

Identification of Implicit Collusion of Investors in Financial Markets by Text Mining Tools

¹Kostyantyn Anatolievich Malysenko, ²Majid Mohammad Shafiee, ¹Vadim Anatolievich Malysenko, ¹Marina Viktorovna Malysenko and ¹Dmitriy Viktorovich Anashkin
¹V.I. Vernadsky Crimean Federal University, Vernadsky Ave, Simferopol, Crimea
²Department of Management, University of Isfahan, Isfahan, Iran

ABSTRACT

Background and Objective: The digitalization of the securities market provides participants of financial markets with new opportunities, which require specific methods to monitor market safety and identify violations. The purpose of this paper was to develop a methodology for identifying implicit collusion of financial market participants based on Text Mining tools. **Materials and Methods:** Concurrent with the traditional comparison of data on the trading volumes and price of individual securities, which let us visualize abnormal dynamics, the extracting sentiment method was utilized to accomplish this objective. The methodology was tested in the form of a case study on the shares of a Russian company currently in circulation. **Results:** The results demonstrated five main combinations of measures, which describe specific exchange situations, based on which it is likely to identify possible collusion of stock market participants. The classification of the information field is presented to determine the type of information that has the greatest impact on the dynamics of the course. **Conclusion:** The proposed methodology solves the problem of identifying investor collusion in financial markets. The scientific significance lies in the development of a new approach to identifying collusion between investors and other participants based on trading data and news flow, through their processing using Text Mining tools.

KEYWORDS

Text-mining, Big Data, manipulation on the stock market, implicit collusion of investors, securities market, financial market

Copyright © 2023 Malysenko et al. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

The development of technology provides new opportunities for participants in financial markets¹. In particular, the digitalization of the securities market and the active introduction of innovative technologies, make it possible to place automated orders, practically withdrawing from personal participation in the course of trading. On the one hand, such innovations are a powerful mechanism for the development of the market, allowing it to expand by attracting investors. On the other hand, new technologies require the development of specific approaches to monitoring the market for the safety of investors. Methods and technologies for detecting certain types of fraud that have worked effectively in the past today are no longer able to identify violations.



The work of many foreign researchers is devoted to the study of the problems of unfair practices, such as manipulation and collusion of investors in the securities market. For example, Biggerstaff *et al.*² in their work, explore the behavior patterns of corporate insiders in the process of trading stocks. The authors have established a stable relationship between their obtaining anomalous profitability, which can be traced in the analysis of their purchases and sales. The studied trading models allow us to assess their impact on the overall course of trading in shares. Kim *et al.*³ and Li *et al.*⁴ also studied the influence of insider behavior on the overall structure of the market. The authors confirm that investors teaming up with insiders will receive higher returns on the trading results in the future. The authors agree that regulation of insider trading helps to improve price efficiency. In the work of Esen *et al.*⁵, also raises the question of the need to detect transactions involving insiders. Among the works of domestic authors, one can note the study of Vladimirovich⁶, Aggarwal and Wu⁷, Ivanitskiy and Tatyannikov⁸, Zagoruiko and Viktorovna⁹.

Okiro and Otiso¹⁰ consider the currently existing methods of fraud in the stock market. Nikolaevich and Igorevna¹¹ examine methods of combating unfair practices in the securities markets, including manipulation, which has increased as a result of the active development of algorithmic trading. The author analyzes foreign experience and assesses the possibility of foreign experience application in domestic practice.

Thus, the study of unfair practices and methods of their detection is relevant to the current stage of the development of financial markets. Thus, one of the serious threats to the financial security of participants in the securities market is the collusion of investors. By uniting in groups, small investors can influence the quotations of securities by submitting orders corresponding to their needs. This practice is manipulative and illegal. As a result, a group of investors in collusion makes a profit that cannot be obtained by natural price formation. In this case, the rest of the bidders incur losses.

The purpose of this study is to explore the impact of collusion among investors on the securities market and to identify effective methods for detecting and preventing such fraudulent practices. With the increasing use of innovative technologies in financial markets, it is essential to develop specific approaches for monitoring the market to ensure the safety of investors. This study aims to contribute to the development of these approaches by examining the nature and extent of collusion among investors and identifying effective strategies for detecting and preventing such practices.

MATERIALS AND METHODS

The main field of this study is investigating the dark side of the financial market, such as the impact of new technologies and trends in investors' behaviors. Investor collusion is often organized through the media. Based on this, price manipulation can be carried out by injecting information through various information channels (information manipulation), or by a technical method (by placing massive orders to form a certain trend).

Information influencing the quotations of securities can be divided into three groups:

- Official data is information published by leading news agencies, state media, confirmed information of enterprises, organizations and firms, officially published on their websites, official statements of authorized persons and other verified, confirmed and officially announced information
- Expert opinions and analytics-forecasts, assumptions, analysis and assessment of experts regarding a particular economic situation, area, events, etc., published in the media
- Rumors-this category includes unofficial information, investors' expectations and information published in chat rooms

Using the methods of information and technical manipulation in combination or separately, investors in the collusion can plan and ensure the receipt of profits for its participants. At the moment, determining the fact of manipulation and its type as a result of collusion of investors is a rather difficult task that requires certain time expenditures. This does not allow a timely response to violations and, accordingly, delays the process of proceedings and determination of liability within the framework of identified violations, which negatively affects the operation of the market since high risks make it unattractive for investors^{12,13}. In this regard, one of the primary tasks today is to find effective methods for identifying collusion of investors, taking into account new practices arising in the process of digitalization of the market infrastructure⁶.

In this paper, which is conducted from November, 2022 to February, 2023 in Russia, a methodology for identifying manipulative activities of investors based on Big Data technologies was proposed¹⁴. Big Data is a series of approaches, tools and methods for processing structured and unstructured data of very large volumes and significant diversity to obtain human-readable results. Such data is efficiently processed using scalable software tools that emerged in the late 2000s and became an alternative to traditional databases. In Russia, Big Data also means processing technologies and in the world-only the object of research itself. Big data analysis provides new, unknown information¹⁵.

In the study, Text Mining technology was used as one of the Big Data technologies that allow us to detect implicit anomalous values in large arrays of text data. The general scheme of the approach proposed by the authors to identify manipulative activity is shown in Fig. 1.

As you can see, for the analysis several sources of information are used. The first (block "1") is an array of data on the main groups of influencing information identified above, namely. As an object of research, the official news on specific security (Fig. 1 block 1.1), expert opinions and analytics (1.2) in the form of text messages were taken. This information is available on websites providing information on financial markets. In this case, it's on Investing.com. The category of rumors, as already mentioned, is any unconfirmed information. Therefore, the formation of the basis for the analysis of this information flow for specific securities was taken from the corresponding section "Forum" on this site, where market participants have the opportunity to communicate and exchange information regarding the asset¹⁶⁻²⁰.

Further, for the analysis was taken data on the shares of the Russian company M. Video (MVID). The specified company issued shares, quotations for which are included in the calculation of the RTS index. After the formation of the base of text messages by types of influencing information, it is necessary to conduct a sentiment analysis using Text Mining tools. To achieve this goal, the corresponding tools of the Orange program, which is in the public domain, were used. Programs of this type are currently not represented by domestic developers, respectively, they have mainly an English interface, as well as databases (dictionaries of sentiment) necessary for analysis in foreign languages. In this regard, before loading the generated information arrays for sentiment analysis, it is necessary to prepare them-to translate them into English.

The program algorithm calculates the sentiment value for each text message in the array. Several news, analytical reviews and several dozen user posts on the forum can be published per day. As a result of processing the initial data through the program, three documents in Excel format were generated, containing the sentiment values for each text message from the original arrays. The algorithm for extracting sentiment using the Orange program can be seen in Fig. 2. For further analysis, it is necessary to reduce the sentiment to a single value for those dates where it is multiple. In other words, get the total sentiment value for a specific date, taking into account the nature of the news. At the same time, it is important that the nature of their influence on the stock price is preserved since it is of key importance

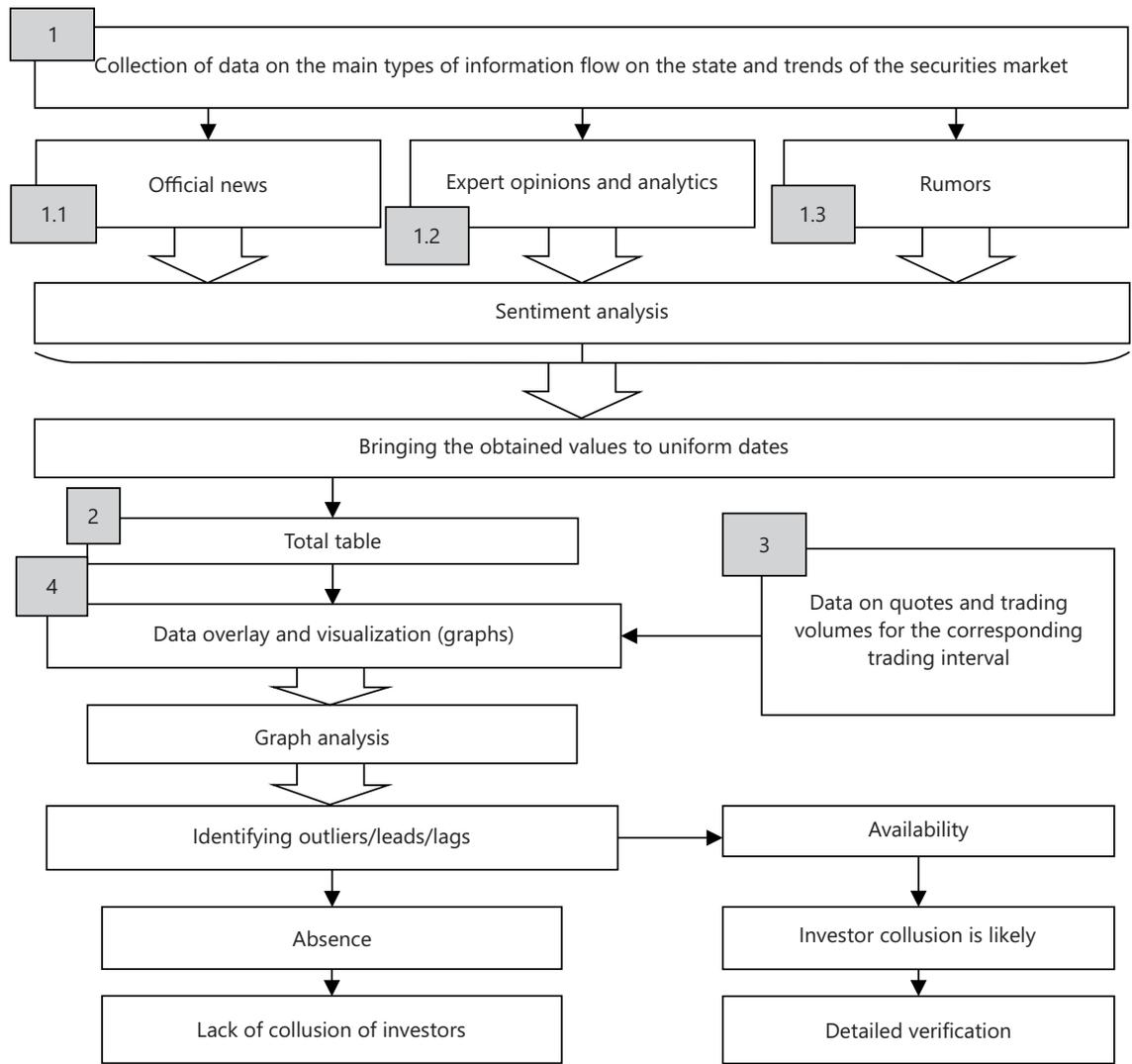


Fig. 1: Procedure for identifying collusion of insiders based on Text Mining technologies

for this study. For this, the resulting files were processed using the SPSS program Ver. 24 (aggregation of messages by date). Sentiment values for specific dates were summed up-this is the optimal method of data processing to comply with the stated condition. The resulting values became possible to combine into a single document for further work (block "2" in Fig. 1).

The next stage of the study consists in superimposing the obtained sentiment values (block "2" in Fig. 1) and data on quotations and trading volumes for the investigated shares (block "3" in Fig. 1) for the corresponding time period. In the current study, it is 6 months-from 01/01/2021 to 07/19/2021.

The Orange software package also has a data visualization add-on, which is necessary for this stage of the study. Overlaying quotes, trading volumes and sentiments allows for visual analysis of the data and identifies inconsistencies that indicate collusion among investors. The visual analysis consists of identifying outliers (sharply differing values of quotes, volumes and sentiment), advancing or lagging their dynamics to the price chart.

Data visualization using the Line Chart tool allowed us to analyze the correspondence of the movement of the volume and price charts to the values of sentiment. Thus, in the absence of unfair practices in the securities market, the change in trading volumes and prices should correspond to the date of the event

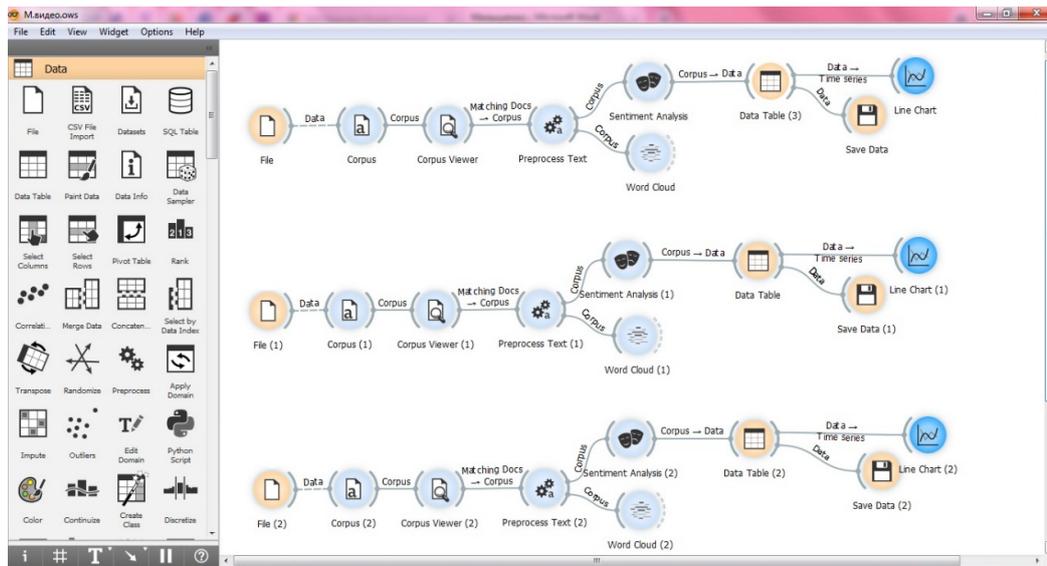


Fig. 2: Result of visualizing the values of quotes, trading volumes and sentiment on the news of the shares of M. video

(the value of sentiment), or be observed after it. If the change in trading volumes and prices precedes the release of the sentiment value, it can be discussed whether unfair practices are present in this event window.

RESULTS

Based on the results of the analysis, five combinations of the indicators can be distinguished, which characterize different exchange situations, respectively:

- The calculated value of a sentiment is insignificant, trading volumes do not change and price change is insignificant, or absent-the normal state of the market
- The calculated value of a sentiment is significant, trading volumes change significantly and there is a price change-the normal state of the market
- The calculated value of a sentiment is insignificant, trading volumes change, the price change is not significant or is not observed and their values are ahead of the value of sentiment-insider activity
- The calculated value of a sentiment is insignificant and trading volumes and/or the price change significantly-manipulative activity as a result of collusion of market participants

The calculated value of a sentiment is significant, the trading volumes do not change or are insignificant, the price change is insignificant, or there is no information-the information is underestimated by the bidders.

The result of this stage of the analysis was shown in Fig. 3. Five main windows were identified that have a significant deviation from the original trend line. According to the basics of the formation of quotations (weighted average) of the stock market, a sharp price change is always associated with a change in trading volumes. So, when news is published, investors begin to actively sell or buy shares, trying to protect themselves from losses, or to make a profit. The volume of trading in security at this moment increases and its price falls or increases. Thus, when analyzing data for the identification of collusion of investors, it is necessary to pay attention to the existence of a relationship between the nature of the news (sentiment), trading volume and price.

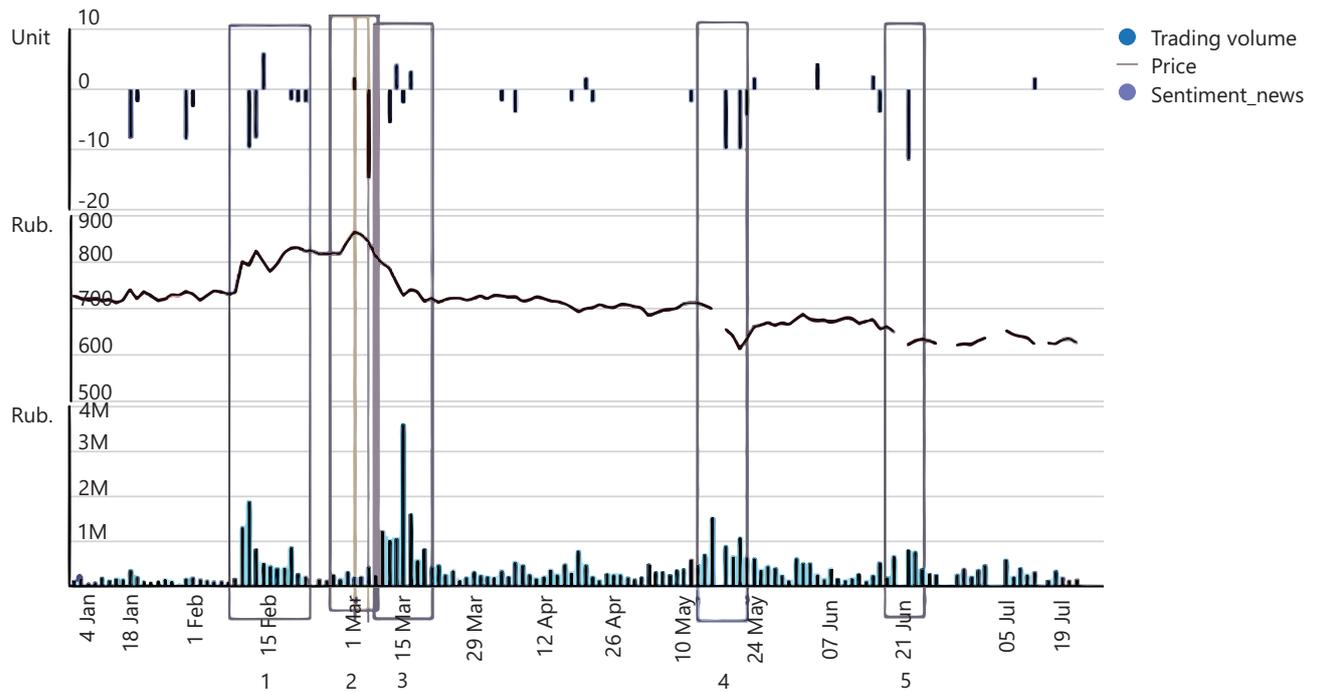


Fig. 3: Result of visualizing the values of quotes, trading volumes and sentiment on the news of the shares of M. video

As follows from Fig. 3, this relationship is preserved in Windows 1, 3, 4 and 5. However, in Window 2, there is a sharp change in the stock price, which is not supported by a change in the trading volume. The sentiment value of a news text message released during this period is also negligible. There is a lag-significant news (a sharp negative value of sentiment) is published later than the price change occurs. This market situation corresponds to the combination number 4, which was presented by us above. This may indicate the existence of collusion among investors and manipulation of the share price.

The visualization of sentiment values by analytics with their imposition on the charts of quotes and trading volumes also demonstrates the presence of inconsistencies in the identified period (Fig. 3, highlighted in a rectangular block). It should be noted that the lag in the market reaction may be acceptable and not exceed three days. This is because exchanges do not trade on weekends. In this case, the news can be published on Friday evening, before the close of trading or after, when investors no longer have the opportunity to react to them. In this case, the reaction to the information will be reflected on the charts as early as Monday. A similar situation was observed with technical manipulation: Investors in collusion place several orders that correct the current rate at the end of the trading session.

In the course of the analysis, it was found that a sharp change in the price of M. video shares occurred in the period from 01.03 (Monday) to 05.03 (Friday) 2021, while there was no corresponding change in trading volumes. According to the sentiment analysis, there is also no publication of significant information in the period preceding or corresponding to this change, which can be seen in the graphs (Fig. 3, 4). Thus, an anomaly is observed for this stock in the period from 26.02 to 05.03.2021, which indirectly indicates the presence of manipulation and requires a more thorough check.

This study examined the impact of categories such as official data and expert opinions and analytics on the formation of stock prices. To check the results obtained, at the next stage, the relationship between the dynamics of the prices of the stocks under study and the values of sentiments of each studied category were analyzed.



Fig. 4: Result of visualizing the values of quotes, trading volumes and sentiment based on the analytics of the shares of M. video

The "rumors" category has the greatest correlation with the course dynamics, compared with official data and analytics. However, the study essentially showed that this type of information flow of the market is more likely a consequence of changes in quotations than their determining factor, i.e. the information published by investors in the forum is their emotional reaction to the course change. This is also confirmed by the high value of the correlation coefficient between the values of sentiment on the news and the forum: it is 0.629.

In this regard, it was concluded that the category "rumors" in this study do not have a significant impact on the change in price dynamics, but it can be of key importance in the study of intraday fluctuations. This hypothesis will be tested by the authors in subsequent studies.

The values of the correlation between quotes and sentiment on news and analytics are also significant and range from 0.55 to 1. The exception is the period from 02.02 to 15.04.2021-in this interval, the correlation coefficient between the values of quotes and sentiment on news varies from -0.14 to -0.30, which indicates that there is practically no connection between these indicators.

The effectiveness of the presented methodology was also confirmed by the fact that in April, 2021 the Central Bank facts of manipulation of several companies. Among them: Are Digital Invest, Region Invest, Stroyzhilinvest, O1 Group Finance, O1 Properties Finance, Prime Finance, Vale Finance, Archer Finance, Moscow Credit Bank, RussNeft, M. video, SAFMAR Financial investments, etc. In total, the manipulation of 34 financial instruments was revealed. The manipulation consisted of "the systematic execution of a series of mutual transactions on the accounts of stable pairs of counterparties-individuals through the highly coordinated filing of counter orders: With identical prices and comparable volumes with a small difference in the time of their issuance, including less than a second. The share of such transactions in the daily trading volume often reached 100%, in several cases, transactions led to significant deviations in the

parameters of trading in financial instruments". Thus, as previously determined, in this case, there is a technical manipulation of the stock price due to the collusion of market participants (traders who speculate on the news).

DISCUSSION

In the current study, a new method has been developed that allows researchers and practitioners to identify the influence of insider trading and manipulation using data from various information flows. The main feature of the technique is that it is based on the use of Text Mining tools, which allow the processing of large arrays of text data and highlighting implicit values-to determine the sentiment of the text (sentiment). The introduced technique makes it possible to automate the process of identifying unfair practices and collusion in the securities market. This would significantly speed up the process of detecting collusion, further investigation and application of appropriate measures. Revealing of implicit collusion of investors in the securities market is carried out based on superposition in the program of the obtained values of sentiment for the three selected groups and trading volumes on the values of quotations, as a result of which it becomes possible to identify outliers, lags or advances. As a result of the correlation analysis, it was determined that the sentiment values normally have a fairly strong relationship with the quotes values. However, its low values signal the presence of manipulation. This distinguishes the current study from previous ones (such as Esen *et al.*⁵) where such a separation is not made. The very use of separated information flows to analyze their impact on the securities market allows us to work on identifying participants engaged in manipulation and insider trading. The data obtained as a result of the study indicated that the proposed method can be used as a basis for solving the problems of the influence of insider activity on the receipt of abnormal profits by insiders, which was described in the study of Kim *et al.*³ and Thanitcul and Srinopnikom²¹. This work also confirmed the assumption that insider transactions are made when the prices and trading volumes of stocks do not correspond to information flows, but at the same time are far from the upcoming prices. These conclusions are concurrent with the study of Li *et al.*⁴. Data mining was used to analyze the securities market, which was similar to the studies of Feuerriegel and Pröllochs²² and Maguluri and Ragupathy²³. However, the current work and its results differ from them in that a more convenient method of analysis was used for ordinary users while obtaining results. The current research also combined methods of intellectual analysis with classical methods of detecting insider trading and manipulation, which has not been done in the works of other authors. The results of this study open up new opportunities to study the impact of information on the securities market, including on decision-making by investors. Thus, it is planned to further explore the possibilities of using technologies when detecting relationships between insiders, as well as when manipulating the market through information stuffing. By shedding light on this important issue, this study will help to promote fair and transparent trading practices in financial markets, thereby enhancing investor confidence and contributing to the overall stability and growth of these markets.

The first implication of the current study states that the development of the new method using Text Mining tools allows for the automation of identifying unfair practices and collusion in the securities market. This implies that the process of detecting collusion can be significantly sped up, leading to quicker investigations and the application of appropriate measures. Moreover, by using the introduced technique, it becomes possible to identify outliers, lags, or advances in trading volumes and sentiment values. This helps in revealing implicit collusion and manipulation, thereby enhancing market integrity and reducing the potential for insider trading.

Thus, the study recommends that researchers and practitioners should consider incorporating Text Mining tools into their analysis of the securities market. These tools can help automate the identification of unfair practices and collusion, making the detection process more efficient and effective. The study also highlights the potential of using technologies, such as data mining and information stuffing analysis, to detect relationships between insiders and market manipulation. It is recommended to continue exploring these technological possibilities to improve market surveillance and prevent fraudulent activities.

There were also some limitations in the current study. The effectiveness of the proposed method heavily relies on the quality and reliability of the data used. Limitations may arise if the data sources contain inaccuracies, biases, or incomplete information. It is important to ensure the validity of the data to avoid misleading results. Moreover, the proposed method may have limitations in terms of generalizability. The effectiveness of the technique may vary across different markets, securities, or time periods. It is recommended to conduct further research and validation in various contexts to assess the method's applicability.

CONCLUSION

This paper developed a methodology for detecting collusion in financial markets using Text Mining tools. The significance of this research lies in its ability to detect collusion and improve market safety in the digitalized securities market. Given the findings, it is recommended that researchers and practitioners consider incorporating Text Mining tools into their analysis of the securities market. These tools can greatly enhance the efficiency and effectiveness of detecting collusion and unfair practices. However, future research should address limitations like reliance on unofficial statements and work on improving specialized dictionaries for more accurate analysis.

SIGNIFICANCE STATEMENT

The paper examines the dynamics of quotations of a group of securities traded on the Russian market and their dependence on the news background. The results of the presented work can serve as the basis for further development of Text Mining technologies to analyze the securities market for the presence of collusion of participants for manipulation.

ACKNOWLEDGMENT

This research was supported by the Russian Science Foundation grant No. 23-28-00946, <https://rscf.ru/en/project/23-28-00946/>.

REFERENCES

1. Shafiee, M.M. and S.I. Najafabadi, 2016. The Interaction of Technological Progress and Tourism Industry Development in the Developing Countries: The Case of Iran's Tourism Industry. 10th International Conference on E-Commerce in Developing Countries: With Focus on E-Tourism (ECDC), 15-16 April, IEEE, Isfahan, Iran, 1-5.
2. Biggerstaff, L., D. Cicero and M.B. Wintoki, 2020. Insider trading patterns. *J. Corporate Finance*, Vol. 64. 10.1016/j.jcorpfin.2020.101654.
3. Kim, D., L. Ng, Q. Wang and X. Wang, 2019. Insider trading, informativeness, and price efficiency around the world. *Asia Pac. J. Financ. Stud.*, 48: 727-776.
4. Li, R., X.W. Wang, Z. Yan and Q. Zhang, 2019. Trading against the grain: When insiders buy high and sell low. *J. Portfolio Manage.*, 46: 139-151.
5. Esen, M.F., M. Singal, H.W. Kot and M.H. Chen, 2019. Can insider trading in U.S. hospitality firms predict future returns? *Int. J. Hospitality Manage.*, 83: 115-127.
6. Vladimirovich, P.V., 2019. On the issue of ensuring the economic security of the Russian stock market based on the diagnosis of insider trading [Russian]. *Sci. Rev. Ser. 1: Econ. Law*, 3-4: 69-80.
7. Aggarwal, R.K. and G. Wu, 2006. Stock market manipulations. *J. Bus.*, 79: 1915-1953.
8. Ivanitsky, V.P. and V.A. Tatyannikov, 2018. Information asymmetry in financial markets: Challenges and threats. *Econ. Reg.*, 14: 1156-1167.
9. Yuryevich, Z.I. and E.Z. Viktorovna, 2019. Modern approaches to unfair competition in the securities market [In Russian]. *Moscow Econ. J.*, 11: 320-327.
10. Okiro, K. and D.O. Otiso, 2021. Detection of fraud in financial statements using Beneish ratios for companies listed at Nairobi Securities Exchange. *Afr. Dev. Finance J.*, 5: 92-126.

11. Nikolaevich, V.S. and E.A. Igorevna, 2017. Combating market manipulation in developing countries: The methods used and the possibility of their application in Russia. *J. Financ. Risk Manage.*, 3: 190-201.
12. Malysenko, K.A., M.M. Shafiee, V.A. Malysenko and M.V. Anashkina, 2021. Dynamics of the securities market in the information asymmetry context: Developing a methodology for emerging securities markets. *Global Bus. Econ. Rev.*, 25: 89-114.
13. Bazargan, N.A. and M. Shafiee, 2017. Customers' electronic trust to online stores: A risk reduction approach. *Karafan*, 13: 113-122.
14. Fong, B., 2021. Analysing the behavioural finance impact of 'fake news' phenomena on financial markets: A representative agent model and empirical validation. *Financ. Innovation*, Vol. 7. 10.1186/s40854-021-00271-z.
15. Hung, C.C., 2019. Analysis of information security news content and abnormal returns of enterprises. *Big Data Cognit. Comput.*, Vol. 3. 10.3390/bdcc3020024.
16. Rahimi, M., M.M. Shafiee and A.A. Tadi, 2020. Group AHP in ranking intangible assets: Study of chemical industry. *J. Financ. Manage. Strategy*, 8: 107-116.
17. Rahimi, M., M.M. Shafiee, A. Ansari and M. Botshekan, 2020. A model for the ranking of intangible assets and brand valuation of listed companies in tehran stock exchange. *Iran. J. Trade Stud.*, 24: 1-18.
18. Chang, L.Y.C., S. Mukherjee and N. Coppel, 2021. We are all victims: Questionable content and collective victimisation in the digital age. *Asian J. Criminology*, 16: 37-50.
19. Rahimi, M., M.M. Shafiee, A.A. Tadi and M. Botshekan, 2019. A comparative study of brand valuation with two approaches of earning per share and price to sales in the tiles, ceramics and cement industries. *J. Brand Manage.*, 6: 65-81.
20. Rahimi, M., M.M. Shafiee, A.A. Tadi and M. Betshekan, 2018. Presenting a model to determine the brand value of companies admitted to the Tehran Stock Exchange [Persian]. *J. Bus. Manage. Perspect.*, 17: 71-90.
21. Thanitcul, S. and T. Srinopnikom, 2019. Monetary penalties: An empirical study on the enforcement of Thai insider trading sanctions. *Kasetsart J. Social Sci.*, 40: 635-641.
22. Feuerriegel, S. and N. Pröllochs, 2021. Investor reaction to financial disclosures across topics: An application of latent dirichlet allocation. *Decis. Sci.*, 52: 608-628.
23. Maguluri, L.P. and R. Ragupathy, 2019. A new sentiment score based improved Bayesian networks for real-time intraday stock trend classification. *Int. J. Adv. Trends Comput. Sci. Eng.*, 8: 1045-1055.