



國立高雄科技大學
企業管理系碩士班
碩士論文

多來源網頁資料對股市投資影響之分析：
以台灣股市為例。

Analysis of the Impact on Stock Market Investment Using
Multi-source Web Data: Taking Taiwan Stock Market as an
Example

研究生：阮泰楠
指導教授：李杭 博士
共同指導教授：陳俊豪 博士

中華民國 112 年 07 月

多來源網頁資料對股市投資影響之分析：
以台灣股市為例。

**Analysis of the Impact on Stock Market
Investment Using Multi-source Web Data: Taking
Taiwan Stock Market as an Example**

研究生：阮泰楠

指導教授：李杭 博士

共同指導教授：陳俊豪 博士

國立高雄科技大學

企業管理系碩士班

碩士論文

A Thesis

Submitted to

Department of Business Administration

National Kaohsiung University of Science and Technology

In Partial Fulfillment of Requirements

For the Degree of Master of Business Administration

June 2023

Kaohsiung, Taiwan, Republic of China

中華民國 112 年 07 月

國立高雄科技大學(燕巢校區)研究所學位論文考試審定書

企業管理系 碩士班

研究生 阮泰楠 所提之論文

論文名稱(中文): 多來源網頁資料對股市投資影響之分析: 以台灣股市為例。

論文名稱(英/日/德文): Analysis of the Impact on Stock Market Investment Using Multi-source Web Data: Taking Taiwan Stock Market as an Example.

經本委員會評審, 符合碩士學位論文標準。

學位考試委員會

召集人

李詠騏

簽章

委員

李詠騏

蘇永輝

傅俊亭

李杭

指導教授

李杭 傅俊亭

簽章

系所主管

王崇星

簽章

中華民國 112 年 月 日

保存期限: 永久

國立高雄科技大學學位論文著作權歸屬協議書

論文名稱：多來源網頁資料對股市投資影響之分析：以台灣股市為例。
Analysis of the Impact on Stock Market Investment Using Multi-source Web Data:
Taking Taiwan Stock Market as an Example.

研究生：阮泰楠 論文種類：碩士論文

系所名稱：企業管理系

指導教授：李杭

共同指導教授：陳俊豪

茲為保障著作人著作權益，並就論文著作權之歸屬及事後權利行使方式，包括論文應如何公開發表、發表時應如何標示著作人姓名、論文事後可作何種修改以及未來應如何授權他人利用等事項，碩、博士生與指導（含共同指導）教授依下列原則達成協議：

- 一、碩、博士生所撰寫之論文，如指導（或共同指導）教授僅為觀念之指導，並未參與內容表達之撰寫，依著作權法規定，學生為該論文之著作人，並於論文完成時，即享有該論文之著作權，指導教授無法於事後主張為共同著作人，亦不得共同掛名為著作人。（著作權法第10條之1）
- 二、如指導（或共同指導）教授不僅為觀念的指導，且參與內容之表達而與學生共同完成論文，且各人之創作，不能分離利用者，則為共同著作，學生與指導教授為論文的共同著作人並共同享有著作權，此等共同著作著作權（包括著作財產權及著作人格權）的行使，即應取得碩、博士生與指導（或共同指導）教授之共同同意後，始得為之。（著作權法第8條、著作權法第40條之1第1項）
- 三、依上述原則，本論文之著作權歸屬：
 - 研究生單獨擁有。
 - 研究生與指導教授共同擁有。
 - 研究生、指導教授及共同指導教授共同擁有。

研究生： 阮泰楠 日期：112年9月6日

指導教授： 李杭 日期：112年9月6日

共同指導教授： 陳俊豪 日期：112年9月7日

多來源網頁資料對股市投資影響之分析： 以台灣股市為例。

學生：阮泰楠
國立高雄科技大學
企業管理系碩士班

指導教授：李杭 博士
國立高雄科技大學
企業管理系 教授

共同指導教授：陳俊豪博士
國立高雄科技大學
資訊工程系 教授

國立高雄科技大學企業管理系碩士班

摘要

近年來，隨著無線網路與行動裝置的迅速發展，投資者愈發容易從多來源網頁資料取得各種資料。雖有眾多的網頁資料來源，但投資者經常受單一來源的資料影響而產生資訊不一致狀況進而誤判趨勢。此外，因為網頁資料的特性複雜且多為非結構化資料，故如何從各種網頁資料來源處理並取得客觀的資訊是一個具挑戰的任務。為解決此問題，本論文提出一個遺傳為基礎的多來源屬性最佳化演算法，旨在從多來源網頁資料找出有效的分類屬性用以建立預測模型分析多來源網頁資料與股市之間的關係及影響。所提方法首先針對三個網頁資料來源，包含：財經新聞、社群媒體與經濟指標，進行資料前處理取得可能的特徵屬性。接著，利用二元編碼方式進行特徵屬性的編碼。在評估函數上，則利用兩個因子計算染色體適合度值，分別為模型準確度與屬性來源多元性。演化後，最佳的分類屬性集合將輸出並用來建立預測模型。運用此預測模型，我們可以提供一個相對客觀的趨勢判斷結果給使用者參考使用。

關鍵詞：多來源網頁資料，情感分析，遺傳演算法，預測模型，自然語言處理。

Analysis of the Impact on Stock Market Investment Using Multi-source Web Data: Taking Taiwan Stock Market as an Example

Student : THAI-NAM NGUYEN

Advisors : Dr. Hang Lee

Co-advisor' s : Chun-Hao Chen

Department of Business Administration
National Kaohsiung University of Science and Technology

ABSTRACT

In recent years, with the rapid development of wireless networks and mobile devices, it has become easier for investors to get various information from multi-source web data. Although there are many sources of web data, investors are often influenced by a single source of data, as a results, inconsistent information and misjudging trends are occurred. In addition, since the characteristics of web data are complex and most of them are unstructured data, how to process and obtain objective information from multi-source web data is a challenging task. In order to solve this problem, this paper proposes a genetic-based multi-source attribute optimization algorithm, which aims to find effective classification attributes from multi-source web data to construct a prediction model to analyze the relationship and influence between multi-source web data and the stock market. The proposed approach first targets three web data sources, including: financial news, social media and economic indicators, and performs data preprocessing to obtain possible classification attributes. Then, the binary encoding schema is utilized to encode those attributes. As to the fitness function, two factors that are accuracy of model and diversity of attribute sources are used to calculate fitness value of a chromosome. After evolution, the best set of classification attributes is output and used to build a prediction model. Using the built prediction model, we expect to provide a relatively objective trend judgment result for users' reference.

Keywords: Multi-source web data, sentiment analysis, genetic algorithms, prediction model, natural language processing.