



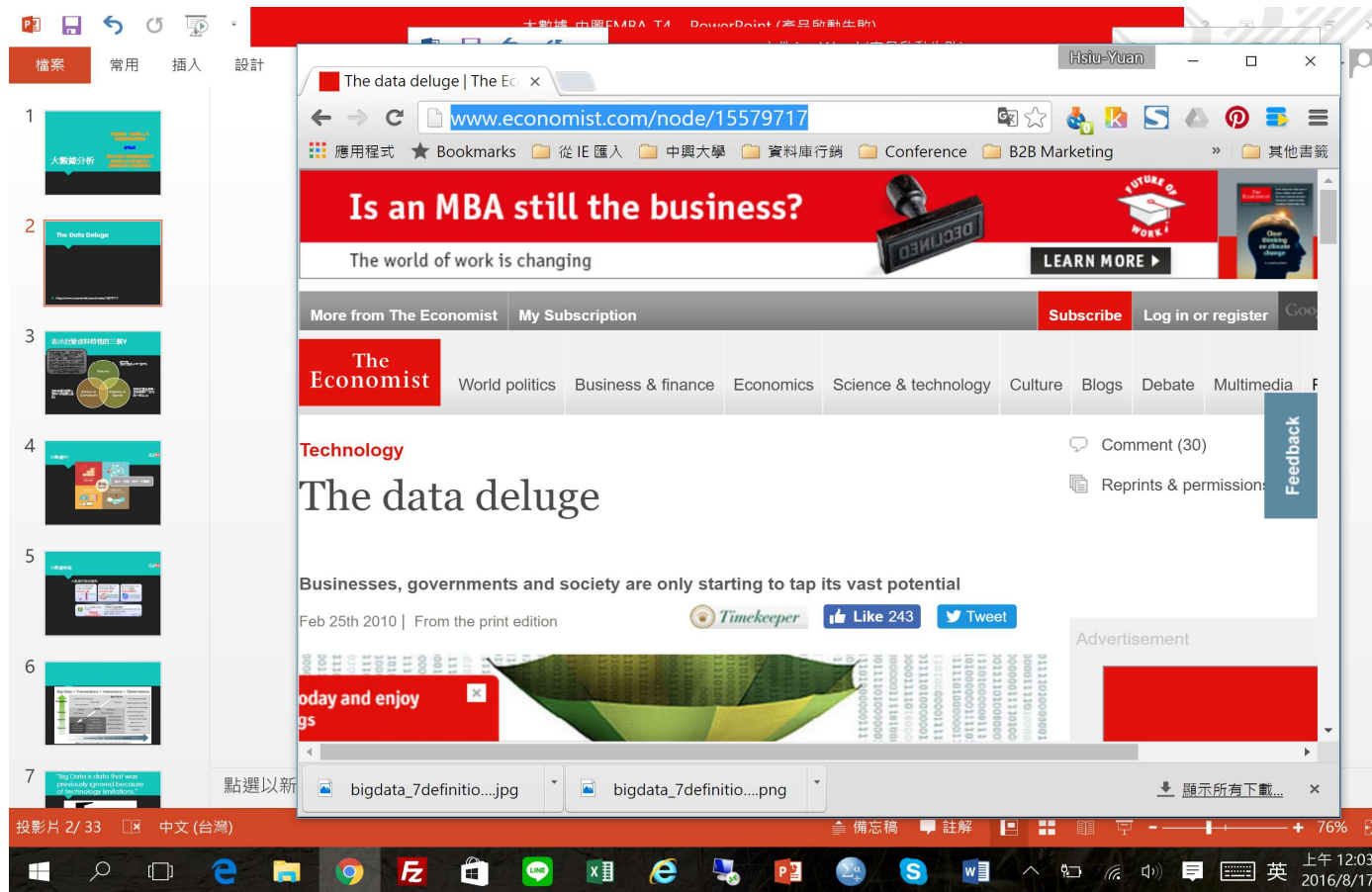
23/12/2023

- **Concept, Technology, and Application**

CONCEPT

Big Data Marketing Analytics

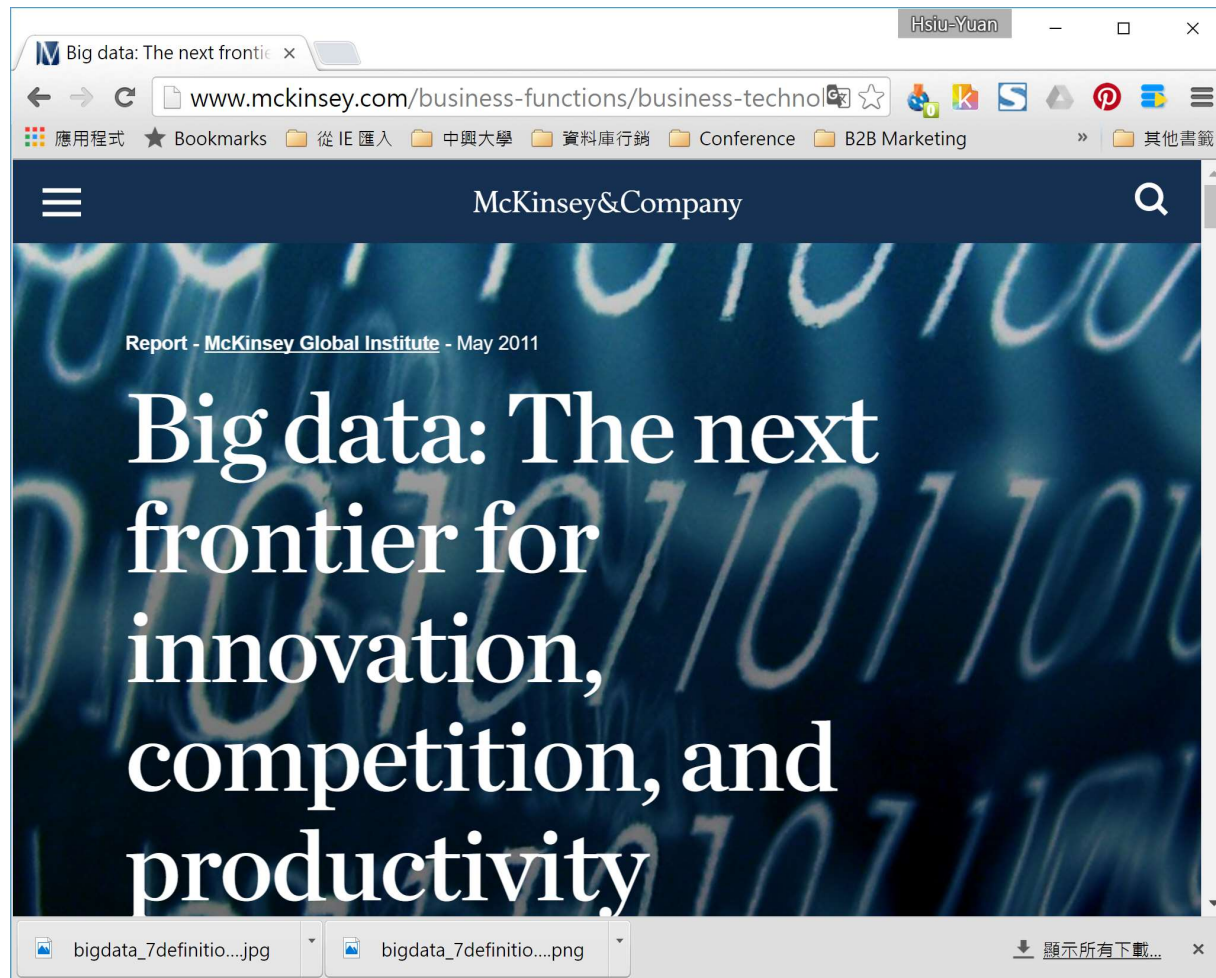
The Data Deluge



The screenshot shows a presentation window on the left with a sidebar containing seven thumbnails. The main window displays a web browser with the URL www.economist.com/node/15579717. The browser window has a red header with the text "Is an MBA still the business?" and "The world of work is changing". Below this is a navigation bar with "More from The Economist" and "My Subscription". The main content area is titled "The data deluge" under the "Technology" category. It includes a sub-headline "Businesses, governments and society are only starting to tap its vast potential", the date "Feb 25th 2010", and the source "From the print edition". There is a "Timekeeper" logo and social media buttons for "Like 243" and "Tweet". A large image of a green boat is visible. The bottom of the browser window shows a taskbar with various application icons and a system clock indicating "上午 12:03 2016/8/17".

- <http://www.economist.com/node/15579717>

»»»» The Big Data



<http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>

Business Intelligent

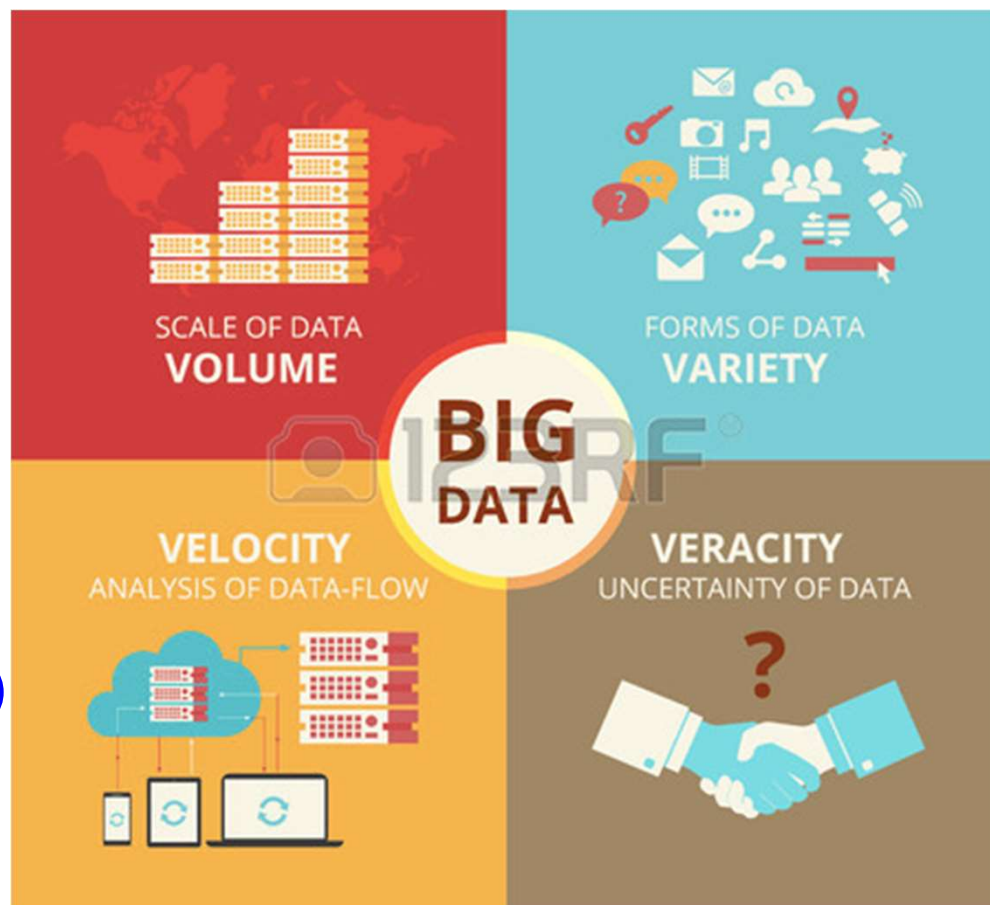
由「分析現狀」進化到「預測未來」，將分析對象擴展至巨量資料，實現更深的分析洞察與更準確的預測。



巨量資料特性的4V

資料量

資料的產生頻率、
更新頻率
(一秒內數十條以上)



資料多樣
(結構化資料+
非結構化資料)

不確定性

APPLICATION

Big Data Marketing Analytics

巨量資料運用案例

- 精準推薦商品或服務
- 行為定位廣告
- 運用地點資訊的行銷
- 揪出盜刷
- 顧客流失分析
- 預測設備故障
- 驗出異常
- 改善服務
- 預測路況
- 預測電力需求
- 預測感冒流行
- 預測股市行情
- 郵資成本的優化

大數據運用 數據驅動決策



以大數據分析為商業核心驅動力，打造以消費者為導向之電商系統



單日RMB**350億**營業額
2013光棍節創下世界紀錄

每分賣**4.8萬**件商品
總商品數超過8億件

超過**5億**會員數
每日訪客數為臺灣人口三倍

大數據運用 預測與推薦



大數據運用 數據分析行為

decide. + ebay

Decide is now part of the eBay family. Thank you to all our supporters who made it possible.



大數據運用 輿論分析與預測

Sentiment v.s Stock Market

The screenshot displays the AlgoTechnology website interface. The top navigation bar includes links for 'AlgoTechnology', 'About', 'Approach', 'Products', and 'Services'. The main content area is titled 'Sentiment Alerts' and features a search bar with options to 'Select Country', 'Select Region', 'Select or Create Portfolio', and 'Search by name or sector'. Below the search bar, there is a list of stocks with their respective sentiment scores and trends. The selected stock, DENTSPLY International Inc, is highlighted with a detailed sentiment analysis, including a 'Sentiment Signal Analysis' table and a 'Sentiment Trend' section. The 'Sentiment Signal Analysis' table shows a 'Spot Score above 90' with a 'Hit Rate' of 100.0%. The 'Sentiment Trend' section includes two charts: 'Price vs Sentiment over the last 60 days' and 'Price vs Sentiment over the last 260 days'.

AlgoTechnology Limited X +

← → ↺ | algotechnology.com

AlgoTechnology About Approach Products Services

Sentiment Alerts

sentiment.algotechnology.com/secure/index.jsp?inst=EQXRAY&refresh_inst=y

Google

TheySay

AlgoTechnology

You are logged in as demo@algotechnology.com. (Logout)

Select Country ... Select Region ... Select or Create Portfolio ... Search by name or sector ...

-- Choose -- Go >> -- Choose -- Go >> -- Choose -- Go >> Search & Aggregate >> Launch Sentiment Monitor >>

26 Jan 2015 14:00 UTC ~1 hour ago Xero Ltd @16.40 NZD Long trade: 15d [5.0% tgr] 88% HR, 3.1% ATR

26 Jan 2015 14:00 UTC ~1 hour ago DENTSPLY International Inc @53.54 USD Long trade: 5d 100% HR, 1.4% ATR

26 Jan 2015 14:00 UTC ~1 hour ago Worldwide Healthcare Trust PLC @17.50 GBP Long trade: 15d [5.0% tgr] 100% HR, 3.6% ATR

26 Jan 2015 13:59 UTC ~1 hour ago Wincor Nixdorf AG @40.26 EUR Short trade: 20d [5.0% tgr] 86% HR, 4.2% ATR

26 Jan 2015 13:59 UTC ~1 hour ago Whirlpool Corp @196.83 USD Long trade: 20d 84% HR, 6.1% ATR

26 Jan 2015 13:59 UTC ~1 hour ago Westfield Corp @9.63 AUD Long trade: 20d 100% HR, 7.8% ATR

26 Jan 2015 13:59 UTC ~1 hour ago Western Digital Corp @110.38 USD Long trade: 10d [5.0% tgr] 83% HR, 4.1% ATR

DENTSPLY International Inc 53.54 USD

Sentiment Scores - Spot +96.2 / 20d Average +19.6

[XRAY](#) | [US](#) | [USD](#) | [Health Care](#) | [Health Care Equipment & Services](#)

This analysis is based on analysing the sentiment scores on 6 news stories over the last week. Sentiment scores of greater than zero indicate net positive sentiment towards DENTSPLY International Inc. Sentiment scores below zero indicate net negative sentiment. A score of zero is neutral.

DENTSPLY International Inc has net bullish sentiment as both the spot score of +96.2 and the average 20d score of +19.6 are both greater than zero. There is no strong price trend for DENTSPLY International Inc but the recent positive sentiment may begin to trigger a rally.

Sentiment Signal Analysis

Direction (Long or Short)	Duration (Days)	Target (%)	Stop Loss (%)	Signal Description (How the signal is derived from the sentiment scores)	Trades (Total)	Winners (Total Wins)	Losers (Total Losses)	Hit Rate (%)	Avg Trade Return (%)
Long	5	None	None	Spot Score above 90	3	3	0	100.0	1.42

Sentiment Trend

The charts below show trends in the sentiment indices over time. The top chart shows the sentiment indices and asset price. The lower chart shows the number of news items that have been used to calculate the sentiment indices and hence gives a feeling for the relative levels of news items now versus history.

Price vs Sentiment over the last 60 days.

Price vs Sentiment over the last 260 days.

http://www.algotechnology.com/#carousel-example-generic

上午 10:30 2016/8/28

http://www.algotechnology.com/

大數據運用深度

1

- 對過去/現況的掌握

2

- 發現行為模式

3

- 預測

4

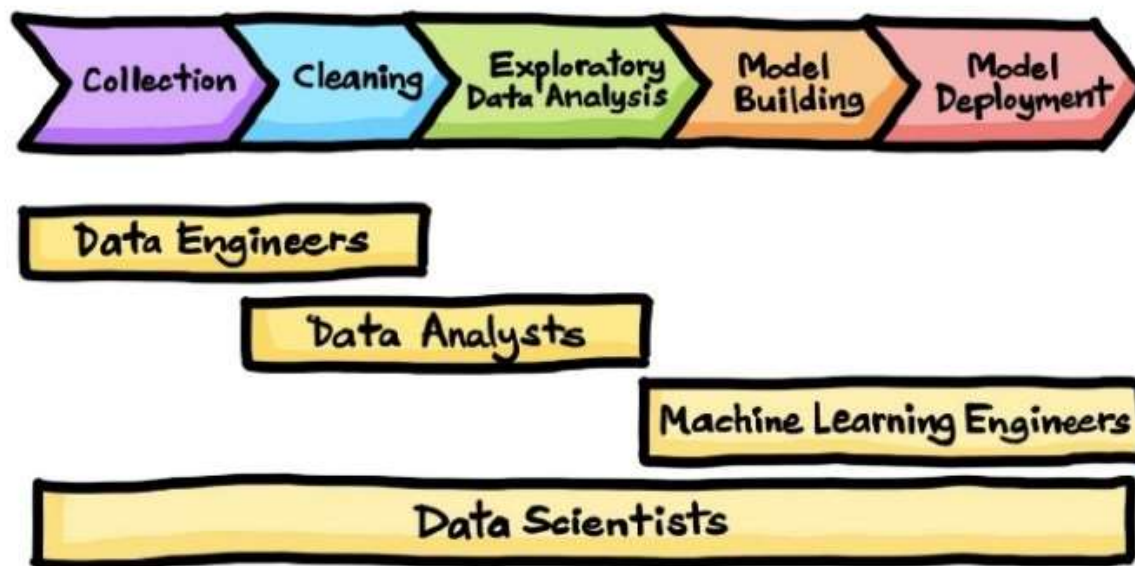
- 優化

TECHNOLOGY

Big Data Marketing Analytics

資料科學處理流程

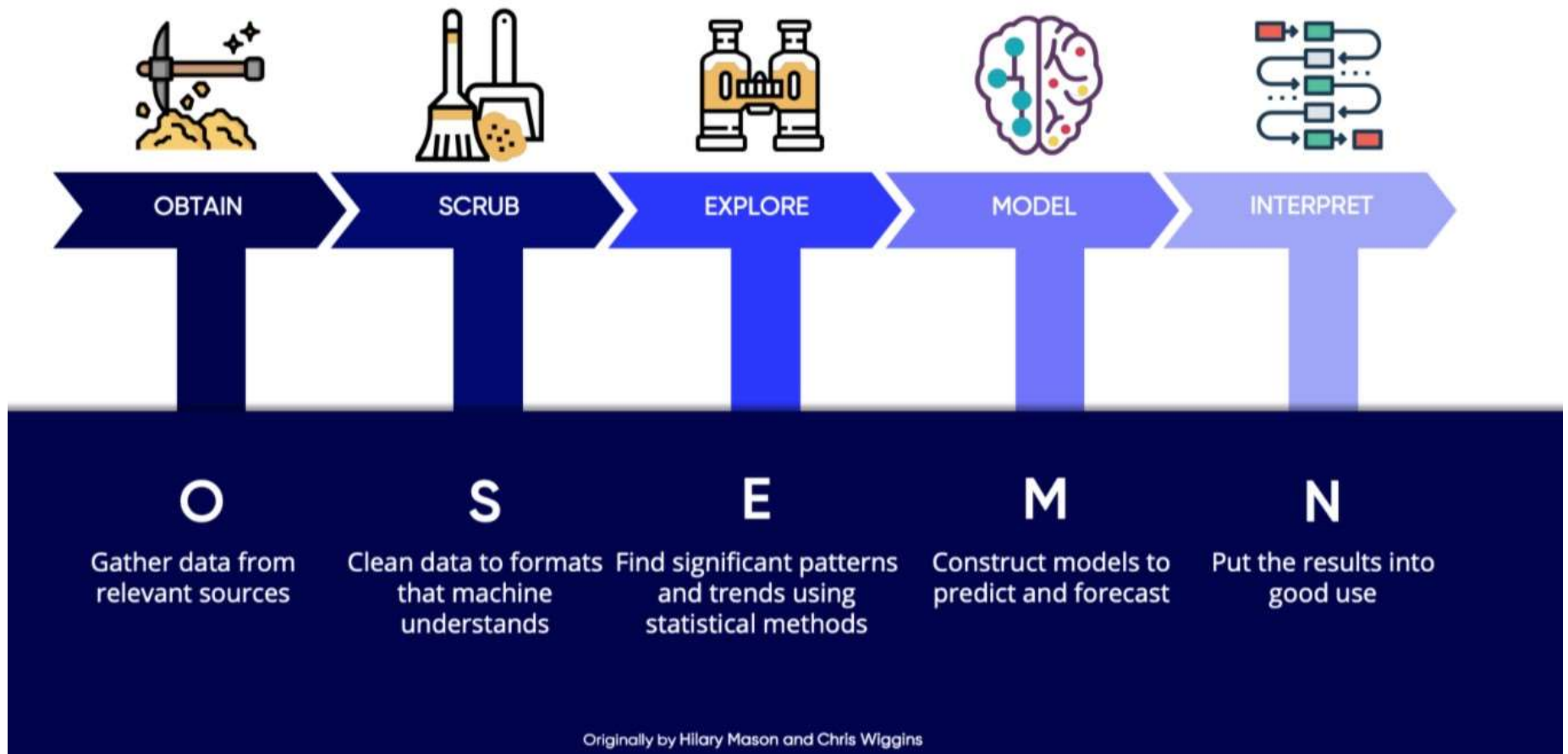
THE DATA SCIENCE PROCESS



Data science life cycle. (Drawn by Chanin Nantasenamat in collaboration with Ken Jee)

資料科學處理流程

Data Science Process



大數據分析相關技術

分群(Clustering)

分類(Classification)

預測(Prediction)

關聯規則(Association)

文字分析(Text Mining)

機器學習(Machine Learning)

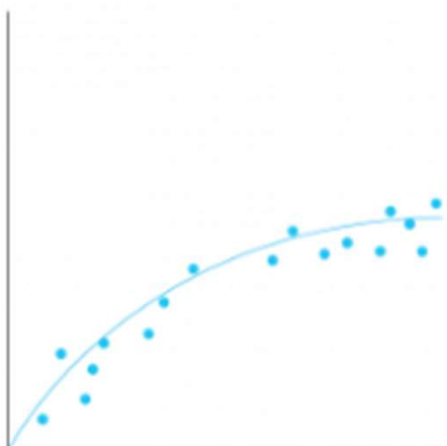


行銷分析 Marketing Analytics

Basic Concept you need to know

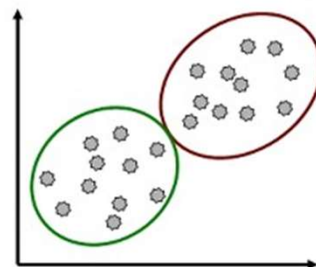
分類還是分群

Regression



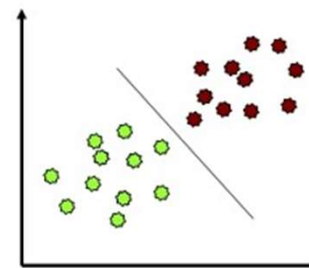
CLUSTERING

- Data is not labeled
- Group points that are "close" to each other
- Identify structure or patterns in data
- Unsupervised learning

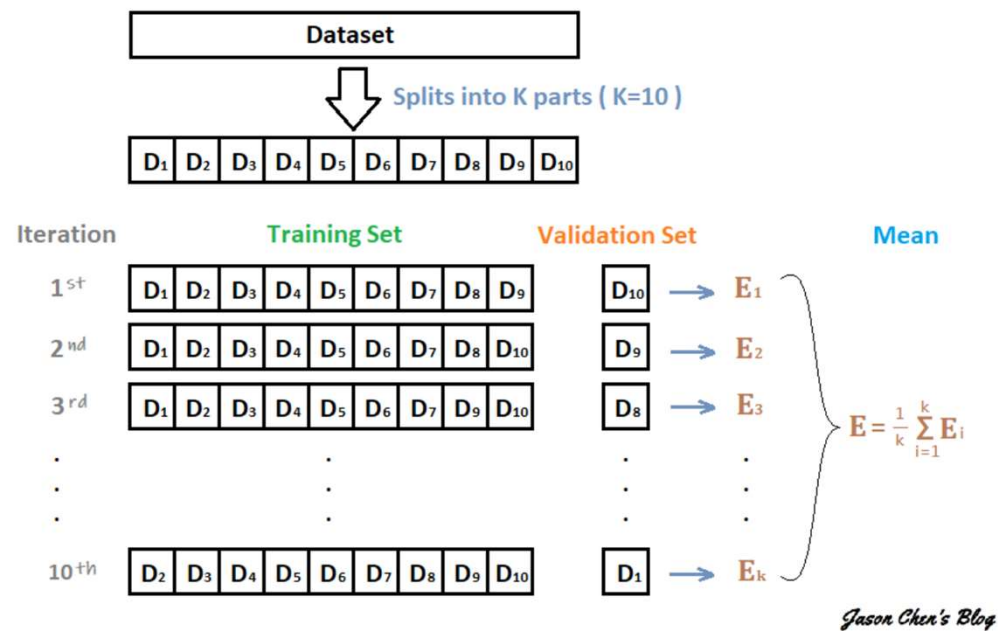
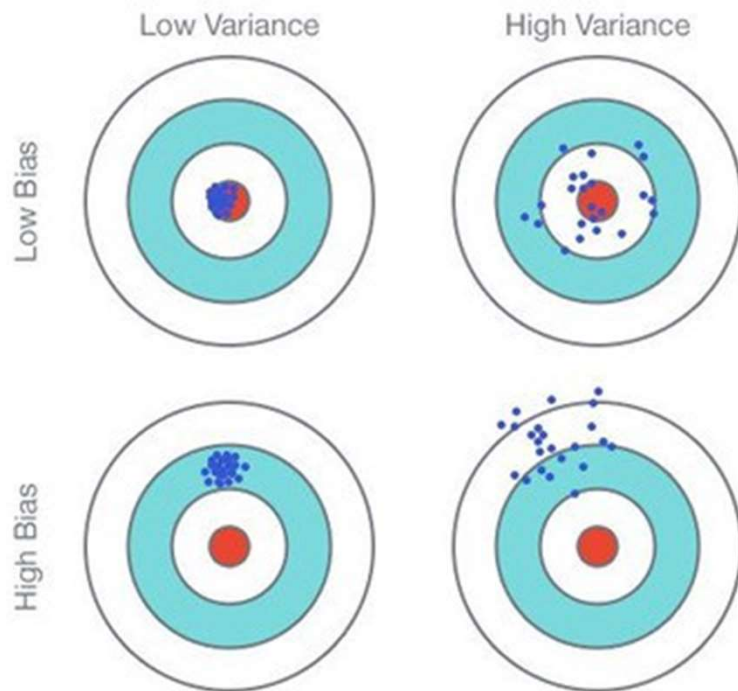


CLASSIFICATION

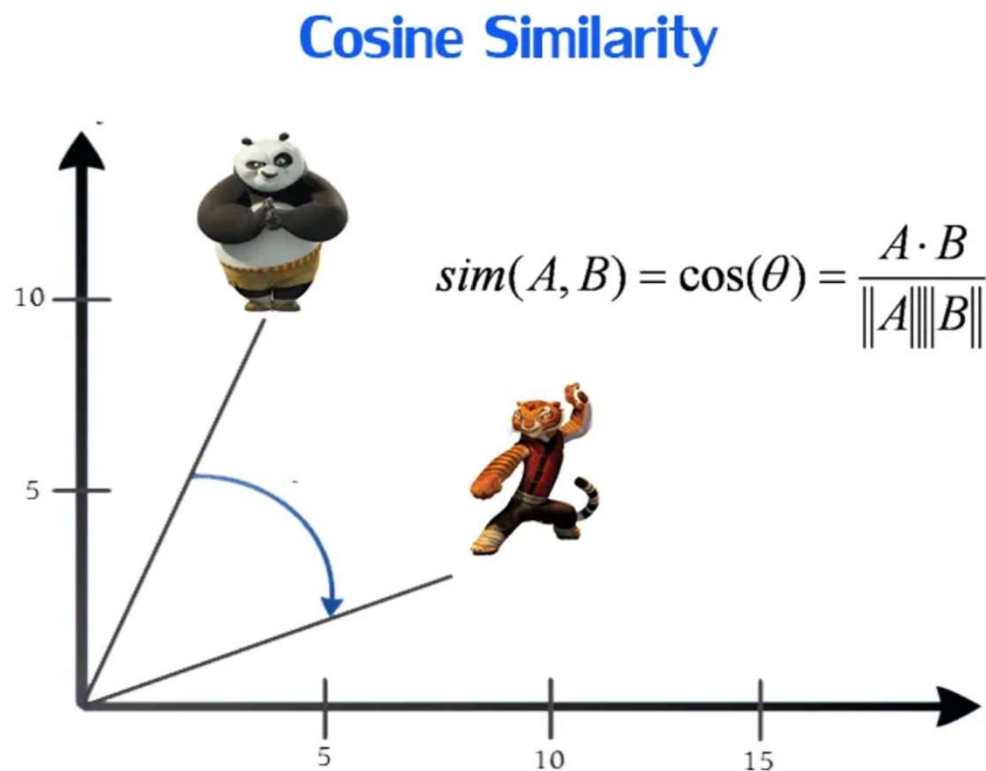
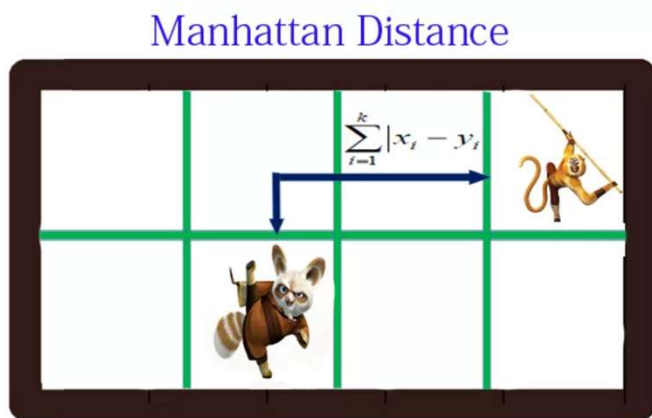
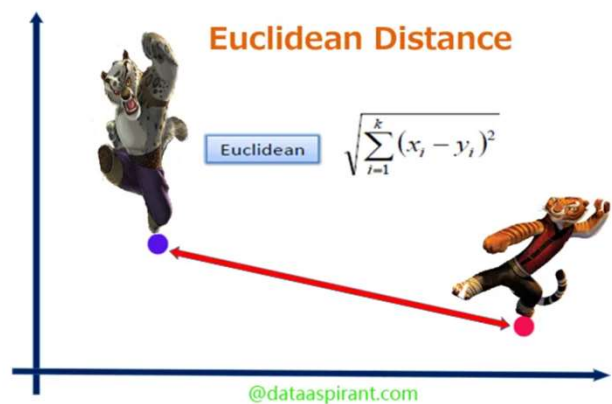
- Labeled data points
- Want a "rule" that assigns labels to new points
- Supervised learning



變異還是偏誤



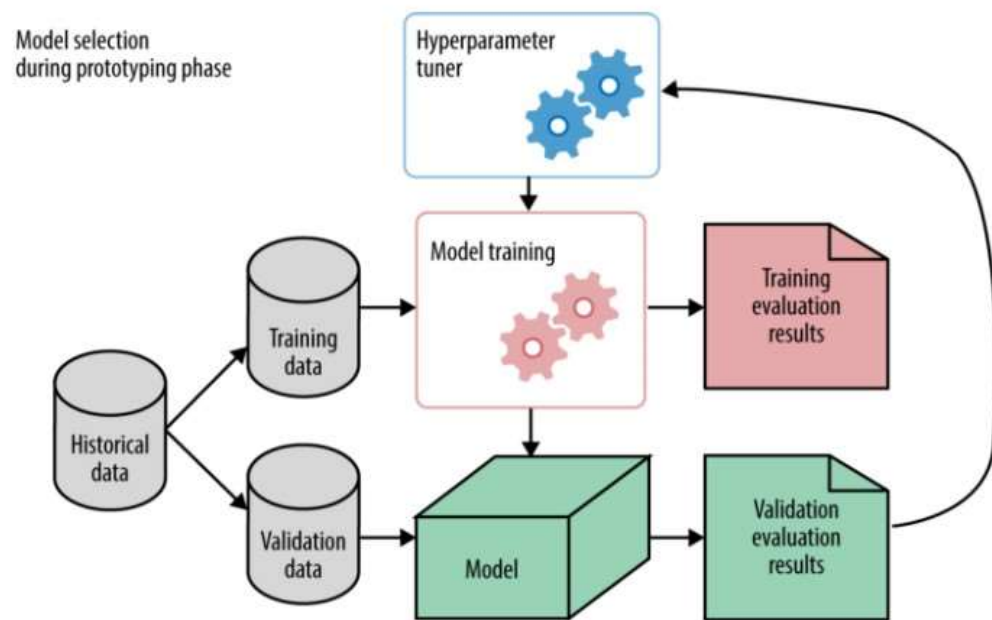
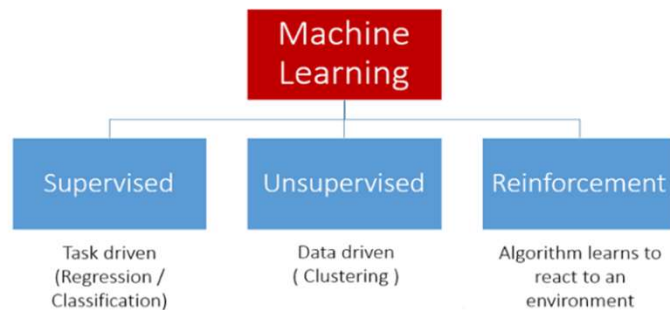
何謂距離



機器學習演算法

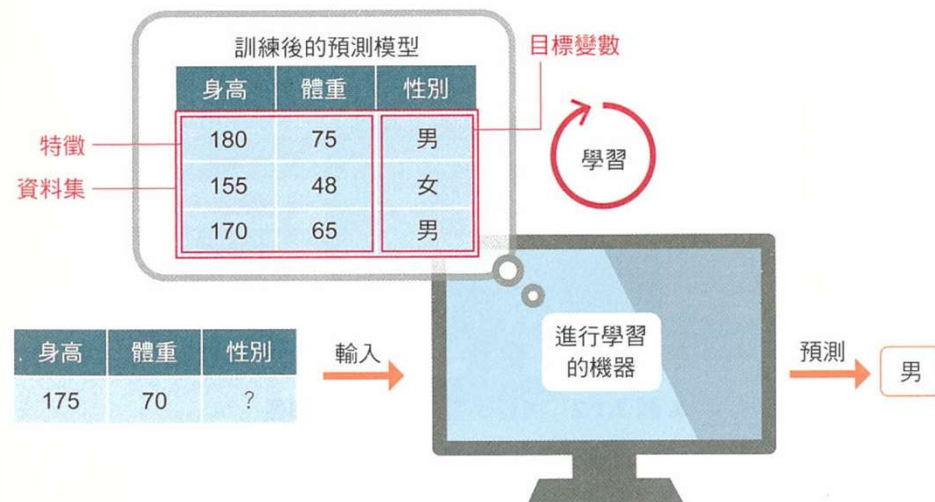
1. Gathering Data.
2. Preparing Data.
3. Model Selection.
4. Training.
5. Evaluation.
6. Prediction.

Types of Machine Learning



分類分群還是預測

Prediction (numerical)	Classification (categorical)	Clustering (categorical)
Supervised	Supervised	Unsupervised
Linear Regression	Logistic Regression	K-Means
Multiple Regression	KNN -K Nearest Neighbor	Hierarchical
Ridge Regression	Naïve Bayes	Neural Networks
Lasso Regression	Decision Trees	
Bayesian	Ensemble Trees	
Neural Networks	SVM -Support Vector Machines	
Quantum Computing	Neural Networks	



非監督式學習

降維

PCA 等

分群

k-means 等

監督式學習

分類問題（二元問題或多元問題）

羅吉斯迴歸等

迴歸問題

線性迴歸

顯著性還是混淆矩陣

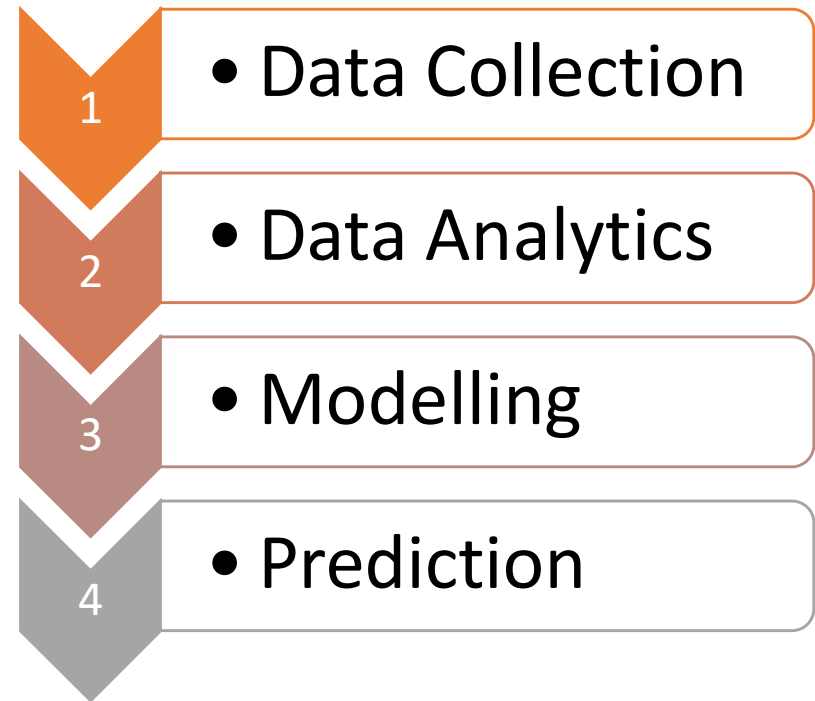
Confusion Matrix- (Also called fitting

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

- 正確率 (Accuracy) : $\frac{TP + TN}{TP + TN + FP + FN}$
- 精確率 (Precision) : $\frac{TP}{TP + FP}$
- 召回率 (Recall) : $\frac{TP}{TP + FN}$
- F1 值 (F1-Score) : $2 \times (\text{精確率} \times \text{召回率}) / (\text{精確率} + \text{召回率})$

大數據分析相關工具與步驟

- 程式語言:
 - R / Python
- Excel Addon:
 - XLSTAT
 - XLMINER
- 分析軟體:
 - IBM SPSS / IBM SPSS Modeller
 - Microsoft Azure Machine Learning
- 機器學習分析平台
 - DataRobot
 - Paperspace Gradient
 - IBM Cognos Analytics



**Ten Machine Learning Algorithms You
Should Know to Become a Data
Scientist**

機器學習演算法

Clustering / Segmentation (範例一)

The image shows the XLSTAT interface with the 'k-means clustering' dialog box open. The background Excel spreadsheet contains data for 15 US cities, including City #, City name, %age Black, %age Hispanic, Unemployment rate, and Per capita income (000's).

XLSTAT - k-means clustering dialog box settings:

- General tab selected.
- Observations/variables table: USCities!\$C:\$H
- Column weights: ☐ (unchecked)
- Row weights: ☐ (unchecked)
- Clustering criterion: Determinant(W)
- Number of classes: from 4 to 4
- Range: ☐ (unchecked)
- Sheet: ☒ (checked)
- Workbook: ☐ (unchecked)
- Column labels: ☒ (checked)
- Row labels: ☒ (checked)

Excel Spreadsheet Data (US Cities):

City #	City	%age Black	%age His	Unemployment rate	Per capita income(000's)
1	Albuquerque		3		
2	Atlanta	67			
3	Austin	12			
4	Baltimore	59			
5	Boston	26			
6	Charlotte	32			
7	Chicago	39			
8	Cincinnati	38			
9	Cleveland	47	5	1	32
10	Columbus	23	1	2	29
11	Dallas	30	21	2	30

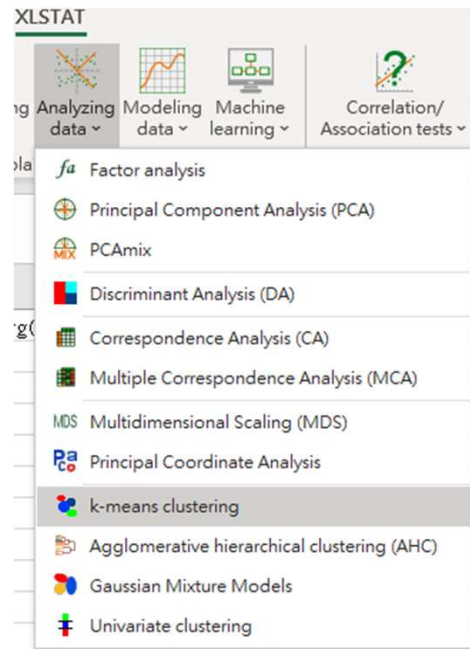
XLSTAT - Results (US Cities):

Class	1	2	3	4
Objects	15	11	20	3
Sum of we	15	11	20	3
Within-cla	180.552	221.255	372.437	241.667
Minimum c	6.881	5.685	7.203	9.809
Average di	12.484	12.988	13.896	12.468
Maximum	17.423	23.940	65.965	15.585

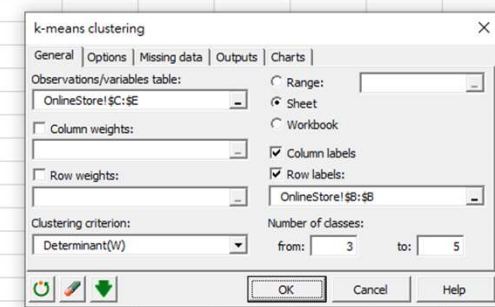
City List by Class:

Class	Cities
1	Albuquerque, Austin, Dallas, Denver, Fort Worth, Fresno, Houston, Long Beach, Los Angeles, NY, Phoenix, Sacramento, San Diego, San Jose, Tucson
2	Atlanta, Baltimore, Chicago, Cincinnati, Cleveland, Detroit, Memphis, New Orleans, Oakland, Philadelphia, St. Louis
3	Boston, Charlotte, Columbus, Honolulu, Indianapolis, Jacksonville, Kansas City, Las Vegas, Milwaukee, Minneapolis, Nashville, Oklahoma City, Omaha, Pittsburgh, Portland, San Francisco, Seattle
4	El Paso, Miami, San Antonio

Clustering / Segmentation(範例二)



A	B	C	D	E	F	G	H	I	J	K	L	M
	CustomerID	TotalSales	OrderCount	AvgOrderValue								
1	12346	77183.6	1	77183.6								
2	12347	4085.18	6	680.8633333								
3	12348	1797.24	4	449.31								
4	12349	1757.55	1	1757.55								
5	12350	334.4	1	334.4								
6	12352	2506.04	8	313.255								
7	12353	89	1	89								
8	12354	1079.4	1	1079.4								
9	12355	459.4	1	459.4								
10	12356	2811.43	3	937.1433333								
11	12357	6207.67	1	6207.67								
12	12358	484.86	1	484.86								
13	12359	6372.58	4	1593.145								
14	12360	2662.06	3	887.3533333								
15	12361	189.9	1	189.9								
16	12362	4697.19	9	521.91								
17	12363	552	2	276								
18	12364	1000.00	5	200.00								



Evolution of variances:			
variance\Class	4	5	6
Within-class	9,873,172.72	8,843,801.88	8,449,023.09
Between-class	61,536,875.05	62,566,245.90	62,961,024.68
Total	71410047.773	71410047.773	71410047.773

Central objects:

Class	TotalSales	OrderCount	AvgOrderValue
1 (14088)	50491.810	13.000	3883.985
2 (14223)	991.130	3.000	330.377
3 (13001)	9227.810	12.000	768.984
4 (17450)	192988.390	45.000	4288.631

What a beautiful world!

Principal Component Analysis

2012MLB - Excel

檔案 常用 插入 版面配置 公式 資料 校閱 檢視 開發人員 XLSTAT 告訴我您想要執行的動作...

Hsiu-Yuan TSAO 共用

F12 849

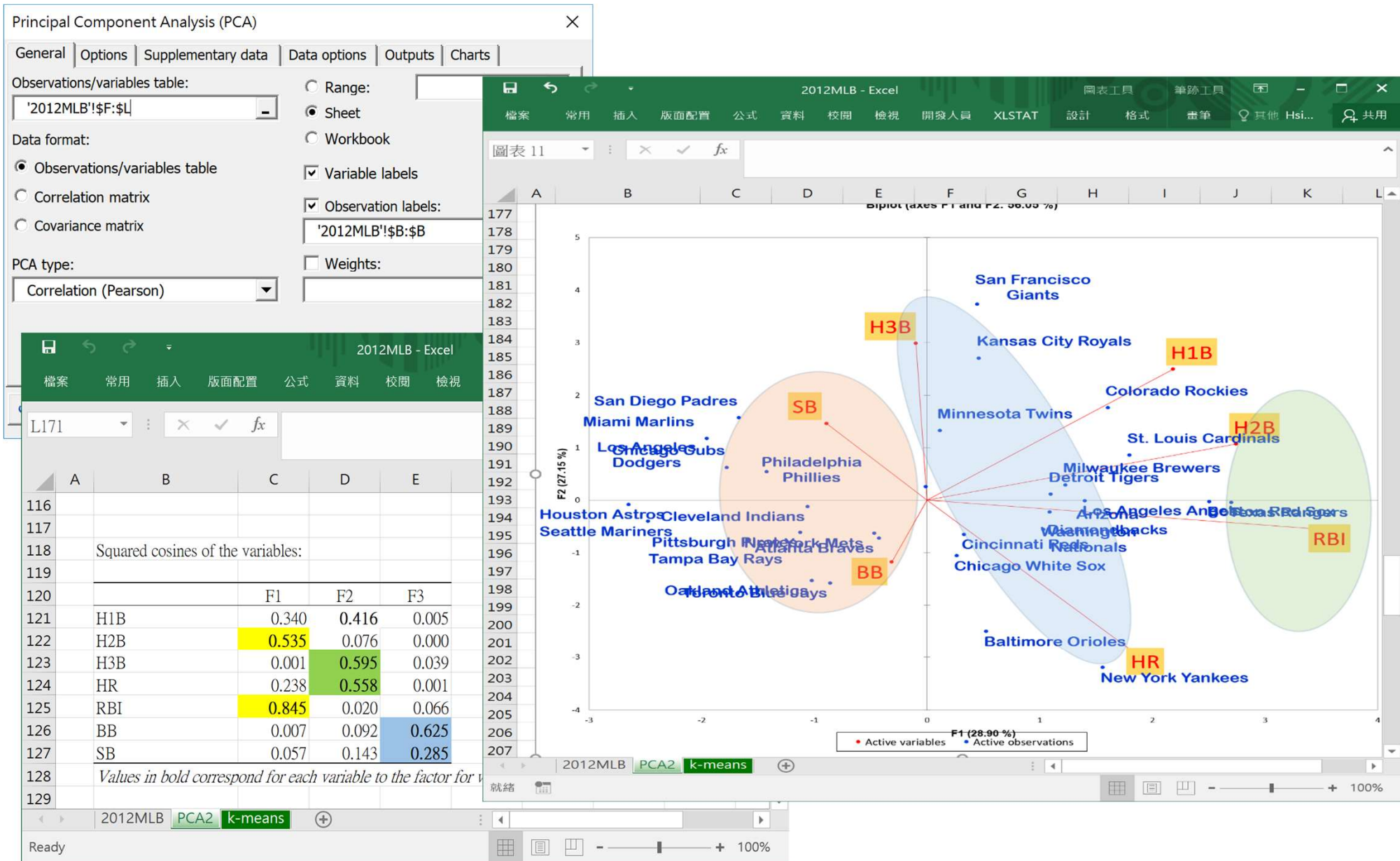
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1		Team	G	R	H	H1B	H2B	H3B	HR	RBI	BB	SB	SO	AVG	OBP			
2		1 Texas Rangers	152	764	1444	941	283	32	188	738	448	89	1030	0.275	0.335			
3		2 Los Angeles Angels	153	726	1424	972	252	20	180	691	423	125	1041	0.273	0.331			
4		3 Colorado Rockies	152	718	1425	932	286	49	158	680	432	96	1129	0.272	0.329			
5		4 St. Louis Cardinals	153	723	1447	983	277	35	152	691	503	89	1128	0.272	0.338			
6		5 San Francisco Giants	153	683	1419	997	273	54	95	642	453	111	1032	0.27	0.327			
7		6 Detroit Tigers	152	689	1379	922	267	37	153	663	490	53	1042	0.268	0.336			
8		7 Kansas City Royals	152	644	1412	971	282	37	122	613	380	126	958	0.267	0.319			
9		8 Boston Red Sox	154	716	1413	907	330	16	160	678	411	92	1129	0.264	0.319			
10		9 Minnesota Twins	153	673	1377	964	258	29	126	640	481	129	1004	0.262	0.326			
11		10 Washington Nationals	152	680	1377	894	281	23	179	640	448	95	1236	0.261	0.322			
12		11 New York Yankees	152	736	1345	849	259	12	225	710	523	90	1115	0.261	0.333			
13		12 Milwaukee Brewers	152	733	1362	854	284	36	188	699	447	150	1185	0.261	0.327			
14		13 Arizona Diamondbacks	152	693	1334	855	293	32	154	673	504	87	1190	0.261	0.329			
15		14 Philadelphia Phillies	153	652	1340	910	253	25	152	628	429	112	1021	0.256	0.317			
16		15 Chicago White Sox	152	702	1321	875	220	29	197	681	423	101	1124	0.255	0.317			
17		16 Cincinnati Reds	153	651	1315	834	284	28	169	618	462	87	1182	0.253	0.318			
18		17 New York Mets	152	612	1276	856	272	19	129	591	482	72	1186	0.25	0.317			
19		18 Atlanta Braves	153	672	1283	860	251	27	145	635	537	93	1215	0.249	0.321			
20		19 Cleveland Indians	153	620	1290	890	246	24	130	590	513	102	1005	0.249	0.321			
21		20 Los Angeles Dodgers	153	591	1270	894	252	21	103	563	462	97	1094	0.248	0.314			
22		21 San Diego Padres	153	617	1269	858	257	41	113	579	506	142	1173	0.248	0.32			
23		22 Baltimore Orioles	152	667	1296	826	260	16	194	635	456	51	1227	0.247	0.312			
24		23 Pittsburgh Pirates	152	622	1250	825	229	34	162	591	413	66	1271	0.245	0.306			
25		24 Miami Marlins	153	587	1260	843	250	37	130	555	458	144	1143	0.245	0.309			
26		25 Toronto Blue Jays	151	669	1242	804	231	21	186	631	439	121	1178	0.243	0.307			
27		26 Chicago Cubs	153	586	1238	821	251	36	130	545	413	92	1171	0.242	0.302			
28		27 Tampa Bay Rays	153	659	1229	802	241	29	157	628	539	131	1241	0.241	0.318			
29		28 Houston Astros	153	551	1211	821	232	27	131	514	443	103	1293	0.237	0.303			
30		29 Oakland Athletics	152	651	1223	767	249	29	178	622	515	117	1296	0.236	0.309			
31		30 Seattle Mariners	153	573	1206	824	218	26	138	543	436	101	1183	0.233	0.294			

2012MLB PCA2 k-means

就緒

下午 10:02 2018/10/25

Principal Component Analysis



Correspondence Analysis (CA) /Positioning

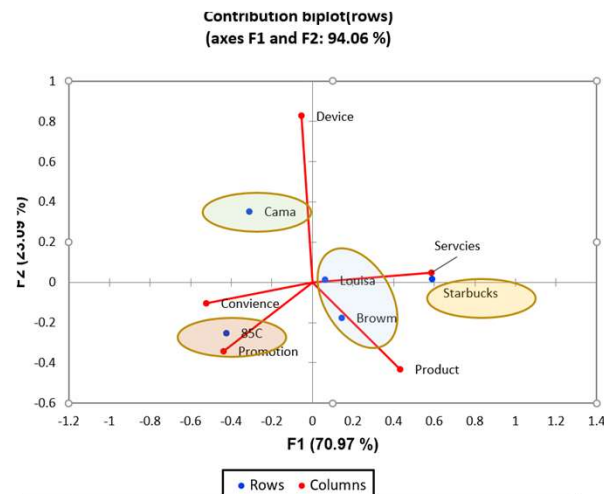
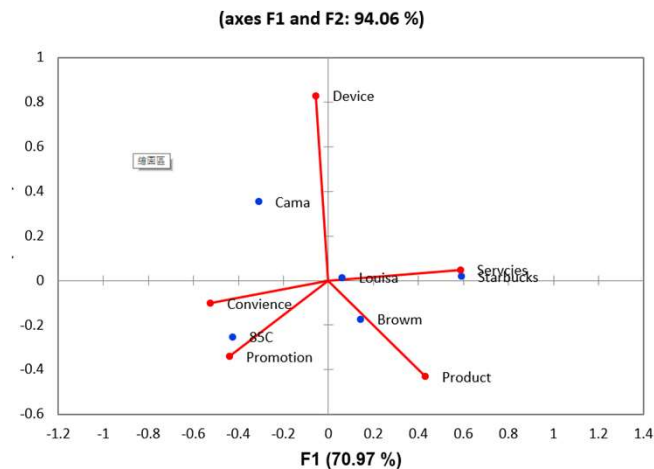
XLSTAT interface showing the Correspondence Analysis (CA) menu and the corresponding data table.

Brand	Product	Servcies	Device	Convience	Promotion
Starbucks	5	5	3	1	1
85C	3	1	2	5	5
Louisa	4	4	4	3	4
Browm	2	3	1	2	2
Cama	1	2	5	4	3

XLSTAT Menu: Analyzing data > Correspondence Analysis (CA)

Correspondence Analysis (CA) dialog box settings:

- Two-way table: CoffeeShop2!\$A\$1:\$F\$6
- Data format: Two-way table
- Labels included: ☒
- Weights: ☐

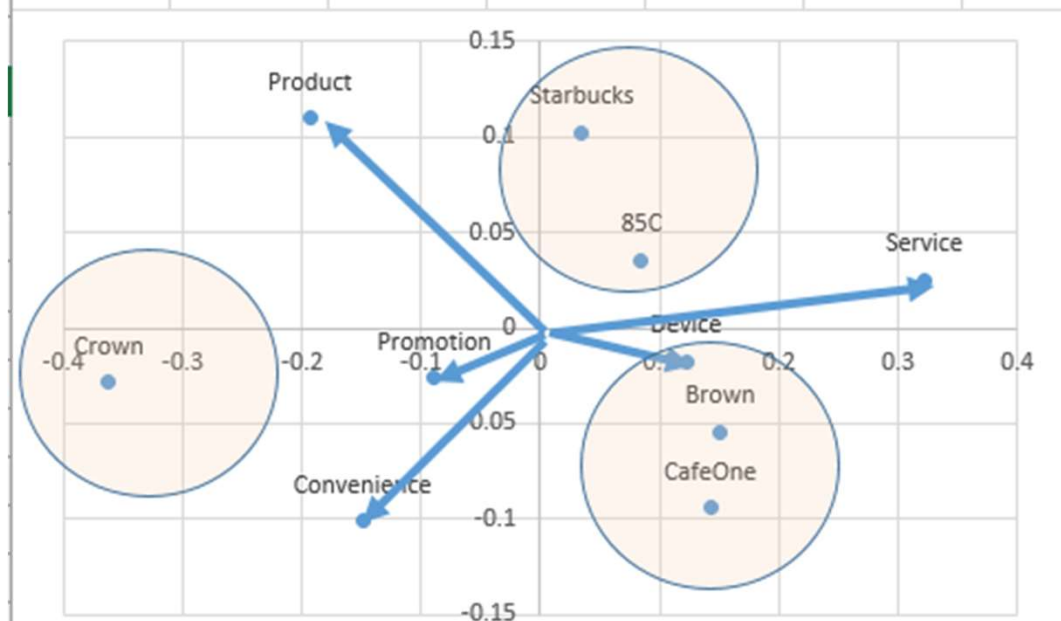


Mult iDimensional Scaling

Brand	Product	Servcies	Device	Convience	Promotion
Starbuck	4.2	3.92	3.98	3.68	4.07
85C	4.1	3.95	3.98	3.65	4.04
Crown	4.18	3.2	3.56	3.7	3.97
Browm	3.05	3.06	3.07	2.77	3.08
CafeOne	2.99	3	3	2.75	3

Configuration:

	Dim1	Dim2
Starbuck	0.605	0.817
85C	0.495	0.825
Crown	0.739	-0.010
Browm	-0.888	-0.755
CafeOne	-0.951	-0.877
Product	-0.531	0.099
Servcies	0.500	0.177
Device	0.139	0.122
Convience	0.174	-0.485
Promotion	-0.281	0.087



Step 1: Compute row and column averages

	Big	Athletic	Friendly	Trainab	Resourc	Animal	Lucky	Average
Dog	80	20	90	90	5	100	40	61
Cat	50	40	40	70	10	100	40	50
Rat	10	70	20	90	80	99	40	58
Cockroach	0	80	2	20	95	20	40	37
Wallaby	35	52	38	47	48	80	40	49
Average	35	52	38	63	48	80	40	51

$35 \times 61 / 52$

Step 2: Compute the expected values

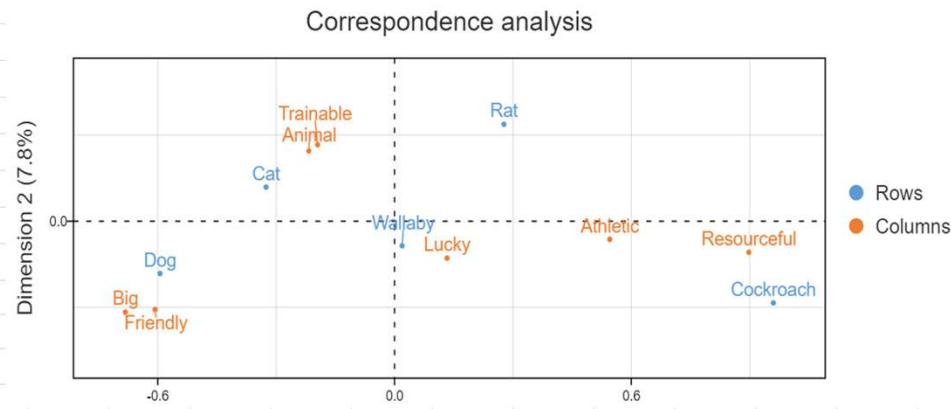
	Big	Athletic	Friendly	Trainab	Resourc	Animal	Lucky
Dog	42	63	45	76	57	95	48
Cat	34	51	37	62	47	78	39
Rat	40	60	44	73	55	92	46
Cockroach	25	38	27	46	34	58	29
Wallaby	33	50	36	61	45	76	38

$38 = 80 - 42$

Step 3: Compute the residuals

Dog	38	-43	45	14	-52	5	-8
Cat	16	-11	3	8	-37	22	1
Rat	-30	10	-24	17	25	7	-6
Cockroach	-25	42	-25	-26	61	-38	11
Wallaby	2	2	2	-14	3	4	2

Step 4: Plotting labels with similar residuals close together



Step 5: Interpreting the relationship between row and column labels

1. Look at the length of the line connecting the row label to the origin. Longer lines indicate that the row label is highly associated with some of the column labels (i.e., it has at least one high residual).
2. Look at the length of the label connecting the column label to the origin. Longer lines again indicate a high association between the column label and one or more row labels.
3. Look at the angle formed between these two lines. Really small angles indicate association. 90 degree angles indicate no relationship. Angles near 180 degrees indicate negative associations.



Multi Dimensional Scaling(MDS)

Describing data Visualizing data Analyzing data Modeling data

- Descriptive statistics
- Variables characterization
- Quantiles estimation
- Histograms
- Normality tests
- Resampled statistics
- Similarity/Dissimilarity matrices (correlation...)
- Biserial correlation
- Multicollinearity statistics
- Reliability analysis
- Contingency table (descriptive statistics)
- Pivot table

Brand	Product	Servcies	Device	Convience	Promotion
Starbuck	4.2	3.92	3.98	3.68	4.07
85C	4.1	3.95	3.98	3.65	4.04
Crown	4.18	3.2	3.56	3.7	3.97
Brown	3.05	3.06	3.07	2.77	3.08
CafeOne	2.99	3	3	2.75	3

Configuration:

	Dim1	Dim2
Starbuck	0.605	0.817
85C	0.495	0.825
	739	-0.010

Similarity/Dissimilarity matrices (correlation...)

General | Missing data | Outputs

Data: MDS!\$A\$1:\$A\$6

Data type: Quantitative

Range: MDS!\$M\$1

Sheet

Workbook

Column labels

Row labels: MDS!\$A:\$A

Proximity type: Euclidean distance

Similarities: Dissimilarities

Compute proximities for the: Columns Rows

OK Cancel Help

Proximity matrix (Euclidean distance):

	Product	Servcies	Device	Convience	Promotion
Product	0	1.030	0.669	0.916	0.256
Servcies	1.030	0	0.366	0.738	0.790
Device	0.669	0.366	0	0.609	0.424
Convience	0.916	0.738	0.609	0	0.732
Promotion	0.256	0.790	0.424	0.732	0



Multi Dimensional Scaling(MDS)

Analysing data ▾ Modeling data ▾ Machine learning ▾ Correlation/ Association tests ▾

- fa Factor analysis
- Principal Component Analysis (PCA)
- PCAmix
- Discriminant Analysis (DA)
- Correspondence Analysis (CA)
- Multiple Correspondence Analysis (MCA)
- MDS Multidimensional Scaling (MDS)**
- Principal Coordinate Analysis
- k-means clustering
- Agglomerative hierarchical clustering (AHC)
- Gaussian Mixture Models
- Univariate clustering

Proximity matrix (Euclidean distance):

	Product	Servcies	Device	Convience	Promotion
Product	0	1.030	0.669	0.916	0.256
Servcies	1.030	0	0.366	0.738	0.790
Device	0.669	0.366	0	0.609	0.424
Convience	0.916	0.738	0.609	0	0.732
Promotion	0.256	0.790	0.424	0.732	0

Multidimensional Scaling (MDS)

General | Options | Missing data | Outputs | Charts

Proximity matrix:

☐ Range:

☒ Sheet

☒ Dissimilarity ☐ Similarities

Select a proximity (dissimilarity or similarity) matrix

Model:

Labels included ☒

Weights: ☐

Dimensions: Minimum: Maximum:

	Dim1	Dim2
Product	0.356	-0.407
Servcies	-0.100	0.521
Device	0.035	0.181
Convience	-0.507	-0.095
Promotion	0.215	-0.200



Multi Dimensional Scaling(MDS)

Proximity matrix (Euclidean distance):

	Starbuck	85C	Crown	Browm	CafeOne
Starbuck	0	0.113	0.840	2.168	2.298
85C	0.113	0	0.868	2.102	2.233
Crown	0.840	0.868	0	1.787	1.901
Browm	2.168	2.102	1.787	0	0.137
CafeOne	2.298	2.233	1.901	0.137	0

	Dim1	Dim2
Starbuck	1.005	0.154
85C	0.961	0.053
Crown	0.336	0.658
Browm	-1.082	-0.434
CafeOne	-1.219	-0.431

Multidimensional Scaling (MDS)

×

General | Options | Missing data | Outputs | Charts

Proximity matrix:

MDS!\$M\$33:\$R\$38

☐ Range:
☐ Sheet
☐ Workbook

☒ Dissimilarities
☐ Similarities

Model:

Absolute

☒ Labels included
☐ Weights:

Dimensions:

Minimum:

2

Maximum:

2

OK

Cancel

Help

	Starbuck	85C	Crown	Browm	CafeOne
Starbuck	0	0.113	0.840	2.168	2.298
85C	0.113	0	0.868	2.102	2.233
Crown	0.840	0.868	0	1.787	1.901
Browm	2.168	2.102	1.787	0	0.137
CafeOne	2.298	2.233	1.901	0.137	0

	Dim1	Dim2
Starbuck	1.005	0.154
85C	0.961	0.053
Crown	0.336	0.658
Browm	-1.082	-0.434
CafeOne	-1.219	-0.431
Product	0.356	-0.407
Servcies	-0.100	0.521
Device	0.035	0.181
Convience	-0.507	-0.095
Promotion	0.215	-0.200

Multi Dimensional Scaling(MDS)

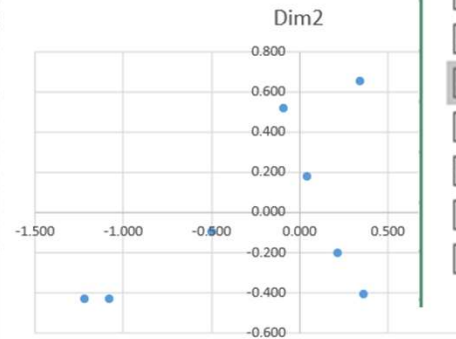
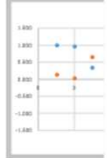
插入圖表

建議的圖表 所有圖表

- 最近
- 範本
- 直條圖
- 折線圖
- 圓形圖
- 橫條圖
- 區域圖
- XY 散佈圖
- 股票圖
- 曲面圖
- 雷達圖
- 矩形式樹狀結構圖
- 放射環狀圖
- 長條圖
- 盒狀圖
- 瀑布圖
- 組合式



散佈圖



圖表項目

- ☒ 座標軸
- ☐ 座標軸標題
- ☒ 圖表標題
- ☒ 資料標籤
- ☐ 誤差線
- ☒ 格線
- ☐ 圖例
- ☐ 趨勢線

資料標籤格式

標籤選項 文字選項

標籤選項

標籤包含

- ☒ 儲存格的值(F)
- ☐ 數列名稱(S)
- ☐ X 值(X)
- ☐ Y 值(Y)

	Dim1	Dim2
Starbuck	1.005	0.154
85C	0.961	0.053
Crown	0.336	0.658
Browm	-1.082	-0.434
CafeOne	-1.219	-0.431
Product	0.356	-0.407
Services	-0.100	0.521
Device	0.035	0.181
Convenience	-0.507	-0.095
Promotion	0.215	-0.200

XLSTAT 2020.5.1.1066

Data: Workbook = Corre

Row labels: Workbook =

Dissimilarity: Euclidean

Compute proximities for

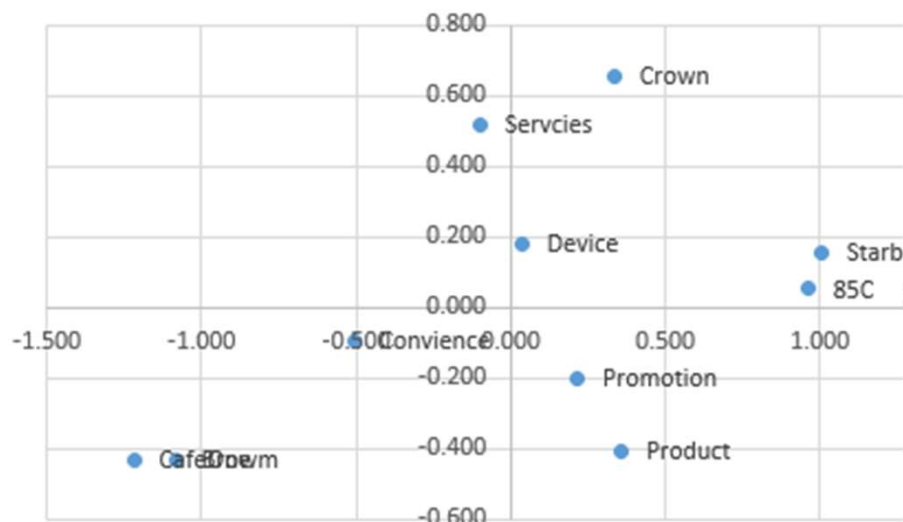
Summary statistics:

Variable Observationwi

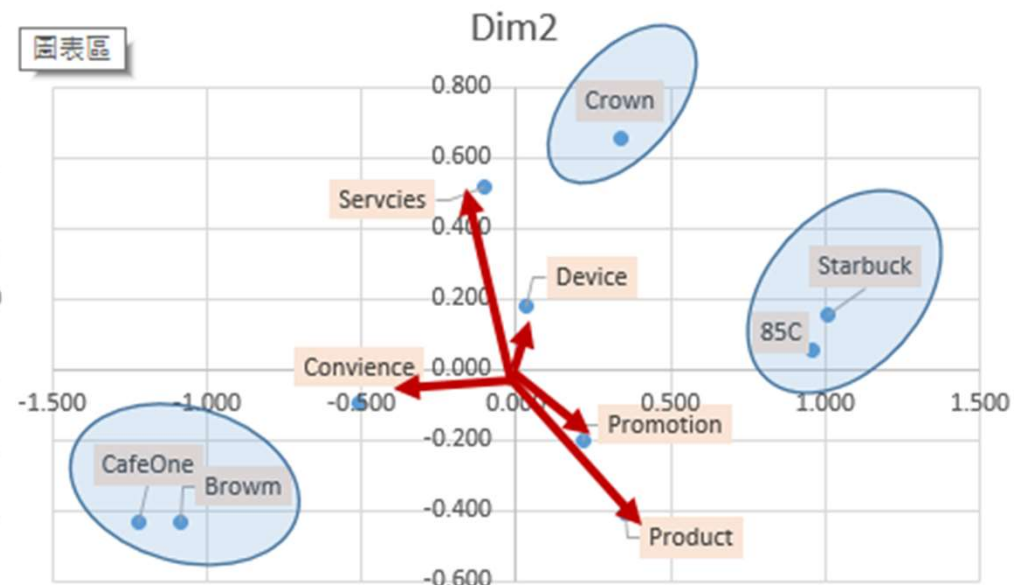
Product 5

資料標籤範圍 ? X
=MDS!\$I\$5:\$I\$14

Dim2



圖表區



Association Rule (關聯法則)

	柳橙汁	碳酸飲料	清潔劑	牛奶	洗衣精	底片	洋芋片
1	1	1	1	0	0	0	0
2	1	0	1	1	0	0	0
3	1	1	0	0	1	0	0
4	0	1	0	0	0	1	0
5	0	1	0	0	0	0	1

分析工具:

[XLSTA](#) ([example](#))

[WEKA](#) ([example](#))

R

Association Rule (關聯法則)

The screenshot shows the XLSTAT interface for performing Association Rule analysis. The main window displays a data table with 5 rows and 6 columns. The 'Association rules' dialog box is open, showing the 'General' tab with the data range 'Sheet1!\$C\$5:\$I\$10' and 'Contingency table' selected for data format. The 'Options' tab is also visible, showing 'Labels included' checked and minimum support/confidence values.

Association rules dialog box settings:

- Items: Sheet1!\$C\$5:\$I\$10
- Data format: Contingency table
- Labels included: ☒
- Minimum support: 0.3
- Minimum confidence: 1
- Min. number antecedent: 1

Summary of association rules:

Antecedent	Consequent	Confidence	Support	Lift
清潔劑	柳橙汁	1.000	0.400	1.667

Exact rules are displayed in bold

Matrix of influence:

	柳橙汁
清潔劑	1.000

Exact rules are displayed in bold

機器學習範例

<https://studio.azureml.net>

Azure Machine Learning Studio

The screenshot displays the Microsoft Machine Learning Studio (classic) interface. On the left, a sidebar contains navigation options: NEW, DATASET, MODULE, PROJECT PREVIEW, and EXPERIMENT. Below these, a search bar shows 'Association', and a list of 'Microsoft Samples' includes 'Discover Association Rules'. The main workspace is titled 'Association Rule' and shows a workflow diagram with two steps: 'Enter Data Manually' and 'Discover Association Rules'. The 'Discover Association Rules' step is highlighted with a red circle and the number '1'. A 'Mini Map' at the bottom left shows a smaller version of the workflow. On the right, the 'Properties' pane is open, showing settings for the 'Discover Association Rules' step. The 'Input Dataset Type' is set to 'Items Matrix'. The 'Columns' section shows 'Selected columns: All columns' and 'Column type: Categorical, All'. The 'Minimal Support' is set to 0.3, 'Minimal Confidence' is 1, 'Minimal Number of Items' is 2, and 'Maximal Number of Items' is 2. The 'Sort By' is set to 'Confidence'. The 'Left Hand Side' and 'Right Hand Side' are empty. The 'Prune Redundancies' checkbox is checked. At the bottom right, a 'Quick Help' section states: 'Discover relations between variables in dataset'.

NEW

DATASET

MODULE

PROJECT
PREVIEW

EXPERIMENT

FROM ZIP PACKAGE

Upload a new custom module from a local file

Association

Microsoft Machine Learning Studio (classic)

Search experiment items

Saved Datasets

My Datasets

BrestCancerDiagnosti...

BrestCancerDiagnosti...

BrestCancerDiagnosti...

BrestCancerDiagnosti...

CustomerValueClean....

mbti1300SENT.csv

mbti8Kboth.csv

OnlineRetail.csv

WA_Fn-UseC_-Marke...

Samples

Adult Census Income...

Airport Codes Dataset

Automobile price dat...

Bike Rental UCI dataset

Bill Gates RGB Image

Blood donation data

Book Reviews from A...

Brest cancer data

Association Rule

In draft

Draft saved at 下午6:38:24

Enter Data Manually

Discover Association Rules

Mini Map

Enter Data Manually

Discover Association Rules

Properties

Project

Input Dataset Type

Items Matrix

Columns

Selected columns:

All columns

Column type:

Categorical, All

Launch column selector

Minimal Support

0.3

Minimal Confidence

1

Minimal Number of It...

2

Maximal Number of It...

2

Sort By

Confidence

Left Hand Side

Right Hand Side

☒ Prune Redundancies

Quick Help

Discover relations between variables in dataset

Recommended System KNN推薦系統

Customer	Age	Income	Purchase Product
1	45	46000	Book
2	39	100000	TV
3	35	38000	DVD
4	69	150000	Car Cover
5	58	51000	???

Step 1: Determine the Distance Formula

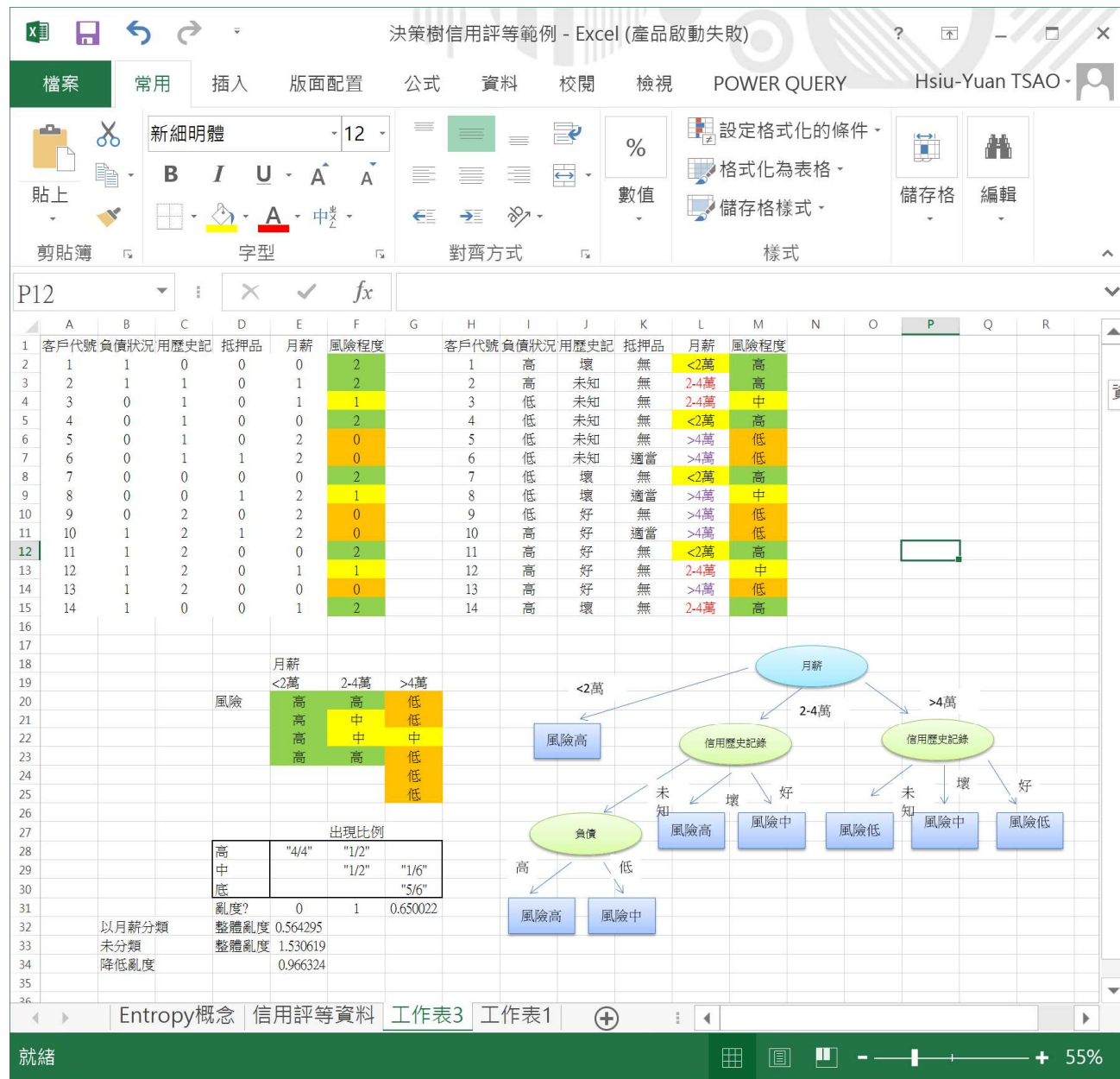
Distance= SQRT(((58- AGE)/(69-35))^2) + (51000- Income) / (150000 - 38000))^2)

Step 2: Calculate the Score				
Customer	Age	Income	Purchase Product	
1	0.38495	46	Book	1
2	0.709711	100	TV	3
3	0.686356	38	DVD	2
4	0.941276	150	Car Cover	4
5	0	51	???	

[KNN Example](#)

Decision Tree (決策樹)

信用評等範例



資訊與火商



對大多數的讀者來說，熱力學是一門抽象難懂的學問，有關熱力學難懂的笑話很多，最常被引用的一段是：
「三個話題：政治、宗教及熱力學是不宜與朋友談論的。」

	$a^x=b$	$\log(x,b)$				
	10^2	2				
	100					
x	0.00001	0.25	0.5	0.75	1	
$\log(x,2)$	-16.60964047	-2	-1	-0.41504	0	
公正銅板出現正反機率為1/2			不公正銅板出現正機率為3/4			
正	"(1/2)*LOG(1/2,2)"	0.5	正	"(3/4)*LOG(3/4,2)"	0.311278	
反	"(1/2)*LOG(1/2,2)"	0.5	反	"(1/4)*LOG(1/4,2)"	0.5	
Entropy		1	Entropy		0.811278	
何者亂度較大?						
何謂亂度?						

不公正銅板出現正機率為100%		
正	"(1)*LOG(1,2)"	0
反	"(0)*LOG(0,2)"	0
Entropy		0

客戶代號	負債狀況	信用歷史記錄	抵押品	月薪	風險程度	風險程度							
						高	中	低					
1	高	壞	無	<2萬	高	"6/14"	"3/14"	"5/14"	1.00				
2	高	未知	無	2-4萬	高	0.43	0.21	0.36					
3	低	未知	無	2-4萬	中								
4	低	未知	無	<2萬	高	Entropy=?							
5	低	未知	無	>4萬	低	0.52388247	0.476226947	0.53051	1.530619	亂度大嗎?	何謂亂度?		
6	低	未知	適當	>4萬	低								
7	低	壞	無	<2萬	高								
8	低	壞	適當	>4萬	中								
9	低	好	無	>4萬	低								
10	高	好	適當	>4萬	低								
11	高	好	無	<2萬	高								
12	高	好	無	2-4萬	中								
13	高	好	無	>4萬	低								
14	高	壞	無	2-4萬	高								

客戶代號	負債狀況	用歷史記	抵押品	月薪	風險程度	客戶代號	負債狀況	用歷史記	抵押品	月薪	風險程度
1	1	0	0	0	2	1	高	壞	無	<2萬	高
2	1	1	0	1	2	2	高	未知	無	2-4萬	高
3	0	1	0	1	1	3	低	未知	無	2-4萬	中
4	0	1	0	0	2	4	低	未知	無	<2萬	高
5	0	1	0	2	0	5	低	未知	無	>4萬	低
6	0	1	1	2	0	6	低	未知	適當	>4萬	低
7	0	0	0	0	2	7	低	壞	無	<2萬	高
8	0	0	1	2	1	8	低	壞	適當	>4萬	中
9	0	2	0	2	0	9	低	好	無	>4萬	低
10	1	2	1	2	0	10	高	好	適當	>4萬	低
11	1	2	0	0	2	11	高	好	無	<2萬	高
12	1	2	0	1	1	12	高	好	無	2-4萬	中
13	1	2	0	0	0	13	高	好	無	>4萬	低
14	1	0	0	1	2	14	高	壞	無	2-4萬	高

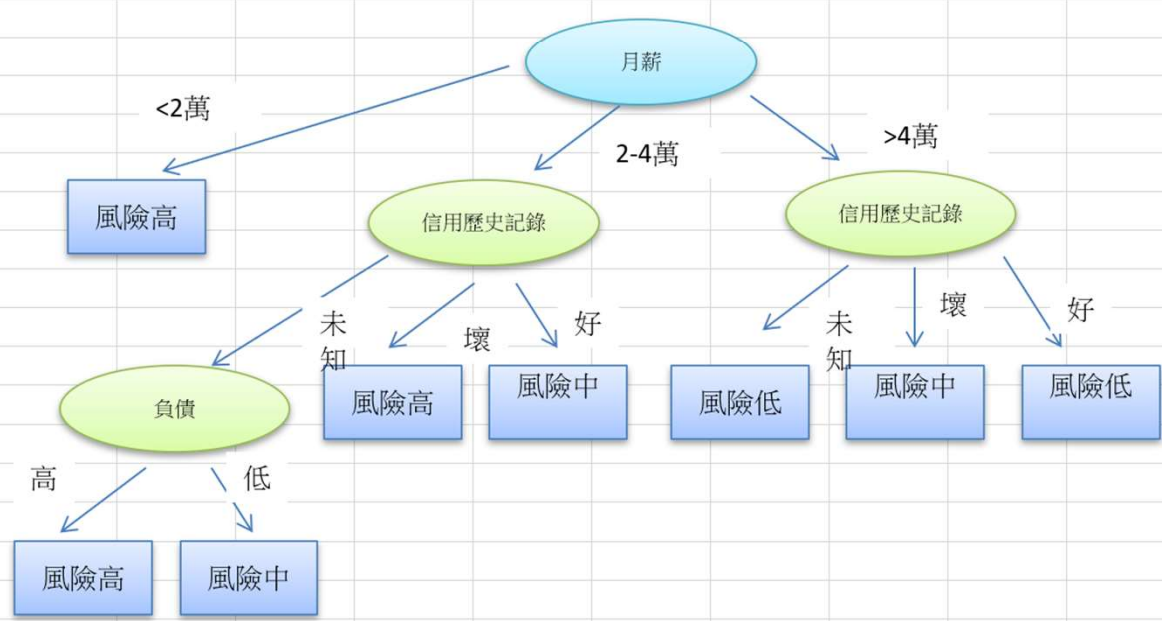


月薪	<2萬	2-4萬	>4萬
風險	高 高 高 高	高 中 中 高	低 低 中 低 低 低

出現比例			
高中底	"4/4"	"1/2"	"1/6"
亂度?	0	1	0.650022

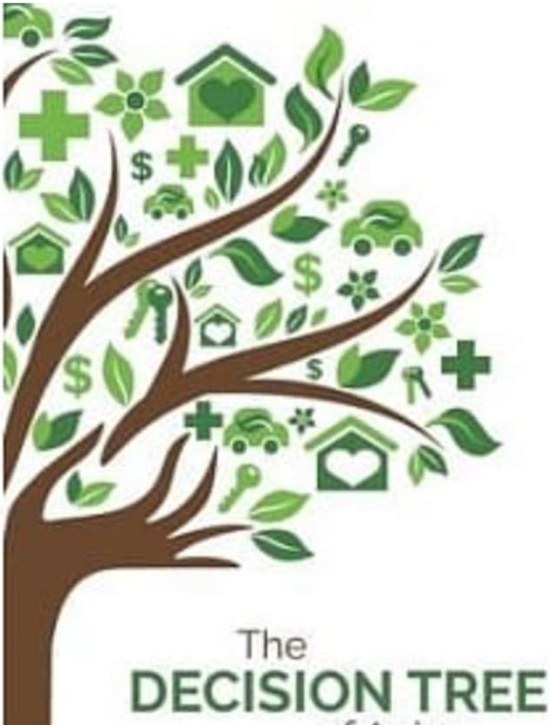
以月薪分類
未分類
降低亂度

整體亂度	0.564295
整體亂度	1.530619
降低亂度	0.966324



常見的決策樹演算法比較

演算法	資料屬性	分割規則	修剪樹規則
ID3	離散型	Entropy, Gain Ratio	Predicted Error Rate
C4.5	離散型	Gain Ratio	Predicted Error Rate
CHAID	離散型	Chi-Square Test	No Pruning
CART	離散與連續型	Gini Index	Entire Error Rate (Training and Predicted)



To give a proper background for `rpart` package and `rpart` method with `caret` package:

1. If you use the `rpart` package directly, it will construct the complete tree by default. If you want to prune the tree, you need to provide the optional parameter `rpart.control` which controls the fit of the tree. R documentation below, eg.:

```
rpart(formula, data, method, control = prune.control)

prune.control = rpart.control(minsplit = 20,
                              minbucket = round(minsplit/3), cp = 0.01,
                              maxcompete = 4, maxsurrogate = 5, usesurrogate = 2,
                              xval = 10, surrogatestyle = 0, maxdepth = 30 )
```

these are the hyper parameters you can tune to obtain a pruned tree.

One followed way is to not provide the `cp` i.e complexity parameter and perform cross validation (`xval`), something like:

```
rpart.control(minsplit = 20, minbucket = round(minsplit/3), xval = 10)
```

complexity parameter (`cp`) can be thought of as a measure of *complexity/ no of splits* of your model and you want to increase complexity until your model generalizes to new observations. i.e. *regularization*

Therefore evaluate the *cross validated error vs cp* and choose the `cp` that gives the good value (`cp_good`).

Finally, add it as your control parameter i.e. `rpart.control(cp = cp_good)` or use the `prune` function i.e. `prune(fit, cp = cp_good)` to get the desired tree.

2. `caret` package on the other hand already implements the `rpart` method with `cp` as the tuning parameter. `caret` by default will prune your tree based on a default run it makes on a default parameter grid (even if you don't supply any `tuneGrid` and `trControl` while training your model:

```
model <- train(data,
               labels,
               method = "rpart")
```

決策樹挑選變數常用的測量值

常見的資訊量（衡量資料純度）：

- Entropy (熵):

$$I_H(t) = - \sum_{i=1}^c p(i|t) \log_2 p(i|t)$$

其中，H代表Homogeneity(同質性)。

當Entropy=0表示completely homogeneous(pure)，而當Entropy=1則表示資料為50%-50%之組成，是不純的(impurity)。

- Gini Impurity (Gini不純度):

$$I_G(t) = \sum_{i=1}^c p(i|t)(1 - p(i|t)) = 1 - \sum_{i=1}^c p(i|t)^2$$

其中，G則代表Gini Impurity。

- Information Gain (資訊增益): 則衡量節點切割前後資料純度的變化。節點的選擇，當選IG值越大的變數為佳。

$$IG = Info(D) - Info_A(D)$$

其中， $Info(D)$ 為原始資料純度，而 $Info_A(D)$ 則表示使用A規則切割後的資料純度。

$$Info_A(D) = \sum_{j=1}^m \frac{N_j}{N_p} Info(D_j)$$

當m=2，即為二元分類時，

$$IG = Info(D) - \frac{N_{left}}{N_p} Info(D_{left}) - \frac{N_{right}}{N_p} Info(D_{right})$$

There are two main ways to do hyper parameter tuning using the `train()` :

1. Set the `tuneLength`
2. Define and set the `tuneGrid`

Grid Search: Automatic Grid

There are two ways to tune an algorithm in the Caret R package, the first is by allowing the system to do it automatically. This can be done by setting the `tuneLength` to indicate the number of different values to try for each algorithm parameter.

This only supports integer and categorical algorithm parameters, and it makes a crude guess as to what values to try, but it can get you up and running very quickly.

The following recipe demonstrates the automatic grid search of the size and k attributes of LVQ with 5 (`tuneLength=5`) values of each (25 total models).

```
1 # ensure results are repeatable
2 set.seed(7)
3 # load the library
4 library(caret)
5 # load the dataset
6 data(iris)
7 # prepare training scheme
8 control <- trainControl(method="repeatedcv", number=10, repeats=3)
9 # train the model
10 model <- train(Species~., data=iris, method="lvq", trControl=control, tuneLength=5)
11 # summarize the model
12 print(model)
```

The final values used for the model were size = 10 and k = 1.

Grid Search: Manual Grid

The second way to search algorithm parameters is to specify a tune grid manually. In the grid, each algorithm parameter can be specified as a vector of possible values. These vectors combine to define all the possible combinations to try.

The recipe below demonstrates the search of a manual tune grid with 4 values for the size parameter and 5 values for the k parameter (20 combinations).

```
Grid search with the caret r package
2 set.seed(7)
3 # load the library
4 library(caret)
5 # load the dataset
6 data(iris)
7 # prepare training scheme
8 control <- trainControl(method="repeatedcv", number=10, repeats=3)
9 # design the parameter tuning grid
10 grid <- expand.grid(size=c(5,10,20,50), k=c(1,2,3,4,5))
11 # train the model
12 model <- train(Species~., data=iris, method="lvq", trControl=control, tuneGrid=grid)
13 # summarize the model
14 print(model)
```

The final values used for the model were size = 50 and k = 5.



機器學習演算法資料前處理心法

To increase the predictive power by **modifying variables** or **creating new variables** from existing variables.

This specific task is called **Feature Engineering**.

This is perhaps the single most important activity in the entire process and decides the success or failure of a classification problem.

This part can not be put into any template as it varies on a case by case basis and one requires a lot of experience and domain expertise to gain a mastery of this skill.



機器學習演算法資料前處理心法

－ 訓練與測試集

```
#===== 產生 訓練資料及測試資料集 =====  
# Create the training and test datasets  
set.seed(100)
```

```
# Step 1: Get row numbers for the training data  
trainRowNumbers <- createDataPartition(credit$Response, p=0.8, list=FALSE)
```

```
# Step 2: Create the training dataset  
trainData <- credit[trainRowNumbers, -1]
```

```
# Step 3: Create the test dataset  
testData <- credit[-trainRowNumbers, -1]
```

```
# Store X (自變數) and Y (依變數) for later use.  
x = trainData[, -4]  
y = trainData$Response
```

```
library(caTools)
```

```
sample <- sample.split(df$Customer, splitRatio = .7)
```

```
trainX <- as.matrix(subset(encodedDF, sample == TRUE))  
trainY <- as.double(as.matrix(subset(df$Engaged, sample == TRUE)))
```

```
testX <- as.matrix(subset(encodedDF, sample == FALSE))  
testY <- as.double(as.matrix(subset(df$Engaged, sample == FALSE)))
```

```
set.seed(123)  
train_sample <- sample(1000, 900)  
credit_train <- credit[train_sample, ]  
credit_test <- credit[-train_sample, ]
```



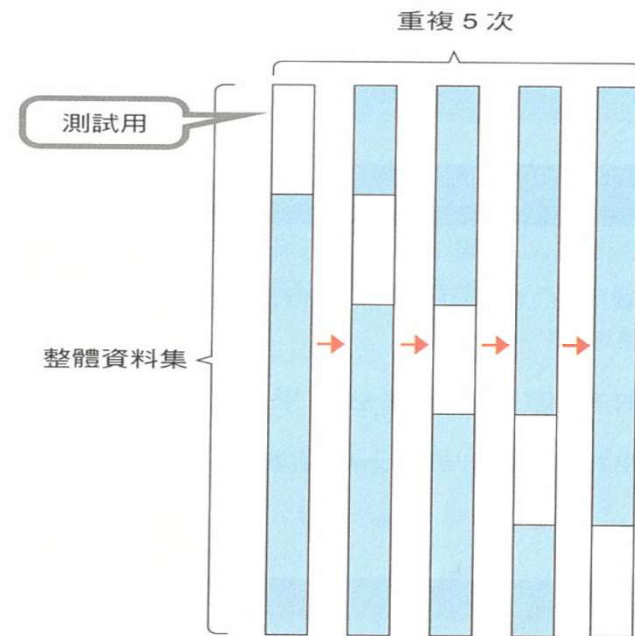
機器學習演算法資料前處理心法

– 交叉驗證(Cross Validation)

```
#-----  
#我們爲了避免模型過度擬合(overfitting)，故要利用k-fold Cross-Validation#  
#的方法進行交叉驗證，我們使用caret這個套件，而K先設定為10次~  
#-----
```

```
train_control <- trainControl(method="cv", number=10)  
train_control.model <- train(default~.,  
                             data=credit_train,  
                             method="C5.0",  
                             trControl=train_control)
```

```
train_control.model
```



▲ 圖 4.1.6 交叉驗證



機器學習演算法資料前處理心法

Categorical

Some algorithms (ex: decision tree) can work with categorical data directly.

Many machine learning algorithms cannot operate on label data directly. They require all input variables and output variables to be numeric.

Logistic regression will not accept categorical variables.

For this we need to transform those variables using “one hot encoding”
In traditional term used in Statistics named “**Dummy Variable**”

Numerical

It is always advisable to standardize the numeric variables before feeding them to a **PCA algorithm/Kmean** or a **Neural Network** model

(Efficient vs. Limitation)



One-Hot-Encoding– option 1

```
# Using model.matrix
new_df <- df
new_df$city <- factor(new_df$city, exclude = NULL)
new_df$hml <- factor(new_df$hml, exclude = NULL)
new_df <- model.matrix(~.-1, data = new_df[, c("city", "hml")],
                      contrasts.arg = list(
                        city = contrasts(new_df$city, contrasts = FALSE),
                        hml = contrasts(new_df$hml, contrasts = FALSE)
                      ))

head(new_df[, 1:3])
```

	cityAlameda	cityAmes	cityApple Valley
## 1	0	0	0
## 2	0	0	0
## 3	0	0	0
## 4	0	0	0
## 5	0	0	0
## 6	0	0	0



One-Hot-Encoding option 2

```
library(caret)
## Loading required package: lattice
## Loading required package: ggplot2
new_df <- df
new_df$city <- factor(new_df$city, exclude = NULL)
new_df$hml <- factor(new_df$hml, exclude = NULL)
new_df$city <- addNA(new_df$city)
new_df$hml <- addNA(new_df$hml)
dv <- caret::dummyVars("~ city + hml", data = new_df)
new_df <- data.frame(predict(dv, newdata = df))
head(new_df[, 1:3])
##      city.Alameda city.Ames city.Apple.Valley
## 1              0              0              0
## 2              0              0              0
## 3              0              0              0
## 4              0              0              0
## 5              0              0              0
## 6              0              0              0
detach("package:caret", unload = TRUE)
```




One-Hot-Encoding option 3

```
library(vtreat)
tz <- vtreat::designTreatmentsZ(df, c("city", "hml"))
## [1] "vtreat 1.5.1 inspecting inputs Mon Feb 03 00:16:33 2020"
## [1] "designing treatments Mon Feb 03 00:16:33 2020"
## [1] " have initial level statistics Mon Feb 03 00:16:33 2020"
## [1] " scoring treatments Mon Feb 03 00:16:33 2020"
## [1] "have treatment plan Mon Feb 03 00:16:34 2020"
new_df <- vtreat::prepare(tz, df, extracols = "target")
head(new_df[, 1:5])
##      city_catP hml_catP city_lev_NA city_lev_x_Alameda city_lev_x_Ames
## 1      0.0090   0.3174           0                0           0
## 2      0.0094   0.3090           0                0           0
## 3      0.0106   0.3090           0                0           0
## 4      0.0116   0.3236           0                0           0
## 5      0.0106   0.3236           0                0           0
## 6      0.0094   0.3090           0                0           0
detach("package:vtreat", unload = TRUE)
```



standardize the numeric variables

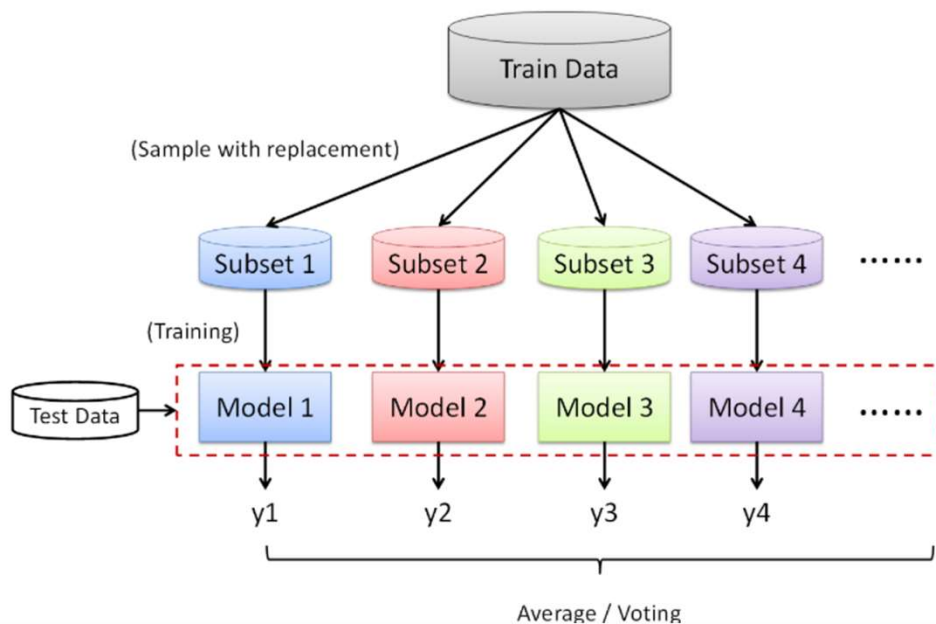
```
rankDF <- customerDF %>%  
  mutate(TotalSales=rank(TotalSales),  
         OrderCount=rank(OrderCount, ties.method="first"),  
         AvgOrderValue=rank(AvgOrderValue))
```

```
normalizedDF <- rankDF %>%  
  mutate(TotalSales=scale(TotalSales),  
         OrderCount=scale(OrderCount),  
         AvgOrderValue=scale(AvgOrderValue))
```

Ensemble Machine Learning

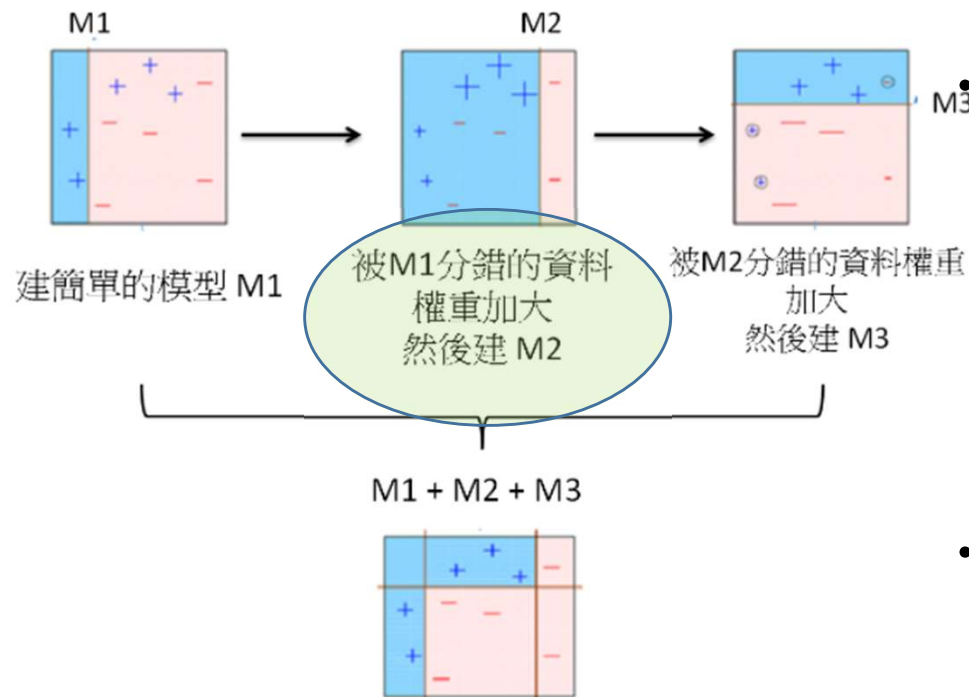
集成學習

Bagging



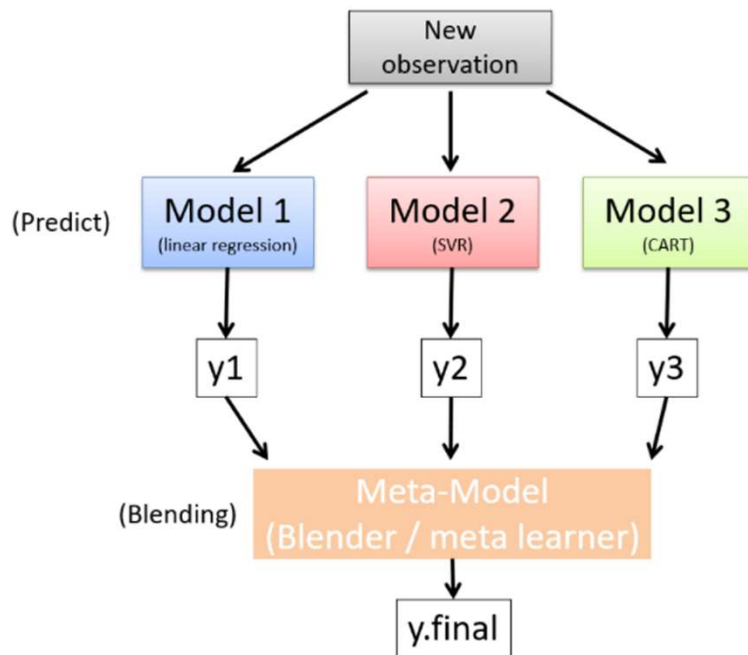
- Bootstrap aggregating (Bagging) , 從字面上來看，就是將資料裝成一個袋子一個袋子(Bag)，然後將每個袋子的結果結合在一起。
- 演算法上，是將樣本重複抽樣(取後放回)，產生多個子資料集(Subsets)後，依序建立多個模型，最後再將所有模型的結果彙整在一起。

Boosting



- 「三個臭皮匠，勝過一個諸葛亮」
- M1 M2 M3 如果太複雜(太強)，那彼此之間就會互相干擾，影響最後預測/分類結果；唯有彼此都是「弱模型」，才能好好專注在自己本身的預測/分類，然後再把彼此的成果結合一起，這就是 Boosting 的概念。
- XGboost 被譽為是「Kaggle 神器」。原因無他，就是因為每年的得名隊伍，使用 XGboost 的團隊基幾乎大多數！

Stacking



- 「在訓練**多個模型**、得到多個預測值/分類結果後，與其使用投票法(hard voting)或平均法(average)將這些結果整合(ensemble)起來，為何不多訓練一個模型來做這樣的整合呢？」

How to Discovery Insight via Text Mining



行銷理論、商業與人生
學術與實務的對話

曹修源 教授

04/2021

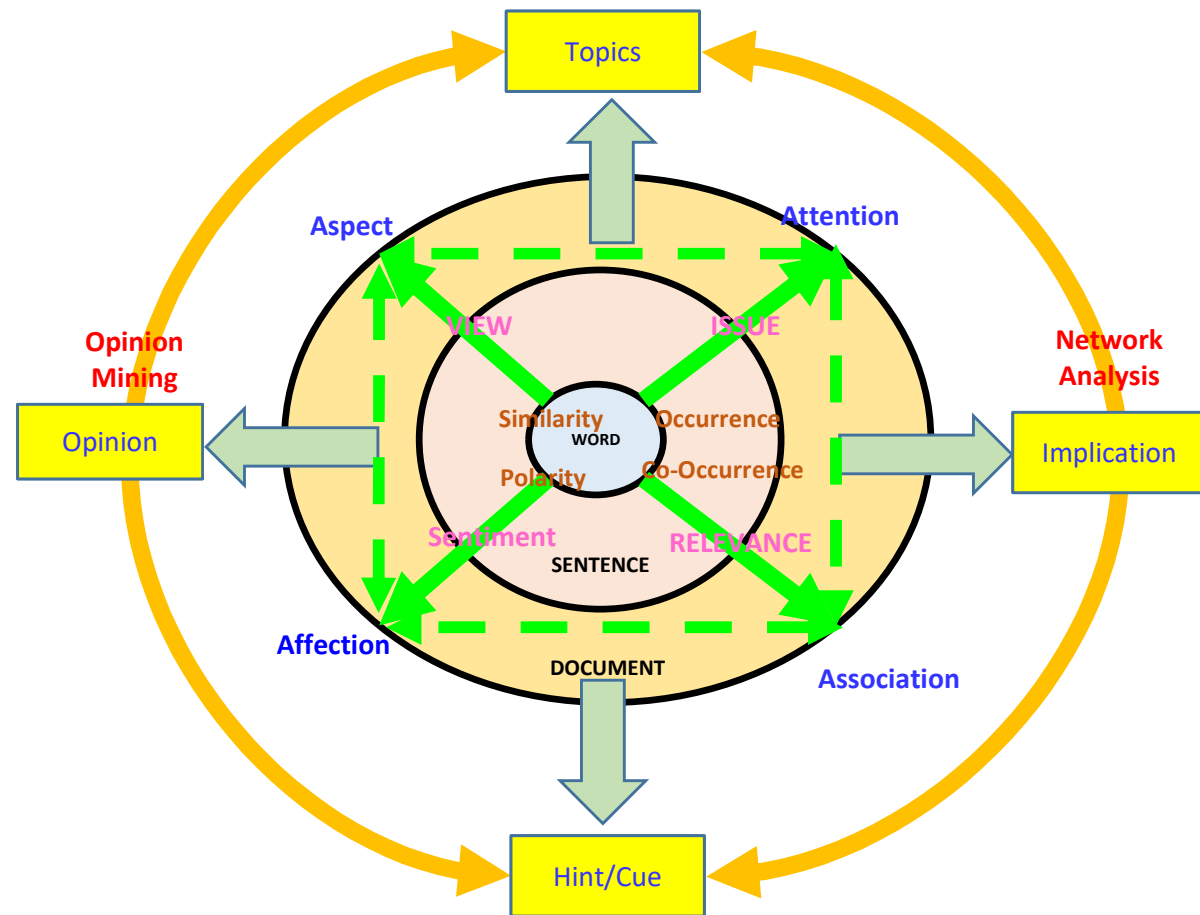
- Concept, Technology, and Application

 **NCHU** 國立中興大學行銷學系
Marketing Department of Marketing, National Chung Hsing University
國立中興大學行銷學系

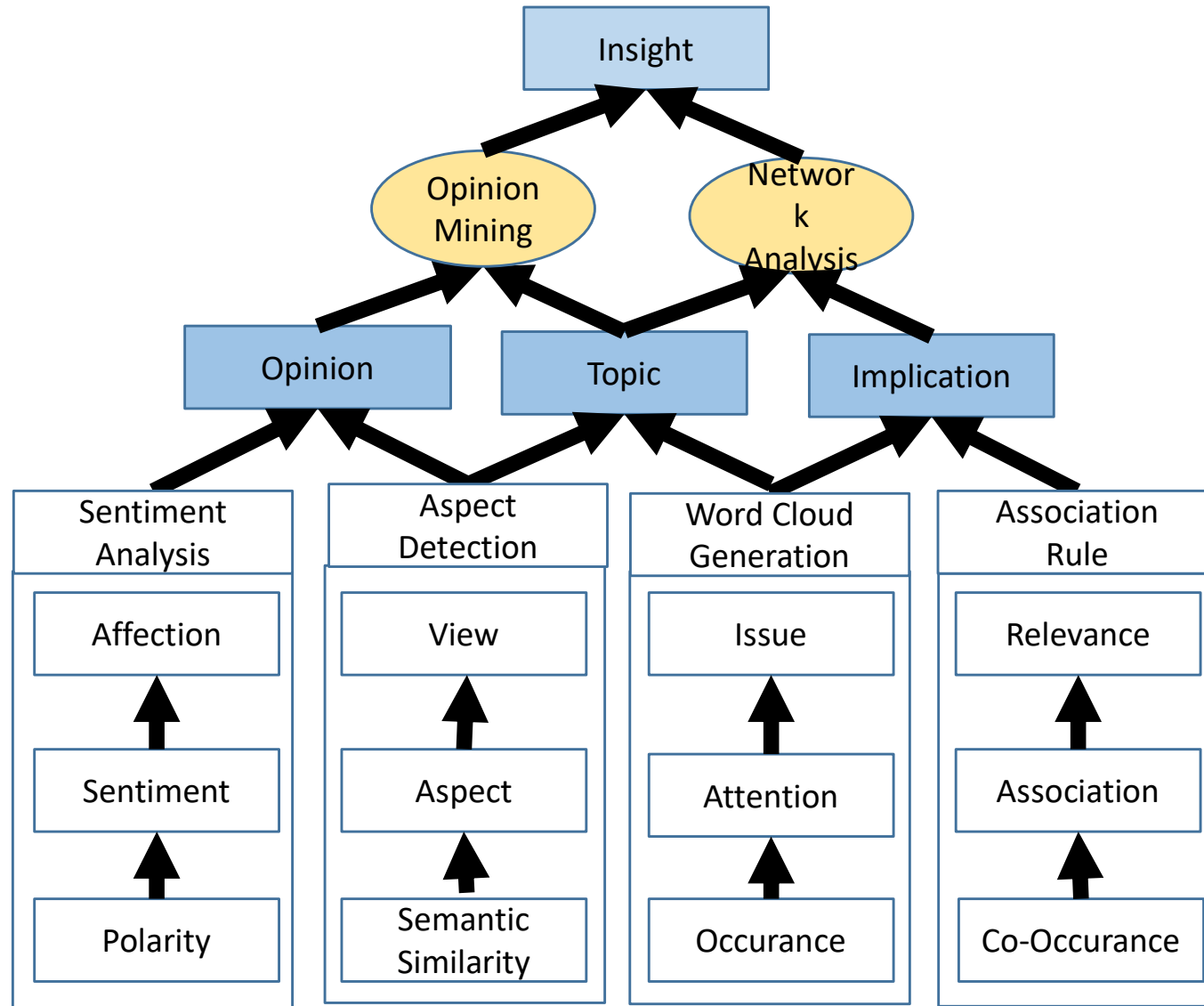




Framework of Textual Analytics



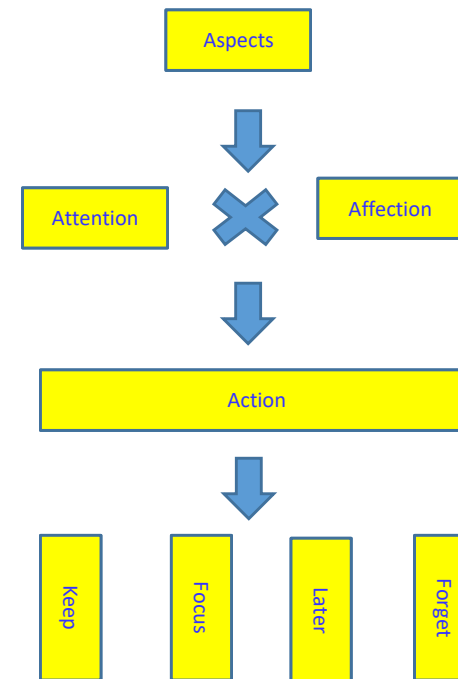
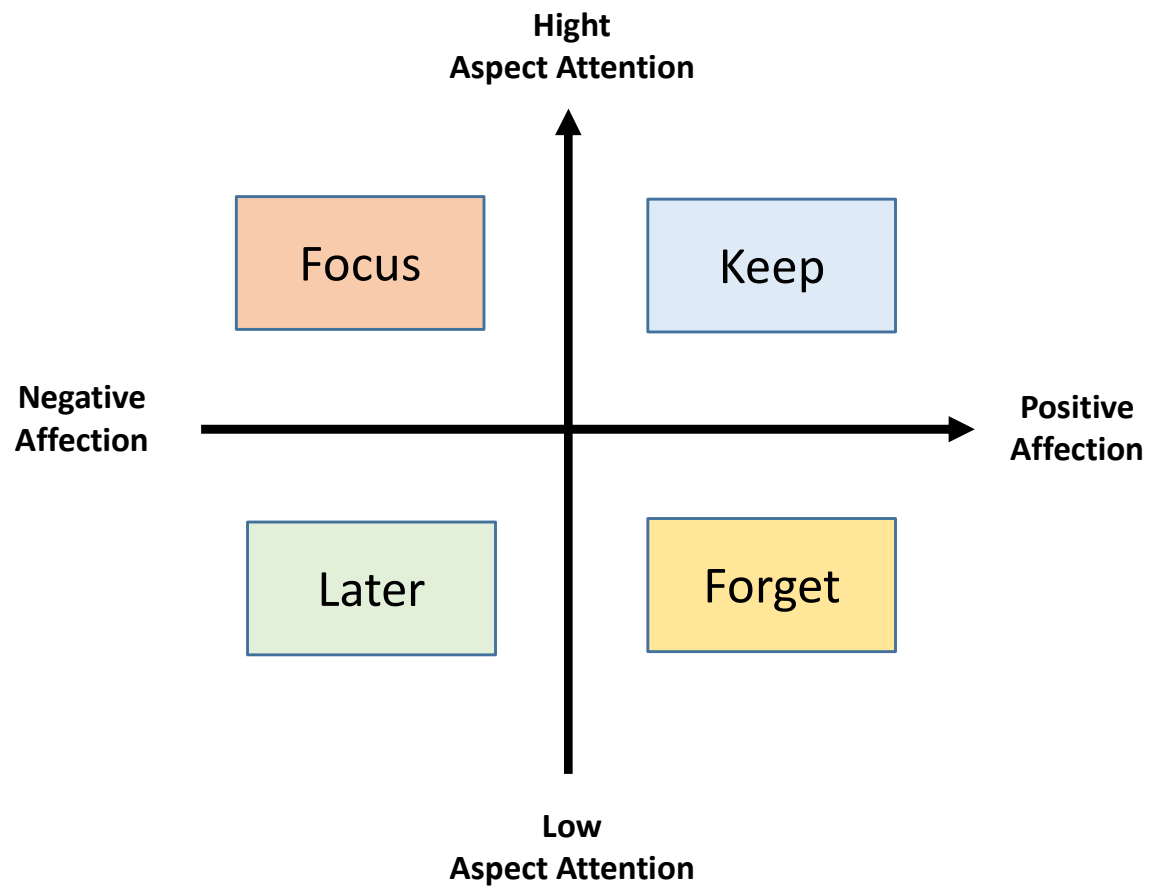
The Structural Textual Data Analytics





From Text to Action(FTTA3A)

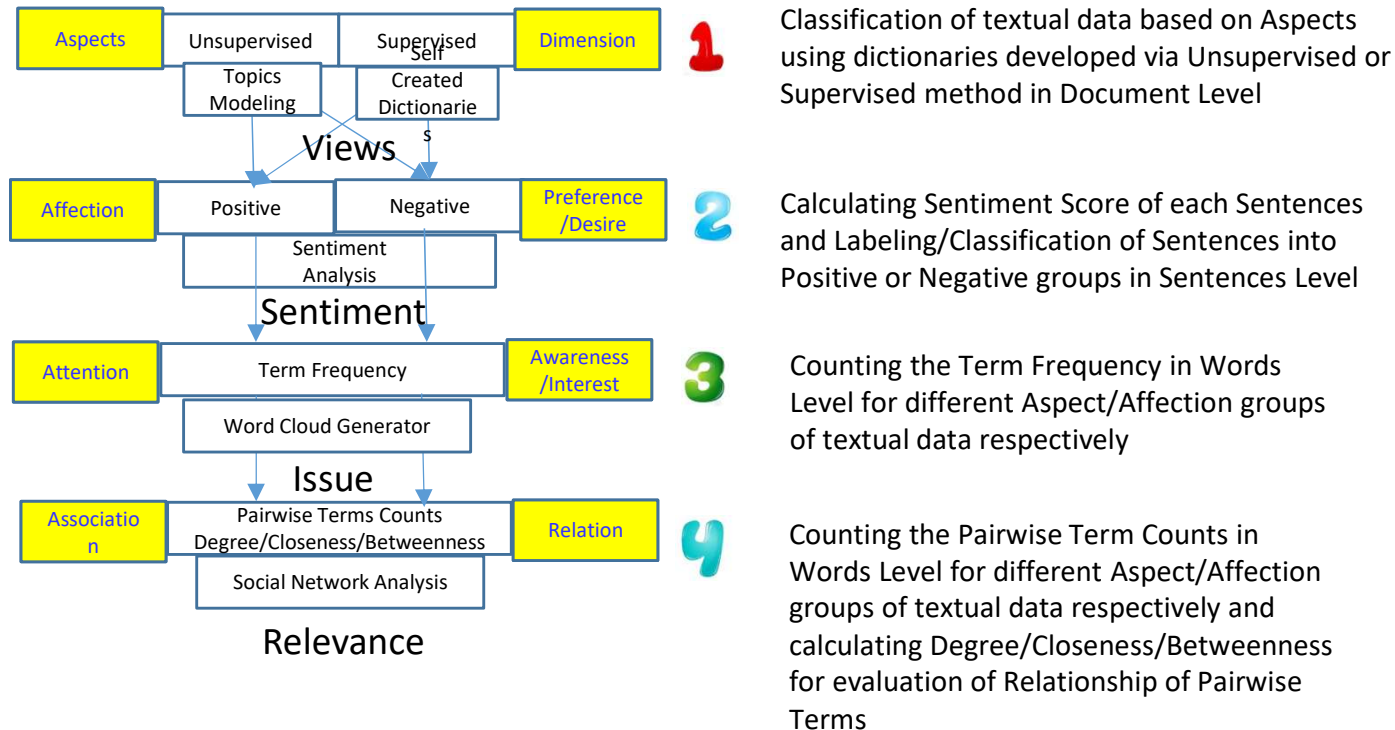
-- Aspect, Affection and Attention





Structural Textual Data Discovery (STDD4A)

-- Aspects, Affection, Attention, and Association



Topics Modeling

Aspect Semi-Auto Detention:

Ex: Sports, Science,
Food Topics

Self Created Dictionaries

Existed Dimensions of
Marketing Scale:

Ex: Competence, Excitement , Ruggedness,
Sincerity, Sophistication of
Brand Personality



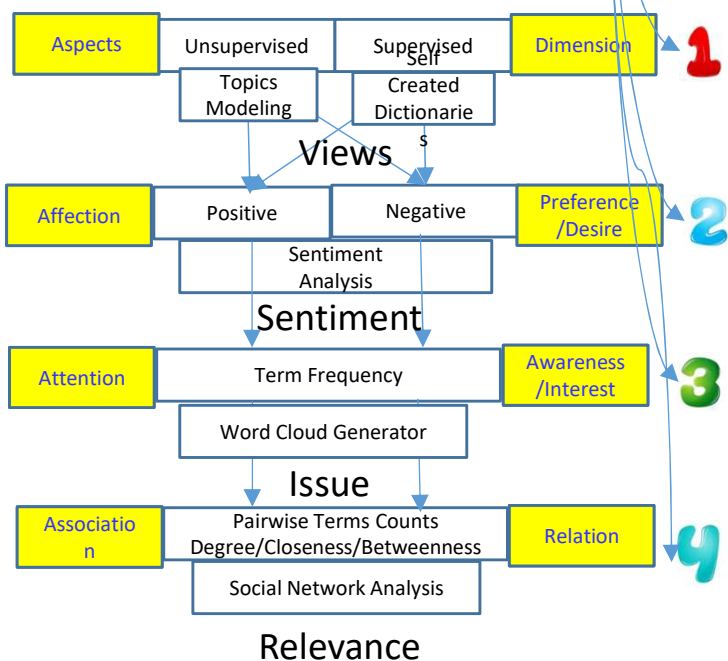
Structural Textual Data Discovery (STDD4A)

-- Aspects, Affection, Attention, and Association

Los Angeles/Spring Break/Family
Travelers/ Accommodation

-地點
-價格
-服務
-

研究問題:
研究對象: Subject E*L*O*P*T 群體
研究方法:
資料來源
資料蒐集
資料準備



1 分析的觀點與角度(Asspect/Features/Dimensions)

- (1)探索就透過電腦演算法Topics Modeling
- (2)驗證就行銷構面建立構面字集

2 針對每個研究對象個別群體，進行情緒分析

- (1)分類並標註正負面
- (2)標註情緒強度(-5~+5)

3 針對每個研究對象個別群體，已標註情緒極性之類別，進行該類別字詞出現頻率(Term Frequency/Occurrence)

4 根據3 所產生之焦點字詞，進行字詞共現性(Co-Occurrence),及社會網路分析，程度中心性、集中性、中介中心性分析



How to Discovery Insight

First, we define the factor **Subject**_{elopt} as the multiple combination of Event, Location, Object, People, and Time for the **Subjects** from which we discovering insight

Subject_{elopt} = f(事件[Event] En * 地點[Location] Ln * 物件[Object] On * 人物[People] Pn * 時間[Time] Tn)

assume:

- 1.E, L, O, P, T for possible focal groups cross Event, Location, Object, People, Time
- 2.n of En, Ln, On, Pn, are index nth element of the list of E, L, O, P, T respectively
- 3.elopt can be any one of combination of list of E,L,O,P,T

Second, we define **Insight**_{eloptAx} as the **Insight** discovering from any one of combination of list of E,L,O,P,T
Insight_{eloptAx} = STDD(Z_{tlpeo})

assume: x of Ax ranging from 1~4, for Aspect, Affection, Attention, Association respectively



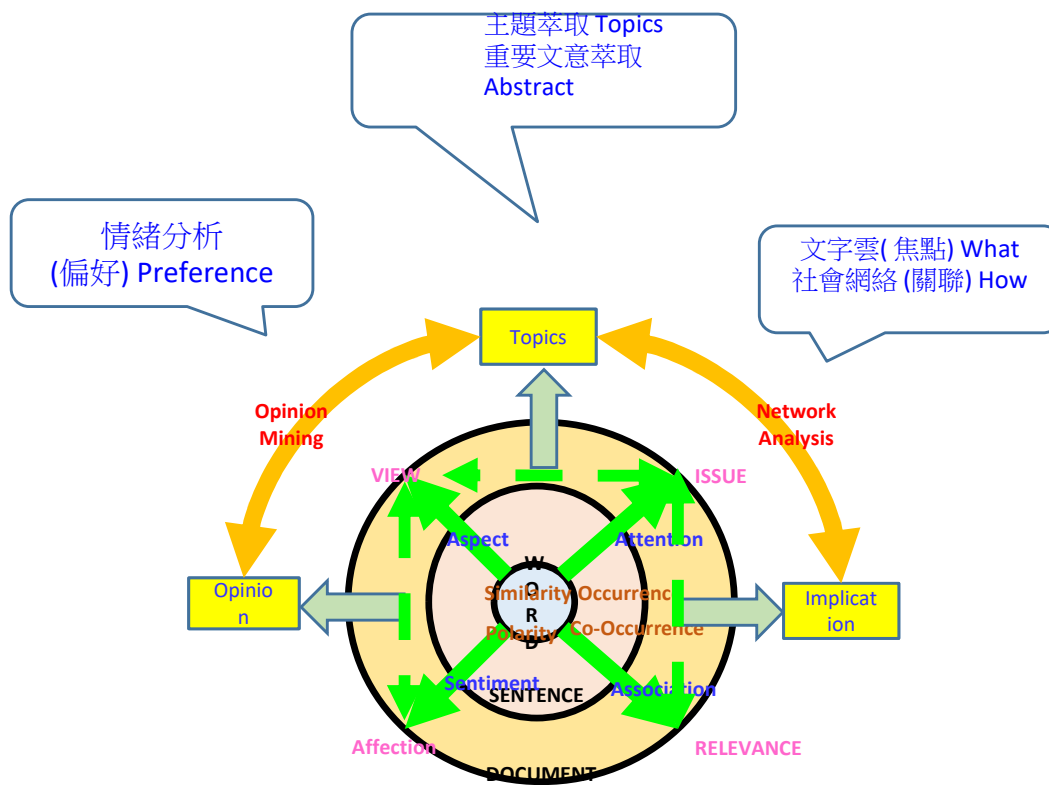
The Example of Deploying the STDD4A in Marketing Discipline

The Application of STDD4A					
Marketing Task	Marketing Measure	Text Analytics	Marketing Analytics	Example	Literatures
Branding	Brand Attitude	Aspect (View) Affection (Sentiment)	Scale(Dimensional) Likers(Preference/Attitude)	Brand Personality	Tsao, J., Campbell, C. L., Ferraro, C., Mavrommatis, A. & Lu, S. (2020). A machine-learning based approach to measuring constructs through text analysis. European Journal of Marketing, 54 (3), pp. 511-524.
Communication	Advertising Effectiveness	Aspect (View) Affection (Sentiment) Attention(Issue)	Scale(Dimensional) Likers(Preference/Attitude) Open Question (What mostly like/ What mostly dislike/	AIDA	
Marketing Research	Satisfaction	Aspect (View) Affection(Sentiment) Attention(Issue) Association(Relevance)	Scale(Dimensional) Likers(Preference/Attitude) Open Question (What mostly like/ What mostly dislike/ Why)	Service Quality Customer Satisfaction	

文字探勘見解剖析

探勘方法

- 文字分析
 - 文字出現頻率 -> 文字雲(焦點)
 - 字詞共線性 -> 社會網絡(關聯)
 - 字詞的正負向情緒 -> 情緒分析(偏好)
- 文件分類
 - 主題萃取
- 摘要自動化
 - 重要文意萃取



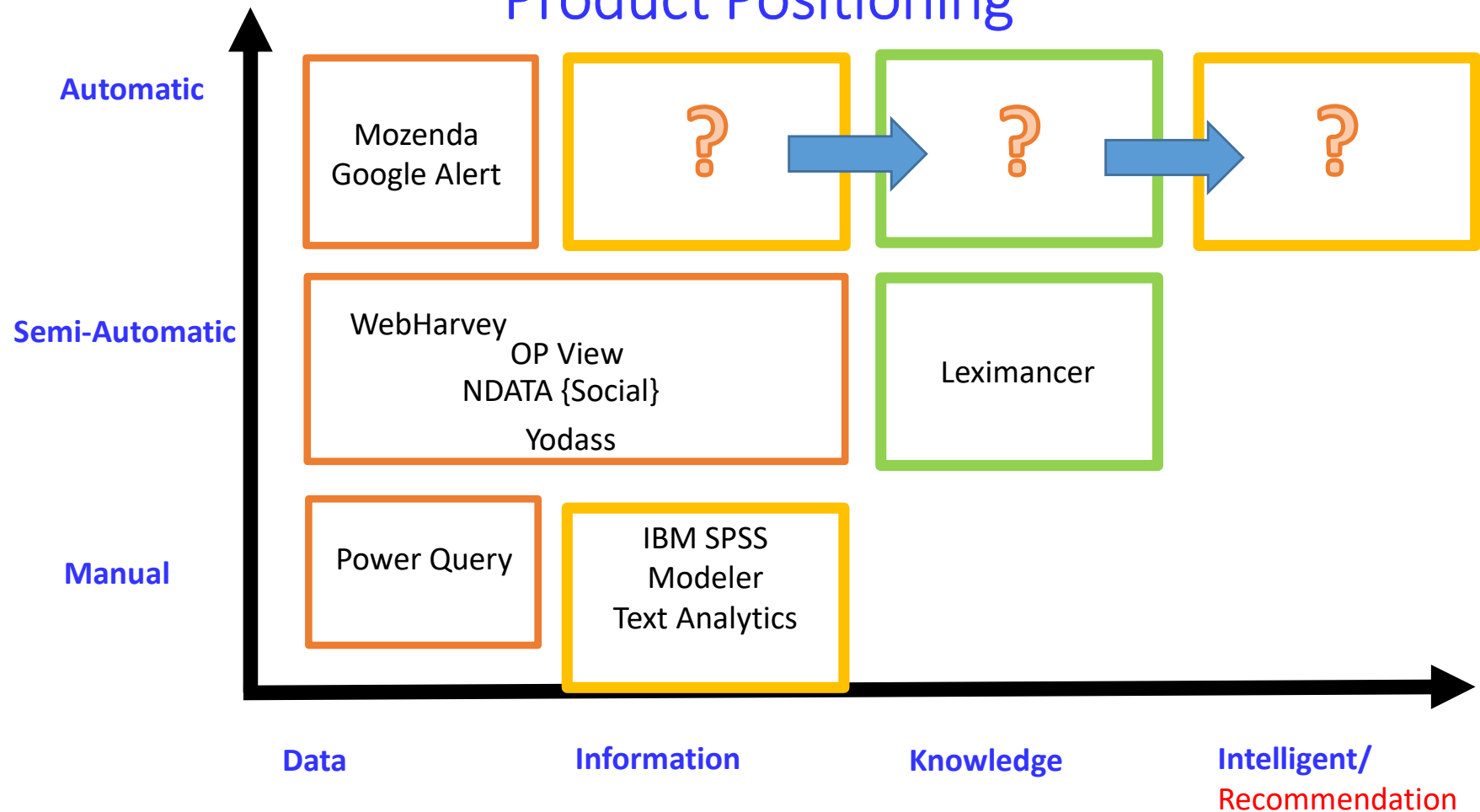
剖析對像

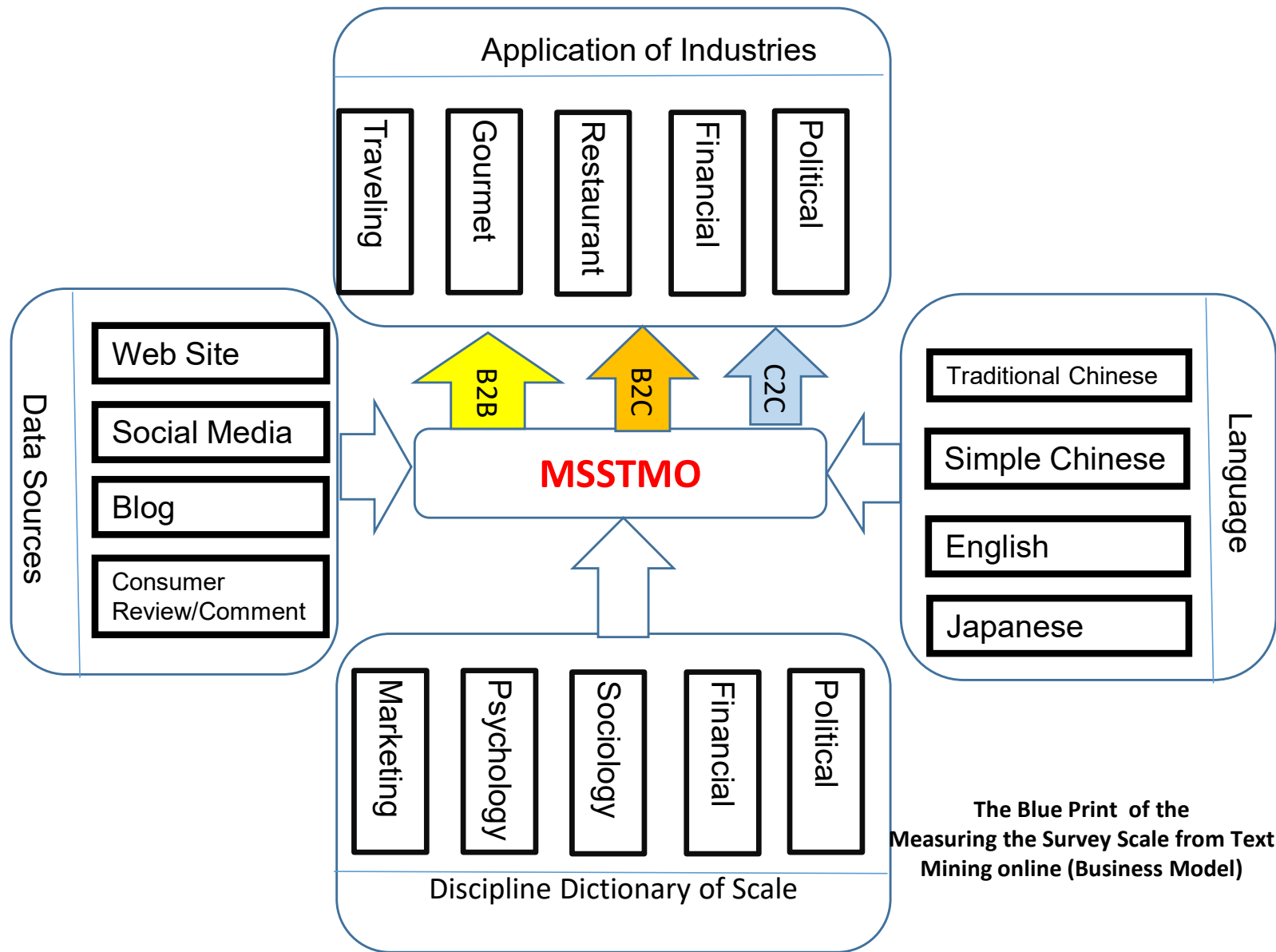
- 時間
- 地點
- 人物
- 事件
- 物件

文字探勘
剖析
行銷問題

- 競爭分析
 - 品牌
 - 定位
 - 市場區隔
- 消費者偏好(情緒分析)
 - 喜歡的是甚麼(What)
 - 不喜歡的是什麼
 - 多喜歡(How)
 - 多不喜歡
- 行銷溝通效果衡量
 - 廣告效果
 - 新產品試銷
- 公關
 - 聲量(焦點)
 - 情緒(偏好)
 - 主題(議題)
 - 趨勢(時間)
 - 預警(危機訊號)

Big Text Analytics Product Positioning





The Application of Big Text Analytics				
Marketing Task	Marketing Measure	Text Analytics	Applied Industries	Literatures
Branding	Brand Personality Brand Experience Brand Attitude	Opinion Mining	Tourism Hospitality	TSAO POKU
Communication	Advertising Effectiveness	Opinion Mining		
Marketing Research	Product Preference Satisfaction	Opinion Mining+ Network Analysis	Government Tourism Restaurant Political and Communication	
Public Relationship	Monitoring, warning, and WOM		E-Commerce Intelligence Institute Non-profit	

請輸入搜尋關鍵字

關鍵字:

(可建立交集、聯集與排除規則，並可依需求加上排除規則，規則需以_AND_、_OR_、_NOT_半形大寫輸入，規則之間不需空格且規則前後需要下底線)

比特幣_AND_特斯拉_NOT_那對夫妻

開始日期: (搜尋區間上限為三個月)

2021-04-01

結束日期: (結束日期需大於等於開始日期)

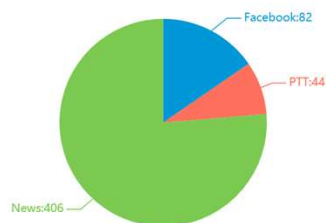
2021-04-30

平台:

☒ Facebook ☒ PTT ☒ News ☐ Forum ☐ Dcard

社群/新聞文章數佔比

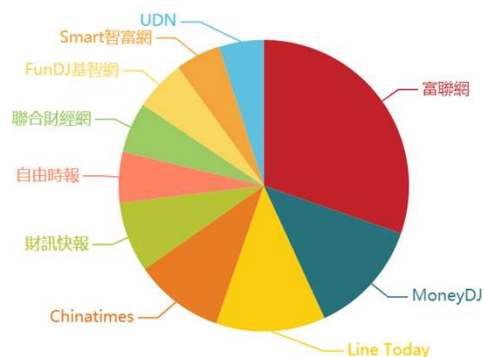
Facebook
PTT
News



總聲量: 532

Top10 新聞媒體文章佔比

富聯網
MoneyDJ
Line Today
Chinatimes
財訊快報
自由時報
聯合財經網
FunDJ智富網
Smart智富網
UDN

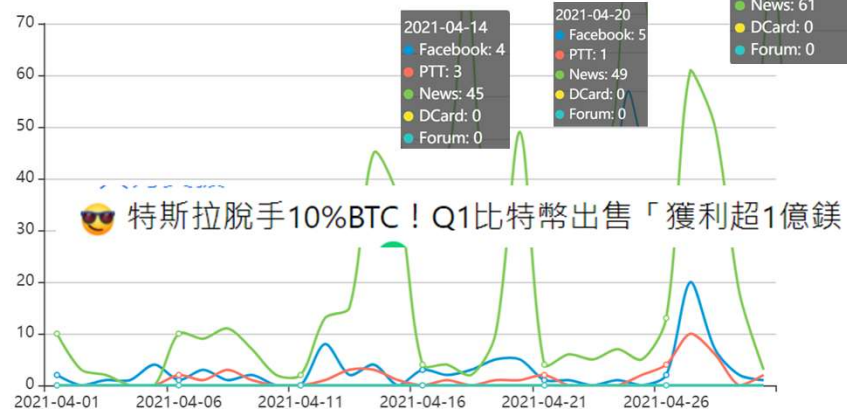


國際金融：馬斯克等揚言要在「狗狗幣日」推上新高，狗狗幣24小時飆升近18%

社群/新聞文章數趨勢

財訊新聞 2021/04/20 15:25

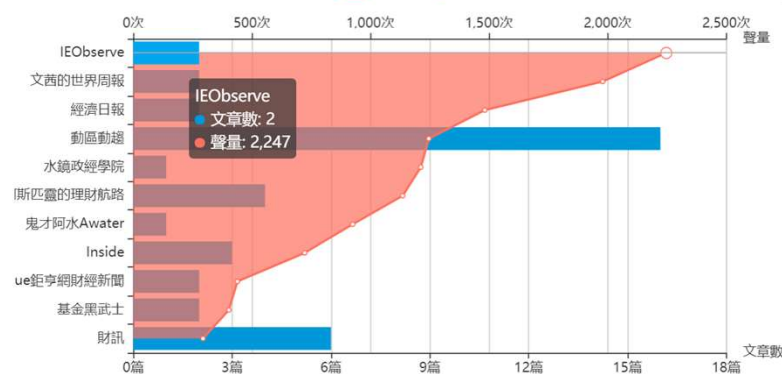
Facebook PTT News Forum DCard



🕶️ 特斯拉脫手10%BTC！Q1比特幣出售「獲利超1億鎊」

Top10 Facebook粉專聲量

文章數 聲量



Facebook



擴散強FB

比特幣 v.s 特斯拉

發文時間	網址	粉專名稱	總互動數	讚數	留言數	分享數
2021-04-27 11:30:15	https://www.facebook.com/179043295447791/posts/4241629512522462/	179043295447791 經濟日報	1463	1071	49	56
2021-04-27 08:20:25	https://www.facebook.com/203839673117112/posts/301918161292819/	203839673117112 IEObserve 國際經濟觀察	1270	1195	12	55
2021-04-18 15:09:18	https://www.facebook.com/2359076411080525/posts/2941987116122782/	2359076411080525 水鏡政經學院	1212	875	23	56
2021-04-17 09:32:47	https://www.facebook.com/322447507832934/posts/4021563741254607/	322447507832934 文茜的世界周報 Sisys World News	1006	958	4	19
2021-04-20 07:38:37	https://www.facebook.com/203839673117112/posts/297710171713618/	203839673117112 IEObserve 國際經濟觀察	977	928	10	24



經濟日報

4d · 🌐

特斯拉財報攤開，居然有逾1億美元賣出比特幣的獲利 !!

#特斯拉 #比特幣 #財報



MONEY.UDN.COM | BY 經濟日報

特斯拉Q1「炒」比特幣 大賺逾1億美元 | 國際焦點 | 國際 | 經濟日報

根據特斯拉 (Tesla) 周一公布的財報，今年第1季根據通用會計準則的淨利達創紀錄的4.38億美...



水鏡政經學院

April 18 at 3:09 PM · 🌐

#加密幣初4啦

- 1.3/4比特幣登上商週封面
- 2.今天傳出美國將針對加密貨幣洗錢進行調查，比特幣從62000下跌到53000
- 3.幣安下單出現問題

一般認為數位幣上漲，部分是洗錢需求

特別是中國外匯管制下，買數位幣拉高數位幣，再到海外換成美元，成為熱門洗錢管道

有人猜測，特斯拉公司買比特幣就是幫忙中國特定族群洗錢出去

也因此被中共官方找麻煩

若美國真查洗錢，隨之而跌的數位幣

台灣除了投資數位幣的受影響，最近飆漲的顯卡（挖礦概念），明天拉回的機會很大

3/4商周封面領先指標再次顯神通

#加百列

熱門討論鄉民聲音 比特幣 v.s 特斯拉

發文時間	發文標題	討論版	總留言數	推文數	箭頭數	噓文數
2021-04-26 23:57:07	[心得] 投資策略分享	stock	269	104	146	19
2021-04-30 17:00:57	[新聞] 蔡明彰觀點：印度疫情全面失控 飆騰股	stock	114	69	40	5
2021-04-21 17:08:50	[新聞] 4 年前鎮壓比特幣，中國如今大改口支持：	Gossiping	113	68	41	4
2021-04-25 01:36:31	Fw: [支付] 第一槍比特幣可以買房了	home-sale	102	48	54	0

批踢踢實業坊 > 看板 Stock

聯絡資訊 關於我們

作者 pgoiwe (nn)
 標題 [心得] 投資策略分享
 時間 Mon Apr 26 23:57:07 2021

看板 Stock

1. 發文前請先詳閱[心得]分類發文規範，未依規範發文將受處份。
2. 若提及股號股名，且屬未來性的預測或分析有推薦意圖，請寫到[標的]分類，否則將以板規4-1砍文處份。

-----按ctrl+y可刪除以上內容。-----

看到很多人在討論綠角的被動投資
 我自認為我有個主動投資的方法比綠角好得多

其實我覺得投資最大的祕密就是一便宜的槓桿
 根據分散時間投資法，年輕的時候要用比較大的槓桿
 而最便宜的槓桿就是一期貨
 期貨的槓桿可以自己控制，不用付利息，比融資好得多
 所以我的策略也是以期貨為中心

以我自身為例，我目前總資產(不含房子)大約2000萬元
 我持有台積電個股期貨12口，實際價值等於24張台積電，約1450萬元，但是我期貨保證金只放300萬元而已，意思是我不用付任何利息就借到了1450-300=1150萬(實際上是負利率啦，因為遠期還有逆價差可以吃，等於借你錢還給你利息)
 可能有人會覺得重押單一個股台積電很不妥
 我倒是很安心XD



熱門焦點新聞

比特幣 v.s 特斯拉

1億美金！「楓之谷」南韓母公司大買1717枚比特幣

2021/04/29 00:21



推出多款熱門遊戲的南韓知名遊戲公司Nexon，宣布花費近1億美金購買比特幣。(路透)

比特幣概念股買啥好？高盛：19檔概念股 今年漲46%

2021/04/28 11:33

回應(0) 人氣(0) 收藏(0)

MoneyDJ新聞 2021-04-28 11:33:52 記者 陳苓 報導



比特幣漲勢驚人，今年以來飆高86%，許多投資人想從此一趨勢獲利，又怕比特幣過於波動，轉向擁抱相關概念股。到底哪些標的可以抱？高盛篩選後，選出19檔純度最高的概念股。

Barron's 27日報導，儘管有好幾家公司接受比特幣、採納比特幣背後的區塊鏈技術、甚至直接敲進比特幣，但是對投資人來說，很難判斷哪些個股和比特幣價格走勢的連動性最強。高盛分析師Ben Snider用三個標準衡量，找出19檔純度最高的個股。

理財

傳臉書最新財報將揭露「持有比特幣」

anue鉅亨網
更新於 2天前 · 發布於 04/28/2021

11圖



© Anue.com 鉅亨網 提供

特斯拉披露：目前持有價值24.8億美元的比特幣

2021/04/29 11:26



美國電動車大廠特斯拉 (Tesla) 28日表示，截至3月31日為止，該公司持有比特幣的公允市場價值為24.8億美元。(路透)



馬斯克碰上比特幣-- 解讀文字雲



只在此山中，雲深不知處



行到水窮處，坐看雲起時

主題分類



情感正負



關注焦點



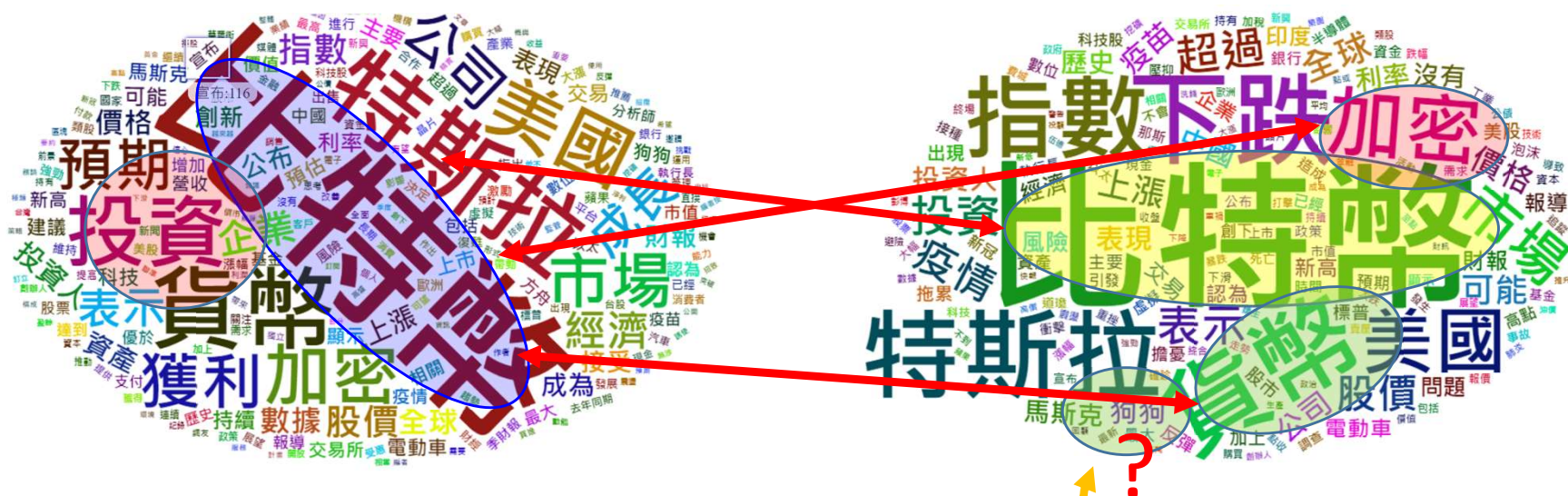
洞見



建議



社群正負面意見焦點 比特幣 v.s 特斯拉



正面的焦點

正負兩面兼具

正面的焦點

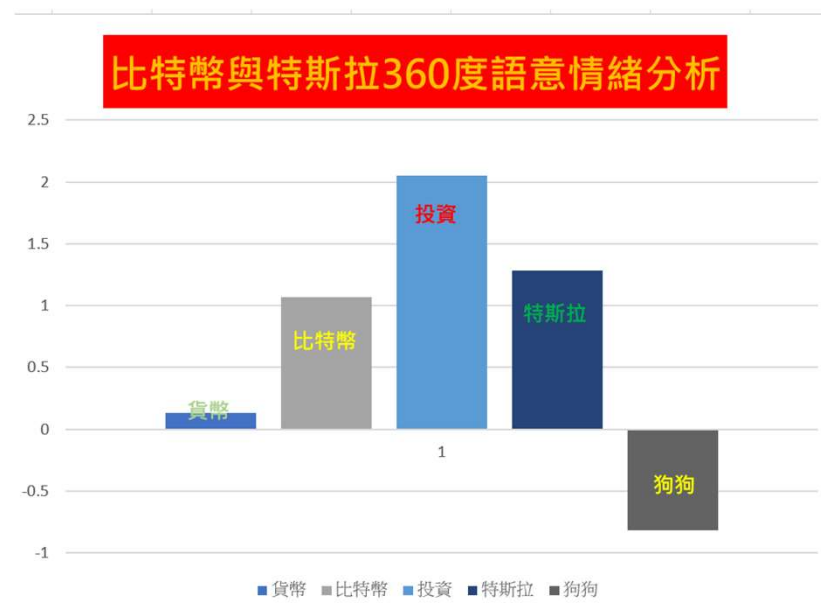
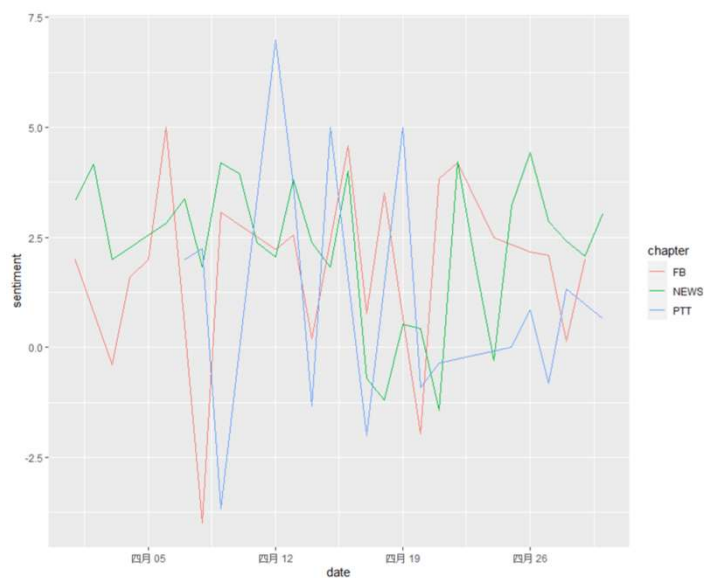
負面的焦點

負面的焦點

讓我們看雲去--比特幣

未被發現的焦點

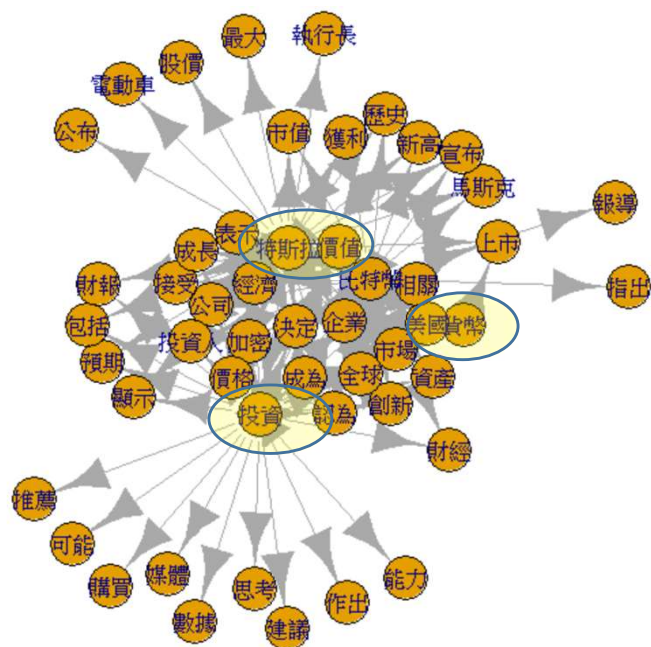
讓我們看雲去--比特幣 v.s 特斯拉



輿論語意情緒變化

多元語意情緒分數

讓我們看雲去--比特幣 v.s 特斯拉



正面

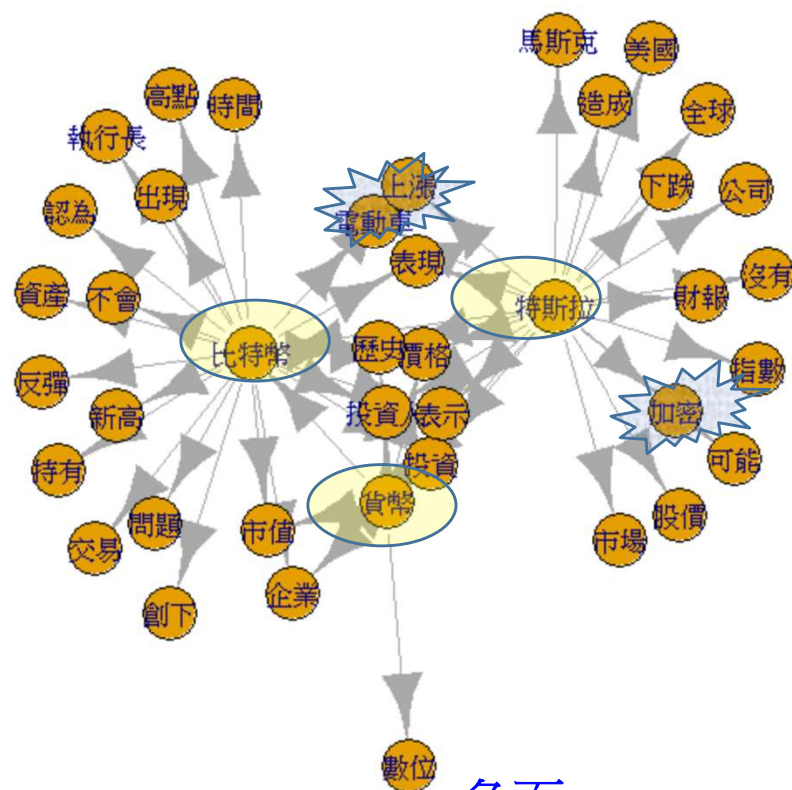
未被發現的焦點

正負兩面兼具

正面的焦點

負面的焦點

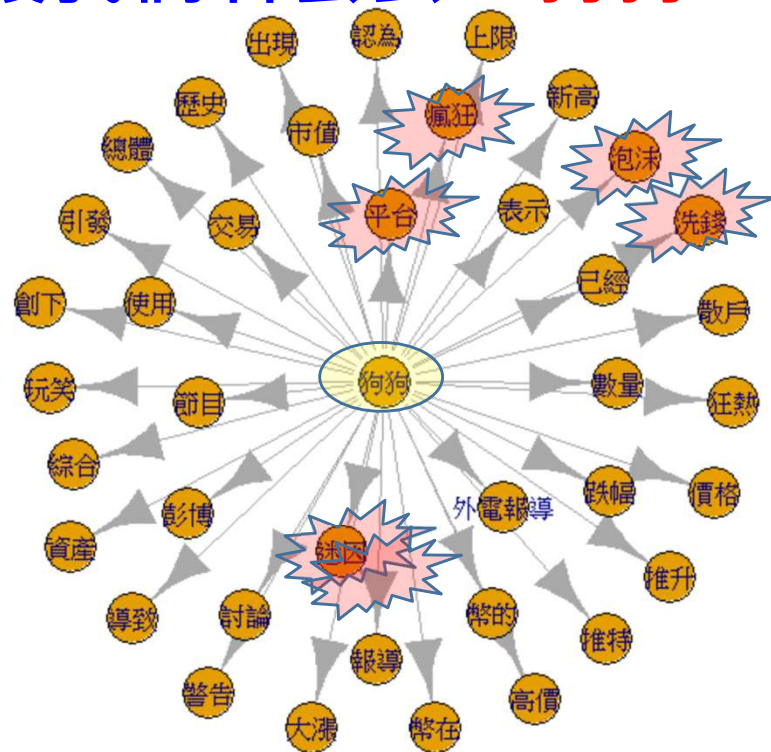
未被發現的焦點



負面

未被發現的焦點

讓我們看雲去——狗狗？



意外的焦點

中文斷字 文字雲產生器



線上展示 / Online Demo

Filename Raw Text

Input Raw Text

message
這個布丁是在無聊的世界中找尋樂趣的一種不能吃的codes，喜愛動畫、遊戲、Coding，以及ABCDV跟世間脫節的生活步調。

START SEGMENTATION

Output Segmentation Result

message
布丁 無聊的世界 找尋 code 1996 種 不能 吃 code
喜愛 動畫 遊戲 code abcdv 世間 脫節 生活 步調

Copy

☒ Remove first line (header)

Draw word cloud

Text

253天無本土病例破功！陳時中13:30親上火線說明

【寵物聖域市集】寵物雲商城全館結帳89折！

4,446

讚

Input Raw Text

國內今（22日）新增3例新冠肺炎確定病例，中央流行疫情指揮中心將於下午1時30分召開臨時記者會，由指揮官陳時中說明病例情况，據了解其中有本土病例，確診病患目前正在醫院接受治療中，也讓台灣連續253天無本土病例破功。

以往指揮中心例行記者會為每周三，一般日的臨時記者會如果是境外移入，僅由發言人莊人祥說明，而今天臨時記者會改由陳時中主持，不排除有本土個案，台灣上一次有本土病例已經是4月12日，是案386的男性，被美國境外移入的同住友人傳染。

指揮中心統計，國內目前累積769例確診，分別為675例境外移入，55例本土病例，36例軟體團隊，2例航空器感染及1例不明；另1例(案530)移除為空號，確診個案中7人死亡，627人解除隔離，其餘住院隔離中。

(醫生)請繼續工作開講

START SEGMENTATION

Output Segmentation Result

國內 22 新增 例 新冠 肺炎 確定 病例 中央 流行 疫情 指揮 中心 下午 30 召開 臨時 記者會 指揮 官 陳 說明 病例 情况 了解 本土 病例 確診 病患 目前 醫院 接受 治療 台灣 連續 253 天 本土 病例 破 功

以往 指揮 中心 例行 記者會 每周 三 臨時 記者會 境外 移入 僅 發言人 莊 人 祥 說明 今天 臨時 記者會 改 由 陳 主持 排除 本土 個案 台灣 一次 本土 病例 已經 12 案 386 男性 美國 境外 移入 住友 傳染

指揮 中心 統計 國內 目前 累積 769 例 確診 675 例 境外 移入 55 例 本土 病例 36 例 軟體 團隊 例 航空 器 感染 例 不明 例 案 530 移除 空 號 確診 個案 死亡 627 解除 隔離 住院 隔離

Copy

Save as text

Save as sheet



SENTIMENT ANALYSIS

文字情緒分析



happiness



joy



proud



sadness



fear



embarrassment



worry



anger



upset

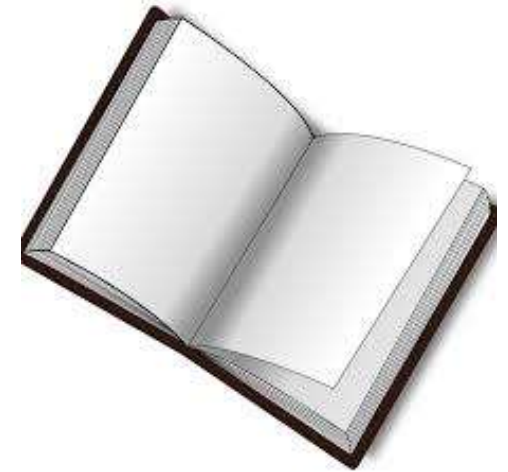
Also known as **opinion mining**: to understand the attitude of a speaker or a writer with respect to some topic

– The attitude may be their judgment or evaluation, their Affection state or the intended emotional communication

- – Most popular classification of sentiment:

• **positive** or **negative**

- unbelievably **disappointing**
- Full of zany characters and **richly** applied satire, and some **great** plot twists
- this is **the greatest** screwball comedy ever filmed
- It was **pathetic**. The **worst** part about it was the boxing scenes.



AFINN Sentiment Analysis

AFINN情緒辭典是一個評分範圍從-5(非常負面)到5(非常正面)的英文單字列表，是附有極性的評分方式。



分數	辭彙
-4	遜、歧視、渣、發脾氣、輸、爛透、不要臉、笨蛋、智障、etc.
-3	可怕、受不了、恐慌、討厭、欺騙、傻眼、厭惡、賠償、etc.
-2	小氣、干涉、不耐煩、反感、欠缺、叫囂、失望、丟人現眼、etc.
-1	不對勁、丟棄、兇、有待加強、吵、取消、延遲、浪費、etc.
1	心安、充足、平淡、安心、安全感、自由、吸引、明晰、etc.
2	及格、友善、心儀、支持、主動、出色、合適、划算、etc.
3	大笑、不賴、心愛、名正言順、好吃、有趣、完美、etc.
4	可圈可點、哇、棒、精彩、豪華、優雅、驚艷、贏、etc.

[TLSSD\(Tsao, Lin & Su Sentiment Dictionary\)](#)



否定詞的加入將會影響到詞彙或句子的意思。像是「好吃」這詞本身就有正面情緒的意思，但是如果在前面加入否定詞「不」來做修飾，變成一個新的字詞「不喜歡」，也使得原本正面情緒的字詞改變為負面情緒。本研究自行編撰否定詞詞庫並將此以-1表示。

否定詞詞庫

不、不及、不太、不必、不用、不再、不如、勿、未、杜絕、沒、拒絕、非、無法、禁止、etc.



利用知網的程度值字別建立程度詞詞庫，依照強度大小分成「級、較、很、稍、無」，五個級別。

級別程度	權重值	程度詞詞庫
極 extreme	5	十分、有夠、卓越、極度、絕對、etc.
較 more	4	太、更、越、較、還要、比較、etc.
很 very	3	出奇、多麼、如此、相當、滿、頗etc.
稍 ish	2	多少、有些、尚可、些微、稍加、etc.
無 least	1	無、沒有、再也、etc.

- Sentiment = Holder + Polarity + Target
 - **Holder**: who expresses the sentiment
 - **Target**: what/whom the sentiment is expressed to
 - **Polarity**: the nature of the sentiment (e.g., positive or negative)

Sentiment = Holder + Polarity + Target

In his recent State of the Union address, US President Bush quite unexpectedly labeled Iran, Iraq, and the DPRK as an "axis of evil" .

Negative

The Example of Analysis of Sentiment Analysis

- Sentiment = Holder + Polarity + Target
- *In his recent State of the Union address, US President Bush quite unexpectedly labeled*
- *Iran, Iraq, and the DPRK as an “**axis of evil**”.*
-



Negative

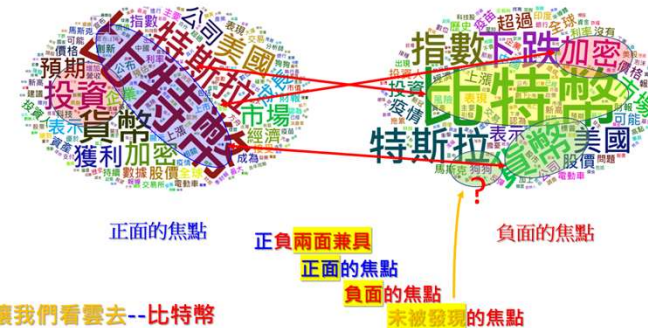
Why sentiment analysis?

- **Movie** : is this review positive or negative?
- **Products** : what do people think about the new iPhone?
- **Public sentiment** : how is consumer confidence?
- Is despair increasing?
- **Politics** : what do people think about this candidate or issue?
- **Prediction** : predict election outcomes, market trends, or stock price from sentiment

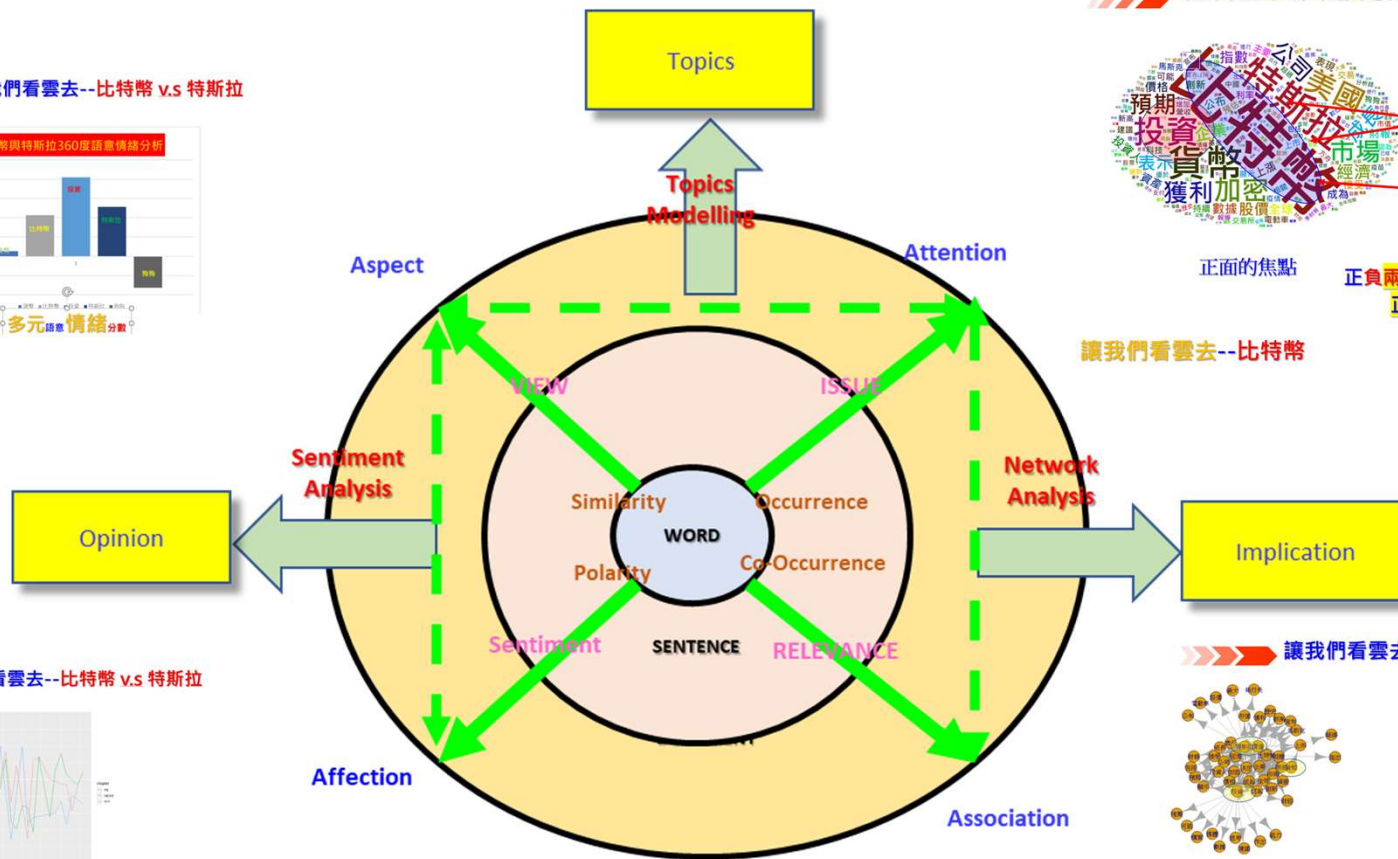
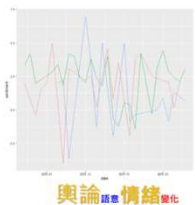
讓我們看雲去--比特幣 v.s 特斯拉



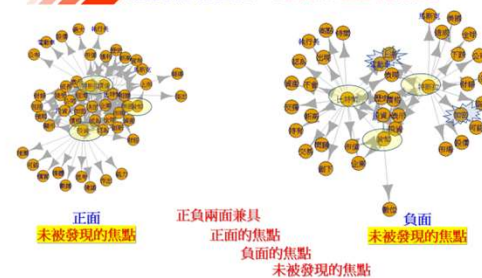
社群正負面意見焦點 比特幣 v.s 特斯拉

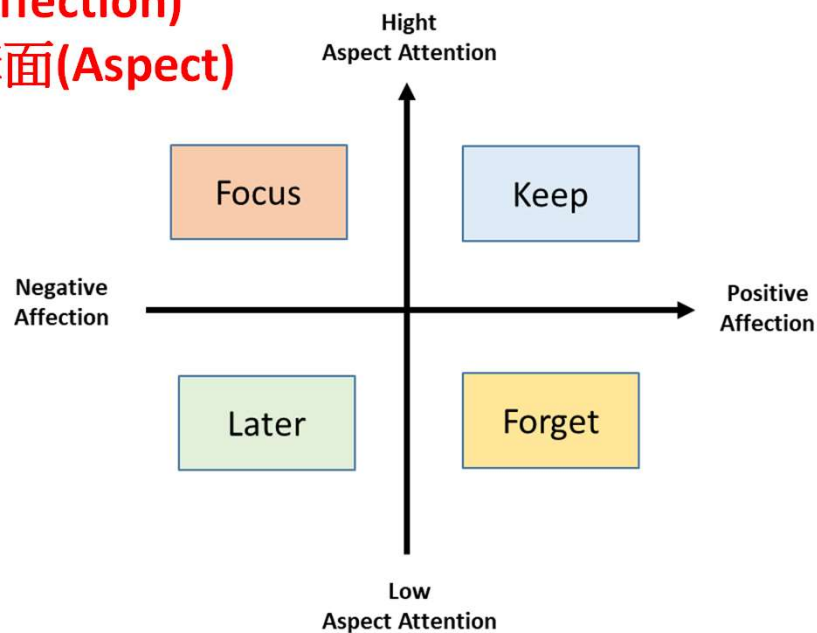
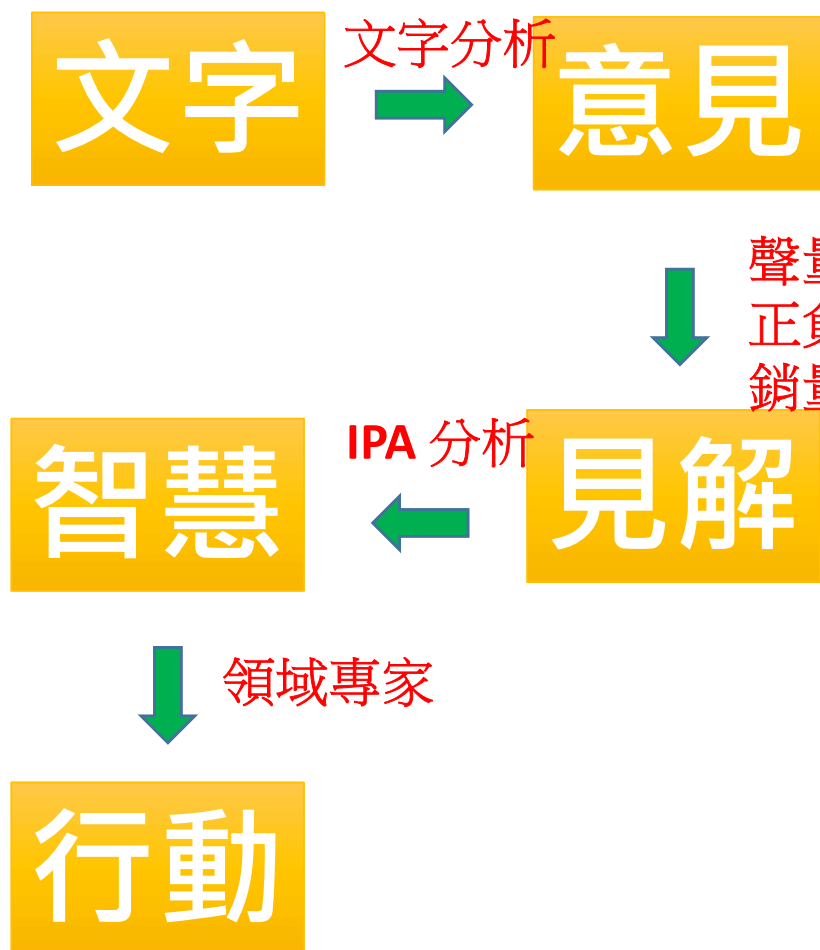


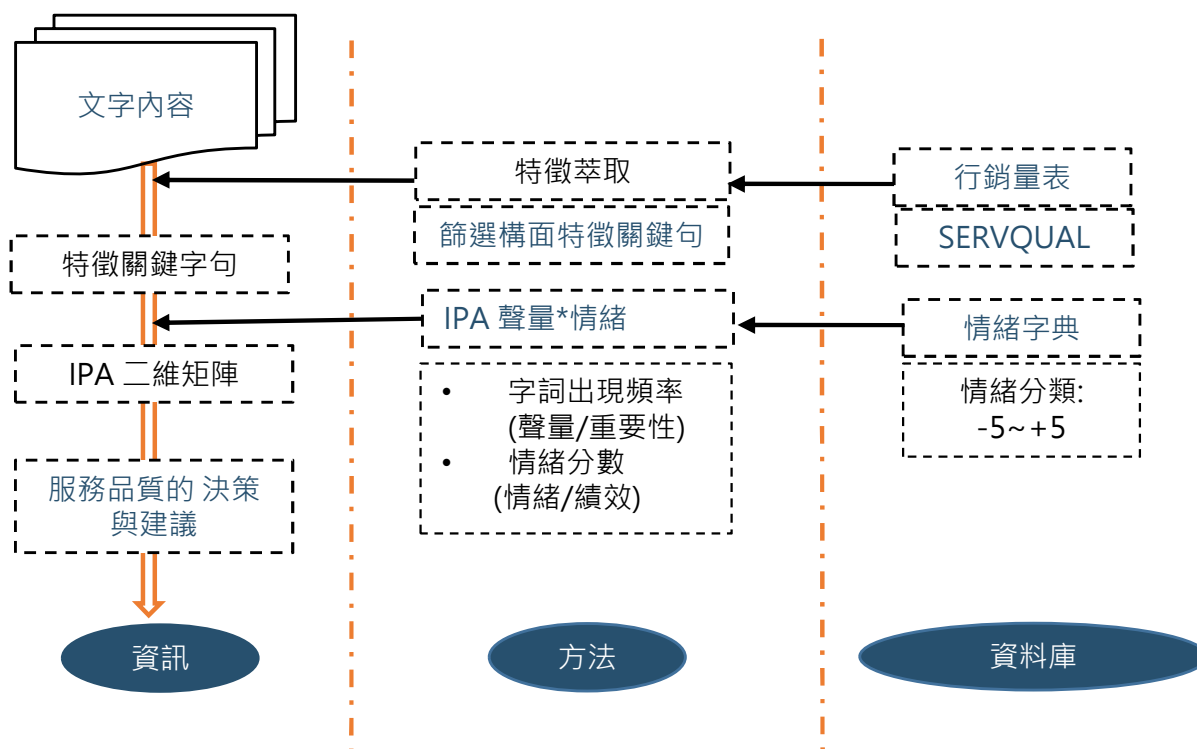
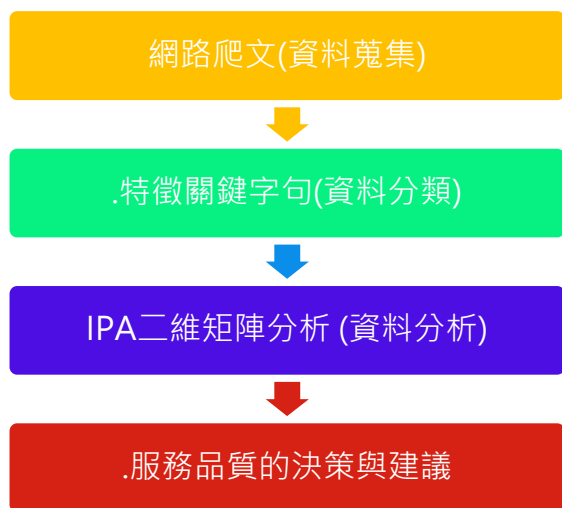
讓我們看雲去--比特幣 v.s 特斯拉



讓我們看雲去--比特幣 v.s 特斯拉









斷字

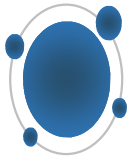
171	1	確認 數據 綠燈 平均速度 低於
172	1	表示 狀況 許久 並非 遠方 基地 近期 故障 影響
173	1	通話 斷斷續續 穩定
174	1	本人 反映 住家 今天 上網 緩慢 通話 正常 體驗 綠燈 熱點 分享 網速 緩慢 聯電 開單
175	1	本人 反映 住家 上網 緩慢 通話 正常 體驗 綠燈 持續 滿格 其他 FET 相同 熱點 分享 網速
176	2	簡述 問題 陽台 才能 使用

斷句

171	1	已確認數據為綠燈且平均速度低於7M
172	1	表示此狀況已許久,並非遠方基地台()近期故障影響
173	1	通話斷斷續續不穩定"
174	1	本人反映住家今天上網緩慢,通話正常,體驗綠燈,3~4格,無熱點分享,網速緩慢<7M,聯電,開單
175	1	本人反映住家上網緩慢,通話正常,體驗綠燈,持續2~3天,滿格,其他FET相同,有熱點分享,網速緩慢<7M位,開單
176	2	簡述客問題:室不佳,要到陽台才能使用

特徵關鍵字 字典

電信業 服務品質構面	有形性 (Tangibles)	可靠性 (Reliability)	回應性 (Responsiveness)	保證性 (Assurance)	同理性 (Empathy)
	地點	承諾	確實	任務	聆聽
	連線	完成	及時	知識	理解
	場所	達成	樂意	能力	心情
	門市	穩定	意願	尊重	建議
	服務	正確	快速	行為	經驗
	速度	準時	負責	技術	分享
	價格	信賴	解決	誠意	處境
	效率	一致	協助	接觸	互動
	功能	準確	授予	支援	道歉
	設施	履行	確認	專業	朋友
	資料	滿意	準確	溝通	公司
	空間	品質	機制	解決	家庭
	環境	管理	資訊	處理	方便
	速率	控管	積極		關心
	上網	訓練	授權		重視
	費率		態度		權益
	網路		熱忱		主動
	合約				需求
	通路				補償
	產品				風險
					傾聽
					關懷
					體驗



1.網路爬文(資料蒐集)

經由輿情系統或爬蟲程式下載

2.特徵關鍵字句(資料分類)

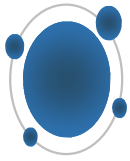
以台中捷運議題話務內容為例

2.特徵關鍵字句(資料分類)

構面分類及情緒分數計算

ID	構面	特徵詞	特徵關鍵句	TLSSD 情緒字詞	情緒分數
2	Tangibility	車廂	車廂非常 清潔與舒適	舒適	+4
2	Assurance	服務態度	人員服務 態度欠佳	欠佳	-2

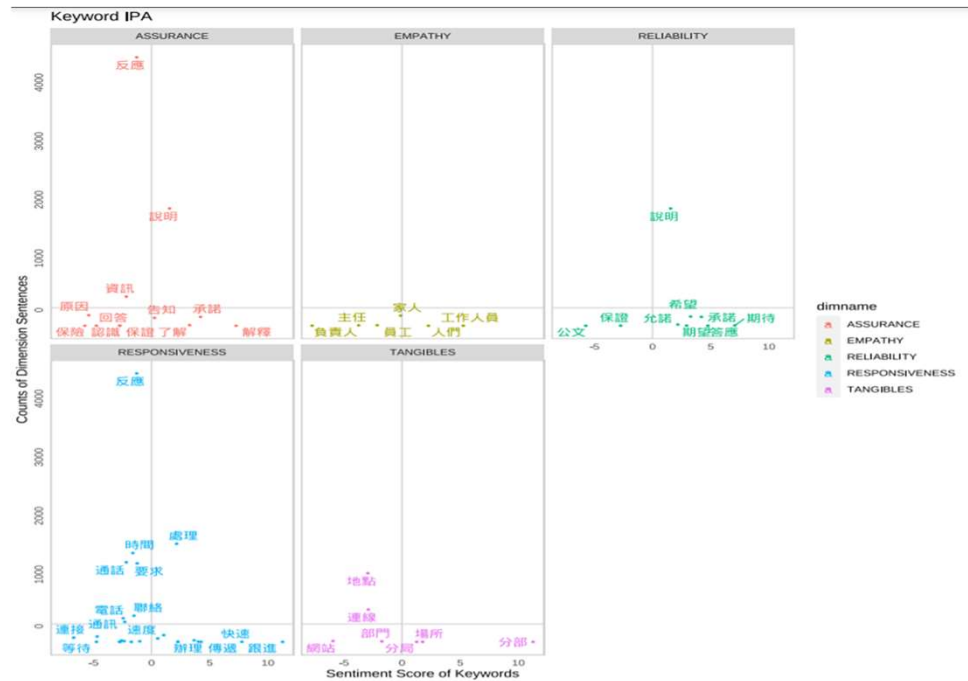
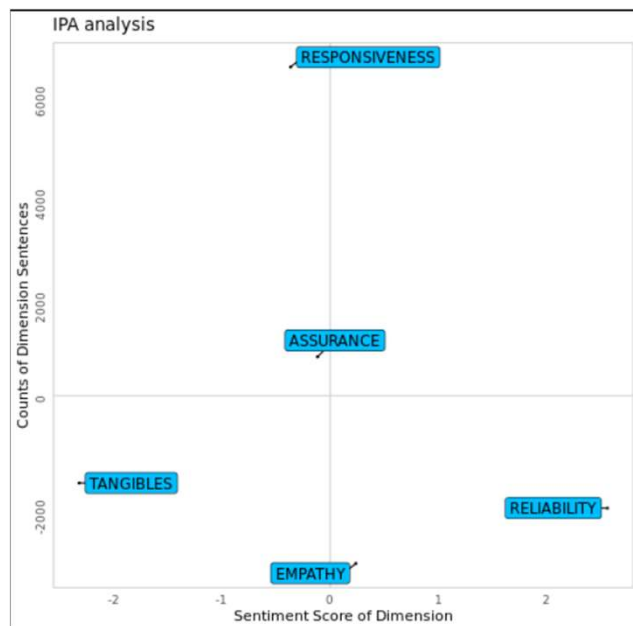
台中捷運服務品質構面	Tangible	Reliability	Responsiveness	Assurance	Empathy
斷裂	問題	報告	安全	公車轉乘	
人行道	服務	川崎	態度	需求	
高鐵	提供	處理	服務	公車	
車廂	正確	服務	信任	乘客	
設備	協助	旅客	能力	重視	
設計	品質	迅速	處理	娛樂	
時間	完善	搭乘	表示	旅遊	
環境	困難	回應	測試	景點	
路線	解決	抱怨	負責	交通	
路網	延誤	規定	知道	接駁	
綠線	承諾	幫助	道歉	商店街	
藍線	資訊	資訊	保證	處理	
橘線	態度	旅遊	確保	抱怨	
運量	重視	即時	交代	表示	
營運	安心	回答	解決	負責	
列車	信任	改善	決定	重要	
車站	信賴	規劃	謹慎	道歉	
軌道	詢問	通車	發現問題	設計	
月台	權益	負責		習慣	
高架	過程	知道		養成	
停車位	方便	發現		方便	
收費	及時	發生		需要	
票證	誤點	搭乘			
	處理				

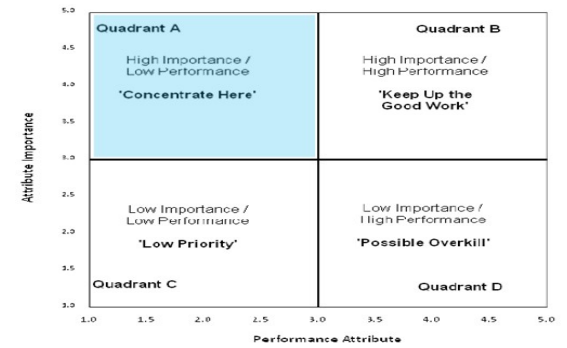
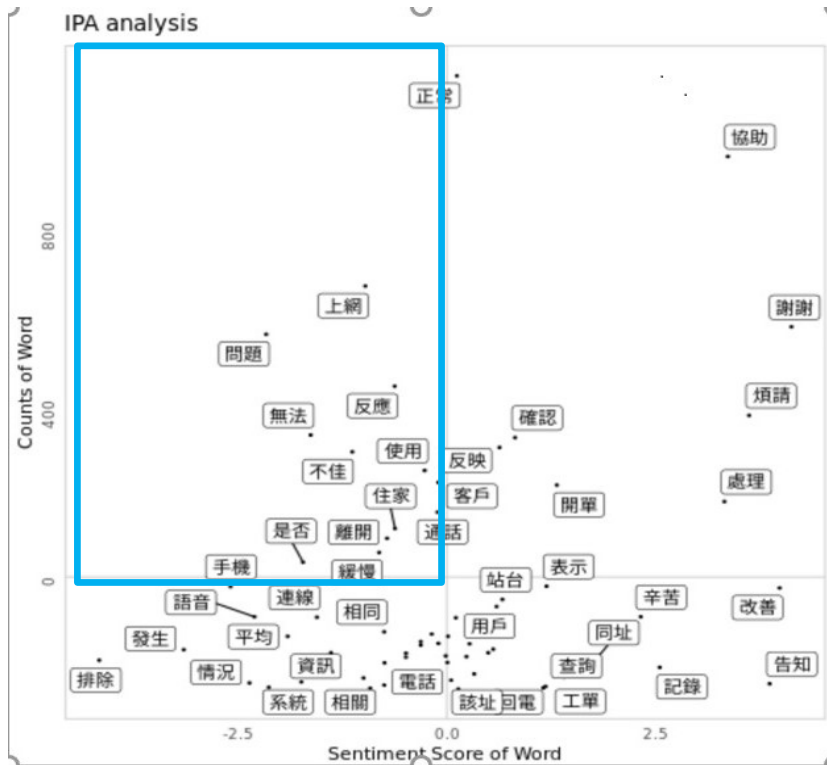


line	sentenceno	sentiment	word	dimname	text
173	1	3	通話	RESPONSIVENESS	通話斷斷續續不穩定"
174	1	-1	通話	RESPONSIVENESS	本人反映住家今天上網緩慢,通話正常,體驗綠燈,3~4格,無熱點分享,網速緩慢
175	1	-3	通話	RESPONSIVENESS	本人反映住家上網緩慢,通話正常,體驗綠燈,持續2~3天,滿格,其他FET相同,有電、GIS無法定位,以常駐站
177	2	-2	反應	ASSURANCE	反應地點：住家(彰化縣),因
177	2	-2	反應	RESPONSIVENESS	反應地點：住家(彰化縣),因
177	2	-2	地點	TANGIBLES	反應地點：住家(彰化縣),因

dimname	sentiment
ASSURANCE	3.007622677
EMPATHY	3.362244898
RELIABILITY	5.688485804
RESPONSIVENESS	2.759298889
TANGIBLES	0.803876853

- 特徵關鍵字句情緒分數(情緒)
- 特徵關鍵字句出現次數(聲量)
- IPA <--- 聲量*情緒





文字分析產品服務市場

分析主題	實例	分析項目	行銷管理內涵	應用產業
品牌個性 品牌經驗 品牌概念	Starbucks v.s 85Cafe 85Cafe vs. 壹咖啡 怡客 vs. 丹堤 海尼根 vs. 老虎	品牌 vs. 競爭品牌 代言人 vs. 競品代言人	網路品牌管理 (品牌定位與差異化) 廣告效果衡量	企業 娛樂業 餐旅業
產品主題情緒分析	85C Café海鹽咖啡 王品菜單 Apple iPhone7 柯P 大巨蛋事件	針高情緒分數與低情緒分數觀察 高關注的項目 產品特性之與高關聯之項目分析	新產品開發 產品滿意度調查 廣告關鍵字之投放建議	餐旅業 政府(觀光局) 3C產業 政治傳播
問卷文字化分析	餐飲業(王品滿意度) 飯店業(雲品滿意度) 1999專線滿意度	廣告效果 滿意度 服務品質	問卷調查	餐旅業 政治傳播
網路口碑管理之流程	電商或Hotel暨競爭者消費者的聲音 Trump vs. Hillary 太陽花學運 個人情緒分析	社群媒體 監測、預警、建議、推薦	網路公關 民意探勘 情資蒐集 高風險個案監測	電商餐旅業 政治傳播 研究調查機構 非營利組織
情緒分數預測民調 情緒分數預測產品市場 成長及股市漲跌	柯P vs. 連勝文 Trump vs. Hillary	民調與情緒分數關連性 預測民調及選舉結果	銷售預測 市場預測	政治傳播 企業/零售/電商 財金

情緒分析網站 (練習範例)

- [AFINN Sentiment](#)
- [lexalytics](#) [簡介PPT](#)

網路爬蟲步驟範例

- [Microsoft Power Query](#) 為例
 - [柯P](#)
 - [Donald Trump](#)
 - [政府開放資料](#)
- Webharvy (練習範例)
- Youtube
- PTT



爬蟲軟體Webharvy Step by Step Demo

東森新聞網
拜登相關新聞





美國大選：習近平祝拜登

美國大選:習近平祝拜登當選的賀電來了。在拜登被...

2020/11/26 08:14

即時 | 熱門 | 娛樂 | 總覽 | 影音 | 關鍵 | 直播 | BBC

Capture

美國大選：習近平祝拜登

Capture

Please provide a name for the item:

Title

OK Cancel

Capture Text

Follow this link

Capture target URL

Set as Next Page link

Capture Image

More Options

選舉結果

表示這

30s AJAX 5s] Miner Options [High,10,2] Mining Threads [4] Zoom [100%] Help for

WebHavy Evaluation

Home Configuration Actions Tools Help

Open Save Edit Start Stop Clear Settings Start-Mine

File Configuration File

Url https://news.ebc.net.tw/Search/Result?type=keyword&value=%25E6%258B%259C%25E7%2599%25BB

EBC東森電視 | 東森Apps | 東森美洲 | 簡體

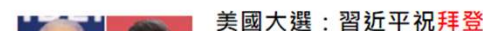


不出國也超好玩！耶誕跨年全攻略

即時 | 熱門 | 娛樂 | 總覽 | 影音 | 關鍵 | 直播 | BBC News 中文 | 話題

首頁 > 搜尋結果

搜尋結果：符合「拜登」的查詢結果共771筆



美國大選：習近平祝拜登

Proxy [OFF] Timeout [PageLoad: 30s AJAX 5s] Miner Options [High,10,2] Mining Threads [4] Zoom [100%] Help for news.ebc.net.tw

Captured Data Preview

Title

美國大選：習近平祝拜登

川普不想認輸！翻轉選舉結果 國務院通知職員展開...

川普附了！拜登創紀錄...

川普「瘋狂退群」卻囑厚台灣？吳斯懷：影響不能輕忽

拜登外交核心三人組：...

美環保署長取消訪台！外交部證實：行程變動

快訊／包機訪台要花700萬！美環保署長取消行程 外交...

白宮批准向拜登提供情...

美國大選：拜登獲進展...

美國大選：川普承認拜登

政權交接啟動後首度現身 川普讓股市破紀錄

快訊／川普終於批准！允許拜登...

川普再踢鐵板！內華達州最高法院確認

拜登、疫苗利多 美股收...

紐時：中國侵略性增強 拜登

搜尋結果：符合「拜登」的查詢結果共771筆



美國大選：習近平祝拜登

美國大選:習近平祝拜登當選的賀電來了。在拜登被...

2020/11/26 08:14

Capture

美國大選：習近平祝拜登
美國大選:習近平祝拜登當選的賀電來了。在拜登被...
2020/11/26 08:14

T Capture Text

Follow this link

Capture target URL

Set as Next Page link

Capture Image

More Options

Title	Content
美國大選：習近平祝拜登	美國大選：習近平祝拜登當選的賀電來了。2020/11/26 08:14 字：字 ▲ (圖/REUTERS/EPA) 在拜選後預測當選下屆美國總統之後兩個多星期，中國最高領導人習近平向拜登發去賀電。美國大選
川普不想認輸！美聯社選舉結果 國務院通知職員開關	...
川普辭職了！拜登的副總統	...
川普「瘋狂總統」卻關照台灣？吳斯懷：影響不能輕忽	...
拜登外交核心三人組：...	...

WebHany Evaluation

Home Configuration Actions Tools Help

Open Save Edit Start Stop Clear Settings Start-Mine

File Configuration is

UH https://news.ebc.net.tw/news/article/237882

即時 | 熱門 | 娛樂 | 總覽 | 影音 | 關鍵 | 直播 | BBC News 中文 | 話題 | 爆料

美國大選：習近平祝拜登當選的賀電來了

2020/11/26 08:14 字級：字

0:00 / 4:37

東森新聞 51頻道 點我看直播

產銷履歷 認真如你

Capture

美國大選：習近平祝拜登當選的賀電來了
2020/11/26 08:14
字級：字

▲ (圖/REUTERS/EPA)

在拜選後預測當選下屆美國總統之後兩個多星期，中國最高領導人習近平向拜登發去賀電。

美國大選：習近平祝拜登當選的賀電來了。在拜登被...

勝出之後不久，包括德國總理默克爾、法國總統馬克龍、英國首相約翰遜在內的多國領導人都在第一時間致電拜登表示祝

Proxy [OFF] Timeout [PageLoad: 30s AJAX: 5s] Miner Options [High, 10.2] Mining Threads [4] Zoom [100%] Help for news.ebc.net.tw

Captured Data Preview

Title

美國大選：習近平祝拜登

川普不想認輸！美聯社選舉結果 國務院通知職員開關

川普辭職了！拜登的副總統

川普「瘋狂總統」卻關照台灣？吳斯懷：影響不能輕忽

拜登外交核心三人組：...

WebHarvy Evaluation

Home Configuration Actions Tools Help

Open Save Edit Start Stop Clear Settings Start-Mine

File Configuration

https://news.ebc.net.tw/Search/Result?type=keyword&value=%25E6%258B%259C%25E7%2599%25BB

WebHarvy Evaluation

Home Configuration Actions Tools Help

Open Save Edit Start Stop Clear Settings Start-Mine

File Configuration

https://news.ebc.net.tw/Search/Result?type=keyword&value=%25E6%258B%259C%25E7%2599%25BB

首頁 > 搜尋結果

搜尋結果：符合「拜登」的1

美國大選：拜登勝過川普

美國大選：2020/11/26 08:43

川普不想開...

訴支持者必主黨的拜登

2020/11/26 08:43

川普掰了

川普掰了！當選人

2020/11/25 20:43

1 2 3 4 5 >

Capture

< > . * T | 手 筆 尺 卷 尺 字典

>

T Capture Text

Follow this link

Capture target URL

Set as Next Page link

節目表 | 新聞自律委員會 | 資訊公開專區 | 取消訂閱推播

東森電視首頁 | 財經新聞台 | 綜合台 | 戲劇台 | 幼幼台 | 電影台 | 洋片台 | 超視 | 美洲衛視 | YOYO幼兒園 | 東

EBC 東森新聞

東森電視事業股份有限公司 10041 台北市忠孝西路一段

服務時間：週一至週五 09:00-20:00 | 電

WebHany Miner

Title	Content
美國大選：習近平視拜登當選的實電來了	美國大選：習近平視拜登當選的實電來了 2020/11/26 08:14 字級：字字 ▲(圖/REUTERS/EPA) ...
川普不想認輸！翻翻轉選結果 國務院通知職員展開過渡	川普不想認輸！翻翻轉選結果 國務院通知職員展開過渡 2020/11/26 06:48 中央社 字級：字字
川普撿了！拜登創紀錄贏8千多萬票 習近平終於致電恭喜	川普撿了！拜登創紀錄贏8千多萬票 習近平終於致電恭喜 2020/11/25 21:52 字級：字字 美國總統...
川普「瘋狂退群」卻讓厚台灣？吳斯懷：影響不能輕忽	川普「瘋狂退群」卻讓厚台灣？吳斯懷：影響不能輕忽 2020/11/25 17:16 東森財經 字級：字字 20...
拜登外交核心三人組：選擇清一色歐巴馬老部下意味著什麼	拜登外交核心三人組：選擇清一色歐巴馬老部下意味著什麼 2020/11/25 09:42 字級：字字 ▲安東...
美環保署長取消訪台！外交部證實：行程變動	美環保署長取消訪台！外交部證實：行程變動 2020/11/25 09:15 東森財經 字級：字字 美國環保署...
快訊／包機訪台要花700萬！美環保署長取消行程 外交部：很惋惜	快訊／包機訪台要花700萬！美環保署長取消行程 外交部：很惋惜 2020/11/25 08:44 中央社 字級：字字 美國政府...
白宮批准向拜登提供情報簡報 支援政權交接工作	白宮批准向拜登提供情報簡報 支援政權交接工作 2020/11/25 08:06 中央社 字級：字字 美國政府...
美國大選：拜登獲准展開交接工作，多位資深前官員獲內閣提名	美國大選：拜登獲准展開交接工作，多位資深前官員獲內閣提名 2020/11/25 08:06 中央社 字級：字字 美國政府...
美國大選：川普承認拜登團隊過渡交接工作可以展開 但警告仍將挑戰選舉結果	美國大選：川普承認拜登團隊過渡交接工作可以展開 但警告仍將挑戰選舉結果 2020/11/25 07:23 中央社 字級：字字 在美國總統...
政權交接動後首度現身 川普讚股市破紀錄	政權交接動後首度現身 川普讚股市破紀錄 2020/11/25 07:23 中央社 字級：字字 在美國總統...
快訊／川普終於批准！允許拜登聽「總統每日簡報」	快訊／川普終於批准！允許拜登聽「總統每日簡報」 2020/11/25 06:52 東森新聞 責任編輯 陳唐誠 字級：字字 日前美國總統...
川普再踴躍板！內閣連州最高法院確認拜登勝選	川普再踴躍板！內閣連州最高法院確認拜登勝選 2020/11/25 06:16 中央社 字級：字字 內閣連州最高法院...
拜登、皮萊利多 美股收高連環首站上3萬點	拜登、皮萊利多 美股收高連環首站上3萬點 2020/11/25 06:14 中央社 字級：字字 美國總統川普...
結時：中國侵略性增強 拜登將延續抗中挺台路線	結時：中國侵略性增強 拜登將延續抗中挺台路線 2020/11/24 22:16 字級：字字 美國總統川普...
快訊／川普輸了！美總統署認拜登勝選：政權交接可開始	快訊／川普輸了！美總統署認拜登勝選：政權交接可開始 2020/11/24 07:31 東森新聞 責任編輯 林孝峰 字級：字字 美國總統...

0 Records mined from 0 Page Mining Page #1 - https://news.abc.net.tw/Search/Result?type=keyword&value=%25E6%2588%259C%25E7%2599%2588

WebHany Miner

Title	Content
美國大選：習近平視拜登當選的實電來了	美國大選：習近平視拜登當選的實電來了 2020/11/26 08:14 字級：字字 ▲(圖/REUTERS/EPA) ...
川普不想認輸！翻翻轉選結果 國務院通知職員展開過渡	川普不想認輸！翻翻轉選結果 國務院通知職員展開過渡 2020/11/26 06:48 中央社 字級：字字
川普撿了！拜登創紀錄贏8千多萬票 習近平終於致電恭喜	川普撿了！拜登創紀錄贏8千多萬票 習近平終於致電恭喜 2020/11/25 21:52 字級：字字 美國總統...
川普「瘋狂退群」卻讓厚台灣？吳斯懷：影響不能輕忽	川普「瘋狂退群」卻讓厚台灣？吳斯懷：影響不能輕忽 2020/11/25 17:16 東森財經 字級：字字 20...
拜登外交核心三人組：選擇清一色歐巴馬老部下意味著什麼	拜登外交核心三人組：選擇清一色歐巴馬老部下意味著什麼 2020/11/25 09:42 字級：字字 ▲安東...
美環保署長取消訪台！外交部證實：行程變動	美環保署長取消訪台！外交部證實：行程變動 2020/11/25 09:15 東森財經 字級：字字 美國環保署...
快訊／包機訪台要花700萬！美環保署長取消行程 外交部：很惋惜	快訊／包機訪台要花700萬！美環保署長取消行程 外交部：很惋惜 2020/11/25 08:44 中央社 字級：字字 美國政府...
白宮批准向拜登提供情報簡報 支援政權交接工作	白宮批准向拜登提供情報簡報 支援政權交接工作 2020/11/25 08:06 中央社 字級：字字 美國政府...
美國大選：拜登獲准展開交接工作，多位資深前官員獲內閣提名	美國大選：拜登獲准展開交接工作，多位資深前官員獲內閣提名 2020/11/25 08:06 中央社 字級：字字 美國政府...
美國大選：川普承認拜登團隊過渡交接工作可以展開 但警告仍將挑戰選舉結果	美國大選：川普承認拜登團隊過渡交接工作可以展開 但警告仍將挑戰選舉結果 2020/11/25 07:23 中央社 字級：字字 在美國總統...
政權交接動後首度現身 川普讚股市破紀錄	政權交接動後首度現身 川普讚股市破紀錄 2020/11/25 07:23 中央社 字級：字字 在美國總統...
快訊／川普終於批准！允許拜登聽「總統每日簡報」	快訊／川普終於批准！允許拜登聽「總統每日簡報」 2020/11/25 06:52 東森新聞 責任編輯 陳唐誠 字級：字字 日前美國總統...
川普再踴躍板！內閣連州最高法院確認拜登勝選	川普再踴躍板！內閣連州最高法院確認拜登勝選 2020/11/25 06:16 中央社 字級：字字 內閣連州最高法院...
拜登、皮萊利多 美股收高連環首站上3萬點	拜登、皮萊利多 美股收高連環首站上3萬點 2020/11/25 06:14 中央社 字級：字字 美國總統川普...
結時：中國侵略性增強 拜登將延續抗中挺台路線	結時：中國侵略性增強 拜登將延續抗中挺台路線 2020/11/24 22:16 字級：字字 美國總統川普...
快訊／川普輸了！美總統署認拜登勝選：政權交接可開始	快訊／川普輸了！美總統署認拜登勝選：政權交接可開始 2020/11/24 07:31 東森新聞 責任編輯 林孝峰 字級：字字 美國總統...

Export Captured Data

Select File Format

XLSX - Excel Workbook

Save as

If file already exists

Append Overwrite Update

Export

WebHany Miner

Title	Content
美國大選：習近平視拜登當選的實電來了	美國大選：習近平視拜登當選的實電來了 2020/11/26 08:14 字級：字字 ▲(圖/REUTERS/EPA) ...
川普不想認輸！翻翻轉選結果 國務院通知職員展開過渡	川普不想認輸！翻翻轉選結果 國務院通知職員展開過渡 2020/11/26 06:48 中央社 字級：字字 美國總統...
川普撿了！拜登創紀錄贏8千多萬票 習近平終於致電恭喜	川普撿了！拜登創紀錄贏8千多萬票 習近平終於致電恭喜 2020/11/25 21:52 字級：字字 美國總統...
川普「瘋狂退群」卻讓厚台灣？吳斯懷：影響不能輕忽	川普「瘋狂退群」卻讓厚台灣？吳斯懷：影響不能輕忽 2020/11/25 17:16 東森財經 字級：字字 20...
拜登外交核心三人組：選擇清一色歐巴馬老部下意味著什麼	拜登外交核心三人組：選擇清一色歐巴馬老部下意味著什麼 2020/11/25 09:42 字級：字字 ▲安東尼·布林肯...
美環保署長取消訪台！外交部證實：行程變動	美環保署長取消訪台！外交部證實：行程變動 2020/11/25 09:15 東森財經 字級：字字 美國環保署...
快訊／包機訪台要花700萬！美環保署長取消行程 外交部：很惋惜	快訊／包機訪台要花700萬！美環保署長取消行程 外交部：很惋惜 2020/11/25 08:44 中央社 字級：字字 美國政府...
白宮批准向拜登提供情報簡報 支援政權交接工作	白宮批准向拜登提供情報簡報 支援政權交接工作 2020/11/25 08:06 中央社 字級：字字 美國政府...
美國大選：拜登獲准展開交接工作，多位資深前官員獲內閣提名	美國大選：拜登獲准展開交接工作，多位資深前官員獲內閣提名 2020/11/25 08:06 中央社 字級：字字 美國政府...
美國大選：川普承認拜登團隊過渡交接工作可以展開 但警告仍將挑戰選舉結果	美國大選：川普承認拜登團隊過渡交接工作可以展開 但警告仍將挑戰選舉結果 2020/11/25 07:23 中央社 字級：字字 在美國總統...
政權交接動後首度現身 川普讚股市破紀錄	政權交接動後首度現身 川普讚股市破紀錄 2020/11/25 07:23 中央社 字級：字字 在美國總統...
快訊／川普終於批准！允許拜登聽「總統每日簡報」	快訊／川普終於批准！允許拜登聽「總統每日簡報」 2020/11/25 06:52 東森新聞 責任編輯 陳唐誠 字級：字字 日前美國總統...
川普再踴躍板！內閣連州最高法院確認拜登勝選	川普再踴躍板！內閣連州最高法院確認拜登勝選 2020/11/25 06:16 中央社 字級：字字 內閣連州最高法院...
拜登、皮萊利多 美股收高連環首站上3萬點	拜登、皮萊利多 美股收高連環首站上3萬點 2020/11/25 06:14 中央社 字級：字字 美國總統川普...
結時：中國侵略性增強 拜登將延續抗中挺台路線	結時：中國侵略性增強 拜登將延續抗中挺台路線 2020/11/24 22:16 字級：字字 美國總統川普...
快訊／川普輸了！美總統署認拜登勝選：政權交接可開始	快訊／川普輸了！美總統署認拜登勝選：政權交接可開始 2020/11/24 07:31 東森新聞 責任編輯 林孝峰 字級：字字 美國總統...

輿情系統

範例介紹




Infominer

忘記密碼?

登入 >

篩選測關注的議題

新冠肺炎本土案例 + 新增關鍵字群組 關鍵字群組使用量 1/100

☒ 關鍵字群組: 新冠肺炎本土案例

1

點選輸入關鍵字，如往右按+新增關鍵字為「OR(或)」例:搜尋新冠肺炎 或 COVID-19

包含關鍵字:

新冠肺炎 x COVID-19 x +

包含關鍵字:

確診 x +

2

往下新增包含關鍵字關係為「AND(且)」例:新冠肺炎 或COVID19 且包含確診

排除關鍵字:

境外移入 x +

3

排除關鍵字，提升分析準確性
(例:排除包含有"境外移入"之文章)

查看報表

擷取公開網站的資料

全頻道報表

是否置頂: ☒ 選擇頻道: ☒ 直接修改關鍵字: ☐ 第二層搜尋: ☐

搜尋標題及主文

最近的

頻道:

Ptt

臉書

新聞

國外新聞

論壇

影音新聞

部落格

臉書社團

instagram

Twitter

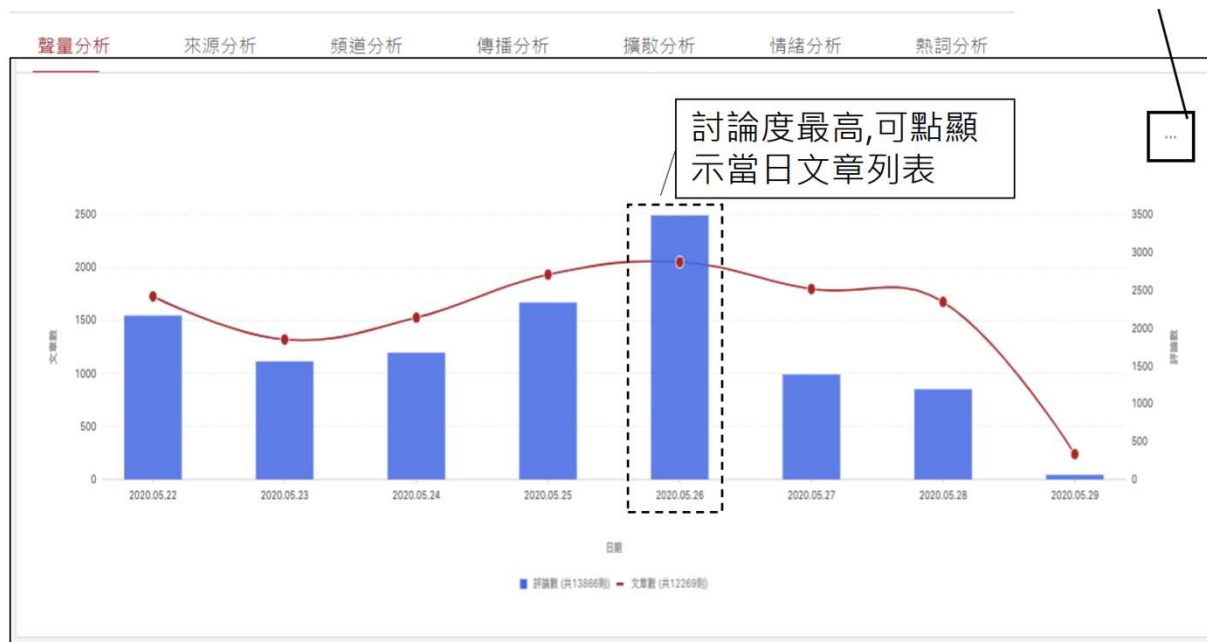
行銷

Youtube網紅

x



聲量(攸關性)



關鍵字文章/評論數(回文數) 趨勢圖

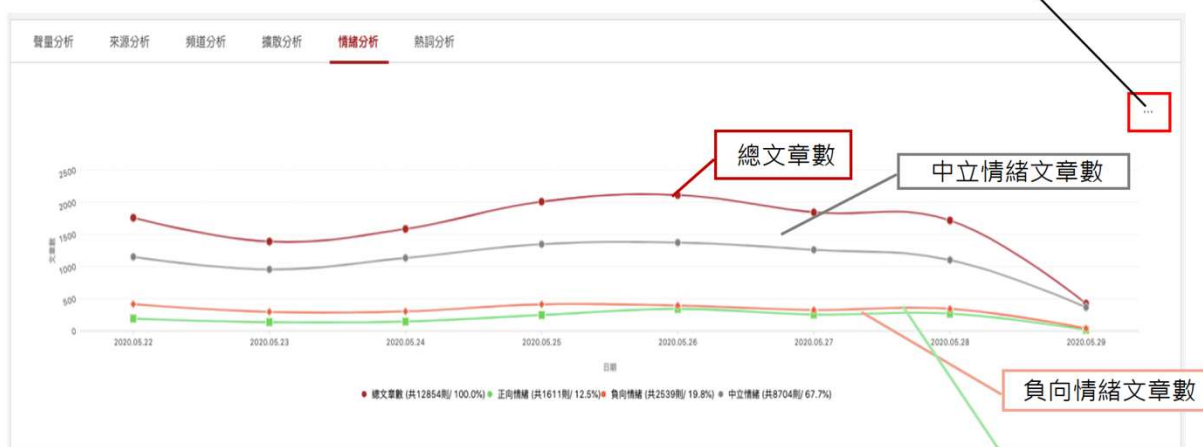
橫軸：日期

縱軸：關鍵字文章數(折線)

文章評論數(回文數)(直條)



情感(篇好與態度)



含關鍵字文章之情緒曲線圖
(橫軸：日期 / 縱軸：文章數)



時間: 2020/05/30 19:00

篩選聲量分數、文章數範圍

本日熱門議題分類

選擇頻道來源

排序方式：聲量分數/文章數

頻道來源：

排序方式：

熱門關鍵字：

點擊關鍵字，進一步搜尋（例：點擊「社會」）

Category	Count
香港	169
美國	223
川普	247
武漢肺炎	177
國際	61
生活	96
政治	167
防疫	75
疫情	44
司法	44
避友宜	35
隔離	43
轉運	43
肥宅	42
總統	48
連勝	60
韓國瑜	93
通商	90
經濟	98
華南昌	98
武漢肺炎	133
通商辣香化	116
韓成昌	98

關鍵字熱度變化:

點擊欄位，關鍵字按照數值大小排序

關鍵字	近24小時文章數	前期數值	熱度變化
柯文哲	31	11	↑ 181.8 %
婚姻	48	26	↑ 84.6 %
運動	60	54	↑ 11.1 %
武漢肺炎	124	123	↑ 0.8 %
侯友宜	36	37	↓ -2.7 %
港版国安法	180	185	↓ -2.7 %
陳冠榮	35	36	↓ -2.8 %
中華職棒	34	35	↓ -2.9 %
美國	223	230	↓ -3.0 %
香港	171	177	↓ -3.4 %



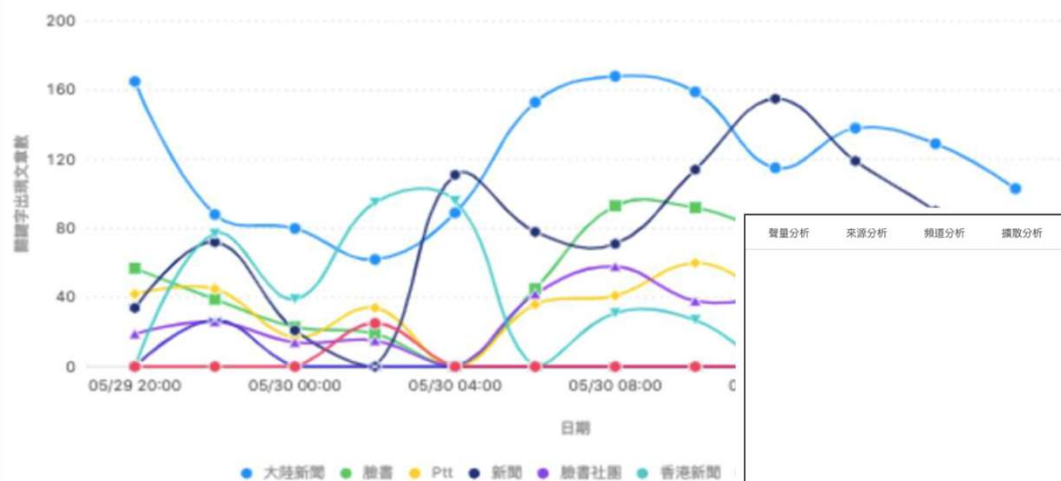
關鍵字搜尋結果分析

搜尋關鍵字: 美國

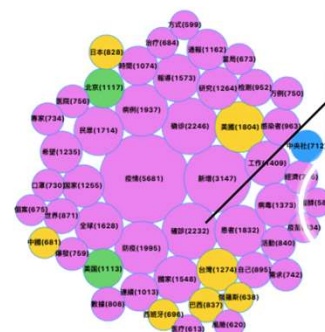
聲量分析

來源分析

文章列表



聲量分析 來源分析 頻道分析 擴散分析 情緒分析 熱詞分析



點選氣泡，觀看含此關鍵字文章

可挑選字詞類別

全部

人物

國家

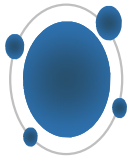
地點

事件

組織

關鍵字

含關鍵字文章之字詞氣泡圖



提供目前台灣地區輿情系統的供應商供參考

1. InfoMiner <https://www.largitdata.com/>
2. LOWI <https://www.chiefappc.com>
3. Opview <https://www.opview.com.tw/>
4. Quickseek <https://www.quickseek.com.tw/>

機器學習範例

Azure Machine Learning Studio

<https://studio.azureml.net>

Azure Machine Learning Studio

- 資料集: iris.csv

-

步驟:

- 1.Data 匯入
- 2.Select Column in dataset
- 3.APPLY SQL transformation
- 4.Split data (將檔案分割)
- 5.(左邊開始寫演算法ex Multiclass Neural Network)
- 6.Train Model
- 7.Score Model
- 8.Evalute Model

-

-

Azure Machine Learning Studio

 PROJECTS

 EXPERIMENTS

 WEB SERVICES

 NOTEBOOKS

 DATASETS

 TRAINED MODELS

 SETTINGS

projects preview

NAME	AUTHOR	CONTENTS	LAST USED	
Create project 				

1. 建立專案(Projects)

 NEW

 DELETE

Azure Machine Learning Studio

PROJECTS

EXPERIMENTS

WEB SERVICES

NOTEBOOKS

DATASETS

TRAINED MODELS

SETTINGS

experiments

MY EXPERIMENTS SAMPLES

	NAME	AUTHOR	STATUS	LAST EDITED	PROJECT	
No experiments found						

0 items selected

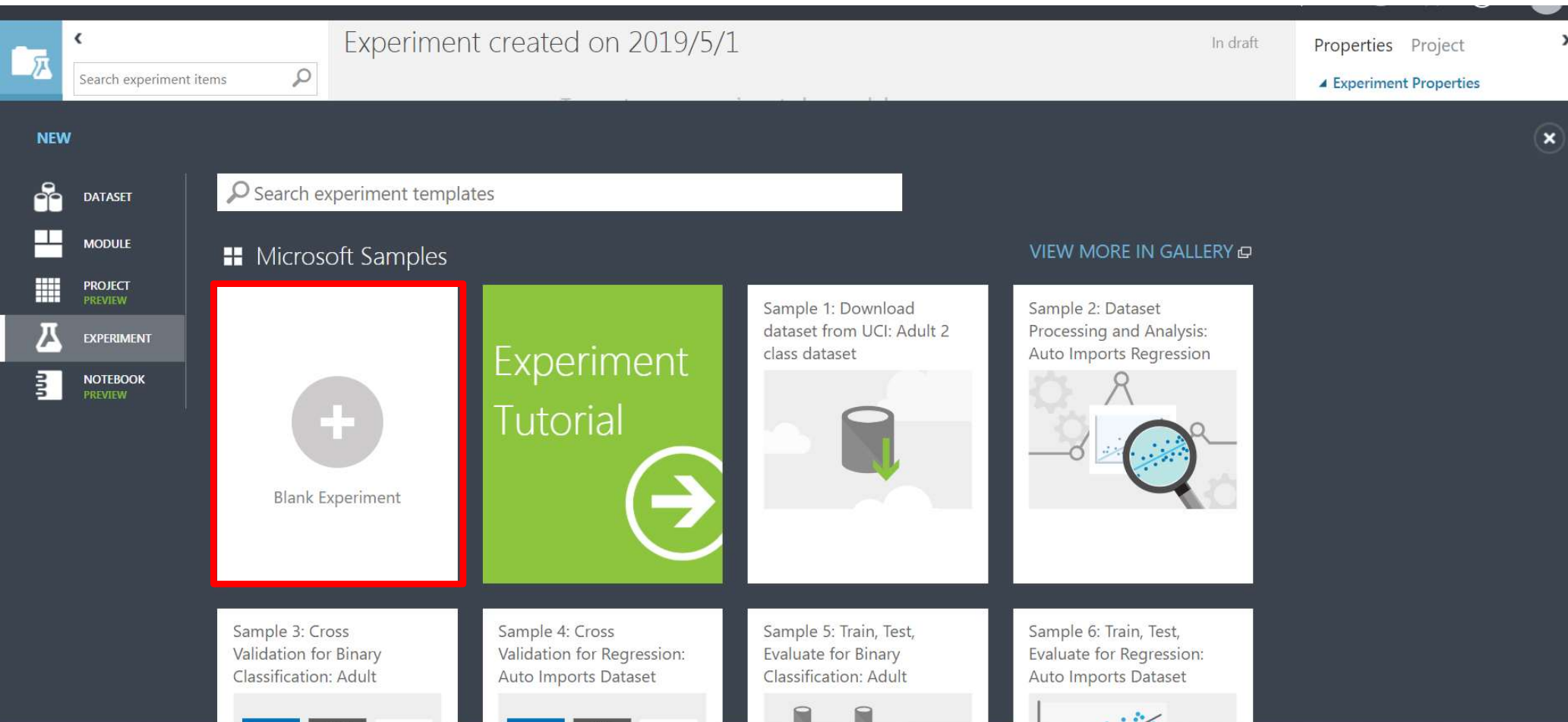
新增實驗experiments

+ NEW

DELETE

ADD TO PROJECT

Azure Machine Learning Studio



The screenshot displays the Azure Machine Learning Studio interface. At the top, a header bar shows the experiment name "Experiment created on 2019/5/1", its status "In draft", and links to "Properties" and "Project". Below the header, a left sidebar contains navigation options: "NEW", "DATASET", "MODULE", "PROJECT", "EXPERIMENT" (highlighted), and "NOTEBOOK". The main workspace features a search bar for "Search experiment templates" and a section titled "Microsoft Samples". The "Blank Experiment" template is highlighted with a red border. Other templates include "Experiment Tutorial", "Sample 1: Download dataset from UCI: Adult 2 class dataset", "Sample 2: Dataset Processing and Analysis: Auto Imports Regression", "Sample 3: Cross Validation for Binary Classification: Adult", "Sample 4: Cross Validation for Regression: Auto Imports Dataset", "Sample 5: Train, Test, Evaluate for Binary Classification: Adult", and "Sample 6: Train, Test, Evaluate for Regression: Auto Imports Dataset". A "VIEW MORE IN GALLERY" link is also present.

Experiment created on 2019/5/1 In draft Properties Project
Search experiment items Search experiment templates
NEW
DATASET
MODULE
PROJECT
EXPERIMENT
NOTEBOOK
Microsoft Samples
Blank Experiment
Experiment Tutorial
Sample 1: Download dataset from UCI: Adult 2 class dataset
Sample 2: Dataset Processing and Analysis: Auto Imports Regression
Sample 3: Cross Validation for Binary Classification: Adult
Sample 4: Cross Validation for Regression: Auto Imports Dataset
Sample 5: Train, Test, Evaluate for Binary Classification: Adult
Sample 6: Train, Test, Evaluate for Regression: Auto Imports Dataset
VIEW MORE IN GALLERY

Azure Machine Learning Studio

The screenshot displays the Azure Machine Learning Studio interface. The central canvas shows a workflow diagram with three dashed boxes connected by arrows, representing a sequence of steps. A red box highlights the left sidebar, which contains a search bar and a list of modules categorized by type: Data Input and Output, Data Transformation, Feature Selection, Machine Learning, OpenCV Library Modules, Python Language Modules, R Language Modules, Statistical Functions, Text Analytics, Time Series, Web Service, and Deprecated. A red box also highlights the right sidebar, which contains the 'Properties' panel with sections for 'Experiment Properties' (Status Code: InDraft), 'Summary' (a text area for a brief description), and 'Description' (a text area for a detailed description). The bottom of the interface features a dark toolbar with icons for 'NEW', 'RUN HISTORY', 'SAVE', 'SAVE AS', 'DISCARD CHANGES', 'RUN', 'SET UP WEB SERVICE', and 'PUBLISH TO GALLERY'. The text 'Experiment created on 2019/5/1' is visible at the top of the canvas area.

Experiment created on 2019/5/1

In draft

Search experiment items

To create your experiment, drag and drop datasets and modules here

基本樣本模式

Drag Items Here

Mini Map

Execute Python Script

Create R Model

Execute R Script

Create R Model

Creates an R model using custom resources

Properties Project

Experiment Properties

STATUS CODE InDraft

Summary

Enter a few sentences describing your experiment (up to 140 characters).

Description

Enter the detailed description for your experiment.

Quick Help

+ NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY

1.資料集(Dataset)上傳

Azure Machine Learning Studio

 PROJECTS

 EXPERIMENTS

 WEB SERVICES

 NOTEBOOKS

 DATASETS

 TRAINED MODELS

 SETTINGS

datasets

MY DATASETS SAMPLES

NAME	SUBMITTED BY	DESCRIPTION	DATA TYPE	CREATED	SIZE	PROJECT
No datasets found						

 NEW

 DOWNLOAD

 DELETE

 OPEN IN NOTEBOOK

 GENERATE DATA ACCESS CODE...

 ADD TO PROJECT



Azure Machine Learning Studio

 PROJECTS

datasets

MY DATASETS SAMPLES

NEW

 DATASET

 MODULE

 PROJECT
PREVIEW

 EXPERIMENT

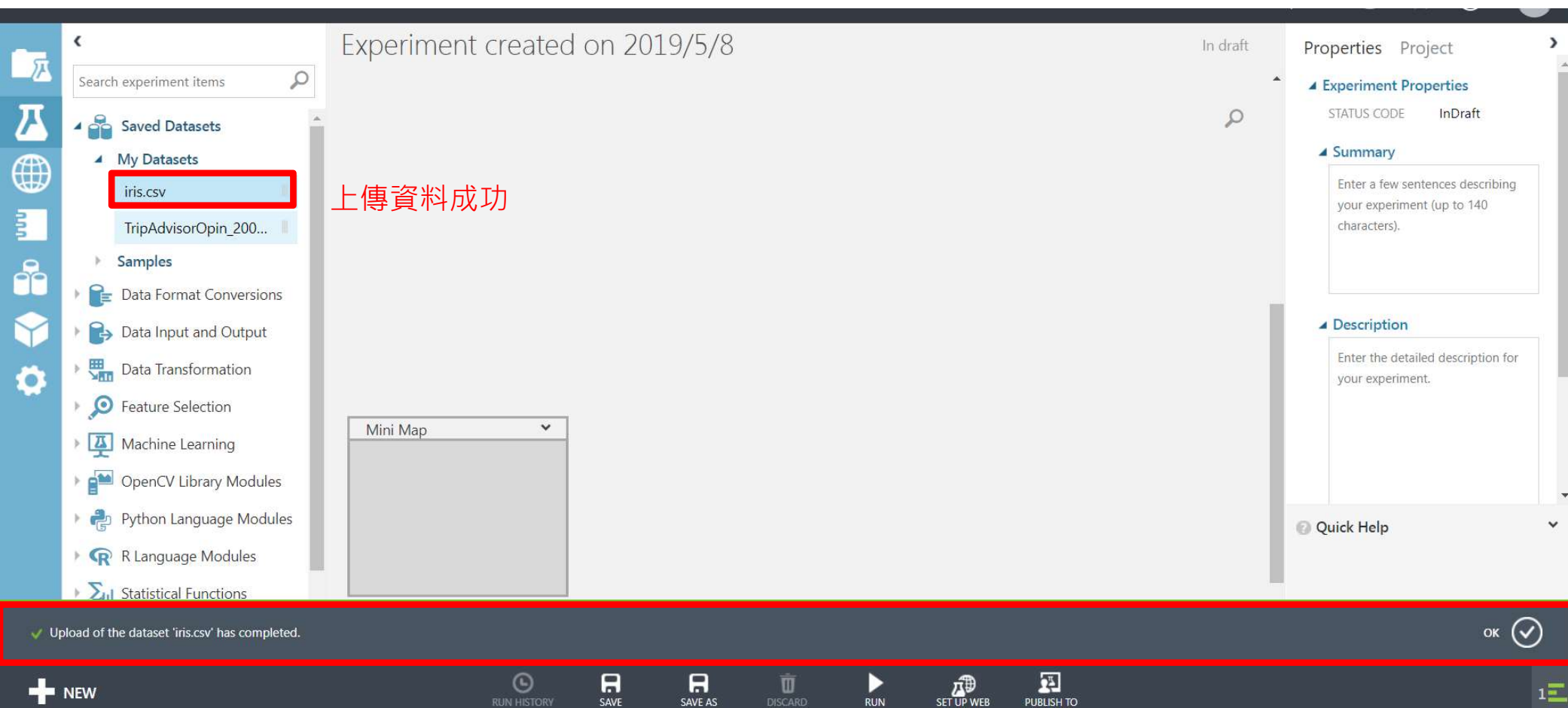
 NOTEBOOK
PREVIEW

 FROM
LOCAL FILE 

Upload a new dataset from a local file

studio.azureml.net/Home/.../83b7c4c75bbd48bd89b4c3b2e4839cda

Azure Machine Learning Studio



The screenshot displays the Azure Machine Learning Studio interface. On the left, the 'My Datasets' section is expanded, and the 'iris.csv' dataset is highlighted with a red rectangle. The central workspace shows the text 'Experiment created on 2019/5/8' and a 'Mini Map' window. On the right, the 'Properties' pane is visible, showing 'Experiment Properties' with a status of 'InDraft'. A red text overlay '上傳資料成功' (Upload data successful) is positioned over the dataset list. At the bottom, a dark blue notification bar with a red border contains the message 'Upload of the dataset 'iris.csv' has completed.' with an 'OK' button and a checkmark icon.

Experiment created on 2019/5/8

In draft

Search experiment items

Saved Datasets

My Datasets

iris.csv

TripAdvisorOpin_200...

Samples

Data Format Conversions

Data Input and Output

Data Transformation

Feature Selection

Machine Learning

OpenCV Library Modules

Python Language Modules

R Language Modules

Statistical Functions

Mini Map

Properties Project

Experiment Properties

STATUS CODE InDraft

Summary

Enter a few sentences describing your experiment (up to 140 characters).

Description

Enter the detailed description for your experiment.

Quick Help

Upload of the dataset 'iris.csv' has completed.

OK

NEW

RUN HISTORY

SAVE

SAVE AS

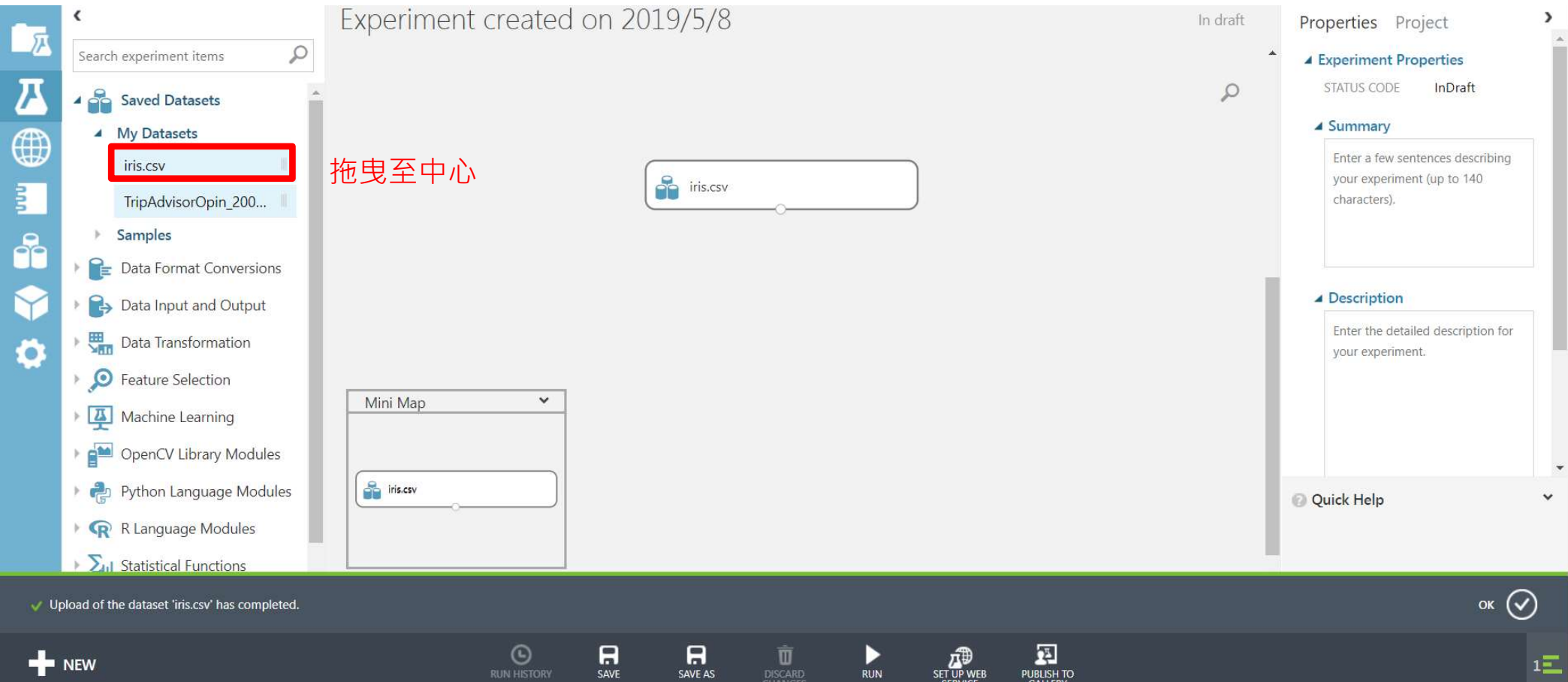
DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY

Azure Machine Learning Studio



The screenshot displays the Azure Machine Learning Studio interface. On the left, a sidebar contains a search bar and a list of categories: Saved Datasets, My Datasets, Samples, Data Format Conversions, Data Input and Output, Data Transformation, Feature Selection, Machine Learning, OpenCV Library Modules, Python Language Modules, R Language Modules, and Statistical Functions. Under 'My Datasets', the file 'iris.csv' is highlighted with a red rectangle. A red text label '拖曳至中心' (Drag to center) is positioned next to it. In the main workspace, the text 'Experiment created on 2019/5/8' is at the top left, and 'In draft' is at the top right. A search bar is also present. The workspace contains a diagram with a box labeled 'iris.csv'. A 'Mini Map' window at the bottom left shows a smaller version of this diagram. On the right, a 'Properties' pane shows 'Experiment Properties' with 'STATUS CODE' set to 'InDraft'. Below this are sections for 'Summary' and 'Description', each with a text input area. At the bottom, a status bar shows a green checkmark and the message 'Upload of the dataset 'iris.csv' has completed.' with an 'OK' button. The bottom toolbar includes icons for '+ NEW', 'RUN HISTORY', 'SAVE', 'SAVE AS', 'DISCARD CHANGES', 'RUN', 'SET UP WEB SERVICE', and 'PUBLISH TO GALLERY'.

Experiment created on 2019/5/8

In draft

Search experiment items

Saved Datasets

My Datasets

iris.csv

TripAdvisorOpin_200...

Samples

Data Format Conversions

Data Input and Output

Data Transformation

Feature Selection

Machine Learning

OpenCV Library Modules

Python Language Modules

R Language Modules

Statistical Functions

拖曳至中心

iris.csv

Mini Map

iris.csv

Properties Project

Experiment Properties

STATUS CODE InDraft

Summary

Enter a few sentences describing your experiment (up to 140 characters).

Description

Enter the detailed description for your experiment.

Quick Help

Upload of the dataset 'iris.csv' has completed.

OK

+ NEW

RUN HISTORY

SAVE

SAVE AS

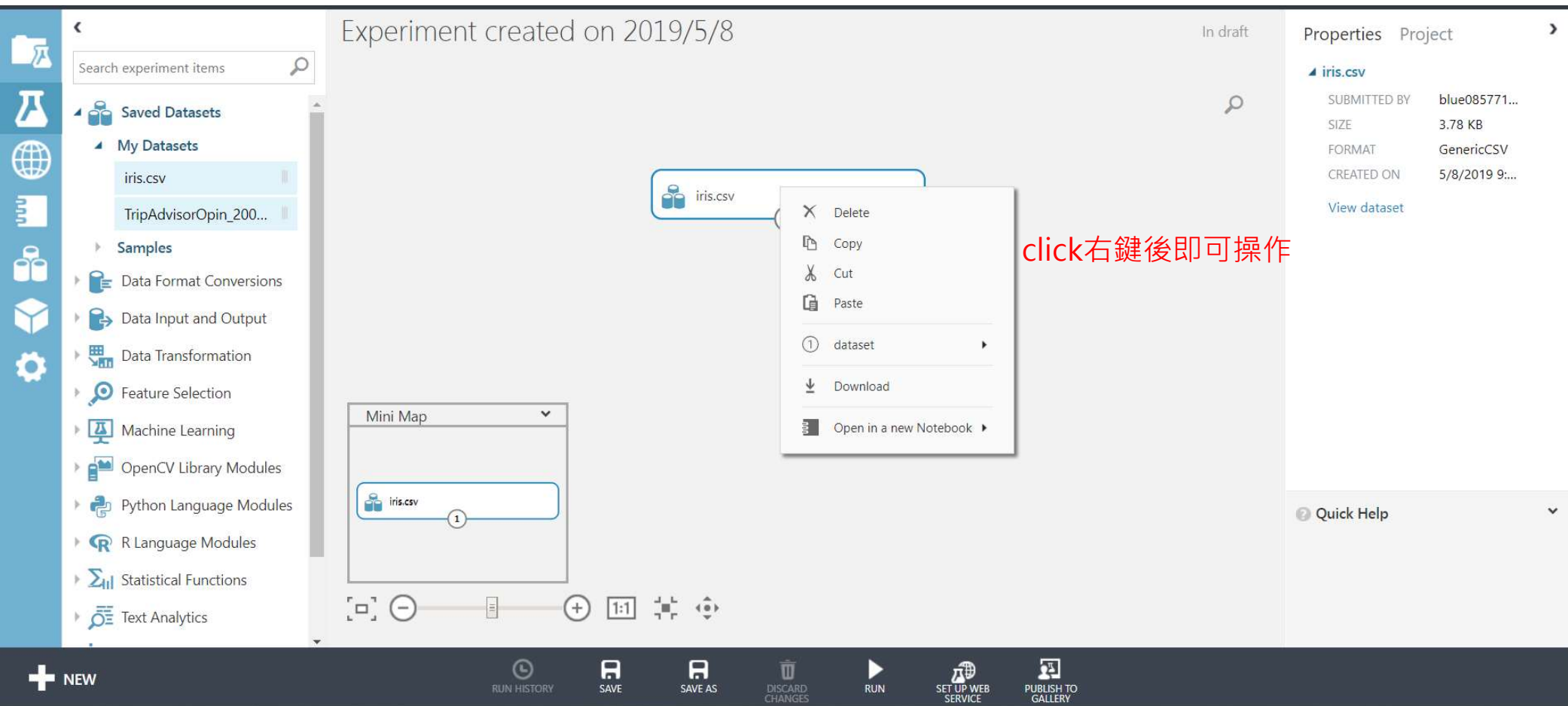
DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY

Azure Machine Learning Studio



The screenshot displays the Azure Machine Learning Studio interface. On the left, a sidebar contains a search bar and a list of categories: Saved Datasets, My Datasets, Samples, Data Format Conversions, Data Input and Output, Data Transformation, Feature Selection, Machine Learning, OpenCV Library Modules, Python Language Modules, R Language Modules, Statistical Functions, and Text Analytics. Under 'My Datasets', 'iris.csv' is selected. The main workspace shows an experiment created on 2019/5/8, currently in draft status. A dataset icon labeled 'iris.csv' is present, and a right-click context menu is open over it, showing options: Delete, Copy, Cut, Paste, dataset (with a sub-menu arrow), Download, and Open in a new Notebook (with a sub-menu arrow). A red text annotation 'click右鍵後即可操作' (click right button then you can operate) points to the context menu. A 'Mini Map' window at the bottom left shows a thumbnail of the experiment flow with the 'iris.csv' dataset at step 1. The bottom toolbar includes icons for NEW, RUN HISTORY, SAVE, SAVE AS, DISCARD CHANGES, RUN, SET UP WEB SERVICE, and PUBLISH TO GALLERY. On the right, the 'Properties' pane for 'iris.csv' shows details: SUBMITTED BY (blue085771...), SIZE (3.78 KB), FORMAT (GenericCSV), and CREATED ON (5/8/2019 9:...), along with a 'View dataset' link and a 'Quick Help' section.

Experiment created on 2019/5/8

In draft

Properties Project

iris.csv

SUBMITTED BY blue085771...

SIZE 3.78 KB

FORMAT GenericCSV

CREATED ON 5/8/2019 9:...

View dataset

Quick Help

click右鍵後即可操作

Search experiment items

Saved Datasets

My Datasets

iris.csv

TripAdvisorOpin_200...

Samples

Data Format Conversions

Data Input and Output

Data Transformation

Feature Selection

Machine Learning

OpenCV Library Modules

Python Language Modules

R Language Modules

Statistical Functions

Text Analytics

Mini Map

iris.csv

1

1:1

NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY

2.Select Columns in Dataset (選擇欄位/構面)



Azure Machine Learning Studio

國立 Experiments - Microsoft Azure Microsoft Professional Program

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/Draft/ViewExp...

應用程式 Google Google Analytics I... Facebook 0514前__DHL_Exp... Research 統計觀念與軟體操... 青年資源讀 - 2019... 電商 研究所 YouTube 至 mp3...

Microsoft Azure Machine Learning Studio Vicky Chen-Free-Workspace

Experiment created on 2019/5/8 In draft

Search experiment items

Data Transformation

- Filter
- Learning with Counts
- Manipulation**
 - Add Columns
 - Add Rows
 - Apply SQL Transformati...
 - Clean Missing Data
 - Convert to Indicator Val...
 - Edit Metadata
 - Group Categorical Values
 - Join Data
 - Remove Duplicate Rows
 - Select Columns in Datas...
 - Select Columns Transfor...
 - SMOTE
- Sample and Split

1. 選擇資料集dataset

iris.csv

2. Select Columns in Dataset

1

Mini Map

iris.csv

Select Columns in Dataset

1

Properties Project

Select Columns in Dataset

Select columns

Selected columns:
Launch the selector tool to make a selection

Launch column selector

Quick Help

Selects columns to include or exclude from a dataset in an operation. Formerly known as Select Columns

Upload of the dataset 'iris.csv' has completed. OK

+ NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY



Azure Machine Learning Studio

國立 Experiments - Microsoft Azure Microsoft Professional Program

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/Draft/ViewExp...

應用程式 Google Google Analytics I... Facebook 0514前__DHL_Exp... Research 統計觀念與軟體操... 青年資源讀 - 2019... 電商 研究所 YouTube 至 mp3...

Microsoft Azure Machine Learning Studio Vicky Chen-Free-Workspace

Experiment created on 2019/5/8 In draft

Search experiment items

Data Transformation

- Filter
- Learning with Counts
- Manipulation**
 - Add Columns
 - Add Rows
 - Apply SQL Transformati...
 - Clean Missing Data
 - Convert to Indicator Val...
 - Edit Metadata
 - Group Categorical Values
 - Join Data
 - Remove Duplicate Rows
 - Select Columns in Datas...
 - Select Columns Transfor...
 - SMOTE
- Sample and Split

iris.csv

Select Columns in Dataset 1

Mini Map

iris.csv

Select Columns in Dataset 1

Properties Project

Select Columns in Dataset

Select columns

Selected columns:
Launch the selector tool to make a selection

Launch column selector

選擇欄位

Quick Help

Selects columns to include or exclude from a dataset in an operation. Formerly known as

Upload of the dataset 'iris.csv' has completed. OK

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY



Azure Machine Learning Studio

Microsoft Azure Machine Learning Studio

Experiment created on 2019/5/8

In draft

Properties Project

Search experiment items

Data Transformation

- Filter
- Learning with Counts
- Manipulation
 - Add Columns
 - Add Rows
 - Apply SQL Transformati...
 - Clean Missing Data
 - Convert to Indicator Val...
 - Edit Metadata
 - Group Categorical Values
 - Join Data
 - Remove Duplicate Rows
 - Select Columns in Datas...
 - Select Columns Transfor...
 - SMOTE
 - Sample and Split

Select columns

BY NAME

WITH RULES

AVAILABLE COLUMNS

All Types search columns

Sepal_Length
Sepal_Width
Petal_Length
Petal_Width
Species

SELECTED COLUMNS

All Types search columns

全部欄位匯入 ->

5 columns available

0 columns selected

Select Columns in Dataset

1

Upload of the dataset 'iris.csv' has completed.

+ NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY

Quick Help

Selects columns to include or exclude from a dataset in an operation. Formerly known as Select Columns

Launch column selector

Selected columns: Launch the selector tool to make a selection

Launch column selector

YouTube 至 mp3...

國立

Experiments - Microsoft Azure

Microsoft Professional Program

+

← → ↺

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...

☆

🔒

🚀

🌈

🔥

📁

🌐

🔦

應用程式

Google

Google Analytics I...

Facebook

0514前_DHL_Exp...

Research

統計觀念與軟體操...

青年資源讀 - 2019...

電商

研究所

YouTube 至 mp3...

»

Microsoft Azure Machine Learning Studio

Vicky Chen-Free-Workspace

?

👤

😊

👤

Experiment created on 2019/5/8

Finished running ✓

Properties

Project

Search ex

Experiment created on 2019/5/8 ▶ Select Columns in Dataset ▶ Results dataset

✕

rows
150

columns
5

view as


Sepal_Length	Sepal_Width	Petal_Length	Petal_Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3	1.4	0.1	setosa

Statistics

Visualizations



To view, select a column in the table.

Run完後
將檔案開啟visualize看檔案是否有匯入

✓ Upload of the dataset 'iris.csv' has completed.

OK

+

NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY

1

#3.APPLY SQL transformation

4.Split data (將檔案分割)



Azure Machine Learning Studio

國立 Experiments - Microsoft Azure Microsoft Professional Program

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...

應用程式 Google Google Analytics I... Facebook 0514前__DHL_Exp... Research 統計觀念與軟體操... 青年資源讀 - 2019... 電商 研究所 YouTube 至 mp3...

Microsoft Azure Machine Learning Studio Vicky Chen-Free-Workspace

Iris_example_0508 In draft Draft saved at 上午9:56:05

Search experiment items

- Data Format Conversions
- Data Input and Output
- Data Transformation
 - Filter
 - Learning with Counts
 - Manipulation
 - Sample and Split
 - Partition and Sample
 - Split Data
 - Scale and Reduce
- Feature Selection
- Machine Learning
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
- Statistical Functions
- Text Analytics

iris.csv

Select Columns in Dataset

Mini Map

Experiment Properties

START TIME	5/8/2019 9:...
END TIME	5/8/2019 9:...
STATUS CODE	InDraft
STATUS DETAILS	None

Summary

Description

Quick Help

Upload of the dataset 'iris.csv' has completed. OK

+ NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY



Azure Machine Learning Studio

國立 Experiments - Microsoft Azure Microsoft Professional Program

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...

應用程式 Google Google Analytics Facebook 0514前__DHL_Exp... Research 統計觀念與軟體操... 青年資源讀 - 2019... 電商 研究所 YouTube 至 mp3...

Microsoft Azure Machine Learning Studio Vicky Chen-Free-Workspace

Iris_example_0508 In draft Draft saved at 上午9:59:16

Search experiment items

- Data Format Conversions
- Data Input and Output
- Data Transformation
 - Filter
 - Learning with Counts
 - Manipulation
 - Sample and Split
 - Partition and Sample
 - Split Data
- Scale and Reduce
- Feature Selection
- Machine Learning
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
- Statistical Functions
- Text Analytics

iris.csv

Select Columns in Dataset

Split Data 1 2

Split Data 分割檔案

Properties Project

Split Data

Splitting mode
Split Rows

Fraction of rows in the first set
0.5

☒ Randomized split

Random seed
0

Stratified split
False

查看是否需要更改設定格式內容

Quick Help

Split the rows of a dataset into two distinct sets

Upload of the dataset 'iris.csv' has completed. OK

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

Azure Machine Learning Studio

Microsoft Azure Machine Learning Studio

Experiments - Microsoft Azure

Microsoft Professional Program

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...

應用程式 Google Google Analytics I... Facebook 0514前__DHL_Exp... Research 統計觀念與軟體操... 青年資源讀 - 2019... 電商 研究所 YouTube 至 mp3...

Microsoft Azure Machine Learning Studio Vicky Chen-Free-Workspace

Iris_example_0508

Finished running ✓

Properties Project

Split Data

Splitting mode
Split Rows

Fraction of rows in the first split
0.5

Randomized split
✓

Random seed
0

Stratified split
False

START TIME 5/8/2019 1...

END TIME 5/8/2019 1...

ELAPSED TIME 0:00:02.133

STATUS CODE Finished

STATUS DETAILS None

iris.csv

Select Columns in Dataset ✓

Split Data 1

Mini Map

Upload of the dataset 'iris.csv' has completed.

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

可以看到會有兩個dataset

Results dataset1
Results dataset2



Azure Machine Learning Studio

國立 Experiments - Microsoft Azure Microsoft Professional Program

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...

應用程序 Google Google Analytics I... Facebook 0514前_DHL_Exp... Research 統計觀念與軟體操... 青年資源讀 - 2019... 電商 研究所 YouTube 至 mp3...

Microsoft Azure Machine Learning Studio Vicky Chen-Free-Workspace

Iris_example_0508 In draft

Search experiment items

- Saved Datasets
- Data Format Conversions
- Data Input and Output
- Data Transformation
 - Filter
 - Learning with Counts
 - Manipulation
 - Sample and Split
 - Partition and Sample
 - Split Data
- Scale and Reduce
- Feature Selection
- Machine Learning
- OpenCV Library Modules
- Python Language Modules
- R Language Modules

iris.csv

Select Columns in Dataset

Split Data 1 2

Mini Map

Properties Project

Split Data

Splitting mode

Split Rows

Fraction of rows in the first set

0.7

更改為0.7 (70%/30%)

Randomized split

Random seed

0

Stratified split

False

Quick Help

Split the rows of a dataset into two distinct sets

Upload of the dataset 'iris.csv' has completed. OK

+ NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

5. 選擇機器學習演算法(Initial Model) Multiclass Neural Network



Azure Machine Learning Studio

國立 Experiments - Microsoft Azure Microsoft Professional Program

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...

應用程式 Google Google Analytics I... Facebook 0514前_DHL_Exp... Research 統計觀念與軟體操... 青年資源讀 - 2019... 電商 研究所 YouTube 至 mp3...

Microsoft Azure Machine Learning Studio Vicky Chen-Free-Workspace

Iris_example_0508 In draft Draft saved at 上午10:03:08

Search experiment items

- Partition and Sample
- Split Data
- Scale and Reduce
- Feature Selection
- Machine Learning
 - Evaluate
 - Initialize Model
 - Anomaly Detection
 - Classification
 - Clustering
 - Regression
 - Score
 - Train
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
- Statistical Functions

選擇演算法

iris.csv

Select Columns in Dataset

Split Data

Mini Map

Properties Project

Split Data

Splitting mode: Split Rows

Fraction of rows in the first...: 0.7

Randomized split: ☒

Random seed: 0

Stratified split: False

START TIME: 5/8/2019 1...

END TIME: 5/8/2019 1...

ELAPSED TIME: 0:00:02.422

STATUS CODE: Finished

STATUS DETAILS: None

View output log

Quick Help

Split the rows of a dataset into two distinct sets

Upload of the dataset 'iris.csv' has completed.

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

6.Train Model



Azure Machine Learning Studio

國立 Experiments - Microsoft Azure

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...

應用程式 Google Google Analytics I... Facebook 0514前__DHL_Exp... Research 統計觀念與軟體操... 青年資源讀 - 2019... 電商 研究所 YouTube 至 mp3...

Microsoft Azure Machine Learning Studio Vicky Chen-Free-Workspace

Iris_example_0508 In draft Draft saved at 上午10:25:27

Search experiment items

- Initialize Model
 - Anomaly Detection
 - Classification
 - Clustering
 - Regression
- Score
- Train
 - Sweep Clustering
 - Train Anomaly Detection Model
 - Train Clustering Model
 - Train Matchbox Recommender
 - Train Model
 - Tune Model Hyperparameters
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
- Statistical Functions
- Text Analytics
- Time Series

iris.csv

Select Columns in Dataset

Multiclass Decision Forest

Split Data

選擇 Train Model

Train Model

1

Properties Project

Train Model

Label column

Selected columns:
Launch the selector tool to make a selection

Launch column selector

Quick Help

Train a previously created classification or regression model
(more help...)

+ NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY



Azure Machine Learning Studio

國立 Experiments - Microsoft Azure

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...

應用程式 Google Google Analytics I... Facebook 0514前__DHL_Exp... Research 統計觀念與軟體操... 青年資源讀 - 2019... 電商 研究所 YouTube 至 mp3...

Microsoft Azure Machine Learning Studio Vicky Chen-Free-Workspace

Iris_example_0508 In draft Draft saved at 上午10:25:40

Search experiment items

- Initialize Model
 - Anomaly Detection
 - Classification
 - Clustering
 - Regression
- Score
- Train
 - Sweep Clustering
 - Train Anomaly Detection Model
 - Train Clustering Model
 - Train Matchbox Recommender
 - Train Model
 - Tune Model Hyperparameters
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
- Statistical Functions
- Text Analytics
- Time Series

iris.csv

Select Columns in Dataset

Multiclass Decision Forest

Split Data

Train Model

Value required.

需要比對的資料 (value required)

1

Properties Project

Train Model

Label column

Selected columns:
Launch the selector tool to make a selection

Launch column selector

Quick Help

Train a previously created classification or regression model
(more help...)

+ NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

國立

Experiments - Microsoft Azure

+

← → ↺

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...

☆

🛡️

🚀

🌈

🔥

📄

🌐

🔦

應用程式

Google

Google Analytics I...

Facebook

0514前_DHL_Exp...

Research

統計觀念與軟體操...

青年資源讀 - 2019...

電商

研究所

YouTube 至 mp3...

»

Microsoft Azure Machine Learning Studio

Vicky Chen-Free-Workspace

?

👤

😊

👤

🔍 Search experiment items 🔍

Initialize Model

Anomaly Detection

Classification

Clustering

Regression

Score

Train

Sweep Clustering

Train Anomaly Detection Model

Train Clustering Model

Train Matchbox Recommender

Train Model

Tune Model Hyperparameters

OpenCV Library Modules

Python Language Modules

R Language Modules

Statistical Functions

Text Analytics

Time Series

Iris_example_0508

In draft

Draft saved at 上午10:26:02

iris.csv

Select Columns in Dataset ✓

Multiclass Decision Forest

Split Data ✓

Dataset (Dataset, DataTableDotNet, GenericCSV, GenericCSVNoHe

Train Model 1

📏

⊖

⊕

1:1

🔍

🔄

Properties Project

Train Model

Label column

Selected columns:
Launch the selector tool to
make a selection

Launch column selector

Quick Help

Train a previously created classification or
regression model
(more help...)

+ NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY



Azure Machine Learning Studio

Microsoft Azure Machine Learning Studio

Vicky Chen-Free-Workspace

Iris_example_0508

In draft

Draft saved at 上午10:26:19

Search experiment items

Initialize Model

- Anomaly Detection
- Classification
- Clustering
- Regression
- Score
- Train
 - Sweep Clustering
 - Train Anomaly Detection Model
 - Train Clustering Model
 - Train Matchbox Recommendation Model
 - Train Model
 - Tune Model Hyperparameters
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
- Statistical Functions
- Text Analytics
- Time Series

BY NAME

WITH RULES

AVAILABLE COLUMNS

All Types search columns

Sepal_Length
Sepal_Width
Petal_Length
Petal_Width

4 columns available

SELECTED COLUMNS

All Types search columns

Species

比對預測的column

1 columns selected

Train Model

Label column

Selected columns:
Launch the selector tool to make a selection

Launch column selector

Quick Help

Train a previously created classification or regression model (more help...)

+ NEW

RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

6.Train Model

7.Score Model

8.Evalute Model

連結Initial Train Score and Evaluate

Azure Machine Learning Studio

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The browser address bar shows the URL: <https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...>. The workspace name is "Vicky Chen-Free-Workspace".

The left sidebar contains a search bar and a list of experiment items under the "Machine Learning" category:

- Machine Learning
 - Evaluate
 - Initialize Model
 - Anomaly Detection
 - Classification
 - Clustering
 - Regression
 - Score
 - Apply Transformation
 - Assign Data to Clusters
 - Score Matchbox Recommender
 - Score Model
 - Train
 - OpenCV Library Modules
 - Python Language Modules
 - R Language Modules
 - Statistical Functions
 - Text Analytics
 - Time Series

The main workspace shows a workflow titled "Iris_example_0508" in draft status. The workflow steps are:

- iris.csv
- Select Columns in Dataset (checked)
- Split Data (checked)
- Multiclass Decision Forest
- Train Model
- Score Model (highlighted with a tooltip: "Score Model. Score a trained classification or regression model")

The right sidebar shows the "Properties" tab for the "Score Model" step, with the option "Append score columns ..." checked.

The bottom toolbar includes icons for: NEW, RUN HISTORY, SAVE, SAVE AS, DISCARD CHANGES, RUN, SET UP WEB SERVICE, and PUBLISH TO GALLERY.

Azure Machine Learning Studio

國立 Experiments - Microsoft Azure

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...

應用程式 Google Google Analytics I... Facebook 0514前_DHL_Exp... Research 統計觀念與軟體操... 青年資源讀 - 2019... 電商 研究所 YouTube 至 mp3...

Microsoft Azure Machine Learning Studio Vicky Chen-Free-Workspace

Iris_example_0508 In draft Draft saved at 上午10:29:44

Search experiment items

- Machine Learning
 - Evaluate
 - Initialize Model
 - Anomaly Detection
 - Classification
 - Clustering
 - Regression
 - Score
 - Apply Transformation
 - Assign Data to Clusters
 - Score Matchbox Recommender
 - Score Model
 - Train
 - OpenCV Library Modules
 - Python Language Modules
 - R Language Modules
 - Statistical Functions
 - Text Analytics
 - Time Series

iris.csv

Select Columns in Dataset

Multiclass Decision Forest

Split Data

Train Model

Score Model

Experiment Properties

START TIME	5/8/2019 10:...
END TIME	5/8/2019 10:...
STATUS CODE	InDraft
STATUS DETAILS	None

Prior Run

Summary

Enter a few sentences describing your experiment (up to 140 characters).

Description

Enter the detailed description for your experiment.

Quick Help

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

Azure Machine Learning Studio

國立 Experiments - Microsoft Azure

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...

應用程式 Google Google Analytics I... Facebook 0514前__DHL_Exp... Research 統計觀念與軟體操... 青年資源讀 - 2019... 電商 研究所 YouTube 至 mp3...

Microsoft Azure Machine Learning Studio Vicky Chen-Free-Workspace

Iris_example_0508 In draft

Search experiment items

- Scale and Reduce
- Feature Selection
- Machine Learning
 - Evaluate
 - Cross Validate Model
 - Evaluate Model
 - Evaluate Recommender
 - Initialize Model
 - Score
 - Train
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
- Statistical Functions
- Text Analytics
- Time Series
- Web Service
- Deprecated

iris.csv

Select Columns in Dataset

Multiclass Decision Forest

Split Data

Train Model

Score Model

Evaluate Model

Draft saved at 上午10:31:03

Experiment Properties

START TIME	5/8/2019 10:...
END TIME	5/8/2019 10:...
STATUS CODE	InDraft
STATUS DETAILS	None

Prior Run

Summary

Enter a few sentences describing your experiment (up to 140 characters).

Description

Enter the detailed description for your experiment.

Quick Help

+ NEW

RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

8.Evalute Model

國立 Experiments - Microsoft Azure UCI Machine Learning Repository

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...

應用程式 Google Google Analytics I... Facebook 0514前__DHL_Exp... Research 統計觀念與軟體操... 青年資源讀 - 2019... 電商 研究所 YouTube 至 mp3...

Microsoft Azure Machine Learning Studio Vicky Chen-Free-Workspace

Iris_example_0508

Finished running ✓

Properties **Project**

Evaluate Model

START TIME	5/8/2019 10:...
END TIME	5/8/2019 10:...
ELAPSED TIME	0:00:03.211
STATUS CODE	Finished
STATUS DETAILS	None

[View output log](#)

Search experiment items

- Saved Datasets
- Data Format Conversions
- Data Input and Output
- Data Transformation**
 - Filter
 - Learning with Counts
 - Manipulation
 - Sample and Split**
 - Partition and Sample
 - Split Data
 - Scale and Reduce
- Feature Selection
- Machine Learning**
 - Evaluate**
 - Cross Validate Model
 - Evaluate Model
 - Evaluate Recommender
 - Initialize Model

iris.csv

Select Columns in Dataset ✓

Split Data ✓

Multiclass Decision Forest ✓

Train Model ✓

Score Model ✓

Evaluate Model

Download

Save as Dataset

Save as Trained Model

Save as Transform ✓

Visualize

Generate Data Access Code...

Open in a new Notebook

Evaluates a scored classification or regression model with standard metrics ([more help...](#))

Delete

Copy

Cut

Paste

① Evaluation results

View Log

Edit Comment

Run selected

Help

NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

RUN

SET UP WEB SERVICE

國立

Experiments - Microsoft Azure

UCI Machine Learning Repository

+

← → ↺

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...

☆

🟢

🔴

🟡

🟠

🟢

🔴

🟡

🟠

🟢

🔴

🟡

🟠

應用程式

Google

Google Analytics I...

Facebook

0514前_DHL_Exp...

Research

統計觀念與軟體操...

青年資源讀 - 2019...

電商

研究所

YouTube 至 mp3...

»

Microsoft Azure Machine Learning Studio

Vicky Chen-Free-Workspace

?

👤

😊

👤

←

Iris_example_0508

Finished running ✓

Properties

Project

➤

Search ex

Iris_example_0508 > Evaluate Model > Evaluation results

✕

Metrics

Overall accuracy0.977778

Average accuracy0.985185

Micro-averaged precision0.977778

Macro-averaged precision0.97619

Micro-averaged recall0.977778

Macro-averaged recall0.974359

Confusion Matrix

Predicted Class

setosa

versicolor

virginica

Actual Class

setosa

versicolor

virginica

100.0%

92.3%

100.0%

7.7%

+

NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY

國立

Experiments - Microsoft Azure

UCI Machine Learning Repository

+

← → ↺

https://studio.azureml.net/Home/ViewWorkspaceCached/49869f3cb79c4c5ba997f5f6b5233dc0#Workspaces/Experiments/Experiment/49869f3cb79...

☆

🛡️

🚀

🌈

🔥

📁

🌐

🔦

應用程式

Google

Google Analytics I...

Facebook

0514前_DHL_Exp...

Research

統計觀念與軟體操...

青年資源讀 - 2019...

電商

研究所

YouTube 至 mp3...

»

Microsoft Azure Machine Learning Studio

Vicky Chen-Free-Workspace

?

👤

😊

👤

Iris_example_0508

Finished running ✓

Properties

Project

Search ex

Iris_example_0508 > Evaluate Model > Evaluation results

✕

Macro-averaged recall

0.974359

Confusion Matrix

Predicted Class

setosa versicolor virginica

Actual Class

setosa

versicolor

virginica

setosa	100.0%		
versicolor	92.3%	7.7%	
virginica			100.0%

+ NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY

機器學習平台範例介紹

DataRobot

人工智慧醫療應用資料集選輯

中文名稱	英文名稱	資料型態	任務簡述	
心臟病資料集	UCI Cleveland Heart Disease Dataset	單一數字表格文件	分類(classification)	https://www.kaggle.com/datasets/cherngs/heart-disease-cleveland-uci
中風資料集	Stroke Prediction Dataset	單一數字文字表格文件	分類(classification)/異常檢測(anomaly detection)	https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset
醫療保險詐欺資料集	Healthcare Provider Fraud Analysis	多個數字文字表格文件	分類(classification)/異常檢測(anomaly detection)	https://www.kaggle.com/datasets/rohitrox/healthcare-provider-fraud-detection-analysis
醫療費用預測資料集	Medical Cost Personal Dataset	單一數字文字表格文件	回歸(regression)	https://www.kaggle.com/datasets/mirichoi0218/insurance
藥品使用反饋資料集	UCI Drug Review	單一數字文字表格文件	自然語言處理(natural language processing)/情緒分析(sentiment analysis)	https://www.kaggle.com/datasets/jessicali9530/kuc-hackathon-winter-2018
新冠肺炎全球趨勢資料集	COVID-19 GobaI Trend	多個數字文字時間表格文件	時間序列分析及預測(time series analysis/forecast)	https://www.kaggle.com/datasets/sudalairajkumar/novel-corona-virus-2019-dataset
胸部 X 光肺炎資料集	Chest X-Ray Images (Pneumonia)	圖像文件(.jpeg)	圖像分類(image classification)	https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia
新冠肺炎電腦斷層掃描資料集	COVID-19 CT Scans	圖像文件(.nii)	圖像分割(image segmentation)	https://www.kaggle.com/datasets/andrewmvd/covid19-ct-scans
心音檢測資料集	Heartbeat Sound	音頻文件(.wav)	音頻分類(audio classification)/音頻分割(audio segmentation)	https://www.kaggle.com/datasets/kinguistics/heartbeat-sounds
藥物交互作用資料集	Drug-Drug Interaction Network	單一文字表格文件	連接預測(link prediction)	http://snap.stanford.edu/biodata/datasets/10001/10001-ChCh-Miner.html
乳癌診斷資料集	Breast Cancer Wisconsin Diagnostic	單一數字表格文件	分類(classification)	https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data

The screenshot shows the Kaggle interface for the 'Breast Cancer Wisconsin (Diagnostic) Data Set'. The left sidebar contains navigation links like Home, Competitions, Datasets, Code, Discussions, Learn, and More. The main header includes a search bar and user profile. The dataset title is prominently displayed, followed by its description: 'Predict whether the cancer is benign or malignant'. Below the title are tabs for Data, Code (2298), Discussion (49), and Metadata. A large purple-tinted image of breast tissue cells is shown on the right. The 'About Dataset' section provides details about the features, which are computed from a digitized image of a fine needle aspirate (FNA). It also mentions the source as the UCI Machine Learning Repository and provides a link to the archive.

The screenshot shows the Kaggle interface for the 'Breast Cancer Wisconsin (Diagnostic) Data Set'. The left sidebar contains navigation links: Home, Competitions, Datasets, Code, Discussions, Learn, and More. Below these is a 'Your Work' section and a 'RECENTLY VIEWED' list. The main content area displays the dataset title, a description ('Predict whether the cancer is benign or malignant'), and a download button. The 'About Dataset' section provides details on the data source (digitized FNA images), availability (UW CS ftp server and UCI Machine Learning Repository), and attribute information (ID number, Diagnosis, and ten real-valued features: radius, texture, perimeter, etc.). The right sidebar shows the dataset's Usability (8.53), License (CC BY-NC-SA 4.0), and Expected update frequency (Not specified).

十大特徵值

1. 半徑(中心到邊緣上的點的平均距離)
2. 紋理(灰階值的標準差)
3. 周長
4. 面積
5. 平滑度(半徑長度的局部變化)
6. 緊密度(周長平方/面積 - 1.0)
7. 凹面(輪廓凹面部分的嚴重度)
8. 凹點(輪廓凹面的數量)
9. 對稱性
10. 分形維度(海岸線近似-1)