



# 智能投資信鴿

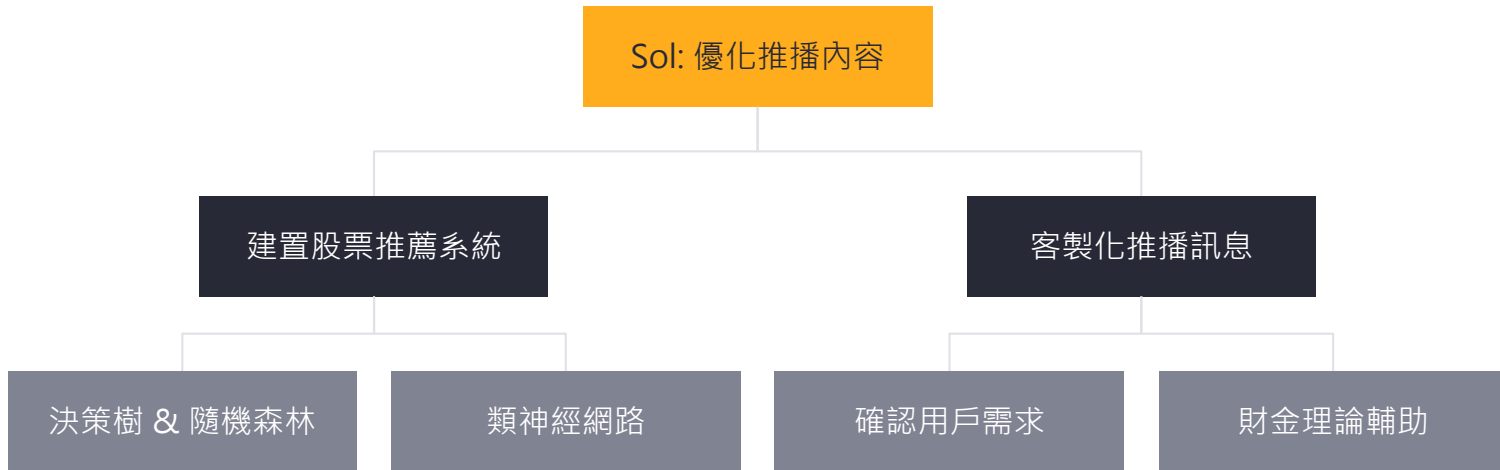


## 0620 期末報告

玉山 2-2 邱子軒 曹千佑 賴永祺 莊雅雯



# 業務痛點：如何提升推播服務滿意度？



# 建置股票推薦系統

透過機器學習的方法，預測用戶感興趣的股票類別，並從中推薦個股給用戶

# 數據集 Dataset (2019/10/27 – 2020/4/23) ---- 點擊紀錄

Unnamed: 0			timestamp	action	symbol_id	card_name	user_id
0	0	2019-10-27 15:58:58.449	open	2353	基本概況	3221	
1	1	2019-10-27 15:58:58.448	open	2353	基本資料	3221	
2	2	2019-10-27 15:58:58.447	open	2353	PTT批踢踢	3221	
3	3	2019-10-27 15:58:58.446	open	2353	股價K線	3221	
4	4	2019-10-27 15:58:58.445	open	2353	新聞	3221	

將點擊紀錄轉成矩陣。

	user_id	1605	2823	2884	2886	000980	0050	00642U	00672L	00757	...	TEX	DHI	JBLU	IEX	3524
0	0	132.0	399.0	95.0	12.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
1	1	0.0	0.0	24.0	0.0	8.0	20.0	8.0	200.0	9.0	...	0.0	0.0	0.0	0.0	0.0
2	2	0.0	0.0	89.0	0.0	0.0	206.0	9.0	84.0	9.0	...	0.0	0.0	0.0	0.0	0.0
3	3	0.0	0.0	0.0	6.0	0.0	60.0	56.0	0.0	63.0	...	0.0	0.0	0.0	0.0	0.0
4	4	18.0	0.0	24.0	108.0	0.0	16.0	0.0	0.0	9.0	...	0.0	0.0	0.0	0.0	0.0

# 數據集 Dataset (2019/10/27 – 2020/4/23) ---- 交易紀錄

	user_id	data_ym	industry	stock_categories	scale	freq
0	354	202003	電子-其他電子	電子中游-NB與手機零組件	小型股	6
1	354	202001	電子-光電	電子中游-LCD-零組件	中型股	1
2	354	202001	電子-其他電子	電子中游-儀器設備工程	小型股	2
3	354	202002	電子-半導體	電子上游-IC-代工	大型股	4
4	354	202004	觀光事業	傳產-觀光	中型股	1

1. 取買最多的交易紀錄。

2. 有購買 = 1 ,  
沒有購買 = 0 的交易紀錄 ,  
定義出 label 。

user_id	傳產-塑膠	傳產-生技	傳產-航運	金融-金控	電子上游-IC-設計	電子上游-LED及光元件	電子下游-電信服務	傳產-其他	傳產-汽車	...	電子下游-資訊通路	傳產-觀光	金融-保險	軟體-系統整合	電子下游-消費電子	傳產-玻璃陶瓷	傳產-高爾夫球	電子下游-顯示器	電子下游-掃描器	電子下游-數位相機
0	1	1	1	1	1	1	1	0	0	...	0	0	0	0	0	0	0	0	0	0
1	2	0	1	1	0	0	1	0	1	1	...	0	0	0	0	0	0	0	0	0
2	5	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
3	6	0	1	0	0	1	1	0	0	0	...	0	0	0	0	0	0	0	0	0
4	8	1	1	0	0	1	0	0	0	0	...	0	0	0	0	0	0	0	0	0

# 決策樹 & 隨機森林 Decision Tree & Random Forest

user_id	1605	2823	2884	2886	000980	0050	00642U	00672L	00757	...	JBLU	IEX	3524	000026	B00028	CTL	GOF	SNHNI	stock_categories
1	0.0	0.0	24.0	0.0	8.0	20.0	8.0	200.0	9.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	傳產-塑膠
2	0.0	0.0	89.0	0.0	0.0	206.0	9.0	84.0	9.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	傳產-生技
5	0.0	0.0	0.0	0.0	0.0	378.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	傳產-塑膠
6	18.0	18.0	433.0	104.0	0.0	954.0	114.0	87.0	81.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	電子上游-IC設計
8	0.0	0.0	29.0	120.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	傳產-生技
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
16270	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	電子下游-太陽能
16297	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	電子中游-其他
16317	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	傳產-紡織纖維
16387	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	傳產-塑膠
16532	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	電子下游-手機製造

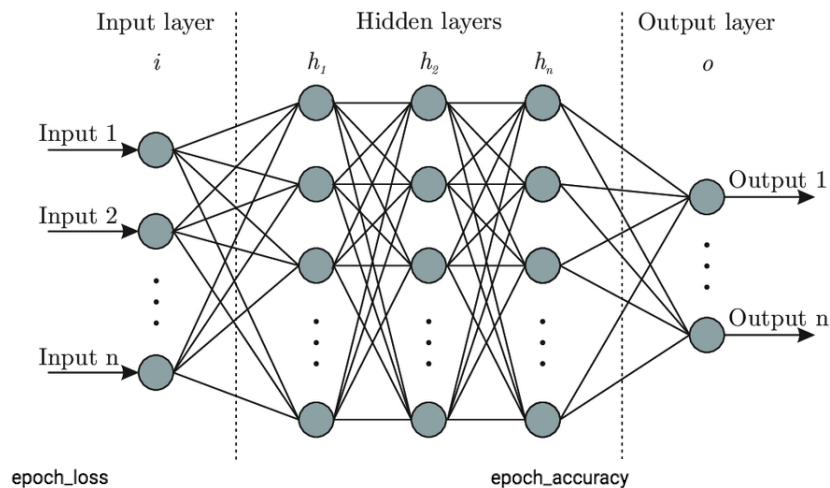
維度太高資料太少，用 decision tree 會容易 overfitting，於是使用 cross validation 來看是否造成 overfitting。

[2023,4115]

Depth	Accuracy
1	0.30879816869386456
2	0.3138070631553217
3	0.3245867908125429
4	0.3234357964499732
5	0.31753206204210616

Depth	Accuracy
6	0.3213466248825448
7	0.31929773566979625
8	0.3214741642590465
9	0.3131699294307054
10	0.3052939520342643

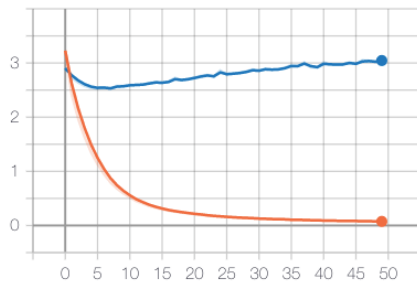
# 類神經網路 Neural Network



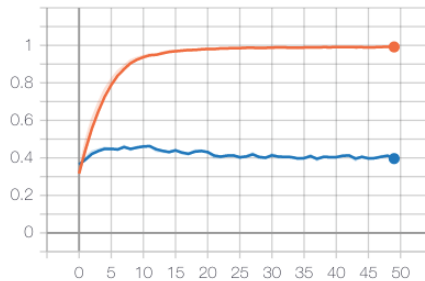
Cross\_entropy

One Hot Label

epoch\_loss



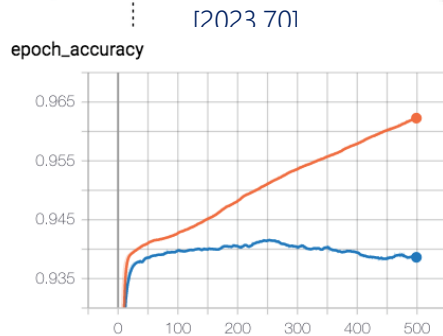
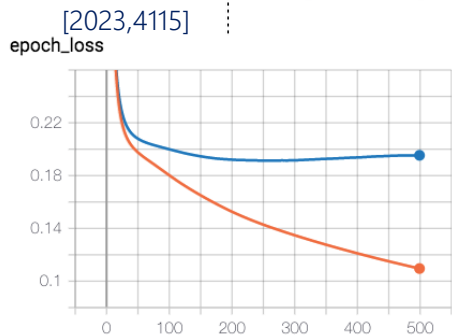
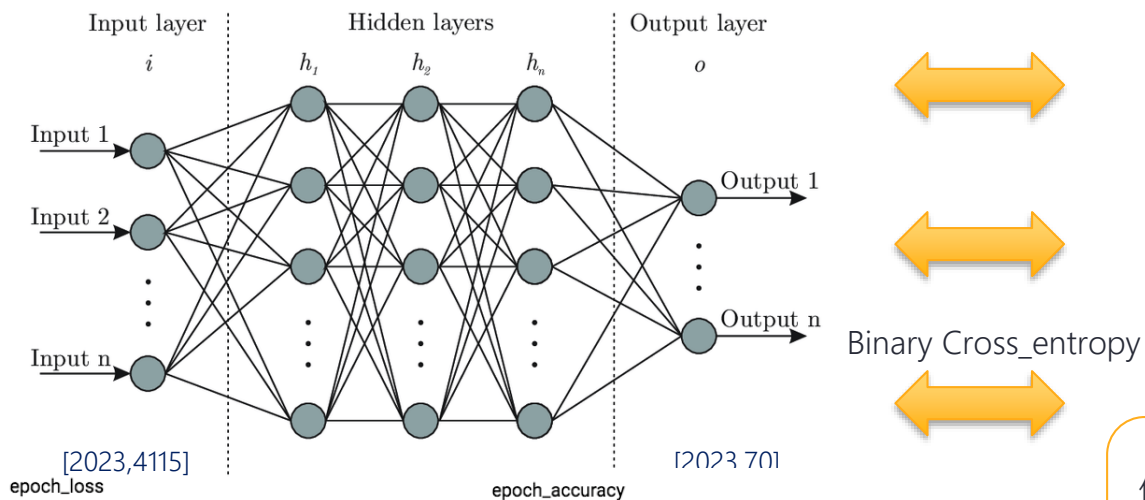
epoch\_accuracy



Dimension太高，因此  
使用 Neural Network 來  
克服此問題，正確率果  
然提升。

Accuracy : 0.46

# 類神經網路 Neural Network



使用 Multi-label來訓練，  
因為猜對 1 個就算對，  
所以正確率提升許多。

Accuracy : 0.94



## 5 個機器訓練的方法之成效比較

Method	Accuracy
SVM	0.303
Decision Tree	0.323
Random Forest + PCA	0.34
Neural Network	0.46
Neural Network (Multi)	0.94

效果最差

先降維度再執行後準確度較高，  
從0.32提升到0.34



## 客製化推播內容

從問卷結果可得知用戶感興趣的股市相關  
訊息，再結合財金理論進行客製化推播



## 股票投資行為需求調查問卷

總共分為 4 個項目設計問卷，包含投資經歷相關調查、投資金額及目的相關調查、投資行為相關調查、基本個人資訊。

截至 6/18 日，回收 209 份有效問卷。

問卷網址：

<https://www.surveycake.com/s/L0QN7>

### 股票投資行為需求調查

本問卷將僅供富果團隊與台大、政大、東吳大學跨校金融科技課程所推出之智能信鴿專案分析專用，資料絕對不會對外公佈，敬請安心填答。

為感謝您的參與，我們將會於下周一（15號）抽出 10 位幸運的回填用戶，並贈送富果獨家指標「券商買賣集中度 30 天使用權限（價值 199元）」喔！謝謝您！

#### 投資經歷相關調查

\*1 從您開始從事基金、股票等理財商品至今，共有多久投資經驗：

☐ 至今尚未

☐ 未滿 5 年

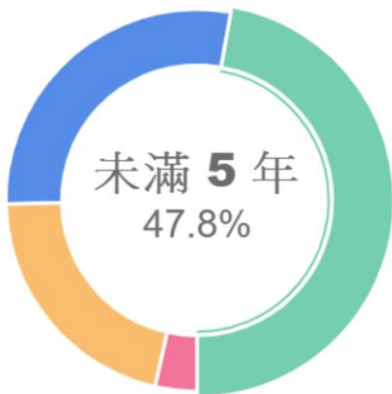
☐ 5 年（含）以上，未滿 10 年

☐ 10 年（含）以上

# 問卷結果

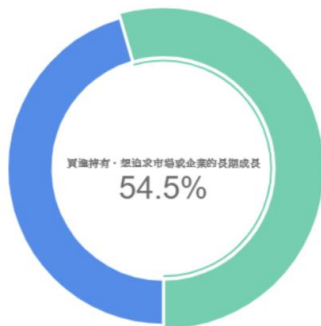
## 用戶投資經驗

- 100 名用戶的投資經驗未滿五年。
- 近 60 名用戶的投資經驗 10 年以上。



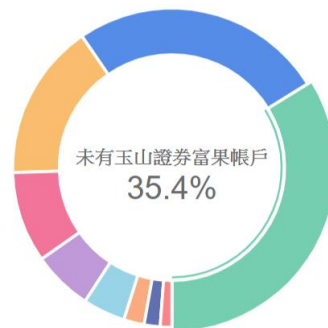
## 用戶投資策略

- 近 115 名用戶的投資策略為：買進持有，想追求市場或企業的長期成長。
- 其餘的投資策略為：買低賣高，波段操作。



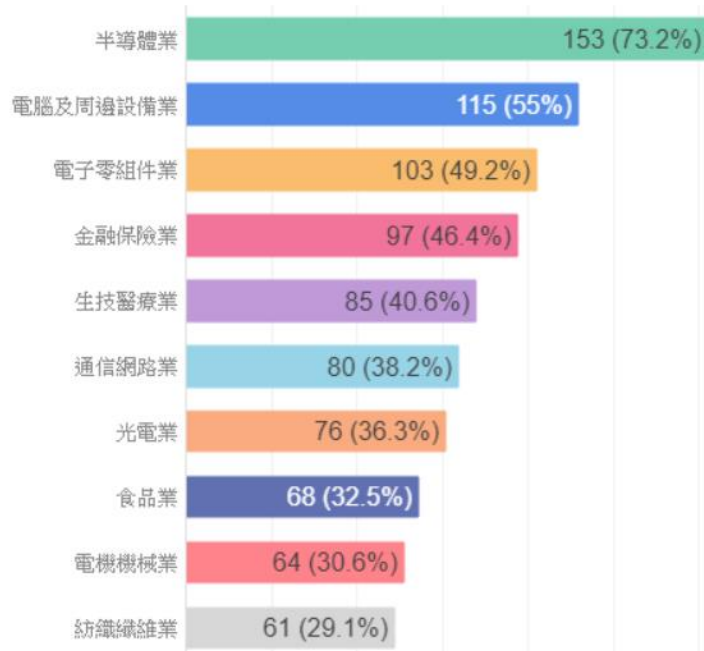
## 使用富果帳戶月平均交易金額

- 近 75 名用戶未有玉山證券富果帳戶。
- 近 55 名用戶的投資金額未滿 5 萬元。

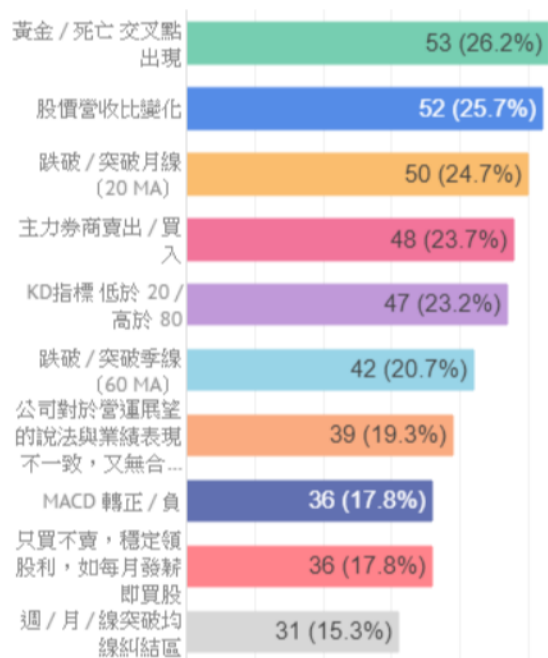


# 問卷結果

## 用戶曾經試圖下單或成交的產業股票



## 用戶通常判斷進出場的訊號



# 財金理論輔助

## 股票風險值

- 讓投資人可以簡單試算他們投資組合的風險。
- 參考股票歷史資料，並以變異數-共變異數法、歷史模擬法計算風險值，提供投資人一個參考的風險指標。

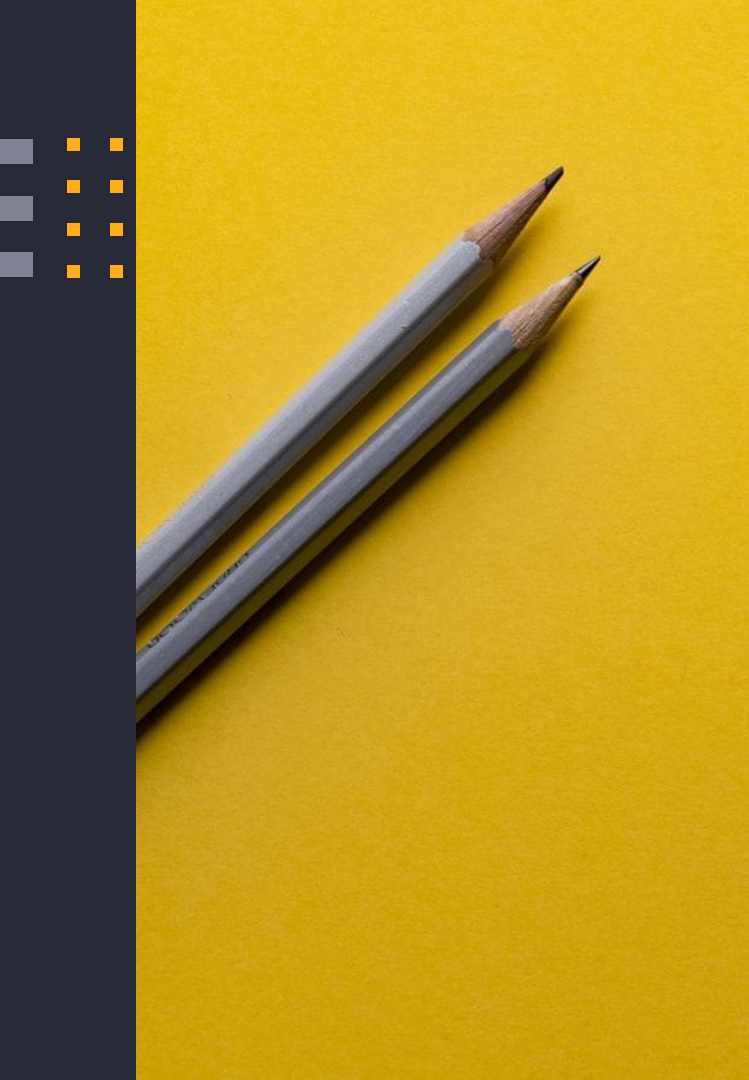
## 股價歷史指標

- 透過機器學習，預測出投資人有興趣的股票。
- 在該股股價到達52週高/低點、歷史新高/低時，推播通知投資人。



## 利差反推違約

- 投資人若想知道一家公司體質的好壞，最客觀的指標是公司債和公債的利差。
- 利差高代表風險高，反之亦然。
- 計算方法參考 Duffie and Singleton 的違約機率推導。(如附錄)



Thank You!

# Appendix

Pre 指的是利差，  
最後算出來的機率是 CDF  
(Cumulative default probability)

$n = 1$ 時<sup>↵</sup>

$$pre \cdot (1 - p_1)D_1 = (1 - R)(p_1 - p_0)D_1 \quad \text{↵}$$

$$\Leftrightarrow pre \cdot D_1 - pre \cdot p_1 D_1 = (1 - R)p_1 D_1 \quad \text{↵}$$

$$\Leftrightarrow (1 + pre - R)p_1 = pre \quad \text{↵}$$

$$\Leftrightarrow p_1 = \frac{pre}{1 + pre - R} \quad \text{↵}$$

$n \geq 2$ 時<sup>↵</sup>

$$pre \sum_{k=1}^n (1 - p_k)D_k = (1 - R) \sum_{k=1}^n (p_k - p_{k-1})D_k \quad \text{↵}$$

$$\Leftrightarrow pre \left[ \sum_{k=1}^{n-1} (1 - p_k)D_k + (1 - p_n)D_n \right] = (1 - R) \left[ \sum_{k=1}^{n-1} (p_k - p_{k-1})D_k + (p_n - p_{n-1})D_n \right] \quad \text{↵}$$

$$\Leftrightarrow pre \left[ \sum_{k=1}^{n-1} (1 - p_k)D_k + D_n \right] - pre \cdot p_n D_n = (1 - R) \left[ \sum_{k=1}^{n-1} (p_k - p_{k-1})D_k - p_{n-1}D_n \right] + (1 - R)p_n D_n \quad \text{↵}$$

$$\Leftrightarrow (1 + pre - R)p_n D_n = pre \left[ \sum_{k=1}^{n-1} (1 - p_k)D_k + D_n \right] - (1 - R) \left[ \sum_{k=1}^{n-1} (p_k - p_{k-1})D_k - p_{n-1}D_n \right] \quad \text{↵}$$

$$\Leftrightarrow p_n = \frac{pre \left[ \sum_{k=1}^{n-1} (1 - p_k)D_k + D_n \right] - (1 - R) \left[ \sum_{k=1}^{n-1} (p_k - p_{k-1})D_k - p_{n-1}D_n \right]}{(1 + pre - R)D_n} \quad \text{↵}$$