



香港浸會大學
HONG KONG BAPTIST UNIVERSITY

School of
BUSINESS
工商管理學院

Department of Economics

ECONOMIC ANALYSIS AND APPLICATION OF BIG DATA

Monique S.K. Wan



AACSB
ACCREDITED



Agenda

- Definition of Big Data
- Industrial Revolution 4.0
- Applications of Big Data
 - Banking and risk management
 - Marketing and recommendation system
 - Government and smart city
- Implications of Big Data
 - Business landscape and job nature
 - Data as asset
 - General Data Protection Regulation
 - Role of economists

Definition of Big Data

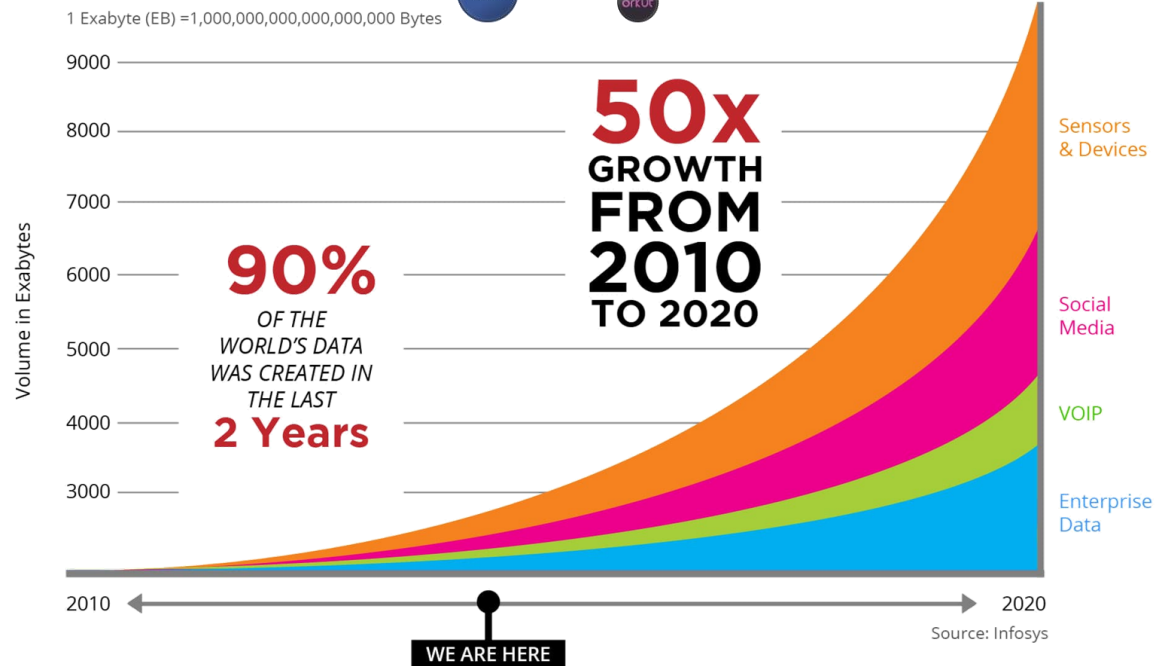
“Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making. ”

– Doug Laney of Gartner, Inc.



Every minute:

- 204 million emails
- 1,800,000 Facebook likes
- 72 hours of videos upload to YouTube



Broader Definition

- **Veracity⁴** – Quality of the data



30,522 of 31,410 people found the following review helpful

★★★★★ **Saved my marriage**, July 30, 2012

By [Mrs Toledo](#)

What can I say about the 571B Banana Slicer that hasn't already been said about the wheel, penicillin, or the iPhone.... this is one of the greatest inventions of all time. My husband and I would argue constantly over who had to cut the day's banana slices. It's one of those chores NO ONE wants to do! You know, the old "I spent the entire day rearing OUR children, maybe YOU can pitch in a little and cut these bananas?" and of course, "You think I have the energy to slave over your damn bananas? I worked a 12 hour shift just to come home to THIS?!" These are the things that can destroy an entire relationship. It got to the point where our children could sense the tension. The minute I heard our 6-year-old girl in her bedroom, re-enacting our daily banana fight with her Barbie dolls, I knew we had to make a change. That's when I found the 571B Banana Slicer. Our marriage has never been healthier, AND we've even incorporated it into our lovemaking. THANKS 571B BANANA SLICER!

Help other customers find the most helpful reviews

[Report abuse](#) | [Permalink](#)

Was this review helpful? 51,565 of 52,218 people found the following review helpful

★★★★★ **No more winning for you, Mr. Banana!**, March 3, 2011

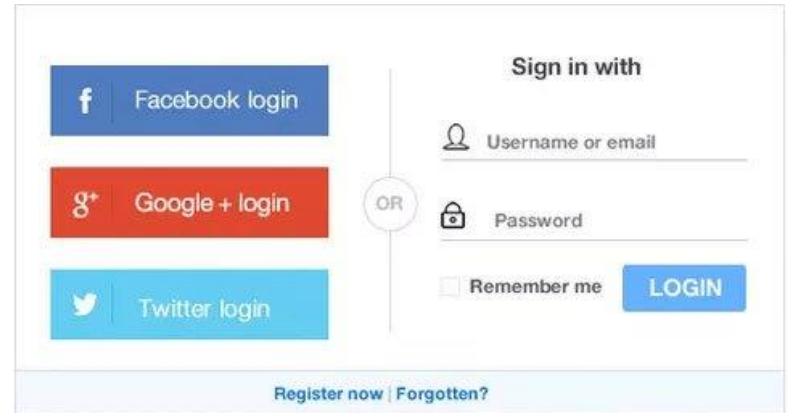
By [SW3K](#) - [See all my reviews](#)

For decades I have been trying to come up with an ideal way to slice a banana. "Use a knife!" they say. Well...my parole officer won't allow me to be around knives. "Shoot it with a gun!" Background check...HELLO! I had to resort to carefully attempt to slice those bananas with my bare hands. 99.9% of the time, I would get so frustrated that I just ended up squishing the fruit in my hands and throwing it against the wall in anger. Then, after a fit of banana-induced rage, my parole officer introduced me to this kitchen marvel and my life was changed. No longer consumed by seething anger and animosity towards thick-skinned yellow fruit, I was able to concentrate on my love of theatre and am writing a musical play about two lovers from rival gangs that just try to make it in the world. I think I'll call it South Side Story.

Banana slicer...thanks to you, I see greatness on the horizon.

Broader Definition

- **Valence⁵**
 - How big data can bond with each other?
 - Measure the ratio of actually connected data items to the possible number of connections that could occur within the collection



Sign in with

f Facebook login

g+ Google + login

OR

Twitter login

Username or email

Password

☐ Remember me

LOGIN

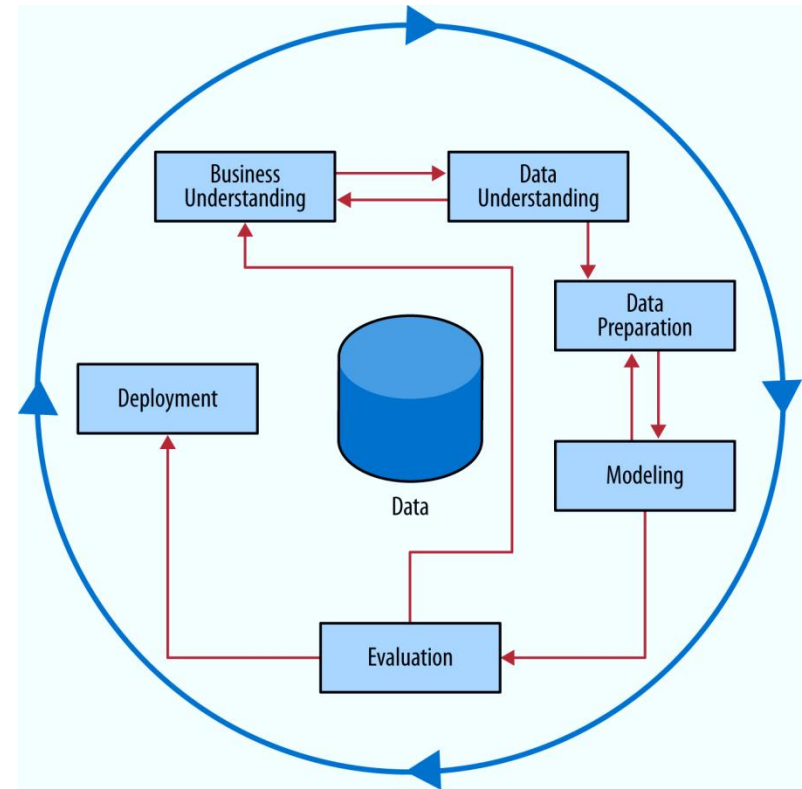
[Register now](#) | [Forgotten?](#)

Google
IS WATCHING
YOU



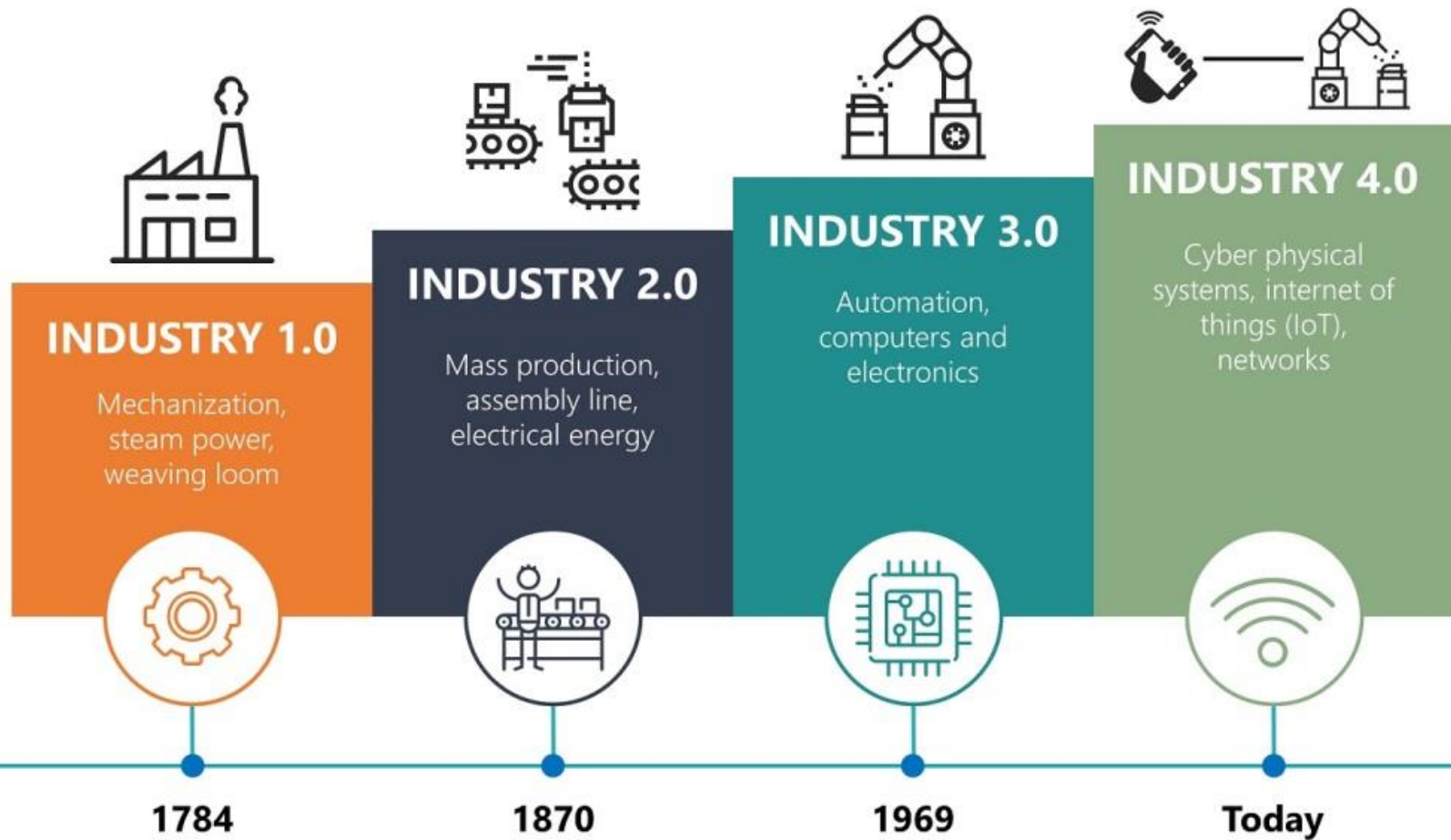
Broader Definition

- **Value⁶**
 - How can big data benefit your organization?
 - Require clear business strategy and data analytics tools
 - Cross Industry Standard Process for Data mining (CRISP-DM; Shearer 2000)



Source: Provost and Fawcett (2013)

Industrial Revolution 4.0



How Big is Big?



- **Company Signet Bank**
 - A small proportion of customers actually account for more than 100% of a bank's profit from credit card operations
- **Business objective**
 - Model profitability – Not just default probability, but also pricing
- **Data understanding/ Data preparation**
 - Use of algorithms to learn the willingness-to-pay and charge-offs
 - Acquire information via 45,000 scientific tests
 - Increase in bad accounts (charge-offs) ← **Cost-and-benefit analysis**

How Big is Big?

- **Deployment/ Action**

- Offer different customers different rates, terms, offers

- **Business value**

- A new spin-off in 1994 (Capital One)
 - Richard Fairbank and Nigel Morris
- Information-based strategy
- Lowest default rate
- Largest credit card issuers



How Fast is Fast?

- Vehicle with 100 telemetry sensors that capture data to improve track performance
- **Data science team**
 - Driver
 - Team principal/Race engineers/Operations room analysts/trackside analysts
- **Real-time data analysis**
 - Lap/split times
 - Tire/brake temp
 - Air pressure
 - Air flow
 - Engine performance
 - GPS information





How complex can it be?

- **Business understanding**
 - Provide streaming movie and TV service
 - One-third of peak-time Internet traffic in US
 - 65M members in over 50 countries
 - 100M hours of TV shows and movies a day
 - Build recommendation engines
- **Data understanding / Data preparation**
 - *Unstructured* data
 - 80,000 features (costly!!!) ← **Cost-and-benefit analysis**
 - Existing data = {Customer ID, movie ID, ratings and the date the movie was watched,...}
 - New data = {Time spent on selecting movies, how often playback is stopped, tags of the movies,...}

Valuable ?

- **Modeling**

- Measure similarity among products
- Measure similarity among customers
- **Predictive analytics** make recommendation
 - Improve user experience
 - Induce to consume more

- **Additional business value**

- **Product Innovation**
 - Develop new business as content creator
 - Outbid HBO: House of Cards directed by David Fincher and starred by Kevin Spacey

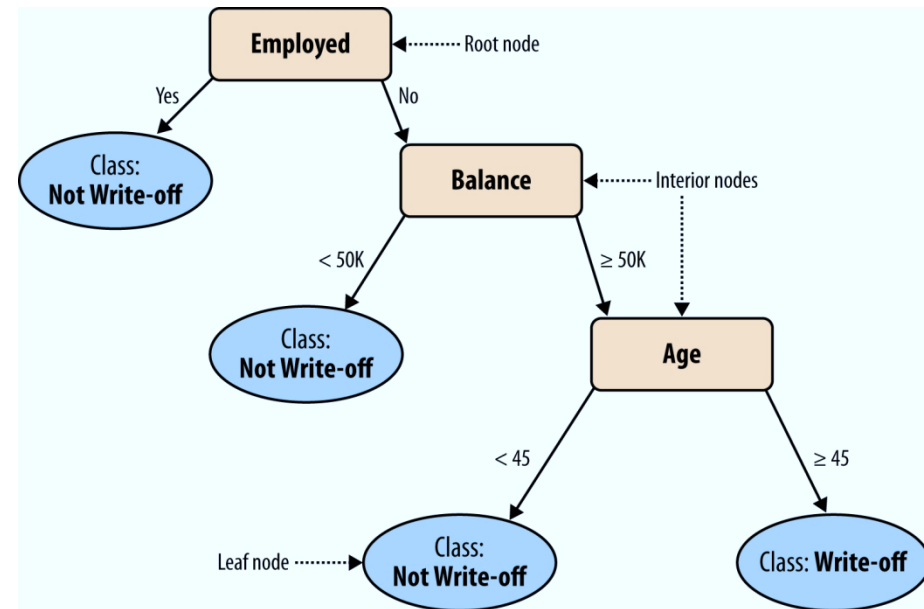


Banking & Risk Management

- **Business question**
 - What is the likelihood of default for this loan applicant ?
- **Data collection**
 - Past loan records: Applicant's profiles and outcome (default or not default)
 - The profile of this loan applicant
- **Modeling (supervised models)**
 - Classification tree
 - Logistic regression
 - Support vector machines
 - Neural networks

Decision Tree For Classification

- **Business decision**
 - Should we write off the loan?
- **Model**
 - Write off the loan if
 - Not employed; and
 - Remaining loan balance is huge; and
 - Older than 45



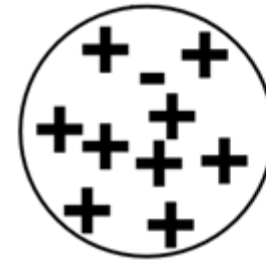
Source: Provost and Fawcett (2013)

Decision Tree

- **Measurement of Purity**

- **Error rate**

- **Definition:** $E = 1 - \max(p_+, p_-)$
 - **Example:** $E = 1 - \frac{9}{10} = 0.1$



$$p_+ = \frac{9}{10}$$
$$p_- = \frac{1}{10}$$

- **Gini index**

- A measure of total variance across the 2 classes
 - **Definition:** $G = p_+(1 - p_+) + p_-(1 - p_-)$
 - **Example:** $G = \frac{9}{10} \frac{1}{10} + \frac{1}{10} \frac{9}{10} = 0.18$

- **Entropy**

- **Definition:** $D = -p_+ \log_2 p_+ - p_- \log_2 p_-$
 - **Example:** $D = -\frac{9}{10} \log_2 \frac{9}{10} - \frac{1}{10} \log_2 \frac{1}{10}$

Logistic Regression

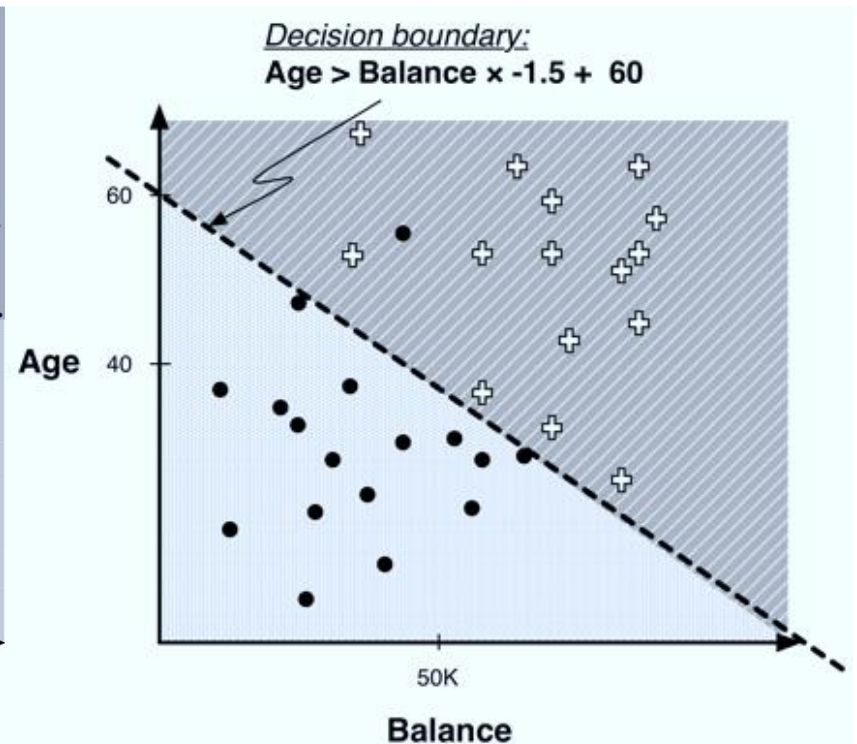
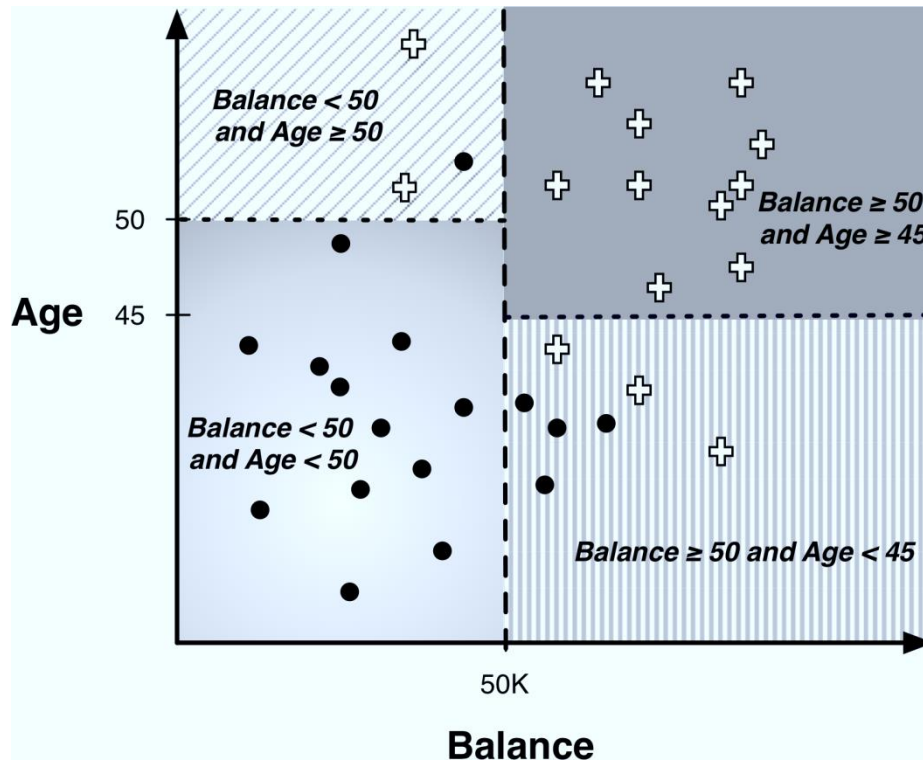
Target variable

$$y_i = \begin{cases} 1 & \text{Write - off} \\ 0 & \text{Not Write - off} \end{cases}$$

Feature variables

- Age_i
- $Balance_i$

$$\text{logit}(y_i) = -60 + Age_i + 1.5Balance_i$$



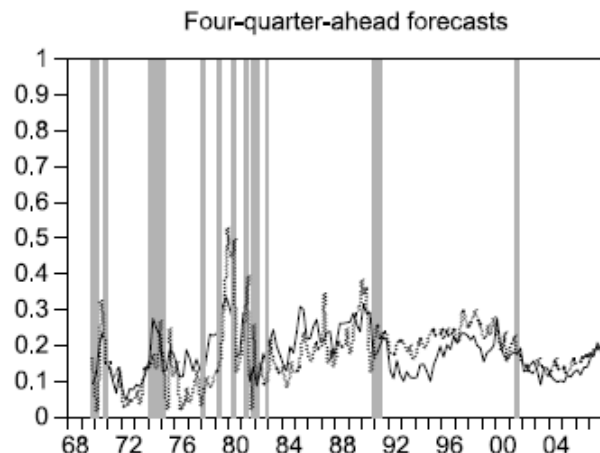
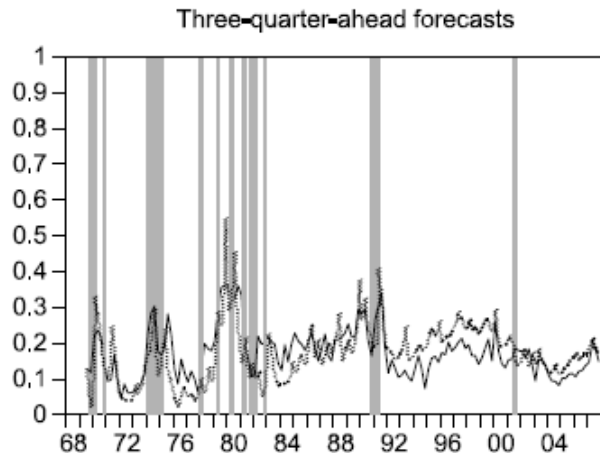
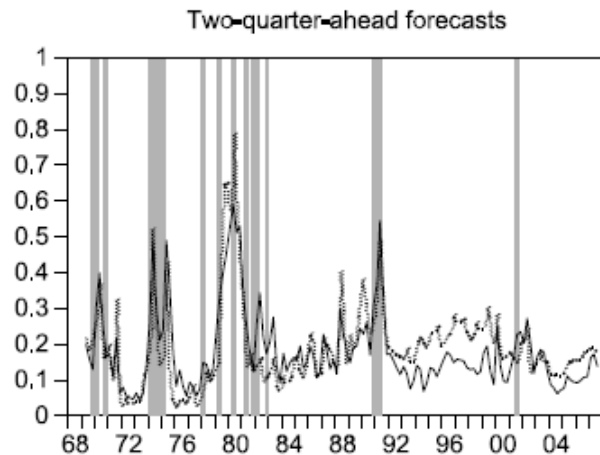
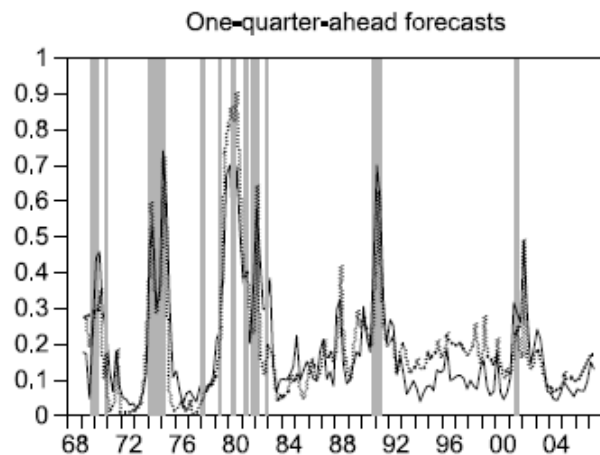
Forecasting Recession

Target variable

$$y_t = \begin{cases} 1 & \text{Recession} \\ 0 & \text{Otherwise} \end{cases}$$

Feature variable

$$\text{Yield curve}_t = LR - SR \text{ interest rate}$$



Marketing Strategy

- **Domain Knowledge**

- Cost = Contact Hours \times Wage Rate = \$25
- Product Price = \$200
- Product Cost = \$100
- **Respond (R)** \rightarrow Benefit = \$200 – \$100 – \$25 = \$75
- **No Respond (NR)** \rightarrow Benefit = –\$25

- **Expected benefit of targeting**

$$EV = p_R \times v_R + p_{NR} \times v_{NR}$$
$$EV = p_R \times \$75 + [1 - p_R] \times (-\$25) > 0$$
$$p_R > 0.25$$

- **Business action**

- Target the customers if their estimated probability of responding to the marketing plan >25%

Data Analytics

Model Prediction (Probability Score)	Action	Class
0.87	Contact	Respond
0.84	Contact	Respond
0.76	Contact	Respond
0.65	Contact	Not Respond
0.61	Contact	Respond
0.54	Contact	Respond
0.47	Contact	Not Respond
0.35	Contact	Not Respond
0.24	Do Not Contact	Respond
⋮	⋮	⋮
⋮	⋮	⋮

Recommendation System

- **Personalization**

- Predict user's interests and recommend products/services
 - Spotify, YouTube, Netflix

- **Data types**

- **Characteristic information**

- Users: User's background information, Preferences
- Items: Categories, Specific features

- **User-item interactions**

- Purchase histories, User ratings, Search engine queries, Browsing sequences

Recommendation System

- **Content-based system**
 - Based on user's profile features and item features
 - New products falling outside the past “interests” will not be recommended
 - New customers with not much information may be ignored
- **Collaborative filtering system**
 - Utilize user-item interaction
 - If a user likes item A, and another user likes A and B. Then, B will be recommended to first user
 - Identify clusters of users and items

Types of Recommendation System

Content-Based



Animated	Yes	Yes	No	No
Marvel	No	No	Yes	Yes
Super Villain	No	Yes	Yes	Yes
Bechdel Test?	Pass	Fail	Pass	Fail
Parks and Rec Alum	Yes (Amy Poehler)	Yes (John Hamm)	No	Yes (Paul Rudd)

Collaborative Filtering



Jason	Yes	Yes	Yes	Yes
Andi	No	Yes	No	Yes
Sarah	Yes	No	Yes	No
Sam	No	No	Yes	Yes
Scaz	Yes	??	??	??

Source: <https://www.youtube.com/watch?v=Eeg1DEeWUjA>

Government & Smart City

- **Technology**

- More information with higher accuracy and faster information flow

- **Applications**

- Transportation: Bus scheduling/ smart parking
- Banking: Cashless and contactless payments
- Waste/ Drainage management system: Robotics and assistive technologies
- Energy conservation: Data-empowered urban environments
- Healthcare: Technology-enabled homes

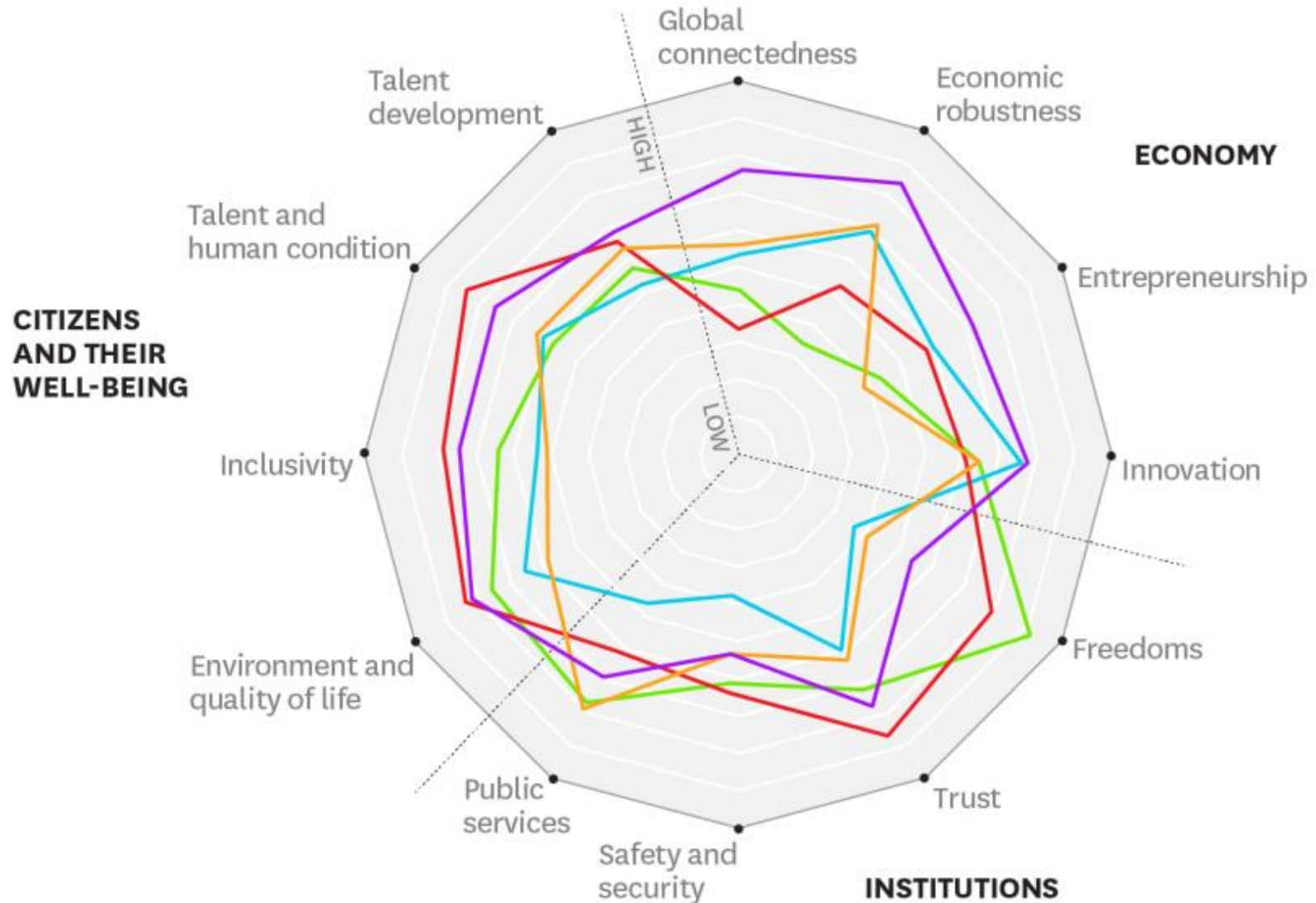
Government & Smart City

- Chakravorti and Chaturvedi (2017)
 - **Citizens/People Components**
 - Inclusivity, environment and quality of life, state of talent and the human condition, talent development
 - **Economy Components**
 - Global connectedness, economic robustness, entrepreneurial ecosystem, innovation capacity.
 - **Institutions Components**
 - Freedoms offline and online, trust, safety and security, public services

How 5 Countries Rank Against Smart Society Benchmarks

Each was rated on how well they use digital technology across 12 competencies in three major areas.

— Benchmark — Estonia — Israel — New Zealand — South Korea — United Kingdom



Business Landscape & Job Nature

Business Models

- **New business/ business models**
 - Data consulting companies
 - Cloud storage
 - Apps developers
 - Automation
 - Sharing economy
- **Traditional business models**
 - Entertainment companies
 - Media
 - Intermediaries

Job Nature

- **Job creation**
 - Data analysts
 - Computer scientists
 - Decision makers with strong domain knowledge
- **Job displacement**
 - Old fashioned sales representatives
 - Jobs with routine tasks

Cloud and Data as Asset



Source: <http://fortune.com/longform/alibaba-tencent-china-internet/>

Alibaba-Tencent struggle disrupts,
invigorates China internet



Tencent 腾讯

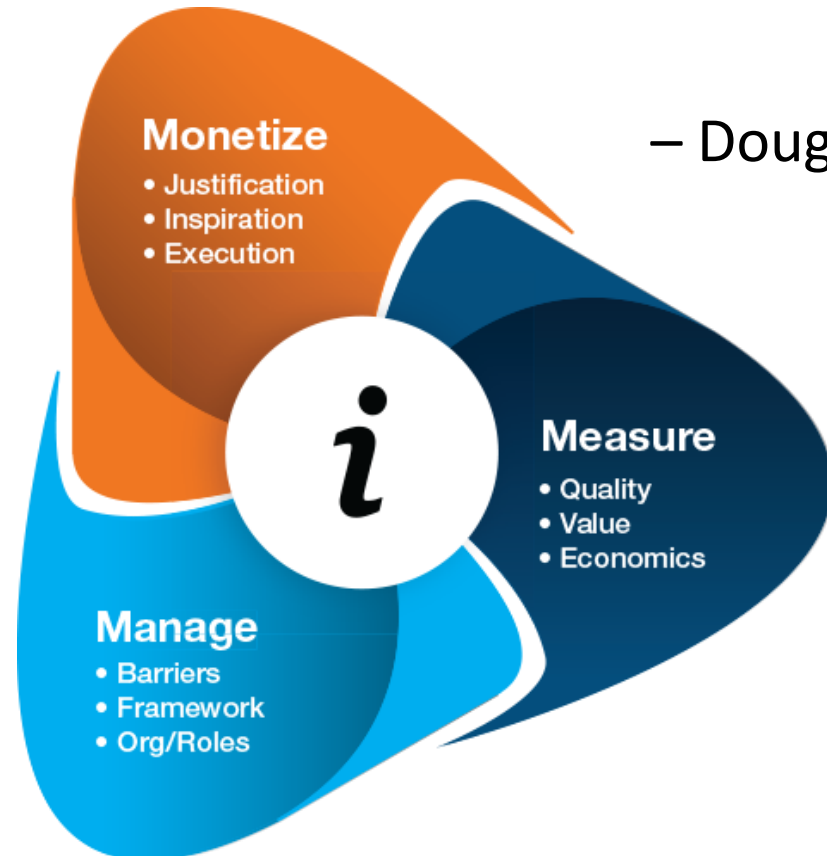
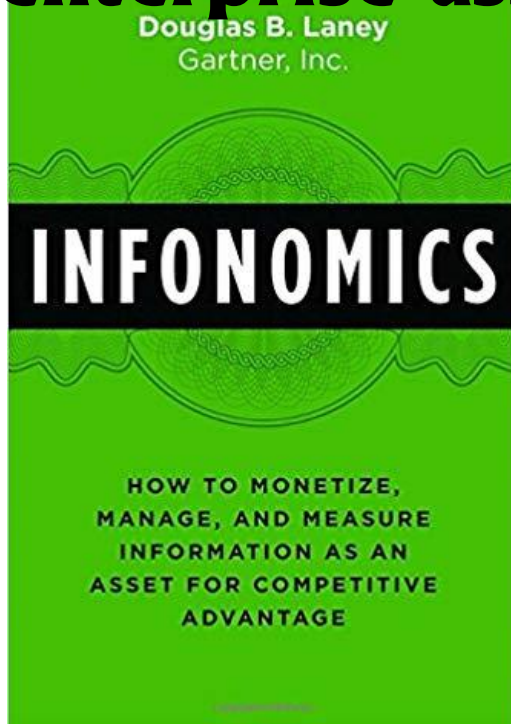


淘宝网
Taobao.com

Source: <https://www.bnnext.com.tw/article/44785/shun-feng-with-alibaba-china-most-strong-logistics-logistics-platform-of-decorum>

Produce and Manage Data Assets

“Infonomics is the concept that information is, or should be, an actual enterprise asset.”



– Doug Laney

Global Data Protection Regulation

- **7 Principles**

- **Purpose** limitation
- Data minimization
- **Accuracy**
- Storage limitation
- Lawfulness, **fairness** and **transparency**
 - Requiring the consent of subjects for data processing
- **Integrity** and **confidentiality**
 - Anonymizing collected data to protect privacy
- **Accountability**
 - Safely handling the transfer of data across borders
 - Providing data breach notifications



Role of Economists

- Macroeconomists
- Microeconomists
- Econometricians

Robert Solow



Robert Lucas



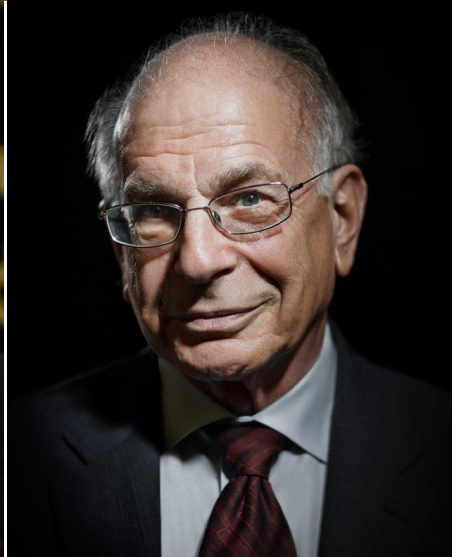
James Heckman



Daniel McFadden



John Nash



Daniel Kahneman

Jobs for Economists

- Examples of tech companies that have hired PhD Economists

Pat Bajari

Hal Varian

Susan Athey

Johnathan Hall



Alibaba
AirBnB
Amazon
AppNexus
CoreLogic
Coursera
Dstillery
Didichuxing
Digonex
eBay
ECONorthwest
Expedia
Facebook

Forkcast
Glassdoor
Google
Granular
Groupon
Houzz
Huawei
IBM
Indeed
ING
Intel
Kensho
Lending Club

LinkedIn
Lyft
Microsoft
Netflix
Nuna
Oath
OpenAI
Pandora
Pinterest
PoliticalSheepdog.com
Prattle
Quantco
Quora

Redfin
Ripple
Rover
Trulia
Uber
Upwork
Vericred
Visa
Walmart
Wealthfront
Yahoo!
Yelp
Zillow

Role of Economists in Tech Firms

- **Empirical Industrial Organization** (Athey and Luca, 2019; Shum, 2016)
 - Estimate demand function and market power
 - Design online advertising strategies and estimate returns to advertising
 - Design review and reputation systems and analyze the effect of reviews
 - Evaluate acquisitions, exclusive deals and strategy
 - Promote incentives in marketplaces

Demand Function

- **Industrial organization**
 - More about supply-side (firm-side)
 - How much market power do firms have?
- **Market power**
 - Markup: $\frac{p-mc}{p}$
 - Marginal cost (mc) is unobserved!
- **Observation:** High price (p) in an industry
 - High market power?
 - High mc ?

Demand Function

- **Monopoly**

$$\max_p pq(p) - C(q(p))$$

- **First-order condition**

$$q(p) + pq'(p) = C'(q(p)) q'(p)$$

- **Optimal price**

$$\frac{p^* - mc(q(p^*))}{p^*} = - \frac{q(p^*)}{q'(p^*)} \frac{1}{p^*}$$

**Price elasticity
of demand**

Traditional Approach to Demand Estimation

- **Consumer demand as a utility maximization problem**

$$\max_{x_1, x_2} U(x_1, x_2) \text{ s.t. } p_1 x_1 + p_2 x_2 = M$$

- p_1 and p_2 are prices of good 1 and good 2
- M is income

- **Solution**

- $x_1^*(p_1, p_2, M)$ and $x_2^*(p_1, p_2, M)$

- **Question**

- Do we really need x_1^* good 1 and x_2^* good 2? **← Discrete choice**

Discrete Choice Modeling

- **Product nature/ Consumption pattern**
 - Many alternatives, too many parameters
 - Automobile, airlines, cereals, toothpaste
 - Consumers only choose one of the available options (discrete choice)
- **Random utility framework** (McFadden, 1978, 1981)

$$\max_{j,z} U_i(x_j, z) \text{ subject to } p_j + p_z z = y_i$$

- x_j is the j^{th} product with price p_j
- z is other product with price p_z
- y_i is the income for the i^{th} individual

Discrete Choice Modeling

- **Indirect utility function**
 - $U_{i,j}^* = V_{ij}(p_j, p_z, y_i) + \varepsilon_{ij}$
 - $V_{ij}(p_j, p_z, y_i)$ is a function that depends on observed variables
 - ε_{ij} is utility shock that is not observed by others
- **Consumer preference**
 - Choose product j when $U_{i,j}^* > U_{i,k}^*$ for $k \neq j$
- **Objective**
 - Estimate the choice probabilities for all goods



Economics

44^{open}
jobs

Leading strategic decision-making with data and technology

Smart decision-making has taken Amazon from a garage-based startup to a leader in the technology, retail and content industries. As the company has grown, so has the need for sophisticated approaches and systems for answering business questions. Economists partner with business, engineering, finance, and science leaders to apply scientific techniques and develop complex, large-scale models that address key business problems faced across Amazon's many organizations, including retail, cloud computing, search, Kindle, streaming video and operations. Whether your background is in applied micro, applied macro/time series, empirical IO, theoretical econometrics, or another specialty area, there are questions to which you can apply your training, and an abundance of data to leverage. We have roles within the Central Economics Team as well as embedded within business organizations. Regardless of where you work at Amazon, you'll collaborate with Chief Economist Pat Bajari and a large community of talented PhD economists to address some of the most challenging projects in tech and ecommerce.

At Amazon, you'll have the opportunity to take your skills and apply them directly to the business with large and immediate business impacts – and never struggle to source high quality data for your analysis again.



Economists at Amazon: Making an Impact

References

- Athey (2017). Beyond prediction: Using big data for policy problems. *Science* 355: 483-485.
- Athey and Luca (2019). Economists (and economics) in tech companies. *Journal of Economic Perspectives* 33(1): 209-230.
- Varian H.R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives* 28(2): 3-28.
- Rudebusch G.D., and Williams J. (2005). Forecasting recessions: The puzzle of the enduring power of the yield curve. *Journal of Business & Economic Statistics* 27(4): 492-503.
- Chakravorti B. and Chaturvedi R.S. (2017) The “smart society” of the future doesn’t look like science fiction. *Harvard Business Review*.
- Provost F. and T. Fawcett (2013). *Data Science for Business*. 1st Ed. O’ Reilly Media.
- Bernard M. (2016). *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*. 1st Ed. Wiley.
- Shum M. (2016). *Econometric Models for Industrial Organization*. World Scientific Lecture Notes in Economics Vol. 3. World Scientific.