



玩家流失預測



前提概要

需求: 每天凌晨預測流失用戶

資料需求: 21天

標的: 流失用戶

實際應用: 為流失用戶發放流失活動 (留存活動、贈送資源、超優惠禮包)

困難點:

1. 每天存活的人比較多、流失的人少，**資料不平衡**
2. 流失玩家跟低活躍玩家可能很難分清楚，**流失的瞬間較難預測**
3. 如果預測準確，**流失用戶收不到活動**，因為預測準確就不會上線

流失定義

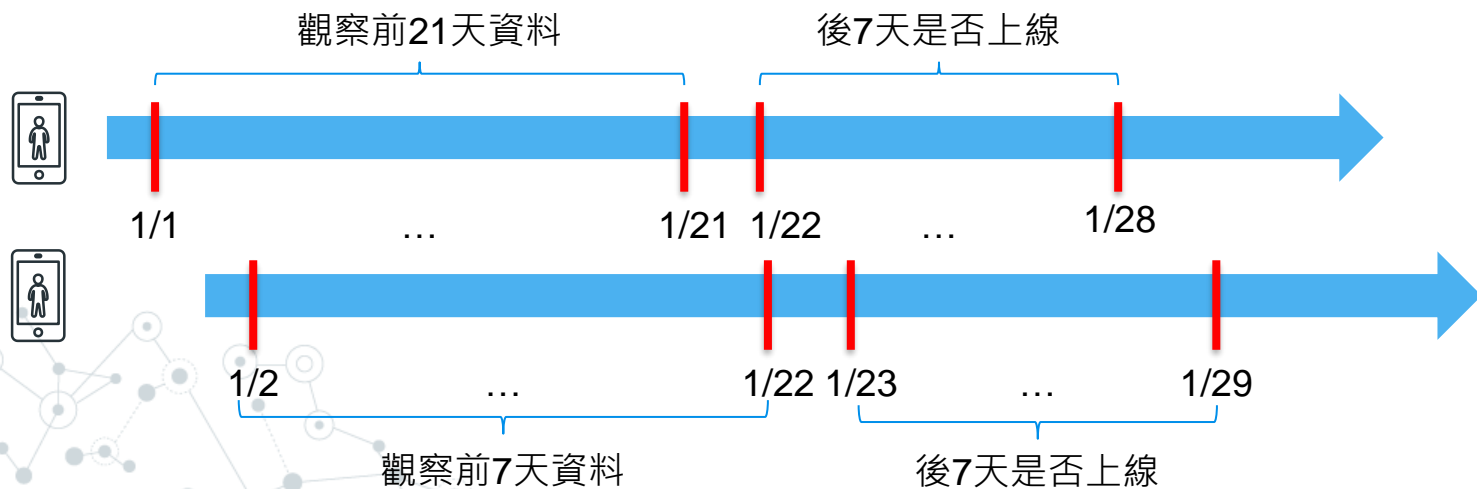
7天未登入

未登入天數	流失比例(%)
1	61.8
3	82.96
5	90.2
7	94.65
9	96.76
11	98.2
13	99.43

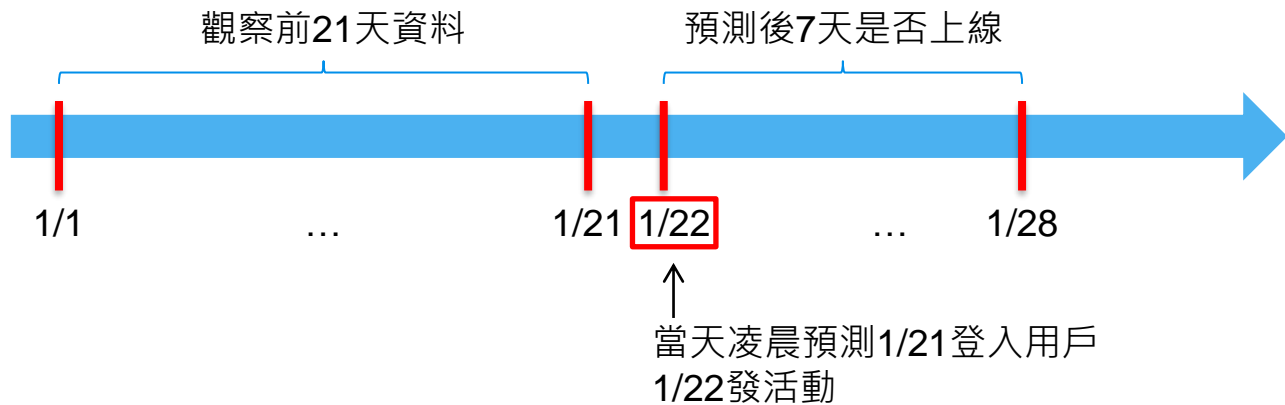
模型預測方式 & 成效評估

資料區間：2023/1/1 - 2023/4/30，4個月

- ◎ 針對註冊滿21天的用戶
- ◎ 預測前一天需上線
- ◎ 使用21天資料，預測後續7天內是否上線
- ◎ 每人每天資料都取



預測流程



建模過程

資料整理

不平衡資料處理

資料前處理

維度縮減 /
特徵挑選

挑選模型

超參數優化

時區修正
新增綜合特徵

UnderSampling

TomekLinks
NearMiss
ENN

OverSampling

SMOTE
BorderlineSMOTE
ADASYN

Normalization

L1 Normalizer
L2 Normalizer
Robust Scaler
Spline Transformer
MinMax Scaler

降維

PCA
t-SNE

特徵選取

RFECV
SF Selector
LGBM importance –
20/30/40/50
XGboost importance –
20/30/40/50

Auto ML(flaml)
LGBM
XGboost

Optuna

Under Sampling

如果數據量很大且冗餘：可以嘗試 欠採樣技術

隨機從多數類樣本中移除部分數據，使多數類和少數類數量接近。

缺點：可能丟失多數類中的有用信息，導致性能下降。

因為數據有包含每人的每天樣本，所有會有較多留存數據

流失數據: 15,000 筆

留存數據: 110,000 筆

數據集不同月份都保持這比例

最後決定留存數據隨機刪除50%

讓流失數據：留存數據 保持 1:3~4左右的比例

模型特徵

已遊玩天數
初始平台



每日數據 (21天)

上線時長
每日任務次數
第一次上線時間
等級
訓練階級

排位賽遊玩時長
排位賽遊玩次數
排位賽間隔時長

普通賽遊玩時長
普通賽遊玩次數
普通賽間隔時長



每週數據總和 (3週)

鑽石消耗
登入時長
排位賽間隔時長
普通賽間隔時長



每週次數加總 (3週)

升級技能
IAA觀看次數
登入天數

共

重要特徵

features	特徵	importance
create_days	創角幾日	1391
online_time_21	預測前一天上線時長	1315
first_hour_21	預測前一天首登時間	1142
online_time_week_3	最後一週上線時長	1052
game_ranked_next_time_20	第20天場次間隔時長	863
online_time_week_1	第一週上線時長	816
game_ranked_next_time_week_3	第3週場次間隔時長	769
game_ranked_next_time_9	第9天場次間隔時長	747
game_ranked_next_time_21	第21天場次間隔時長	724
quest_21	最後一天任務次數	713
online_time_week_2	第2週上線時長	637
game_ranked_next_time_week_2	第2週場次間隔時長	627

⋮

LGBM importance – 20/30/40/50

XGboost importance – 20/30/40/50

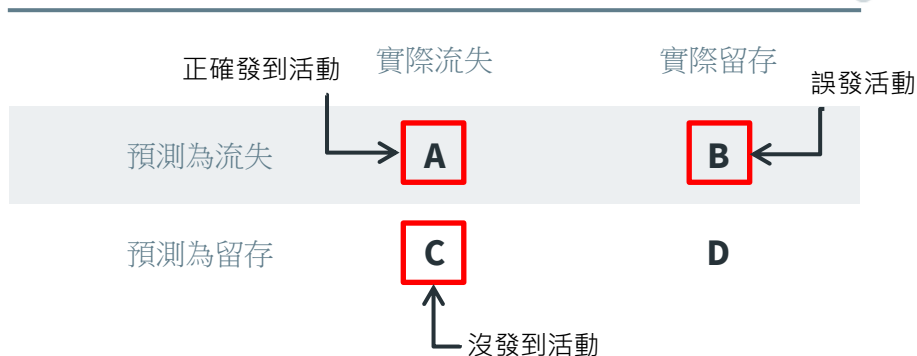
經過多次嘗試LGBM 前30重要特徵表現最好

模型表現

◎ **Accuracy** : $(A+D) / (A+B+C+D)$

◎ **Recall** : $(A) / (A+C)$

◎ **Precision** : $(A) / (A+B)$



Accuracy : 準確率

Recall : 流失用戶中，有多少比例能被模型正確預測出來

Recall越高，模型越能抓到流失用戶

Precision : 模型預測為流失的人中，有多少比例為真的流失用戶

Precision越高，模型誤發獎勵的機率越低

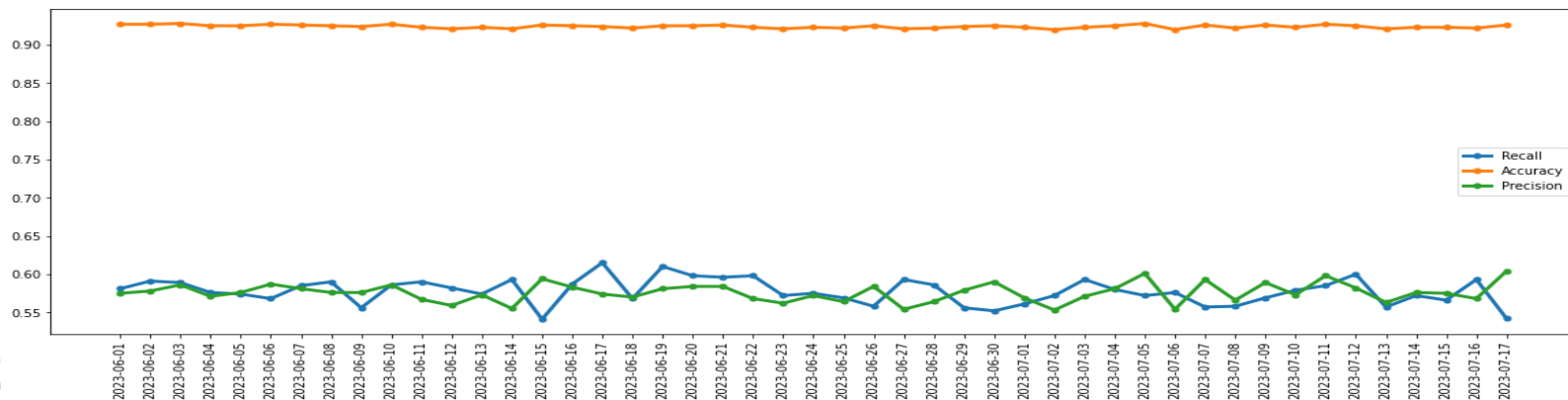
模型表現 – testing data

測試時間：2023/06/01 – 2023/07/17

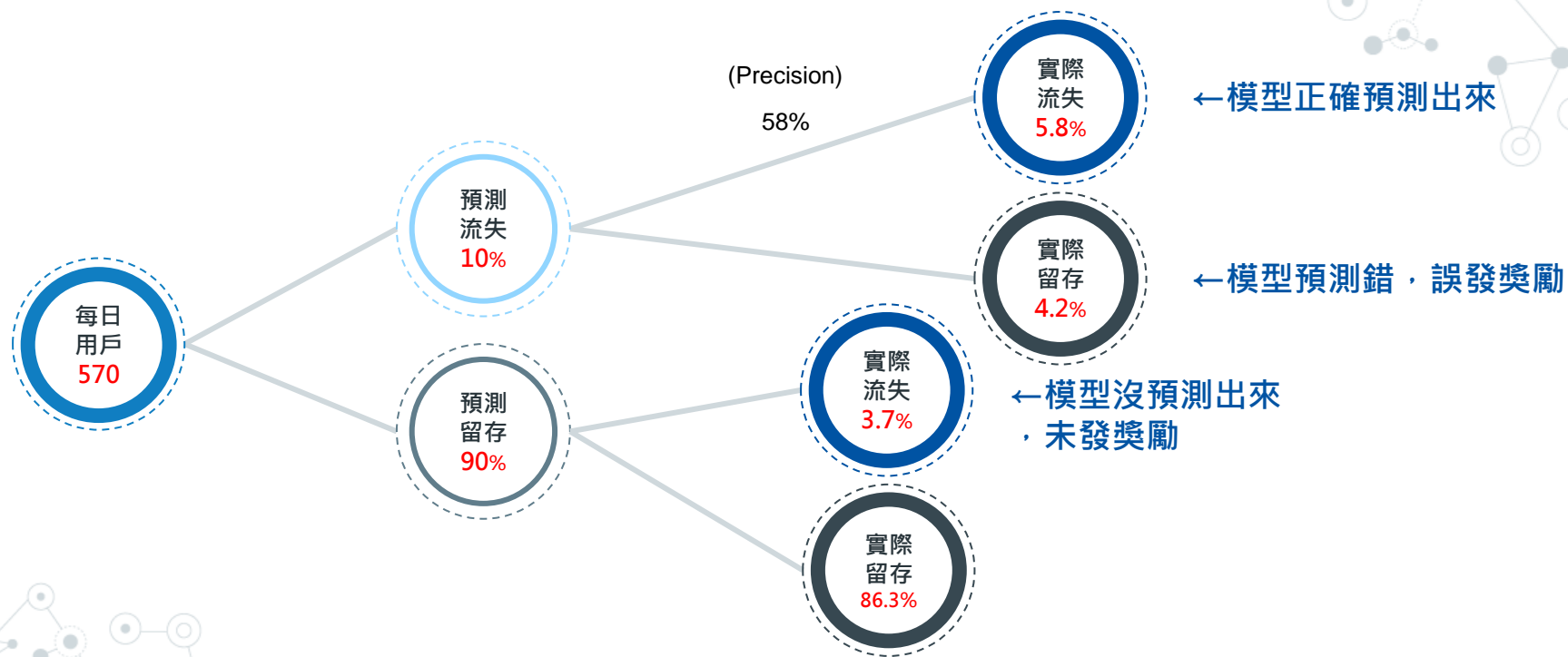
Accuracy : 92%

Recall : 58%

Precision : 58%



模型表現 – 實際預測



模型表現 – 實際預測

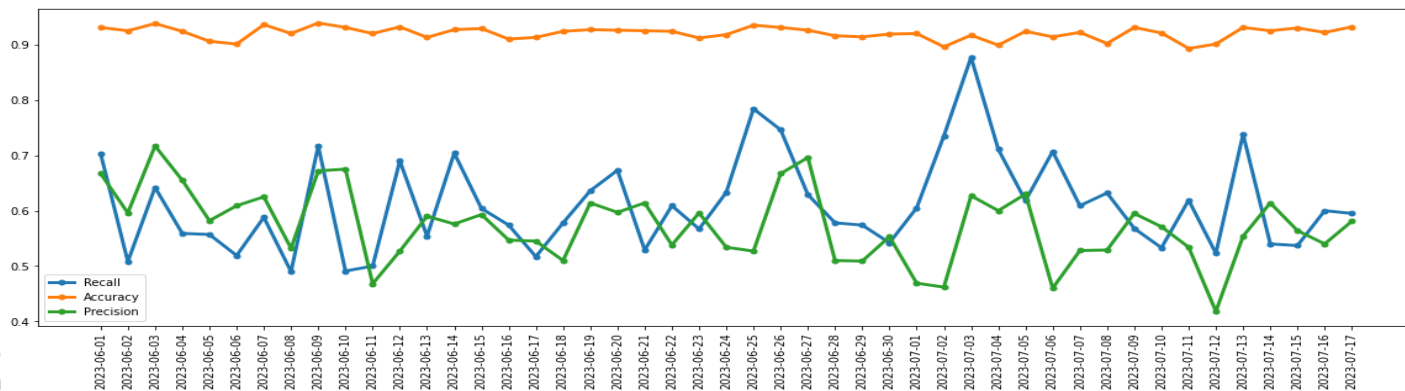
Training data time span: Last 4 months

Accuracy : 92%

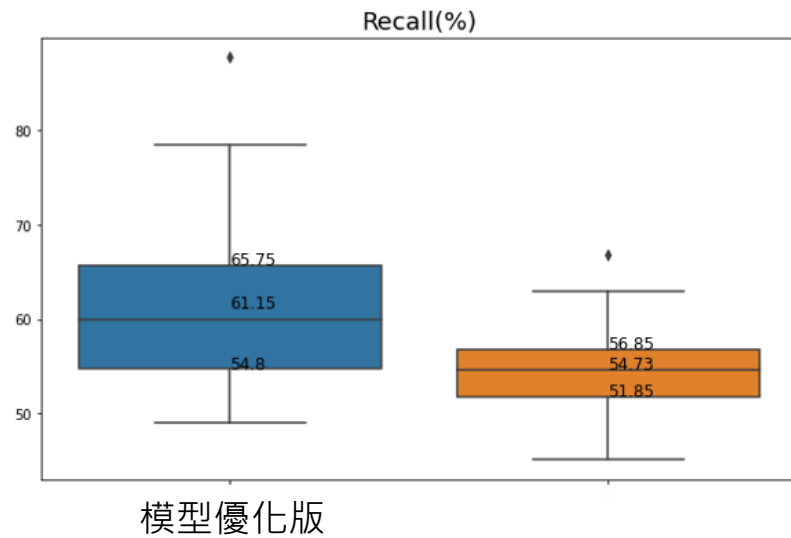
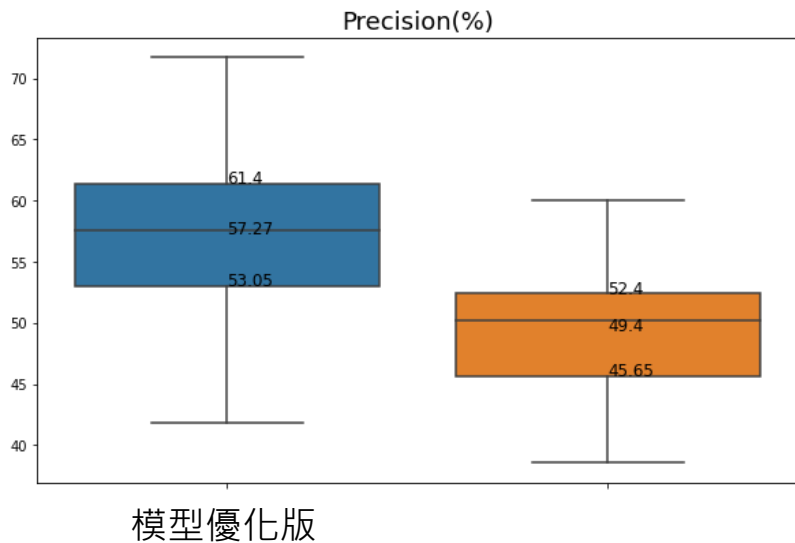
Recall : 61%

Precision : 57%

	Actual Churn	Actual Retention
Predicted Churn	5.8%	4.2%
Predicted Retention	3.7%	86.3%



模型表現 – 實際預測



模型表現 – 實際預測

預測時間：2023/06/01 – 2023/07/17

※每日平均資料

免費玩家

Accuracy：91%

Recall：63%

Precision：58%

共 人

實際流失

實際留存

預測為流失

預測為留存

付費玩家

Accuracy：94%

Recall：57%

Precision：56%

共 人

實際流失

實際留存

預測為流失

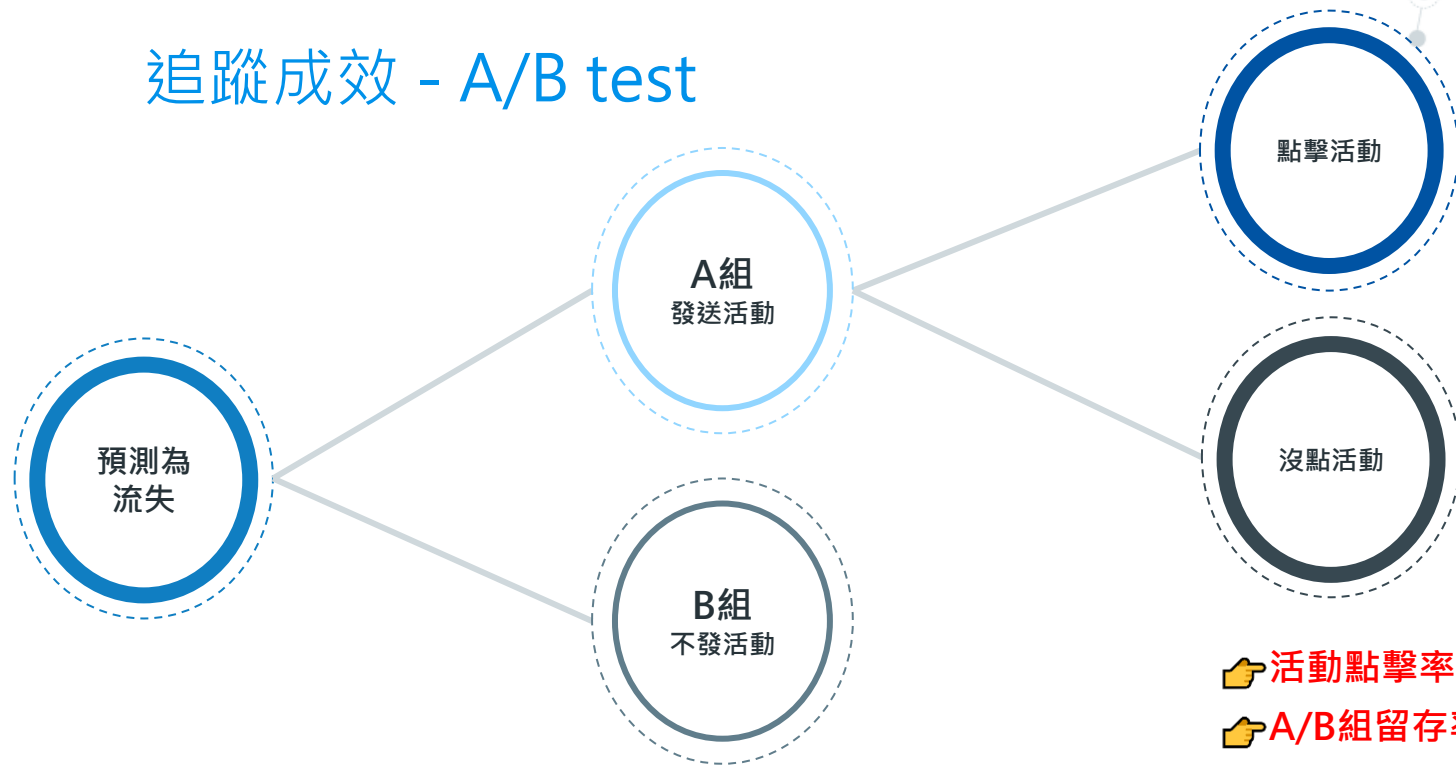
預測為留存

預測結果

帳號 流失風險 流失風險類別 活動AB組

	A	B	C	D
1	account_id	pred_probap	pred_class	group
2	111	0.946869	high	A
3	112	0.766034	medium	A
4	113	0.513532	low	A
5	221	0.849331	high	B
6	222	0.778911	medium	B
7	223	0.569415	low	B

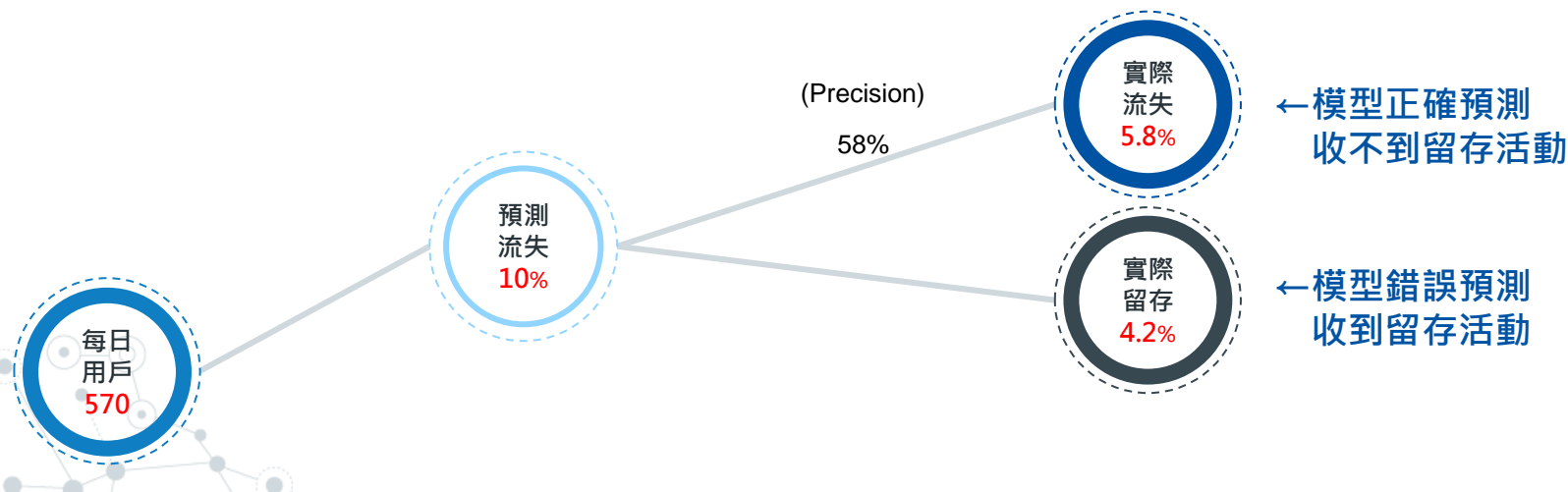
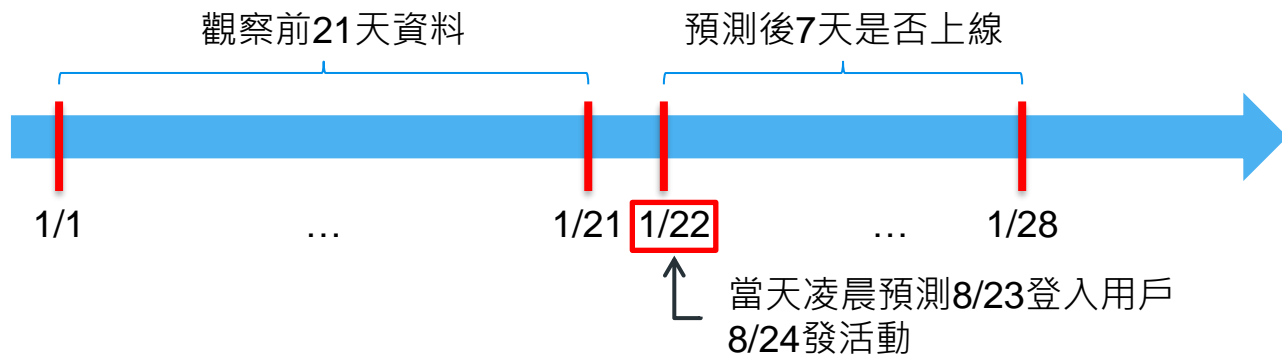
追蹤成效 - A/B test



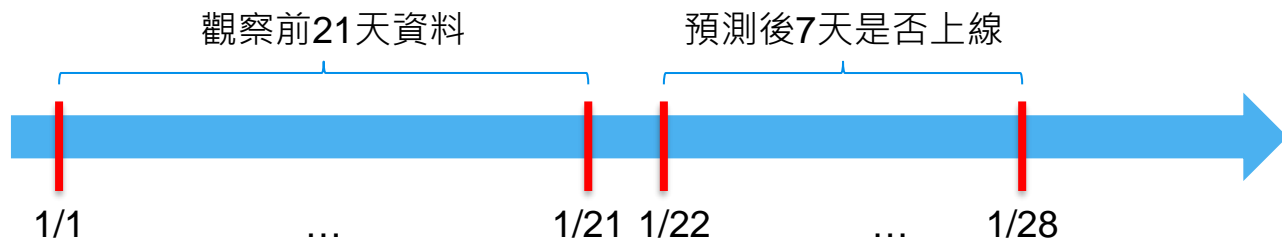
👉 活動點擊率 → 活動頁面設計

👉 A/B組留存率 → 活動獎勵設計

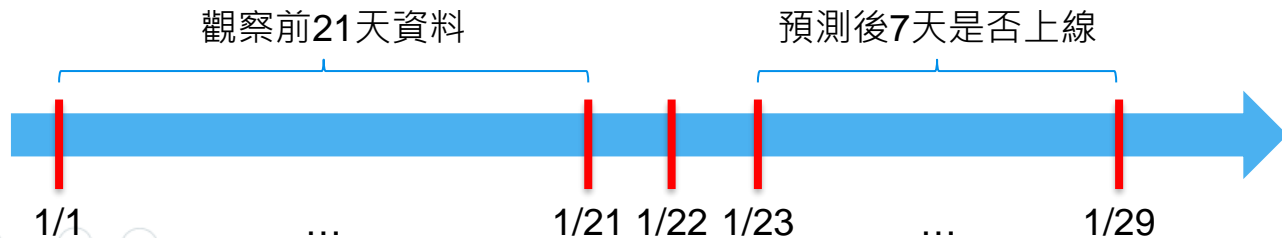
補充



改進方向



觀察與預測中間隔一日有上線並建模，
預測時使用21天資料做預測，讓用戶有機會在流失前看到活動



追蹤活動成效

整體 A組留存 (有給獎勵 - 流失預測)

7日 | 留存

日期	首次A組用戶數	當日	次日	3日	4日	5日	6日	7日
階段值	12440	2873 23.09%	2185 17.60%	1932 15.58%	1717 13.86%	1628 13.16%	1549 12.53%	1447 11.73%
	22	4 18.18%	1 4.55%	-	-	-	-	-
	21	4 19.05%	6 28.57%	1 4.76%	-	-	-	-
	10	2 20.00%	2 20.00%	1 10.00%	0 0.00%	-	-	-
	12	0 0.00%	0 0.00%	2 16.67%	1 8.33%	1 8.33%	-	-
	15	4 26.67%	4 26.67%	2 13.33%	2 13.33%	2 13.33%	1 6.67%	-

預測日有上7日留存率 = 50.4% (1447/2873)

整體 B組留存 (沒給獎勵 - 流失預測)

7日 | 留存

日期	首次B組用戶數	當日	次日	3日	4日	5日	6日	7日
階段值	12476	2778 22.27%	2146 17.24%	1771 14.24%	1610 12.97%	1494 12.04%	1454 11.73%	1339 10.81%
	22	2 25.93%	3 11.11%	-	-	-	-	-
	16	3 18.75%	3 18.75%	0 0.00%	-	-	-	-
	16	1 6.25%	2 12.50%	2 12.50%	2 12.50%	-	-	-
	9	1 11.11%	0 0.00%	2 22.22%	2 22.22%	0 0.00%	-	-
	9	1 11.11%	0 0.00%	0 0.00%	1 11.11%	1 11.11%	0 0.00%	-

預測日有上7日留存率 = 48.2% (1339/2146)

流失預測：有發活動且預測當日有上線

整體留存率提高 2.2% (48.2% > 50.4%)

追蹤活動成效

預測流失，有發活動：

當日有上7日留存率
高風險 = 30.7% (118/384)
低風險 = 51.8% (343/662)

pred_proba ▾	首次A組用戶數 ▾	當日 ▾	次日 ▾	3日 ▾	4日 ▾	5日 ▾	6日 ▾	7日 ▾
high	2803	384 13.70%	252 8.99%	204 7.28%	172 6.14%	154 5.49%	140 4.99%	118 4.21%
medium	1695	631 37.23%	422 24.90%	349 20.59%	300 17.70%	300 17.70%	252 14.87%	241 14.22%
low	1323	662 50.04%	517 39.08%	447 33.79%	411 31.07%	378 28.57%	337 25.47%	343 25.93%

預測流失，沒發活動：

當日有上7日留存率
高風險 = 41.1% (144/350)
低風險 = 44.6% (274/614)

pred_proba ▾	首次B組用戶數 ▾	當日 ▾	次日 ▾	3日 ▾	4日 ▾	5日 ▾	6日 ▾	7日 ▾
high	2778	350 12.60%	243 8.75%	186 6.70%	142 5.11%	133 4.79%	127 4.57%	144 5.18%
medium	1767	640 36.22%	422 23.88%	342 19.35%	300 16.98%	282 15.96%	250 14.15%	242 13.70%
low	1264	614 48.58%	503 39.79%	397 31.41%	362 28.64%	307 24.29%	295 23.34%	274 21.68%

結論：

流失機率高，收到活動7日留存反而下降 (41.1% > 30.7%)，推論玩家已不想玩過多的活動反而提升厭煩程度，需要再理解玩家流失的原因，或者對於高流失玩家過多的處理都沒有用

流失機率低，收到活動7日留存上升 (44.6% > 51.8%)，推論對遊戲還有興趣，提高活動力度顯著提升留存