Data Analytics: FinTech & Real-World Applications

Prof. 廖世偉, NTU
liao@csie.ntu.edu.tw

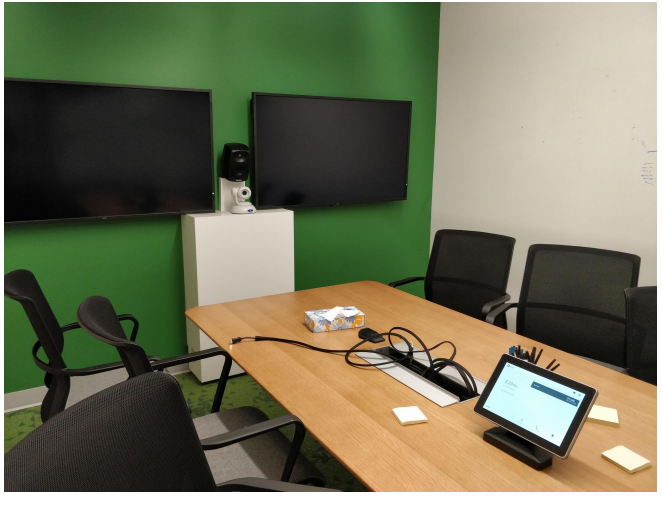
Instructor:

廖世偉
Steve Liao

Speaker Intro:

- Half & Half: 1969-1991: Taiwan. Afterwards: USA.
  Ex-Googler since 2014.
  - Went to Stanford for PhD, after receiving BS from NTU.
  - Left Google now, after receiving Google Founders' award.
- Recent impacts:
  - 3 blockchains: One was described in "Personalized Difficulty Adjustment for Countering the Double-Spending Attack in Proof-of-Work Consensus Protocols," published in IEEE Blockchain conference 2018 (15% acceptance rate). Second was presented in an invited Google Tech talk on YouTube. The last one is tangerine.
  - Creator of smart explorer (Google's machine-learning based data-center optimizer): Published in IEEE Supercomputing (SC).
  - Android tech lead on low-level Virtual Machine & Toolchains
  - Creator of RenderScript frontend & backend, linking loader.
  - Open-source leadership: MCLinker co-founder and committer, LLVM committer, and SUIF committer. The Android code above has been open-sourced.
  - Global citizenship: Solve world-class problems. Contribute to open source.
- Publication:
  - Author of ~75 papers and 25 patents.

International Collaboration



Outline of the Talk

- Digital transformation (DX)
- Data analytics
- Single-variable model (單變量模型) (未考慮價格因素，只考慮前n期的需求量)
- Multi-variable model (多變量模型) (考慮價格因素及前n期的需求量)
- Who are good customers?
- FinTech ABCD
- How it works in real world?

Digital Transformation (DX): 數位轉型

Definition: Using digital tools to boost:

- 效率 (efficiency) and
- 效能 (effectiveness);

Definition in Chinese: 用數位工具，為產業帶來效率及效能的提昇;

emphasizing：

- Customer-oriented (以客為尊)
- System integration (整合)

API型人才請上座

「所有數位轉型的企業，多數都困在IT上，」大聯大控股執行長葉福海跟我分享，「我花了整整一年的時間對IT部門洗腦。」沒想到，這位出身老宏碁的資訊大將，跟我這個IT門外漢的痛點一致。

我們都同意，再好的轉型策略，如果沒有科技賦能，哪裡都去不了，「IT已經從後勤單位變成第一線的作戰單位了。」但為什麼這一步這麼難？

尤其在傳統的中年企業裡，當IT部門舉目望去都是老員工，他們過去工作單純，滿足內部需求即可，但現在商業模式全變了，所有人都必須以用戶為核心，當用戶需求超越你目前所能提供的範疇，你就得放棄本位主義，盡快到外頭的市場找到解決方案。

此時，若內部沒有一個敏捷而開放的系統架構（包括作業流程等），隨時能與外部無縫對接，新的生意是進不來的。尤其是IT架構，要從一個原本封閉的獨家系統，轉型成海納百川的開放架構，關鍵在IT人的心態，願不願意從一個「驕傲的工程師」轉型成為「謙卑的整合者」，從技術導向轉向服務導向。

「不要再想著什麼都要自己來了，市面上有那麼多現成的API（應用程式介面，Application Programming Interface，扮演應用程式和應用程式之間的橋樑），要想的是，如何善用API對外連結，」葉福海苦口婆心的傳教。

過去工程師追求技術領先，但此刻技術早已成熟，市面上可選擇的技術方案非常多，市場根本不缺工程師，因為最難的是應用，一個能夠真正理解用戶需求、讓技術落實在商業場景、對接內外資源的「API型人才」，才是最稀缺的。

在我看來，不只資訊人員得善用API，所有人都要練習把自己變成API。以程式開發的角度來看，一組漂亮的API模組，必須具備低耦合（Coupling）、高內聚（Cohesion）性。前者是指不同程式模組間不應有太高的依賴性，避免牽一髮而動全身；後者則是指每一組程式要極大化可重複使用性，如此才能讓整個系統的運

數位轉型

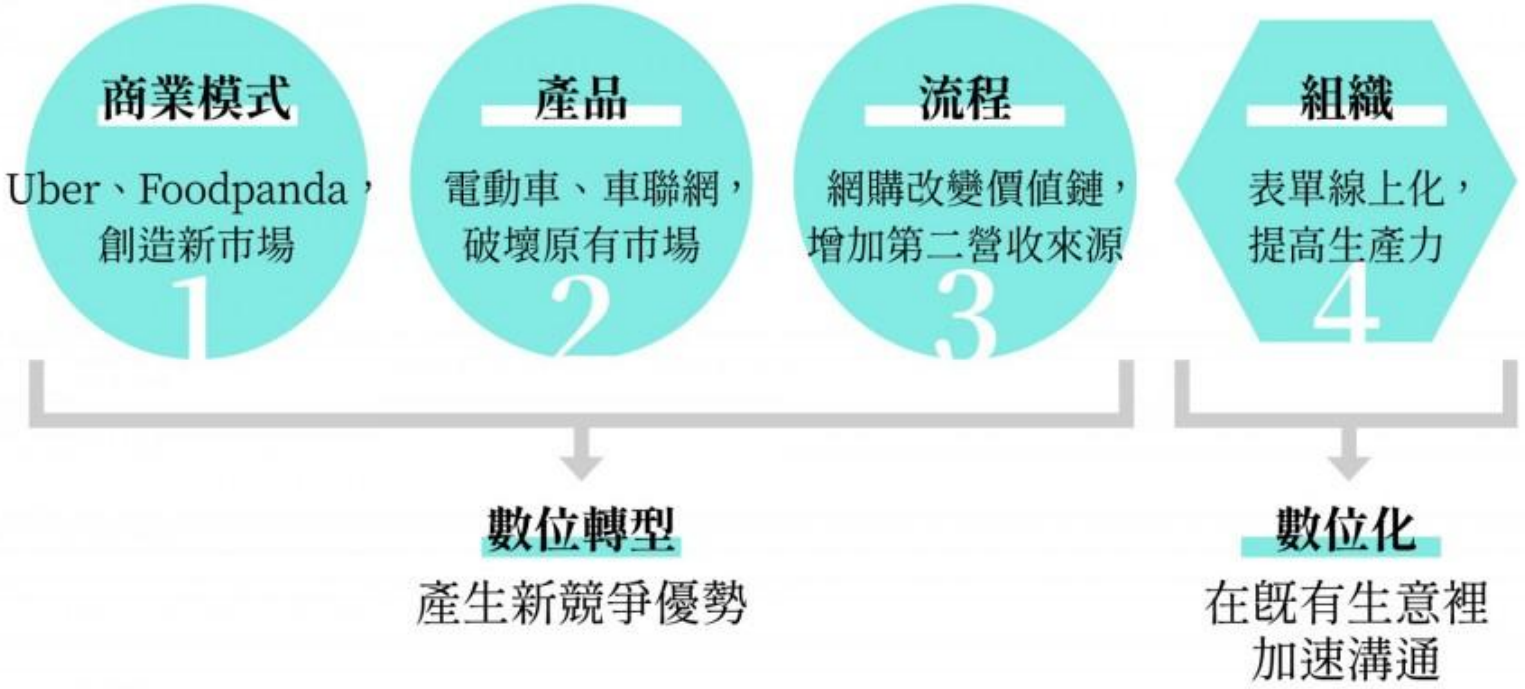
商業週刊
2020/6/13

DX in real world:

- Especially after the COVID-19 outbreak, business is going digital, internet, agile and speedy. DX embodies itself in:
  - 營運模式與生產管理 (Operation & Production management)
  - 行銷與業務 (Sales & Marketing)
  - 資訊科技 (Information Technology)
  - 創新研發 (Innovation & R&D)

數位化≠數位轉型，別再搞錯

科技的 4 種影響，並非都是數位轉型……



3 Steps in Digital Transformation (DX)

3 Steps: 數位轉型三部曲:

1. 數位化 (Digitize)
    - E.g., call 台灣大車隊 (Taxi hotline)

2. 數位優化 (AI optimizations)
    - E.g., Taxi route optimization

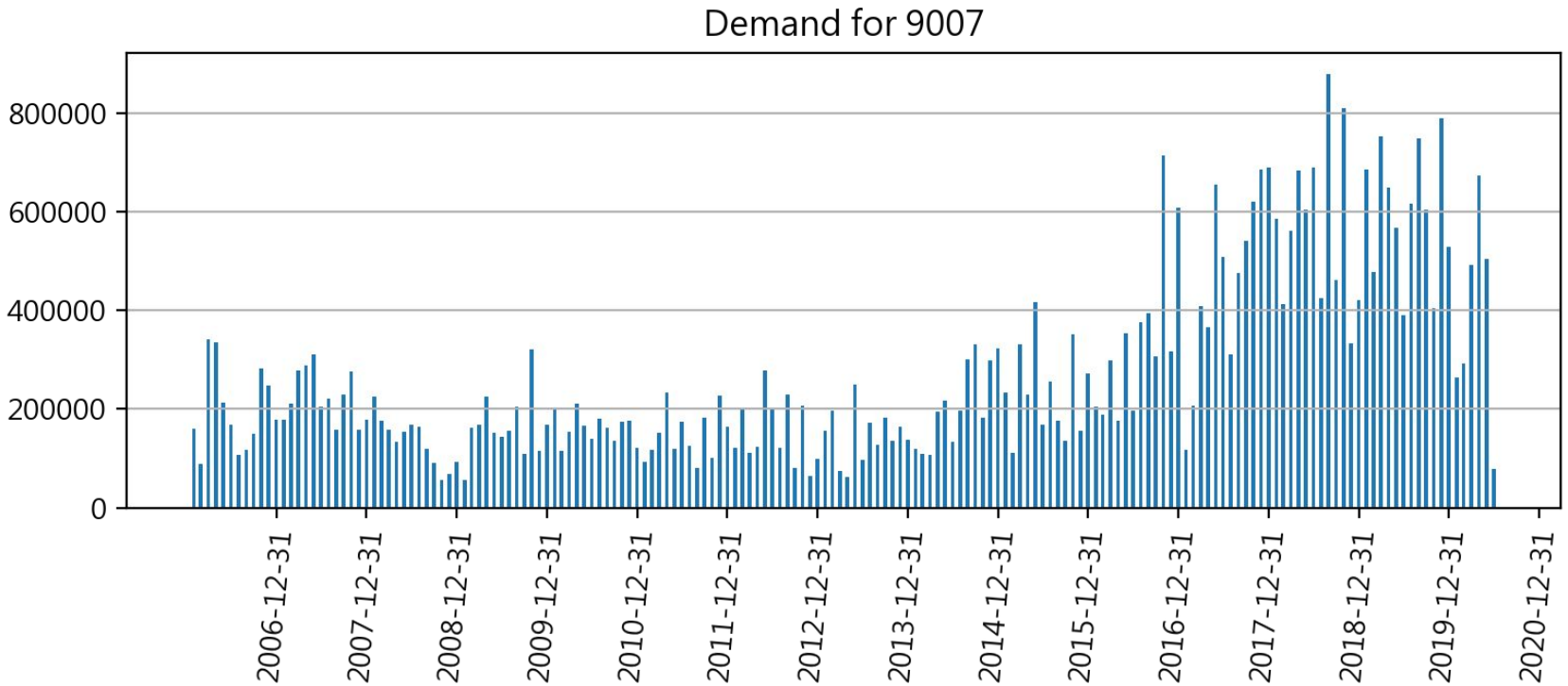
3. 數位轉型 (Digital transformation)
    - E.g., Uber replacing taxi.

Outline of the Talk

- Digital transformation (DX)
- Data analytics
- Single-variable model (單變量模型) (未考慮價格因素，只考慮前n期的需求量)
- Multi-variable model (多變量模型) (考慮價格因素及前n期的需求量)
- Who are good customers?
- FinTech ABCD
- How it works in real world?

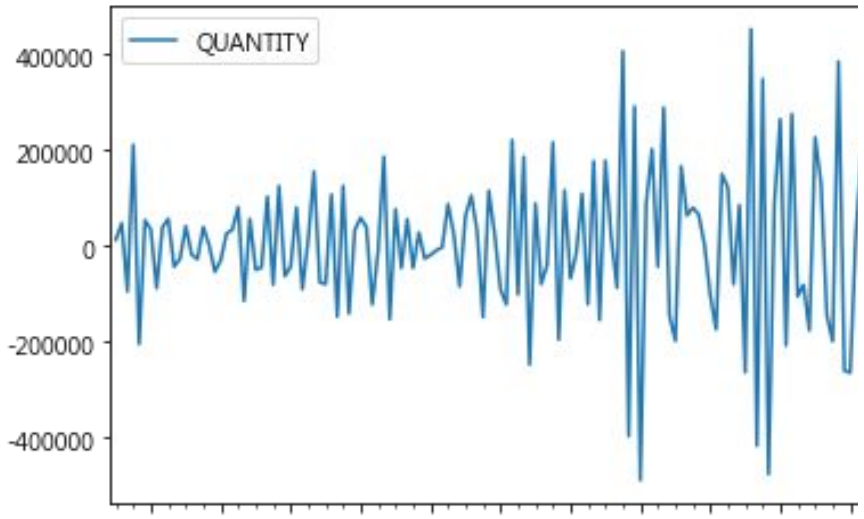
“Putting-it-all-together”, Real Case: Sales/Demand Forecasting for 9007

Data Analytics - Overview



Data Analytics: 一階差分: First-order difference equation

- 波動很大
- 標準差不固定
- 自2016波動加劇

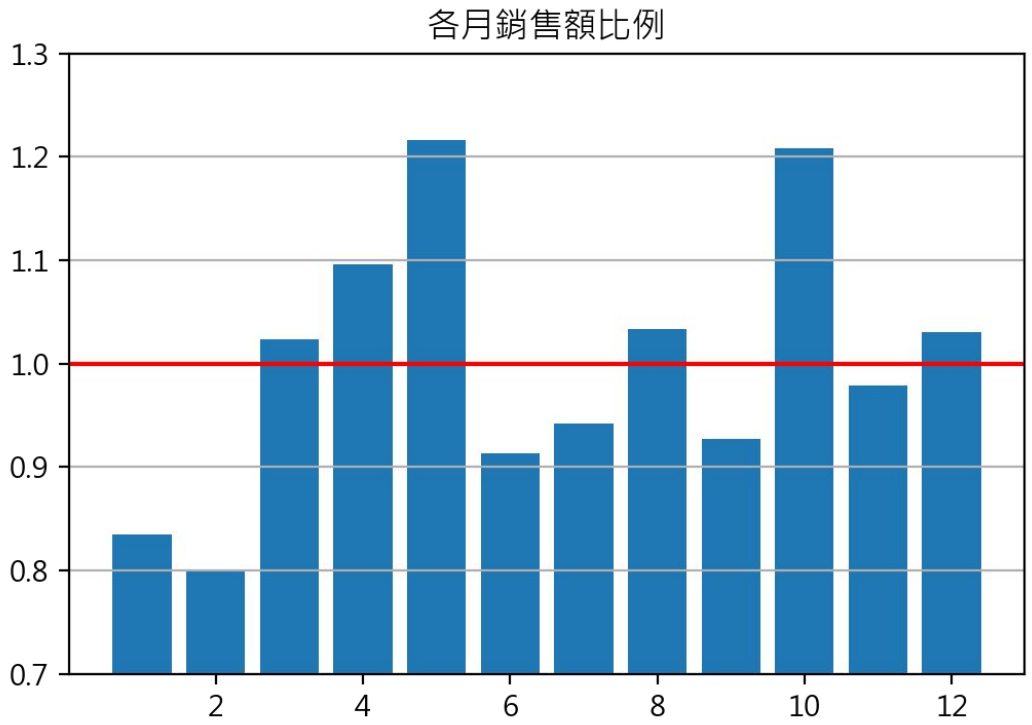


Data Analytics: 季節性分析 Seasonal Analysis

統計2006年至今的每月銷售量

旺季: 五月、十月

淡季: 一、二月、六、七月



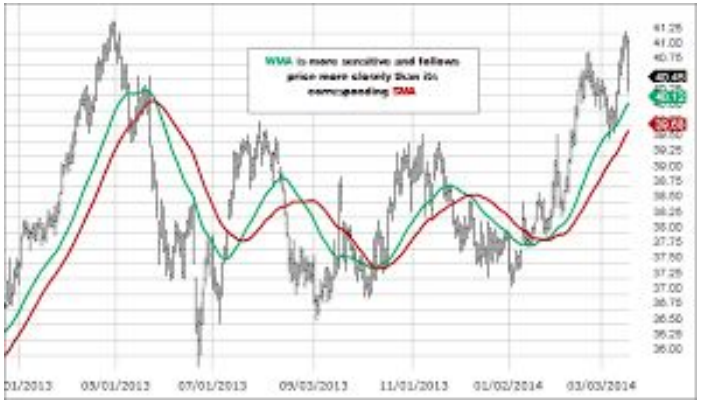
Data Analytics: 季節性分析 Seasonal Analysis

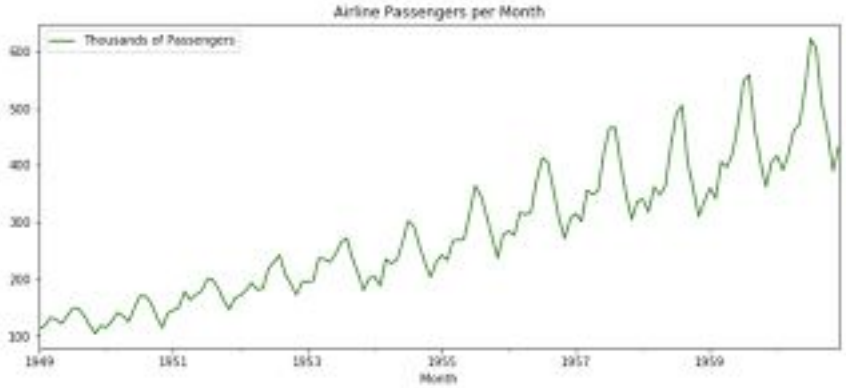
- 價格與需求量稍微正相關，但相關性很低

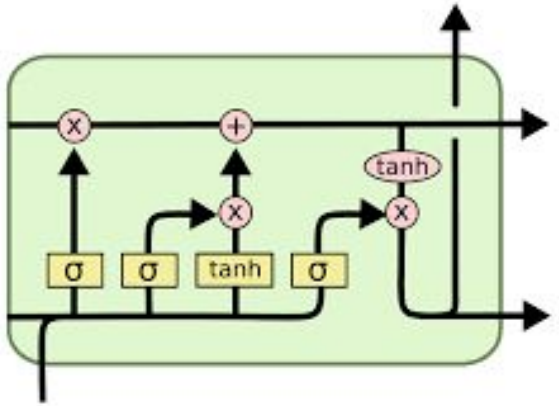
相關係數	HDPE film [CFR中國] (USD/MT)	HDPE inj [CFR中國] (USD/MT)	HDPE blow [CFR中國] (USD/MT)	HDPE yarn [CFR中國] (RMB/MT)
需求量	0.181136	0.073959	0.195248	0.054707

單變量模型 (未考慮價格因素，只考慮前n期的需求量)

- Moving Average Approach (3MA)
- Econometric Approach (ARIMA)







Moving Average Approach - 3MA

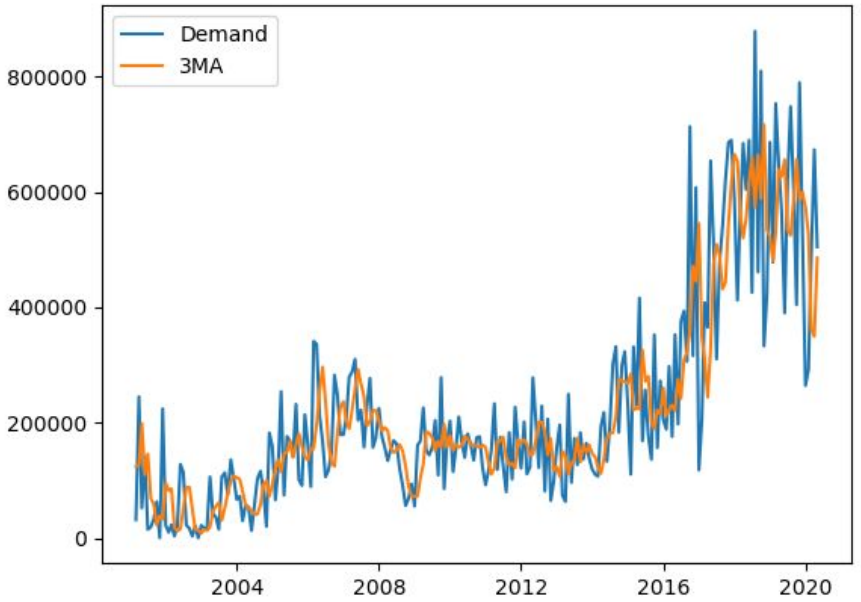
均方根誤差約為107,087

(從2001至今)

假設3MA為 500,000

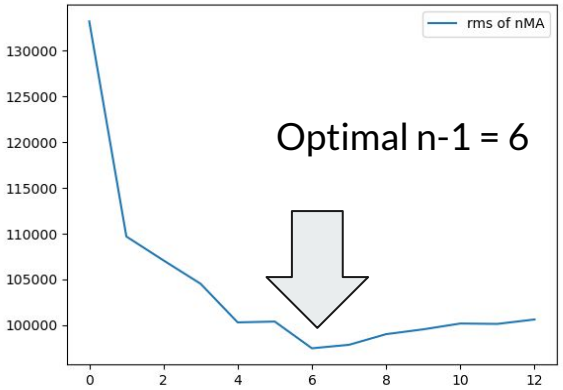
則需求量68%信心水準的信賴區間為400,000 ~ 600,000

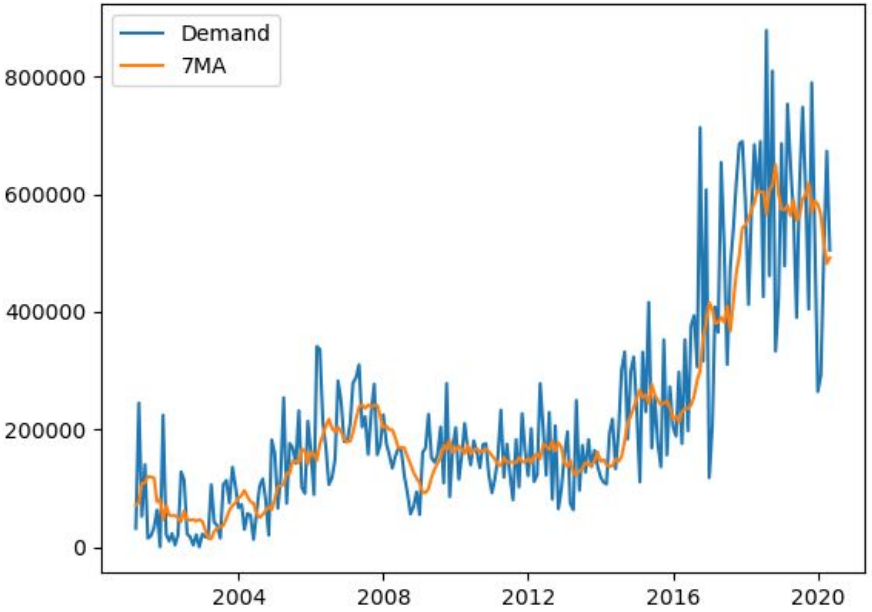
需求量95%信心水準的信賴區間為300,000 ~ 700,000



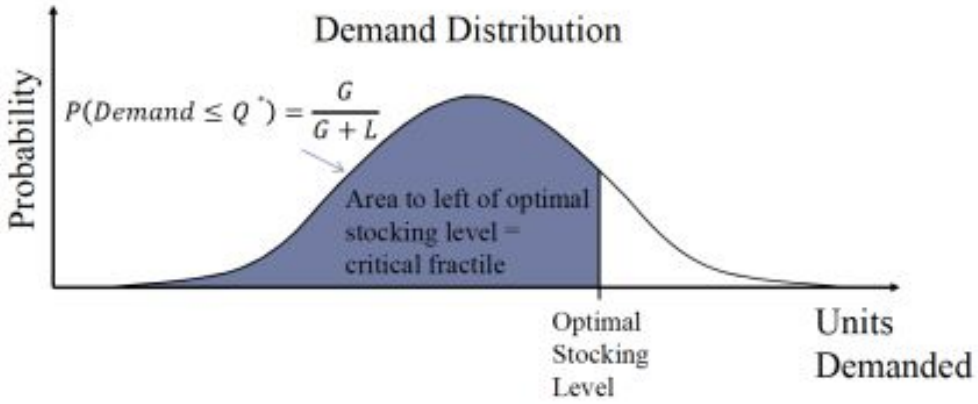
Moving Average Approach - nMA

- 使用近七個月銷量平均的優點
  - 平均會有最小的RMSE (97,484)
  - 比3MA 好了(10,000)
  - 誤差減少了4個百分點





Moving Average Approach - 7MA 之最適生產量



The critical fractile is the probability that demand is *less than or equal* to the **optimal stocking level**

假設7MA為 500,000

G = 每單位倉儲成本

L = 每單位利潤

則當月之最適生產量為

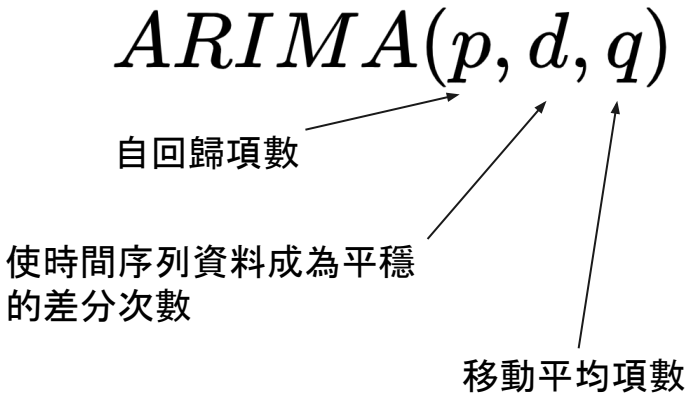
P = G/(G+L)

Q = 500,000 + 97,484 * how many sigmas

ARIMA模型 (Autoregressive Integrated Moving Average model)

一種時間序列的預測分析方法

用於非定態的資料



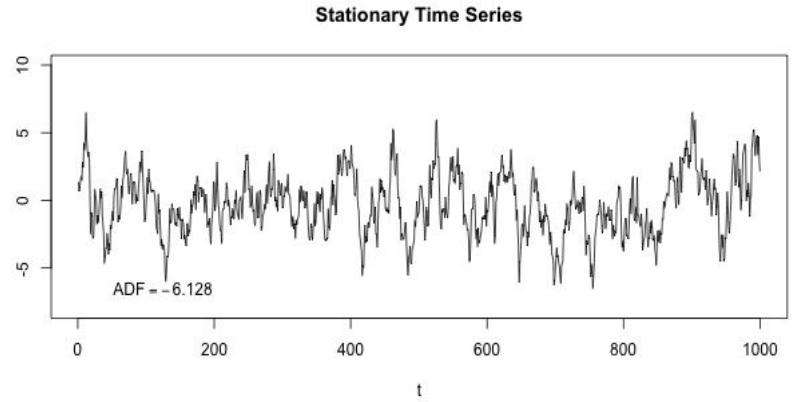
定態性

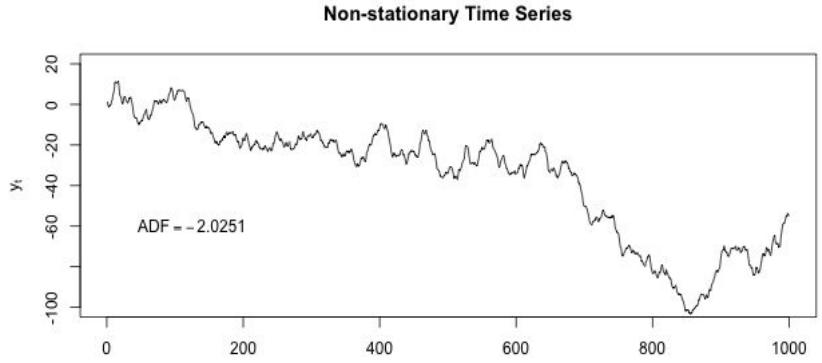
一組時間序列資料的統計特性不會隨著時間而改變

強定態性: 資料在每一個時點的機率分配完全相同並且獨立。That is, (i.i.d)

弱定態性: 僅要求資料的平均數、變異數以及自我共變異數為有限的常數項

一般通常採用弱定態的定義





https://mropengate.blogspot.com/2015/04/stationary-and-non-stationary-time.html

AR

$$y_t = \alpha + \sum_{i=1}^{p} \gamma_i y_{t-i} + \epsilon_t$$

常數項

模型的係數

誤差

描述當期的數值與前p期的數值之間的關係

用變量自身的歷史時間數據對自身進行預測

必須滿足定態性的要求

限制:

必須有定態性、自相關性，自相關係數<0.5時不宜使用

只適用與自身前期相關的現象

MA

關注誤差項的累加

能有效的消除預測中的隨機波動

$$y_t = \mu + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t$$

序列的均值

模型的係數

誤差

ARMA

AR, MA 兩個模型的結合

$$y_t = \alpha + \sum_{i=1}^{p} \gamma_i y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t$$

常數項

模型求解的係數

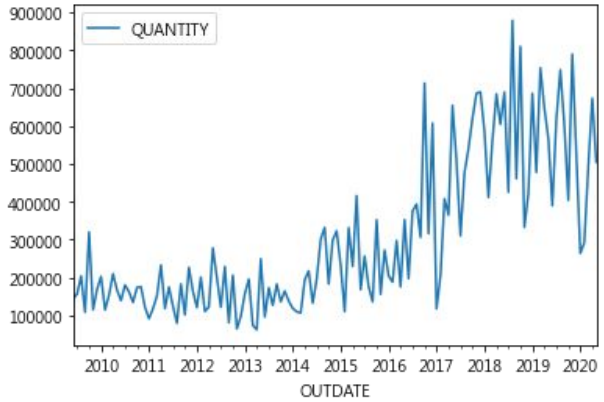
誤差

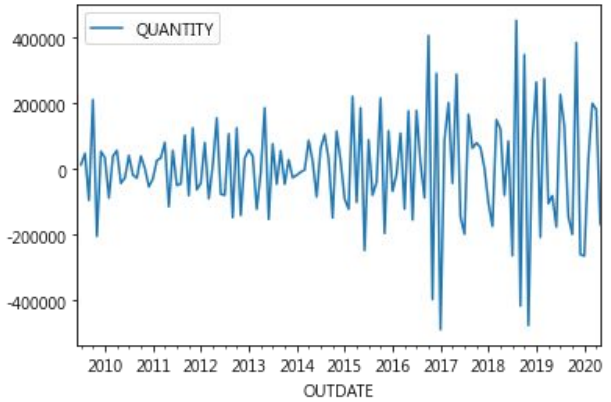
差分

用於減輕數據之間的不規律波動，使其變的平穩

一階差分: 以當期的值減掉前一期的值

二階差分: 對一階差分後的值再做一次差分





Augmented Dickey-Fuller (ADF) test

一種統計學的檢測方法，用於檢測一個有某種趨勢的時間序列的平穩性。

虛無假設為該序列不穩定

統計量越小越容易拒絕虛無假設，也就是該序列很穩定

評估模型的參數: ACF, PACF

自相關係數 (AutoCorrelation Function, ACF)

反映了同一序列在不同時序中數值之間的關聯性

$$ACF(k) = \rho_k = \frac{Cov(y_t, y_{t-k})}{Var(y_t)}$$

範圍為[-1, 1]

偏自相關係數 (Partial AutoCorrelation Function, PACF)

PACF就是ACF剔除掉其他隨機變數的干擾

評估模型的參數: ACF, PACF

模型	ACF	PACF
AR(p)	拖尾	p階後截尾
MA(q)	q階後截尾	拖尾
ARMA(p,q)	拖尾	拖尾

評估模型的參數: ACF, PACF

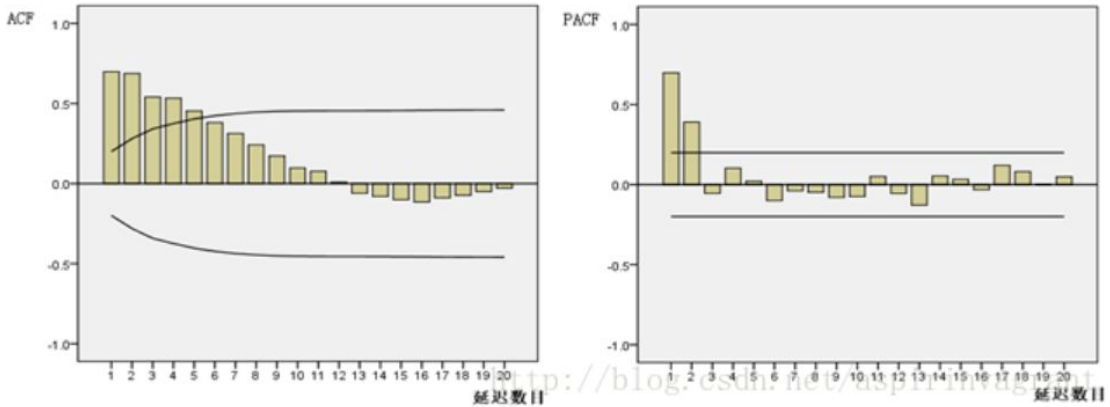
拖尾 (Tail Off): 相關係數隨落後期數遞減

截尾 (Cut Off): 在某落後期數以後相關係數皆為0

右圖呈現ACF拖尾

PACF 2階後截尾

所以使用AR(2)模型



ACF PACF

評估模型的參數: AIC

AIC (Akaike Information Criterion, AIC)

$$AIC = 2 * k - 2 * ln(L)$$

模型越簡單越好，也就是說AIC越小越好

k：模型參數個數

n：樣本數量

L：似然函數

評估模型的參數

	優點	缺點
AIC	1. 估計值較精準 2. 可以在正確配適的模型中擇優	計算量大
ACF/PACF	1. 直觀 2. 計算量小	可能出現好幾種配適的方法

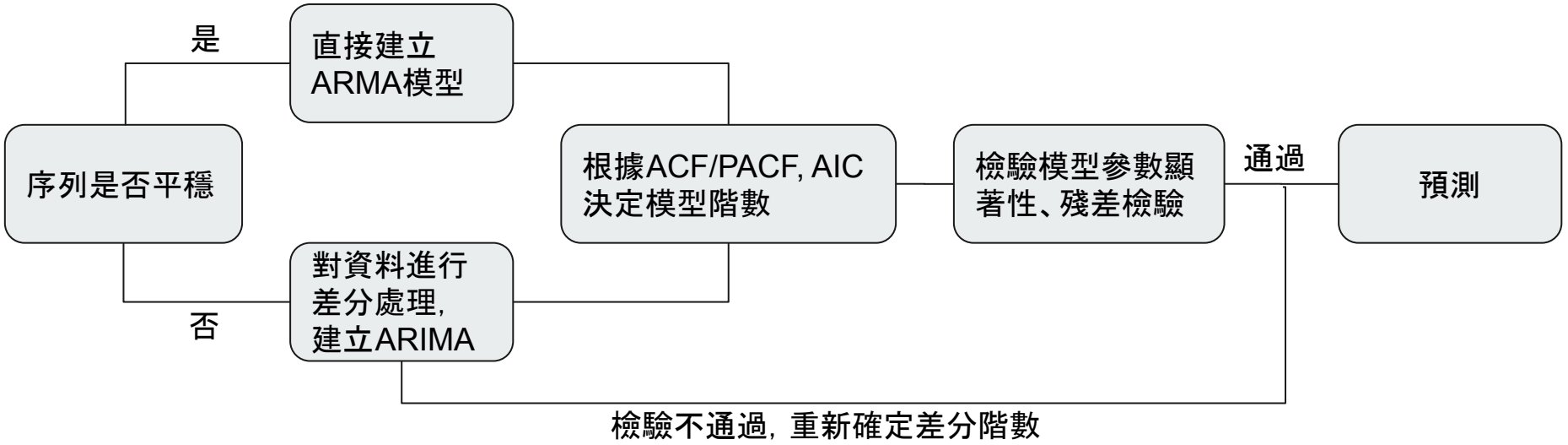
通常會混用，例如先用 ACF/PACF找出幾個適合的參數，再用AIC選擇誤差較小的

檢驗殘差 (Residual)

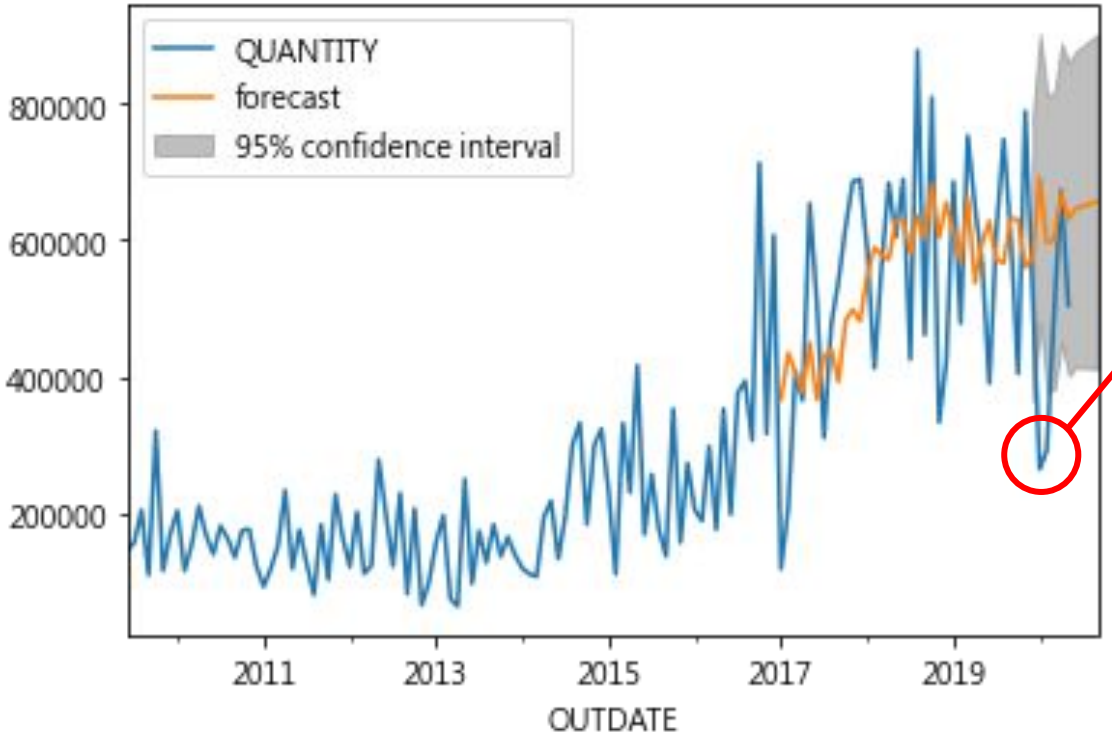
殘差: 實際觀察值與估計值之間的差

為了確保模型的參數適合，我們需要進行殘差檢驗，如果殘差符合常態分配且不自相關，代表有用的資訊都已經被提取到ARMA模型之中。

Putting it altogether: 流程



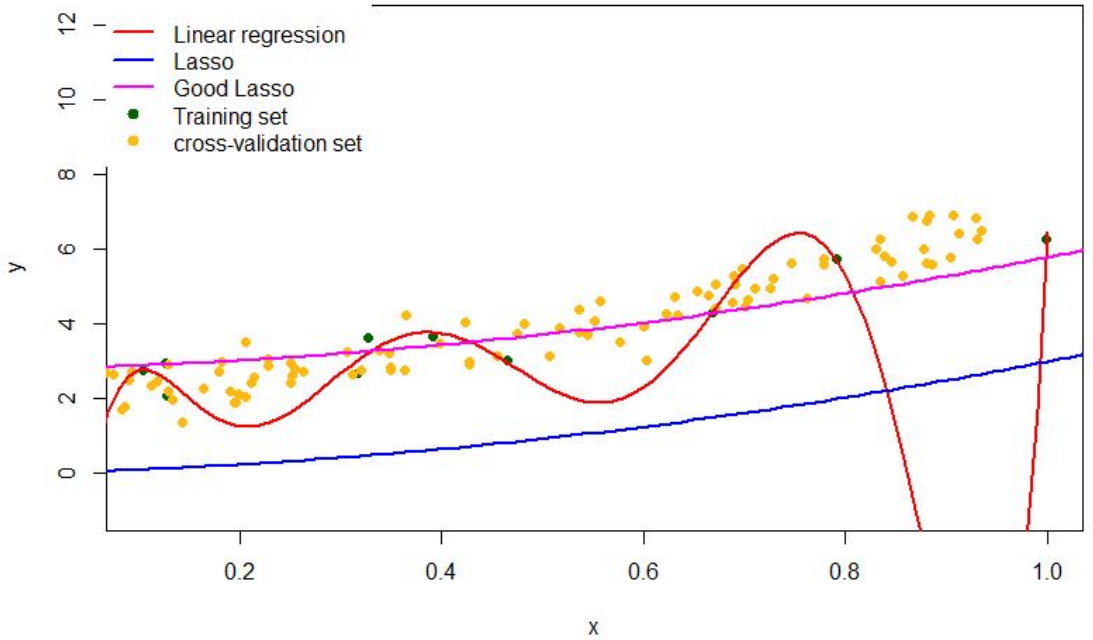
使用ARIMA(0,1,7)預測銷售量



使用過去的銷售量無法推測出的現象
猜測與COVID-19相關

多變量模型 (考慮價格因素及前n期的需求量)

- Regression Approach (Lasso/Ridge)
- Machine Learning Approach (XGBoost)



模型變數:前10期的銷售量、前3期原物料價格、前3期月均HDPE價格、一階價格差分、當月月份、指對數成長率、匯率 ...等，共63個變數。

Regression to Ridge / LASSO

- Linear Regression

  最小化誤差-->可能over fitting

  解法:L1/ L2 Regularization

- LASSO (L1)

  篩選特徵:減少模型參數

- Ridge (L2)

  平滑:降低模型複雜度

L1 Regularization

$$\text{Cost} = \sum_{i=0}^{N}(y_i - \sum_{j=0}^{M} x_{ij}W_j)^2 + \lambda \sum_{j=0}^{M} |W_j|$$

L2 Regularization

$$\text{Cost} = \sum_{i=0}^{N}(y_i - \sum_{j=0}^{M} x_{ij}W_j)^2 + \lambda \sum_{j=0}^{M} W_j^2$$

Loss function

Regularization Term

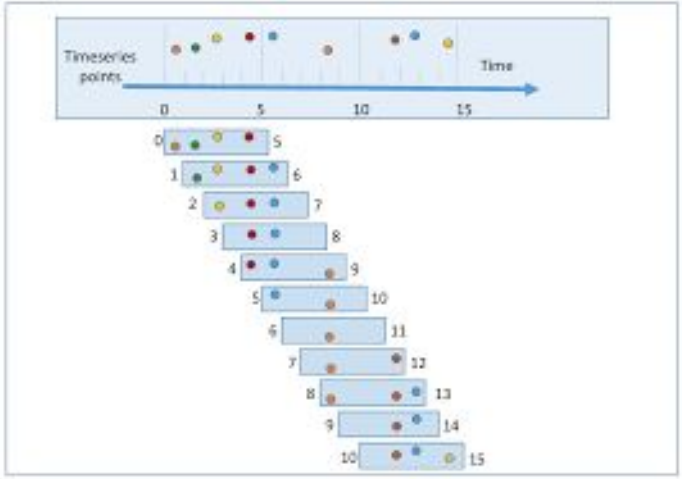
L1 vs. L2

https://www.youtube.com/watch?v=TmzzQoO8mr4

多變量模型效果評估

利用2009至2016的資料進行模型訓練，並以滾動式方法進行回測2016-2020

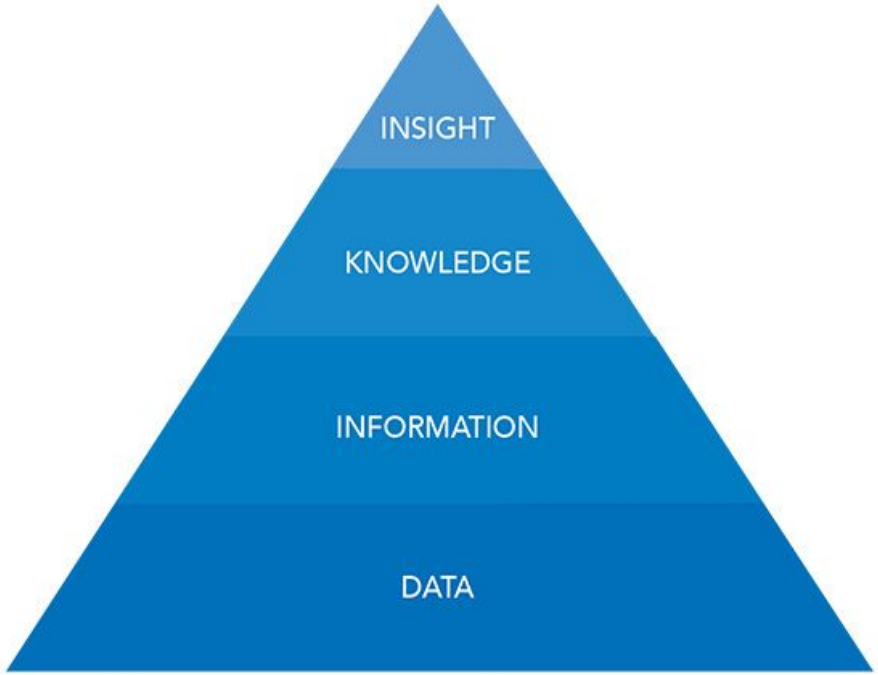
Approach/RMSE	Test Set 過去3年的均方根誤差	名次
3MA	182,549	6
7MA	**159,966**	**1**
Ridge	174,618	3
LASSO	176,574	4
XGBoost	181,464	5



FROM DATA TO INSIGHT

經LASSO挑選較有影響之變數及對應係數

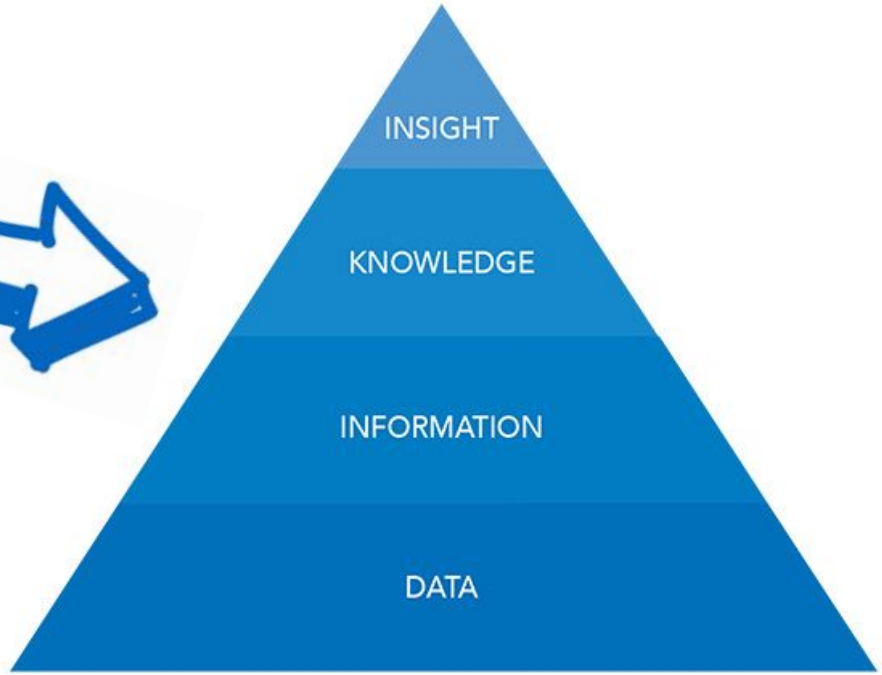
- 1 二月　　-10000
- 7 八月　　32841
- 9 十月　　98545
- t-1 前一期數量 193620
- t-4 前四期數量 213110
- A12-1 前一期的布蘭特原油　34165.781



FROM DATA TO INSIGHT

經LASSO挑選較有影響之變數及對應係數

- 1 二月　　-10000 >>>> 淡季
- 7 八月　　32841 >>>> 旺季
- 9 十月　　98545 >>>> 旺季
- t-1 前一期數量 193620
- t-4 前四期數量 213110
- A12-1 前一期的布蘭特原油　34165.781



FROM DATA TO INSIGHT

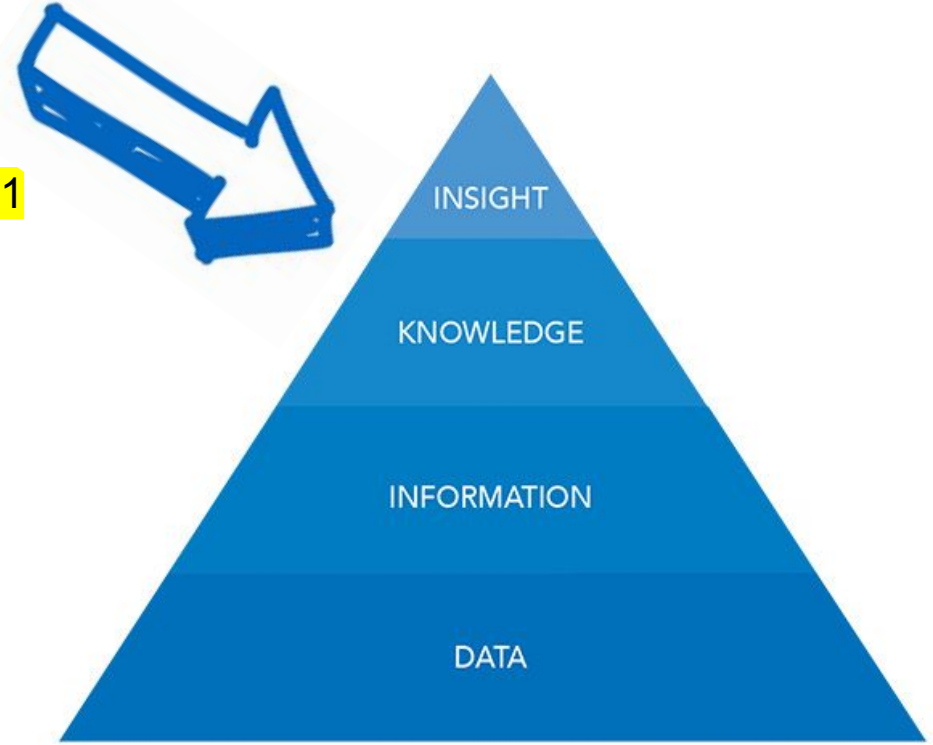
經LASSO挑選較有影響之變數及對應係數

- A12-1 前一期的布蘭特原油 34165.781

  >>> 廠商預期心理

  原物料漲價, 產品即將漲價

  趁現在趕快買



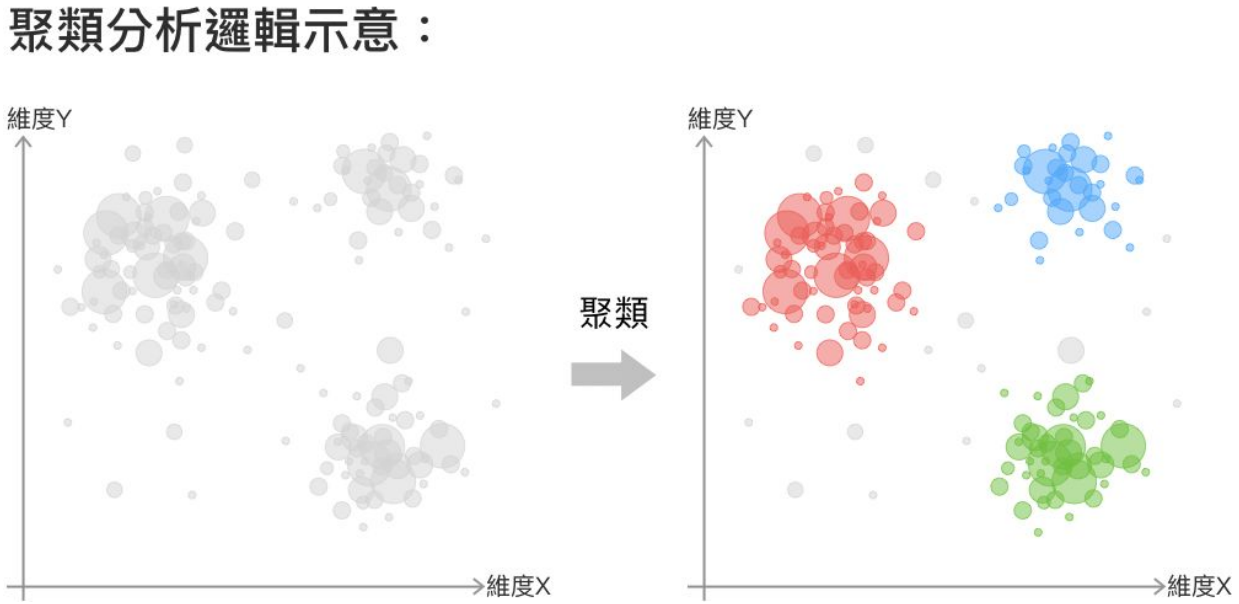
Who are good customers? Customer Clustering

將您的客戶做分群

以非監督式學習(KMEANS)

將客戶分群

- 購買量穩定的客戶
- 購買量不穩定的客戶
- 流失客戶



內銷vs外銷

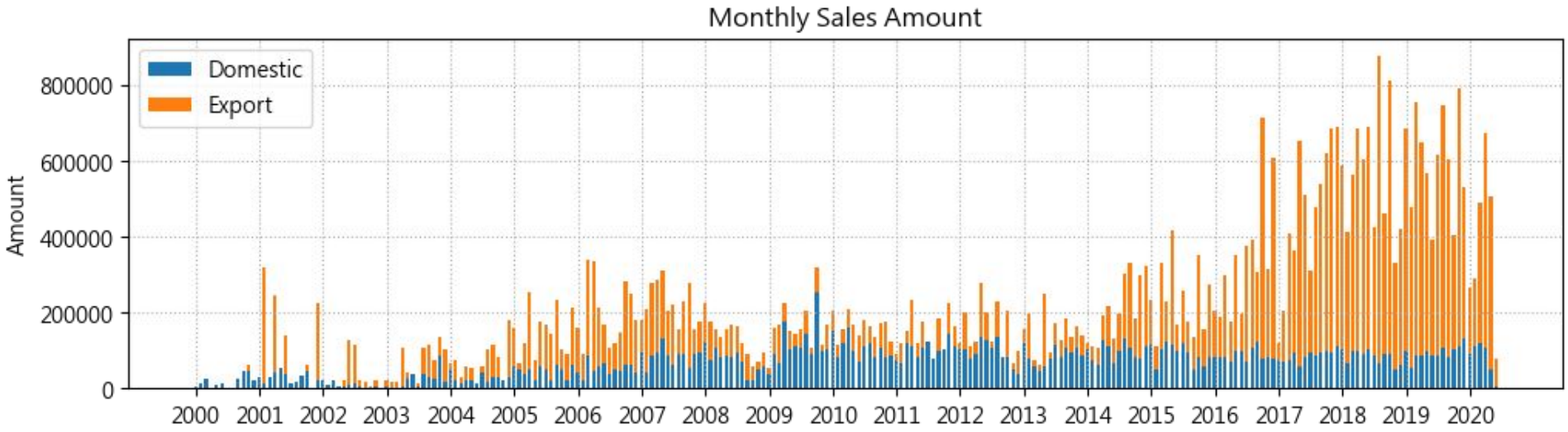
特徵:購買次數、平均購買量、歷史購買紀錄的標準差、是否為外銷。

Index	0	times	avg	std	isM	label
188	SKSY2	1	100	-10000	1	0
192	YFKSA	9	5333.33	1658.31	1	0
193	HFCY4	1	3000	-10000	1	0
194	LWKF	1	250	-10000	1	0
195	YUKF1	2	1000	0	1	0
196	ZHYH1	1	25	-10000	1	0
197	SYYH6	1	150	-10000	1	0
3	SACO1	48	15756.2	4306.39	0	1
12	LGHK	20	15300	4673.24	0	1
15	WEWU	90	15488.9	4031.46	0	1
30	YUCO	1	17000	-10000	0	1
37	YFCBO	403	15776.7	4255.35	0	1
38	UNCH	23	15521.7	4066.15	0	1
39	PECH	15	13606.7	6235.56	0	1

分類別	均單次購買量	購買量標準差	是否為外銷
class 1(不穩定)	高	高	大部份為外銷
class 2(穩定)	中	低	大部份為內銷
class 3(次數少)	低	中	都有

內銷vs外銷

從總銷售量也可以看出內外銷的波動幅度差別



Key: Asking the next set of questions:

1. 銷售量受產能限制，無法完全反應真實需求量

2. 寡佔市場銷售量受定價策略之影響

3. 影響外銷波動的因素(疫情、品質)?

Recap

- Data analytics
- 單變量模型 (未考慮價格因素，只考慮前n期的需求量)
  - Moving Average Approach (3MA)
  - Econometric Approach (ARIMA)
- 多變量模型 (考慮價格因素及前n期的需求量)
  - Regression Approach (Lasso/Ridge)
  - Machine Learning Approach (SVR)
- Who are good customers?
  - Customer Clustering

Outline of the Talk

- Digital transformation (DX)
- Data analytics
- Single-variable model (單變量模型) (未考慮價格因素，只考慮前n期的需求量)
- Multi-variable model (多變量模型) (考慮價格因素及前n期的需求量)
- Who are good customers?
- FinTech ABCD
- How it works in real world?

# 金融科技 = ABCD [Deloitte 2017], where
A = AI, B = Blockchain, C=Cloud, D = Data

Prof. 廖世偉, National Taiwan University
liao@csie.ntu.edu.tw

We try to 註明出處 of photos and charts etc., which are owned by their original authors. 如有未盡／未及處, they are still owned by their respective authors.

https://www.middlebury.edu/institute/people/moyara-ruehsen

MENU

DEGREES

Moyara Ruehsen

Associate Professor, Financial Crime Intelligence

Professor Moyara Ruehsen oversees the Financial Crime Management program, which offers a specialization for master's degree candidates as well as a stand-alone certificate for mid-career professionals. She has published articles and book chapters on a variety of topics related to threat finance and is a Certified Anti-Money Laundering Specialist and a Certified Financial Crime Specialist. Professor Ruehsen teaches financial crime-related courses on a variety of topics including money laundering, trade-based financial crime, corruption, proliferation financing, terrorist financing and cyber-enabled financial crime.

Before coming to MIIS, Professor Ruehsen received three graduate degrees from Johns Hopkins University, and spent a post-doc year at the University of California, Berkeley, to study international organized crime.Her regional areas of interest include South Asia, Southeast Asia and the Middle East, where she spent a year as a Fulbright scholar.

Professor Ruehsen also consults for the U.S. government, multilateral organizations and the private sector.She served for several years on the Editorial Advisory Board of *Money Laundering Alert,* and the Middle East Task Force of ACAMS (Association of Certified Anti-Money Laundering Specialists).

Virtual Asset RegTech

https://www.youtube.com/watch?v=0rsOLJ1RmqA

Extension: 金融科技 @ 台大進修推廣學院 (廖世偉, 萬幼筠, 林建隆, 朱師右, 江炯聰)

https://www.ntuspecs.ntu.edu.tw/specs/TC/classStudyListContent.aspx?id=1693&Chk=af288a16-aff5-4e50-a42b-29787bb2c1c7&cid=22&cchk=

Next: Fintech ABCD, Vision, Challenges

- Fintech (F) = ABCD: AI, Blockchain, Cloud, Data [Deloitte 2017]:
  - 台大金融科技4部曲: 數位轉型出寬客 (F), 黑客 (CD), 極客 (A), 極客 (B): 共4門課
  - Example: This class in October: 普世普惠金融 (WoFi) 的量化交易 (Quantitative Trading) 比賽
    i. WoFi is our VISION: World Financial Inclusion (WoFi). See next 11 slides.
- Fintech's 4 Challenges
  - 數位化 → 數位優化 → 數位轉型: Challenging. Sol. Must understand 本質!
  - Fintech 跨領域: Social Science vs. Science/Engineering Schools even biology!
    i. 金融人不懂科技，法律；科技人不懂金融，法律。Even worse, cultures are different:
    ii. Social science is physics envy: Many hypotheses such as EMH, AMH, …
    iii. Worst: Unknown unknowns → 高階決策 instead of 戰逃反應 → 寬客。
    iv. Summary: Those quotients can be orthogonal: IQ, EQ (AQ), SQ, FQ. Need to be 寬客。
  - 大家FQ低: Sol. This class will boost your FQ (Finance Quotient), which requires both IQ and EQ (AQ)!
  - Traditional finance doesn't want to be disrupted: 如何推出 Fintech 產品?

Despite challenges above,

Fintech thrives (勢不可擋):

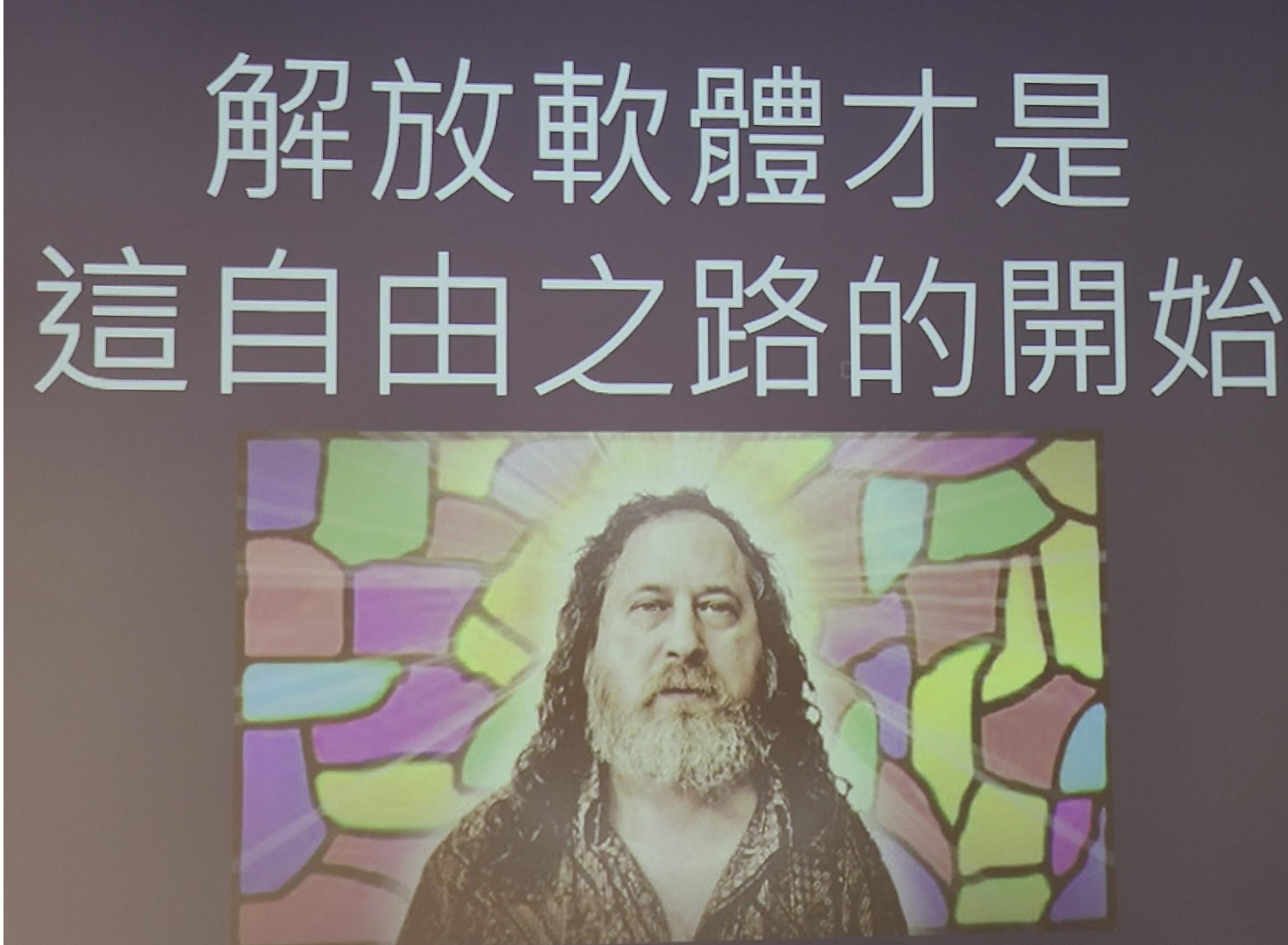
This class is about “Sexual Enlightenment” in finance

VISION: World Financial Inclusion (WoFi)

After COVID-19 lockdown, finally, Software Freedom Day, 2020/9/19:

Keynote speech: “~35 years ago when I was at NTU, RMS totally enlightened me. Today it’s my turn to enlighten you about Finance Liberation in digital society.

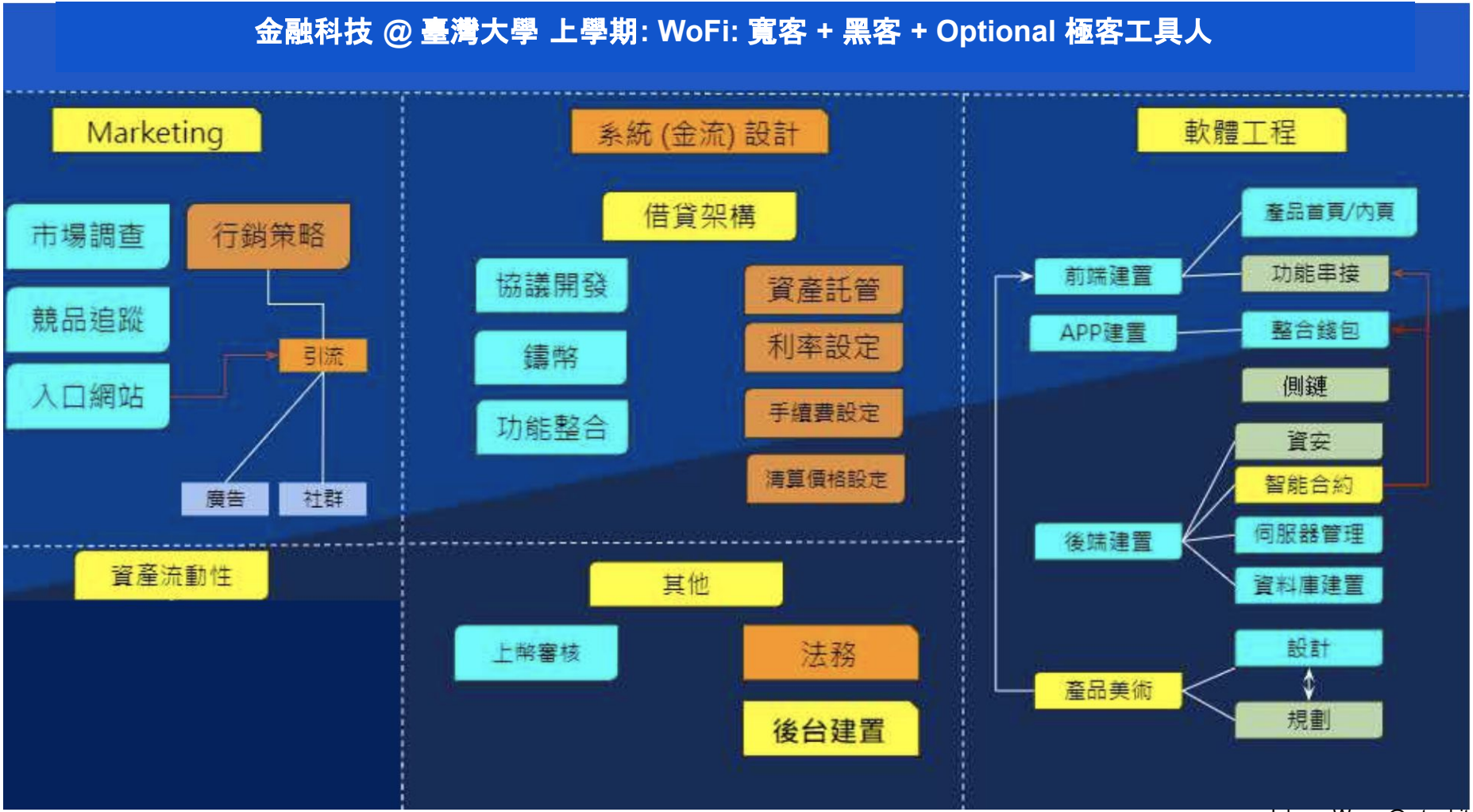
It’s time.”



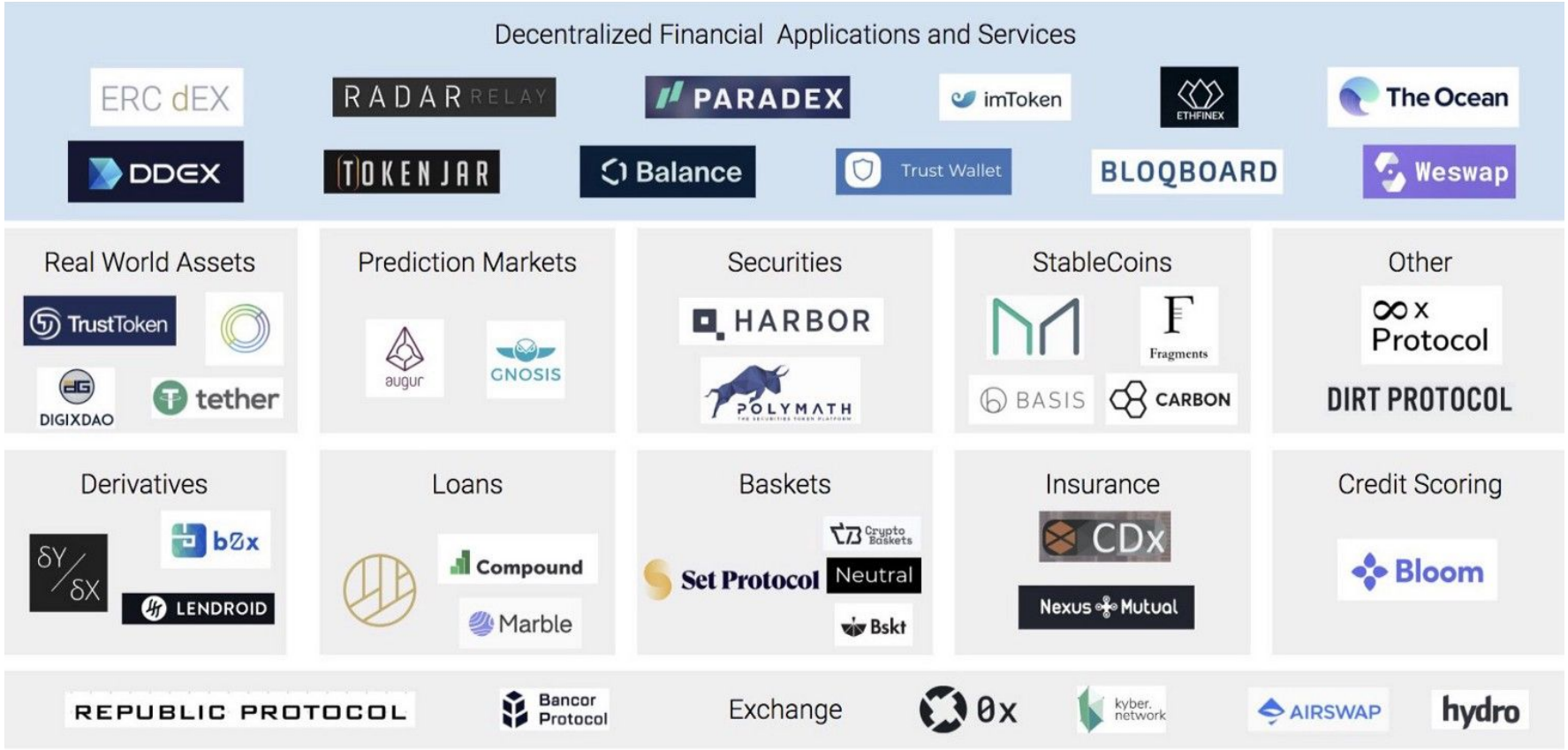
Richard Stallman (RMS) @ MIT (1953 -):
Liberating software leads the road to freedom

- “When you buy a house, you won’t tolerate having to ask a tyrant for permission to fix something in the house. He can screw you any time.” How about software?
- RMS established GNU project in 1983
- RMS established FSF (Free Software Foundation), aka Foundation for Software Freedom
  - Free as in liber, not free as in beer or in “free food alert.”
- ~40 years later, now you have Linux, Android phones everywhere, and GCC, thanking to RMS.
- Now it’s about time to liberating finance!
  - Liberating finance (a.k.a. WoFi, 普世普惠金融) is a key milestone in digital society freedom.
  - Just like sex liberation is a key milestone in traditional society freedom.
    - 領主不再擁有你的初夜權 (Right of the lord)
    - Just like: “領主” (a single bank) should not monopolize the money in the digital society.
    - Fintech: Using software (technology) to liberate finance in digital society.

金融科技 @ 臺灣大學 上學期: WoFi: 寬客 + 黑客 + Optional 極客工具人

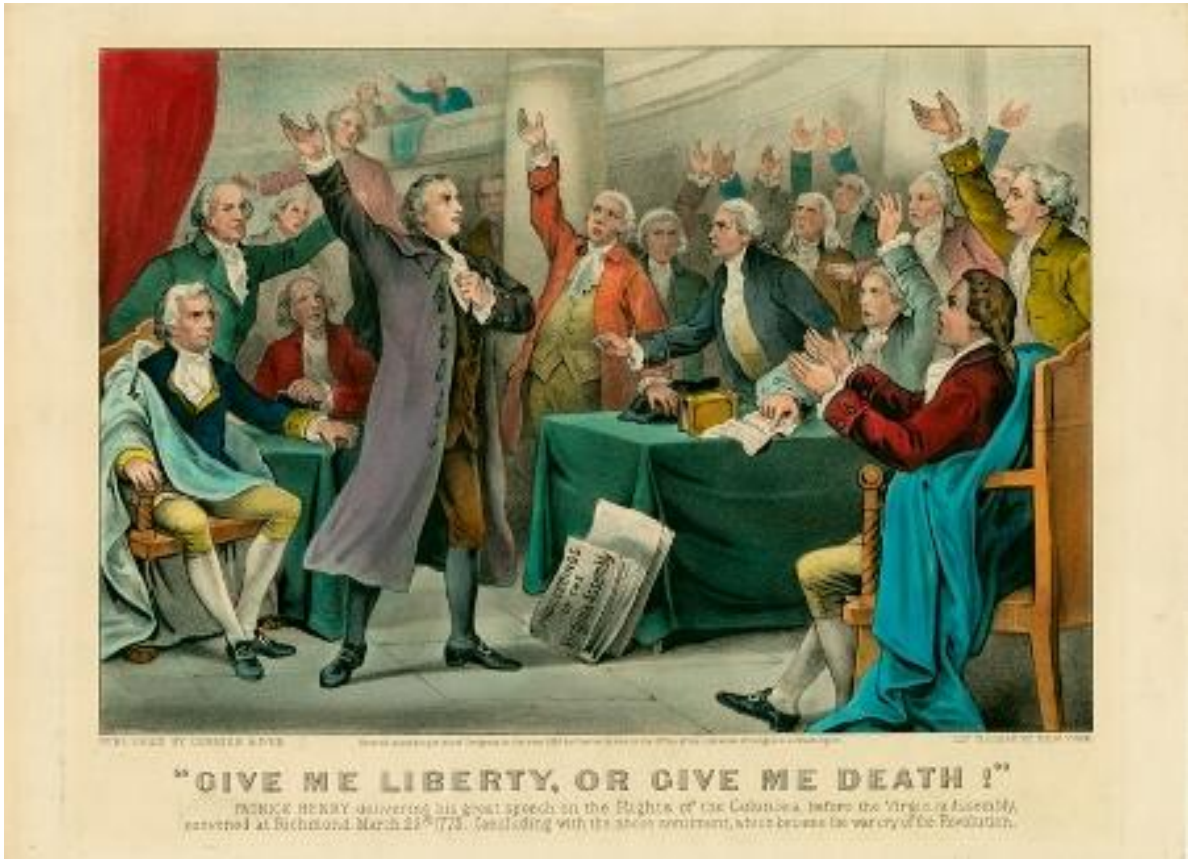


source: Johnny Wang @ star bit


Decentralized Financial Applications and Services
ERC dEX
RADAR RELAY
PARADEX
imToken
ETHFINEX
The Ocean
DDEX
TOKENJAR
Balance
Trust Wallet
BLOQBOARD
Weswap
Real World Assets
TrustToken
DIGIXDAO
tether
Prediction Markets
augur
GNOSIS
Securities
HARBOR
POLYMATH
StableCoins
Fragments
BASIS
CARBON
Other
0x Protocol
DIRT PROTOCOL
Derivatives
bZx
LENDROID
Loans
Compound
Marble
Baskets
Crypto Baskets
Set Protocol
Neutral
Bskt
Insurance
CDx
Nexus Mutual
Credit Scoring
Bloom
REPUBLIC PROTOCOL
Bancor Protocol
Exchange
0x
kyber network
AIRSWAP
hydro

Just like “性教育”: 鴕鳥心態 won’t work: Sex will not be controlled by lords forever

Freedom or Death:



A Monetary History
of the
United States
1867-1960

MILTON FRIEDMAN
ANNA JACOBSON SCHWARTZ

A STUDY BY THE
NATIONAL BUREAU OF ECONOMIC RESEARCH, NEW YORK

PUBLISHED BY
PRINCETON UNIVERSITY PRESS, PRINCETON
1963

用 Python 理財：打造小資族選股策略

1 個章節，33 段單元、5 項作業，共 470 分鐘

當前單元 試看

▶ 課程介紹 03:08

第 1 章 本課程不分章節

單元 1 – 程式環境架設｜盤古開天！ 12:53

單元 2 – 程式基本教學1｜股票買賣簡單計算 15:11

單元 3 – 程式基本教學2｜Python的 12:10

7,308 位同學 | 470 分鐘 | 問題討論 | 作業批改

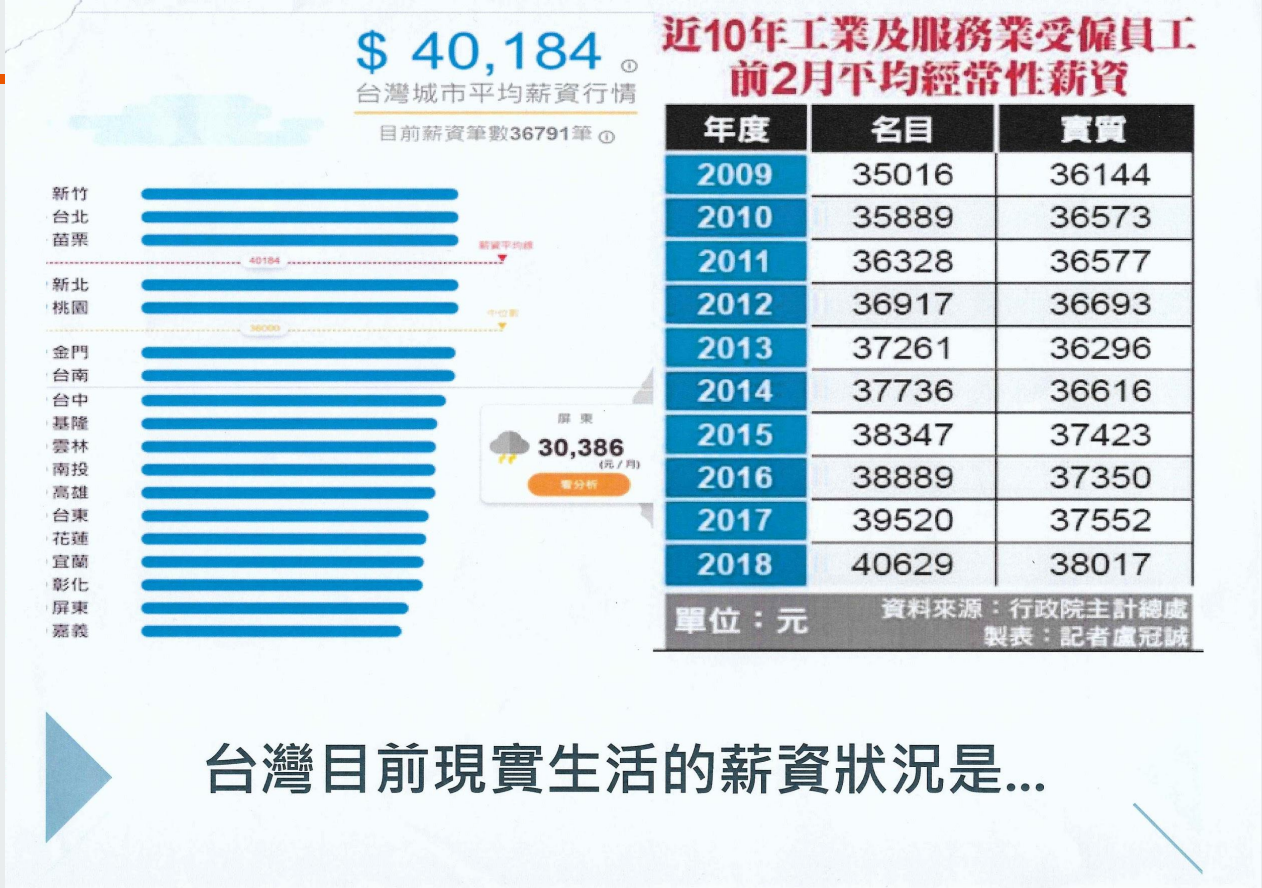
大賣課優惠折抵活動進行中！

NT$3,200

★★★★★ (563) 贈送課程 收藏

馬上購買

Back-to-First-Principle: 金錢 != 財富. First:



年度	名目	實質
2009	35016	36144
2010	35889	36573
2011	36328	36577
2012	36917	36693
2013	37261	36296
2014	37736	36616
2015	38347	37423
2016	38889	37350
2017	39520	37552
2018	40629	38017

台灣目前現實生活的薪資狀況是...

Many Students Dislike Fatcat:

每個月薪資NT$50,000,
40年不失業
年終2個月
等於年薪NT$700,000.

努力工作40年
可以獲得70萬*40年 =
2800萬。
+300萬(勞退)
= 3100萬

= 3100萬

想拼湊一個基礎現實的人生需要多少花費？

買房1500萬、40年4台國產車各100萬、每個月孝親費2萬20年、生兩個孩子月開銷2萬養到成人、夫妻兩人每月生活花費2萬元就好...

平凡生活需要 = 4820萬

優質的收入，卻養不起最普通的人生。

→ 寬客的逆襲 and 世代差異

Total Value Locked (USD)

$9.27B

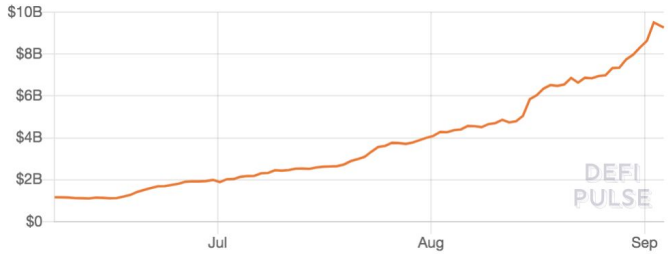
Uniswap Dominance

19.13%

Total Value Locked (USD) in DeFi

TVL (USD) | ETH | BTC

All | 1 Year | 90 Day | 30 Day



Long or short assets with up to 20x leverage and lower slippage with Perpetual Protocol. Learn more

DEFI PULSE DATA

1000 Free Credits per MONTH!

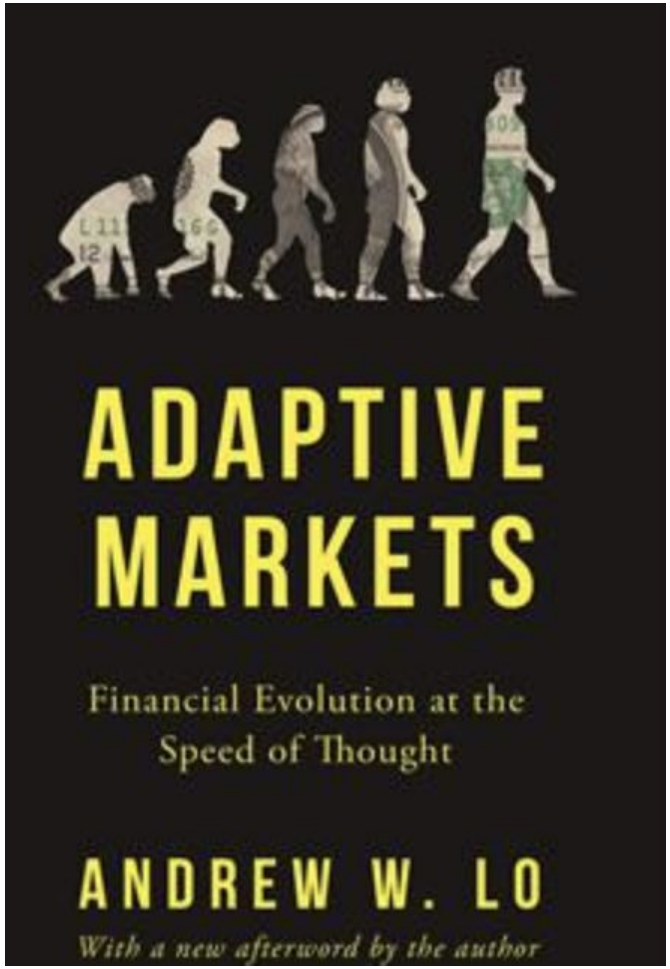
POOL OWNER DATA | MONEY FLOWS | DAPP USER COUNTS | ARBITRAGE TRANSACTIONS | DEFI INTEREST RATES | AMM TRADE HISTORIES | FAST GAS PRICES | TOTAL VALUE LOCKED | DAPP

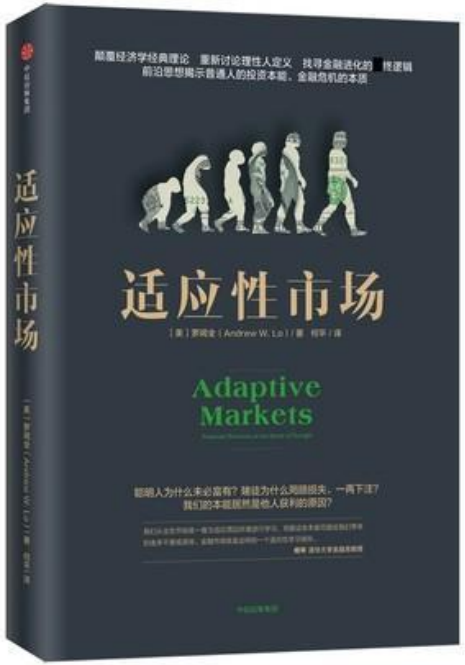
ALL | LENDING | DEXES | DERIVATIVES | PAYMENTS | ASSETS

DEFI PULSE	Name	Chain	Category	Locked (USD) ▼	1 Day %
1.	Uniswap	Ethereum	DEXes	$1.77B	8.45%
2.	Maker	Ethereum	Lending	$1.57B	-0.09%
3.	Aave	Ethereum	Lending	$1.47B	-2.30%
4.	Curve Finance	Ethereum	DEXes	$1.19B	10.77%
5.	yearn.finance	Ethereum	Assets	$952.9M	17.69%
6.	Synthetix	Ethereum	Derivatives	$862.9M	-5.90%

Next: Fintech ABCD, Vision, Challenges

- Fintech (F) = ABCD: AI, Blockchain, Cloud, Data [Deloitte 2017]:
  - 台大金融科技4部曲: 數位轉型出寬客 (F), 黑客 (CD), 極客 (A), 極客 (B)
  - Example: This class in October: 普世普惠金融 (WoFi) 的量化交易 (Quantitative Trading) 比賽
    i. WoFi is our VISION: World Financial Inclusion (WoFi). See next 11 slides.
- Fintech's 4 Challenges
  - 數位化 → 數位優化 → 數位轉型: Challenging. Sol. Must understand 本質!
  - Fintech 跨領域: Social Science vs. Science/Engineering Schools even biology!
    i. 金融人不懂科技，法律；科技人不懂金融，法律。Even worse, cultures are different:
    ii. Social science is physics envy: Many hypotheses such as EMH, AMH, …
    iii. Worst: Unknown unknowns → 高階決策 instead of 戰逃反應 → 寬客。
    iv. Summary: Those quotients can be orthogonal: IQ, EQ (AQ), SQ, FQ. Need to be 寬客。
  - 大家FQ低: Sol. This class will boost your FQ (Finance Quotient), which requires both IQ and EQ (AQ)!
  - Traditional finance doesn't want to be disrupted: 如何推出 Fintech 產品?


ADAPTIVE
MARKETS
Financial Evolution at the
Speed of Thought
ANDREW W. LO
With a new afterword by the author


适应性市场
适应性市场
Adaptive
Markets

Adaptive Markets: IQ, FQ, SQ
Intelligence, Finance, Sex Quotients

第1章　現在我們是否都是經濟人？
悲劇與群眾的智慧
漫步歷史
效率市場之誕生
透視效率市場
所謂的理性預期
效率市場之應用

第2章　如果你那麼聰明，為什麼卻沒錢？
否定隨機漫步論
風險vs.不確定性和艾斯伯格矛盾
贏的感覺雖然爽，輸的傷害更痛
無限注德州撲克、流氓交易員與監理官員
機率對應與三月瘋
人類作為預測機器
打倒一個理論要靠另一個理論
文化衝擊

第3章　如果你那麼有錢，為什麼卻不聰明？
探索人腦的運作
神經科學的顯微鏡
恐懼
痛楚
愉悅與貪婪
測量交易員的生理反應
優秀交易員的素質
以神經通貨思考金錢
我全部都要，現在就要

Efficient Market vs. Behaviorist

- Efficient Market Hypothesis (EMH): Eugene Fama @1970
- Behaviorist:
  - 炒作是人的天性？
    - Driven by animal spirit (fear & greed)
  - People behave irrationally.
  - Traditional finance framework is flawed
  - Not wrong, but incomplete (physics envy)
  - Stable environment ⇒ stable financial policies (EMH)
  - Dynamic environment ⇒ dynamic financial policies (AMH)
  - The current environment is highly dynamic
  - We must adapt to changing market conditions
  - “It’s the economy, stupid” → “It’s the environment, stupid”

- **Adaptive Markets Hypothesis is a framework for post-COVID financial market dynamics**

EMH vs. AMH (Adaptive Markets Hypothesis)

“Nothing makes sense in biology except in the light of evolution,” Dobzhansky (1973)

“Nothing makes sense in the financial industry except in the light of the Adaptive Markets Hypothesis,” Lo (2017)

1. **Individuals act in their own self-interest**
2. **Individuals make mistakes (“satisfice”)**
3. **Individuals learn and adapt (heuristics)**
4. **Competition drives adaptation and innovation**
5. **Evolution determines market dynamics**

A New Finance Paradigm is Emerging:

Adaptive Markets

- **Long/short strategies (EMH: Long-only constraints)**
- **Diversify across more asset classes and strategies**
- **Passive transparent indexes (EMH: Market-cap-weighted indexes)**
- **Manage risk via active volatility scaling algorithms**
  - **(EMH: via asset allocation)**
- **Alphas ⇒ multiple betas**
- **Markets are adaptive vs. Markets are efficient**
  - **“In the long run we’re all dead,” but make sure the short run doesn’t kill you first**
  - **COVID-19 era: Build 水位**

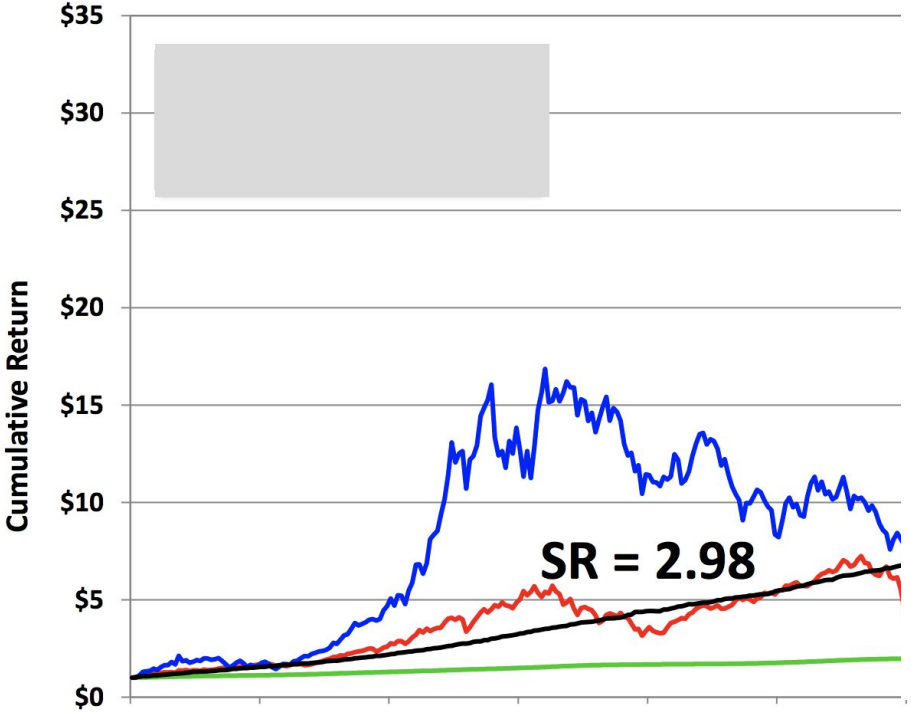
EMH Relies on Wisdom of Crowds (Independent Crowds!)

$$
\begin{aligned}
x_i &= x + \epsilon_i \\
\overline{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
&= x + \frac{1}{n}\sum_{i=1}^{n} \epsilon_i \\
&\approx x
\end{aligned}
$$

0

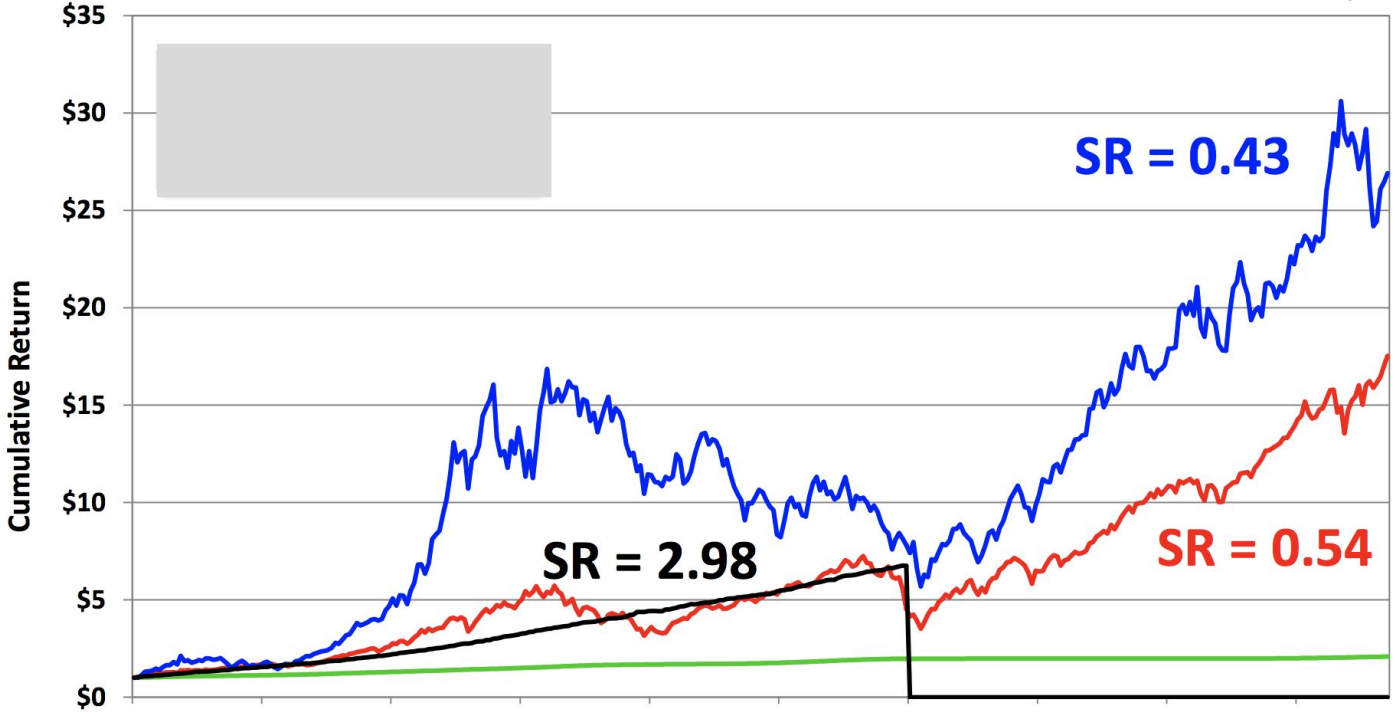
FQ: What do you want?

$$\text{Sharpe Ratio} \equiv \frac{\mathrm{E}[R] - R_f}{\mathrm{SD}[R]}$$



FQ: What do you want?

$$\text{Sharpe Ratio} \equiv \frac{\mathrm{E}[R] - R_f}{\mathrm{SD}[R]}$$



Fairfield Sentry Fund: “歸零幣” after 2008

- Black: Fairfield Sentry is a feeder fund to Bernard Madoff
  - Bernie is serving 150-year sentence
  - 削峰填谷
    i. Sharpe Ratio = 2.98: Is it too good to be true?
- Blue: Pfizer
- Red: Composite index fund

What Do Investors Fear?

Urn A contains 100 balls:

- 50 red, 50 black
- Pick a color, then draw a ball
- If you draw your color, $10,000 prize
- Which color would you prefer?
- How much would you pay to play?

What Do Investors Fear?

Urn B contains 100 balls:

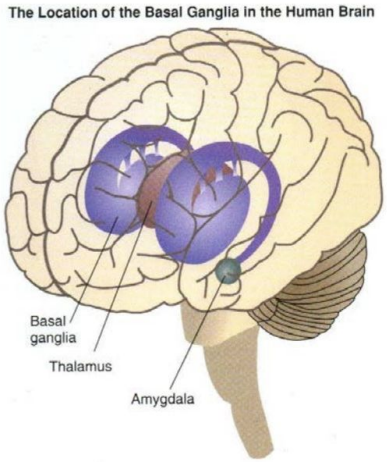
- Unknown proportion of black and/or red balls
- Pick a color, then draw a ball
- If you draw your color, $10,000 prize
- Which color would you prefer?
- How much would you pay to play?

The Unknown Unknowns

Wisdom of Crowd vs. Madness of Mobs

The Neuroscience of Fear

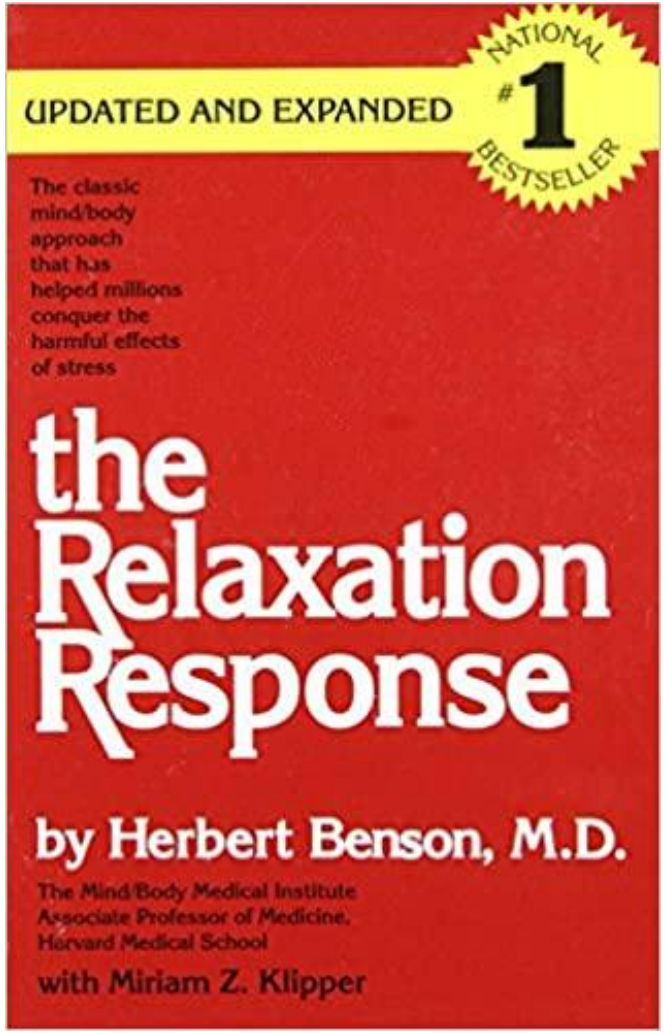
The Amygdala and "Fight or Flight"



- Increased breathing, heart rate, blood pressure
- Constriction of blood vessels to certain body parts, including **reduced blood flow to prefrontal cortex**
- Dilation of blood vessels to muscles, pupils
- Release of nutrients (fat, glucose); slowed digestion
- Reduced hearing, tunnel vision, accelerated reflexes
- Inhibited hunger and sex drive

戰逃反應 vs. The Relaxation Response

- To counteract “Fight or Flight Response”, Benson (1975) proposes TM (Transcendental Meditation)
  - Repetition of word or phrase; mantra; focus on breath; let go of other thoughts
  - See http://www.massgeneral.org/bhi/basics/rr.aspx
- Mindfulness:
  - 正念冥想
- Stress reduction
  - Exercise, yoga, extra sleep, comedy, reduce caffeine, declutter your home, etc.


UPDATED AND EXPANDED
NATIONAL
#1
BESTSELLER
The classic mind/body approach that has helped millions conquer the harmful effects of stress
the Relaxation Response
by Herbert Benson, M.D.
The Mind/Body Medical Institute
Associate Professor of Medicine,
Harvard Medical School
with Miriam Z. Klipper

How To Deal With

- Yoga
- Tai Chi
- Mindfulness

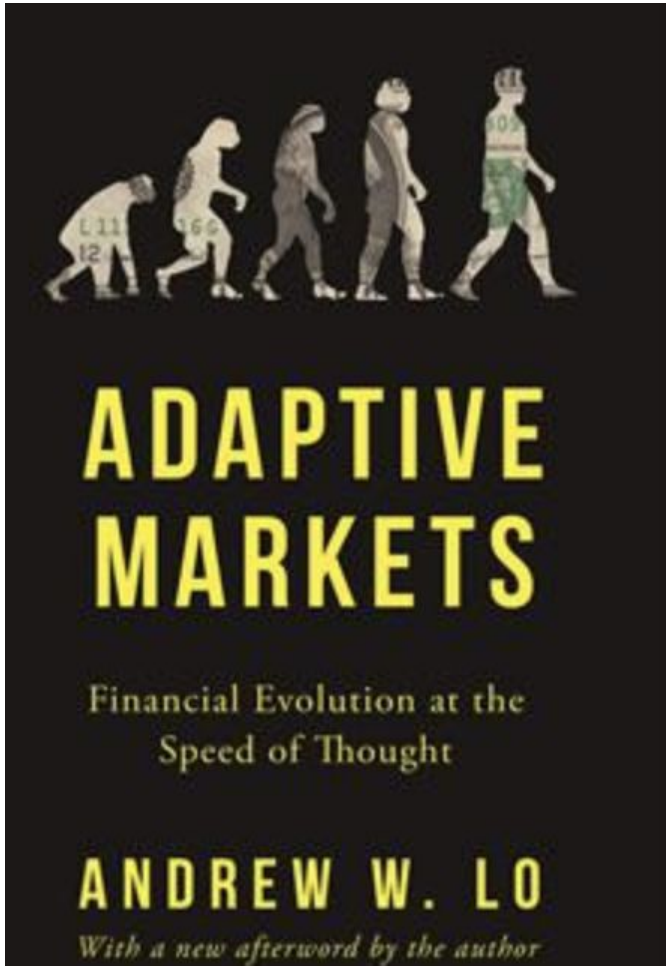
Will you commit to 15 minutes a day to one of these activities?

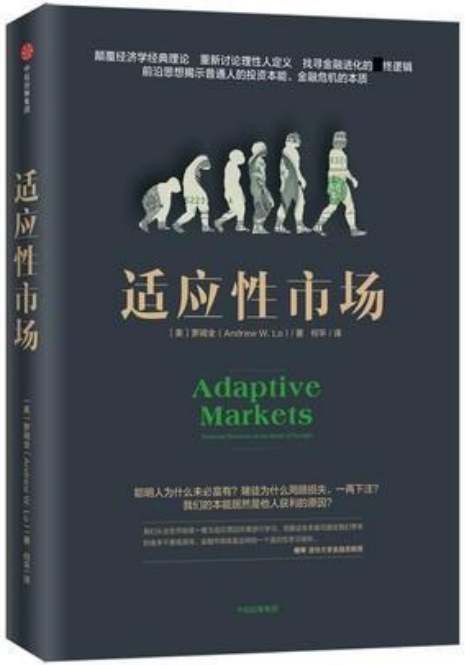
But What About Coordinated Fear??

Nightmarish unknown unknowns

2008 Financial Market crash

Behaviorists win (Not EMH):
Read Adaptive Markets





Next: Fintech ABCD, Vision, Challenges

- Fintech (F) = ABCD: AI, Blockchain, Cloud, Data [Deloitte 2017]:
  - 台大金融科技4部曲: 數位轉型出寬客 (F), 黑客 (CD), 極客 (A), 極客 (B): 共4門課
  - Example: This class in October: 普世普惠金融 (WoFi) 的量化交易 (Quantitative Trading) 比賽
    i. WoFi is our VISION: World Financial Inclusion (WoFi). See next 11 slides.
- Fintech's 4 Challenges
  - 數位化 → 數位優化 → 數位轉型: Challenging. Sol. Must understand 本質!
  - Fintech 跨領域: Social Science vs. Science/Engineering Schools even biology!
    i. 金融人不懂科技, 法律 ; 科技人不懂金融, 法律。Even worse, cultures are different:
    ii. Social science is physics envy: Many hypotheses such as EMH, AMH, …
    iii. Worst: Unknown unknowns → 高階決策 instead of 戰逃反應 → 寬客。
    iv. Summary: Those quotients can be orthogonal: IQ, EQ (AQ), SQ, FQ. Need to be 寬客。
  - 大家FQ低: Sol. This class will boost your FQ (Finance Quotient), which requires both IQ and EQ (AQ)!
  - Traditional finance doesn't want to be disrupted: 如何推出 Fintech 產品?

Adaptive Markets: IQ, FQ, SQ
Intelligence, Finance, Sex Quotients

第1章　現在我們是否都是經濟人？
悲劇與群眾的智慧
漫步歷史
效率市場之誕生
透視效率市場
所謂的理性預期
效率市場之應用

第2章　如果你那麼聰明，為什麼卻沒錢？
否定隨機漫步論
風險vs.不確定性和艾斯伯格矛盾
贏的感覺雖然爽，輸的傷害更痛
無限注德州撲克、流氓交易員與監理官員
機率對應與三月瘋
人類作為預測機器
打倒一個理論要靠另一個理論
文化衝擊

第3章　如果你那麼有錢，為什麼卻不聰明？
探索人腦的運作
神經科學的顯微鏡
恐懼
痛楚
愉悅與貪婪
測量交易員的生理反應
優秀交易員的素質
以神經通貨思考金錢
我全部都要，現在就要

寬客

- Use 人腦高階決策過程
  - Automate the 決策過程 with programs: Program Trading
- Not 低階決策過程(戰逃反應)。

- In the near future, program trading won't be 100% replacing human.
  - Hybrid mode: AI-as-assistant
    - Before each final decision, we have run through many many strategies: Program Trading
    - But the final decision may still be gut feeling (戰逃反應) for a few seconds.

  vs.

  - 100% AI mode:
    - Program trading instead of human.

- Note: Program trading: No guarantee for making more $. But 延年益壽 (relieve your body from 戰逃反應) and use 高階決策過程 more and more.

Agenda Today: Fintech ABCD, Vision, Challenges

- Fintech (F) = ABCD: AI, Blockchain, Cloud, Data [Deloitte 2017]:
  - 台大金融科技4部曲: 數位轉型出寬客 (F), 黑客 (CD), 極客 (A), 極客 (B): 共4門課 above
  - Example: This class in October: 普世普惠金融 (WoFi) 的量化交易 (Quantitative Trading) 比賽
    i. WoFi is our VISION: World Financial Inclusion (WoFi). See next 11 slides.
- Fintech's 4 Challenges
  - 數位化 → 數位優化 → 數位轉型: Challenging. Sol. Must understand 本質!
  - Fintech 跨領域: Social Science vs. Science/Engineering Schools even biology!
    i. 金融人不懂科技, 法律; 科技人不懂金融, 法律。Even worse, cultures are different:
    ii. Social science is physics envy: Many hypotheses such as EMH, AMH, …
    iii. Worst: Unknown unknowns → 高階決策 instead of 戰逃反應 → 寬客。
    iv. Summary: Those quotients can be orthogonal: IQ, EQ (AQ), SQ, FQ. Need to be 寬客。
  - 大家FQ低: Sol. This class will boost your FQ (Finance Quotient), which requires both IQ and EQ (AQ)!
  - Traditional finance doesn't want to be disrupted: 如何推出 Fintech 產品?
- Tools: Adaptive market, technical analysis, quants, program trading in Python, private key & cryptography, Blockchain, Smart contract, AI, Data analytics, DeFi. Arbitrage

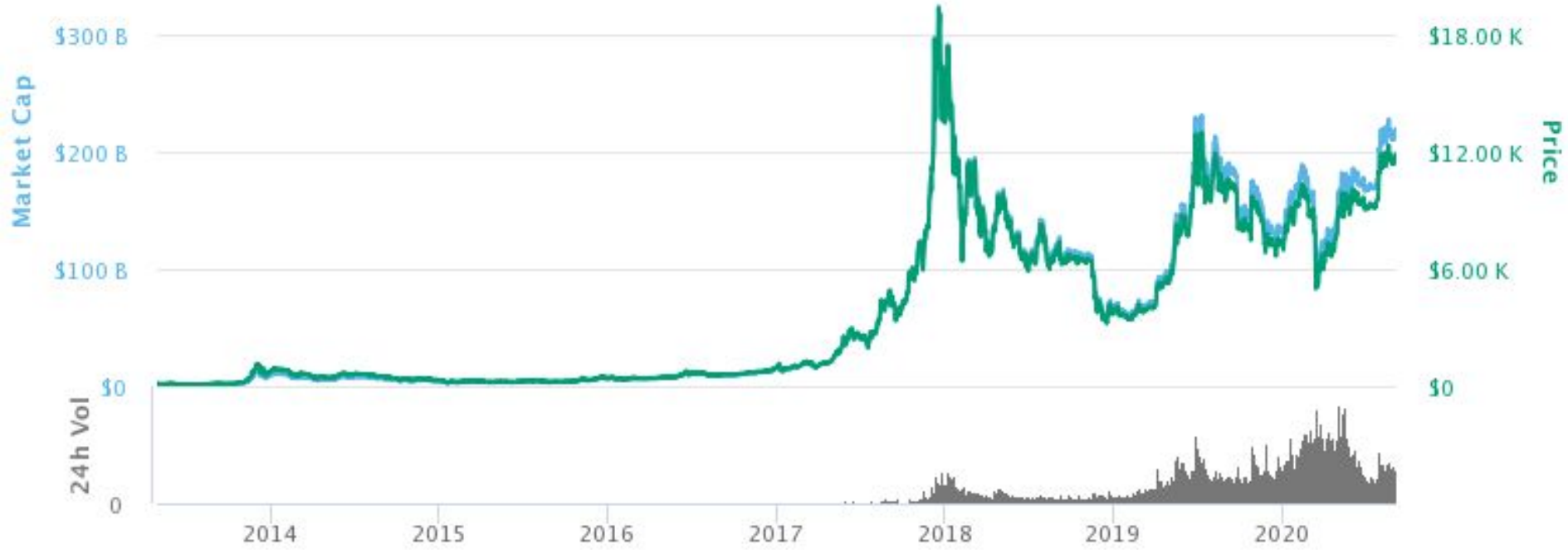
Outline of the Talk

- Digital transformation (DX)
- Data analytics
- Single-variable model (單變量模型) (未考慮價格因素，只考慮前n期的需求量)
- Multi-variable model (多變量模型) (考慮價格因素及前n期的需求量)
- Who are good customers?
- FinTech ABCD
- How it works in real world?

大數據監理 with Visualization

Virtual Asset RegTech

Best Example of FinTech ABCD in 20 minutes:
Bitcoin Tracing: 金融科技的詐欺, 治理, 審計, 鑑識


Market Cap
$300 B
$200 B
$100 B
$0
24 h Vol
0
$18.00 K
$12.00 K
$6.00 K
$0
Price
2014
2015
2016
2017
2018
2019
2020

Motivation

Cryptocurrencies are prevalent in recent years as they become more and more valuable. Bitcoin is the pioneer and is the most representative one.

- Activities - payment, investment, gambling, money laundering, etc.
- Identification - pseudo-anonymous.

We aim to identify the usage of given addresses.

Related Work

There are several publications that aimed at Bitcoin network analysis:

- Transaction flow
- Address clustering
- Graph pattern
- Unsupervised learning
- Supervised learning

Bitcoin Network

- Decentralized, immutable, secure, trust machine
- Proof-of-Work consensus algorithm
- Private key, public key, and address
- User, miner, and full node
- Transaction -> block -> blockchain
- Formed of a lot of transactions

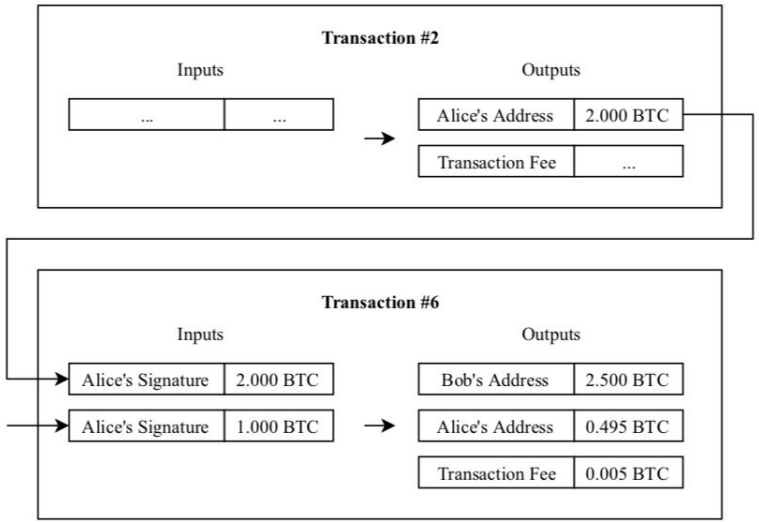
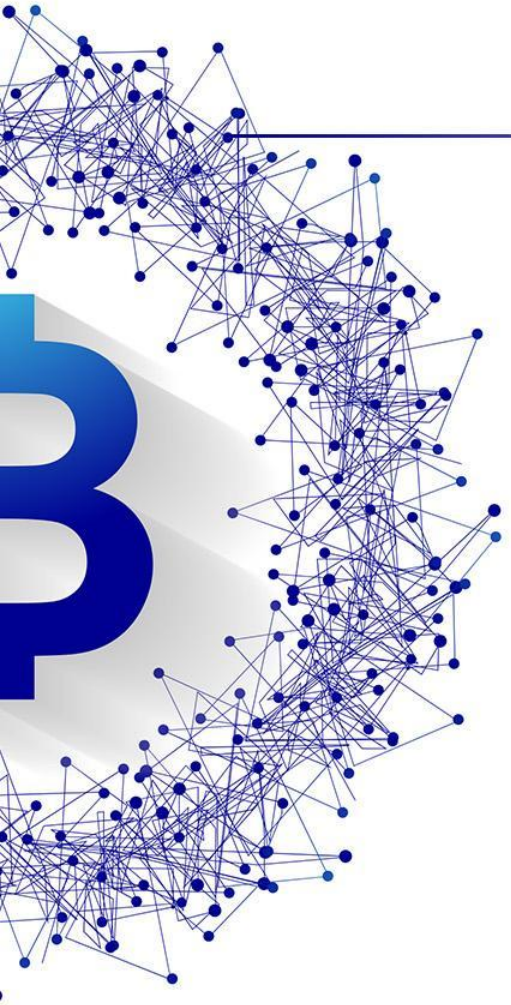


Fig. 2: An example of Bitcoin transactions.



Parser

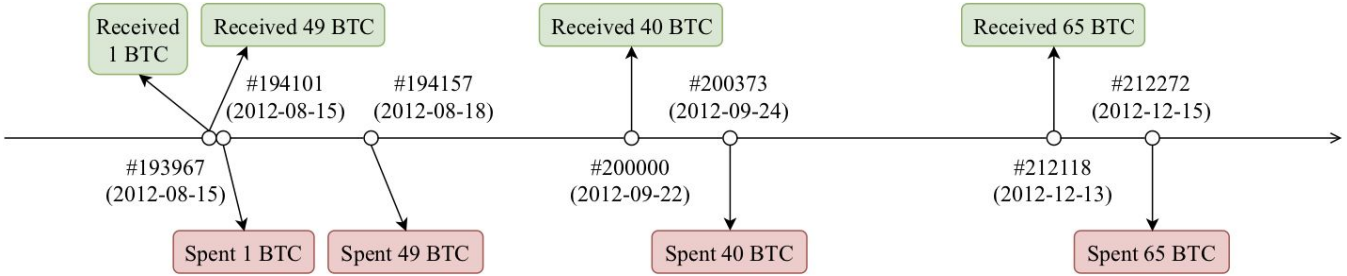
- Full node to retrieve block data
  - Millions of transactions per day
  - Up to 1000 to 10000 blocks per .dat

blk00456.dat	blk00928.dat	blk01400.dat	rev00456.dat	rev00928.dat	rev01400.dat
blk00457.dat	blk00929.dat	blk01401.dat	rev00457.dat	rev00929.dat	rev01401.dat
blk00458.dat	blk00930.dat	blk01402.dat	rev00458.dat	rev00930.dat	rev01402.dat
blk00459.dat	blk00931.dat	blk01403.dat	rev00459.dat	rev00931.dat	rev01403.dat
blk00460.dat	blk00932.dat	blk01404.dat	rev00460.dat	rev00932.dat	rev01404.dat
blk00461.dat	blk00933.dat	blk01405.dat	rev00461.dat	rev00933.dat	rev01405.dat
blk00462.dat	blk00934.dat	blk01406.dat	rev00462.dat	rev00934.dat	rev01406.dat
blk00463.dat	blk00935.dat	blk01407.dat	rev00463.dat	rev00935.dat	rev01407.dat
blk00464.dat	blk00936.dat	blk01408.dat	rev00464.dat	rev00936.dat	rev01408.dat
blk00465.dat	blk00937.dat	blk01409.dat	rev00465.dat	rev00937.dat	rev01409.dat
blk00466.dat	blk00938.dat	blk01410.dat	rev00466.dat	rev00938.dat	rev01410.dat
blk00467.dat	blk00939.dat	blk01411.dat	rev00467.dat	rev00939.dat	rev01411.dat
blk00468.dat	blk00940.dat	blk01412.dat	rev00468.dat	rev00940.dat	rev01412.dat
blk00469.dat	blk00941.dat	blk01413.dat	rev00469.dat	rev00941.dat	rev01413.dat
blk00470.dat	blk00942.dat	blk01414.dat	rev00470.dat	rev00942.dat	rev01414.dat
blk00471.dat	blk00943.dat	**index**	rev00471.dat	rev00943.dat	

Transaction History

Characterize an address or an entity by extracting features from its **transaction history**, which is a set of transactions relevant to itself.

Bitcoin Transaction History

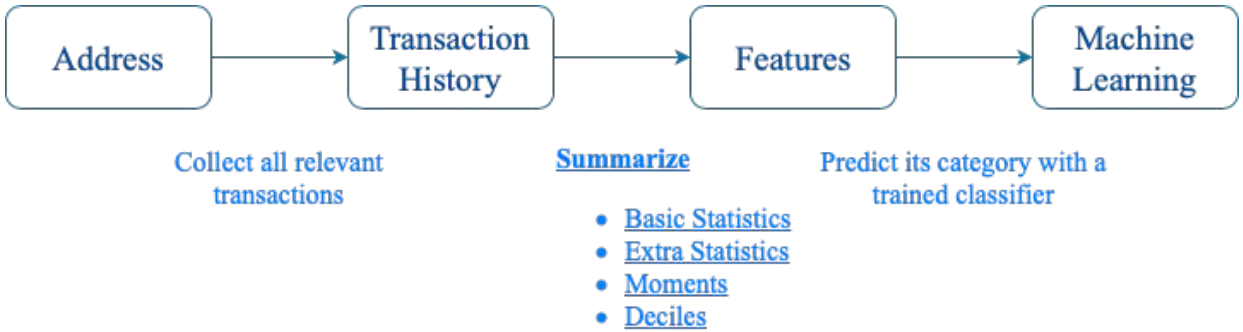


The transaction history of the address *15L23mj1TnFa9trXdpQ83iXrGzVd1byKUG*.

Basically, there are received transaction, spent transaction, coinbase transaction, and payback transaction. Each transaction is identified as only one of the four types exclusively.

Propose Method

Transaction History Summarization with four types of features: Basic Statistics, Extra Statistics, Moments, and Deciles.



Feature	Description
f_{TX}	The frequency of transactions, defined as number of all transactions per day in the address/entity's lifetime.
r_{received}	The ratio of received transactions to all transactions.
r_{coinbase}	The ratio of coinbase transactions to all transactions.
$f_{\text{spent}}(10^i)$	The frequency of digit i in USD appeared in spent transactions, where $i \in (10^{-3}, 10^{-2}, \ldots, 10^6)$.
$f_{\text{received}}(10^i)$	The frequency of digit i in USD appeared in received transactions, where $i \in (10^{-3}, 10^{-2}, \ldots, 10^6)$.
r_{payback}	Payback ratio defined as the ratio of Bitcoin addresses that appear in both inputs and outputs.
$\bar{N}_{\text{inputs}}$	The mean value of the numbers of inputs in the spent transactions.
$\bar{N}_{\text{outputs}}$	The mean value of the numbers of outputs in the spent transactions.
Basic Statistics	

Feature	Description
lifetime	The duration between the first transaction and the last transaction in terms of days.
BTC_{spent}	Total spent BTC.
BTC_{received}	Total received BTC.
USD_{spent}	Total spent USD, which are converted based on daily BTC/USD rates from *Coinmarketcap.com* [2].
USD_{received}	Total received USD, which are converted based on daily BTC/USD rates from *Coinmarketcap.com* [2].
n_{TX}	The number of transactions.
n_{spent}	The number of spent transactions.
n_{received}	The number of received transactions.
n_{coinbase}	The number of coinbase transactions.
n_{payback}	The number of payback transactions.
$\mu_{\text{balance\_btc}}$	The mean value of balance in BTC after each transaction.
$\sigma_{\text{balance\_btc}}$	The standard deviation of balance in BTC after each transaction.
$\mu_{\text{balance\_usd}}$	The mean value of balance in USD after each transaction.
$\sigma_{\text{balance\_usd}}$	The standard deviation of balance in USD after each transaction.
$\sigma_{N_{\text{inputs}}}$	The standard deviation of the numbers of inputs in the spent transactions.
$\sigma_{N_{\text{outputs}}}$	The standard deviation of the numbers of outputs in the spent transactions.
Extra Statistics	

Feature	Description
$m_{\text{n,overall}}$	The moments of overall transaction distribution.
$m_{\text{n,spent}}$	The moments of spent transaction distribution.
$m_{\text{n,received}}$	The moments of received transaction distribution.
$m_{\text{n,coinbase}}$	The moments of coinbase transaction distribution.
$m_{\text{n,payback}}$	The moments of payback transaction distribution.
$m_{\text{n,interval}}$	The moments of transaction interval distribution.
	Moments

Feature	Description
$d_{\text{n,overall}}$	The n^{th} deciles of overall transaction distribution. n ranges from 1 to 9.
$d_{\text{n,spent}}$	The n^{th} deciles of spent transaction distribution. n ranges from 1 to 9.
$d_{\text{n,received}}$	The n^{th} deciles of received transaction distribution. n ranges from 1 to 9.
$d_{\text{n,coinbase}}$	The n^{th} deciles of coinbase transaction distribution. n ranges from 1 to 9.
$d_{\text{n,payback}}$	The n^{th} deciles of payback transaction distribution. n ranges from 1 to 9.
$d_{\text{n,interval}}$	The n^{th} deciles of transaction interval distribution. n ranges from 1 to 9.
	Deciles

Transaction Moment

1). First moment

$$m_1 = \mathrm{E}[X]$$

2). Second moment

$$m_2 = \mathrm{E}[(X - \mu)^2]$$

3). Third standardized moment

$$m_3 = \mathrm{E}[(\frac{X - \mu}{\sigma})^3]$$

4). Fourth standardized moment

$$m_4 = \mathrm{E}[(\frac{X - \mu}{\sigma})^4]$$

Transaction Moment

1). First moment

$$m_1 = \mathrm{E}[X]$$

2). Second moment

3). T$m_2 = \mathrm{E}[(X-\mu)^2]$ ıent

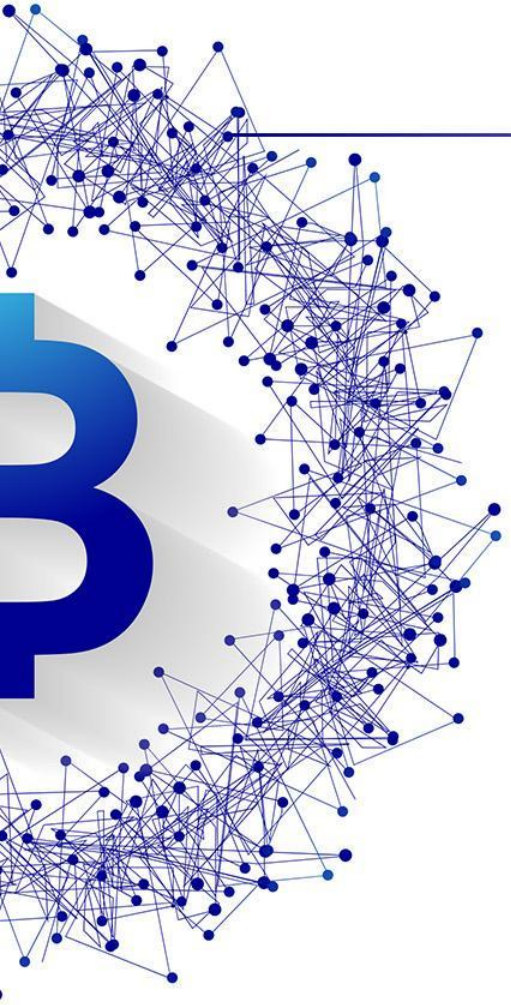
4). F

$$m_3 = \mathrm{E}[(\frac{X-\mu}{\sigma})^3]$$

$$m_4 = \mathrm{E}[(\frac{X-\mu}{\sigma})^4]$$

TABLE I
Moments.

Name	Meaning
1^{st} *moment*	measure of location
2^{nd} *moment*	measure of spread
3^{rd} *moment*	measure of symmetry
4^{th} *moment*	measure of peakedness



Parser

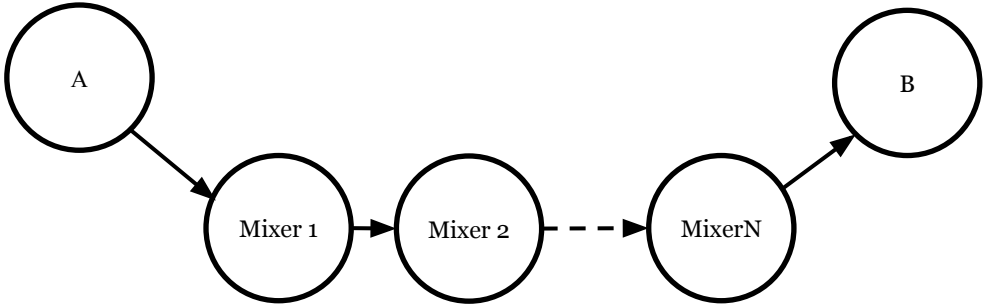
- Multi-purpose bitcoin address
  - Mixer Service
  - HYIP (High-Yield Investment Program)
  - Gambling
  - Exchange Wallet
  - Market
  - Faucet
  - Mining Pool

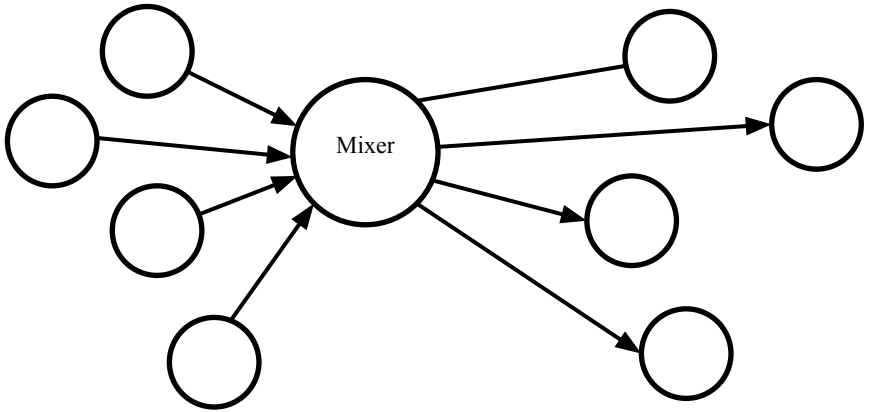
Parser


HOW DOES THE BITCOIN MIXER WORK?
Step 1
You make a deposit
Step 2
BestMixer.io mixes your coins
Step 3
You receive clean, untraceable coins

Parser

Two types of mixer

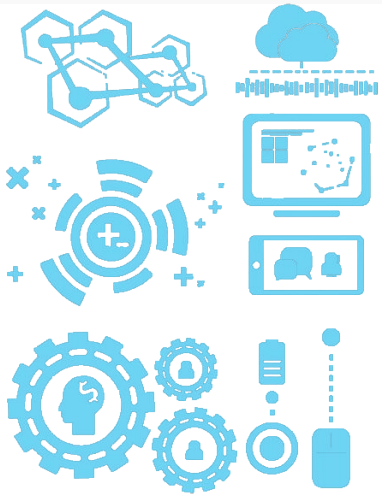




Machine Learning Algorithms

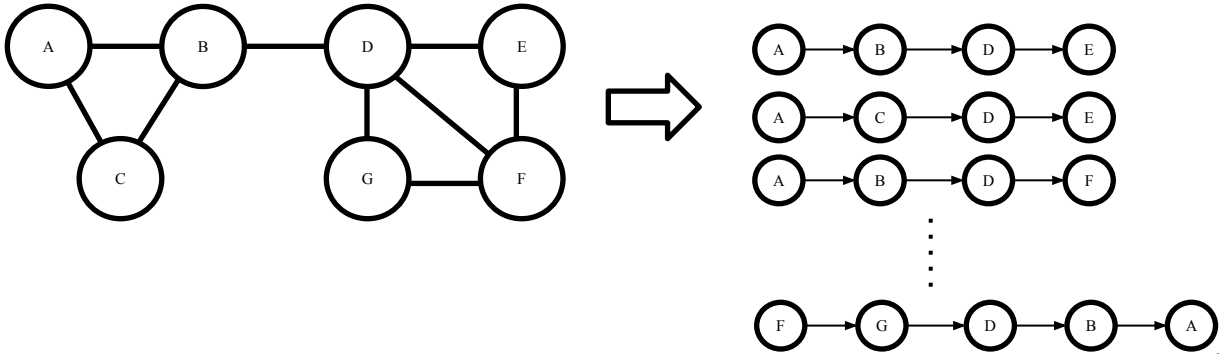
Trained 8 classifiers using grid search cross validation:

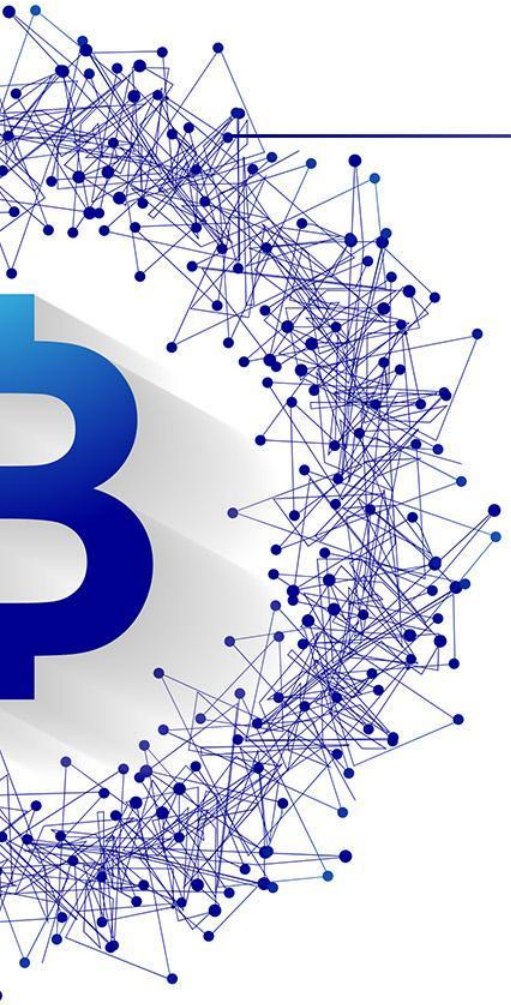
- Logistic Regression
- Perceptron
- Support Vector Machine
- Adaboost-SAMME
- Random Forest
- XGBoost
- LightGBM
- Neural Network



Unsupervised Analyzer

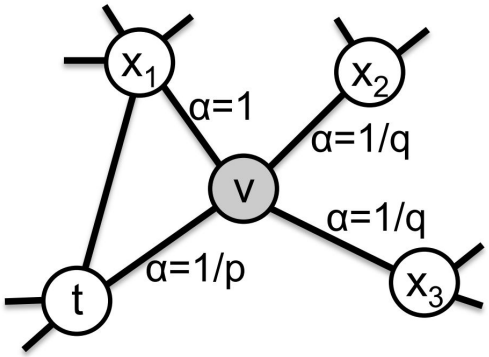
- Network Representation Learning
  - DeepWalk is one of NRL algorithm, random neighborhood to N vectors

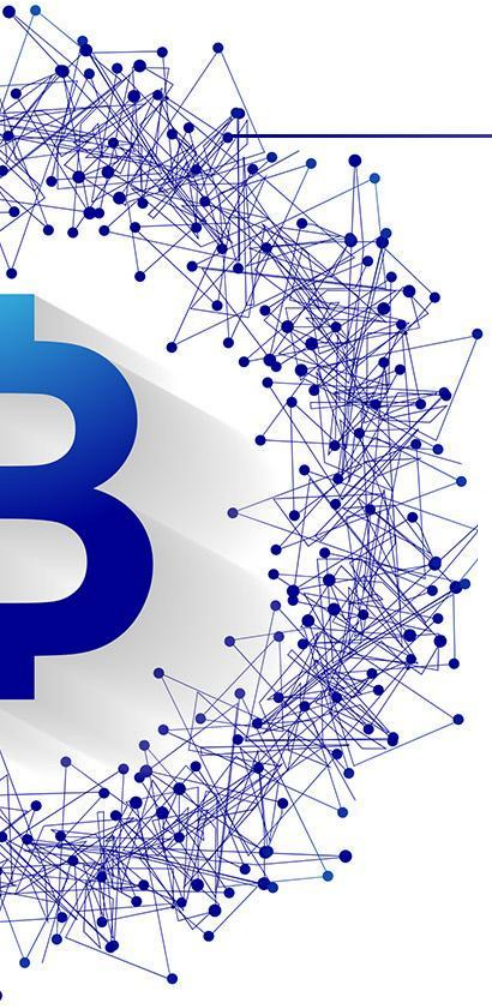




Unsupervised Analyzer

- Social network analysis (SNA)
  - Girvan-Newman algorithm
  - Node2vec algorithm





Unsupervised Analyzer

- 2015/12/25, 2016/2/16 transactions
- Optimized Node2Vec parameter

  p = 1.0, q = 2.0

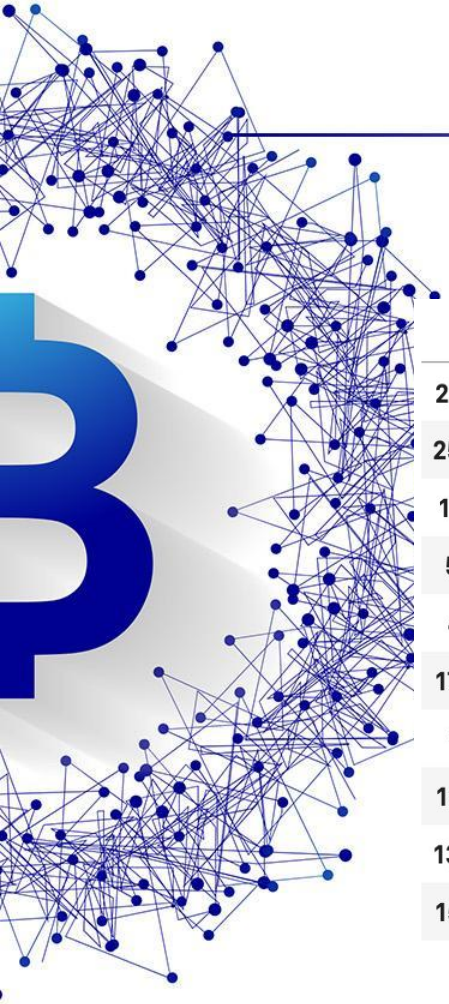
P \ q	0.2	0.4	0.8	1.0	2.0	4.0	8.0
0.5	0.82743	0.81599	0.81239	0.79383	0.81846	0.81063	0.82433
1.0	0.81467	0.82406	0.81471	0.79457	0.86185	0.83049	0.82399
2.0	0.85169	0.81616	0.81509	0.85359	0.43675	0.70773	0.81027
4.0	0.83021	0.79571	0.78691	0.81029	0.80053	0.82483	0.80587
8.0	0.81059	0.28053	0.83294	0.82109	0.25279	0.24713	0.81529

Unsupervised Analyzer

- K-means to 1000 clustering
  - More than 160,000 bitcoin address
  - 52 known mixer address similarity

分群 968	310321	3/52
	341831	
	355853	
分群 989	111374	
	179234	
	227283	
	235883	7/52
	241178	
	278002	
	359957	

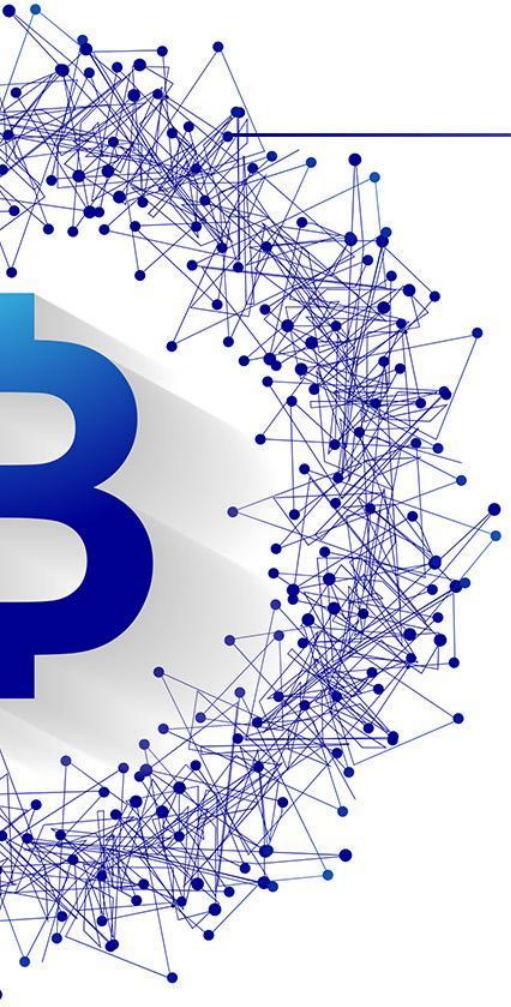
Supervised Analyzer



- 26313 of labeld bitcoin address

	name	address	class	owner
20132	Vircurex.com	1GhMQg4KGaSAsWZwrYjDD5KqmGxwkogDdA	Exchange	58523298
25552	AllCrypt.com	18aoEXYv8JjQKDgdh6Kv2ucmfPTgyXM35m	Exchange	14847183
11799	DiceNow.com	17ZDaTSNzhAc51nn178et751dJ6iUKwHrp	Gambling	58273210
5806	HelixMixer-old21	192cdtuk5GzhXPwiAkm3zKbBLGj9L6zxSX	Mixer	77600611
4527	HelixMixer-old8	1MFAGpHfK1rhkSRw9DsPQLRzWMoEUV7Kmo	Mixer	52557781
17893	Kraken.com	1KRaR3G7hSagLVjm5tB6u5w1NdkZumWtrg	Exchange	40915984
1339	PrimeDime - Automated Investments (Total wage...	1H8r7FykKGBL6iCwWhMazEWUZKberuHN7G	HYIP	39065049
12473	SilkRoad2Market	1EBE2JUvWtfPKwu2nsgzuwuK2GZU7ANxJ1	Market	15231866
13398	SheepMarketplace	1BmxMh8EY1AH29KvsddD71j9RDB1fxa4Su	Market	10577289
15760	Just-Dice.com	1B1is1hbR7T4CqvNJe4Lsp9txYG9LgsTGF	Gambling	169513

Retrieved from https://www.walletexplorer.com



Supervised Analyzer

- Bitcoin features extraction

Feature	Description
f_tx	平均每日交易次數
r_received	接受交易的比例
r_coinbase	挖到礦的比例
f_spent	支出交易的金額數量級統計
f_received	接受交易的金額數量級統計
r_payback	找錢給自己的比例
n_inputs	支出交易的平均 input 數
n_outputs	接受交易的平均 output 數
std_interval	平均交易間隔的標準差

Evaluation

Method	Entity-based Scheme		Address-based Scheme	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1
Logistic Regression	0.76	0.62	0.47	0.45
Perceptron	0.63	0.53	0.37	0.34
SVM	0.59	0.46	0.43	0.41
AdaBoost-SAMME	0.38	0.30	0.38	0.37
Random Forest	0.90	0.72	0.82	0.80
XGBoost	0.90	**0.76**	0.84	0.83
LightGBM	**0.91**	**0.76**	**0.87**	**0.87**
Neural Network	0.89	**0.76**	0.81	0.79

Evaluation

- LightGBM overall accuracy rate at 86%

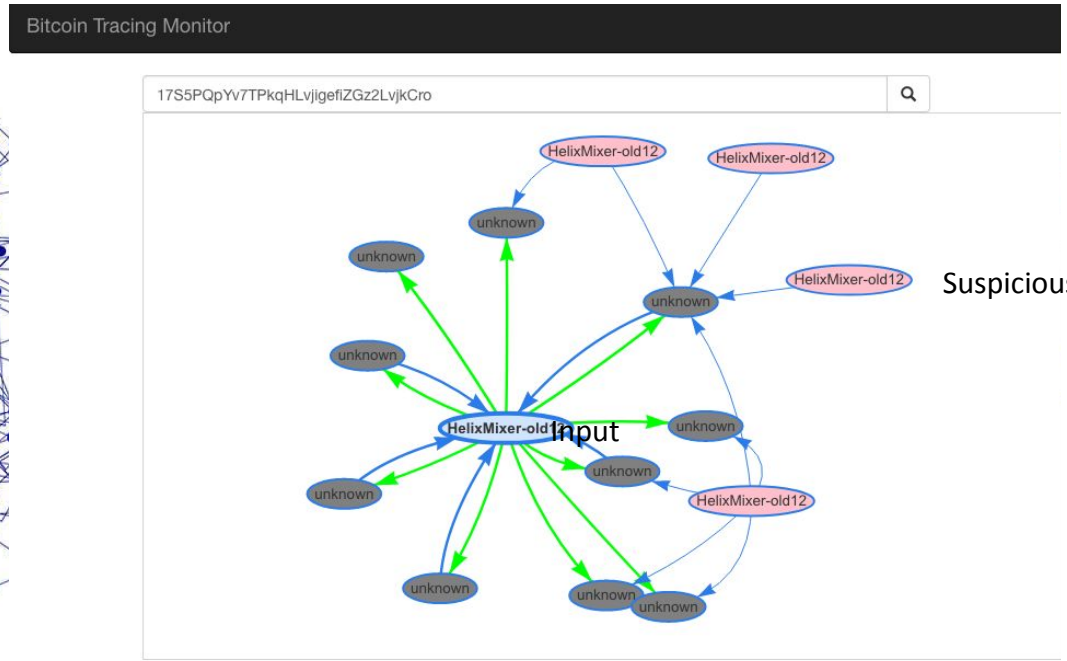
Lin, Y. J., Wu, P. W., Hsu, C. H., Tu, I. P., & Liao, S. W. (2019, May). An Evaluation of Bitcoin Address Classification based on Transaction History Summarization.

Evaluation

Method				Results							Results	
B	**E**	**M**	**D**	**Exchange**	**Faucet**	**Gambling**	**HYIP**	**Market**	**Mixer**	**Pool**	**Micro Avg.**	**Macro Avg.**
√				0.80	0.66	0.73	0.82	0.69	0.96	0.78	0.79	0.78
	√			0.84	0.70	0.76	0.78	0.76	0.96	0.79	0.82	0.80
		√		0.71	0.19	0.57	0.53	0.56	0.92	0.53	0.66	0.57
			√	0.70	0.14	0.53	0.54	0.46	0.91	0.53	0.64	0.54
√	√			0.87	**0.81**	0.82	0.87	0.81	**0.98**	**0.87**	**0.87**	0.86
√	√	√		**0.88**	**0.81**	**0.83**	**0.88**	0.82	**0.98**	**0.87**	**0.87**	**0.87**
√	√	√	√	**0.88**	**0.81**	**0.83**	**0.88**	**0.83**	**0.98**	**0.87**	**0.87**	**0.87**

Basic+Extra+Moments slightly improve the performance on Exchange, Gambling, HYIP, and Market.
Basic+Extra+Moments+Deciles slightly improve the performance on Market.

Monitor



Suspicious address

Input

Related address