Investing Mapping 101: Visualization of the Industry through Keywords & Sensitivity Analysis

Daehyun Kim1

*1*Hankuk University of Foreign Studies(s)

Kdh1834@hufs.ac.kr1

# *Abstract*

I was thinking about what information people needed to start investing in stocks, but when I started investing in stocks now, I faced the complexity of the information and the large amount of information in the process of collecting it, and the information was difficult to organize. To solve this problem, we started this project for convenient information access, understanding, and utilization when investing in stocks. In order to easily understand, utilize & access complex stock information and data, we have developed an integrated platform that can combine both network graphs and vast amounts of information in the business sector into a single dashboard. By developing this platform, domestic individual investors find promising investment areas by receiving optimized information, reduce investment risk by helping consumers make effective investment decisions, and contribute to investors' understanding of market stocks.

***Keywords:*** *Investing, Stock, Sensibility Analysis, Trend, Mapping,*

1. Introduction

*When I started investing in stocks, I started thinking about what people needed to invest for the first time, and in the process, I found that a lot of information and inaccessible information needed to invest in stocks was difficult for those who started investing in stocks and that there was a lot of information to know. Based on this awareness of the problem, we started the Investment Mapping 101 project. The purpose of the project is to integrate and provide the information needed for investment so that investors can easily understand and utilize it.*

*In general, those who start investing in stocks for the first time spend a lot of time interpreting and analyzing difficult stock information and not knowing how to use it, which is a major drawback for those who invest in stocks for the first time in a modern investment environment where they have to adapt to rapid changes in the stock market. So it was important for us to synthesize the responses of news and communities and calculate their proportions appropriately to accurately analyze and indicate whether or not the keyword is a promising field. And we wanted to show how the actual stock price was affected by public opinion in the field through correlation analysis*[1].

*The development goal was to get investors to know information at a glance. By dividing into nine major industries and using network graphs to visually express keywords within a single business area and providing various information about the business area in the form of a dashboard, investors can access the information they need without spending much time studying and gathering information. We believe that these platforms can help investors understand market trends and understand the flow of public opinion about the market through emotional analysis.*

**2. Methodology**



**Figure 1**.Project implementation details

2.1. Data collection & preprocessing

News and community data collect: We collected about 500,000 crore stock-related data from korea’s major community sites such as Bboomboo, Naver Café, Eppem Korea and Naver Stock Investment site. It has secured an additional 1.85 million stock-related news data for the year on a site called Big Cainz[2].

Collects corporate classification data: The data collected by webcrolling corporate-related data from the corporate disclosure channel KIND and Naver Stock sites is used to classify and analyze companies[3].

Remove special characters and URLs: Improve data quality by removing meaningless special characters and URLs from the collected data.

Remove missing and duplicated values: Missing values and duplicated values have been removed to ensure the accuracy of the data.

2.2. Class Classification

Keyword extraction using TD-IDF and n-gram: In order to establish a classification system for business sectors, key keywords were extracted using TD-IDF from the company overview, and keyphrases were generated using n-gram, and clustering was carried out in the process [4].In addition, the generated keyphrases were referred to the classification of GICS and WICS, and categories were selected using GPT.

Subsequently, we established a taxonomy and categorized the enterprise overview into up to four levels, large/medium/small/minimum, by sector using GPT.

2.3. Building an emotional analysis model & keyword emotional analysis

Fine-tuning and performance evaluation for each model: KoBERT, KcELECTRA, and KoELECTRA models were fine-tuned with financial sentence datasets to evaluate the accuracy of each model. Among them, KoBERT showed the best performance, so we finally chose it [5][6].

Table 1. Keyword Sensitivity Analysis Accuracy(%)

|  |  |
| --- | --- |
| **Method** | **Accuracy** |
| KoBERT | 0.8870 |
| KcELECTRA | 0.8685 |
| KoELECTRA | 0.8230 |

Emotional analysis using the KoBERT model: We constructed a positive emotional analysis model for news and community data using fine-tuned KoBERT model and analyzed emotions.

Analyzing the amount of comments and positive rates: We also identified the amount of comments made by each keyword, and analyzed the percentage of positive public opinion based on sentiment analysis data to identify current trends.

2.4. Structured data analysis

Financial statement data and yield data were analyzed to simplify the relevance and each key characteristic.

2.5. Visualization and user interface

Web page creation: We created web pages using React and JavaScript, and we implemented web pages by visualizing them with easy UI considering design and composition using Figma.

Data visualization: Full network graphs were created using D3 libraries, time series, and radar graphs used Chart.js. All data were visualized by converting to JSON format[7].

Implementation of graphs: A dynamic and interactive graph was implemented using the React-d3 library. Both node and edge data are entered in JSON format, and all information is managed through a back\_end built into Express.

**3. Result**

*3.1 Development Results*

*Our Mapping 101 provides differentiated features from existing financial sites. Key features include the ability to graphically visualize monthly positive indices, the ability to statize and provide various financial indicators, the ability to analyze companies related to specific keywords based on ESG scores and market capitalization*[8], *and the ability to compare similar price trend industries with the same and different themes. In addition, users can explore related topics through interactive node selection, and in-depth information can be obtained through positive indices and radial data for each keyword. All of these features provide users with a rich, multidimensional experience of analyzing financial data.*

A screenshot of a computer

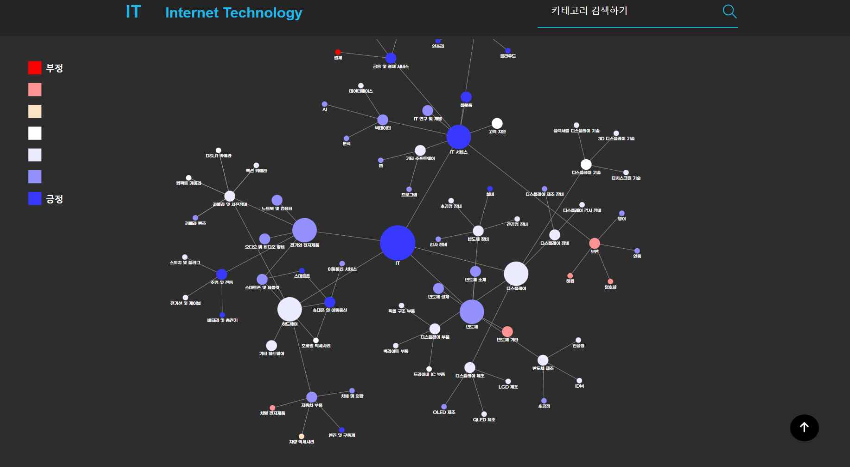
Description automatically generated

Figure 2. Positive, Excess Profit Persistence, Figure 3. Network Graph

ESG Score By industry

***3.2*** *Positive Index Reliability*

*I once drew a graph of the reliability of the positive index to express the accuracy of the service. As you can see at a glance, the graph looks similar. Blue shows our positive index graph, and short blue shows the ETF returns of keywords. We can see that the positive index precedes the yield index a little more, and as the keyword changes positively, the yield also rises subsequently. Through this graph, we can see that our positive index precedes the yield indicator in a large direction.*

***A graph with lines and numbers

Description automatically generated*** ***A graph with lines and numbers

Description automatically generated***

Figure . Network and Infrastructure Figure 5. Financial & Payment Service Positive Index Reliability

***4. Conclusion & Discussion***

*Investment mapping 101 provides optimized information for users to easily find the key information they need from complex data. The service visualizes monthly positive indices and management indicators to accurately assess the growth potential and stability of stocks and offers a variety of investment options through recommendations for stocks with similar price trends. With the latest trends and market trend analysis, users can respond quickly to market changes and make active investment decisions. By reducing potential investment risks and increasing users' understanding of the market and stock, we expect this service to help support professional investment activities and improve the investment performance of individual investors.*

*However, there were some regrets that due to the limited project period, various models could not be tested, data collection was limited, and the team members who participated in the project had experience in data pre-processing, clustering, and natural language processing, but there were opinions that the implementation was slightly insufficient in terms of visualization because there were no people who had experience implementing the web, front, UX, and UI. In the future, I think these problems should be solved by the existing personnel developing more and supplementing them or by recruiting new personnel.*

**Acknowledgement**

*I would like to thank the team members of Hoyoung, Hyeongjun, Chanhyuk, Dongkyu, and Wonseok who participated in this project, and I sincerely thank the officials of Hankuk University of Foreign Studies, including Professor Oh Se-jong, Professor Park Min-guk of the AI Institute of Education, researchers who came to give lectures, professors, and Manager Kim Hyun-jin who gave feedback.*

**References**

1. Kang Jang-gu, Kwon Kyung-yoon, and Shim Myung-hwa. " Investor sentiment and stock returns of individual investors." Financial Management Research 3 (2013): 35-68.
2. Kim Yu-shin, Kim Nam-gyu, and Jung Seung-ryul. "News and Stock Price: The Decision Model of Intelligent Investment through Sensitivity Analysis of Big Data." Intelligent Information Research, Korea Intelligent Information System Society 2 (2012).
3. Kim Young-min, Jung Seok-jae, and Lee Seok-jun. "A study on the prediction of stock price fluctuations through social media sentiment analysis." Entrule Journal of Information Technology 13.3 (2014): 59-69.
4. Ko Jae-chang, Cho Geun-tae, and Cho Yun-ho. "Recent research trends in technology management examined through keyword network analysis." Intelligent Information Research 2 (2013): 101-123.
5. Cho Su-ji, Kim Heung-gyu, and Yang Cheol-won. "Building a Korean-language emotional dictionary for corporate financial analysis." Journal of the Korea Securities Society 2 (2021): 135-170.
6. Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
7. Ahn Seong-won and Cho Sung-bae. " stock price prediction using news text mining and time series analysis." Korean Society of Information Sciences Academic Presentation Papers 1C (2010): 364-369.
8. Sung Tae-eung, et al. "Study on an intelligent value chain network system based on corporate information." Intelligent Information Research 3 (2018): 67-88.