**Frontiers in Business, Economics and Management**
ISSN: 2766-824X | Vol. 16, No. 2, 2024

# Research on Behavioral Finance on the Platform of Financial Technology: An Exploration of Investors' Emotional Indicators Based on Text Analysis

**Chenyu He**

University of California, Irvine 92612, The United States

**Abstract:** This study discusses the behavioral financial analysis of investor sentiment on the financial technology platform, and uses text analysis technology to construct investor sentiment indicators. With the rapid development of financial technology, financial technology platform not only provides convenient trading channels, but also enriches market information through technologies such as big data and cloud computing. As an important factor affecting the financial market, investor sentiment has become the focus of research on the financial technology platform. This study collects text data from financial technology platform, uses natural language processing technology to extract investors' emotional tendency, and constructs effective emotional indicators. The research results show that the constructed investor sentiment index can accurately reflect the changes of investor sentiment, and has a certain correlation with the market trend. In addition, the index also shows the forward-looking in market forecasting, which provides a new method for monitoring and forecasting the market sentiment of the financial technology platform. This study not only deepens the understanding of investors' behavior patterns and psychological dynamics, but also provides a new perspective and tool for financial market forecasting and decision-making.

**Keywords:** Behavioral Finance, Financial Technology, Investors' Emotional Indicators, Text Analysis.

## 1. Introduction

With the rapid development of financial technology, the traditional financial service model is undergoing profound changes. Financial technology platform, as an important carrier of this change, not only provides investors with more convenient and efficient trading channels, but also greatly enriches the information sources and analysis methods of financial markets through advanced technologies such as big data and cloud computing [1-2]. However, these technological changes not only changed the trading behavior of investors, but also had a far-reaching impact on investors' psychology.

Among many factors that affect the financial market, investor sentiment is an important factor that cannot be ignored. It reflects the expectations and confidence of market participants in the future market, and then affects the market trend [3]. On the platform of financial technology, this kind of emotion is often reflected by investors' speeches and trading behaviors. Therefore, how to effectively capture and analyze this emotional information has become an important research topic in the field of financial technology.

The purpose of this study is to explore the emotional indicators of investors on the financial technology platform through text analysis. Based on a large number of text data collected on the platform, this paper uses natural language processing and other technologies to extract and analyze investors' emotional tendencies, and tries to construct effective emotional indicators. This not only helps to understand investors' behavior patterns and psychological dynamics, but also provides new perspectives and tools for financial market forecasting and decision-making. Therefore, this study has important theoretical and practical significance.

## 2. Research Method

### 2.1. Data Source

In order to deeply explore the investor's sentiment on the financial technology platform, this study chooses a representative financial technology platform, which brings together a large number of investors who have in-depth research and enthusiasm for the financial market, and they actively discuss market trends, investment strategies and personal views on the market in the forum. Therefore, the text data on this platform can provide rich investor's sentiment information for this study.

From the platform, the discussion posts published by investors in the forum are captured, which contain investors' views on the current market, some stocks or financial products, and their investment strategies. These contents are an important source of information for analyzing investors' emotions. In addition to the main post, I also collected comments and replies below the post. These comments and replies often contain the approval, opposition or supplement to the main post, which can more comprehensively reflect investors' emotions and attitudes. Some users will share their investment experiences and feelings of market dynamics in their personal dynamics, which is also a window to understand investors' emotions.

### 2.2. Text Preprocessing

After collecting the text data on the financial technology platform, in order to ensure the data quality and the accuracy of subsequent analysis, text preprocessing is needed. Firstly, remove the content unrelated to investor sentiment analysis, such as advertisements, non-text content (pictures, video links, etc.), duplicate data and malformed text [4-5]. In addition, typos and spelling mistakes in the text are corrected.

Word segmentation is to divide a continuous text into independent lexical units. Because there is no obvious separator between words in Chinese sentences, word segmentation is a key step in Chinese text processing. Use the professional Chinese word segmentation tool jieba Word Segmentation, which can accurately segment words according to Chinese grammar rules and thesaurus.

Stop words refers to words that frequently appear in the text but make little contribution to the meaning of the text, such as "de", "Shi" and "zai". These words often cause noise in text analysis, so they need to be removed. A stop words list is constructed, and stop words in the text is deleted in the preprocessing process [6].

## 2.3. Text Analysis

When conducting text analysis, natural language processing techniques are used to extract features from the text. For the purpose of this study, which is to explore sentiment indicators of investors on financial technology platforms, the paper designs a text analysis method based on the BoW (Bag of Words) model and TF-IDF (Term Frequency-Inverse Document Frequency) [7-8].

Firstly, the preprocessed text data is converted into BoW representation. BoW regards text as a collection of words, ignoring grammar and the order of words, but retaining the frequency information of words. Build a vocabulary and map each word in the text to a unique index. Then, for each text, count the number of occurrences of each word in the vocabulary to form a lexical frequency vector.

Suppose that the vocabulary contains the following words: {buy, stock, feeling, good, hope and rise}, and the TF and IDF values of each word have been calculated. For the text "I bought XX shares today, and I feel good. I hope I can go up a bit!" According to the TF-IDF formula, the TF-IDF value of each word is calculated, and a feature vector is obtained.

TF calculation formula:

$$TF(t,d)=\frac{f_{t,d}}{\sum_{t^{'}\in d} f_{t^{'},d}} \tag{1}$$

Where $f_{t,d}$ represents the number of occurrences of the term $t$ in the document $d$, and the denominator represents the sum of the occurrences of all the terms in the document $d$.

DF calculation formula:

$$IDF(t)=\log_e \frac{N}{df(t)} \tag{2}$$

Where $N$ is the total number of documents in the document collection, and $df(t)$ is the number of documents containing the term $t$.

TF-IDF value calculation formula:

$$TF-IDF(t,d)=TF(t,d)\times IDF(t) \tag{3}$$

This value takes into account the frequency of words in a specific document and the rarity in the whole document collection. Calculate the TF-IDF value of each word and represent these values as the characteristics of the text. These characteristics will be used for the subsequent construction of investor sentiment indicators.

## 2.4. Construction of Investor Sentiment Index

After extracting the text features, the next step is to construct investor sentiment indicators. According to the actual situation of this study, supervised learning method is used to construct this index. First, a labeled data set is needed to train the model. This can be achieved by inviting experts to mark a part of text data as positive, negative or neutral. Use the text features extracted by methods such as TF-IDF before. Further, through feature selection techniques (chi-square test, mutual information), the features most related to investor sentiment are screened out.

Considering that investor sentiment classification is a multi-classification problem (positive, negative and neutral), multi-classification algorithms are adopted, such as logistic regression, support vector machine (SVM), random forest or neural network [9-10]. In this study, the logistic regression model is chosen as the baseline model, because it is simple, interpretable and performs well in many text classification tasks.

Suppose there are $m$ training samples $(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)$, where $x_i$ is the text feature vector and $y_i$ is the emotional label (positive, negative or neutral). The logistic regression model estimates the parameter $\theta$ by maximizing the likelihood function:

$$L(\theta)=\prod_{i=1}^{m} P(y_i|x_i;(\theta)) \tag{4}$$

Use the labeled data set to train the selected model. The performance of the model is evaluated by 5-fold cross-validation, and the main evaluation indexes include accuracy, accuracy, recall and F1 score. According to the evaluation results, the model parameters are adjusted to achieve optimal performance.

For the new unlabeled text data, the trained model is used to predict and get the emotional classification (positive, negative or neutral) of each text. Further, the classification results are converted into numerical indicators, and the positive emotions are assigned to 1, the negative emotions to -1, and the neutral emotions to 0. For new text data $x_{new}$, its emotional index $s_{Index}$ can be defined as:

$$s_{Index}(x_{new})=\arg\max_{c\in positive,negative,neutral} P(c|x_{new};\theta) \tag{5}$$

In order to get a finer-grained emotional index, the probability predicted by the model is used as the emotional strength, and the logistic regression model will output the probability of each category. Use the probability value as the emotional intensity $s_{stre}$:

$$s_{stre}(x_{new},c)=P(c|x_{new};\theta) \tag{6}$$

Where $c$ stands for emotional category (positive, negative or neutral).

# 3. Experimental Results and Analysis

The supervised learning method is adopted, and the logistic regression model is specifically selected to classify investors' emotions. After the model is optimized on the training set, it has achieved good performance on the test set, and the accuracy, precision, recall and F1 score all reach a satisfactory level. By applying the model to unlabeled text data, the emotional classification results (positive, negative or neutral) of each text are obtained. Further, these classification results are converted into numerical indicators for quantitative analysis. Specifically, positive emotions are assigned to 1, negative emotions to -1, and neutral emotions to 0. The results of the constructed investor sentiment indicators are shown in Table 1.

**Table 1.** The results of the constructed investor sentiment index

| Text ID | Text content | Predictive emotion classification | Emotional index value |
|---|---|---|---|
| T1 | Today, the market soared and made a lot of money! | frontage | 1 |
| T2 | Why has this stock been falling? | negative side | -1 |
| T3 | I think it is too risky to enter the market now. | negative side | -1 |
| T4 | What will the market trend be like tomorrow? | neutral | 0 |
| T5 | The recent investment income is not bad! | frontage | 1 |
| T6 | It's bad luck to be trapped again. | negative side | -1 |
| T7 | I don't know what stock to buy, and I'm in a dilemma. | neutral | 0 |
| T8 | This investment strategy seems to have failed. | negative side | -1 |
| T9 | Be optimistic about this industry and prepare to hold it for a long time. | frontage | 1 |
| T10 | The stock market is really unpredictable! | neutral | 0 |

In order to analyze the correlation between investor sentiment and market trend, the market index data in the corresponding time period were obtained and compared with investor sentiment indicators. The results show that there is a certain correlation between investor sentiment fluctuation and market trend. When the market sentiment is positive, the market index tends to show an upward trend; On the contrary, when the market sentiment is negative, the market index may fall (Figure 1).

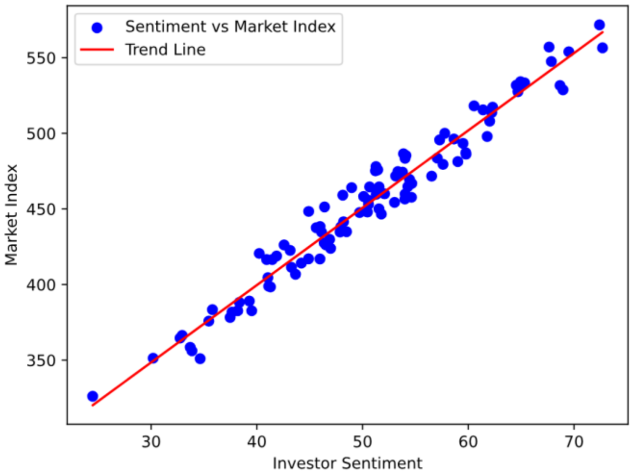


**Figure 1.** Correlation between investor sentiment and market trend

The influence of investor sentiment on financial technology platform is reflected in many aspects. First of all, the high or low mood will directly affect the activity and transaction volume of users on the platform. Secondly, changes in mood may also lead to market fluctuations, which in turn will affect the stability and revenue of the platform (Table 2).

**Table 2.** The influence of investor sentiment on financial technology platform

| Emotional state | User activity (daily average) | Transaction volume (average daily transaction volume) | Number of new user registrations (daily average) | User retention rate (weeks) |
|---|---|---|---|---|
| frontage | 150,000 | $5,000,000 | 3,000 | 75% |
| neutral | 100,000 | $3,000,000 | 2,000 | 65% |
| negative side | 70,000 | $1,500,000 | 1,000 | 55% |

In the aspect of forecasting ability, the investor sentiment index constructed by the research shows a certain foresight. Through the analysis of historical data, it is found that when the investor's emotional indicators remain at a high or low level, the market will tend to rise or fall in a corresponding period of time. This provides a certain reference for the short-term forecast of the market.

In order to evaluate the advantages and limitations of the indicators constructed in this study, it is compared with several other common investor sentiment indicators, including sentiment analysis indicators based on social media and sentiment indicators based on market transaction data. The results show that the indicators constructed in this study have high sensitivity and accuracy in capturing investors' emotional changes. Compared with other indicators, it can reflect the change of market sentiment more timely and provide more useful information for market forecast. However, this indicator also has some limitations, such as high dependence on text data, which may be affected by text quality and labeling accuracy (Table 3).

**Table 3.** Comparison with other investor sentiment indicators

| Indicator type | Sensitivity (%) | Accuracy (%) | Reaction time (hours) | data source |
|---|---|---|---|---|
| The construction index of this study | 85 | 90 | 1 | text data |
| Emotional analysis of social media | 70 | 75 | 3 | social media |
| Emotional indicators of market transaction data | 60 | 80 | 6 | Market data |

To sum up, this study provides an effective method for monitoring and forecasting market sentiment on the financial technology platform by constructing investor sentiment indicators and analyzing their correlation with market trends. In the future, we will further optimize the model algorithm and improve the data quality to improve the accuracy and forecasting ability of the indicators.

## 4. Conclusion

Through in-depth analysis of the text data released by investors on the financial technology platform, this study successfully constructed an investor sentiment index based on text analysis. This index can not only effectively capture investors' emotional tendencies, but also provide a new perspective for understanding investors' behavior patterns and psychological dynamics. In addition, the results show that there is a significant correlation between investor sentiment and market trend. When investor sentiment is high, the market tends to show an upward trend, and vice versa. This discovery provides a new basis and tool for financial market forecasting and decision-making. However, this study also has some limitations. First of all, because the index is highly dependent on text data, it may be affected by text quality and annotation accuracy. Secondly, this study only collects data from a single financial technology platform, so the results may be biased. Future research can further optimize the model algorithm, improve the data quality, and expand the scope of data collection, so as to further improve the accuracy and prediction ability of indicators.

# References


[1] Sun, W., Zhao, C., Wang, Y., & Cho, C. H. (2018). Corporate social responsibility disclosure and catering to investor sentiment in china. Management Decision, 56(9), 1917-1935.

[2] Liu, Z., Jiang, X., Zhu, P., Li, X., & Zhu, M. X. (2023). The influences of livestreaming on online purchase intention: examining platform characteristics and consumer psychology. Industrial Management & Data Systems, 123(3), 862-885.

[3] Anderson, A. G., & Larkin, Y. (2019). Does noninformative text affect investor behavior. Financial management, 48(1), 257-289.

[4] Zhang, W., Xu, Y., & Zheng, H. (2019). The antecedents and consequences of crowdfunding investors' citizenship behaviors: an empirical study of motivations and stickiness. Online Information Review, 43(4), 584-599.

[5] Bernard, D., Cade, N. L., & Hodge, F. (2018). Investor behavior and the benefits of direct stock ownership. Journal of Accounting Research, 56(2), 431-466.

[6] Frydman, C., & Wang, B. (2020). The impact of salience on investor behavior: evidence from a natural experiment. The Journal of Finance, 75(1), 229-276.

[7] Davis, L. E. (2018). Financialization and the non-financial corporation: an investigation of firm-level investment behavior in the united states. Metroeconomica, 69(1), 270-307.

[8] Wang, D., Huang, Q., Ye, T., & Tian, S. (2021). Research on the two-way time-varying relationship between foreign direct investment and financial development based on functional data analysis. Sustainability, 13(11), 6033.

[9] Hsiao, H. F., Zhong, T., & Dincer, H. (2019). Analysing managers' financial motivation for sustainable investment strategies. Sustainability, 11(14), 3849.

[10] Yin, X., Hai, B. L., & Chen, J. (2019). Financial constraints and r&d investment: the moderating role of ceo characteristics. Sustainability, 11(15), 4153.