



Winner winner, chicken dinner

誰才是最後贏家——
以資料探勘預測誰能吃雞

指導教授：李鍾斌

學生：劉律奇、曾士育、阮之群、邱邦旭

蔡宗宏、李奇學、游宏文



目錄

1 | 前言

2 | 文獻探討

3 | 研究方法

4 | 實證分析

5 | 結論

前言



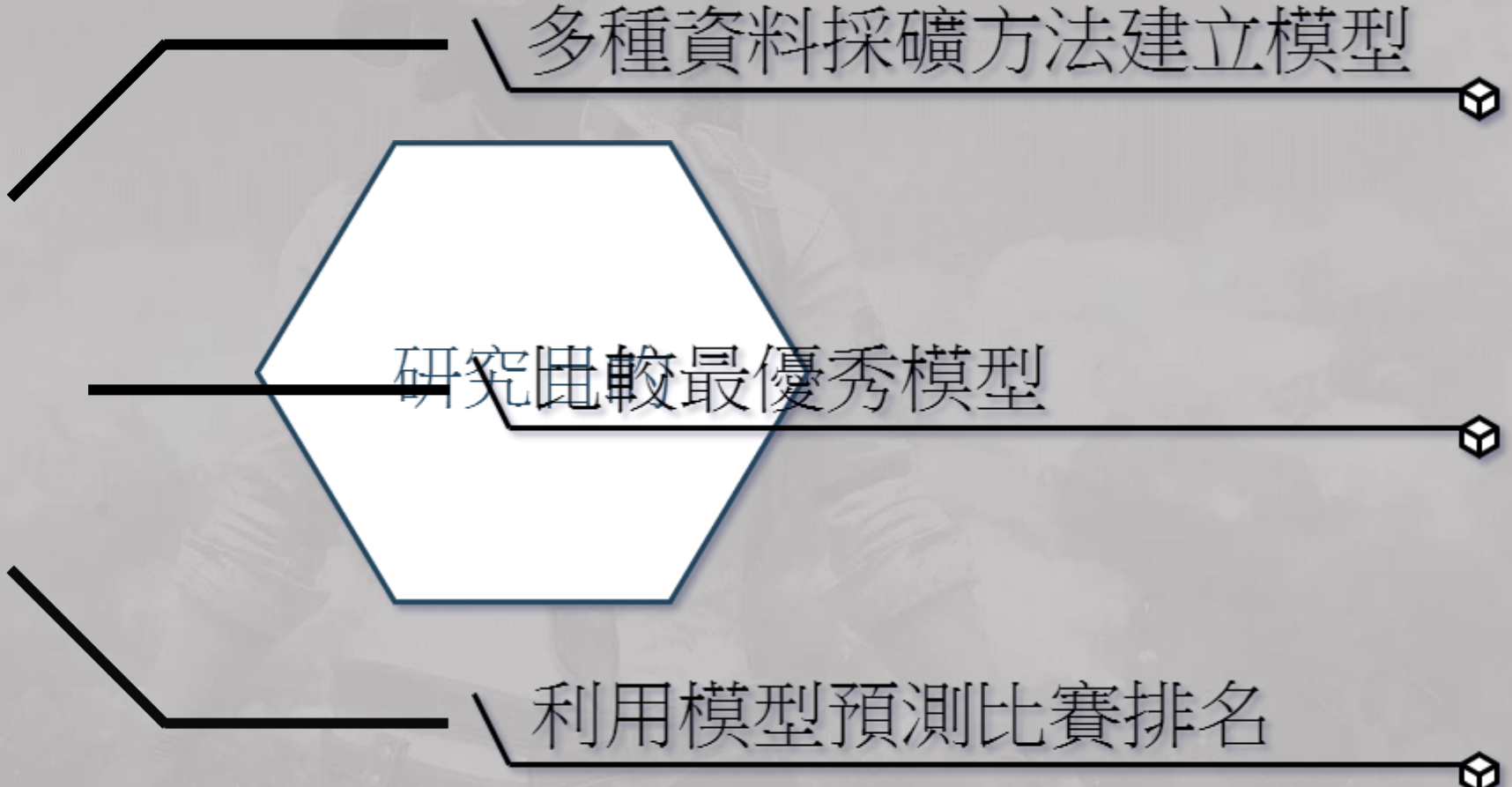
研究背景與動機

- 網路遊戲普及

WE ARE GAMER

選擇「絕地求生」的原因





多種資料採礦方法建立模型

研究比較最優秀模型

利用模型預測比賽排名

研究架構

1

界定研究主題與目的

2

探討相關文獻

3

擬定研究方法

4

資料分析與解釋

文獻探討



電競產業



遊戲



選手

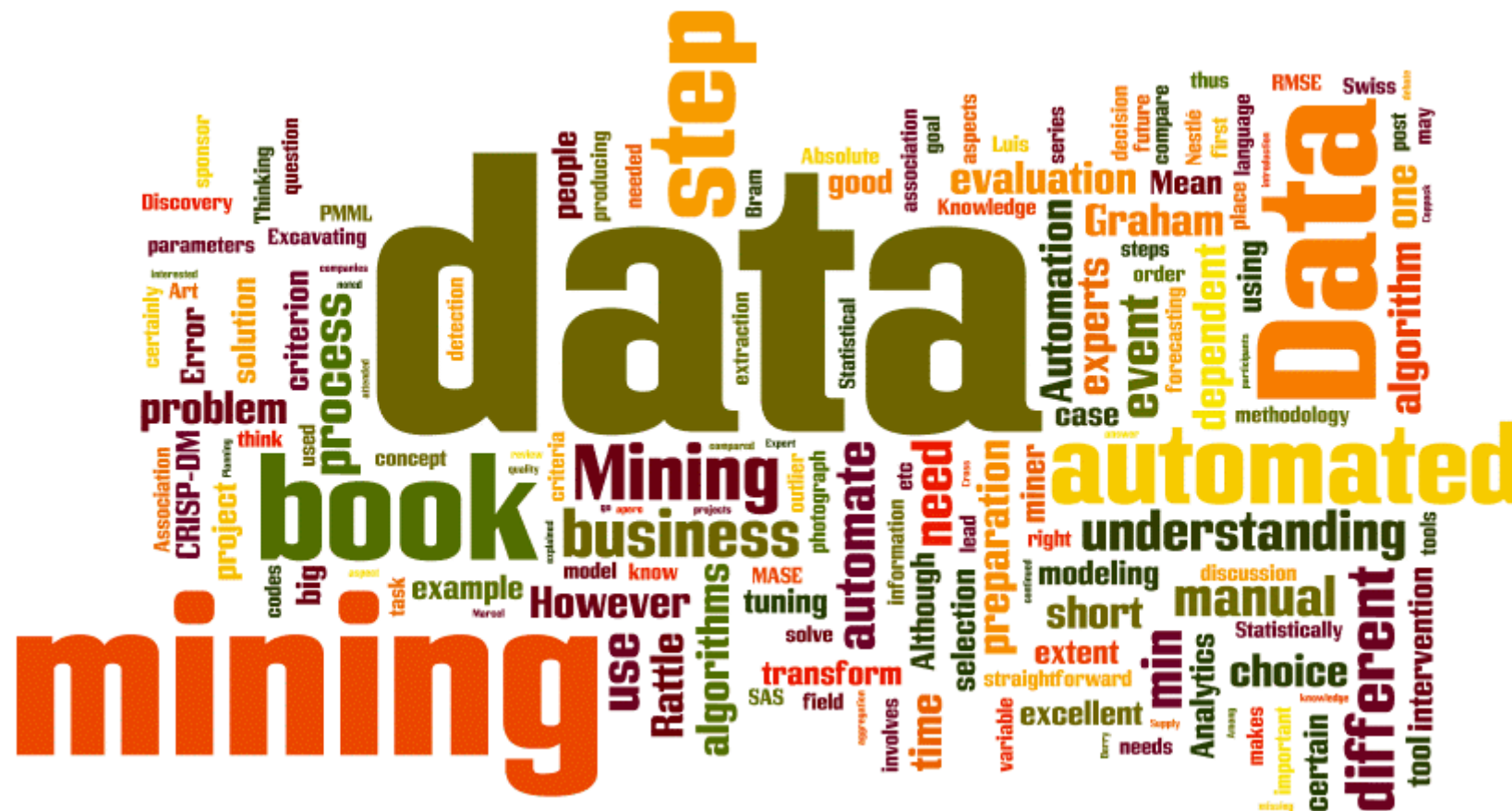


設備



賽事

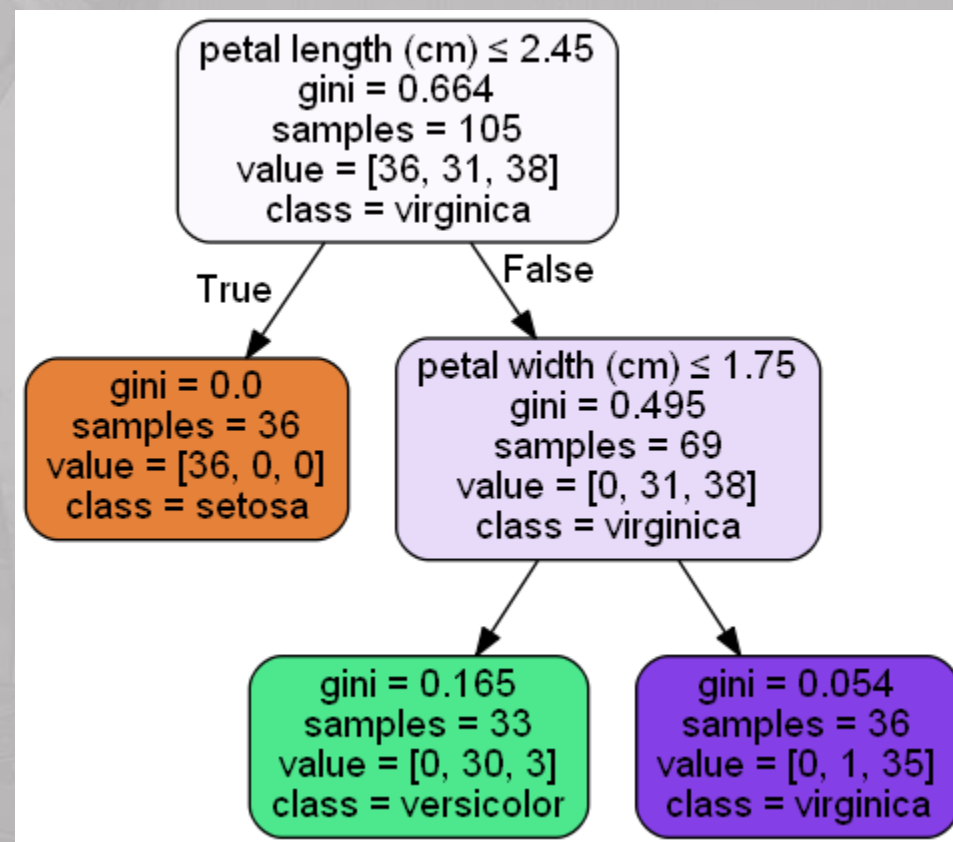
資料探勘



決策樹

決策樹優點：

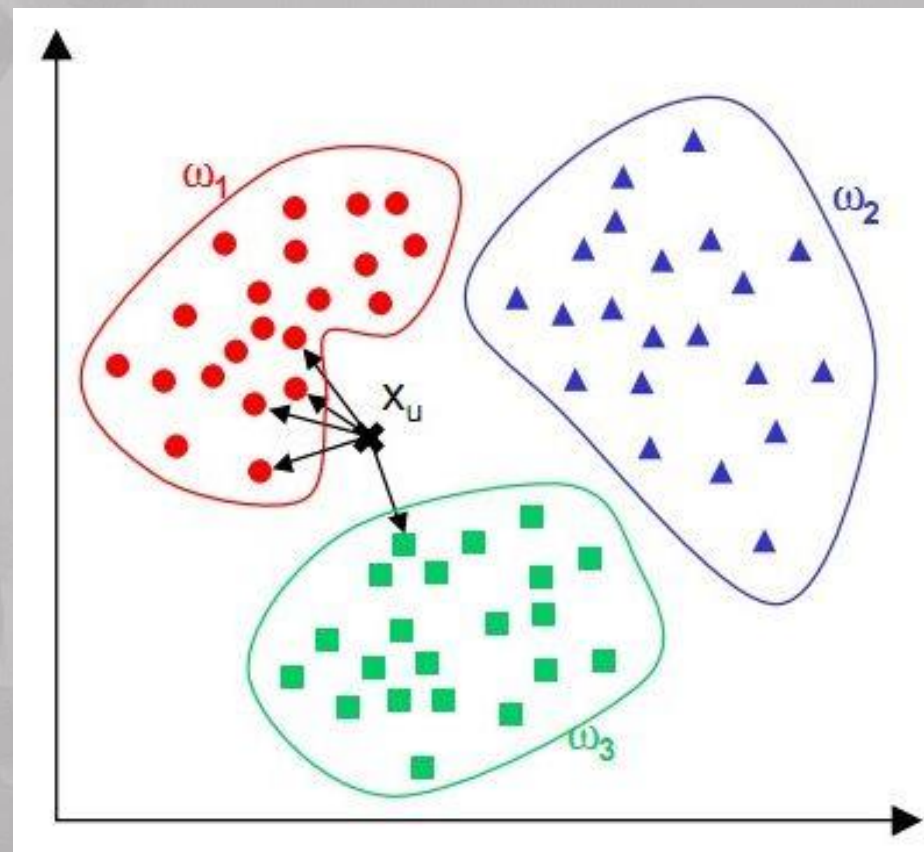
- 能用最佳變數切割樣本
- 概念簡單



KNN

KNN優點：

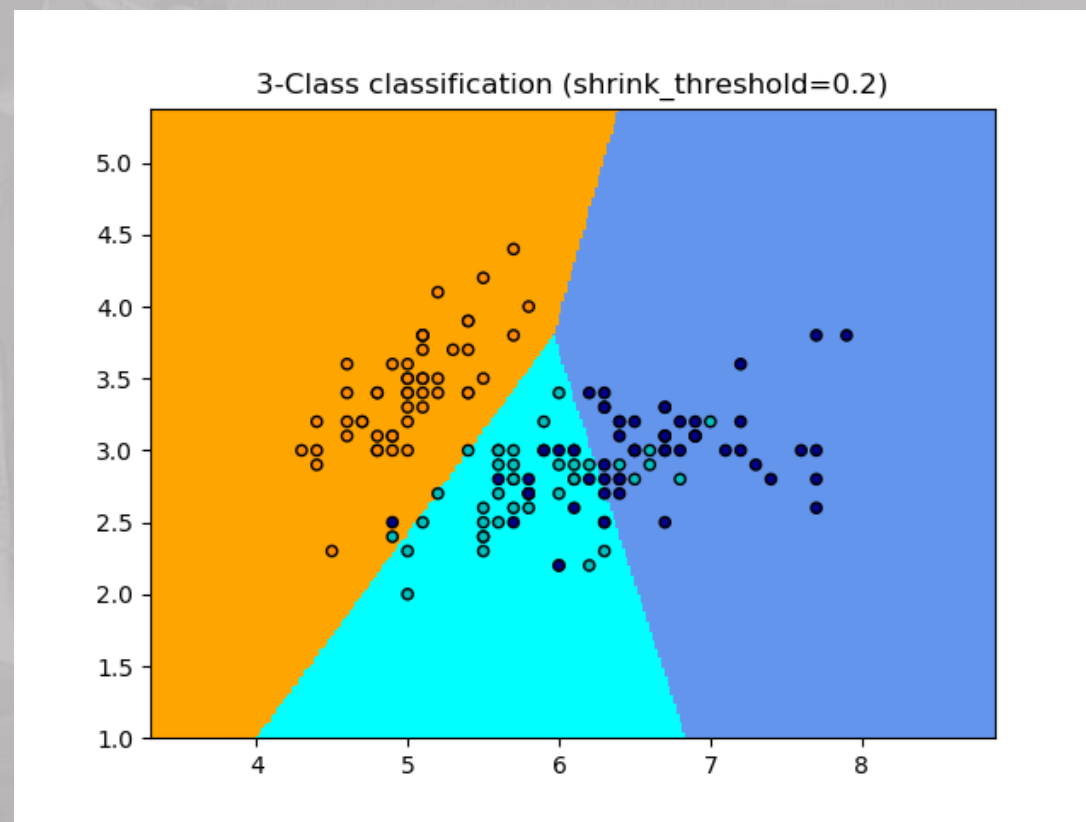
- 模型可以做分類也可以做回歸。
- 可以通過距離確定訓練樣本和測試樣本的相似性。



NCN

NCN優點：

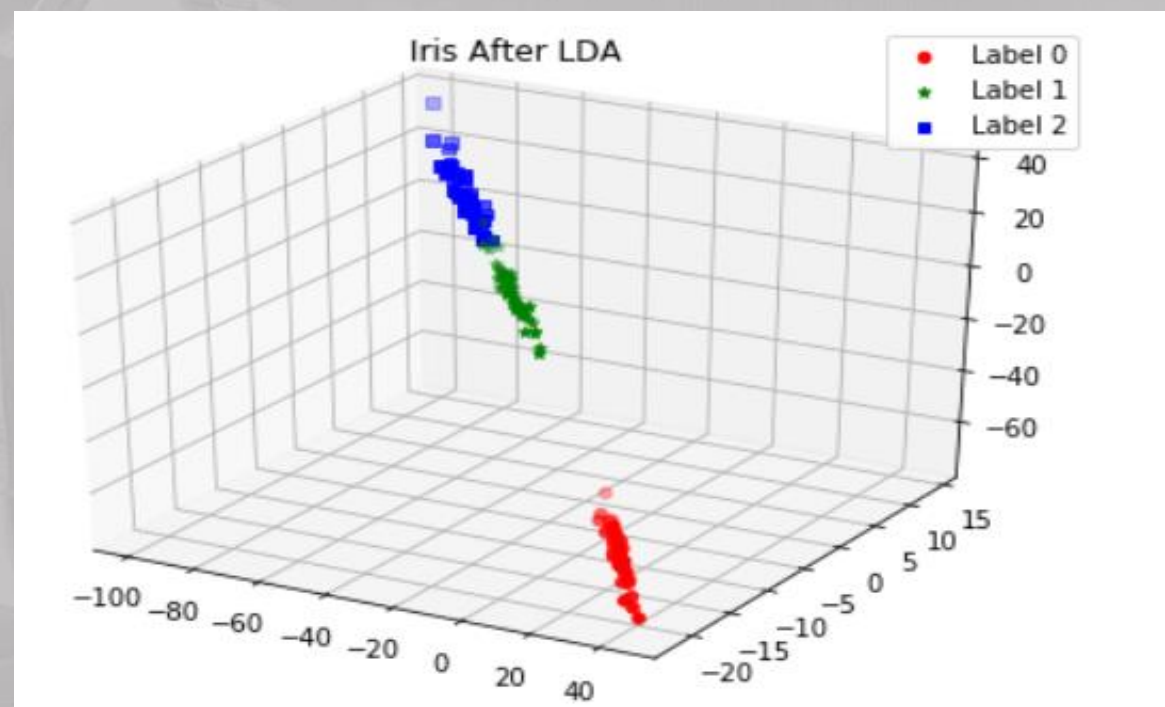
- 汲取了KNN的優點。
- 選取的k個近鄰點盡可能分佈在測試樣本點的周圍。



LDA

LDA優點：

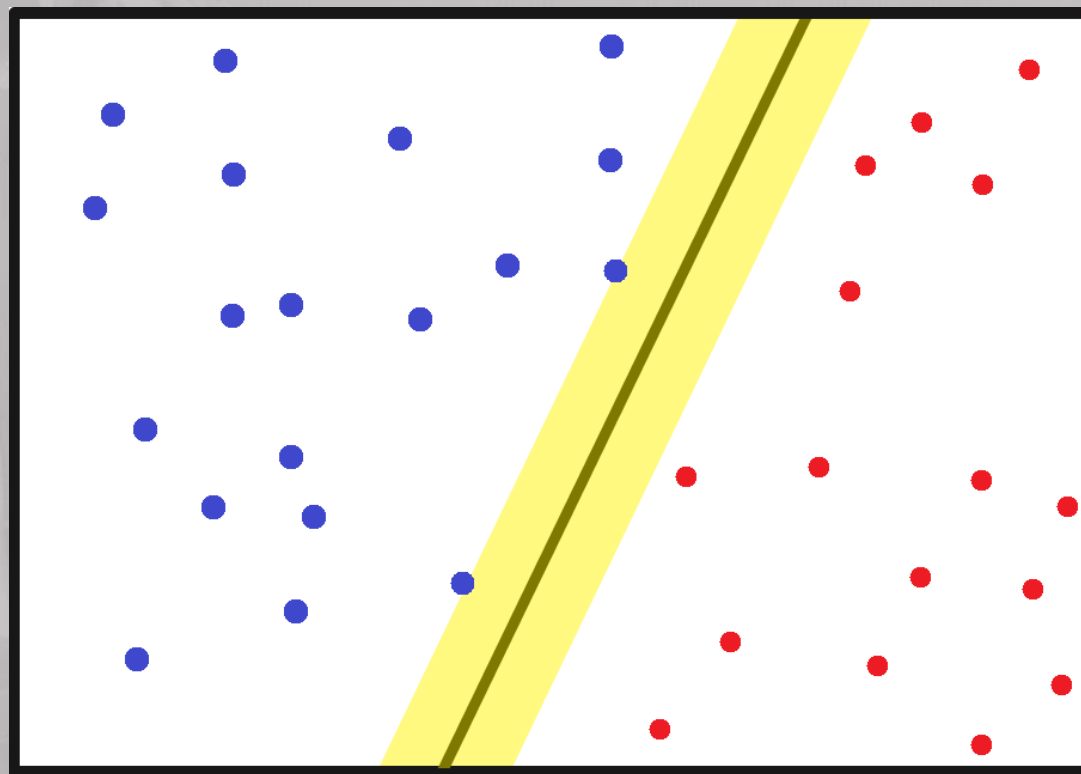
- 使同類樣本的投影點在同一條直線上。
- 3D立體視覺化模型便於分析。



SVM

SVM優點：

- 在解決小樣本、非線性、高維度模式識別問題有優勢。
- 能夠將兩個類別之間の間隙區隔開來以便讓分類器更精準地進行判斷與預測。



資料探勘方法統整

決策樹

KNN

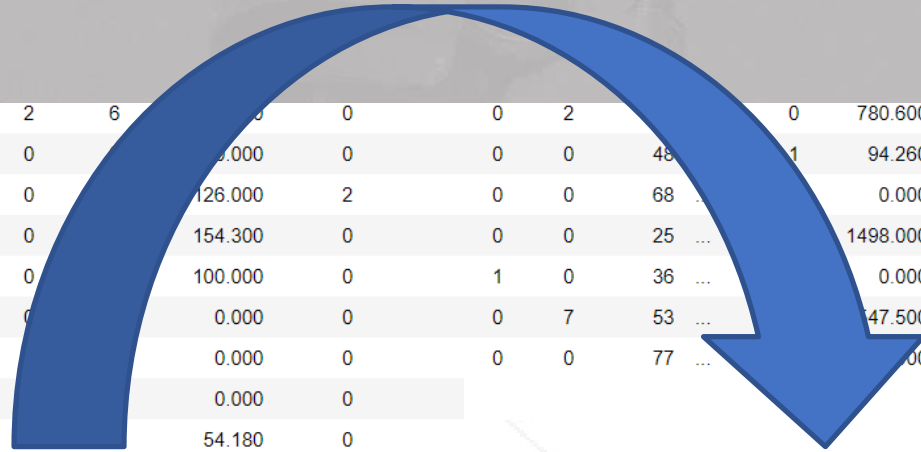
NCN

LDA

SVM

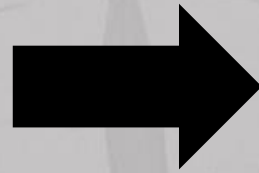
AWS EMR

9972	4a7e5ee394e22f	5cc1c847df2a73	004d71d7e987ef	2	6	0	0	2	0	780.6000
9973	808842ff8734f3	39aef5ef489a9	007b41f0dbfe07	0	0	0	0	48	1	94.2600
9974	12c7d532bc6245	5e1e527e9f2851	3b7ae7db7aaed5	0	126.000	2	0	0	68	0.0000
9975	0625260c720cf1	d49543fce96b1a	e26491499dcd7b	0	154.300	0	0	0	25	1498.0000
9976	db05cf72ff35ee	261fd07fbd7315	eb8be6252aa04f	0	100.000	0	1	0	36	0.0000
9977	8706859dfa58cc	4e3ff848c23f84	e3c029153563ab	0	0.000	0	0	7	53	47.5000
9978	1c12718796496e	6799e5ba6412b4	706eff109a1b3c	0	0.000	0	0	0	77	0.0000
9979	cbf26c309cd8a9	cf3089a936d9f1	a5fae1738f38c7	0	0.000	0	0	0	0	0.0000
9980	758ab6e0bd07e1	d67e7533fd30fe	09e20fa42dcd49	0	54.180	0	0	0	0	0.0000
9981	ebdf8f0eb48ccf	752c1174abe6e5	924783d6dae94a	0	4	0.000	0	0	0	0.0000
9982	6565fb6a8c4d97	5ac63af01785d5	63	0	0	0	0	0	0	0.0000
9983	9f323a6b9c1665	f5ce1694bf13a6	7a	0	0	0	0	0	0	0.0000
9984	64c673cc6d3fee	a7722328409e94	fe	0	0	0	0	0	0	0.0000
9985	f0541a53780b32	2dfab0d7de1b16	b5	0	0	0	0	0	0	0.0000
9986	7963a03596f1e5	8c29e672a1ae7c	de	0	0	0	0	0	0	0.0000
9987	9bcbcc34245dc5	f7b737e84a7e55	9	0	0	0	0	0	0	0.0000
9988	f37ca2626c88b6	1a7db2049d808d	6	0	0	0	0	0	0	0.0000
9989	cf1b2938ea2c0a	3ed9c7447d557a	30	0	0	0	0	0	0	0.0000
9990	19e2263f4fd8c4	c78	0	0	0	0	0	0	0	0.0000
9991	03f7ac2b9b4ce4	ed0	0	0	0	0	0	0	0	0.0000
9992	6f6a85fc91cdbc	609	0	0	0	0	0	0	0	0.0000
9993	60ecea8f4adf5	71c	0	0	0	0	0	0	0	0.0000
9994	e3da7eaaae176d	b5	0	0	0	0	0	0	0	0.0000
9995	c4402a9036ed91	2ca	0	0	0	0	0	0	0	0.0000





+



amazon
EMR

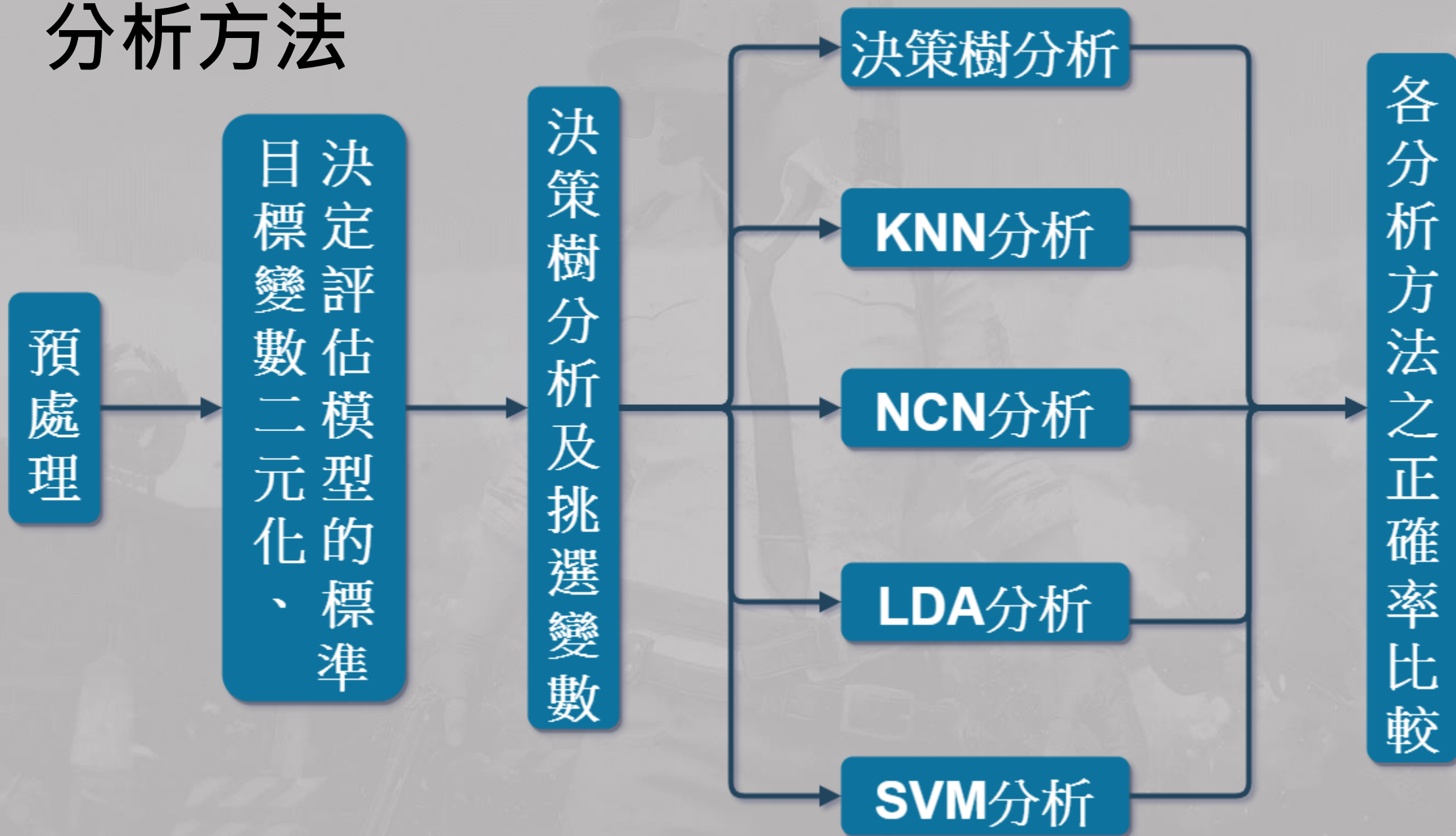
Amazon Web Services(AWS)



研究方法



分析方法



資料來源

1	DBNOs
2	assists
3	boosts
4	damageDealt
5	headshotKills
6	heals
7	ld
8	killPlace
9	killPoints
10	killStreaks
11	kills
12	longestKill
13	matchPoints
14	winPlacePerc
15	maxPlace
16	rideDistance
17	roadKills
18	swimDistance
19	teamKills
20	vehicleDestroys
21	walkDistance
22	weaponsAcquired
23	winPoints
24	groupId
25	numGroups
26	maxPlace
27	winPlacePerc

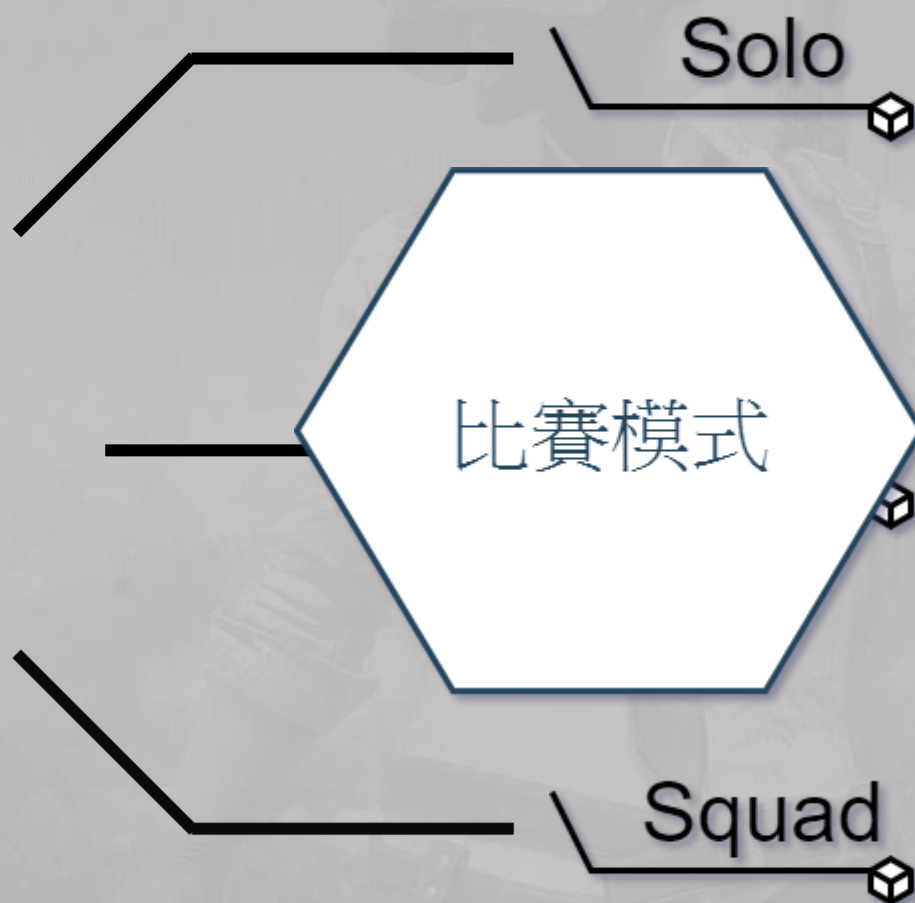
kaggle

winPlacePerc



預處理

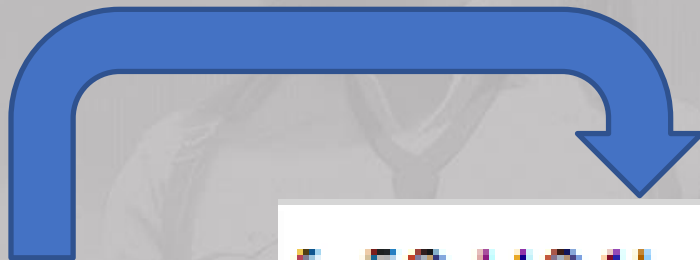
1	Id	groupId	matchId	assists	boosts	damageDealt	DBNOs	headshotKills	heals	killPlace	killPoints	kills	killStreaks	longestKill	matchDuration	matchType
2	7f96b2f878858a	4d4b580de459be	a10357fd1a4a91	0	0	0	0	0	0	60	1241	0	0	0	1306	squad
3	eef90569b9d03c	684d5656442f9e	aeb375fc57110c	0	0	91.47	0	0	0	57	0	0	0	0	1777	squad
4	1eaf90ac73de72	6a4a42c3245a74	110163d8bb94ae	1	0	68	0	0	0	47	0	0	0	0	1318	duo
5	4616d365dd2853	a930a9c79cd721	ff1f14ef412d7e	0	0	32.9	0	0	0	75	0	0	0	0	1436	squad
6	315c96c26c9aac	de04010b3458dd	6dc8ff871e21e6	0	0	100	0	0	0	45	0	1	1	58.53	1424	solo-f
7	ff9c12f326506	289a6836a88d27	bac52627a12114	0	0	100	1	1	0	44	0	1	1	18.44	1395	squad
8	95959be0e21ca3	2c485a1ad3d0f1	a8274e903927a2	0	0	0	0	0	0	96	1262	0	0	0	1316	squad
9	311b84c6ff4390	eaba5fcb7fc1ae	292611730ca862	0	0	8.538	0	0	0	0	1000	0	0	0	1967	solo-f
10	1a68204ccf9891	47cfbb04e1b1a2	df014fbee744	0	0	51.6	0	0	0	0	0	0	0	0	1375	squad
11	e5bb5a43587253	759bb6f7514fd2	3d2b0e93305b	0	0	27	0	0	0	0	0	0	0	0	1930	squad
12	2b574d43972813	c549efede67ad2	abdddb8320fc1	0	0	28.5	0	0	0	0	0	0	0	0	1811	squad
13	8de328a74658a9	f643df9df387	80170383d90003	0	0	137.9	1	0	0	0	0	0	0	0	1384	duo-f
14	ce4f6ac165705e	da24cdb593cc	535b51965a94	0	0	0	0	0	0	0	0	0	0	0	1774	squad
15	b7807186e3f679	3c08e4774749	2c30d831c52d	0	0	24.2	0	1	5	4	0	4	1	49.83	1886	solo-f
16	8e244ac61b6aab	d40d0c93573a1	94e1c148c85	0	0	22.8	1	0	2	1	0	1	1	37.91	1458	squad
17	12d8d4bd94312c	fe52d41ae68b	6fd9e7dd0c5	0	0	80.71	1	0	0	0	0	0	0	0	1313	duo-f
18	62f2f0917d84b2	f61b6982149f5	1d6cfe1f23b0	0	0	81.71	1	0	14	1	0	1	1	9.158	1882	squad
19	92022479b92ce7	2f2c33f54b09	07948123b9c0f	0	0	254.3	0	0	0	12	0	0	1	36	1371	solo-f
20	7bd224781f064b	6dde607d1511	733af30cc00099	0	0	0	0	0	0	0	0	0	0	0	1301	squad
21	71cbdbc3b263e5	7b61f74b51906c	2b0e099449ad7	0	1	0	0	0	1	134	0	0	0	0	1322	squad
22	02ace8c6e58461	a4bc548028f800	80f2b841e1	0	4	269.1	0	1	8	0	0	2	1	7.438	1890	duo-f
23	00341b1caa5420	d661a2d19e7ae9	f3956286eb39a5	0	0	158.7	1	0	0	0	0	0	0	0	1853	squad
24	9b2961d4d51f91	799d0a4d61dc3c	e833ca2282169d	0	1	192.3	1	2	3	15	1022	2	1	280.6	1775	duo-f
25	0b6f9dbf59c994	7a75c3e86934f6	8b0a78c005cea0	0	6	1011	6	2	2	2	1564	9	2	186.6	1857	duo
26	736eda9b9c20b3	d35e80e4e64dd4	62f8e726028662	0	3	327.6	4	1	1	3	1346	4	2	167.8	1454	squad
27	4c45dc732689ec	8e0a0ea95d3596	37f43ba55ec0a4	1	4	558.6	3	0	4	11	1265	3	2	237.5	1841	duo-f
28	91f5da9c3628eb	090fd12f3ca8a8	e8b8cf5d9231d3	0	0	44.28	0	0	0	78	0	0	0	0	1879	squad
29	dbf611495bfda3	10cbb86844dee0	2cdae31ee18601	0	4	381.2	2	1	2	7	0	4	2	42.13	1896	duo-f
30	f9473c4f1cfdc4	8483976f3ba230	6057f846f3ed12	0	6	345.6	2	1	1	6	0	4	1	105.2	1339	squad
31	ac5b57ff39979c	857cc55b2b6001	e019e04dee4f19	0	0	0	0	0	0	87	0	0	0	0	1530	duo
32	3a4e362d28f411	b4fbf11bf33ac3	cf3d655571b26b	0	0	36.73	0	0	0	62	0	0	0	0	1901	duo
33	4d16222292e9c8	539c5ba5a4ca96	9d090204befa18	0	0	0	0	0	0	80	0	0	0	0	1371	squad
34	e050d1f179d289	9e9fc09bcdab3f	bf3065e01dd3f9	0	3	0	0	0	1	61	0	0	0	0	1913	duo
35	4b604176d374e0	0be145f90609b1	b2f1bb79ad4298	0	2	65.47	0	0	2	61	1095	0	0	0	1348	squad
36	3d1735f01b602	b6fca1e1d1db5	022d41e64c4234	0	1	120.0	1	0	1	21	0	1	1	11.26	1885	squad



預處理

groupid
4d4b580de459be
684d5656442f9e
6a4a42c3245a74
a930a9c79cd721
de04010b3458dd
289a6836a88d27
2c485a1ad3d0f1
eaba5fcb7fc1ae
47cfbb04e1b1a2
759bb6f7514fd2
c549efede67ad3
f643df9df3877c
da24cdb91969cc
3c08e461874749
d401877d857814
fe52d481bae68b
fe52d481bae68b
212c331348c4b9
6dde607d151819
7b61f74b51906c
a4bc548028f800
d661a2d19e7ae9
799d0a4d61dc3c
7a75c3e86934f6
d35e80e4e64dd4
8e0a0ea95d3596
090fd12f3ca8a8
10cbb86844dee0
8483976f3ba230
857cc55b2b6001
b4fbf11bf33ac3

一樣的！



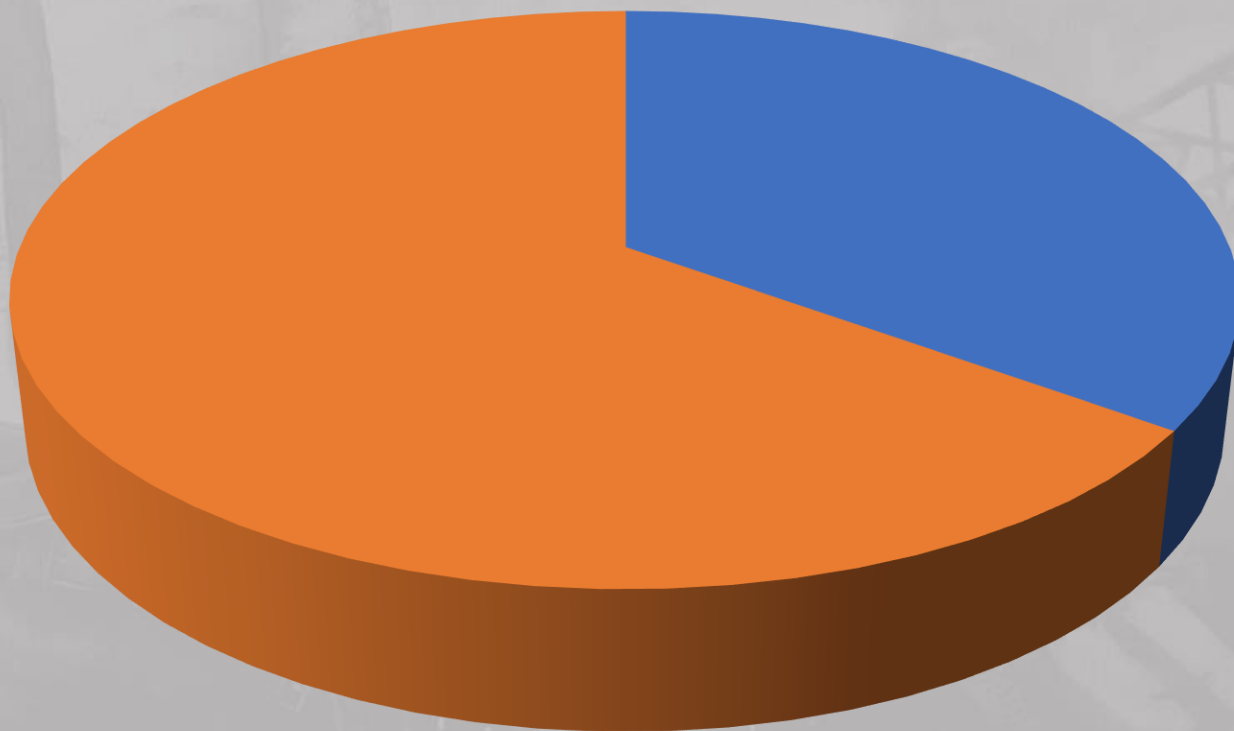
fe52d481bae68b

fe52d481bae68b

目標變數二元化

試驗三：前50%名次為win；後50%名次為lose

參考每周蘭賽積分獎勵機制，前35%為可獲得積分之玩家



評估模型

正確率(Accuracy)

- 精確率(Precision)
- 召回率(Recall)
- F1-score

混淆矩陣	預測正確	預測錯誤
真實正確	True Positive (TP)	False Negative(FN)
真實錯誤	False Positive (FP)	True Negative(TN)

$$F1 - score = \frac{2 * 精確率 * 召回率}{精確率 + 召回率}$$

研究限制及範圍

New!!!



實證分析

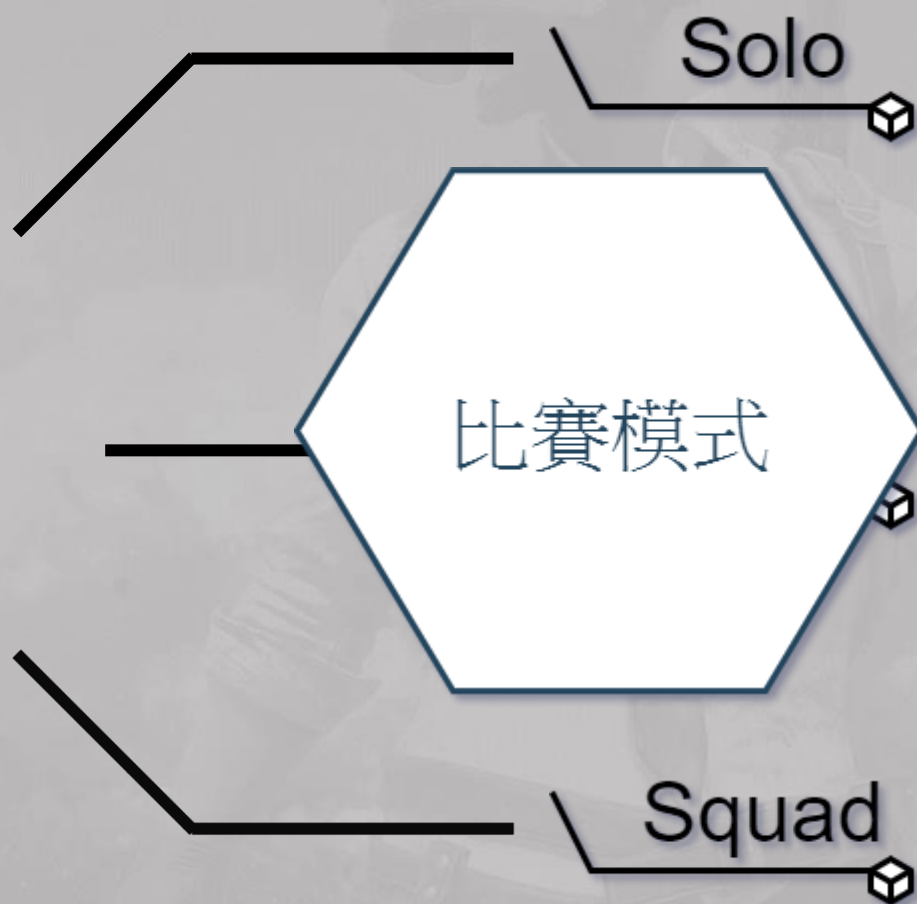


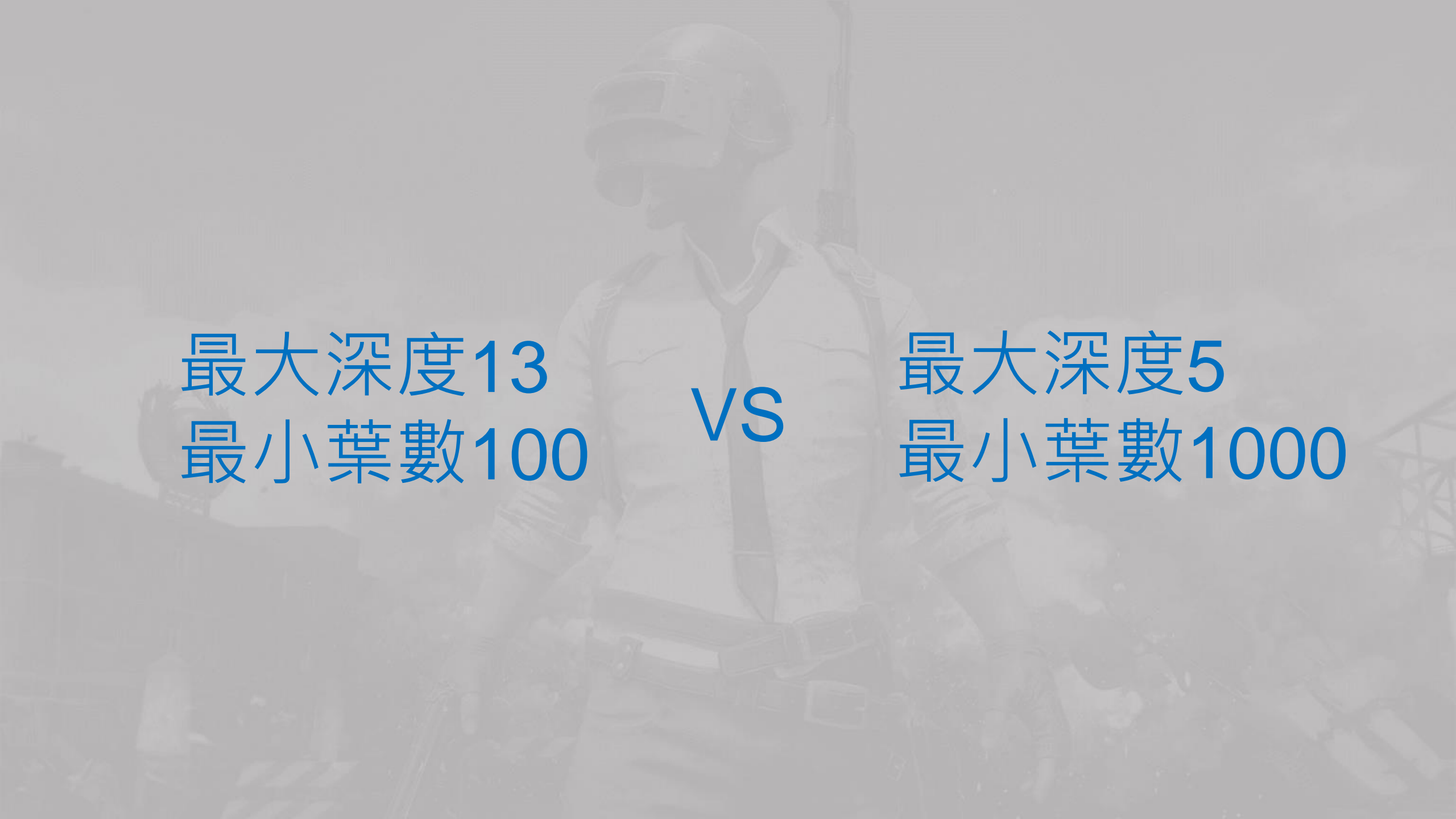


排名前**50%**
VS
排名後**50%**

排名前**10%**
VS
排名後**90%**

排名前**35%**
VS
排名後**65%**





最大深度13
最小葉數100

VS

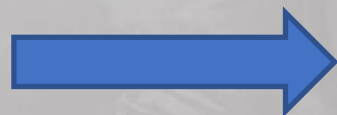
最大深度5
最小葉數1000



CART vs C4.5

挑選變數

1	DBNOs
2	assists
3	boosts
4	damageDealt
5	headshotKills
6	heals
7	Id
8	killPlace
9	killPoints
10	killStreaks
11	kills
12	longestKill
13	matchDuration
14	matchId
15	matchType
16	rankPoints
17	revives
18	rideDistance
19	roadKills
20	swimDistance
21	teamKills
22	vehicleDestroys
23	walkDistance
24	weaponsAcquired
25	winPoints
26	groupId
27	numGroups
28	maxPlace
29	winPlacePerc



walkDistance

Boost?

killPlace

matchDuration

目標歸一

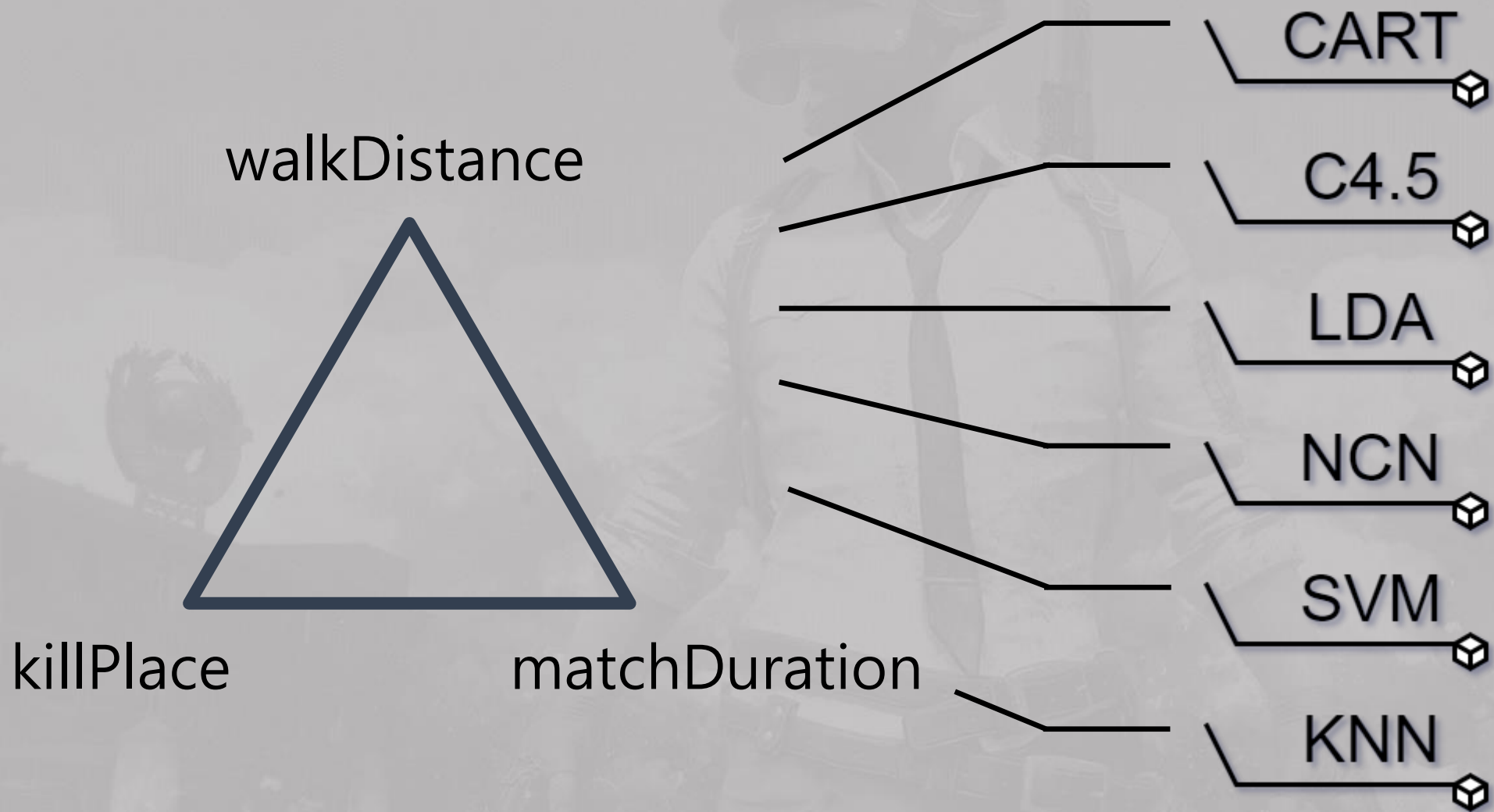
個人模式				
正確率	CART		C4.5	
目標分類	第一組	第二組	第一組	第二組
50 : 50	0.944	0.936	0.944	0.929
35 : 65	0.94	0.928	0.939	0.928
10 : 90	0.927	0.926	0.928	0.925



排名前**50%**
VS
排名後**50%**

排名前**10%**
VS
排名後**90%**

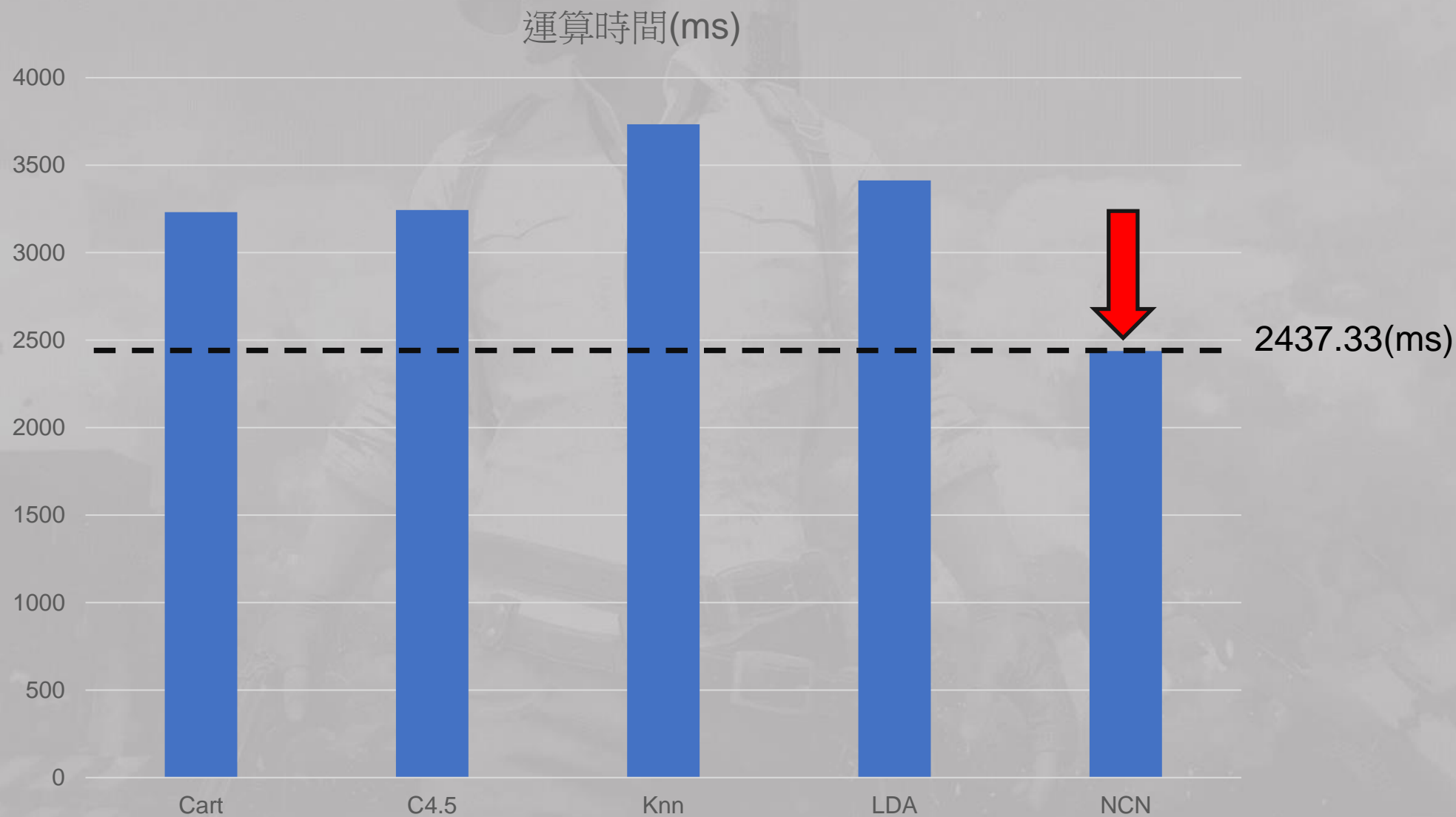
排名前**35%**
VS
排名後**65%**



分組正確率比較

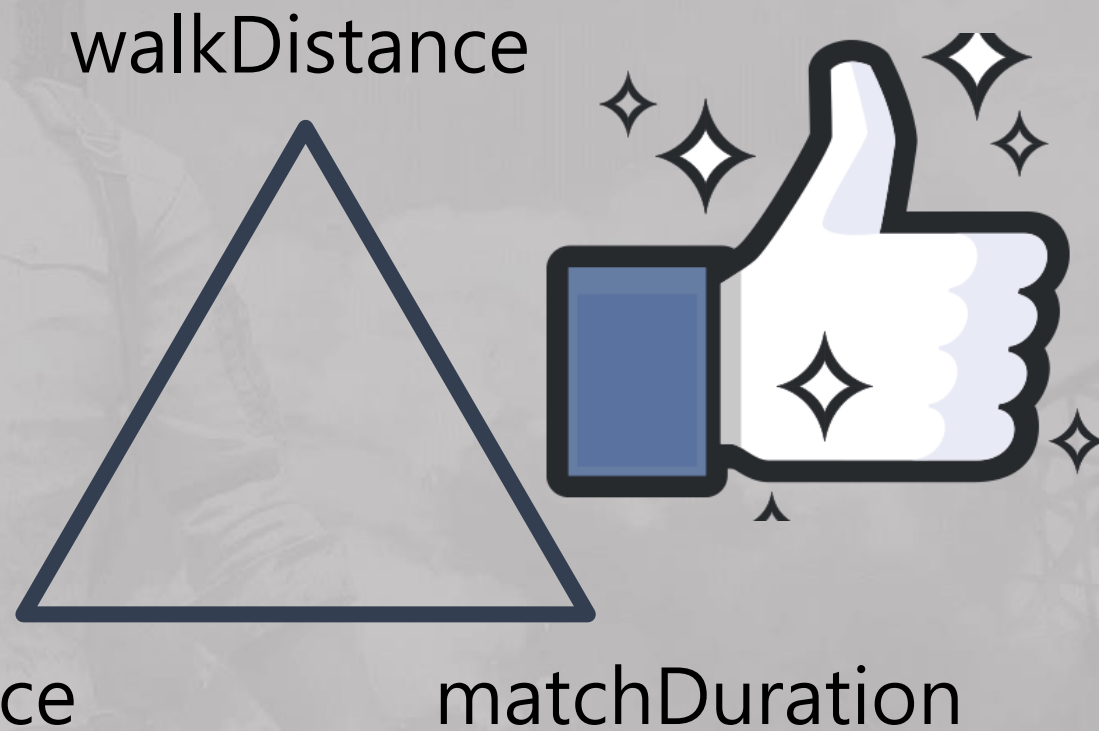
模型名稱	CART	C4.5	KNN	LDA	NCN	SVM
個人模式	0.940	0.939	0.922	0.914	0.905	0.658
雙人模式	0.930	0.930	0.923	0.921	0.906	0.643
四人模式	0.909	0.909	0.902	0.905	0.883	0.636

分組時間比較



得出重要變數

- 正確率高，解釋能力強
- 各個模型正確率相當



結論



個人模式

正確率	CART		C4.5	
目標分類	第一組	第二組	第一組	第二組
50 : 50	0.944	0.936	0.944	0.929
35 : 65	0.94	0.928	0.939	0.928
10 : 90	0.927	0.926	0.928	0.925

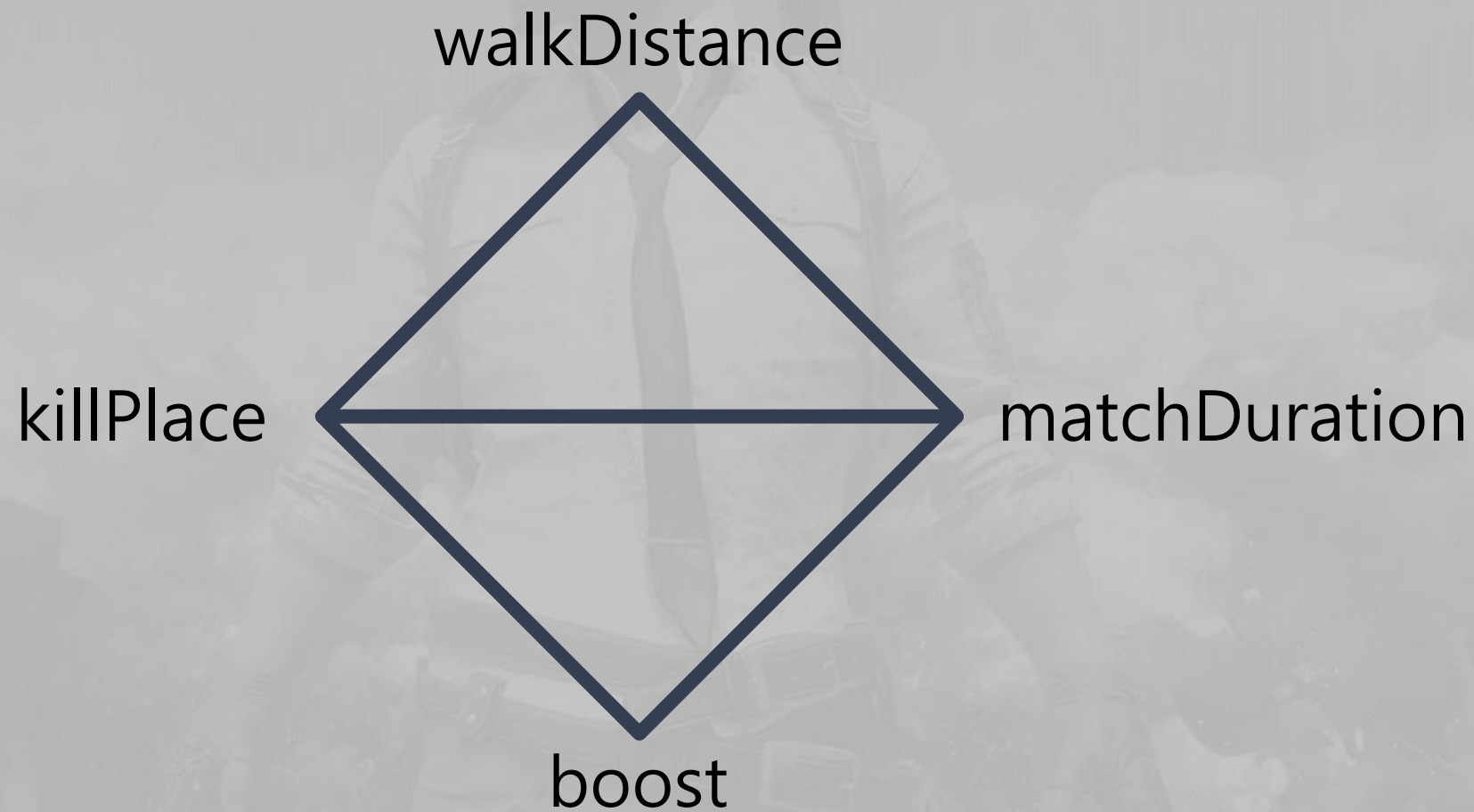


排名前**50%**
VS
排名後**50%**

排名前**10%**
VS
排名後**90%**

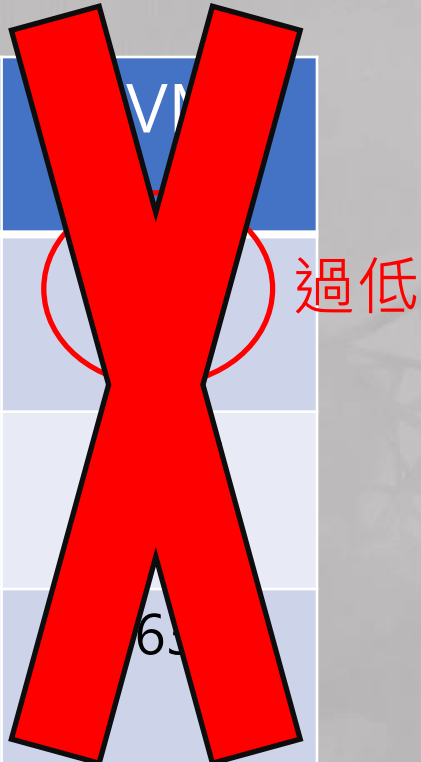
排名前**35%**
VS
排名後**65%**

重要變數

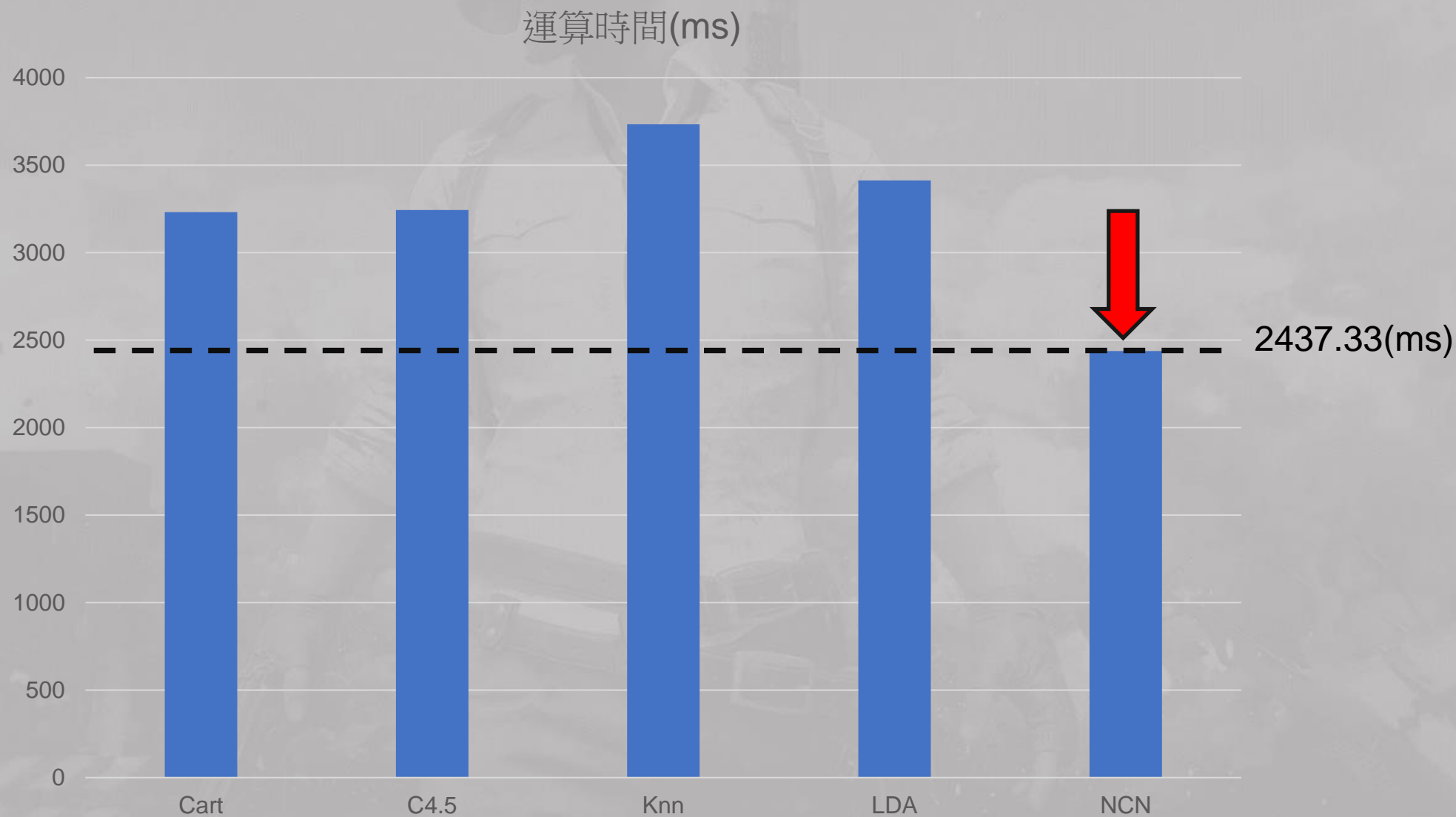


分組正確率比較

模型名稱	CART	C4.5	KNN	LDA	NCN	VM
個人模式	0.940	0.939	0.922	0.914	0.905	0.883
雙人模式	0.930	0.930	0.923	0.921	0.906	0.883
四人模式	0.909	0.909	0.902	0.905	0.883	0.863



分組時間比較



未來用途



or





報告結束

