

# 野村投信 | 運用熱門主題及字組 提升行銷轉換成率之探索性研究

指導業師：野村投信 張孝璿、陳景堯

指導老師：蔡芸琇老師

小組成員：謝天青、莊子揚、簡辰芳、簡君哲、鄭羽健

# 報告大綱

1. 專案背景與目標

2. 商業發想

3. 技術說明

4. 驗證成效

5. 優化建議

6. 未來展望

7. 團隊分工

# 1 專案背景與目標

# 專案背景與目標

## 目前現況

採用 TOP-DOWN 行銷流程

## 專案預期

採用 BOTTOM-UP 行銷流程，由熱門關鍵字進行產品推薦

## 專案目標

藉由輿論及新聞得出之熱門關鍵字，協助投信公司推薦  
績效更優良之產品給其客戶

## 2 商業發想

# 商業發想內容

## 建立熱門關鍵字基金推薦系統

在野村銷售管道，如：野村官網、野村理財通app，藉由熱門關鍵字推薦相關之基金給客戶，並提供用戶基金績效表及視覺化回測圖



# 網站呈現



# 網站呈現

FUND PILOT

新聞 社群 回頂部

關鍵字精選基金快查

請輸入要查詢週數 (1~20周)  
例：欲查找第3周的精選基金，請輸入 03  
本網站涵蓋 自2020/12/28至2021/05/16共20周之精選基金

使用者輸入其關心週數



FUND PILOT

野村投信專題—  
台灣大學：財金四 | 莊子揚 財金四 | 簡辰芳 財金碩一 | 簡君哲 會研碩二 | 鄭羽健  
東吳大學：巨資二 | 謝天青



# 網站呈現

FUND PILOT

新聞

社群

回頂部

關鍵字精選基金快查

01

請輸入要查詢週數 (1~20周)

例：欲查找第3周的精選基金，請輸入 03

本網站涵蓋 自2020/12/28至2021/05/16共20周之精選基金

## 推薦基金之績效表現

## 基金淨值 vs 大盤走勢圖



週數	區間	基金代碼	基金名稱	周報酬%	月報酬%	夏普指數
01	2020/12/28~2021/01/03	881	國泰台灣5G+	6.07	12.47	1.31
01	2020/12/28~2021/01/03	T3201S	野村優質S	7.12	11.45	1.66
01	2020/12/28~2021/01/03	T3228Y	野村高科技	5.31	5.6	1.57

FUND PILOT

野村投信專題一

台灣大學：財金四 | 莊子揚 財金四 | 簡辰芳 財金碩一 | 簡君哲 會研碩二 | 鄭羽健

東吳大學：巨資二 | 謝天青

# 3 技術説明

# 技術概述

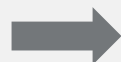
- ✓ 使用 AWS EC2 服務開啟一部雲端 instance 搭載 ubuntu server 18.04 ，並使用 tasksel 軟體包實現以 LAMP 架構部屬線上網站，其中包含前端的 HTML 網頁、後端的 PHP 網頁以及以 MYSQL 為基礎的 phpmyadmin 資料庫
- ✓ 使用python語言來進行資料的爬取、正規化、文字探勘與深入分析工作

# 文字探勘流程

資料蒐集



資料清理



資料分析

Python 爬蟲套件

- BeautifulSoup
- selenium

# 爬蟲資料來源

## 基金資訊

- TEJ
- 投信投顧公會

## 社群輿論

- ptt 股
- mobile01
- Dcard
- Emoney
- 股市爆料同學會
- 財報狗討論區

## 新聞內容

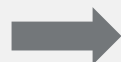
- 鉅亨
- MoneyDJ
- Yahoo 新聞
- Pchome 股市

# 文字探勘流程

資料蒐集



資料清理



資料分析

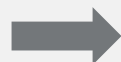
- 正規表達式
- 結巴分詞

# 文字探勘流程

資料蒐集



資料清理



資料分析

關鍵字萃取模型

- 卡方
- TFIDF
- Text Rank

# 1. 萃取每週熱門關鍵字

2021/1/1	['台積電', '長榮', '鴻海', '陽明', '國際', '電池', '電動車', '手機', '基金', '產能', '特斯拉', '半導體', '海運', '5g', '蘋果', '中芯', '運價', '製程', 'ptt', 'eps']
2021/1/8	['台積電', '長榮', '汽車', '水手', '特斯拉', '電動車', '疫苗', '高檔', '水怕', '奈米', '貨櫃', '英特爾', '晶片', '水水', '運價', '航運', '產能', '市值', '海運', '風險']
2021/1/15	['台積電', '台積', '疫苗', '長榮', '晶片', 'gg', '汽車', '5g', '半導體', '股王', '陽明', '未亡', '蘋果', '手機', '電動車', '主力', '當沖', '成本', '風險', 'etf']
2021/1/22	['台積電', '晶片', '英特爾', 'gg', '產能', '半導體', '台積', '生產', '疫苗', '三星', '車用', '特斯拉', '汽車', '電動車', '聯電', '製程', '拜登', '代工', '風險', '美股']
2021/1/29	['選項', '台積電', 'gme', '第一', 'gamestop', '疫苗', '基金', 'robinhood', '晶片', '機構', 'reddit', 'ptt', '華爾街', '車用', '風險', '做空', '產能', '放空', '美股', '鄉民']
2021/2/5	['台積電', '富驊', '連生', '楊梅', '本件', '鄭正忠', '遠百', '晶片', '疫苗', '基金', '全特', '盡量', '特斯拉', '能不能', '濫竽充數', '點擊', '標明', '行文', '過長', '過短']
2021/2/12	['開春', '不羈', '基金', '台積電', '大吉', 'ptt', '33m', 'm4', 'm3', '蘋果', '團圓飯', 'm2', '晶片', '特斯拉', '半導體', '美股', '盡量', '汽車', '行文', '能不能']
2021/2/19	['台積電', '蘋果', '疫苗', '半導體', '行文', '能不能', '汽車', '盡量', '晶片', '過短', '點擊', '段落', '過長', '濫竽充數', '標明', '特斯拉', '生產', '現金', '成本', '電動車']
2021/2/26	['台積電', '晶片', '美股', '疫苗', '能不能', '半導體', '盡量', '行文', '過短', '點擊', '過長', '標明', '段落', '濫竽充數', '鴻海', '電動車', '成本', '汽車', '國際', '必需']
2021/3/5	['台積電', '特斯拉', '能不能', '盡量', '段落', '標明', '殖利率', '行文', '過長', '過短', '點擊', '濫竽充數', '電動車', '電池', '疫苗', '成本', '美股', '半導體', '基金', '現金']
2021/3/12	['露西', '股拉', '台積電', '時乃空', '能不能', '疫苗', '盡量', '現金', '過短', '濫竽充數', '標明', '過長', '行文', '點擊', '段落', '鴻海', 'etf', '好姨', '工具', '風險']
2021/3/19	['台積電', '現金', '英特爾', '代工', '公積', '晶片', 'of', '鴻海', '運河', '電動車', '長榮', '配股', '半導體', '疫苗', '能不能', '盡量', '轉增資', 'in', '點擊', '蘇伊士']



## 2. 利用關鍵字找出主題與基金

區間	主題	對應基金	weekly return	monthly return	sharpe ratio
2021/1/1~2021/01/07	電動車，電池，5G	合庫AI電動車及車聯網創新基金(台幣)	3.9491	4.8713	0.3877
		日盛全球智能車	1.7072	5.6467	0.3973
		統一全球智聯網AIoT基金(台幣)	-1.2238	5.4513	0.6099
2021/1/8~2021/01/14	航運類股，長榮，貨櫃	元大中型100	-2.5988	2.1522	0.3212
		元大臺灣ESG永續	3.5973	5.3232	0.4203
		富邦公司治理	5.2792	6.1736	0.4221
2021/1/15~2021/01/21	晶片，5G，台積電	野村高科技基金	-0.1076	5.3125	0.4317
		國泰科技生化基金(新台幣)	-2.0006	5.7187	0.4518
		元大高科技基金	-2.6358	7.7167	0.3346
2021/1/22~2021/01/28	台積電，晶片，英特爾	野村高科技基金	1.5987	5.4113	0.4317
		國泰科技生化基金(新台幣)	1.4467	4.4018	0.4518
		元大高科技基金	2.6745	6.1644	0.3346
2021/1/29~2021/2/4	GME，晶片，美股	瀚亞美國高科技基金A類型	3.5627	-8.5267	0.3908
		國泰標普北美科技ETF基金	1.716	-6.6568	0.2758
		JPM美國科技(美元)	4.9522	-7.4229	2.04
2021/2/5~2021/2/11	台積電，晶片，富驊	野村高科技基金	7.0551	0.7349	0.532
		國泰科技生化基金(新台幣)	5.4459	4.12	0.5513
		元大高科技基金	5.6095	0.6268	0.4294

### 3. 提供基金之淨值變化與夏普比例

區間	主題	對應基金	weekly return	monthly return	sharpe ratio
2021/1/1~2021/01/07	電動車，電池，5G	合庫AI電動車及車聯網創新基金(台幣)	3.9491	4.8713	0.3877
		日盛全球智能車	1.7072	5.6467	0.3973
		統一全球智聯網AIoT基金(台幣)	-1.2238	5.4513	0.6099
2021/1/8~2021/01/14	航運類股，長榮，貨櫃	元大中型100	-2.5988	2.1522	0.3212
		元大臺灣ESG永續	3.5973	5.3232	0.4203
		富邦公司治理	5.2792	6.1736	0.4221
2021/1/15~2021/01/21	晶片，5G，台積電	野村高科技基金	-0.1076	5.3125	0.4317
		國泰科技生化基金(新台幣)	-2.0006	5.7187	0.4518
		元大高科技基金	-2.6358	7.7167	0.3346
2021/1/22~2021/01/28	台積電，晶片，英特爾	野村高科技基金	1.5987	5.4113	0.4317
		國泰科技生化基金(新台幣)	1.4467	4.4018	0.4518
		元大高科技基金	2.6745	6.1644	0.3346
2021/1/29~2021/2/4	GME，晶片，美股	瀚亞美國高科技基金A類型	3.5627	-8.5267	0.3908
		國泰標普北美科技ETF基金	1.716	-6.6568	0.2758
		JPM美國科技(美元)	4.9522	-7.4229	2.04
2021/2/5~2021/2/11	台積電，晶片，富驊	野村高科技基金	7.0551	0.7349	0.532
		國泰科技生化基金(新台幣)	5.4459	4.12	0.5513
		元大高科技基金	5.6095	0.6268	0.4294

# 文字探勘流程

## 資料蒐集

Python 爬蟲套件

- BeautifulSoup
- selenium

## 資料清理

- 正規表達式
- 結巴分詞

## 資料分析

關鍵字萃取模型

- 卡方
- TFIDF
- Text Rank

# 網頁端-頁面跳轉按鈕

FUND PILOT

新聞

社群

回頂部

關鍵字精選基金快查

請輸入要查詢週數 (1~20周)

例：欲查找第3周的精選基金，請輸入 03

本網站涵蓋 自2020/12/28至2021/05/16共20周之精選基金

使用簡單的[<a href>](#)元素搭配 css 中的  
transition & hover 指令來實現當鼠標移至  
聯結文字時發生動畫與點擊跳轉頁面

```
13     </div>
14     <div class="headsplitt col-5">
15       <div class="col-3">
16         &nbsp;
17       </div>
18       <div class="set1 col-3">
19         <a href="/news.php" class="a full">
20           <span class="headfont">新聞</span>
21         </a>
22       </div>
23       <div class="set1 col-3">
24         <a href="/social.php" class="a">
25           <span class="headfont">社群</span>
26         </a>
27       </div>
28       <div class="set1 col-3">
29         <a href="#top" class="a">
30           <span class="headfont">回頂部</span>
31         </a>
```



運用 `<input>` 元素打造出搜尋框並  
使用 `oninput` 屬性來達成當 `input`  
中不同的值可對應到不同的圖表

# 網頁端 - 搜尋列聯結後台 php

## 關鍵字精選基金快查

請輸入要查詢週數 (1~20周)

例：欲查找第3周的精選基金，請輸入 03

本網站涵蓋 自2020/12/28至2021/05/16共20周之精選基金

運用 JQuery 中的 AJAX 架構來實現  
非同步向資料庫呼叫需求，並由資料  
庫回傳相符數據給前端 html 網頁

```
103 <div class="col-2">&nbsp;&nbsp;&nbsp;</div>
104 </div>
105
106
107 <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.4.0/jquery.min.js"></script>
108 <script>
109 $(document).ready(function() { //文件準備
110   $("input.search").keydown(function(){ //填上標籤選擇器後的search，事件為鍵盤按下
111     $("input.search").css("background-color", "#d3f9fc");
112     $("#load").html("<i class='fa fa-spinner fa-spin fa-fw'></i>");
113   });
114   $("input.search").keyup(function() {
115     $("input.search").css("background-color", "#ffffff");
116     $("#load").html("");
117     //使用ajax將傳給後端來做回應
118     $.ajax({
119       type : "GET",
120       url : "news_address.php",
121       data : {
122         txt : $("#username").val(), //將id存入txt參數，傳給後端做接收
123       },
124       dataType : 'html' //設定該網頁回應的會是 html 格式
125     }).done(function(data) {
126       //成功會秀出
127       if(data == ""){
128         $("#result").hide();
129       }else{
130         $("#result").html(data).show();
131       }
132       console.log(data);
133     }).fail(function(jqXHR, textStatus, errorThrown) {
134       //失敗會秀出
135       // alert("有錯誤，請至 console log查看原因");
136       // console.log(jqXHR.responseText);
137     });
138   });
139 });
140 </script>
141 <style>
142 #chart1,#chart2,#chart3,#chart4,#chart5,#chart6,#chart7,#chart8,#chart9,#chart10,#chart11,#chart12,#
```

# 網頁端 - 後台 table

週數	區間	基金代碼	基金名稱	周報酬%	月報酬%	夏普指數
01	2020/12/28~2021/01/03	881	國泰台灣5G+	6.07	12.47	1.31
01	2020/12/28~2021/01/03	T3201S	野村優質S	7.12	11.45	1.66
01	2020/12/28~2021/01/03	T3228Y	野村高科技	5.31	5.6	1.57

透過 html 網頁中的 AJAX 架構聯結後端 php 語法，運用php呼叫 phpmyadmin 資料庫加上 MYSQL 語言篩選出符合前端 html 網頁內

```
17 if($_GET['txt'] != "") {
18     $n = $_GET['txt'];
19     /*從score資料庫，擷取所有欄位(用*米字號)，當所傳來的值是字首、中、尾有符合(LIKE)資料name欄位的名字，
20     就會呼叫出來，並且透過id欄位來設定為降序(ASC是升序(小跑到大)，DESC是降序(大跑到小))*/
21     $sql = "SELECT * FROM `social_data` WHERE week LIKE '%$n%' ORDER BY week ASC";
22     $score_result = $kt_conn->query($sql);
23 }
24
25 echo "<table border='1'>
26     <tr>
27         <th>週數</th>
28         <th>區間</th>
29         <th>基金代碼</th>
30         <th>基金名稱</th>
31         <th>周報酬%</th>
32         <th>月報酬%</th>
33         <th>夏普指數</th>
34     </tr>";
35
36 if(!empty($score_result)) {
37     echo "RESULT";
38 }
39 else{
40     while($row = $score_result->fetch_assoc() )
41     {
42         echo "<tr>";
43         echo "<td>" . $row['week'] . "</td>";
44         echo "<td>" . $row['date'] . "</td>";
45         echo "<td>" . $row['code'] . "</td>";
46         echo "<td>" . $row['name'] . "</td>";
47         echo "<td>" . $row['week_remuneration'] . "</td>";
48         echo "<td>" . $row['month_remuneration'] . "</td>";
49         echo "<td>" . $row['sharp_index'] . "</td>";
50         echo "</tr>";
51     }
52     echo "</table>";
53 }
54
55 //關閉資料庫
56 $kt_conn->close();
```



# 網頁端 - 前端圖表



使用 javascript 內的 switch 判斷式搭配 `getElementById()` 語法控制所輸入 `<input>` 內的週數可對應至相符的圖表

```

378 <script>
379 function a() {
380     var day=document.getElementById('username').value;
381     switch (day) {
382         case '01':
383             document.getElementById("chart1").style.display="block";
384             document.getElementById("chart2").style.display="none";
385             document.getElementById("chart3").style.display="none";
386             document.getElementById("chart4").style.display="none";
387             document.getElementById("chart5").style.display="none";
388             document.getElementById("chart6").style.display="none";
389             document.getElementById("chart7").style.display="none";
390             document.getElementById("chart8").style.display="none";
391             document.getElementById("chart9").style.display="none";
392             document.getElementById("chart10").style.display="none";
393             document.getElementById("chart11").style.display="none";
394             document.getElementById("chart12").style.display="none";
395             document.getElementById("chart13").style.display="none";
396             document.getElementById("chart14").style.display="none";
397             document.getElementById("chart15").style.display="none";
398             document.getElementById("chart16").style.display="none";
399             document.getElementById("chart17").style.display="none";
400             document.getElementById("chart18").style.display="none";
401             document.getElementById("chart19").style.display="none";
402             document.getElementById("chart20").style.display="none";
403             document.getElementById("chart21").style.display="none";
404
405             break;
406         case '02':
407             document.getElementById("chart1").style.display="none";
408             document.getElementById("chart2").style.display="block";
409             document.getElementById("chart3").style.display="none";
410             document.getElementById("chart4").style.display="none";
411             document.getElementById("chart5").style.display="none";
412             document.getElementById("chart6").style.display="none";
413             document.getElementById("chart7").style.display="none";
414             document.getElementById("chart8").style.display="none";
415             document.getElementById("chart9").style.display="none";
416             document.getElementById("chart10").style.display="none";
417             document.getElementById("chart11").style.display="none";

```



## 4 驗證成效

# 驗證成效

## 驗證目標

精選基金是否能提供更好績效表現

## 驗證方式

利用大盤績效做為基準，檢驗60檔精選基金中績效優於大盤之數量

## 驗證結果

精選基金優於大盤之比率

- 周報酬：45% (27/60)
- 月報酬：53% (30/57)

→ 精選基金表現無法顯著優於大盤

# 分析原因

1. 爬蟲資料：網站以台股市場為主，忽略國際市場或是總體經濟之趨勢影響
2. 文字探勘：並無進行情緒分析或建立漲跌訊號語意庫，單以熱門作為關鍵字選擇標準，無法判斷該關鍵字對於市場是具正面/負面/沒有影響
3. 關鍵字連結精選基金：未建立利用關鍵字連結基金之標準，造成精選基金無法有效代表關鍵字所連結到之標的
4. 其他：實驗樣本數不足 (關鍵字數量、推薦基金數量)、實驗期間過短 (6個月)

# 5 優化建議

# 專案研究優化建議

1. 爬蟲資料來源：專案以台股的社群及新聞網站為主要資料來源，考量基金的產品類型及市場，可擴大爬蟲資訊管道，例：[investing.com](https://www.investing.com)
2. 文字探勘模型：專案以卡方、TFIDF、Text Rank等方法產出熱門關鍵字，但缺乏考量關鍵字對於股價的漲跌影響及程度，建議可分為新聞及社群兩種資料類別，參考以下方法優化關鍵字分析：

# 專案研究優化建議

新聞：

(a) 參考 Mittermayer ( 2004 ) 之 Support Vector Machine, SVM , 可使用SVM 進行資訊的二元分類：助漲、助跌，其餘則分類為無影響

(b) 參考 Lavrenko, Schmill, Lawrie, Ogilvie, Jensen & Allan ( 2000 ) 之 Language Model , 將資訊內文所連結的字眼進行漲跌影響分析，例如「破產」對於股價有跌幅影響，「併購」則有漲幅影響等

# 專案研究優化建議

社群：

(a) 建立市場情緒指標 ( Sentiment Ratio , SR ) ，建立語意庫分析輿論的樂觀或悲觀程度，量化市場情緒以協助預測標的漲跌，同時可搭配其他機構所發佈的情緒指標優化，如：美國散戶投資人情緒指數、消費者信心指數、VIX恐慌指數等

# 專案研究優化建議

3. 熱門關鍵字連結主題基金：專案將當周之社群及新聞20~40個關鍵字中篩選出之1~3個代表性關鍵字以選擇對應基金，但未建立利用關鍵字選擇基金之標準，若是以專案研究的台股市場為例，建議可利用以下方式優化：

- (a) 利用代表性關鍵字連結到相關m檔個股
- (b) 尋找與相關m檔個股占基金資產比最高之n檔對應基金
- (c) 利用基金績效篩選出前3檔對應基金



## 6 未來展望

# 未來展望

專案以驗證「熱門關鍵字」可協助投信公司進行更高績效的基金推薦為主，因此是利用基金歷史資料進行回測，建立在熱門關鍵字確實可協助推薦更佳報酬的基金之前提，小組建議公司官網及app可新增以下服務：

# 未來展望

## 熱門關鍵字排行榜

利用新聞及輿論文本進行資料探勘得出之關鍵字列出排行榜，讓消費者可以快速掌握目前市場的熱門投資主題

## 市場情緒溫度計

參考股市爆料同學會，讓大眾針對本日市場看漲或看跌發表意見，量化市場的輿論情緒以協助投資人進行交易

## 關鍵字基金推薦

利用關鍵字推薦對應之基金產品，並提供用戶基金績效表及視覺化回測圖

# 7 團隊分工

# 團隊分工

- 台大財金四 簡辰芳  
驗證成效、優化建議、未來展望、PPT製作與美編、影片製作
- 台大財金碩一 簡君哲  
社群資料爬蟲、斷詞、關鍵字擷取 ( TF-IDF )
- 東吳巨資二 謝天青  
社群資料爬蟲、斷詞、網頁架設、整合資料庫
- 台大會研碩二 鄭羽捷  
財經新聞爬蟲、斷詞、關鍵字擷取 ( TF-IDF、TextRank )、PPT製作、影片製作
- 台大財金四 莊子揚  
財經新聞爬蟲、斷詞、關鍵字擷取 ( 卡方 )、報酬表格及圖表產出