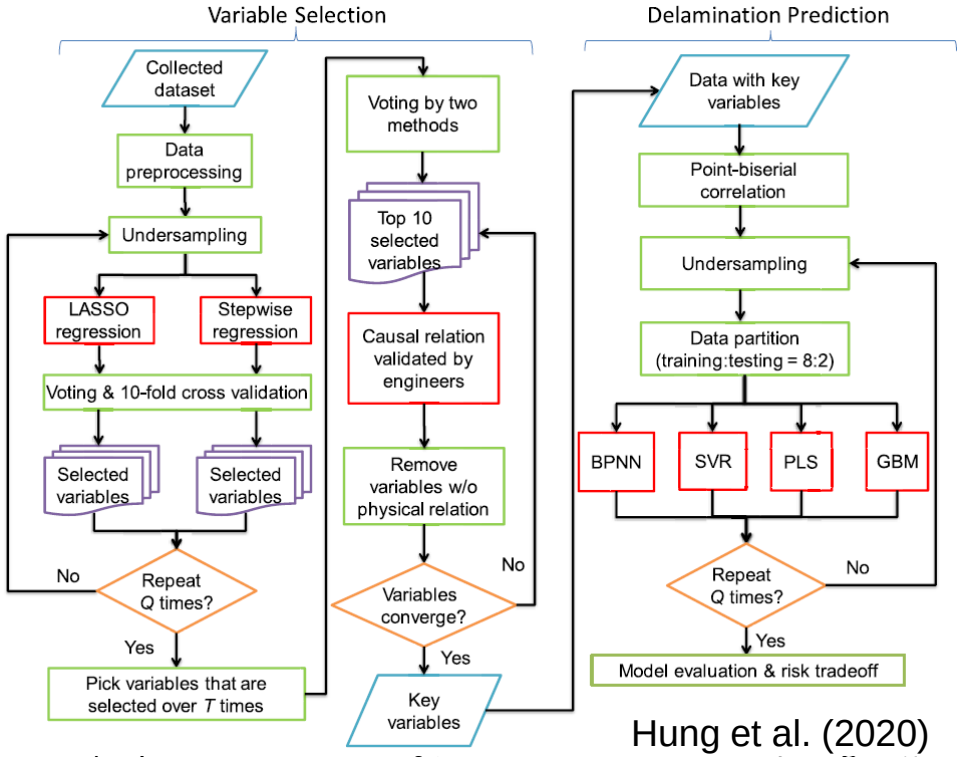
- CIM & production control system



Waldner, J.-B. (1992). Principles of Computer-Integrated Manufacturing. London: John Wiley & Sons. pp. 128–p132.

CIM and Automation

- Automatic Framework
  - Data integration: consolidate data from multiple data sources
  - Algorithm design: design ML/DS algorithms to analyze the dataset
  - Machine learning: systems gain knowledge from dataset
  - Process automation: system provides decision/solution to complete the process
  - Modularization: customize and modularize the processes to generate a total solution
- Step-by-Step Description for Automation: flow chart, pseudocode, …



Hung et al. (2020)

Algorithm *Random Forest for Regression or Classification.*

1. For $b = 1$ to B:
   (a) Draw a bootstrap sample $\mathbf{Z}^*$ of size N from the training data.
   (b) Grow a random-forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size n_{min} is reached.
      i. Select m variables at random from the p variables.
      ii. Pick the best variable/split-point among the m.
      iii. Split the node into two daughter nodes.
2. Output the ensemble of trees $\{T_b\}_1^B$.

To make a prediction at a new point x:

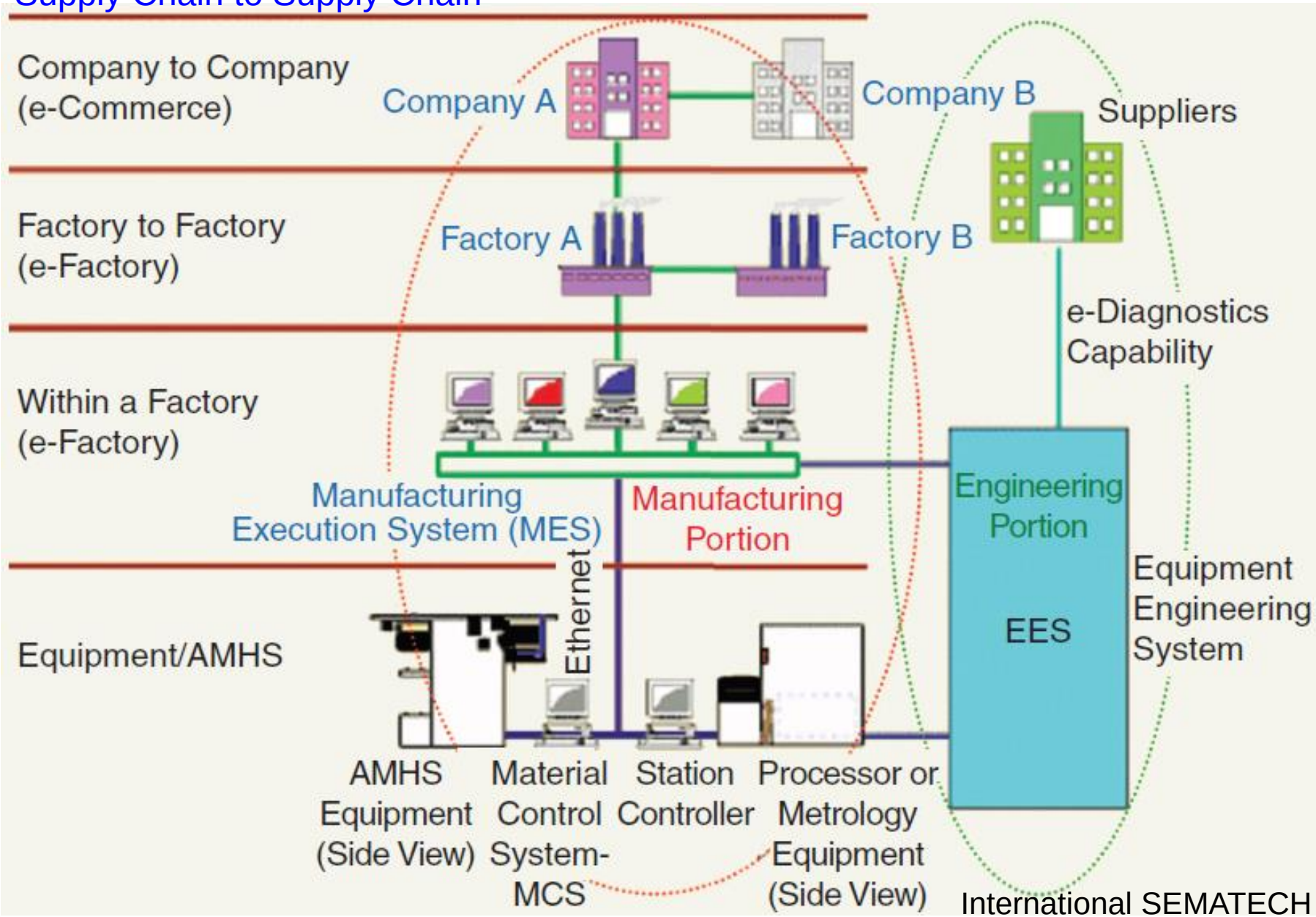
Regression: $\hat{f}_{\text{rf}}^B(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$.

Classification: Let $\hat{C}_b(x)$ be the class prediction of the bth random-forest tree. Then $\hat{C}_{\text{rf}}^B(x) = \textit{majority vote}\ \{\hat{C}_b(x)\}_1^B$.

Hastie et al. (2009)

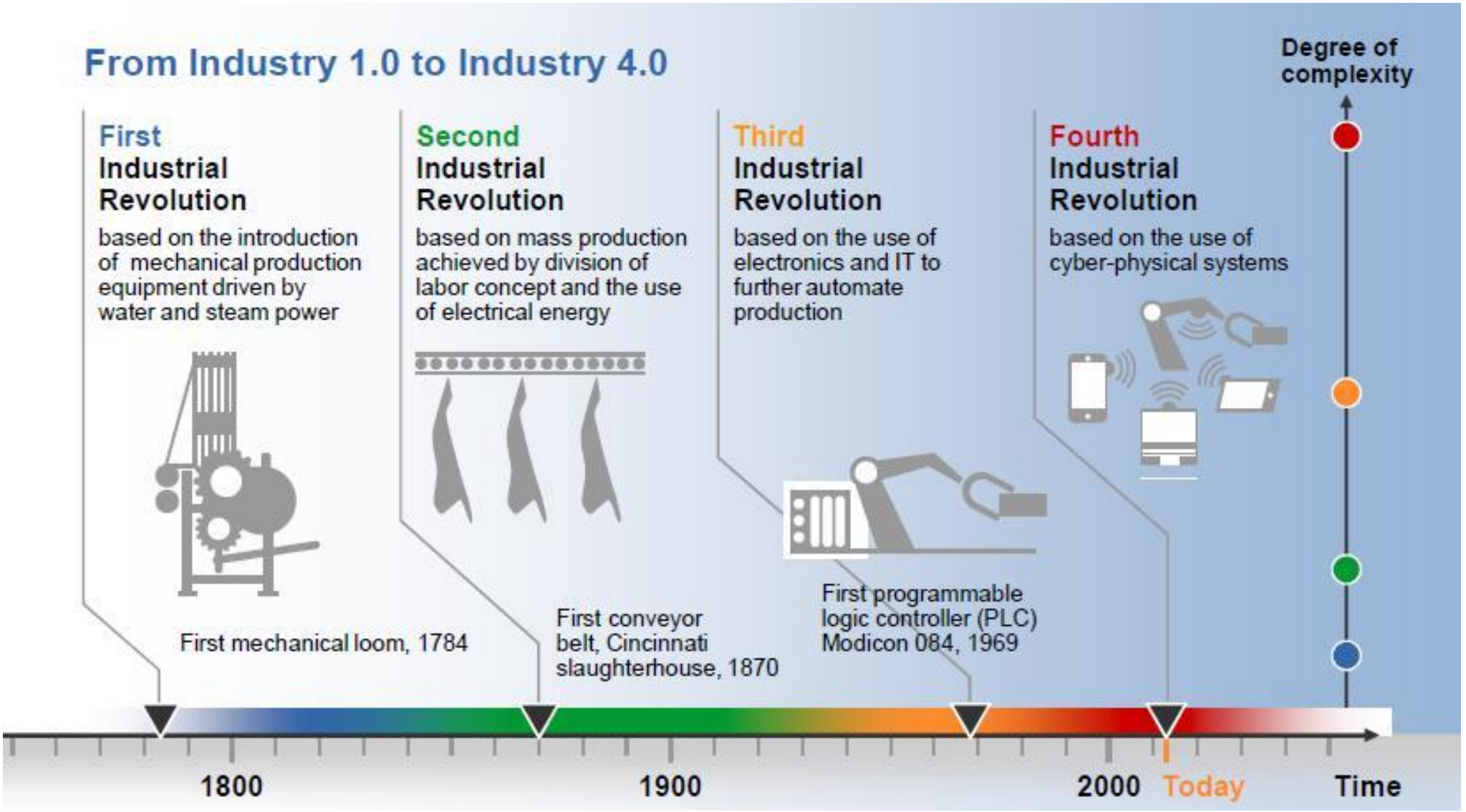
Automation Hierarchy

Supply Chain to Supply Chain



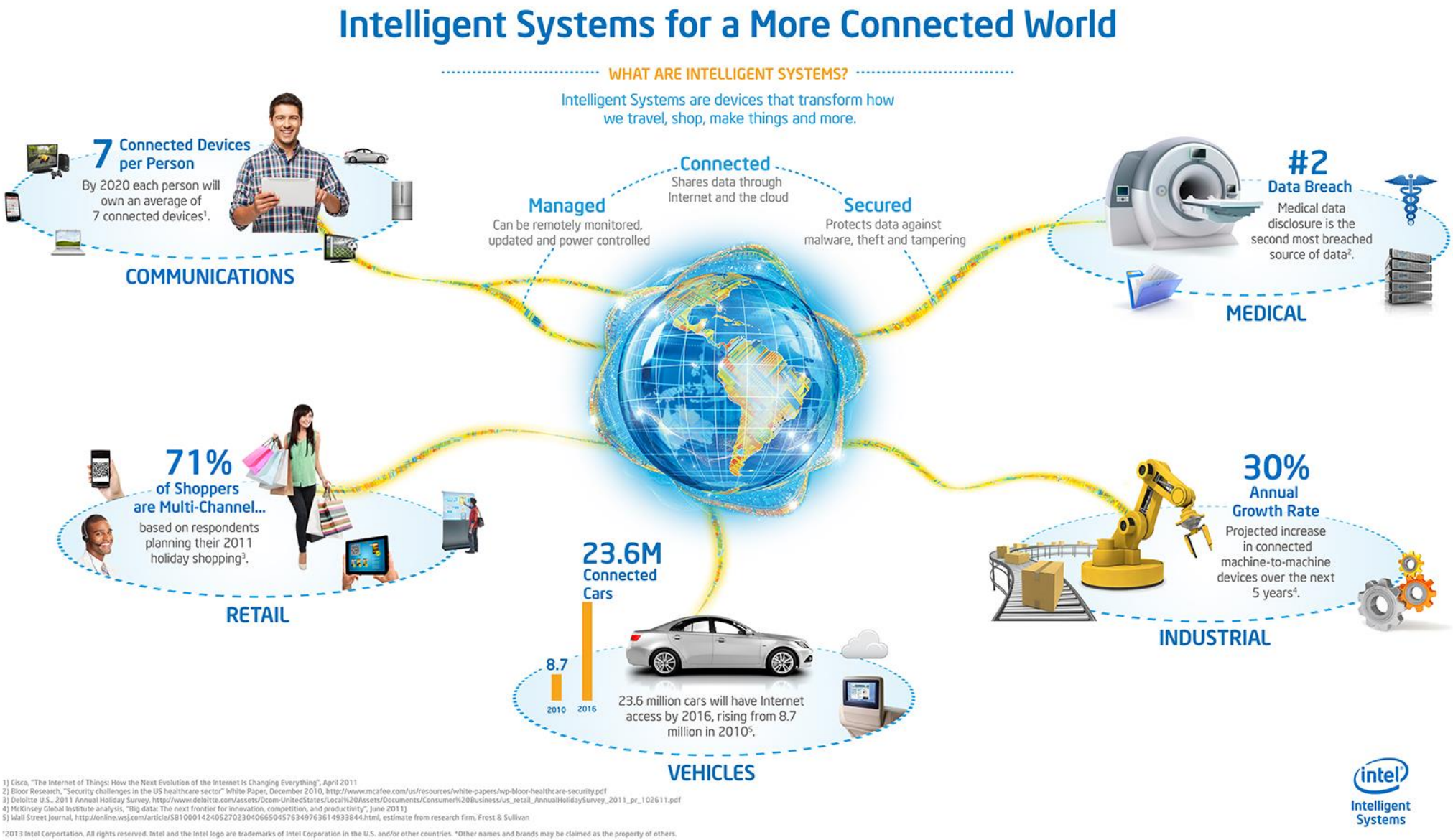
International SEMATECH (2002)

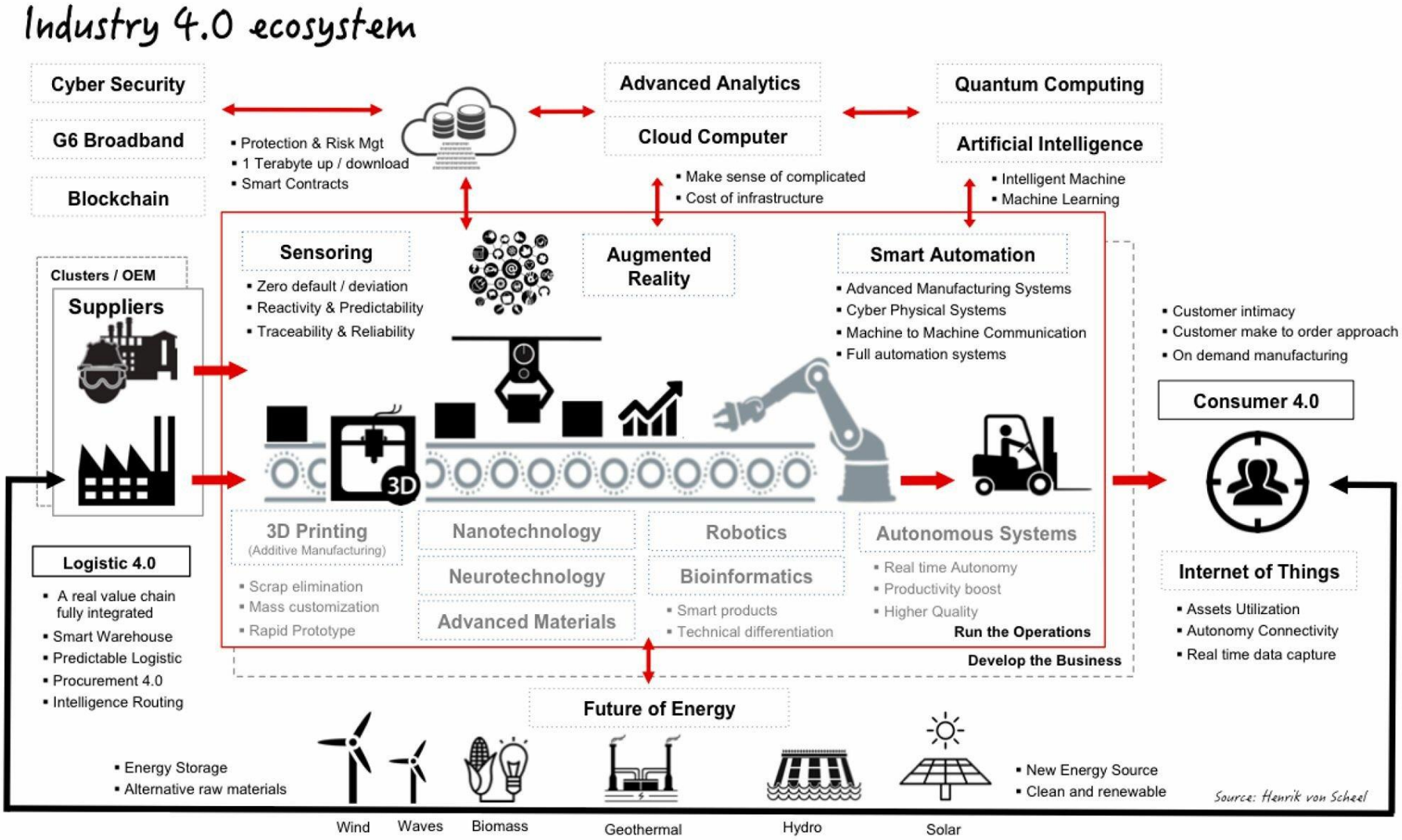
Industry 4.0



engineersjournal.ie (2014)

Internet of Things (IoT)





Henrik von Scheel, 2018. Industry 4.0 made easy. One view of the Industry 4.0 ecosystem. https://www.youtube.com/watch?v=7dU8unYYWjA

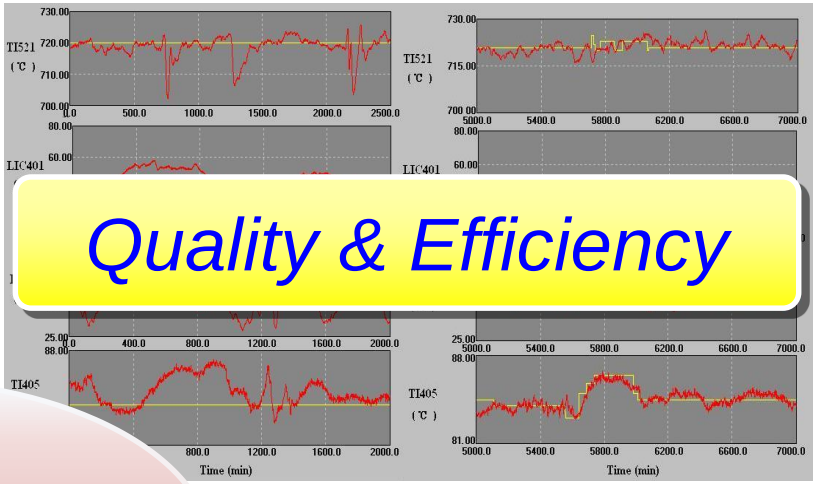
Data Science Functions & Applications

CIM and Automation

MES

On-Time Delivery

APC



Quality & Efficiency

Big Data Analytics

AMHS

Cycle Time Reduction

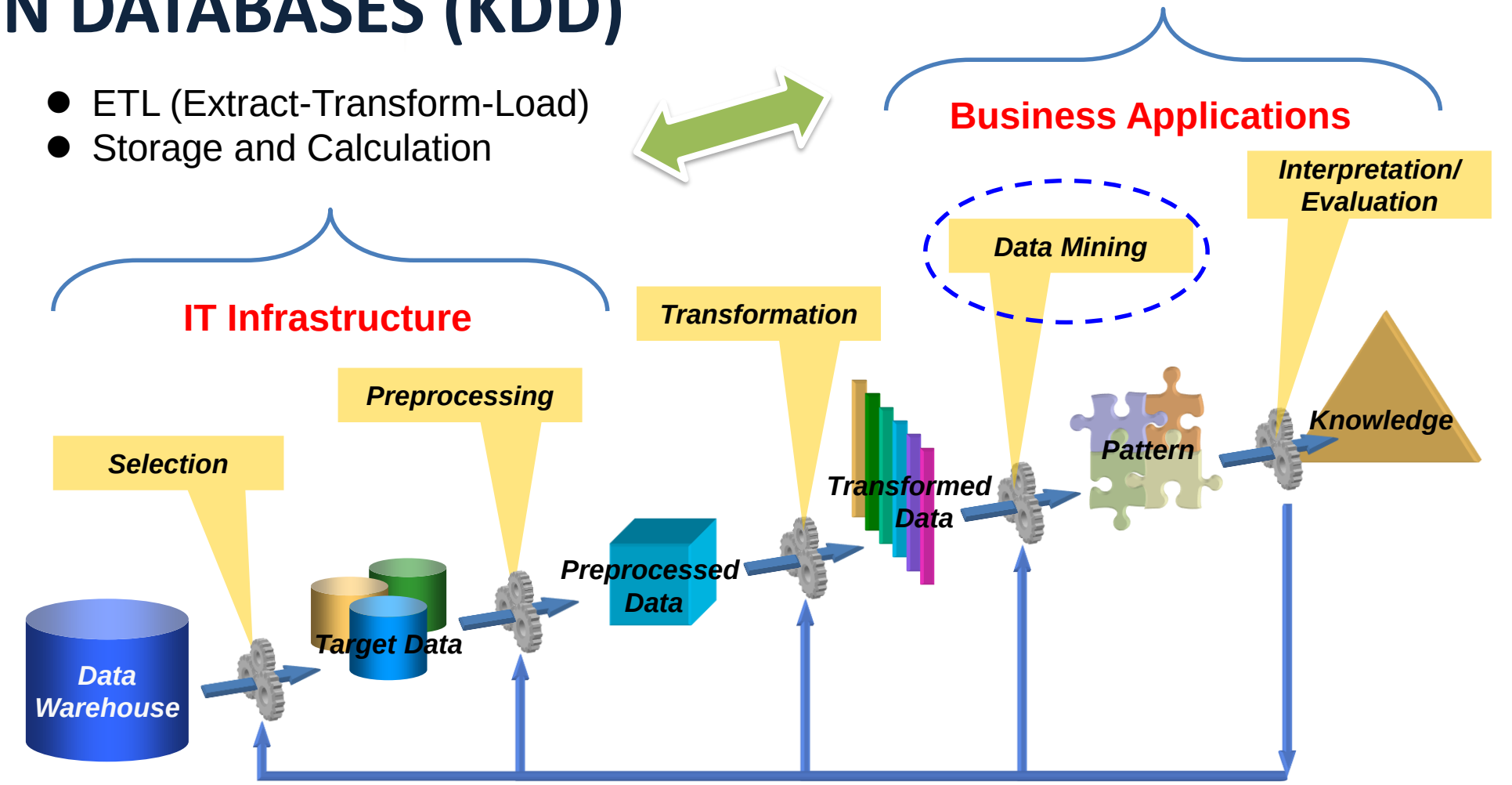
FMEC

Green Environment

KNOWLEDGE DISCOVERY IN DATABASES (KDD)

- ETL (Extract-Transform-Load)
- Storage and Calculation

- Pattern Extraction
- Value Interpretation for **Profitability**



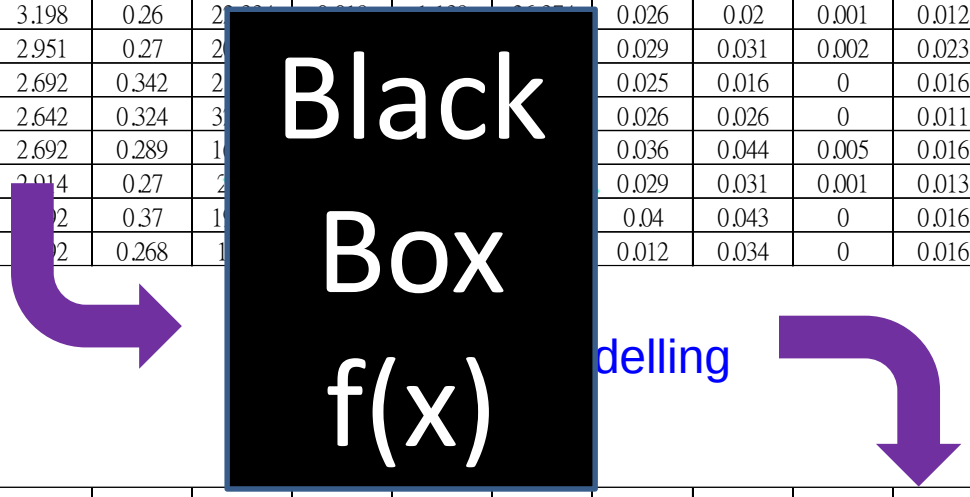
hadoop Spark python SPSS sas R

Source :"From Data Mining to Knowledge Discovery: An Overview", Advances in Knowledge Discovery and Data Mining, AAAI Press/The MIT Press 1996.

Prediction of ML/DS

- Generally, the Prediction of ML/DS…
  - Supervised Learning, Unsupervised Learning, Reinforcement Learning

	Var_1	Var_2	Var_3	Var_4	Var_5	Var_6	Var_7	Var_8	Var_9	Var_10	Var_11	Var_12	Var_13	Var_14	Var_15	Var_16	Var_17	Class
0 bs1	-5349	19.8	20.2	0	2.713	0.324	24.069	0.011	2.092	24.301	0.024	0.031	0.002	0.01	6.653	24.478	295.528	1
0 bs2	-5597	11.682	28.318	0	2.713	0.319	22.302	0.013	3.949	23.668	0.029	0.032	0	0.01	5.942	23.592	90.394	1
0 bs3	-5440.5	22.435	17.566	0	2.713	0.35	21.646	0.016	0.685	23.346	0.023	0.013	0	0.01	3.073	19.719	101.866	1
0 bs4	-5614.25	17.163	22.837	0	2.713	0.29	17.521	0.016	1.066	27.508	0.021	0.032	0.002	0.01	2.635	27.749	224.542	1
0 bs5	-5534.5	25.457	14.543	0	3.198	0.35	22.798	0.018	1.619	29.305	0.042	0.024	0.001	0.012	3.533	18.54	162.479	1
0 bs6	-2649.25	20.551	19.449	0	2.914	0.324	20.481	0.022	1.411	21.722	0.03	0.022	0.001	0.013	2.526	14.507	0	1
0 bs7	-5554.25	18.683	21.317	0	3.198	0.38	33.628	0.023	1.641	19.756	0.046	0.034	0	0.012	3.513	24.921	147.607	1
0 bs8	-6566	25.443	14.557	0	2.951	0.403	16.265	0.016	1.947	21.162	0.029	0.04	0	0.023	5.361	21.856	0	1
0 bs9	-5613.75	17.637	22.363	0	2.914	0.247	20.281	0.031	1.183	16.635	0.039	0.022	0	0.013	7.098	23.817	913.882	1
0 bs10	-6546.5	24.351	15.649	0	2.951	0.403	15.373	0.015	4.217	30.47	0.023	0.033	0.006	0.023	4.796	32.937	183.158	1
0 bs11	-5652	12.063	27.937	0	2.951	0.27	22.971	0.022	4.47	25.754	0.026	0.029	0	0.023	3.714	24.679	526.739	1
0 bs12	-6656.75	9.523	30.477	0	2.914	0.287	20.734	0.02	1.612	13.433	0.026	0.033	0	0.013	5.745	16.663	219.026	1
0 bs13	-5681.5	22.925	17.076	0	3.198	0.26	2				0.026	0.02	0.001	0.012	2.813	12.986	163.435	1
0 bs14	-5537.5	9.334	30.666	0	2.951	0.27	2				0.029	0.031	0.002	0.023	4.219	24.275	150.967	1
0 bs15	-5349.25	12.539	27.461	0	2.692	0.342	2				0.025	0.016	0	0.016	3.356	23.889	168.404	1
0 bs16	-5495.75	12.431	27.569	0	2.642	0.324	3				0.026	0.026	0	0.011	7.639	25.813	208.284	0
0 bs17	-6518.25	28.636	11.365	0	2.692	0.289	1				0.036	0.044	0.005	0.016	3.703	32.604	329.604	1
0 bs18	-4556	8.693	31.307	0	2.914	0.27	2				0.029	0.031	0.001	0.013	3.831	38.056	149.578	1
0 bs19	-5432.5	-5.306	25.306	0	2	0.37	1				0.04	0.043	0	0.016	3.584	33.733	977.346	1
0 bs20	-5677.75	-8.749	28.749	0	2	0.268	1				0.012	0.034	0	0.016	2.849	23.108	771.212	1
0 bs101	-5384.25	2.583	37.417	0	2.828	0.378	33.6	0.012	2.439	19.77	0.018	0.059	0	0.039	4.417	31.293	226.108	?

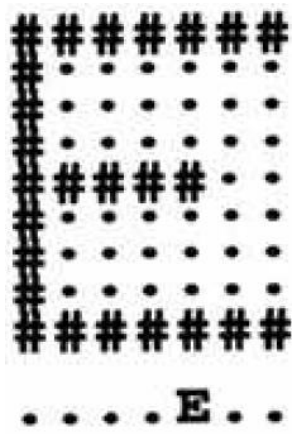


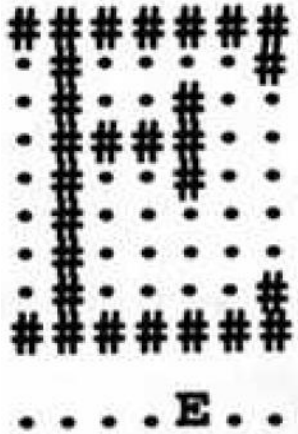
M. McCann, Y. Li, L. Maguire, A. Johnson, Causality Challenge: Benchmarkingrelevant signal components for effective monitoring and process control,Journal of Machine Learning Research: Workshop and ConferenceProceedings, 6 (2008) 277–288.

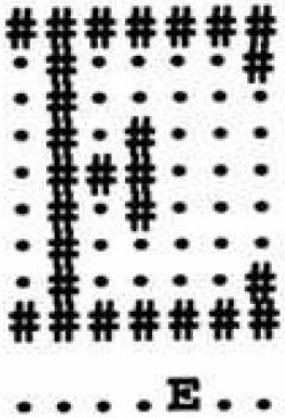
小試身手- Pattern Recognition

Handwritten Alphabet/Digit Recognition (Fausett, 1994)

- A perceptron to classify letters from different fonts: one output class
- 7 x 9 = 63 grid of pixels

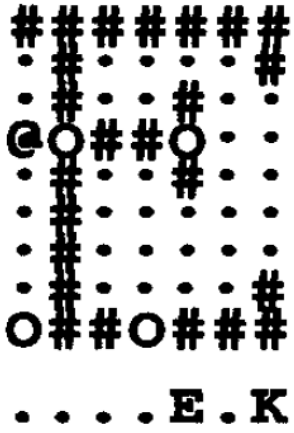


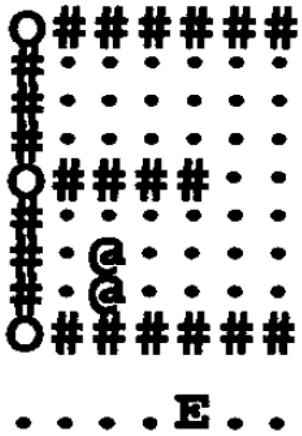




#: 1
•: 0

- With noise (@) and missing data (O)





If noise exists, use bipolar better!

: 1
• : -1
Noise: 0

小試身手- Pattern Recognition

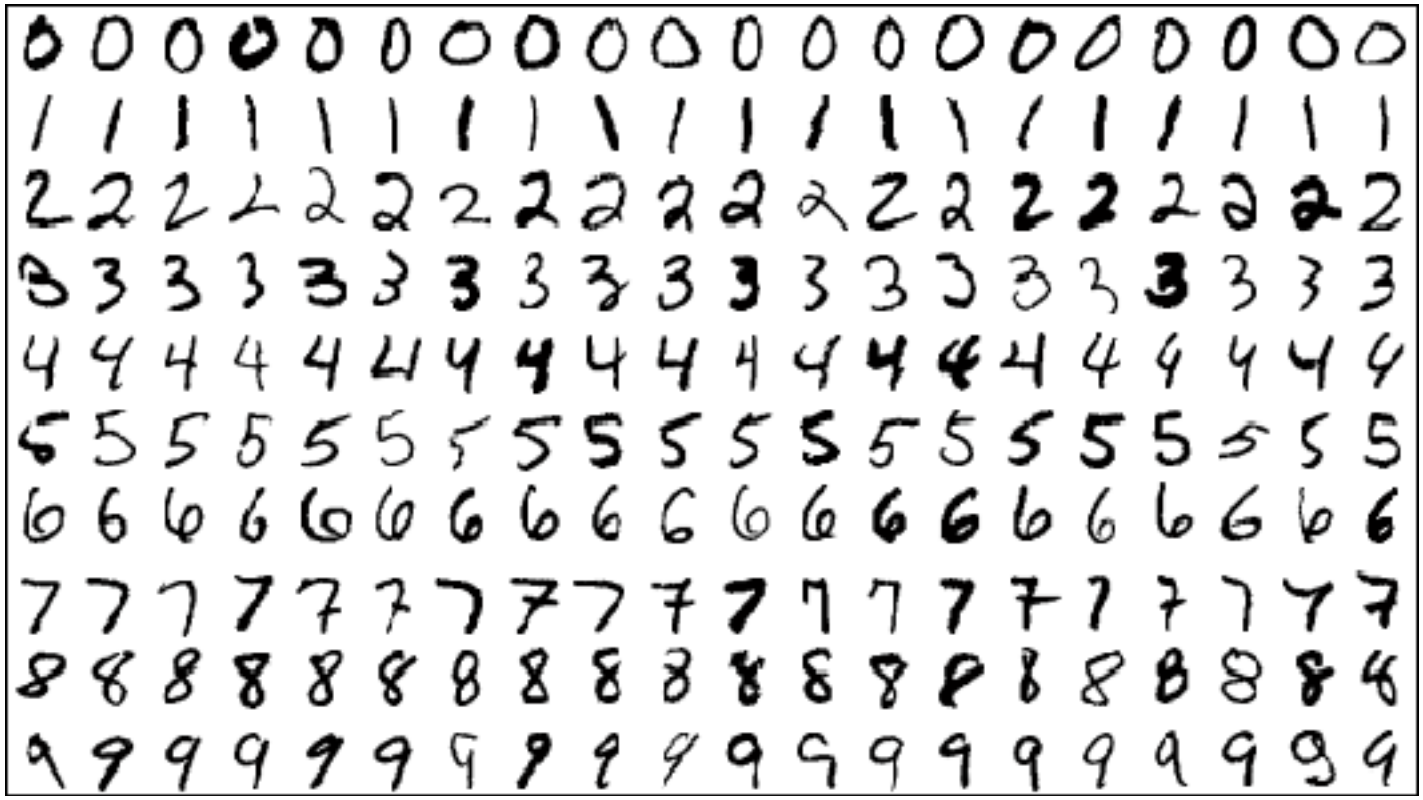
- Handwritten Alphabet/Digit Recognition
  - 16 x 16 = 256 grid of pixels

Pixel															
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64
65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96
97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112
113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128
129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144
145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160
161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176
177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192
193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208
209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224
225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256

1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0
0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0
0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
0	0	1	1	1	0	0	0	0	0	0	0	1	1	1	1
0	0	1	1	1	1	0	0	0	1	1	1	1	1	1	0
0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0

小試身手- Pattern Recognition

- Handwritten Alphabet/Digit Recognition
  - Data Collection- MNIST dataset



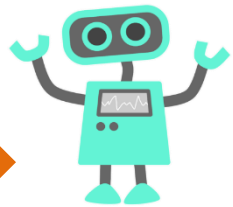
http://yann.lecun.com/exdb/mnist/

小試身手- Pattern Recognition

	A	B	C	D	E	F	G	H	I	J	K	IW	IX	IY	IZ	JA	JB	JC	JD	JE
1	No.	Pixel001	Pixel002	Pixel003	Pixel004	Pixel005	Pixel006	Pixel007	Pixel008	Pixel009	Pixel010	…	Pixel256	Target0	Target1	Target2	Target3	Target4	Target5	Target6
2	1	0	0	0	0	0	0	1	1	1	1	…	0	1	0	0	0	0	0	0
3	2	0	0	0	0	0	1	1	1	1	1	…	0	1	0	0	0	0	0	0
4	3	0	0	0	0	0	0	0	0	0	1	…	0	1	0	0	0	0	0	0
5	4	0	0	0	0	0	0	1	1	1	1	…	0	1	0	0	0	0	0	0
6	5	0	0	0	0	0	0	0	0	0	1	…	0	1	0	0	0	0	0	0
7	6	0	0	0	0	1	1	1	1	1	1	…	0	1	0	0	0	0	0	0
8	7	0	0	0	0	0	1	1	1	1	1	…	0	1	0	0	0	0	0	0
9	8	0	0	0	0	1	1	1	1	1	1	…	0	1	0	0	0	0	0	0
10	9	0	0	0	0	0	1	1	1	1	1	…	0	1	0	0	0	0	0	0
11	10	0	0	0	1	1	1	1	1	1	1	…	0	1	0	0	0	0	0	0
12	11	0	0	0	0	1	1	1	1	1	1	…	0	1	0	0	0	0	0	0
13	12	0	0	0	0	0	0	0	0	0	1	…	0	1	0	0	0	0	0	0
14	13	0	0	0	0	0	1	1	1	1	0	…	0	1	0	0	0	0	0	0
15	14	0	0	1	1	1	1	1	1	1	1	…	0	1	0	0	0	0	0	0
16	15	0	0	0	0	0	0	1	1	1	1	…	0	1	0	0	0	0	0	0
17	16	0	0	0	0	1	1	1	1	1	1	…	0	1	0	0	0	0	0	0
18	17	0	0	0	0	0	0	1	1	1	1	…	0	1	0	0	0	0	0	0
19	18	0	0	0	0	0	0	1	1	1	1	…	0	1	0	0	0	0	0	0
20	19	0	0	0	1	1	1	1	1	1	1	…	0	1	0	0	0	0	0	0
21	20	0	0	0	0	0	0	0	0	1	1		0	1	0	0	0	0	0	0

training

[0, 0, 0, 0, 0, 1, 1, 0, …, 1, 0, 0, 0]



f(x)

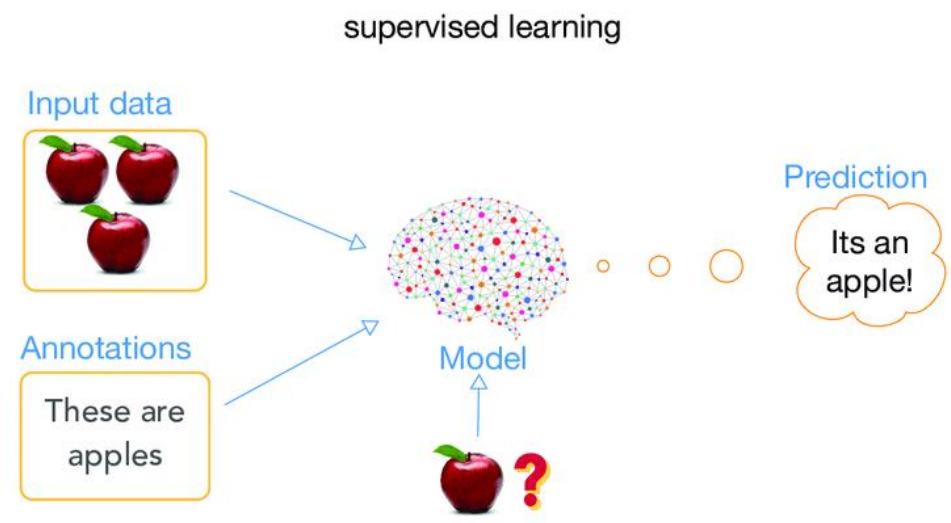
Machine Learning

- Kaplan and Haenlein (2019) defined AI
  - "…a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation"
- Machine Learning

	Supervised	Unsupervised	Reinforcement
Continuous	Prediction eg. Regression regression tree, neural network, deep learning	Clustering eg. hierarchical clustering, k-means, DBSCAN Dimensionality eg. SVD, PCA	Algorithms learns to react to an environment Optimal Policy eg. actor-critic, policy gradients
Categorical	Classification eg. KNN, decision trees, logistic regression, SVM, Naïve Bayes	Association Rule eg. Apriori, FP-Growth Parameter Estimation eg. Hidden Markov Model	Optimal Policy eg. Markov Decision Process, Q-learning

Machine Learning

- Find a "Function f(x)"!
- Supervised Learning
  - Data pair (X,Y)
  - Y is label
    - continuous → prediction
    - categorical → classification
- Unsupervised Learning
  - Data (x) without label Y
  - Clustering, dimension reduction
- Reinforcement Learning
  - Elements: State, Action, Transition Probability, Rewards, …
  - Optimal Policy: mapping from State to Action



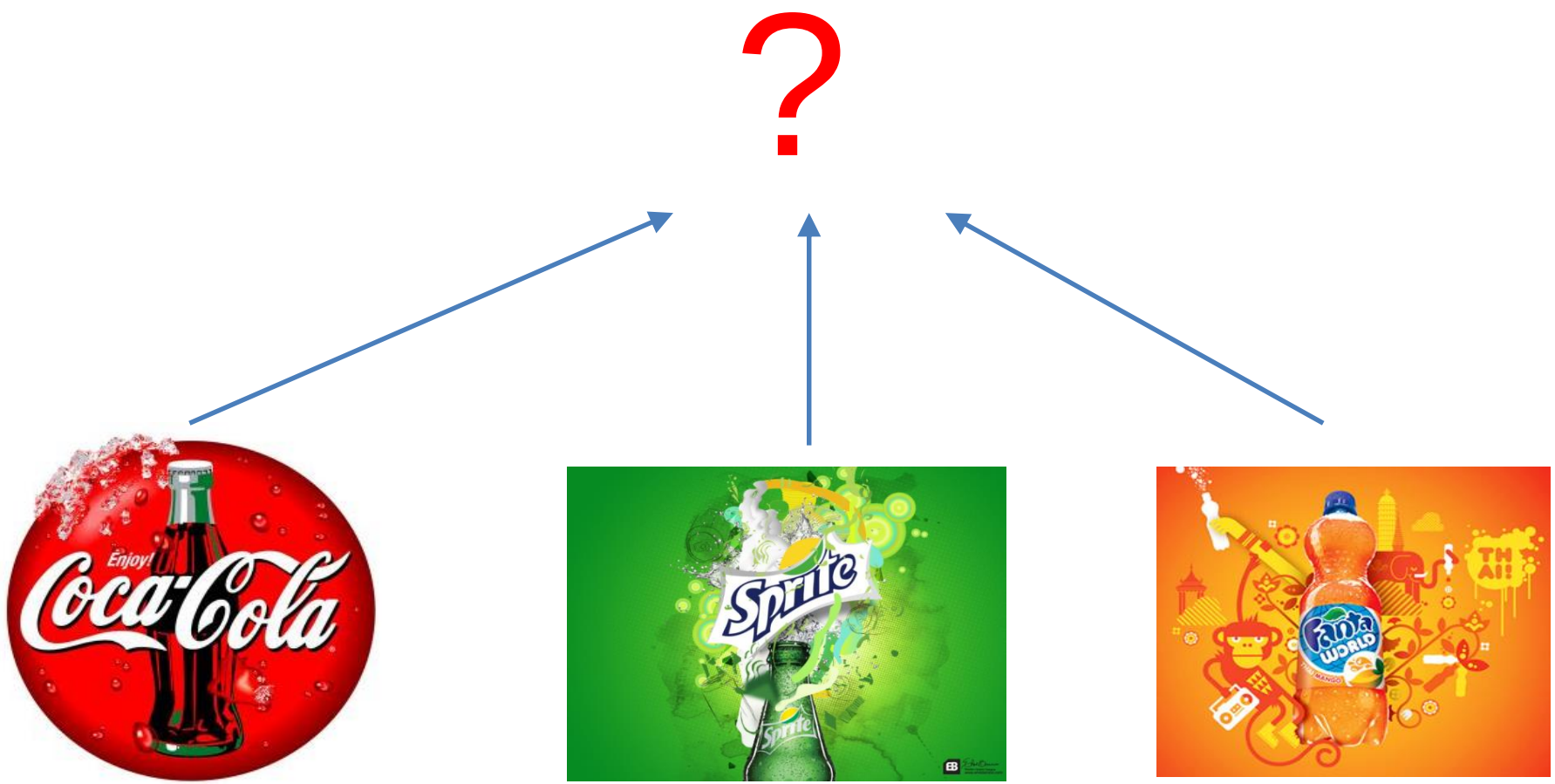
Yan et al., 2018. Background Augmentation Generative Adversarial Networks (BAGANs): Effective Data Generation Based on GAN-Augmented 3D Synthesizing. Symmetry, vol. 10, no. 12, 734.

ML/DS Functionalities

- Generalization (一般化) (B)
- Feature Selection (特徵選取/重要參數篩選) (B)
- Association Rules (關聯法則) (U)
- Classification (分類) (S)
- Prediction (預測) (S)
- Time Series Prediction (時間序列) (B)
- Clustering (集群) (U)
- Outlier (離群值) (U)
- Optimization (最佳化) (B)
- Visualization (視覺化)
- Interpretation (解釋)
- …etc.

S: Supervised Learning
U: Unsupervised Learning
B: Both

ML/DS Functionalities: Generalization



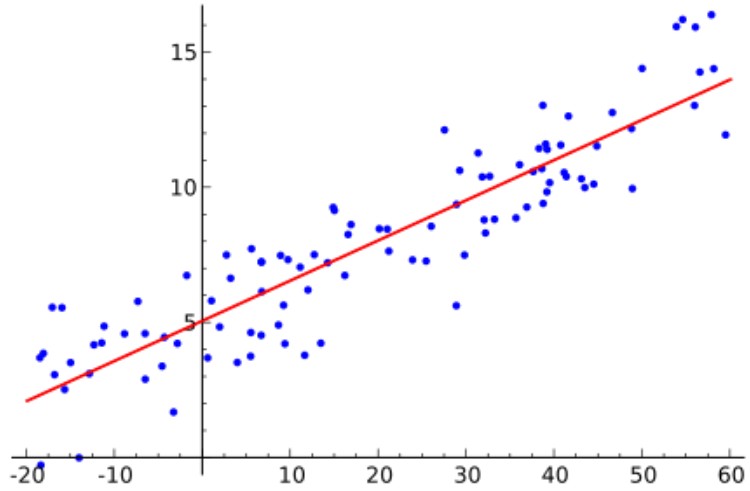
ML/DS Functionalities: Generalization

- Linear Regression
  - Show the average behavior

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

or

$$y = X\beta + \varepsilon$$



歸納法：Specification (特殊化) → Generalization (一般化)

ML/DS Functionalities: Feature Selection

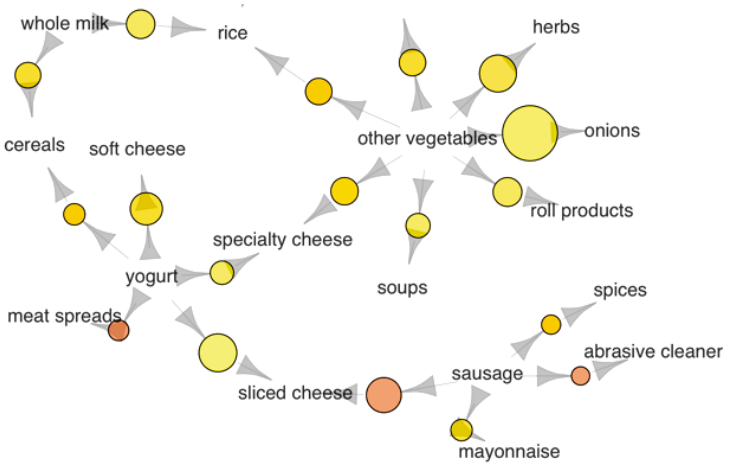
- 為什麼要選重要參數/因子?
  - 易於決策判斷
    - 看"_____"就知道要不要決定出去旅遊
    - 看"_____"就知道這衣服適不適合
    - 量"_____"就知道身體健康狀況

- 製造現場篩選重要因子的目的
  - Diagnostics and Troubleshooting
    - 掌握影響機台品質變異的主要因子、上下游因子的交互作用
    - Engineering Process Control (EPC)
  - 建立管理機制、簡單法則
    - 看Bottleneck就可推估現場WIP level
  - 精度預測?
    - 提升預測準確度
    - On-line real-time prediction
  - 監控Monitoring → 用較低成本/較少管制圖
    - 環境因子監控、機台參數監控

Data Mining Functionalities: Association Rules

Association Rule (關聯規則)

- A rule-based machine learning method for discovering interesting relations between variables in large databases.
- It identifies the relation and frequency among items appearing at the same time to extract the association that the products will be purchased at the same time.
- To derive a set of strong association rules in the form of IF-THEN from in a database. 顧客若買了X之後，很有可能會再買Y。
- Applications: market basket analysis, layout design, cross selling…



https://www.kdnuggets.com/2016/04/association-rules-apriori-algorithm-tutorial.html

Data Mining Functionalities: Association Rules

- Example

Transactional data for an *AllElectronics* branch.

TID	*List of item_IDs*
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

	I1	I2	I3	I4	I5
T100	1	1			1
T200		1		1	
T300		1	1		
T400	1	1		1	
T500	1		1		
T600		1	1		
T700	1		1		
T800	1	1	1		1
T900	1	1	1		

Han et al. (2011)

Data Mining Functionalities: Classification

Classification

- Supervised learning with label
- According to the defined class group (label), conduct ML to establish a estimated function or relationship to identify the pattern that can classify the data to the known category.
- Classifier (分類器): the model can be used immediately to classify/ predict when new observations are added

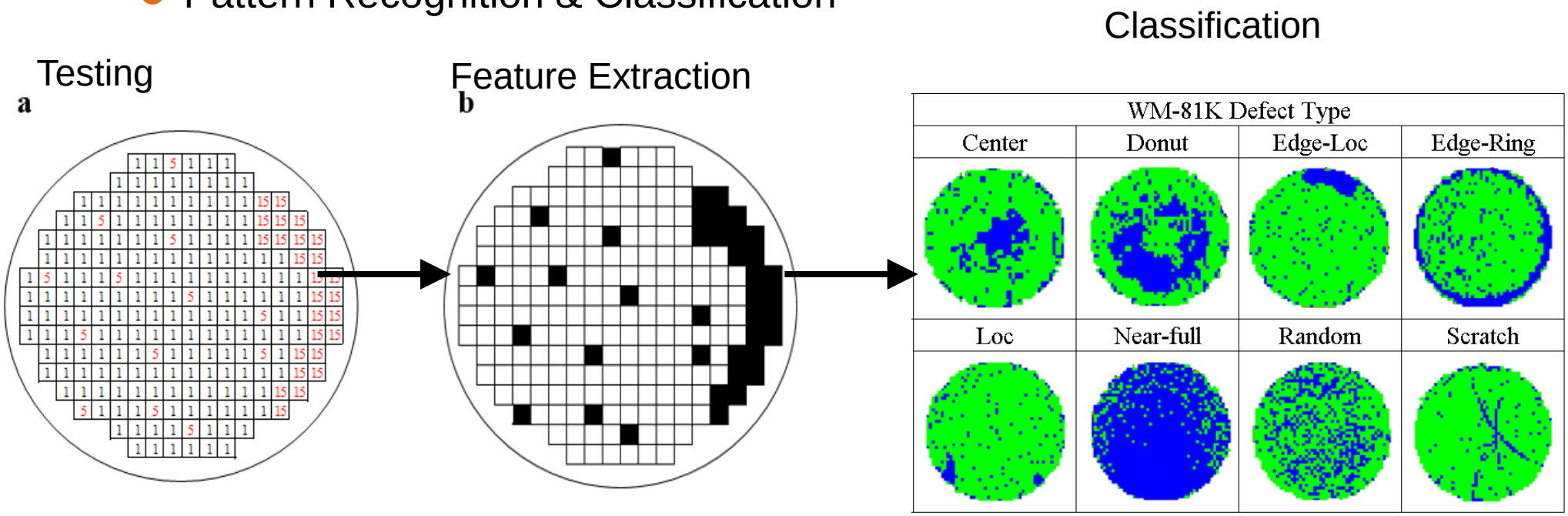
Example (可以label好都可以學)

- Defect bin code/ Automated optical inspection (AOI)
  - Center, ring, edge, scratch, random…
- Market segmentation/ Consumers classification
  - student group, working group, pink collar group, geek group, …
- Risk of credit card applicants
  - high risk, medium risk, and low risk, to reduce the risk of bad debts.

ML/DS Functionalities: Classification

Process Diagnostics

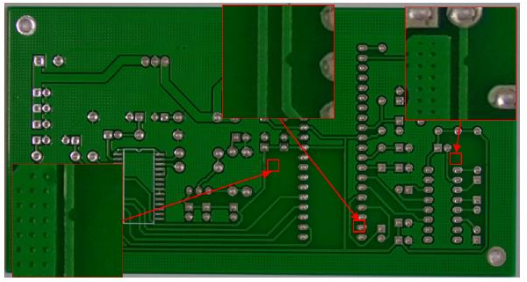
- Wafer Bin Map (WBM) problem
- In the semiconductor fabrication process, the circuit probe (CP) test is used to detect specific failures of each die and thus indicate the test results with the corresponding bin values.
- Pattern Recognition & Classification



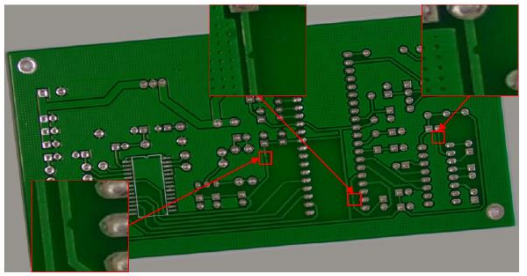
Hsu, S. and Chien, C. (2007), "Hybrid Data Mining Approach for Pattern Extraction from Wafer Bin Map to Improve Yield in Semiconductor Manufacturing," International Journal of Production Economics, 107, 88-103.

ML/DS Functionalities: Classification

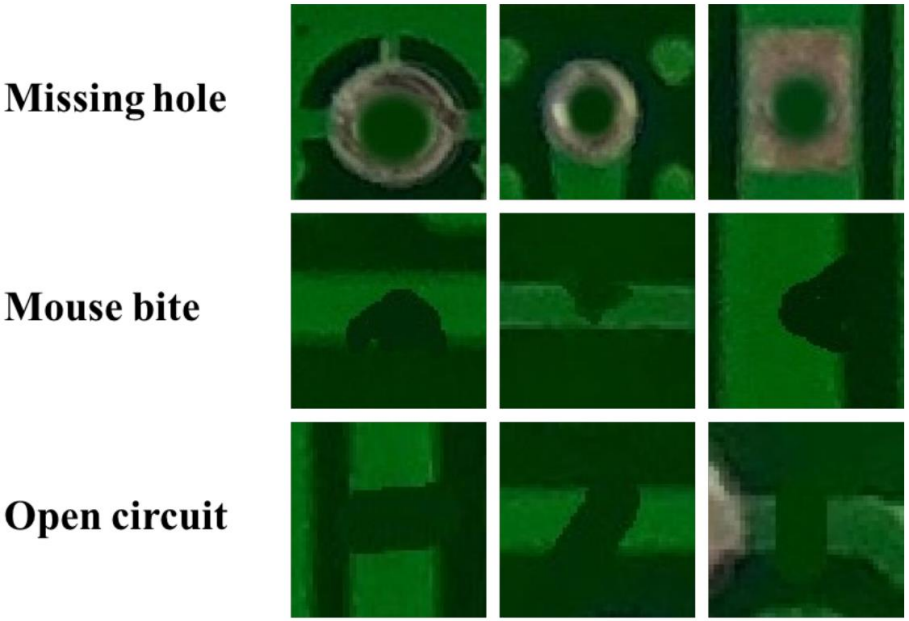
- Deep Learning
  - 有看過才會，沒看過不一定會
  - → 需要大量數據(?)
  - → 需要大量且多元/異質的代表性數據
    - Rotation, flip, brightness, contrast, …
    - Elevation angle, depression angle, …

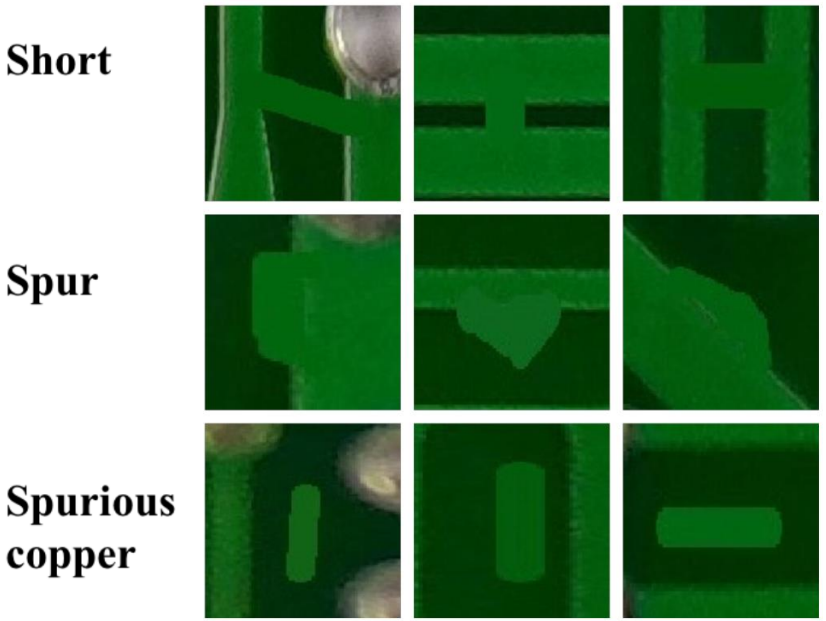


(a) PCB with mouse bite.



(b) PCB with mouse bite and rotation.



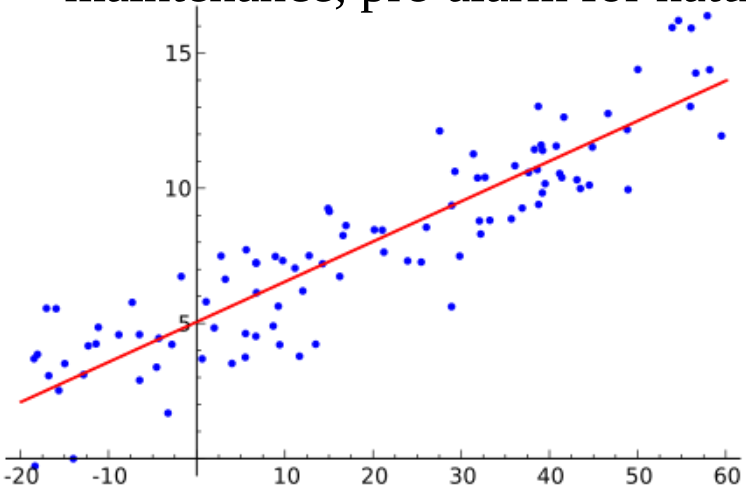


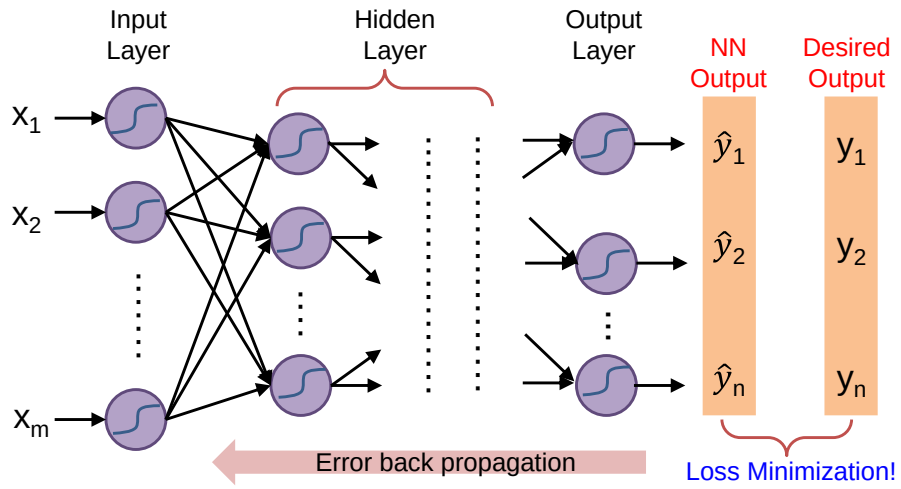
Huang, W., and Wei, P. (2018). A PCB Dataset for Defects Detection and Classification. https://arxiv.org/pdf/1901.08204v1.pdf

ML/DS Functionalities: Prediction

Prediction (預測)

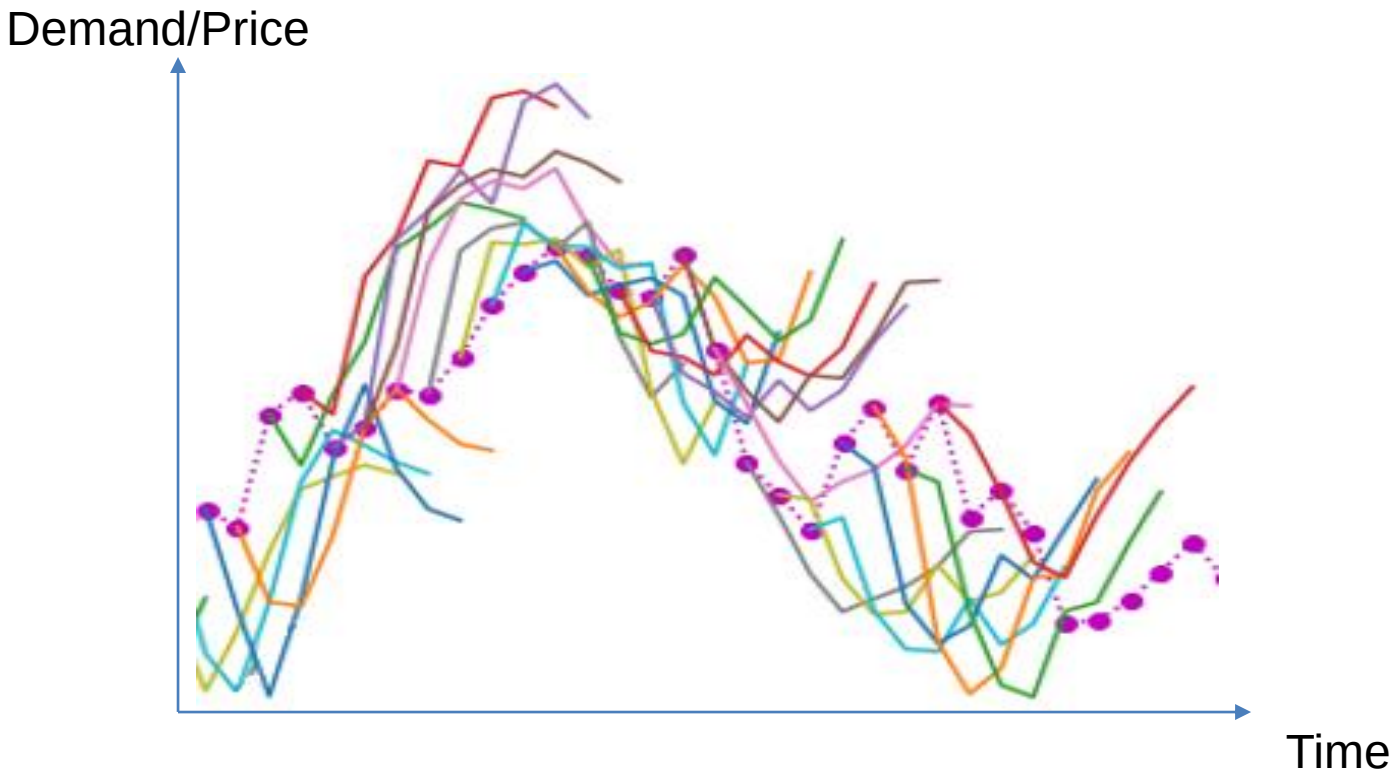
- Estimate the future value of the attribute based on the past observations with one or several factors/variables/attributes. It builds the relationship between dependent variable (i.e. response) and independent variable (i.e. regressor) in the data.
- Statistical regression techniques: given a functional form and estimate the parameters for data generalization. $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$
- Neural network: an interconnected group of natural or artificial neurons (nodes) that uses a mathematical or computational model for information processing based on a connectionistic approach to computation.
- Applications: stock market, demand forecast, price prediction, predictive maintenance, pre-alarm for natural disaster…





ML/DS Functionalities: Time Series Prediction

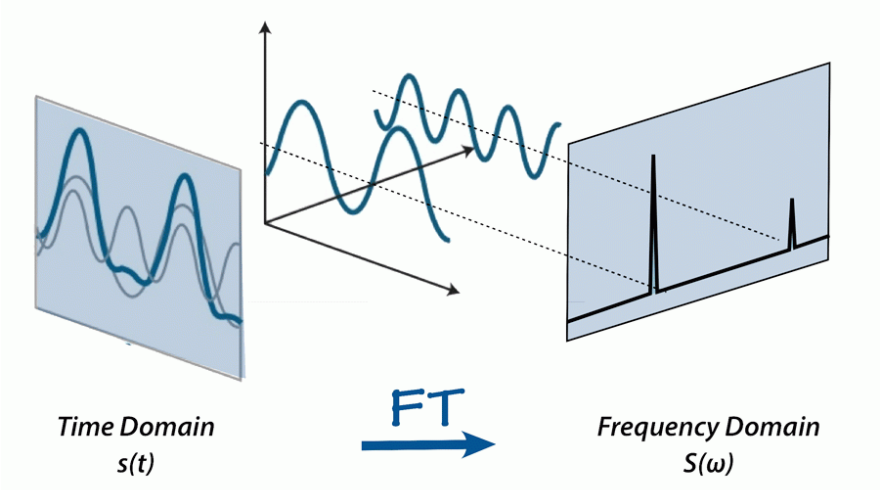
- Brent crude oil prices
  - Time-series data of price



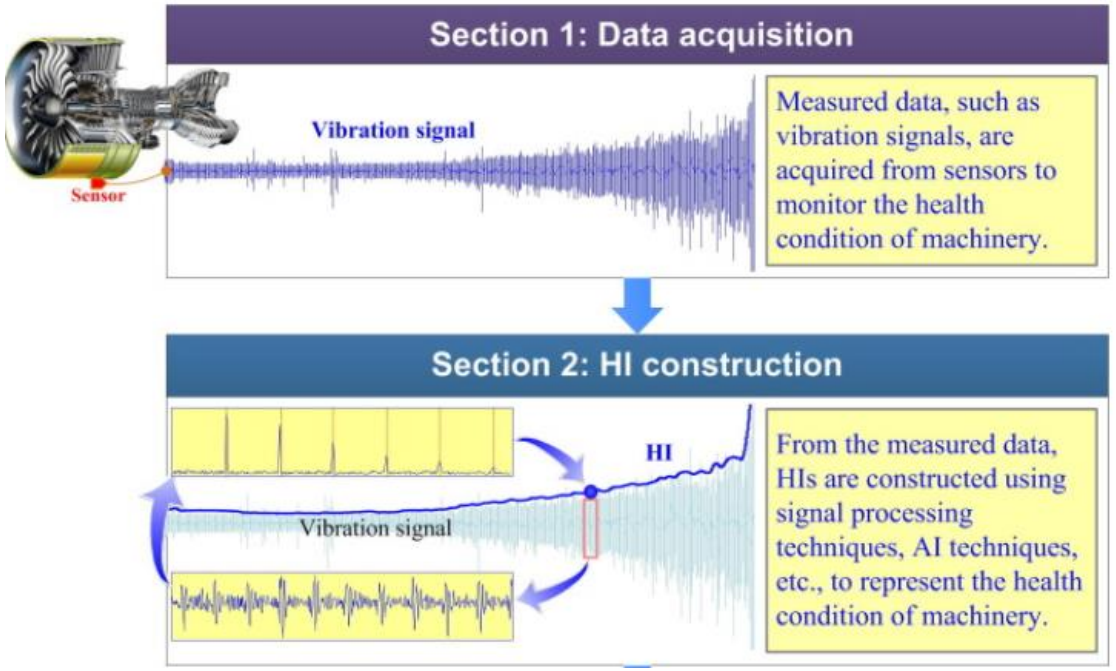
European Central Bank (2015). Forecasting the price of oil. Economic Bulletin, 4, 87-98.

ML/DS Functionalities: Time Series Prediction

- Signal Analysis (special case of time series)
  - Frequency-domain Characteristics (Fourier Transform)
  - Prognostic and Health Management (PHM)



https://en.wikipedia.org/wiki/Fourier_transform

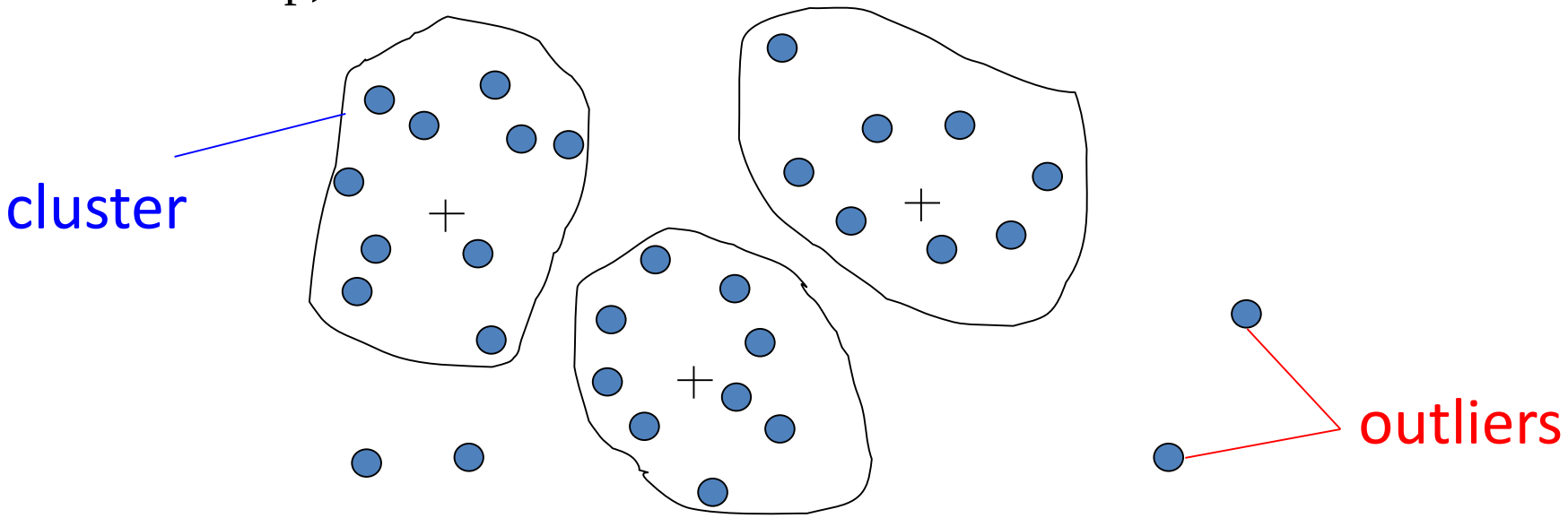


Remaining useful life (RUL) prediction

Lei, et al. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical Systems and Signal Processing, 104, 799-834.

ML/DS Functionalities: Clustering

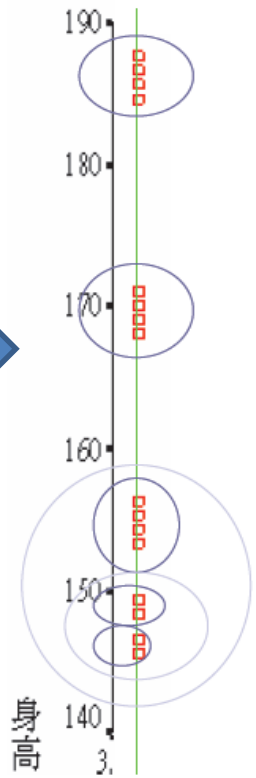
- Clustering (集群/ 群集/ 聚集/ 分群)
  - grouping a set of objects so that objects in the same group (called a cluster) are more similar to each other than to those in other groups.
  - Unknown label. The clustering depends on the attributes of objects.
  - Principle: maximize intra-class similarity & minimize interclass similarity
  - Measure: correlation coefficient, distance, density, …
  - Applications: market segmentation, customer relationship management, wafer bin map, outlier detection.



ML/DS Functionalities: Clustering

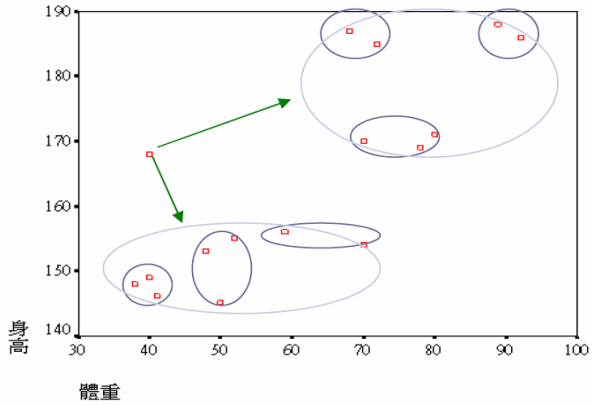
1-D Clustering

身高
145
146
148
149
155
156
154
153
168
169
170
171
185
187
188
186



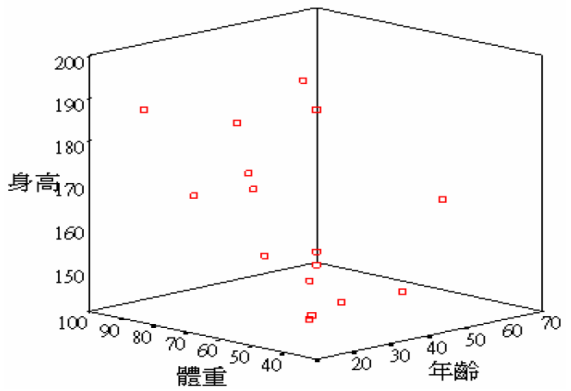
2-D Clustering

身高	體重
145	50
146	41
148	38
149	40
155	52
156	59
154	70
153	48
168	40
169	78
170	70
171	80
185	72
187	68
188	89
186	92



3-D Clustering

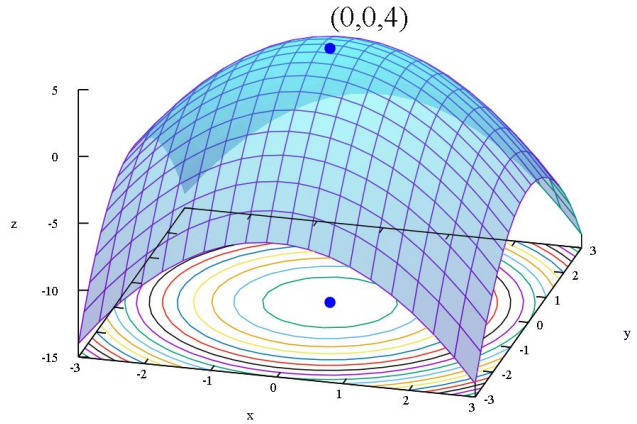
身高	體重	年齡
145	50	50
146	41	18
148	38	16
149	40	25
155	52	29
156	59	35
154	70	31
153	48	24
168	40	52
169	78	19
170	70	28
171	80	35
185	72	46
187	68	22
188	89	15
186	92	60



ML/DS Functionalities: Optimization

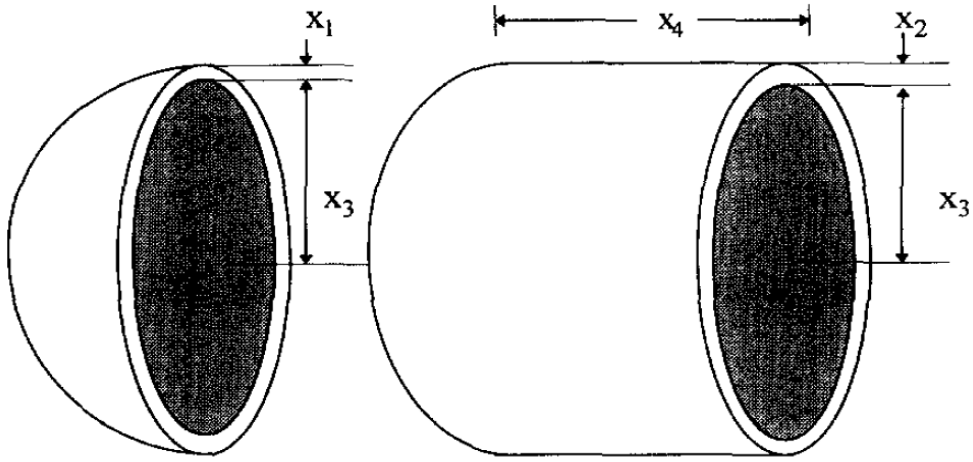
Optimization (最佳化/優化)

- Algorithm is developed for controlling the parameters to maximize/minimize some KPI.
- Mathematical programming is to optimize the objective function (eg. cost) with respect to some decision variables (eg. number of product generated) in the presence of constraints (eg. resource).
- Design of experiments (DOE) is a statistical technique for quickly optimizing performance of systems with known input variables. It screens experimental design which involves all of the known factors that are suspected to affect the system's performance.
- Applications: parameter optimization, production scheduling, capacity planning and allocation, portfolio, transportation routing,…



ML/DS Functionalities: Optimization

- Mechanical Design Problem
  - Tube and Pressure Vessel
  - Objective function
    - Minimize the total volume



Sandgren (1990)

$$\text{Minimize} \quad 0.6224x_1x_3x_4 + 1.7781x_2x_3^2 + 3.1661x_1^2x_4 + 19.84x_1^2x_3$$

$$\text{subject to} \quad -x_1 + 0.0193x_3 \le 0,$$

$$-x_2 + 0.00954x_3 \le 0,$$

$$-\pi x_3^2 x_4 - \tfrac{4}{3}\pi x_3^3 + 1296000 \le 0,$$

$$-240 + x_4 \le 0,$$

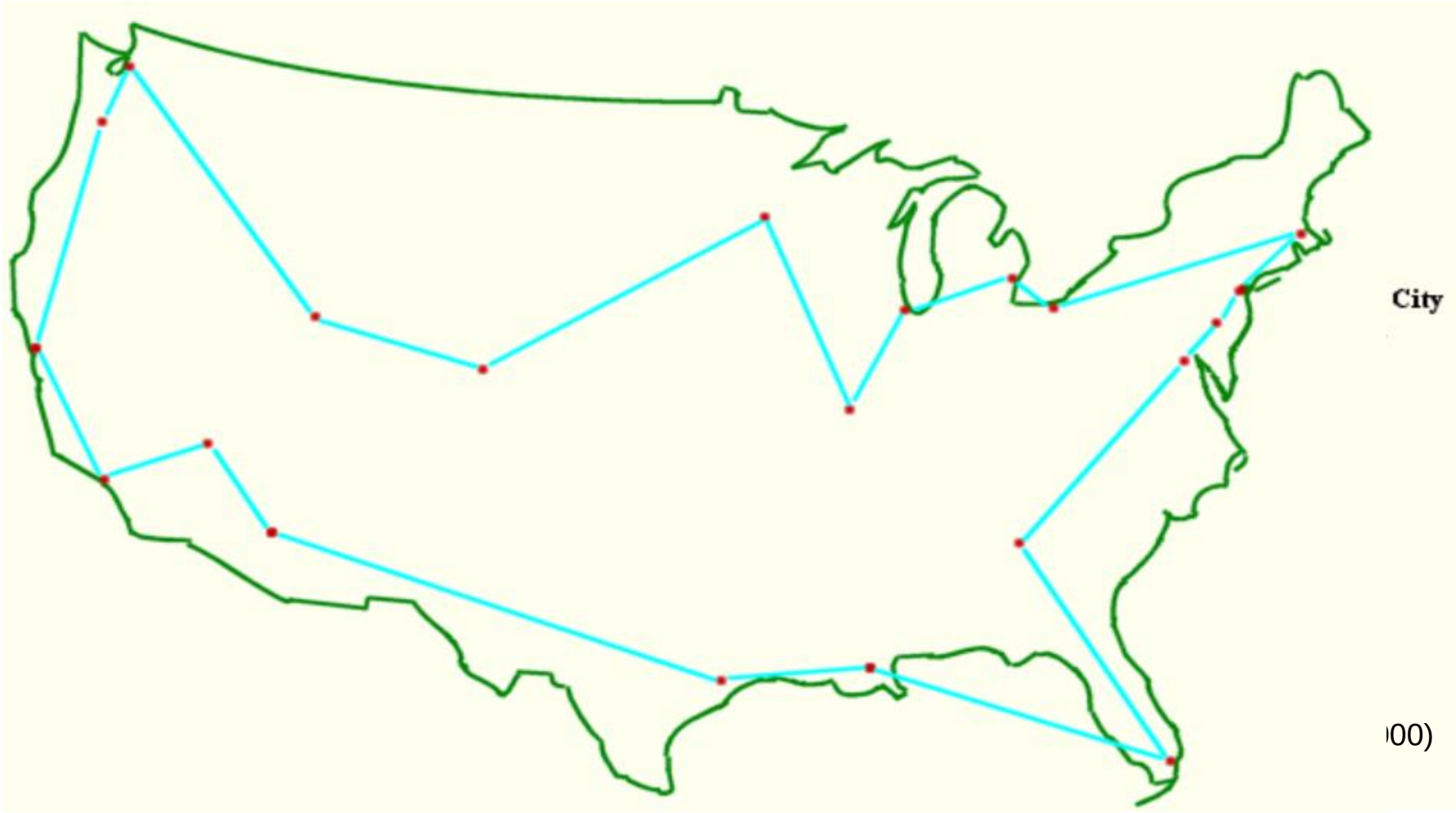
$$1 \le x_1 \le 1.375,$$

$$0.625 \le x_2 \le 1, 48 \le x_3 \le 52, \quad 90 \le x_4 \le 112,$$

Sandgren, E., 1990. Nonlinear integer and discrete programming in mechanical design optimization. Journal of Mechanical Design 112(2) 223–229.

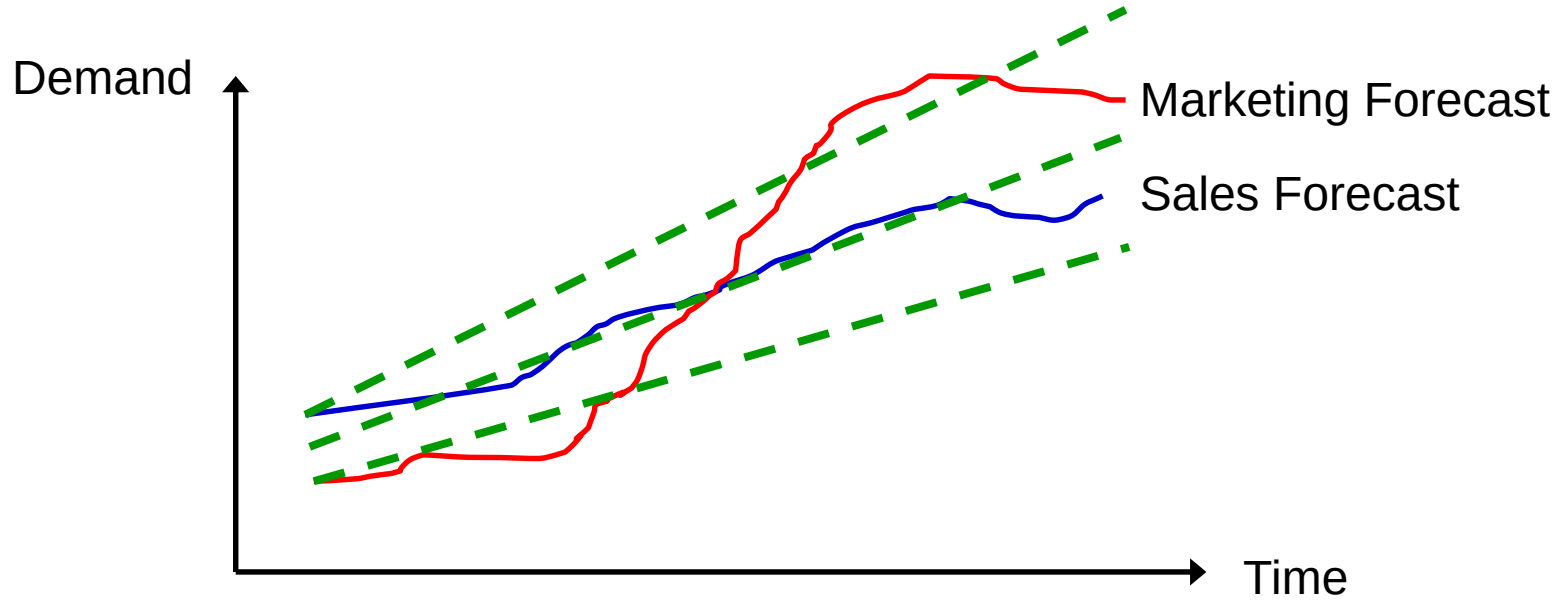
ML/DS Functionalities: Optimization

- Travel Salesman Problem (TSP)
  - Given a set of cities and pairwise distances, to find the shortest possible route that visits each city exactly once and returns to the origin city



ML/DS Functionalities: Optimization

Capacity Planning Problem (Newsboy problem 報童問題)



- Demand forecast: Marketing vs. Sales
  - Should capacity level follow marketing or sales, or …?
- Capacity Shortage (loss of sales) vs. Capacity Surplus (inventory holding cost)

ML/DS Functionalities: Optimization

Markowitz's Mean-Variance Model (1952, 1959)

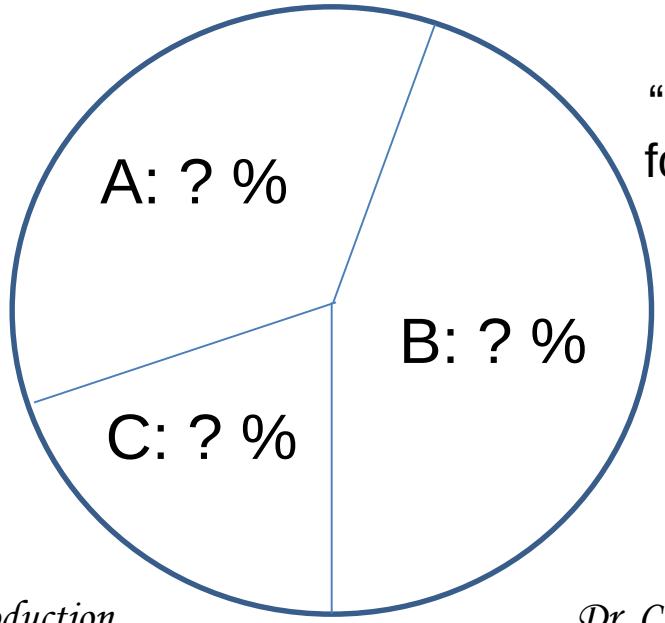
- Given the required return parameters R,

$$\min_{w} \sigma^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij}$$

$$\text{s.t. } \sum_{i=1}^{N} w_i E(r_i) \geq R$$

$$\sum_{i=1}^{N} w_i = 1$$

- Solve the quadratic programming get the minimum variance frontier by tuning different parameters R sequentially.

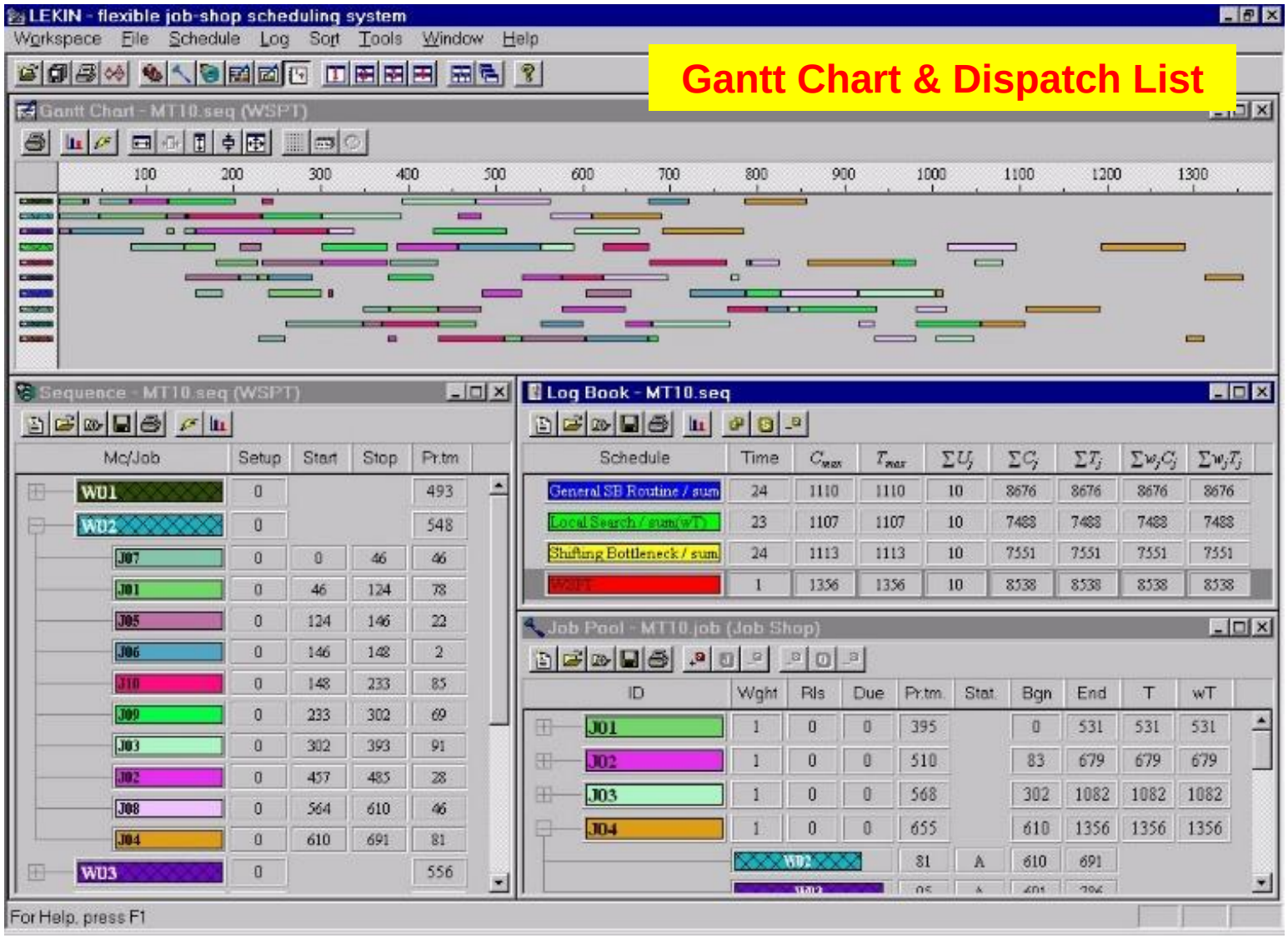


"Pie Chart" for portfolio

ML/DS Functionalities: Optimization

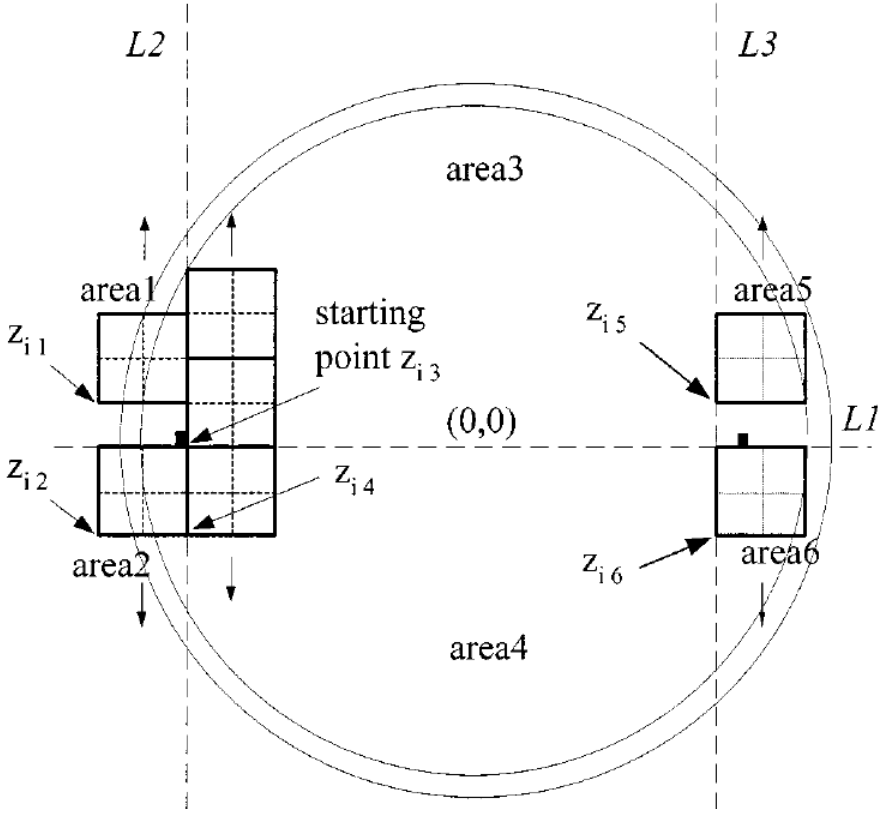
Sequencing & Scheduling

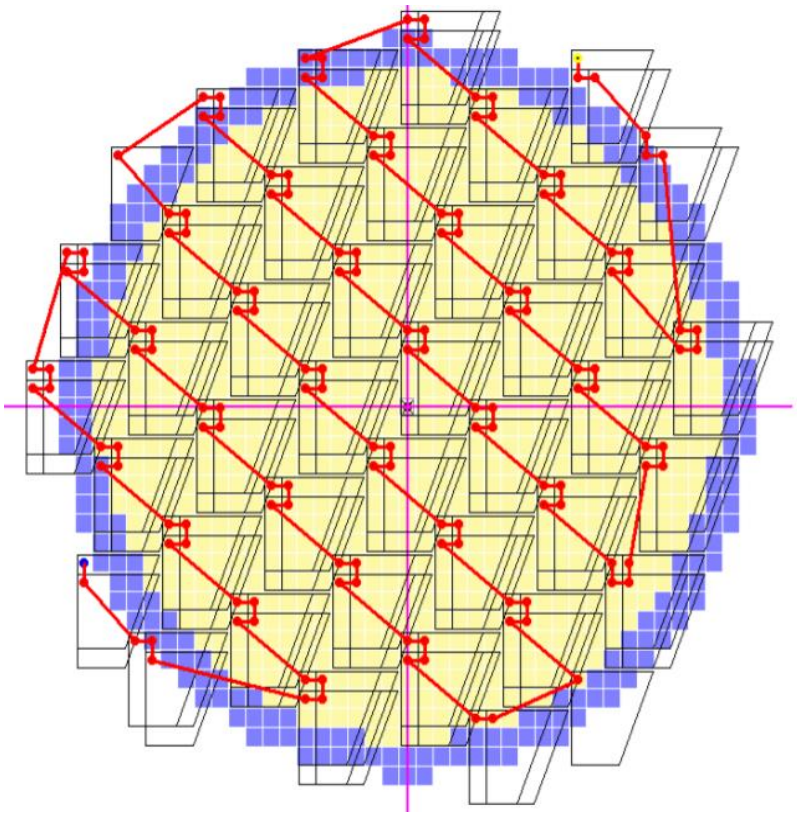
- LEKIN (http://web-static.stern.nyu.edu/om/software/lekin/)



ML/DS Functionalities: Optimization

- Die cutting and scanning route optimization
  - Cutting algorithm to maximize the gross die yields
  - Route algorithm to optimize the probe card stepping off wafer





Chien, C.-F.., Hsu, S., & Deng, J. (2001), "A cutting algorithm for optimizing the wafer exposure pattern," IEEE Transactions on Semiconductor Manufacturing, 14(2), 157-162.

Koski, R., & Fresquez, D. 2015. Probe route optimization and its effects on the efficiency of test. Semiconductor Wafer Test Workshop, San Diego, California, June 7-10, 2015.

ML/DS Functionalities: Optimization

Parameter Optimization

- DOE/Taguchi Method

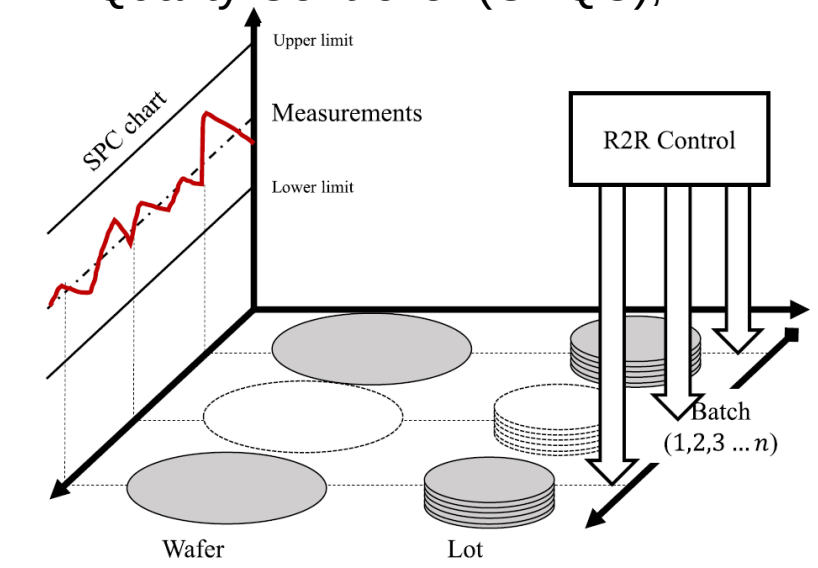
Run	A	B	C	Labels	A	B	C
1	−	−	−	(1)	0	0	0
2	+	−	−	*a*	1	0	0
3	−	+	−	*b*	0	1	0
4	+	+	−	*ab*	1	1	0
5	−	−	+	*c*	0	0	1
6	+	−	+	*ac*	1	0	1
7	−	+	+	*bc*	0	1	1
8	+	+	+	*abc*	1	1	1

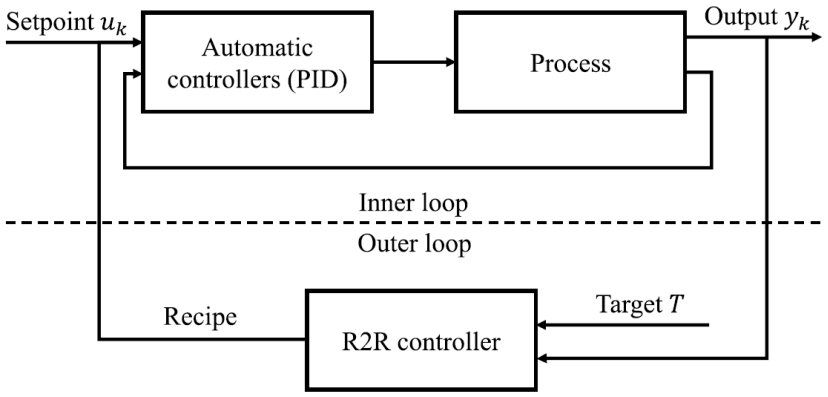
Table 6-6 Analysis of Variance for the Fill Height Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	*P*-Value
Percent carbonation (*A*)	36.00	1	36.00	57.60	<0.0001
Pressure (*B*)	20.25	1	20.25	32.40	0.0005
Line speed (*C*)	12.25	1	12.25	19.60	0.0022
AB	2.25	1	2.25	3.60	0.0943
AC	0.25	1	0.25	0.40	0.5447
BC	1.00	1	1.00	1.60	0.2415
ABC	1.00	1	1.00	1.60	0.2415
Error	5.00	8	0.625		
Total	78.00	15			

R2R Control

- PID, EWMA, Optimizing Adaptive Quality Controller (OAQC), ANN





Montgomery, D. C. 2012. Design and Analysis of Experiments 8ed.: Wiley.

Liu, K., Chen, Y., Zhang, T., Tian, S., & Zhang, X. 2018. A survey of run-to-run control for batch processes. ISA Transactions, 83, 107-125.

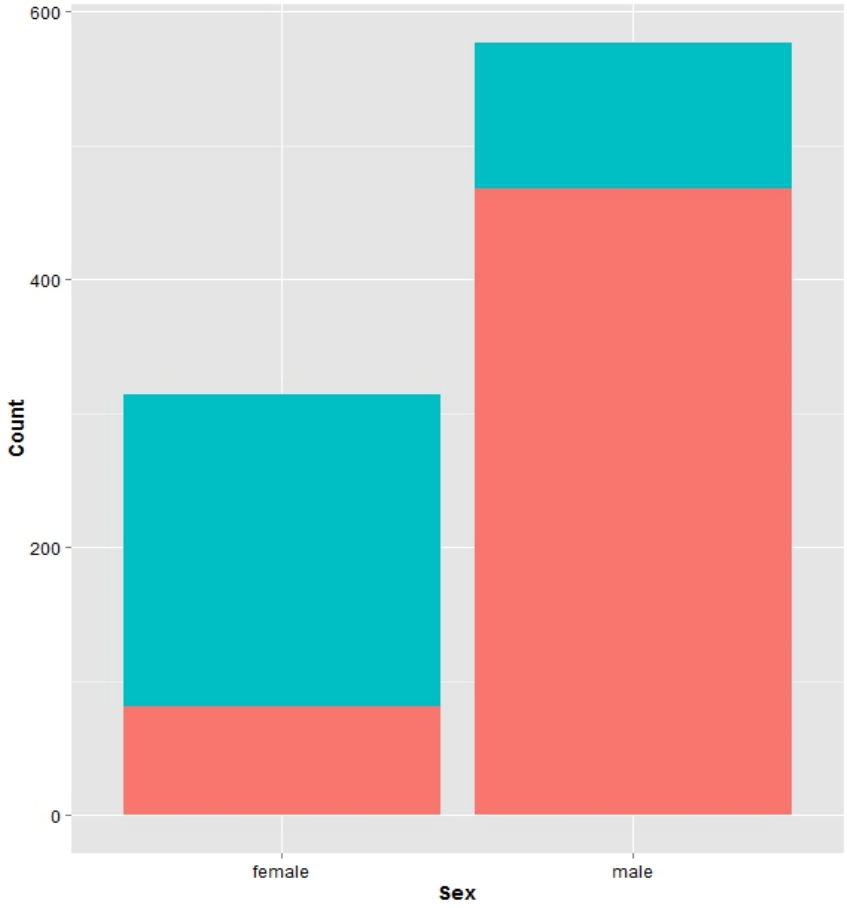
Visualization and Interpretation

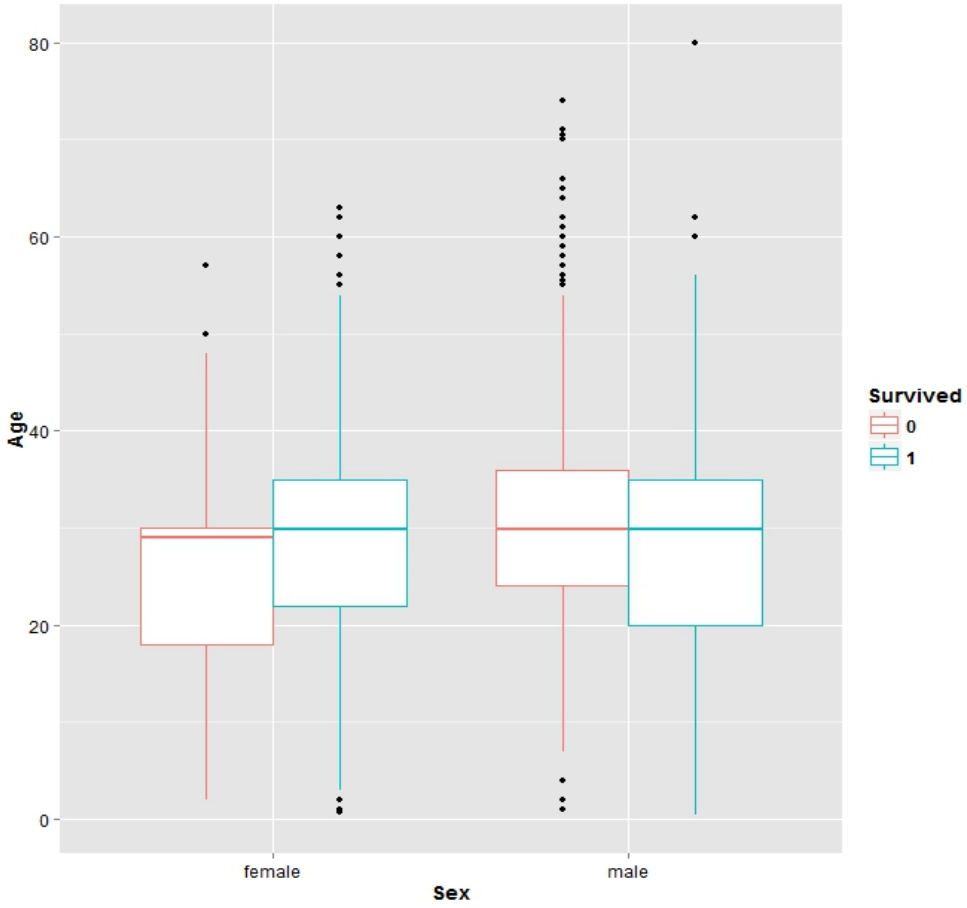
資料收集後的第一項工作(極重要!!)

- Data Descriptive Characteristics (敘述統計!!)
  - To better understand the data: central tendency, variation and spread
- Data dispersion characteristics
  - median, max, min, quantiles, outliers, variance, etc.
- 資料探索與視覺化
  - 類型：
    - 呈現資料次數分布：直方圖、條形圖與圓餅圖
    - 表現資料分布離勢情形：莖葉圖、盒鬚圖與常態機率圖 (box plot)
    - 顯示時間序列變化：趨勢圖 (trend chart)
    - 顯現兩變數相對變化關係：散佈圖 (scatter plot)
  - 資料型態的不同，所適用的統計圖亦有所不同
    - 離散資料：適用長條圖
    - 連續資料：適用直方圖與圓餅圖

ML/DS Functionalities: Visualization

- Passengers – Sex & Age





https://www.kaggle.com/c/titanic

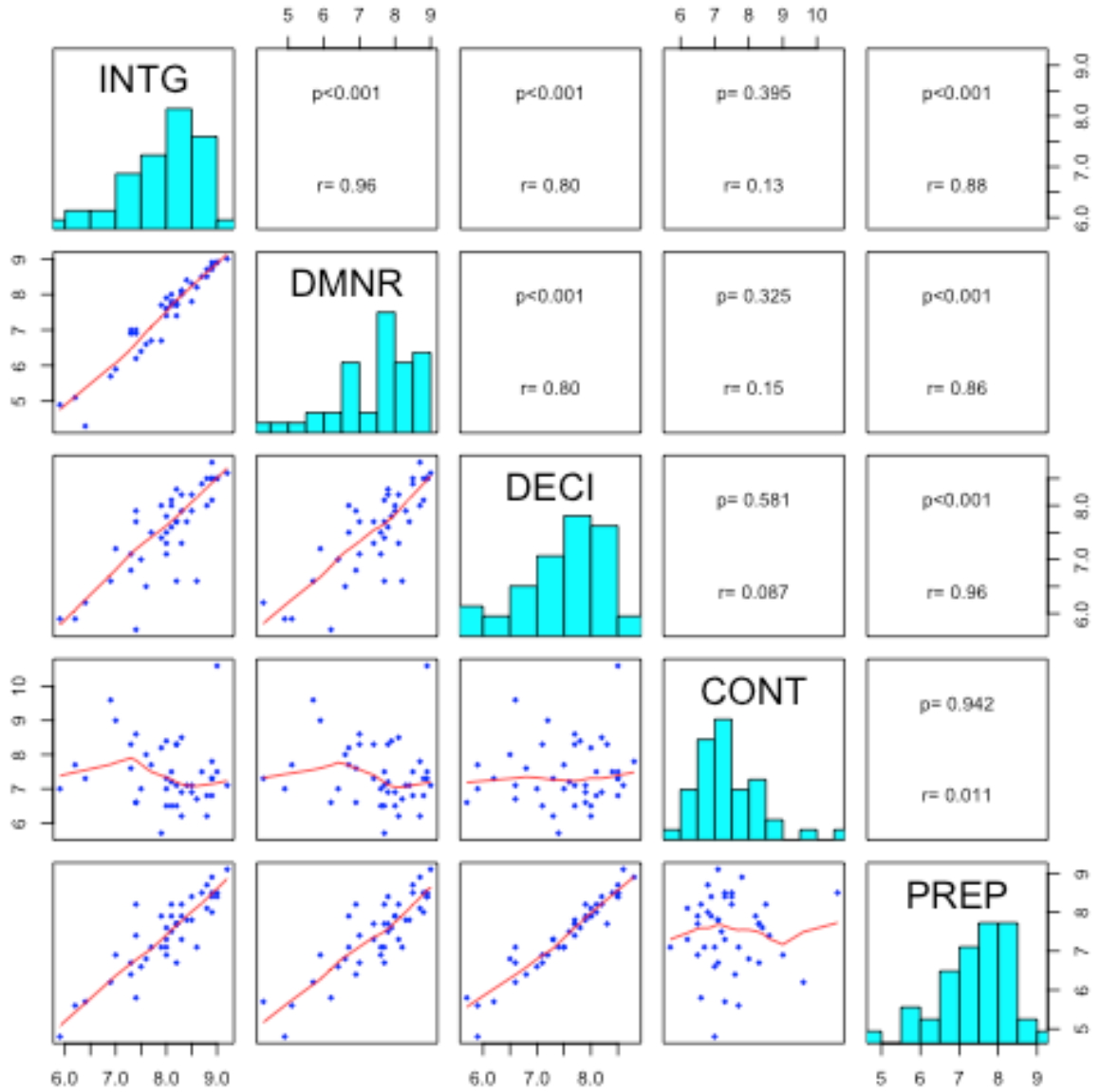
Descriptive Statistics

```
summary(data[,1:8])
```

```
##   Class             Time                          Feature1       Feature2
##  pass:1463   Min.   :2008-07-19 11:55:00   Min.   :2743   Min.   :2159
##  fail: 104   1st Qu.:2008-08-22 00:55:30   1st Qu.:2966   1st Qu.:2452
##              Median :2008-09-11 08:06:00   Median :3011   Median :2499
##              Mean   :2008-09-09 18:37:39   Mean   :3014   Mean   :2496
##              3rd Qu.:2008-09-29 11:33:00   3rd Qu.:3057   3rd Qu.:2539
##              Max.   :2008-10-17 06:07:00   Max.   :3356   Max.   :2846
##                                            NA's   :6      NA's   :7
##     Feature3       Feature4       Feature5            Feature6
##  Min.   :2061   Min.   :   0   Min.   :   0.6815   Min.   :100
##  1st Qu.:2181   1st Qu.:1082   1st Qu.:   1.0177   1st Qu.:100
##  Median :2201   Median :1285   Median :   1.3168   Median :100
##  Mean   :2201   Mean   :1396   Mean   :   4.1970   Mean   :100
##  3rd Qu.:2218   3rd Qu.:1591   3rd Qu.:   1.5257   3rd Qu.:100
##  Max.   :2315   Max.   :3715   Max.   :1114.5366   Max.   :100
##  NA's   :14     NA's   :14     NA's   :14          NA's   :14
```

Jeff Hung (2017)，半導體製造資料分析， http://rpubs.com/jeff_datascience/Semiconductor_Manufacturing

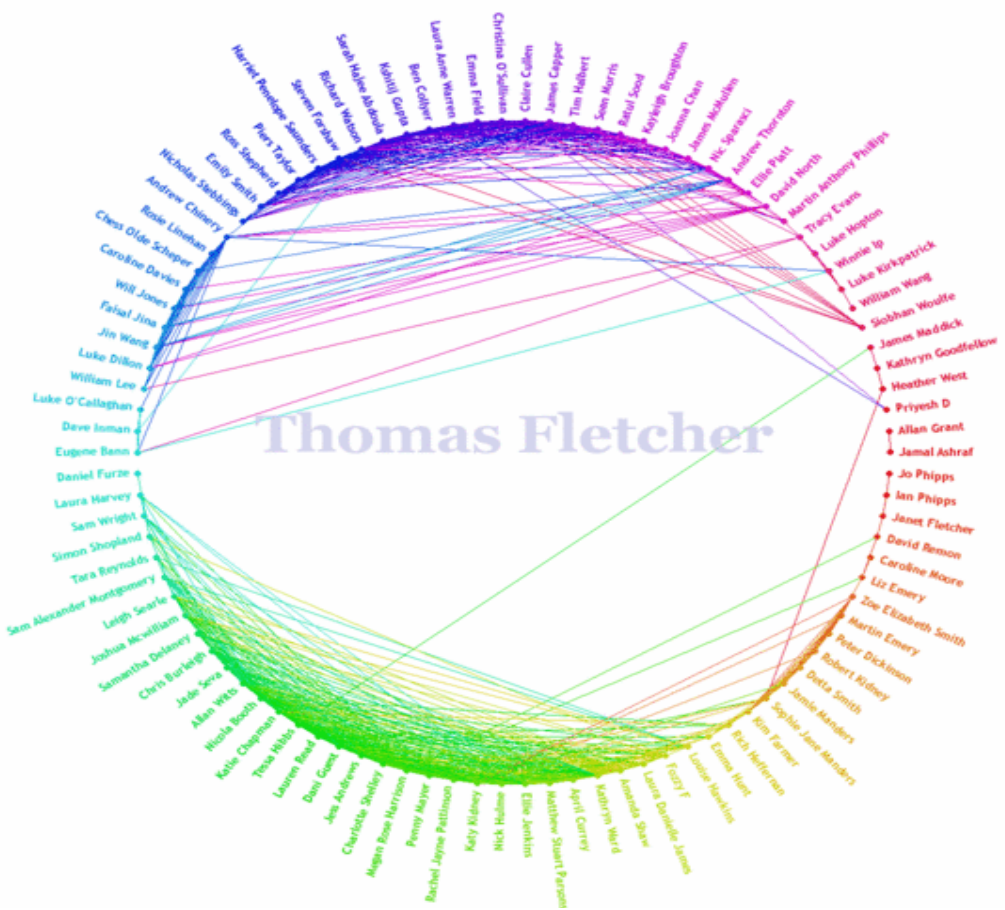
Descriptive Statistics



https://stackoverflow.com/questions/15271103/how-to-modify-this-correlation-matrix-plot

Visualization

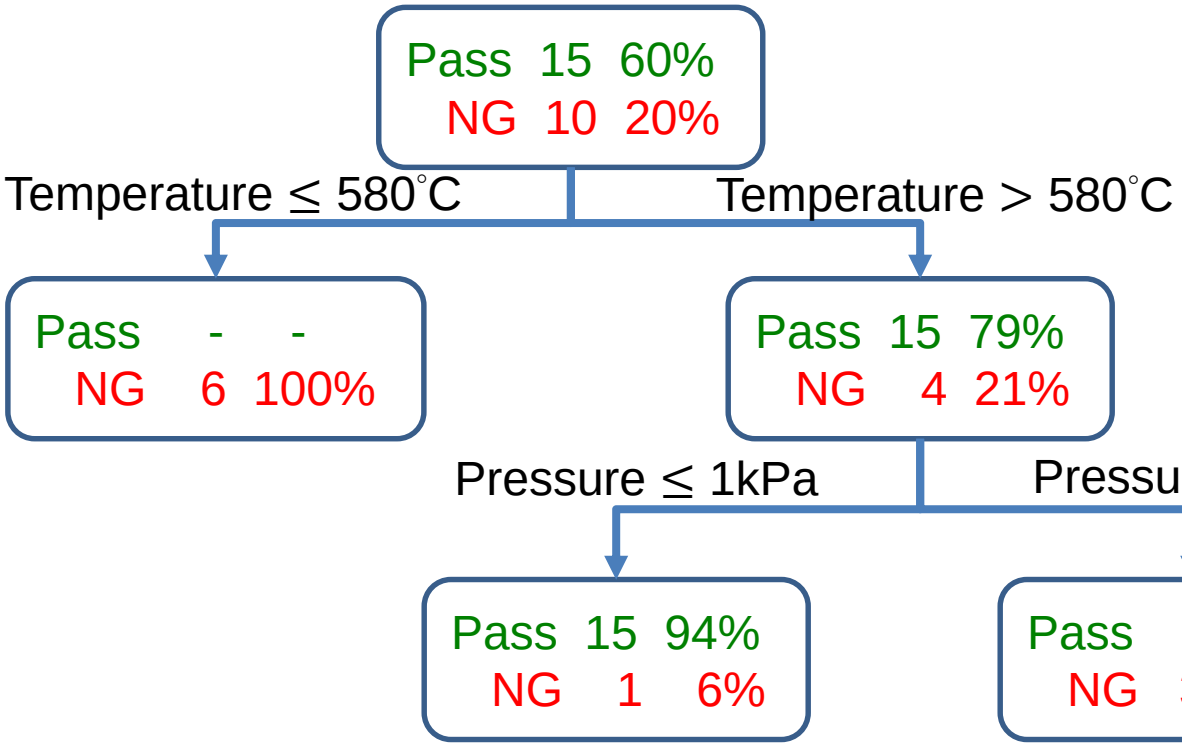
- Visual representation of data and knowledge may facilitate interactive knowledge mining in databases.
- Friend Wheel: Facebook social networks



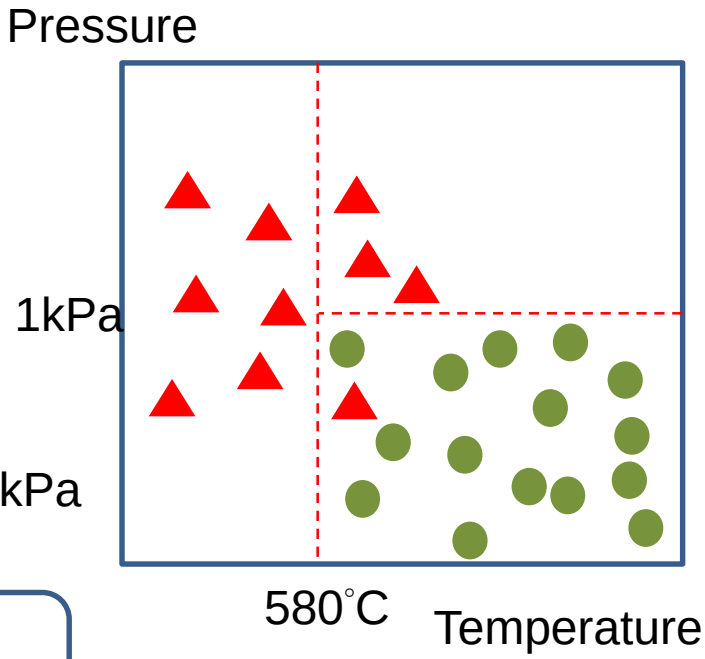
Visualization

Decision Tree

Tree Structure



Rectangle Partitions

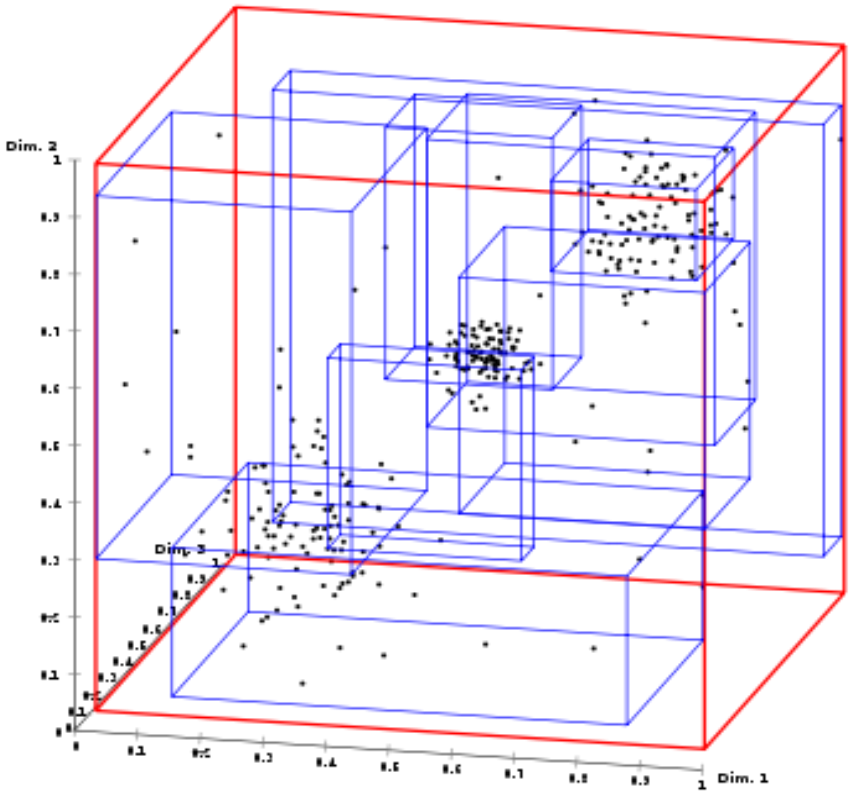


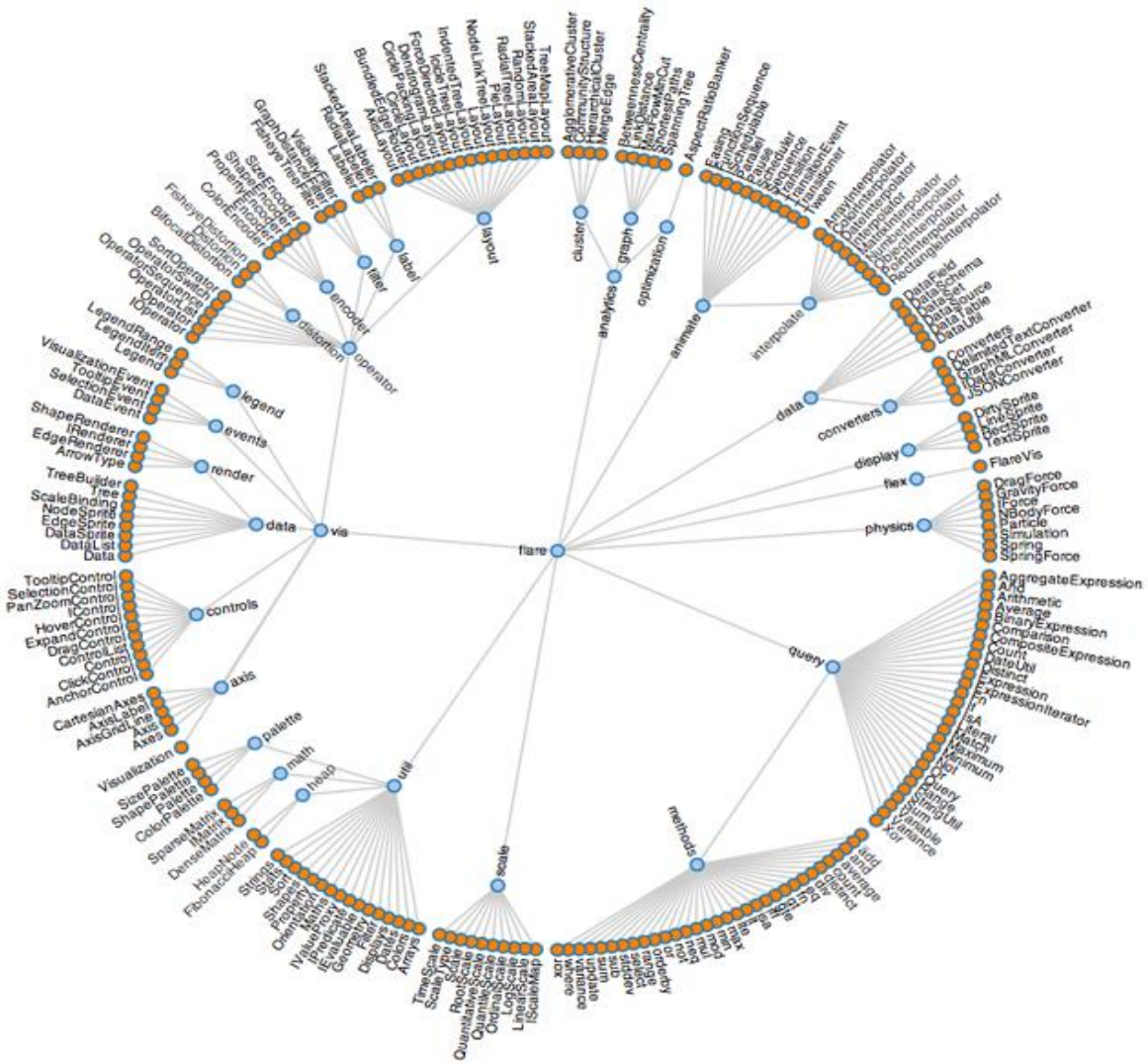
Rule Extraction

- If Temperature ≤ 580˚C, then NG. (Accuracy: 100%)
- If Temperature > 580˚C and Pressure > 1kPa, then NG. (Accuracy: 100%)
- If Temperature > 580˚C and Pressure ≤ 1kPa, then Pass. (Accuracy: 94%)

Lee, C.-Y., Hung, Y.-H., and Chen, Y.-W. 2020. Hybrid data science and reinforcement learning in data envelopment analysis. Working paper.

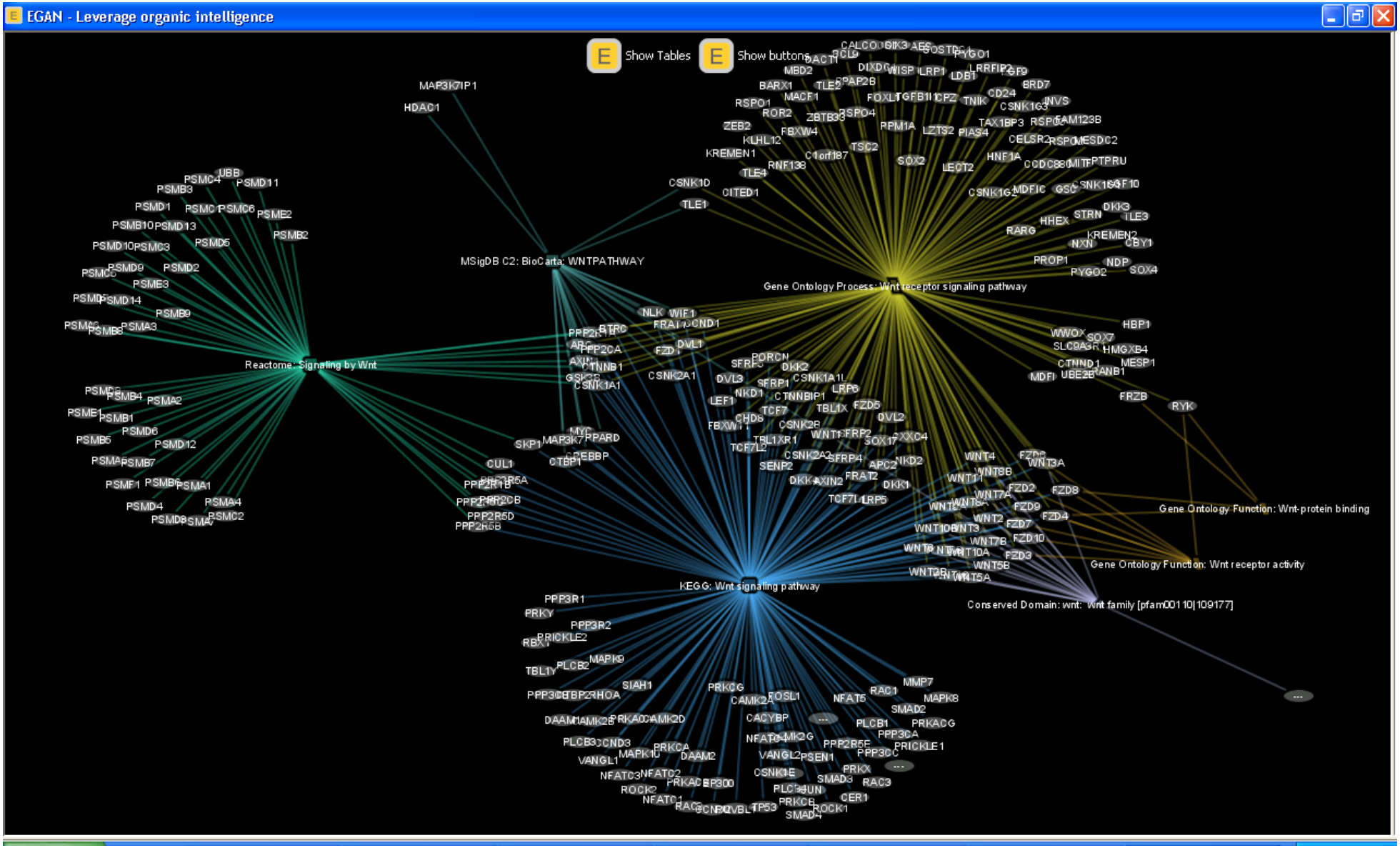
Visualization





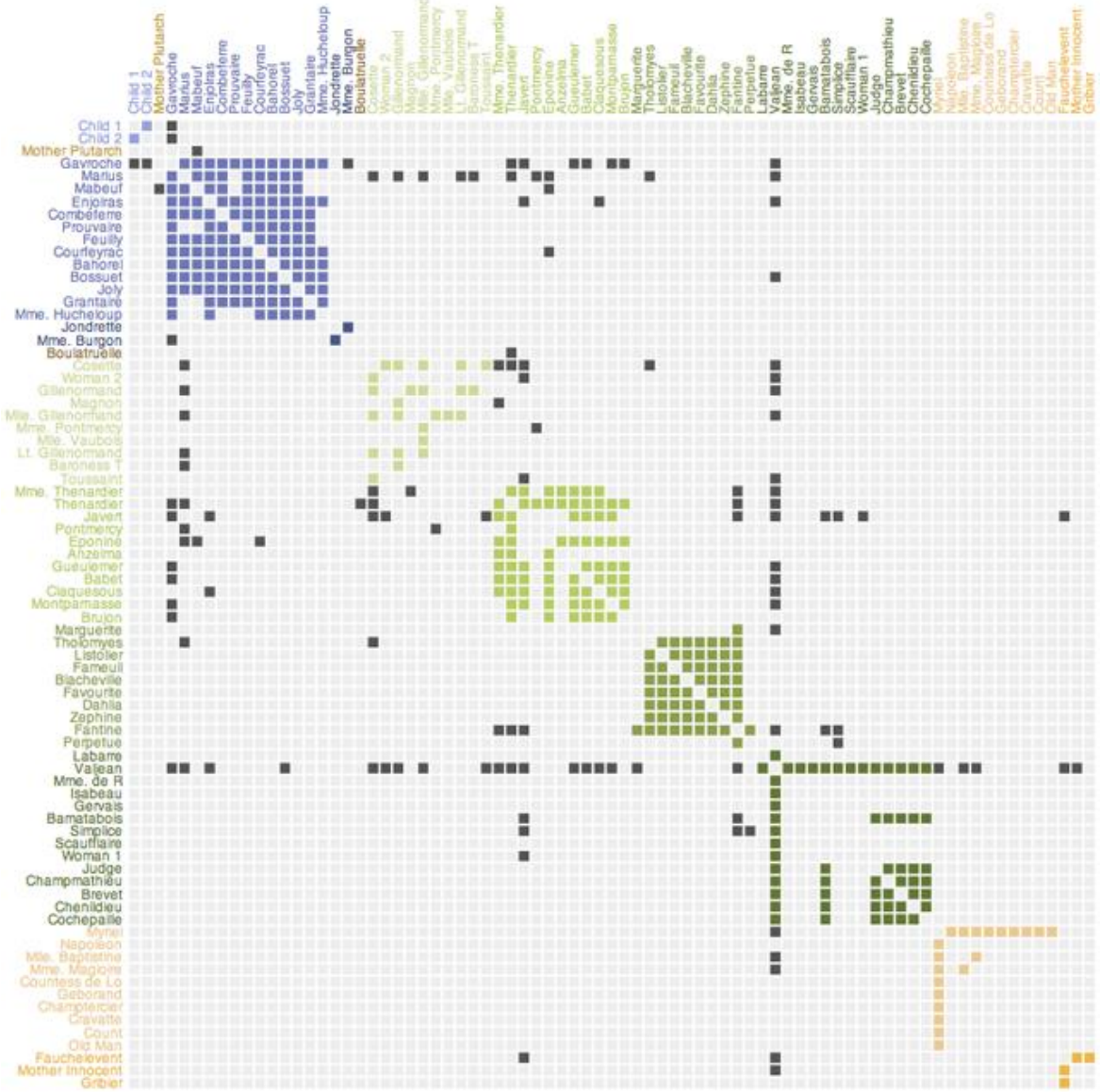
A Tour Through the Visualization Zoo (2010). http://homes.cs.washington.edu/~jheer/files/zoo/

Visualization (Gene Association)



Exploratory Gene Association Networks: http://akt.ucsf.edu/EGAN/screenshots.php

Visualization (Gene Association)

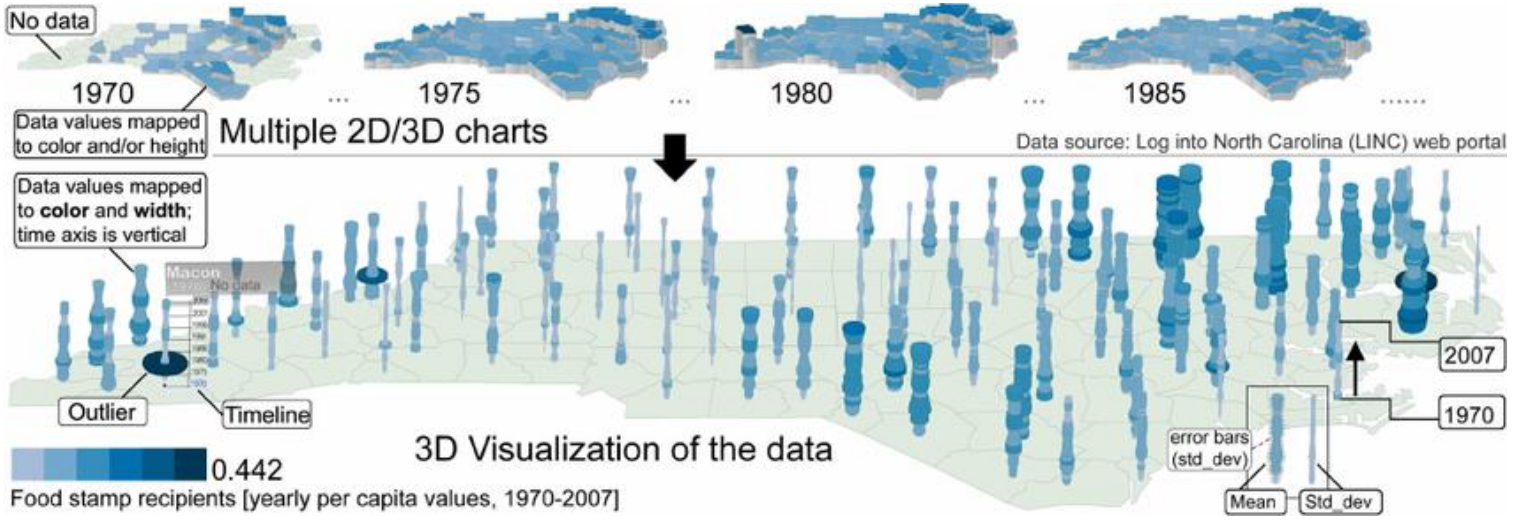


A Tour Through the Visualization Zoo (2010). http://homes.cs.washington.edu/~jheer/files/zoo/

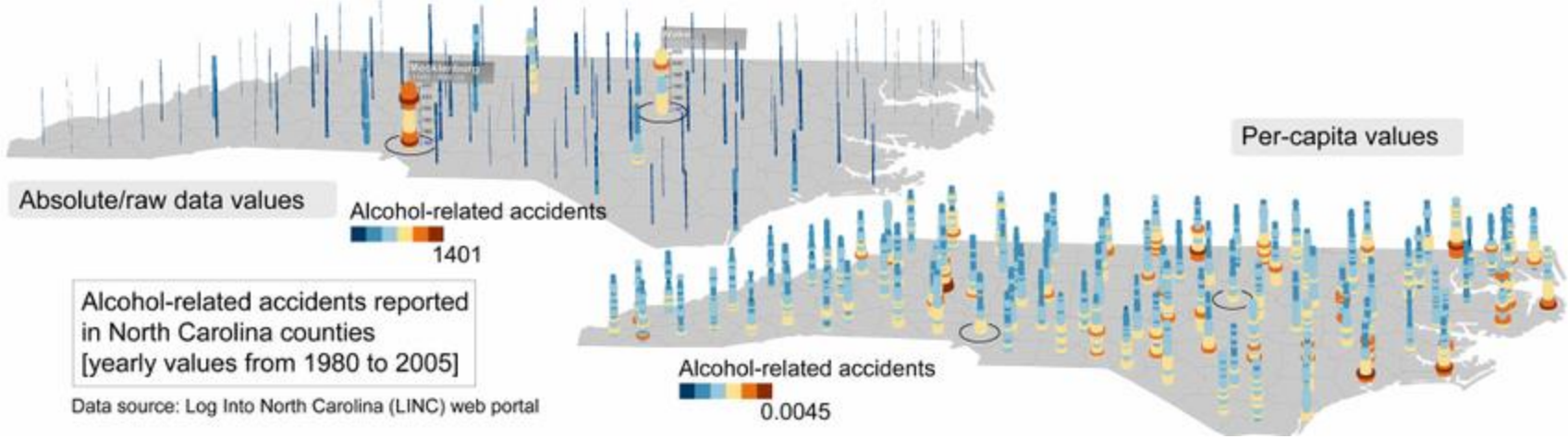
Visualization

Temporal-Spatial Visualization

- Food stamps issued in the 100 counties in North Carolina



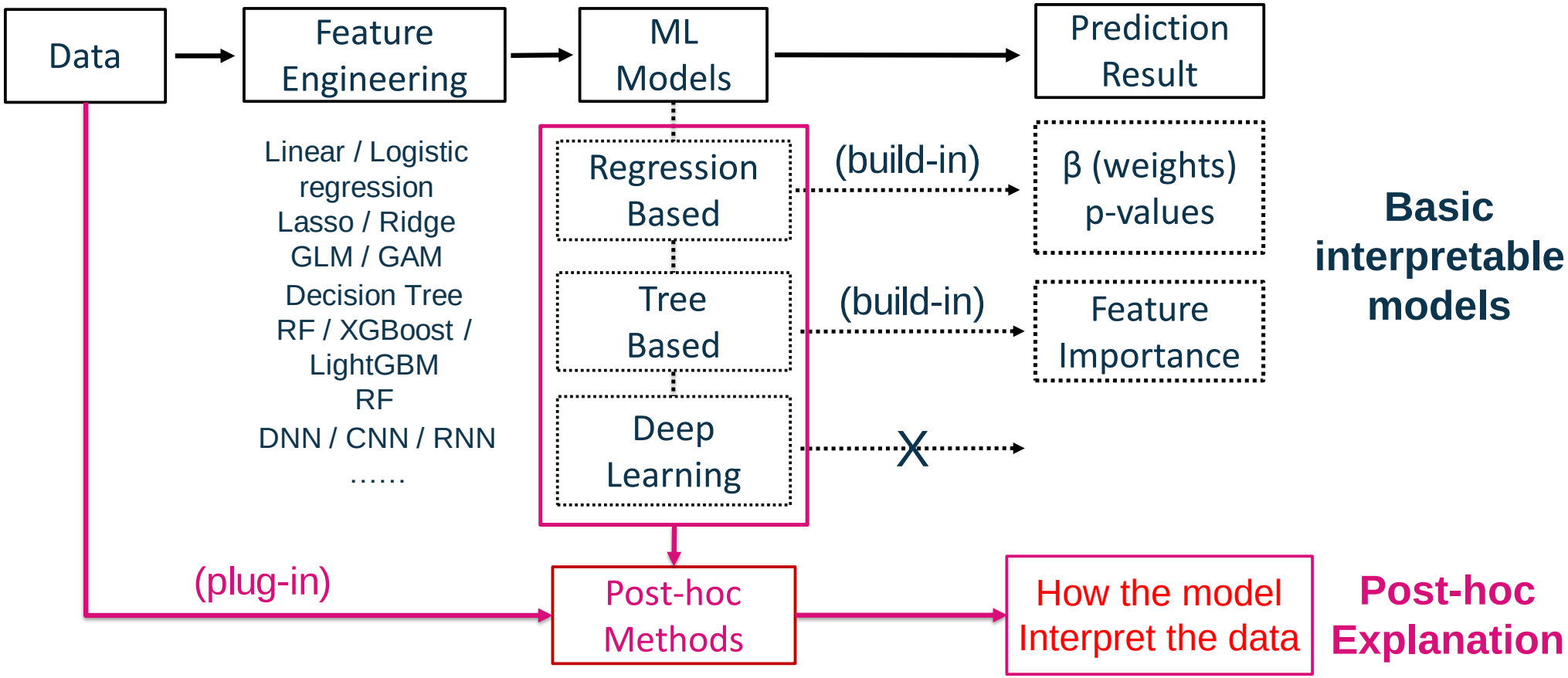
- Alcohol-related accidents



https://www.researchgate.net/figure/Visualizations-of-the-spatial-and-temporal-distributions-of-food-stamps-issued-in-the-100_fig3_216819851

Visualization and Interpretation

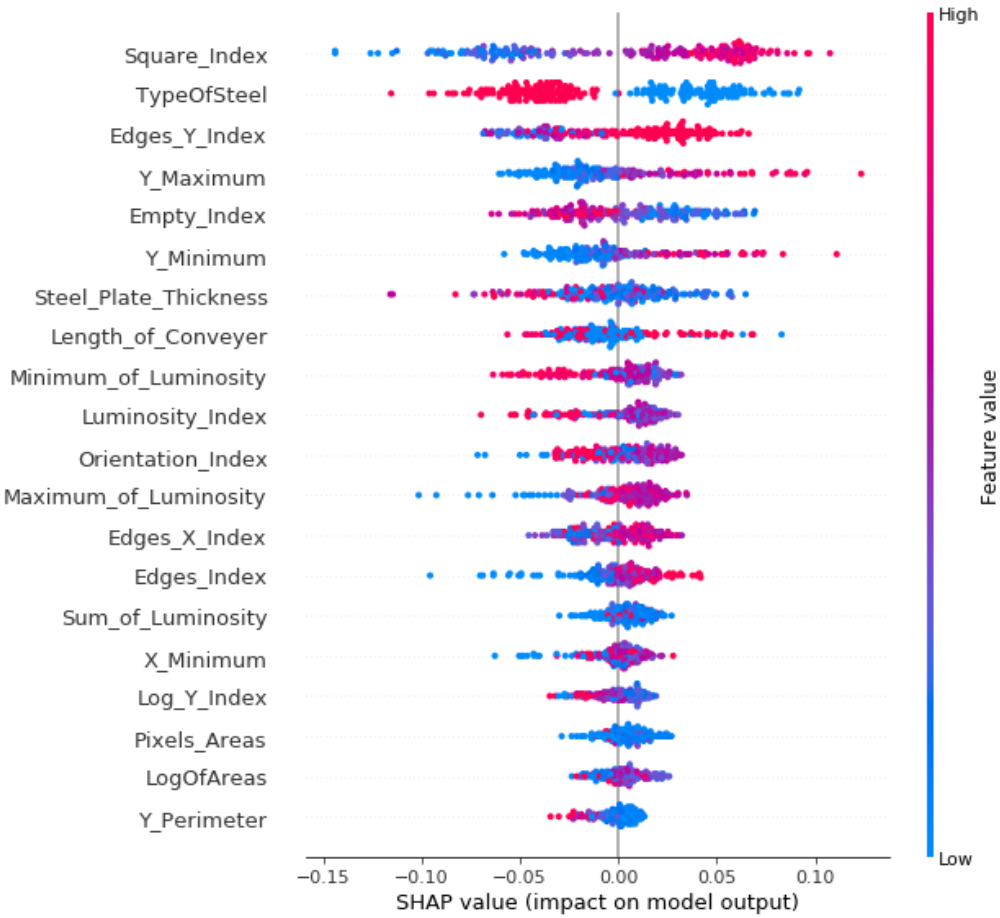
- AI is black box? Explainable AI (XAI)?



Hung, S.-Y. (洪紹嚴), 2019. **XAI (Explainable AI)-Introduction**.
https://speakerdeck.com/skydome20/xai-explainable-ai

Visualization and Interpretation

- Explainable AI (XAI)
- SHAP
  - Shapley value is a solution concept in cooperative game theory, which is used to divide the reward for each player according to their contributions. In machine learning model, Shapley value can be viewed as average marginal contribution to calculate the importance of a feature by comparing what a model predicts with and without the feature. The order in which a model sees features can affect its predictions, hence every possible order should be considered when calculate shapley values.



Source: Kaggle-Datasets
Dataset: Faulty Steel Plates
Features(x): 27 (25 continuous; 2 category)
Target(y): 7 classes of defects (six types + "other")
Sample size: 1941

Hung, I.-T. (洪翊庭), 2019. Introduction-to-Explainable-AI. https://github.com/ITingHung/Introduction-to-Explainable-AI

Concluding Remarks

The Sexiest Job in the 21st Century…?

Data Scientist

- *"Data Scientist" by Yahoo News 4/4/2013*
  - "PEHub reports pay for data scientists is upwards of US$225,000 even for people straight out of graduate school, up from $125,000 just a few years ago. For someone with a few years of experience working in the field, pay can reach much higher. One Seattle software-company CEO describes candidates with these skillsets “almost like unicorns.” One got away from the executive when Microsoft approached the data scientist with a $650,000 salary plus bonuses.“

http://finance.yahoo.com/blogs/daily-ticker/sexiest-job-21st-century-122238562.html

50 Best Jobs in America for 2020

	Job Title	Median Base Salary	Job Satisfaction	Job Openings	
#1	Front End Engineer	$105,240	3.9/5	13,122	View Jobs
#2	Java Developer	$83,589	3.9/5	16,136	View Jobs
#3	Data Scientist	$107,801	4.0/5	6,542	View Jobs
#4	Product Manager	$117,713	3.8/5	12,173	View Jobs
#5	DevOps Engineer	$107,310	3.9/5	6,603	View Jobs
#6	Data Engineer	$102,472	3.9/5	6,941	View Jobs
#7	Software Engineer	$105,563	3.6/5	50,438	View Jobs
#8	Speech Language Pathologist	$71,867	3.8/5	29,167	View Jobs
#9	Strategy Manager	$133,067	4.3/5	3,515	View Jobs
#10	Business Development Manager	$78,480	4.0/5	6,560	View Jobs

Glassdoor, 2020. https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm

50 Best Jobs in America for 2021

#1	Java Developer	$90,830	4.2/5	10,103	View Jobs
#2	Data Scientist	$113,736	4.1/5	5,971	View Jobs
#3	Product Manager	$121,107	3.9/5	14,515	View Jobs
#4	Enterprise Architect	$131,361	4.0/5	10,069	View Jobs
#5	Devops Engineer	$110,003	4.0/5	6,904	View Jobs
#6	Information Security Engineer	$110,000	4.0/5	5,621	View Jobs
#7	Business Development Manager	$82,182	4.1/5	8,827	View Jobs
#8	Mobile Engineer	$94,301	4.1/5	4,631	View Jobs
#9	Software Engineer	$110,245	3.8/5	40,564	View Jobs
#10	Dentist	$134,122	4.0/5	4,315	View Jobs

Glassdoor, 2021. https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm

Data Scientist

"*Data scientists are the people who understand how to fish out answers to important business questions from today's tsunami of unstructured information*."

Davenport and Patil (2012)

Data Scientist:

The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data.
by Thomas H. Davenport and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

70 Harvard Business Review October 2012

http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/

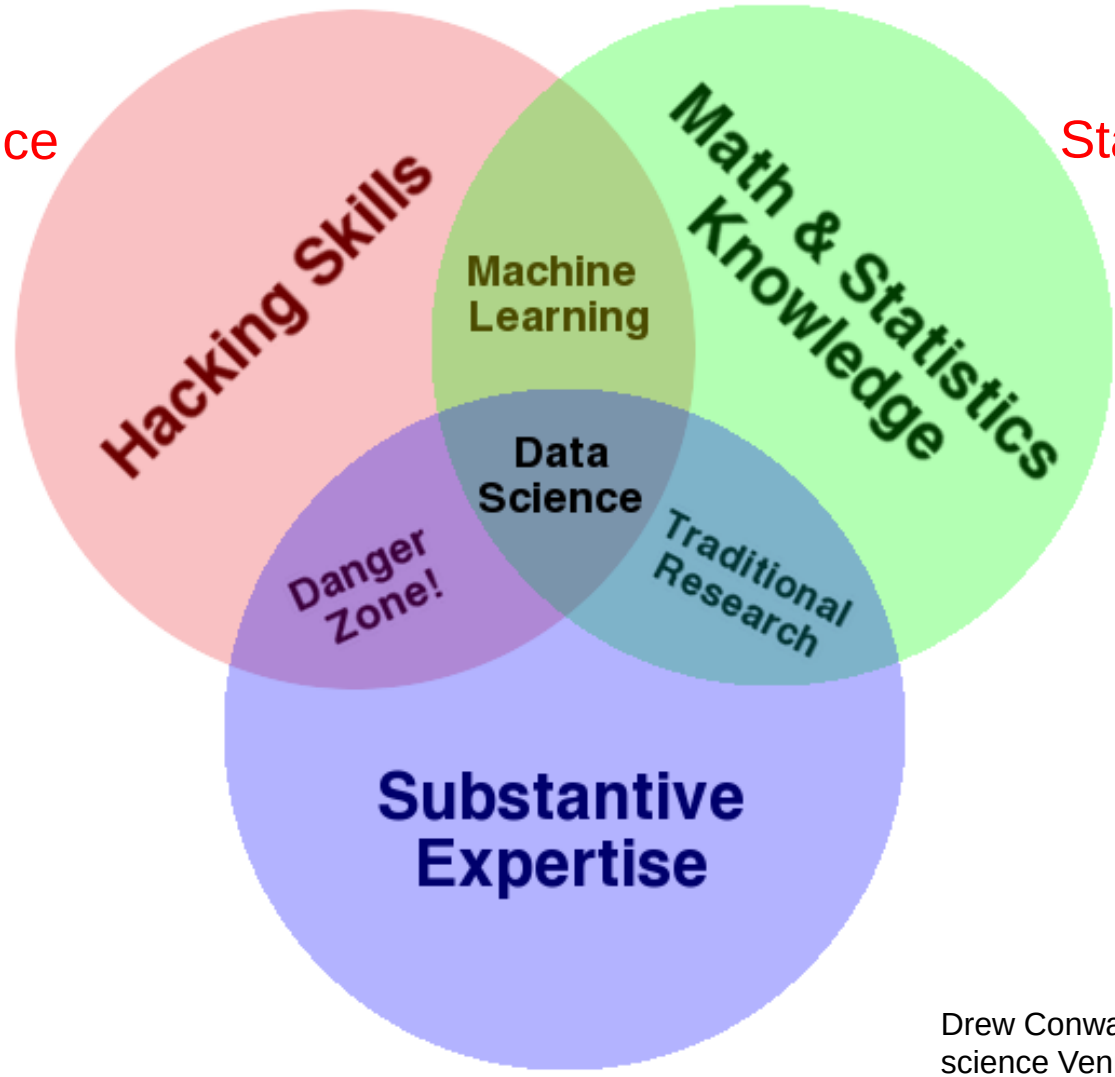
Some Highlights…

- Highlights from HBR
  - "…As companies rush to capitalize on the potential of big data, the largest constraint many face is the scarcity of this special talent."
  - "…The shortage of data scientists is becoming a serious constraint in some sectors."
  - "…Data scientists need autonomy but want to be "on the bridge," responding to management issues with their managerial colleagues in real time."
  - "…Data scientists want to build things, not just give advice. One describes being a consultant as "the dead zone."
  - "…Data scientists today are akin to the Wall Street "quants" of the 1980s and 1990s."

Data Scientist: nerd or geek !?

Computer Science

Statistical Theory



Domain Knowledge

Drew Conway Data Consulting (2013). The data science Venn Diagram. http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Data Scientist Should Have …

- Necessary Skills for Candidates…
  - " …user groups devoted to data science tools. The R User Groups (for an open-source statistical tool favored by data scientists) and Python Interest Groups (for PIGgies) …"
  - "…hang out with data scientists at the Strata, Structure: Data, and Hadoop World conferences and similar gatherings..."
  - "… host a competition on Kaggle or TopCoder, the analytics and coding competition. Follow up with the most-creative entrants…"
  - "… Coding skills don't have to be at a world-class level but should be good enough to get by. Look for evidence, too, that candidates learn rapidly about new technologies and methods…"

Data Scientist Should Have …

- Necessary Skills for Candidates…
  - "…Make sure a candidate can find a story in a data set and provide a coherent narrative about a key data insight…"
  - "…Be wary of candidates who are too detached from the business world. When you ask how their work might apply to your management challenges, are they stuck for answers?"
  - "…Ask candidates about their favorite analysis or insight and how they are keeping their skills sharp. Have they gotten a certificate in the advanced track of Stanford's online Machine Learning course, contributed to open-source projects, or built an online repository of code to share (for example, on GitHub)?"

Road to Data Scientist …

Dept. of Statistics

Probability & Statistics

Operations Research

Multivariate Analysis

Dept. of Computer Science

Data Structure

Algorithm

Database Management

Graduate Level

Production(產)
Marketing(銷)
HR(人)
R&D(發)
Finance(財)
Information(資)

Management Insight

Data Mining

Cloud/Edge Computing

Statistical Learning Theory

Data-Driven Decision Making

Applications

Signal Processing/ Time Series

Natural Language Processing/ Text Mining

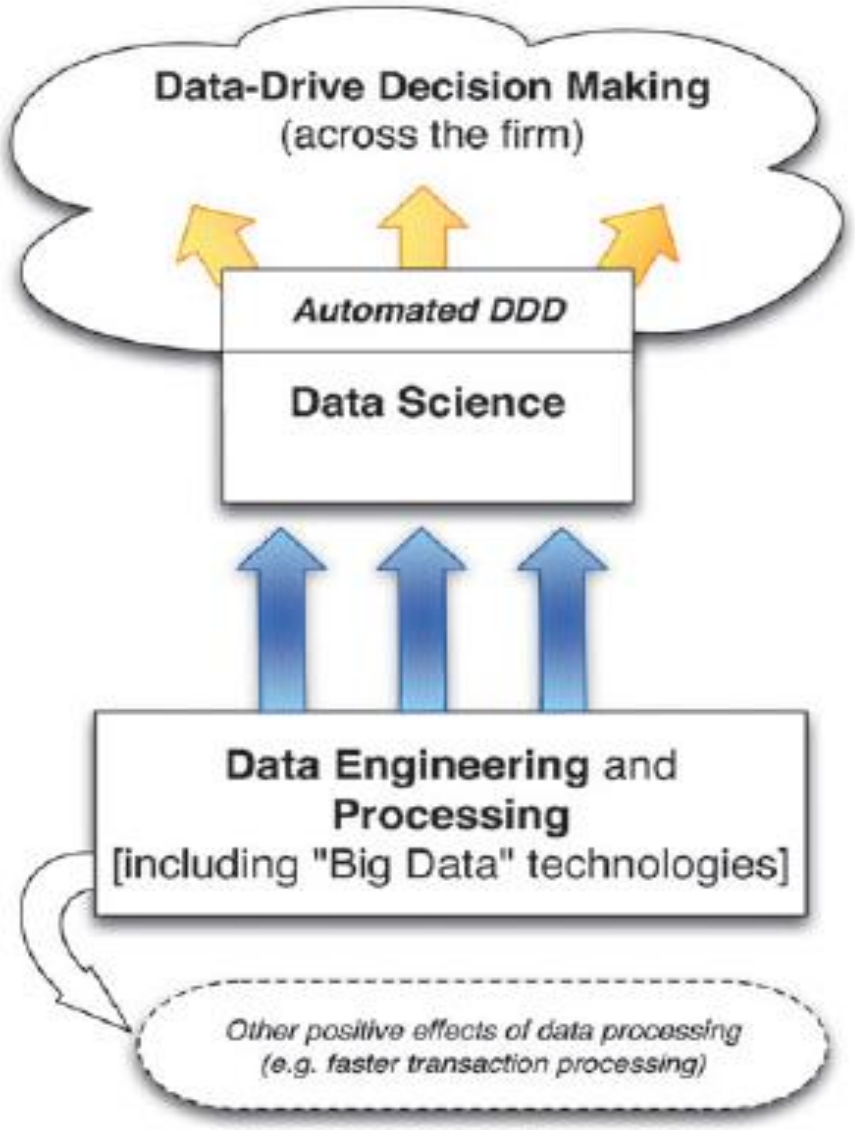
Domain Knowledge

Computer Vision/ Image Processing

Speech/Voice Processing

Source: Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, *1*(1), 51-59.

Data Science in the Organization



Data Science Team!!

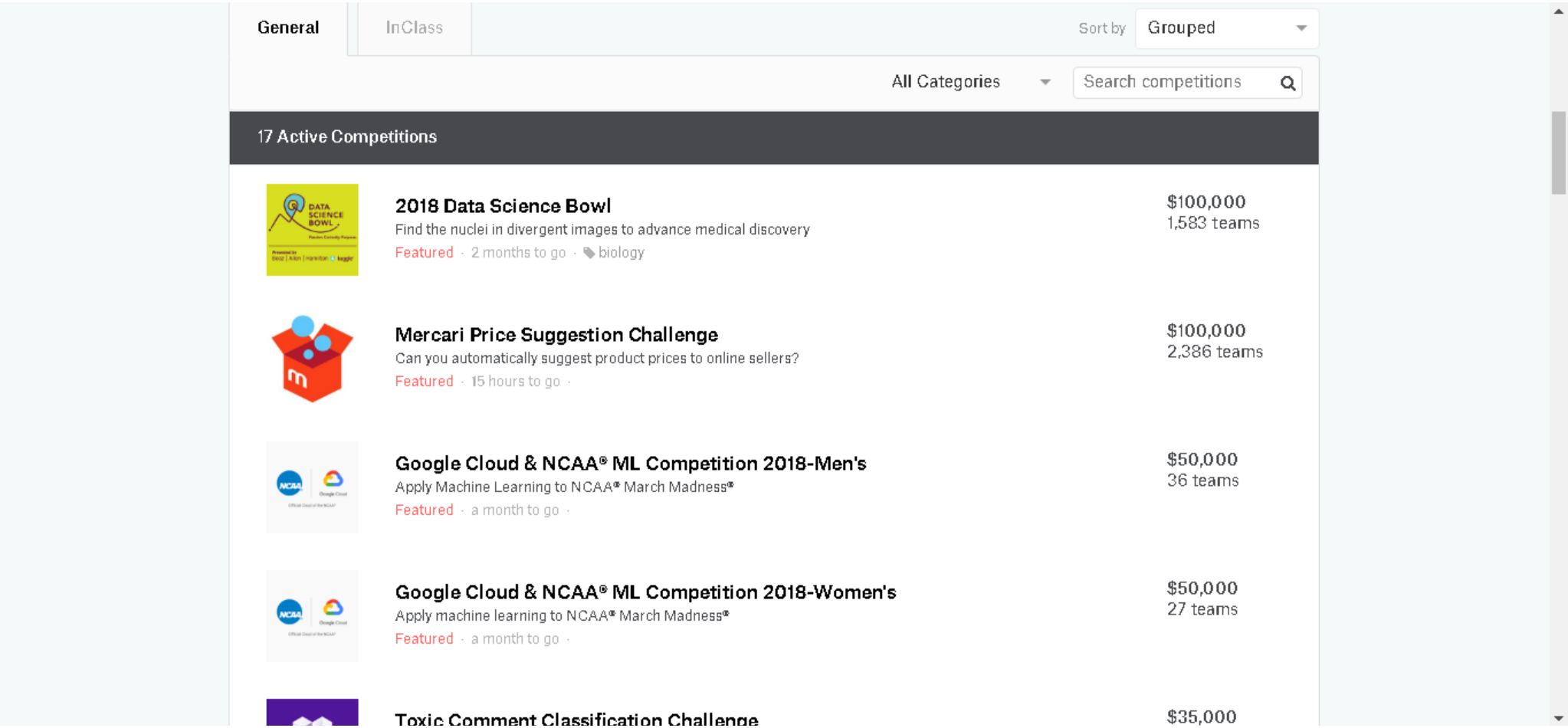
Source: Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, *1*(1), 51-59.

ML/DS Complexity and Issues

- Volume of Data
  - Clever algorithms needed for reasonable performance
  - Performance: efficiency, effectiveness, and scalability
- Interest measures and Interpretation
  - How do we ensure algorithms select “interesting” results?
  - How do we interpret results
- “Statistics, Programming, and Domain Knowledge” skill required
  - How to select methodology, prepare data?
  - Domain expert for different applications
- Data Quality
  - Garbage-in Garbage-out
- Data Source Heterogeneity
  - How do we combine data from multiple sources?
  - Structured and unstructured data
  - Protection of data security, integrity, and privacy

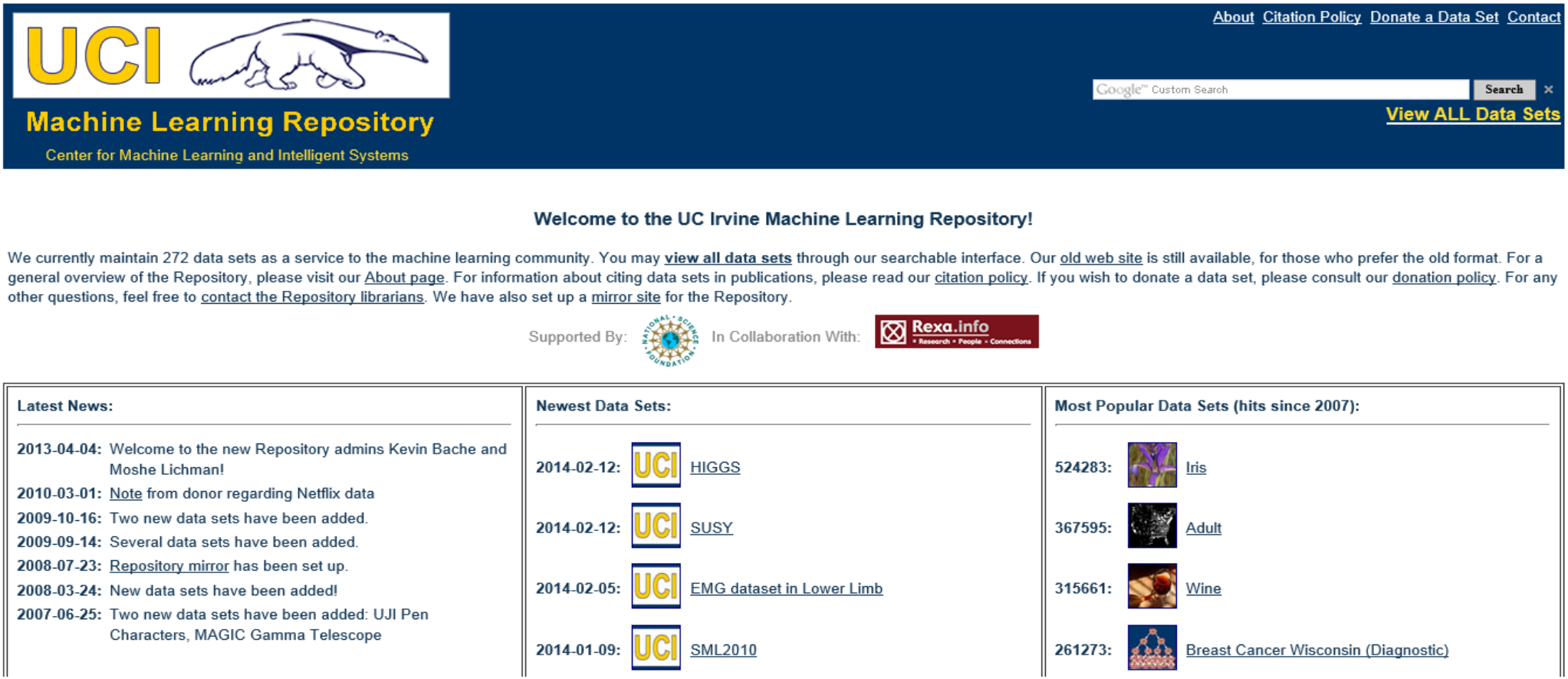
Kaggle

- Kaggle (https://www.kaggle.com/)



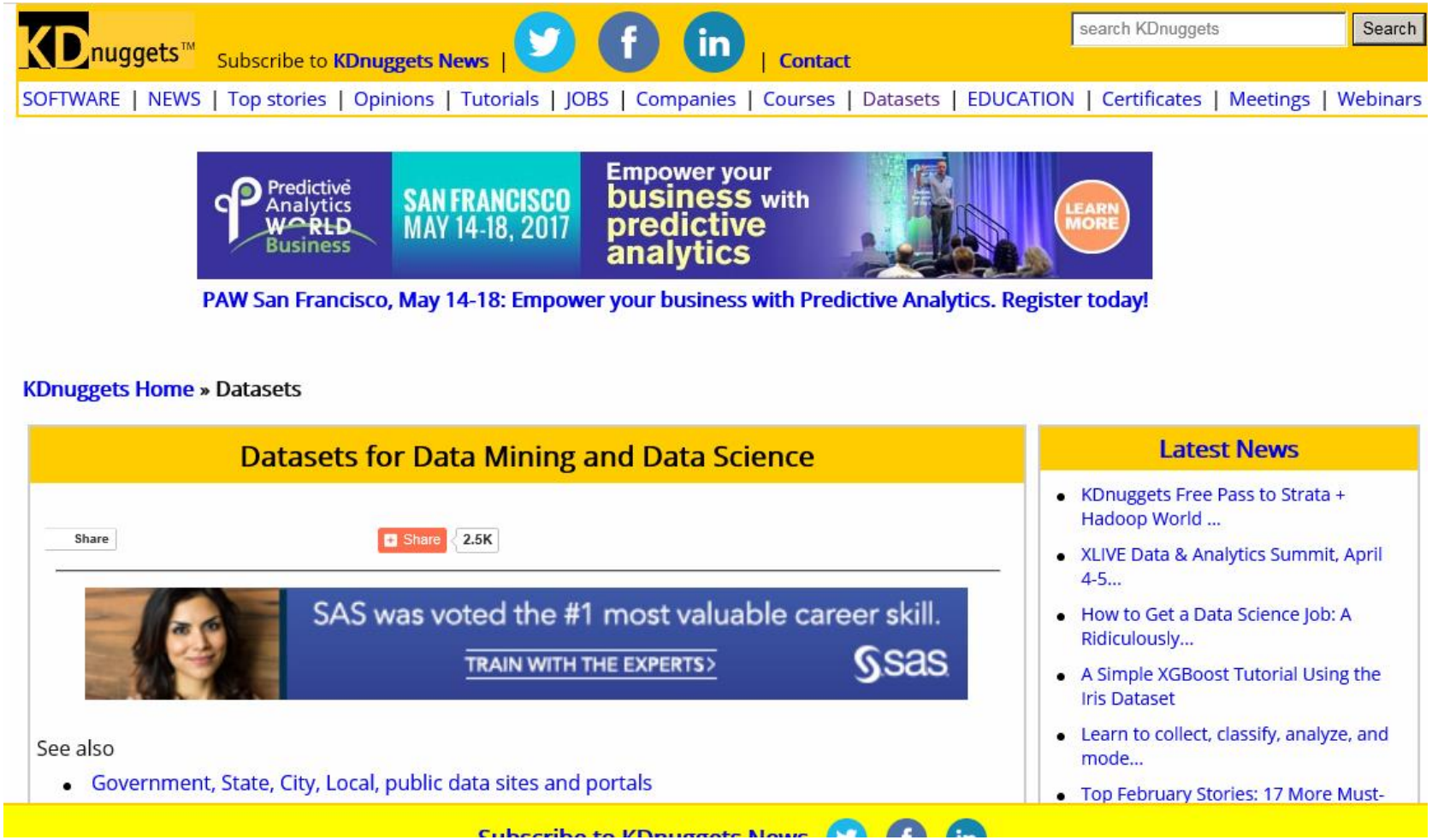
UCI Machine Learning Repository

- UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/)



Kdnuggets

- Kdnuggets (http://www.kdnuggets.com/datasets/index.html)

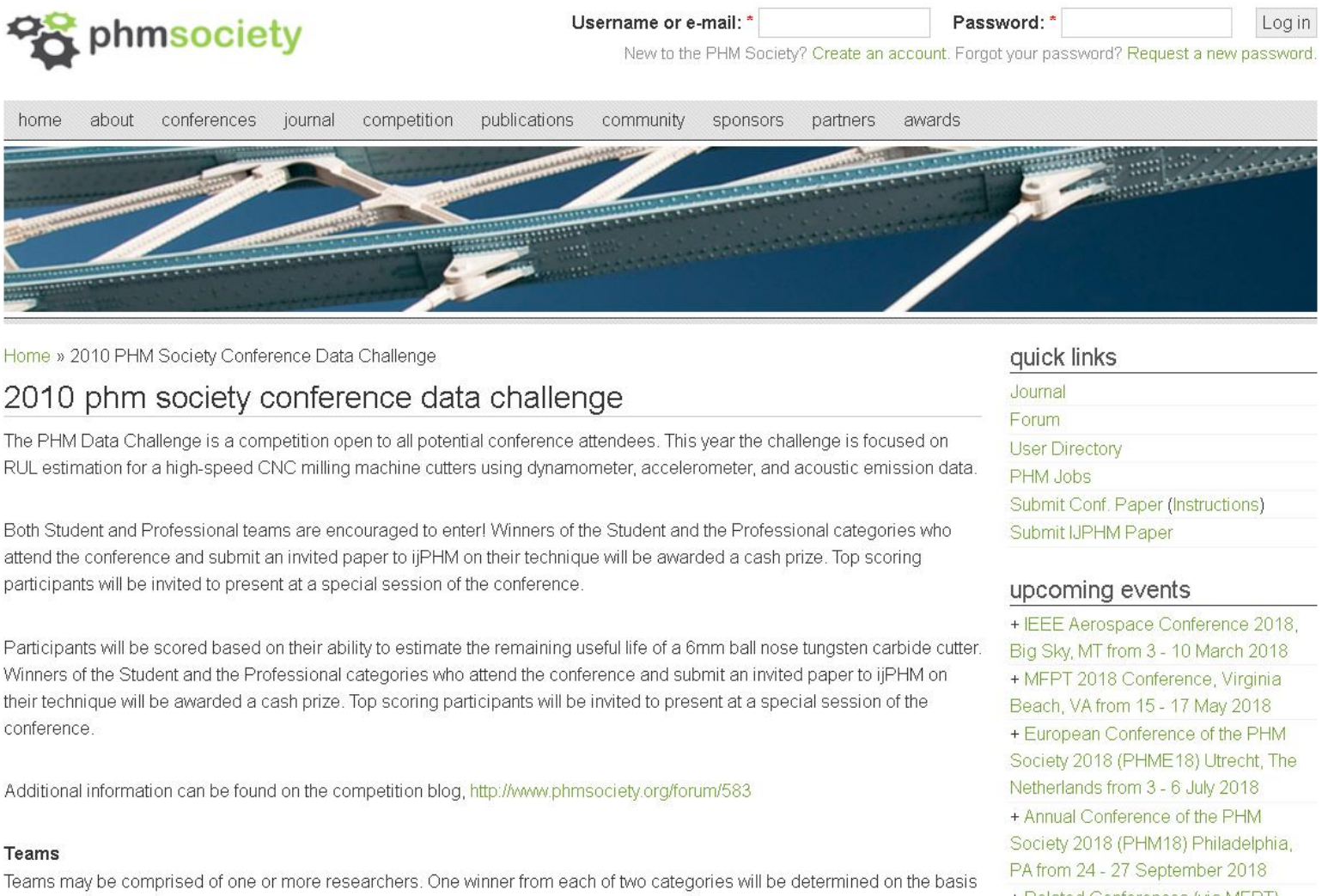


政府資料開放平台

- 政府資料開放平台 (https://data.gov.tw/)

Prognostics and Health Management (PHM)

- https://www.phmsociety.org/competition/phm/10



phmsociety

Username or e-mail: * Password: * Log in

New to the PHM Society? Create an account. Forgot your password? Request a new password.

home about conferences journal competition publications community sponsors partners awards

Home » 2010 PHM Society Conference Data Challenge

2010 phm society conference data challenge

The PHM Data Challenge is a competition open to all potential conference attendees. This year the challenge is focused on RUL estimation for a high-speed CNC milling machine cutters using dynamometer, accelerometer, and acoustic emission data.

Both Student and Professional teams are encouraged to enter! Winners of the Student and the Professional categories who attend the conference and submit an invited paper to ijPHM on their technique will be awarded a cash prize. Top scoring participants will be invited to present at a special session of the conference.

Participants will be scored based on their ability to estimate the remaining useful life of a 6mm ball nose tungsten carbide cutter. Winners of the Student and the Professional categories who attend the conference and submit an invited paper to ijPHM on their technique will be awarded a cash prize. Top scoring participants will be invited to present at a special session of the conference.

Additional information can be found on the competition blog, http://www.phmsociety.org/forum/583

Teams

Teams may be comprised of one or more researchers. One winner from each of two categories will be determined on the basis of score. The categories are:

quick links

Journal
Forum
User Directory
PHM Jobs
Submit Conf. Paper (Instructions)
Submit IJPHM Paper

upcoming events

+ IEEE Aerospace Conference 2018, Big Sky, MT from 3 - 10 March 2018
+ MFPT 2018 Conference, Virginia Beach, VA from 15 - 17 May 2018
+ European Conference of the PHM Society 2018 (PHME18) Utrecht, The Netherlands from 3 - 6 July 2018
+ Annual Conference of the PHM Society 2018 (PHM18) Philadelphia, PA from 24 - 27 September 2018
+ Related Conferences (via MFPT)

Summary

- Intelligent Manufacturing Systems (IMS)
  - Decision-oriented system: process, resources, function
- Manufacturing Systems and Automation
  - VPC analysis, manufacturing vs. service, CIM
- Data Science Functions and Applications
  - KDD process includes data preprocessing, features selection, transformation, modelling, pattern evaluation, and knowledge presentation
  - Generalization, feature selection, association rules, classification, prediction, clustering, time series, optimization, visualization, interpretation
  - Letter recognition, mechanical optimization, portfolio, capacity planning, shortest path, etc.
- Data Scientist
  - Statistics and Optimization, Programming Skills, Domain knowledge
  - ML/DS Complexity and Issues