



## Investigating the Impact of Trading Signals Published in Telegram Channels on Capital Market Trading Value Using Machine Learning Algorithms

**Mohammad Ebrahim Raei Ezabadi (Corresponding Author)** 

Assistant Prof., Department of Finance and Accounting, ST.C Branch, Islamic Azad University, Tehran, Iran. E-mail: meraei68@iau.ac.ir

**Seyyed Hossein Alavi** 

MSc., Department of Finance Management, ST.C. Branch, Islamic Azad University, Tehran, Iran. E-mail: s.hossein.alavi@gmail.com

**Ali Kianifar** 

Ph.D. Candidate, Department of Finance, ST.C. Branch, Islamic Azad University, Tehran, Iran. E-mail: ali.kianifar@iau.ac.ir

### Abstract

#### Objective

The expansion of social networks and messaging platforms in recent years has fundamentally transformed the structure of information circulation in financial markets. Due to its high sensitivity to information, expectations, and collective investor behavior, the capital market is among the most affected domains by this transformation. In this context, Telegram, as one of the most widely used platforms among Iranian capital market participants, plays a significant role in the dissemination of analyses, news, rumors, and trading signals. Many retail investors, alongside fundamental and technical analyses, rely on the content published on Telegram channels when making decisions about buying or selling stocks. Therefore, an empirical and scientific examination of the relationship between trading signals published in these channels and actual market behavior can contribute to a deeper understanding of how social media influence market dynamics. The main objective of this study is to investigate the relationship between the number of trading signals published in stock-market-related Telegram channels and the daily trading value of the Iranian capital

**Citation:** Raei Ezabadi, Mohammad Ebrahim; Alavi, Seyyed Hossein & Kianifar, Ali (2026). Investigating the impact of trading signals published in Telegram channels on capital market trading value using machine learning algorithms. *Media Management Review*, 5(1), 31-54. (in Persian)



market. The key innovation of this research lies in its focus on “trading signals” rather than general news, its use of Persian-language data from Telegram, and its examination of “daily trading value” as an indicator of market dynamism, whereas most previous studies have primarily focused on price, return, or the overall market index.

### **Research Methodology**

This study is applied in terms of purpose and descriptive-analytical and correlational in terms of method. It was conducted using data mining, text mining, and machine learning approaches. The research data consists of two main categories: first, textual data extracted from Telegram channels active in the field of the stock market; and second, numerical data related to the daily trading value of the stock market. The textual dataset includes messages published in stock-market-related Telegram channels during the year 1401 in the Iranian calendar. The selected sample consisted of channels that regularly published analyses and buy/sell recommendations for stocks. In the numerical section, the study used the daily value of retail stock and pre-emptive rights transactions in the Tehran Stock Exchange and Iran Fara Bourse, excluding block trades, over 237 trading days in 1401. The textual data were extracted using crawlers and automated data collection tools. In total, more than 3.5 million messages were included in the analysis process. Among these data, approximately 5,000 messages were manually labeled by an individual familiar with capital market terminology. The messages were classified into three categories: “buy signal,” “sell signal,” and “neutral/news related.” The main criterion for labeling was the presence of an explicit or implicit recommendation to buy or sell a stock in the text of the message.

In the preprocessing stage, tokenization, noise removal, deletion of unnecessary characters, letter normalization, stop-word removal, and stemming were performed on the texts so that raw social media data could be transformed into standardized and processable data. Subsequently, numerical representation of the texts was carried out using methods such as Bag-of-Words, TF-IDF, and transformer-based models, including BERT. In the modeling stage, supervised machine learning algorithms, including Naive Bayes, Support Vector Machine, and a BERT-based model, were used to classify the messages. The labeled data were divided into training, validation, and test sets. After model evaluation, the trained model was applied to classify all extracted messages. Then, the daily number of buy signals, sell signals, and neutral messages was calculated and matched with daily trading value data. Pearson’s correlation coefficient was used to examine the relationship between the variables. In addition, the performance of the classification model was evaluated using the confusion matrix and the indices of accuracy, precision, and recall.

### **Findings**

The modeling results showed that the designed algorithm was able to classify Telegram messages into the three categories of buy, sell, and neutral with acceptable accuracy. In the final evaluation, with a focus on the two main classes of “buy” and “sell,” the model achieved an overall accuracy of 96.6%. Moreover, the model’s precision and recall were reported to be approximately 89.15% and 90.82%, respectively, indicating a balanced performance in identifying trading signals. In particular, the model performed very well in detecting sell signals, with precision of about 98% and recall of about 97%. This finding suggests that the linguistic pattern of selling signals in Telegram messages is relatively

clearer and that the machine learning model was able to identify these patterns successfully. In the statistical analysis section, Pearson's correlation coefficient indicated a positive and relatively strong relationship between the number of sell signals and the daily trading value of the market, with a correlation coefficient of 0.80. The number of buy signals also showed a positive correlation of 0.72 with daily trading value. Neutral or news-related messages had a weaker correlation with trading value, equal to 0.58. The difference between the number of buy and sell signals showed a correlation of 0.61 with trading value. These results indicate that trading signals, particularly sell signals, have a stronger association with increased levels of trading activity in the market compared with neutral or news-related messages. In other words, whenever signaling activity in Telegram channels intensified, the daily trading value of the market also tended to increase.

### **Discussion & Conclusion**

The findings of this study indicate that unstructured data published on social networks can, if properly processed, be transformed into quantitative and analytical indicators. The study demonstrated that Telegram trading signals can be regarded as reflections of the sentiments and expectations of capital market participants and that they have a statistically significant relationship with the level of daily trading dynamism. From a theoretical perspective, the results are consistent with the principles of behavioral finance and with approaches that emphasize the role of informal information, collective expectations, and imitative behaviors in investors' decision-making processes. Due to their high speed of dissemination and broad user accessibility, social networks can play an influential role in shaping collective perceptions of market conditions and, through this mechanism, affect trading intensity. Nevertheless, it should be emphasized that the findings of this study indicate correlation rather than a definitive causal relationship. Therefore, it cannot be concluded with certainty that Telegram signals alone caused an increase in trading value. Rather, it can be stated that increased signaling activity on Telegram was accompanied by an increase in trading value. From an applied perspective, the results of this study may be useful for capital market supervisory institutions, the Securities and Exchange Organization, analysts, and investors. The development of intelligent social media monitoring tools can help identify abnormal behaviors, track waves of trading signals, reduce the possibility of market manipulation, and enhance informational transparency. Furthermore, investors can make more informed decisions by understanding both the role and the limitations of trading signals published on social networks.

**Keywords:** Social networks, Telegram, Trading value, Stocks, Data mining.



## بررسی تأثیر سیگنال‌های معاملاتی منتشر شده در کانال‌های تلگرامی بر ارزش معاملات

### بازار سرمایه با به‌کارگیری الگوریتم‌های یادگیری ماشین

محمد ابراهیم راعی عزآبادی (نویسندهٔ مسئول)

استادیار، گروه مالی و حسابداری، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران. رایانامه: meraei68@iau.ac.ir

سید حسین علوی

کارشناس ارشد، گروه مدیریت مالی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران. رایانامه: s.hossein.alavi@gmail.com

علی کیانی‌فر

دانشجوی دکتری، گروه مالی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران. رایانامه: ali.kianifar@iau.ac.ir

### چکیده

**هدف:** در این پژوهش، با استفاده از داده‌های متنی مرتبط با بازار سرمایه در کانال‌های تلگرامی و همچنین، داده‌های عددی مربوط به ارزش معاملات بازار سهام، تأثیر سیگنال‌های معاملاتی منتشر شده در این کانال‌ها بر ارزش معاملات بررسی شده است.

**روش:** این پژوهش با استفاده از رویکرد متن‌کاوی و تجزیه و تحلیل خودکار محتوای متنی، از طریق مدل ایجاد شده توسط الگوریتم‌های یادگیری ماشین، اجرا شده است. به‌منظور درک بهتر مسئله، ابتدا پژوهش‌های پیشین بررسی شد؛ سپس با توجه به خلأهای موجود در آن‌ها و در نظر گرفتن تغییرات روزانه در شبکهٔ اجتماعی تلگرام از نظر نوع فعالیت و همچنین تولید محتوا، مدلی طراحی شد که از نظر به‌روز بودن، از مدل‌هایی که در گذشته ارائه شده کارایی و دقت بیشتری دارد. صحت این مدل ۹۶ درصد و دقت و فراخوانی آن حدود ۹۰ درصد است. گام‌های اجرای این پژوهش عبارت بودند از: بارگیری داده‌های متنی از پیام‌رسان تلگرام، برچسب‌گذاری پیام‌ها، پیش‌پردازش و بازنمایی آن‌ها، ایجاد مدل توسط الگوریتم‌های یادگیری ماشین و پایش کل پیام‌ها در بازهٔ زمانی مد نظر با مدل. خروجی مدل، تعداد سیگنال‌ها در هر روز است که با ارزش معاملات بازار در آن روز مقایسه و هم‌بستگی آن بررسی شده است.

**یافته‌ها:** نتایج به‌دست‌آمده، گویای هم‌بستگی ۷۰ تا ۸۰ درصدی تعداد سیگنال‌های خرید و فروش با میزان ارزش معاملات روزانه سهام در بورس اوراق بهادار تهران است. به‌عبارتی، بیشترین هم‌بستگی با ۸۰ درصد بین تعداد سیگنال‌های فروش و ارزش معاملات است. در مقابل، پیام‌های خنثی یا خبری، هم‌بستگی ضعیف‌تری (۵۸ درصد) با ارزش معاملات دارند که می‌تواند از ماهیت غیرسیگنال‌محور این دسته از پیام‌ها نشئت گرفته باشد.

**نتیجه‌گیری:** یافته‌های این پژوهش می‌تواند برای سازمان بورس و نهادهای ناظر در جهت پایش فعالیت‌های سیگنال‌دهی در شبکه‌های اجتماعی استفاده شود. با توجه به نقش شایان توجه این سیگنال‌ها در افزایش فعالیت معاملاتی، توسعهٔ ابزارهای تحلیلی مبتنی بر داده‌کاوی، می‌تواند به شناسایی رفتارهای غیرعادی، کاهش احتمال دست‌کاری بازار و بهبود شفافیت اطلاعاتی کمک کند.

**کلیدواژه‌ها:** شبکه‌های اجتماعی، تلگرام، ارزش معاملات، سهام، داده‌کاوی.

**استناد:** راعی عزآبادی، محمد ابراهیم؛ علوی، سید حسین و کیانی‌فر، علی (۱۴۰۵). بررسی تأثیر سیگنال‌های معاملاتی منتشر شده در کانال‌های تلگرامی بر ارزش معاملات بازار سرمایه با به‌کارگیری الگوریتم‌های یادگیری ماشین. *بررسی‌های مدیریت رسانه*، ۱۵(۱)، ۳۱-۵۴.

## مقدمه

در سال‌های اخیر، تحول دیجیتال در نحوه تولید، انتشار و دریافت اطلاعات در بازارهای مالی تغییرات اساسی ایجاد کرده است؛ به‌ویژه با گسترش شبکه‌های اجتماعی، جریان اطلاعات مالی از حالت متمرکز و رسمی خارج و به بستری تعاملی و غیرمتمرکز تبدیل شده است. در این میان، بازار سرمایه به‌عنوان یکی از حساس‌ترین حوزه‌ها به اطلاعات، بیش از سایر بخش‌ها تحت تأثیر این تغییرات قرار گرفته است.

امروزه بخش شایان توجهی از سرمایه‌گذاران، به‌ویژه سرمایه‌گذاران خرد، برای دریافت اطلاعات و اتخاذ تصمیم‌های معاملاتی به شبکه‌های اجتماعی متکی هستند. در این فضا کاربران تحلیل‌ها، دیدگاه‌ها و به‌ویژه سیگنال‌های معاملاتی را به اشتراک می‌گذارند که می‌تواند بر رفتار سایر فعالان بازار تأثیرگذار باشد. انتشار سریع این اطلاعات، صرف‌نظر از صحت یا عدم صحت آن‌ها، می‌تواند به نوسان‌های چشمگیر در بازار منجر شود و اهمیت بررسی این پدیده را دوچندان سازد.

این شبکه‌ها تبادل عمومی اطلاعات را تسهیل می‌کنند و حاوی مقدار زیادی اطلاعات مربوط به سهام هستند. این داده‌ها، به‌طور فزاینده‌ای برای پیش‌بینی تحولات بازار سهام تجزیه و تحلیل می‌شوند. نتایج تحقیقات نشان می‌دهد که احساسات مالی، بیشترین تأثیر را دارد و باید توسط سرمایه‌گذارانی که تصمیم‌گیری می‌کنند، به‌دقت در نظر گرفته شود (بنتهاوس و بک<sup>۱</sup>، ۲۰۱۵). امروزه اخبار مربوط به بازار سهام در عرض چند ثانیه پخش می‌شود. هنگام مقایسه نوسان‌ها و درصد تغییرات، رسانه‌های اجتماعی تأثیر بیشتری نسبت به رسانه‌های جمعی دارند (بنتولیل و تامپکینز<sup>۲</sup>، ۲۰۲۲). پیشرفت‌های اخیر در زبان‌شناسی محاسباتی<sup>۳</sup>، پردازش زبان طبیعی<sup>۴</sup>، یادگیری ماشین<sup>۵</sup> و اقتصاد سنجی<sup>۶</sup>، همراه با بهبود دسترسی به داده‌ها به مقاله‌های رسانه‌ای، بحث‌های آنلاین و شبکه‌های اجتماعی، محققان را قادر ساخته است تا از چنین تکنیک‌هایی در زمینه‌های متعدد تحقیقات مالی استفاده کنند (لوگران و مک‌دونالد<sup>۷</sup>، ۲۰۱۶).

بازار سهام پرنوسان است؛ زیرا به عوامل سیاسی، مالی، محیطی و عوامل مختلف داخلی و خارجی همراه با داده‌های تاریخی سهام وابستگی دارد (مقبول، آگاروال، کاتور، میتال و غنایی<sup>۸</sup>، ۲۰۲۳). بازار تحت تأثیر سیاست‌های ملی، اقتصاد جهانی و منطقه‌ای و همچنین عوامل روانی و انسانی است، بنابراین عوامل خارجی مانند رسانه‌های اجتماعی و اخبار مالی، بر قیمت سهام تأثیرهای مثبت یا منفی دارند (پنگ، ژو، ونگ، لین و چنگ<sup>۹</sup>، ۲۰۲۰).

به‌طور سنتی، برای پیش‌بینی قیمت سهام، از دو رویکرد استفاده می‌شود. تحلیل بنیادی که به تحقیق درباره عوامل بنیادی شرکت مانند نرخ‌های رشد سالانه، موقعیت بازار، قراردادهای جدید شرکت، درآمد و هزینه‌ها می‌پردازد (نتی،

1. Benthau & Beck
2. Bentolila & Thompkins
3. Computational Linguistics
4. NLP (natural language processing)
5. machine learning
6. Econometrics
7. Loughran & McDonald
8. Maqbool, Aggarwal, Kaur, Mittal & Ganaie
9. Pang, Zhou, Wang, Lin & Chang

آدکویا و ویوری<sup>۱</sup>، (۲۰۲۳) و تحلیل تکنیکال که از نمودارهای قیمت تاریخی برای یافتن الگوها و پیش‌بینی استفاده می‌کند (وانگ و چن<sup>۲</sup>، ۲۰۰۷). با این حال خیلی از افراد فقط به کمک نظرات دیگران سرمایه‌گذاری می‌کنند. یکی از این روش‌ها، استفاده از رسانه‌ها و شبکه‌های اجتماعی است.

همان‌طور که اشاره شد، یکی از موضوعات تعاملی در شبکه‌های اجتماعی، مسائل مربوط به بازارهای مالی است. اینستاگرام به‌عنوان یک رسانه و تلگرام به‌عنوان یک شبکه اجتماعی، در رقابت با دیگر رسانه‌ها و شبکه‌ها گوی سبقت را در زمینه بازار سرمایه ربوده‌اند. بین ۳۴ تا ۷۰ درصد سرمایه‌گذاران در تصمیم‌گیری برای خرید و فروش سهام، از محتوای شبکه‌های اجتماعی استفاده می‌کنند. برای اینکه بتوان حال و هوای بازار یا در اصطلاح احساسات بازار را از شبکه‌های اجتماعی استخراج کرد، باید تحلیل داده‌ها را در مقادیر زیاد انجام داد. استخراج متون از شبکه‌های اجتماعی، به دلیل عدم ساختارمند بودن و داشتن اصطلاحات خاص و مختص اهالی بازار اغلب دشوار است (درخشان و بیگی<sup>۳</sup>، ۲۰۱۹).

با وجود گسترش مطالعات بین‌المللی در زمینه تحلیل احساسات و تأثیر رسانه‌های اجتماعی بر بازارهای مالی، در ادبیات داخلی پژوهش‌های محدودی به بررسی تجربی نقش سیگنال‌های معاملاتی در شبکه‌های اجتماعی فارسی پرداخته‌اند. علاوه بر این، بیشتر مطالعات بر شاخص‌هایی نظیر قیمت یا بازده سهام تمرکز داشته‌اند، در حالی که متغیر «ارزش معاملات روزانه»، به‌عنوان شاخصی از میزان فعالیت و مشارکت در بازار، کمتر مورد توجه قرار گرفته است.

در این پژوهش، تمایز مفهومی میان «محتوای منتشرشده»، «سیگنال معاملاتی» و «احساسات بازار» به‌صورت مشخص در نظر گرفته شده است. بدین ترتیب، پیام‌های منتشرشده در کانال‌های تلگرامی، به‌عنوان داده خام اولیه در نظر گرفته می‌شوند که پس از پردازش، به سیگنال‌های معاملاتی شامل پیشنهاد خرید، فروش یا محتوای خنثی تبدیل خواهند شد. این سیگنال‌ها، در واقع بازتابی از احساسات حاکم بر فعالان بازار هستند و به‌عنوان شاخصی از احساسات بازار مورد استفاده قرار می‌گیرند.

در این راستا، پژوهش حاضر، در پی پاسخ به این پرسش است که آیا سیگنال‌های معاملاتی منتشرشده در کانال‌های تلگرامی، می‌توانند بر ارزش معاملات روزانه بازار سرمایه تأثیرگذار باشند. برای این منظور، پیام‌های منتشرشده در کانال‌های تلگرامی مرتبط با بورس اوراق بهادار تهران، استخراج و با استفاده از الگوریتم‌های یادگیری ماشین، به سیگنال‌های خرید، فروش و خنثی طبقه‌بندی شده‌اند. سپس با تجمیع این سیگنال‌ها در مقیاس روزانه، رابطه میان تعداد آن‌ها و ارزش معاملات بازار مورد بررسی قرار گرفته است.

تمرکز بر «سیگنال‌های معاملاتی»، به‌جای اخبار عمومی به‌عنوان بازنمایی دقیق‌تر از رفتار و تصمیم‌های سرمایه‌گذاران و بررسی تأثیر این سیگنال‌ها بر «ارزش معاملات روزانه» به‌عنوان شاخصی از پویایی بازار، در حالی که اغلب مطالعات بر قیمت یا بازده تمرکز داشته‌اند و به‌کارگیری الگوریتم‌های یادگیری ماشین برای استخراج و طبقه‌بندی خودکار سیگنال‌های معاملاتی و بررسی تجربی رابطه میان سیگنال‌های رسانه‌ای فارسی که در مقایسه با مطالعات داخلی کمتر مورد توجه قرار گرفته است، از جمله جنبه‌های نوآورانه این پژوهش محسوب می‌شود.

## پیشینه نظری پژوهش

### رسانه‌های اجتماعی و تحول در جریان اطلاعات

رسانه‌های اجتماعی به‌عنوان یکی از مهم‌ترین تحولات عصر دیجیتال، ساختار ارتباطات و جریان اطلاعات را به‌طور بنیادین دگرگون کرده‌اند. این رسانه‌ها به فناوری‌های تعامل محور مبتنی بر رایانه گفته می‌شود که تولید و به‌اشتراک‌گذاری محتوا، اطلاعات و نظرها را برای افراد آسان می‌کند (کیتزمن، هرمنکنز، مک‌کارتی و سیلوستره<sup>۱</sup>، ۲۰۱۱). رسانه‌های اجتماعی به فعالیت‌های برخاسته از مجموعه‌هایی گفته می‌شود که از این طریق می‌توانند دانش، اطلاعات و عقاید خود را با استفاده از رسانه‌های ارتباطی به اشتراک بگذارند. به کمک این ابزار مبتنی بر وب، افراد می‌توانند اطلاعات، تصویرها، ویدئوها، متن‌ها و... را به راحتی برای یکدیگر بفرستند (سافکو و بریک<sup>۲</sup>، ۲۰۰۹). رسانه اجتماعی، در واقع به‌عنوان گروهی از برنامه‌های اینترنت محور تعریف می‌شود که بر بستر وب ایجاد می‌شود و به کاربران خود اجازه می‌دهد محتوا ایجاد کنند و آن‌ها را به‌اشتراک‌گذاری بگذارند (کاپلان و هانلین<sup>۳</sup>، ۲۰۱۰).

### شبکه‌های اجتماعی و انتشار اطلاعات

شبکه‌های اجتماعی ساختارهایی متشکل از افراد یا سازمان‌ها هستند که از طریق روابط متقابل به یکدیگر متصل شده‌اند. این شبکه‌ها می‌توانند در قالب پلتفرم‌های آنلاین نظیر توئیتر، اینستاگرام و تلگرام شکل بگیرند و بستری برای تعامل، تبادل اطلاعات و شکل‌گیری روابط اجتماعی فراهم کنند.

در این شبکه‌ها، محتوای تولیدشده توسط کاربران شامل متن‌ها، نظرها، تصویرها و سایر اشکال داده، نقش محوری دارد و جریان ارتباطی برخلاف رسانه‌های سنتی، دوسویه و تعاملی است. کاربران رسانه‌های اجتماعی از طریق ایجاد حساب کاربری منحصر به خود، در این بستر به فعالیت می‌پردازند (بوید و الیسون<sup>۴</sup>، ۲۰۰۷).

در سال‌های اخیر، الگوهای دریافت و انتشار خبر، به سمت پلتفرم‌های اجتماعی و پیام‌رسان‌ها سوق یافته‌اند و سازوکارهای الگوریتمی این پلتفرم‌ها در انتخاب و برجسته‌سازی محتوا نقش تعیین‌کننده‌ای یافته‌اند (الوندی، خانیکی و اکبرزاده جهرمی، ۱۴۰۱). در این چارچوب، پیام‌ها و سیگنال‌های منتشرشده در بسترهایی مانند تلگرام، می‌توانند در شکل‌دهی ادراک جمعی و در نتیجه تصمیم‌گیری‌های اقتصادی کاربران نقش‌آفرین باشند. از محبوب‌ترین رسانه‌های اجتماعی، می‌توان به فیسبوک، یوتیوب، اینستاگرام، توئیتر، ویجت، تلگرام و... اشاره کرد.

### رابطه محتوای خبری با بازار سهام

محتوای اخبار مالی بازار سهام با روان‌شناسی و جامعه‌شناسی سهام‌داران در ارتباط است (تتلاک<sup>۵</sup>، ۲۰۰۷). اطلاعات منتشر شده در خصوص صورت‌های مالی شرکت‌ها و سود آن‌ها، روزبه‌روز در حال بیشتر شدن است و حجم پرداختن به

1. Kietzmann, Hermkens, McCarthy & Silvestre
2. Online
3. Safco & Brake
4. Kaplan & Hanlin
5. Boyd & Ellison
6. Tetlock

اخبار مالی نیز در رسانه‌ها افزایش داشته است. در این خصوص، دو سؤالی که مطرح می‌شود، این است که آیا این اخبار سیگنالی درباره آینده شرکت‌ها به بازار می‌دهد؟ و آیا بازار به این سیگنال‌های احتمالی عکس‌العمل نشان می‌دهد؟ دیویس، پیگر و سدور<sup>۱</sup> (۲۰۱۲) برای پاسخ به این دو پرسش نشان دادند که مدیران از طریق ادبیات و کلمات تلاش دارند تا توقعاتشان به آینده بنگاه را منعکس کنند و بازار هم به این سیگنال‌ها واکنش نشان می‌دهد.

تحلیل بار احساسی متن‌ها نیز در تبیین رفتار بازار جایگاه مهمی دارد. چن و همکاران<sup>۲</sup> (۲۰۱۲) با سنجش واژگان دارای بار منفی در اخبار منتشرشده در پایگاه سیکینگ آلفا<sup>۳</sup> نسبت به کل واژگان، نشان دادند که افزایش واژگان منفی در اخبار با کاهش بازده سهام همراه است. در همین راستا، تتلاک، سار تسه کانسکی و مک‌اسکاسی<sup>۴</sup> (۲۰۰۸) با بررسی واژگان منفی در اخبار شاخص داو جونز<sup>۵</sup> و وال استریت طی سال‌های ۱۹۸۰ تا ۲۰۰۴، نشان دادند که محتوای زبانی اخبار اخبار می‌تواند به صورت نظام‌مند بر قیمت‌ها اثرگذار باشد. این یافته‌ها بیانگر آن است که لحن و محتوای متنی، همانند سایر متغیرهای اقتصادی، می‌تواند در شکل‌دهی انتظارات و تصمیمات سرمایه‌گذاران نقش ایفا کند.

## تأثیر انواع اخبار بر رفتار بازار سرمایه

### اخبار صنعتی

شرایط صنایع مختلف می‌تواند بر واکنش بازار به اخبار تأثیرگذار باشد. در تحقیقی اسپرنگر و ولپ<sup>۷</sup> (۲۰۱۱) به بررسی واکنش بازار به رویدادهای مرتبط با صنایع گوناگون پرداختند. نتایج تحقیق آنان نشان داد که بازار به پیشامدهای مختلف متناسب با آنکه روی چه صنعتی اثر می‌گذارد، اهمیت می‌دهد. آن‌ها از روش‌های زبان‌شناسی محاسباتی برای تمایز بین اخبار خوب و بد با کنترل احساسات (یعنی لحن مثبت در مقابل منفی) اخبار مختلف استفاده کردند و به این نتیجه رسیدند که قدر مطلق بازده قیمتی قبل از یک رویداد خبری برای اخبار مثبت بیشتر از اخبار منفی است و این امر نشان‌دهنده نشت اطلاعات گسترده‌تر قبل از اخبار خوب است؛ بنابراین واکنش بازار به انواع مختلف رویدادهای خبری متفاوت است.

### اخبار اقتصادی

بحث و بررسی روی این مطلب که آیا نوسان‌های قیمت سهام، تحت تأثیر اعلان رویدادهای اقتصادی در بین رسانه‌های جمعی است یا خیر، از مباحث کلیدی و به‌روز صنعت مالی و رسانه است. اخبار اقتصادی از جمله نرخ تورم، بیکاری و سیاست‌های پولی، می‌توانند بر بازار سرمایه اثرگذار باشند. پیرس و رولی<sup>۸</sup> (۱۹۸۴) واکنش‌های روزانه قیمت سهام به اطلاعاتی‌های مربوط به عرضه پول، تورم، فعالیت اقتصادی واقعی و نرخ تنزیل را تحلیل و بررسی کردند. نتایج نشان داد که اخبار سیاست‌های پولی، بر قیمت‌های سهام مؤثر به‌شدت است؛ اما در ارتباط با تورم، نتایج قوی نبود. همچنین،

1. Davis, Piger & Sedor

2. Chen et al

3. Seeking Alpha

4. Tetlock, Saar-Tsechansky & Macskassy

5. Dow Jones

6. Wall Street

7. Sprenger & Welpe

8. Pearce & Roley

رابطه‌ای بین فعالیت‌های واقعی اقتصادی و قیمت سهام در روز اعلان دیده نشد. دیاز و جارنو<sup>۱</sup> (۲۰۰۹) واکنش کوتاه‌مدت قیمت‌های روزانه سهام به اخبار مرتبط با تورم در بازار سهام اسپانیا را بررسی کردند و به این نتیجه رسیدند که واکنش بازار به اخبار تورمی، به ماهیت خبر و شرایط اقتصادی بستگی دارد.

در ادبیات داخلی، پیرایی و شهسوار (۱۳۸۸) نشان دادند که شاخص قیمت سهام در ایران با تولید ناخالص داخلی و سطح عمومی قیمت‌ها رابطه مستقیم و با حجم پول و نرخ ارز رابطه معکوس دارد. همچنین هیبتی و زندیه (۱۳۹۰) در بررسی واکنش بازار سهام ایران به اخبار بحران مالی جهانی، نشان دادند که برخی صنایع به انتشار این اخبار واکنش معنادار نشان داده‌اند و بازدهی‌های غیرعادی‌ای را در بازه ۳۰ روزه، پس از انتشار خبر تجربه کرده‌اند.

### اخبار سیاسی

اخبار سیاسی به‌عنوان یکی از عوامل ریسک سیستماتیک، می‌تواند بر بازار سرمایه تأثیر چشمگیری داشته باشد. به‌طور کلی، در شرایط نااطمینانی سیاسی مانند انتخابات، تغییرات دولت یا بحران‌های ساختاری، سرمایه‌گذاران بازده و نوسان‌های بازار را با حساسیت بیشتری دنبال می‌کنند. در بسیاری از این مطالعات، گزارش شده است که اخبار خوب سیاسی می‌تواند موجب افزایش بازده بازار و کاهش نوسان‌ها شود، در حالی که اخبار منفی سیاسی، موجب کاهش بازده و افزایش نوسان‌ها می‌شود.

تأثیر انتخابات و اخبار سیاسی پس از انتخابات ریاست جمهوری، بر نوسان‌های بازدهی بازار سهام موضوعی بود که کشاورز حداد و حیدری، در سال ۱۳۹۰ بررسی کردند و نتیجه گرفتند که تلاطم بازدهی سهام در روزهای قبل از انتخابات، به دلیل شرایط عدم قطعیتی که بر بازار حاکم است، افزایش می‌یابد. همچنین این افزایش تلاطم برای روزهای پس از انتخابات با به قدرت رسیدن برنده انتخابات، به دلیل شوک‌های سیاسی همچنان پایدار است (کشاورز حداد و حیدری، ۱۳۹۰).

أندر و سیمگاموگان<sup>۲</sup> (۲۰۰۶) پژوهشی در خصوص تأثیر اخبار سیاسی و اقتصادی بر شاخص‌های بازار سهام در دو بازار نوظهور آرژانتین و ترکیه انجام دادند. آن‌ها نتیجه گرفتند که اخبار بر بازارهای سهام تأثیر می‌گذارند. به نظر می‌رسد در هر دو بازار، رویدادهای سیاسی داخلی بر نوسان‌های بازده تأثیر می‌گذارد.

### سایر اخبار

یک‌سری اخبار، محتوا و رخدادهایی نظیر اخبار فرهنگی، اجتماعی، زیست محیطی و هنری نیز می‌توانند به‌صورت غیرمستقیم از طریق تأثیر بر شرایط اقتصادی و روانی جامعه، بر رفتار سرمایه‌گذاران اثرگذار باشند.

### داده‌کاوی و استخراج دانش از داده‌های مالی

رشد فناوری‌های اطلاعاتی و گسترش داده‌های دیجیتال، موجب شکل‌گیری حوزه داده‌کاوی به‌عنوان رویکردی

1. Diaz & Jareno

2. Onder & Simga-Mugan

میان‌رشته‌ای شده است. داده‌کاوی با بهره‌گیری از روش‌های آماری، یادگیری ماشینی<sup>۱</sup> و هوش مصنوعی<sup>۲</sup>، به استخراج الگوهای پنهان از داده‌های حجیم می‌پردازد. در شرایطی که روابط میان متغیرها به صورت صریح مشخص نیست، این رویکرد امکان کشف ساختارهای نهفته در داده‌ها را فراهم می‌کند (هان، کمبر و پی<sup>۳</sup>، ۲۰۱۰). در چارچوب فرایند استخراج دانش از پایگاه‌های داده، مراحل انتخاب داده، آماده‌سازی، کاهش ابعاد، مدل‌سازی و ارزیابی به صورت نظام‌مند اجرا می‌شود. این فرایند، به‌ویژه در تحلیل داده‌های مالی که حجم و پیچیدگی ساختاری زیادی دارند، اهمیت دارد.

### متن‌کاوی و تحلیل احساسات در داده‌های شبکه‌های اجتماعی

بخش عمده داده‌های تولیدشده در رسانه‌های اجتماعی، ماهیت غیرساختاریافته دارند. متن‌کاوی به‌عنوان شاخه‌ای از داده‌کاوی، به استخراج دانش از داده‌های متنی از طریق روش‌های پردازش زبان طبیعی و یادگیری ماشینی می‌پردازد (عباسی، ۱۳۹۵). هدف اصلی متن‌کاوی، تبدیل متن‌های غیرساختاریافته به نمایش‌های ساختاریافته و قابل تحلیل است. در فرایند متن‌کاوی، پیش‌پردازش متن‌ها (شامل حذف ایست‌واژه‌ها، ریشه‌یابی، نشانه‌گذاری و انتخاب ویژگی)، کاهش ابعاد و سپس طبقه‌بندی یا استخراج الگو انجام می‌شود. در بسیاری از کاربردهای مالی، روش‌های کیسه‌واژگان، مدل‌سازی موضوعی و الگوریتم‌های یادگیری با نظارت برای طبقه‌بندی متون به کار می‌روند. یکی از کاربردهای مهم متن‌کاوی در حوزه اقتصاد و مالی، «تحلیل احساسات» یا «عقیده‌کاوی» است. در این رویکرد، بار مثبت یا منفی متن‌ها استخراج و به شاخص‌های کمی تبدیل می‌شود. این شاخص‌ها می‌توانند نمایانگر فضای روانی حاکم بر بازار باشند و در پیش‌بینی رفتار سرمایه‌گذاران استفاده شوند.

### چارچوب نظری پژوهش

بر اساس مبانی نظری فوق، استدلال‌های زیر مطرح است:

۱. رسانه‌های اجتماعی بستری برای انتشار سریع اطلاعات و شکل‌گیری انتظارات جمعی هستند؛
۲. زبان و محتوای متنی منتشرشده، می‌تواند حاوی سیگنال‌های اطلاعاتی باشد؛
۳. احساسات و ادراکات جمعی، مطابق نظریه مالی رفتاری، بر رفتار معاملاتی اثر می‌گذارند؛
۴. افزایش فعالیت سیگنال‌دهی در شبکه‌های اجتماعی، می‌تواند با افزایش پویایی بازار (نظیر ارزش معاملات) همراه باشد.

پژوهش حاضر در این چارچوب نظری، با استخراج سیگنال‌های خریدوفروش از پیام‌های تلگرامی فارسی و بررسی رابطه آن با ارزش معاملات روزانه، به آزمون تجربی این پیوند می‌پردازد.

### پیشینه تجربی پژوهش

در سال‌های اخیر، نقش رسانه‌های اجتماعی و تحلیل احساسات در پیش‌بینی بازارهای مالی به‌طور گسترده مورد توجه

1. machine learning  
2. Artificial intelligence  
3. Han, Kamber & Pei

پژوهشگران قرار گرفته است. مقبول و همکاران (۲۰۲۳) در پژوهشی با عنوان «پیش‌بینی سهام با ادغام احساسات اخبار مالی و رویکرد یادگیری ماشین (رگرسیون MLP)»، به بررسی امکان بهبود پیش‌بینی بازار از طریق ترکیب داده‌های احساسی اخبار مالی و مدل‌های یادگیری ماشین پرداختند. نتایج این پژوهش حاکی از آن است که ادغام داده‌های متنی با مدل‌های یادگیری، می‌تواند دقت پیش‌بینی را بهبود بخشد.

در پژوهشی دیگر، جیانگ و ژنگ<sup>۱</sup> (۲۰۲۳) به بررسی نقش مدل‌های یادگیری عمیق در بهبود پیش‌بینی بازار سهام پرداختند. آن‌ها با استفاده از مدل FinBERT و شبکه LSTM نشان دادند که ترکیب احساسات متنی با داده‌های تاریخی قیمت، دقت پیش‌بینی روند بازار را به‌طور معناداری افزایش می‌دهد.

اسمیت و اوهر<sup>۲</sup> (۲۰۲۲) نیز در پژوهشی به بررسی این موضوع پرداختند که در دوران شیوع ویروس کووید، آیا احساسات اخبار روزانه چندین شرکت و احساسات توئیتر مدیرعاملان آن‌ها، بر عملکرد بازار آن‌ها تأثیر می‌گذارد. نتایج نشان دادند که هم‌بستگی بین احساسات توئیتر و تغییرات قیمت سهام محدود است، در حالی که احساسات اخبار مالی ارتباط قوی‌تری با عملکرد بازار دارد.

ژانگ، چن، رونگ، و تان<sup>۳</sup> (۲۰۲۲) در پژوهشی با هدف تحلیل اثر شایعه‌های رسانه‌ای بر نوسان‌های بازار سهام، از مدل‌های اقتصادسنجی برای بررسی سازوکار انتقال اطلاعات استفاده کردند. نتایج نشان داد که شایعه‌های رسانه‌های اجتماعی، در تشدید نوسان‌های بازار نقش مهمی دارند.

وانگ، هو، لی، هو و کمبری<sup>۴</sup> (۲۰۲۱) در مطالعه‌ای احتمال دوام پیش‌بینی روند بازار سهام مبتنی بر یادگیری را با ترکیب تحلیل احساسات رسانه‌های اجتماعی بررسی کردند. پنج روش یادگیری ماشین شامل<sup>۵</sup> MLP،<sup>۶</sup> SVM،<sup>۷</sup> LR،<sup>۸</sup> رندوم‌فارست<sup>۹</sup> و نایویز<sup>۹</sup> به‌عنوان مدل پایه انتخاب شد. نتایج نشان داد که ترکیب داده‌های رسانه‌های اجتماعی با مدل‌های یادگیری، می‌تواند احتمال پیش‌بینی موفق را افزایش دهد.

در پژوهشی دیگر که پادمانایانا و باویا<sup>۱۰</sup> (۲۰۲۱) انجام دادند، تأثیر احساسات بیان شده از طریق توئیتهای توئیتر در پیش‌بینی قیمت سهام بررسی شد. در این پژوهش، توئیتهای با استفاده از طبقه‌بندی نایویز، به سه دسته مثبت، منفی و خنثی تقسیم شد. داده‌های موجود همراه با داده‌های توئیتهای، به‌عنوان ورودی به مدل یادگیری ماشین برای به‌دست آوردن نتیجه داده شد. نتایج به‌دست‌آمده پس از فرایند پیش‌بینی، به‌وضوح مشخص می‌کند که با دقت ۸۹/۸ درصد، می‌توان از داده‌های توئیتر برای پیش‌بینی قیمت در زمان آینده استفاده کرد.

1. Jiang & Zeng
2. Smith & O'Hare
3. Zhang, Chen, Rong, Wang & Tan
4. Wang, Hu, Li, Ho & Cambria
5. Multi-Layer Perception
6. Support Vector Machine
7. Logistic Regression
8. Random Forest
9. Naive Bayes
10. Padmanayana & Bhavya

مهتا، پانديا و دوشي<sup>۱</sup> (۲۰۲۰) نیز به بررسی ارتباط تغییرات قیمت سهام یک شرکت با نظرات (احساسات) مردم در مورد آن شرکت پرداختند. در این پژوهش یک ابزار پیش‌بینی قیمت سهام با در نظر گرفتن احساسات عمومی طراحی شد. الگوریتم پیشنهادی، احساسات عمومی، نظرات، اخبار و قیمت‌های تاریخی سهام را برای پیش‌بینی قیمت‌های آتی سهام در نظر می‌گیرد. طی این آزمایش‌ها، از روش‌های یادگیری ماشین و یادگیری عمیق از جمله SVM، رگرسیون خطی، نایویز و LSTM<sup>۲</sup> استفاده شد. نتایج ارائه شده حاکی از موفقیت روش پیشنهادی بوده است.

وو، ونگ، ما و یه<sup>۳</sup> (۲۰۱۷)، پژوهشی با هدف بررسی تأثیر رسانه‌های اجتماعی بر نوسان‌های سهام انجام دادند. این پژوهش یک مدل پیش‌بینی بر اساس عوامل اجتماعی و احساسات عمومی برای پیش‌بینی نوسان‌های سهام معرفی کرده است. نتایج نشان داد که تأثیر داده‌های رسانه‌های اجتماعی بر نوسان‌های روز بعد بیشتر است؛ اما با گذشت زمان کاهش می‌یابد.

در ایران نیز پژوهش‌هایی در زمینه کاربرد متن‌کاوی و تحلیل احساسات در بازار سرمایه انجام شده است. ابراهیمیان، عباسی، عالم تبریز و محمدزاده (۱۴۰۰) در پژوهشی با عنوان «پیش‌بینی روند روزانه قیمت سهام با استفاده از متن‌کاوی احساسات کاربران شبکه اجتماعی و داده‌کاوی نشانگر (اندیکاتور)های تکنیکال» به پیش‌بینی روند آتی قیمت سهام پرداختند. آن‌ها با ترکیب تحلیل احساسات و شاخص‌های تکنیکال و استفاده از الگوریتم‌های درخت تصمیم<sup>۴</sup>، تصمیم<sup>۵</sup>، نایویز و SVM، مدلی با دقت حدود ۷۳ درصد ارائه کردند و نشان دادند ماشین بردار پشتیبان، نسبت به سایر الگوریتم‌ها عملکرد بهتری دارد و بین حجم معاملات و تعداد نظرهای کاربران هم‌بستگی معناداری وجود دارد.

هاتفی قهقرخی و شمس فرد (۱۳۹۹)، پژوهشی با عنوان «پیش‌بینی بورس اوراق بهادار تهران با استفاده از تحلیل احساسات نظرات متنی آنلاین» انجام دادند. آن‌ها با طراحی واژگان احساسی تخصصی برای زبان فارسی و استفاده از تحلیل رگرسیون چندگانه، اثر نظرهای کاربران شبکه سهام‌یاب را بر پیش‌بینی بورس تهران بررسی کردند. نتایج نشان داد که پیش‌بینی‌پذیری سهام مختلف در اوراق بهادار تهران، بسته به ویژگی‌های آن‌ها متفاوت است.

در برخی پژوهش‌های دانشگاهی نیز کاربرد متن‌کاوی در بازار سرمایه بررسی شده است. فروغی‌مجد (۱۳۹۸) با استفاده از مدل ترکیبی داده‌های متنی و سری‌های زمانی قیمت سهام و الگوریتم SVM، نشان داد که افزودن داده‌های متنی موجب بهبود دقت پیش‌بینی روند روز بعد می‌شود، هرچند کارایی مدل در افق‌های بلندمدت کاهش می‌یابد.

عزیزی، عبدالوند، قالیباف اصل و رجائی هرندي (۱۳۹۹) نیز با استفاده از تکنیک‌های متن‌کاوی و رگرسیون به بررسی تأثیر اخبار بر شاخص سهام پرداختند و نشان دادند دسته‌بندی خودکار اخبار، می‌تواند در پیش‌بینی تغییرات شاخص مؤثر باشد.

در نهایت، راعی، حسینی و کیانی هرچگانی (۱۳۹۵) در مطالعه‌ای، توانایی نظرهای کاربران شبکه‌های اجتماعی را بر پیش‌بینی جهت و قیمت سهام در بورس اوراق بهادار تهران بررسی کردند. در این مطالعه از داده‌های دوره‌های زمانی

1. Mehta, Pandya & Doshi  
2. Long Short-Term Memory  
3. Wu, Wang, Ma & Ye  
4. Indicator  
5. Decision Tree

ابتدای تیرماه ۱۳۹۲ تا انتهای خرداد ۱۳۹۳ استفاده شده است. برای پیش‌بینی، شبکه‌ی مصنوعی پیش‌خور<sup>۱</sup> و بهینه‌سازی شبکه با ۳ تا ۱۰ وقفه و یک لایه پنهان تا ۲۰ نرون به کار گرفته شده است. نتایج نشان می‌دهد که در پیش‌بینی قیمت سهام در دو روش تفاوت معناداری وجود ندارد؛ اما در پیش‌بینی جهت قیمت سهام با استفاده از قیمت سهام و پیشنهادهای خریدوفروش (ترکیب داده‌های قیمت و شبکه اجتماعی) نسبت به پیش‌بینی فقط با قیمت سهام در دوره رونق برای شرکت‌های پربیننده و در دوره رکود برای ۱۰ سهم فعال تفاوت معناداری وجود دارد.

مرور مطالعات نشان می‌دهد که در سطح بین‌المللی، نقش داده‌های متنی و احساسات در تحلیل بازارهای مالی، به‌خوبی بررسی شده است؛ اما در ادبیات داخلی همچنان خلأهایی شایان توجهی وجود دارد. بسیاری از پژوهش‌های داخلی بر پیش‌بینی قیمت یا بازده تمرکز داشته‌اند و کمتر به بررسی متغیرهایی مانند «ارزش معاملات روزانه» به‌عنوان شاخصی از پویایی بازار پرداخته‌اند. همچنین، مطالعات محدودی به تحلیل سیگنال‌های معاملاتی در شبکه‌های اجتماعی فارسی، به‌ویژه تلگرام، پرداخته‌اند، در حالی که این بستر، در انتشار اطلاعات و شکل‌دهی رفتار سرمایه‌گذاران در بازار سرمایه ایران، نقش مهمی ایفا می‌کند. بر این اساس، پژوهش حاضر با تمرکز بر استخراج سیگنال‌های معاملاتی از داده‌های متنی فارسی و بررسی رابطه آن‌ها با ارزش معاملات روزانه، در تلاش است تا این خلأ پژوهشی را پوشش دهد.

### روش‌شناسی پژوهش

پژوهش حاضر از نظر هدف، کاربردی و از نظر روش، توصیفی - تحلیلی از نوع هم‌بستگی است که با رویکرد داده‌کاوی و متن‌کاوی انجام شده است. هدف اصلی این پژوهش، بررسی تأثیر سیگنال‌های منتشرشده در شبکه اجتماعی تلگرام روی ارزش معاملات روزانه بورس اوراق بهادار تهران است.

جامعه آماری پژوهش در بخش داده‌های متنی، شامل کلیه شبکه‌های اجتماعی فعال در حوزه بازار سرمایه ایران است که از میان آن‌ها، شبکه پیام‌رسان تلگرام به دلیل فراگیری آن در میان فعالان بازار سرمایه و نقش مؤثر در انتشار تحلیل‌ها و سیگنال‌های معاملاتی، برای نمونه انتخاب شد. زیرنمونه پژوهش، شامل کانال‌های بورسی فعال با حداقل ۲۰۰ عضو بود که به‌طور مستمر به انتشار تحلیل‌ها و توصیه‌های خریدوفروش سهام می‌پرداختند. در بخش داده‌های عددی، جامعه آماری شامل کلیه معاملات انجام‌شده در بورس اوراق بهادار تهران و متغیر مورد بررسی، ارزش معاملات روزانه کل بازار در دوره زمانی مدنظر است.

در این پژوهش، متغیر مستقل، احساسات استخراج‌شده از سیگنال‌های معاملاتی منتشرشده در کانال‌های تلگرامی فعال در حوزه بازار سرمایه و متغیر وابسته، ارزش معاملات روزانه سهام در بورس اوراق بهادار تهران است. داده‌های این پژوهش شامل دو دسته داده‌های عددی و متنی است. داده‌های عددی به ارزش معاملات روزانه بازار سرمایه مربوط است که از سایت شرکت مدیریت فناوری بورس تهران استخراج شده است. در این راستا، به‌منظور ارائه تصویری دقیق‌تر از رفتار واقعی معاملات، فقط ارزش معاملات خرد سهام و حق تقدم‌های بورس و فرابورس (بدون احتساب معاملات بلوکی) در بازه زمانی ابتدا تا انتهای سال ۱۴۰۱ در نظر گرفته شد که شامل ۲۳۷ روز معاملاتی است.

داده‌های متنی نیز از کانال‌های تلگرامی منتخب با استفاده از ربات‌های خزنده استخراج شد. در مجموع، تعداد شایان توجهی پیام متنی از این کانال‌ها جمع‌آوری شد که پس از پردازش، کل داده‌های مورد تحلیل شامل بیش از ۳/۵ میلیون پیام در بازه زمانی سال ۱۴۰۱ است. از میان این داده‌ها، حدود ۵۰۰۰ پیام به‌صورت دستی و توسط فرد آشنا به ادبیات بازار سرمایه برچسب‌گذاری و برای آموزش مدل استفاده شد. در فرایند برچسب‌گذاری، پیام‌ها بر اساس محتوای خود به سه دسته «سیگنال خرید»، «سیگنال فروش» و «خنثی/خبری» تقسیم شدند. معیار اصلی در این طبقه‌بندی، وجود توصیه صریح یا ضمنی برای خرید یا فروش سهام در متن پیام‌ها بود. به‌منظور کاهش سوگیری و افزایش یکنواختی، برچسب‌گذاری بر اساس دستورالعمل مشخص انجام شد و با بازبینی مکرر همراه بود؛ به‌گونه‌ای که در موارد ابهام، تصمیم‌گیری بر اساس غالب محتوای پیام صورت گرفت. پس از آموزش مدل، از آن برای طبقه‌بندی خودکار سایر پیام‌ها استفاده شد.

به‌منظور آماده‌سازی داده‌های متنی برای تحلیل، در مرحله پیش‌پردازش، عملیات توکن‌سازی، حذف نویز و کاراکترهای غیرضروری، حذف ایست‌واژه‌ها، نرمال‌سازی حروف و ریشه‌یابی روی متن‌ها انجام شد تا داده‌ها به فرم استاندارد و قابل پردازش تبدیل شوند. سپس به‌منظور بازنمایی عددی متون، از روش‌های بردارسازی نظیر Bag-of-words، TF-IDF و مدل‌های مبتنی بر ترنسفورمر از جمله BERT استفاده شد.

در مرحله مدل‌سازی، از الگوریتم‌های یادگیری ماشین با ناظر شامل ماشین بردار پشتیبان (SVM)، نایویز و مدل مبتنی بر BERT برای طبقه‌بندی پیام‌ها بهره گرفته شد. داده‌های برچسب‌خورده به سه بخش آموزش (۷۰ درصد)، اعتبارسنجی (۱۵ درصد) و آزمون (۱۵ درصد) تقسیم شد. هدف از به‌کارگیری این الگوریتم‌ها، استفاده از توانمندی‌های متفاوت آن‌ها در تحلیل داده‌های متنی و افزایش دقت طبقه‌بندی پیام‌ها بوده است و تمرکز پژوهش بر مقایسه عملکرد مدل‌ها نیست. پس از آموزش و ارزیابی اولیه، از مدل‌های توسعه‌یافته برای طبقه‌بندی کل پیام‌های استخراج‌شده استفاده شد. در ادامه، تعداد سیگنال‌های خرید، فروش و خنثی در هر روز معاملاتی، بر اساس خروجی مدل‌ها محاسبه و به‌عنوان ورودی تحلیل‌های آماری مورد استفاده قرار گرفت.

در نهایت، به‌منظور بررسی رابطه میان متغیرهای پژوهش، از ضریب هم‌بستگی پیرسون برای تحلیل ارتباط بین تعداد سیگنال‌های روزانه و ارزش معاملات روزانه بازار استفاده شد. همچنین عملکرد مدل‌های طبقه‌بندی بر اساس ماتریس درهم‌ریختگی و شاخص‌های صحت، دقت و فراخوانی ارزیابی شد.

## یافته‌های پژوهش

### توصیف داده‌های مورد استفاده

در گام نخست، داده‌های مورد استفاده در پژوهش، شامل داده‌های متنی استخراج‌شده از کانال‌های تلگرامی فعال در حوزه بازار سرمایه و داده‌های عددی مربوط به ارزش معاملات روزانه بازار بورس اوراق بهادار تهران گردآوری و سازمان‌دهی شد. جدول ۱ مشخصات داده‌های مورد استفاده در پژوهش را نشان می‌دهد.

## جدول ۱. مشخصات داده‌های مورد استفاده پژوهش

شرح	منبع	نوع داده
پیام‌های مرتبط با سیگنال خریدوفروش	کانال‌های تلگرامی بورسی	داده‌های متنی
ارزش معاملات روزانه بازار	tsetmc.com	داده‌های عددی

همان‌گونه که جدول ۱ نشان می‌دهد، پژوهش حاضر بر ترکیب داده‌های متنی و عددی استوار است و رویکردی تلفیقی در تحلیل اطلاعات بازار اتخاذ کرده است.

## پیش‌پردازش و آماده‌سازی داده‌های متنی

پس از گردآوری داده‌های متنی، مرحله پیش‌پردازش با هدف استانداردسازی و کاهش نویز انجام شد. این مرحله نقش تعیین‌کننده‌ای در بهبود عملکرد مدل‌های یادگیری ماشین دارد؛ زیرا داده‌های خام شبکه‌های اجتماعی مانند تلگرام معمولاً شامل نشانه‌ها، نگارش‌های متفاوت و واژگان زائد هستند. جدول ۲ مراحل پیش‌پردازش داده‌های متنی را نشان می‌دهد.

## جدول ۲. مراحل پیش‌پردازش داده‌های متنی

شرح عملیات	مرحله
تبدیل متن به واحدهای واژگانی	توکن‌سازی
حذف نشانه‌ها و کاراکترهای غیرضروری	حذف نویز
یکسان‌سازی حروف هم‌ارز	نرمال‌سازی
حذف واژگان فاقد بار معنایی	حذف ایست‌واژه‌ها
تبدیل کلمات به بن‌واژه	ریشه‌یابی

اجرای مراحل فوق، موجب کاهش ابعاد داده و افزایش انسجام معنایی متون شد و زمینه لازم را برای بازنمایی عددی و مدل‌سازی فراهم کرد. برای پیاده‌سازی و انجام پژوهش، از زبان برنامه‌نویسی SQL به‌منظور کار با داده‌های عددی و برای کار متن‌کاوی از پایتون استفاده شد.

## نتایج مدل‌سازی

در این بخش، خروجی مدل طبقه‌بندی روی داده‌های متنی استخراج شده ارائه می‌شود. پس از آموزش و بهینه‌سازی مدل، تمامی پیام‌های منتشرشده در کانال‌های منتخب تلگرامی پردازش و در سه طبقه «خرید»، «فروش» و «خنثی» دسته‌بندی شدند. به‌منظور هم‌ترازسازی زمانی با داده‌های معاملاتی بازار، پیام‌های هر روز در بازه زمانی ساعت ۹ صبح تا ۹ صبح روز بعد گردآوری و تعداد سیگنال‌های خریدوفروش برای هر روز معاملاتی محاسبه شد. برای بررسی رابطه میان فعالیت سیگنال‌دهی و ارزش معاملات بازار، سه شاخص مستقل از خروجی مدل استخراج شد:

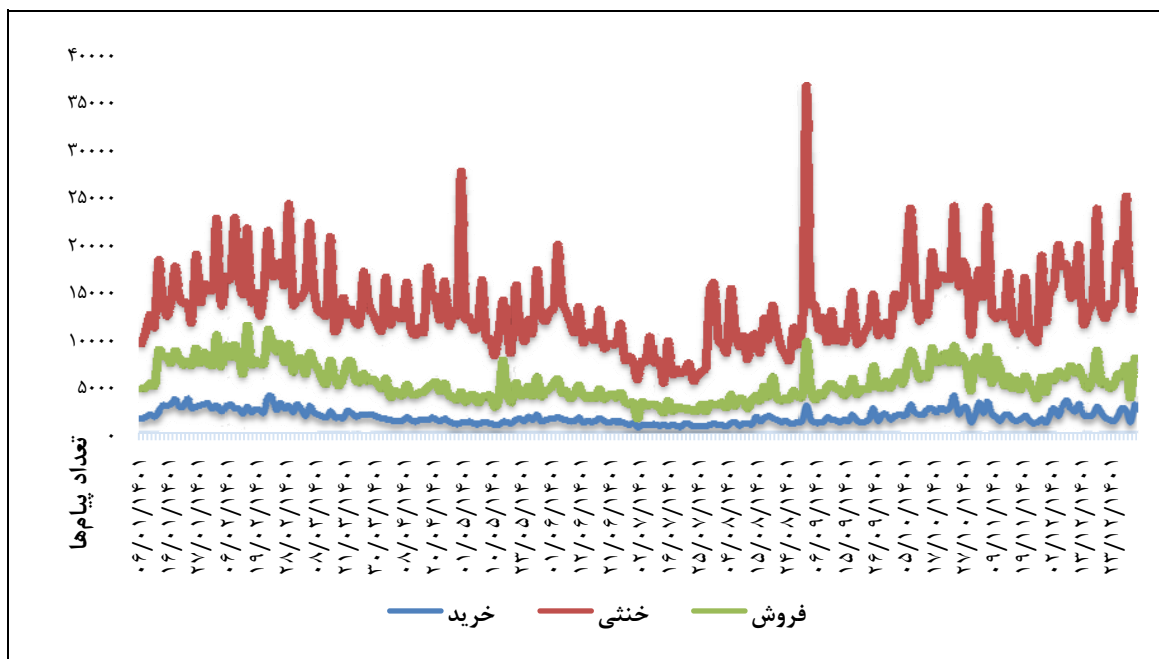
- تعداد روزانه پیام‌های خرید؛
- تعداد روزانه پیام‌های فروش؛
- تفاضل تعداد پیام‌های خرید و فروش.

این مقادیر پس از تجمیع روزانه، مبنای تحلیل‌های آماری بعدی قرار گرفتند. جدول ۳ نمونه‌ای از نتایج مدل‌سازی پژوهش برای داده‌های ورودی را نشان می‌دهد.

جدول ۳. نمونه‌ای از نتایج مدل‌سازی پیام‌ها

فروش	خشتی	خرید	تاریخ	فروش	خشتی	خرید	تاریخ
۸۳۸۳	۲۲۱۰۴	۲۶۹۲	۰۴/۰۳/۱۴۰۱	۴۶۹۲	۹۴۶۳	۱۳۷۱	۰۶/۰۱/۱۴۰۱
۷۳۹۱	۱۵۲۲۰	۲۰۱۱	۰۷/۰۳/۱۴۰۱	۴۸۵۶	۱۰۹۴۹	۱۵۷۰	۰۷/۰۱/۱۴۰۱
۶۷۴۹	۱۲۹۲۳	۱۷۹۷	۰۸/۰۳/۱۴۰۱	۵۲۴۹	۱۲۴۸۲	۱۷۸۶	۰۸/۰۱/۱۴۰۱
۵۶۶۸	۱۲۶۸۹	۱۶۰۰	۰۹/۰۳/۱۴۰۱	۵۰۴۲	۱۱۲۷۷	۱۵۹۵	۰۹/۰۱/۱۴۰۱
۵۲۰۳	۱۲۳۶۴	۱۵۱۹	۱۰/۰۳/۱۴۰۱	۸۷۵۴	۱۸۱۶۹	۲۰۳۶	۱۰/۰۱/۱۴۰۱
۷۵۸۴	۲۰۶۰۱	۲۱۲۵	۱۱/۰۳/۱۴۰۱	۸۱۷۷	۱۴۰۴۵	۲۷۳۷	۱۴/۰۱/۱۴۰۱
۵۹۶۷	۱۰۷۲۸	۱۴۹۵	۱۶/۰۳/۱۴۰۱	۸۰۳۸	۱۲۳۶۶	۲۷۵۵	۱۵/۰۱/۱۴۰۱
۵۰۱۴	۱۱۳۱۸	۱۵۱۸	۱۷/۰۳/۱۴۰۱	۷۴۱۸	۱۳۹۹۸	۲۸۷۶	۱۶/۰۱/۱۴۰۱
۵۷۹۶	۱۴۲۱۰	۱۴۵۷	۱۸/۰۳/۱۴۰۱	۸۶۸۳	۱۷۴۸۹	۳۴۰۴	۱۷/۰۱/۱۴۰۱
۷۳۴۰	۱۲۴۸۱	۲۱۷۲	۲۱/۰۳/۱۴۰۱	۷۹۰۰	۱۴۲۵۸	۲۷۱۰	۲۰/۰۱/۱۴۰۱
۷۴۵۰	۱۲۹۶۱	۱۹۵۴	۲۲/۰۳/۱۴۰۱	۷۲۵۲	۱۳۷۱۱	۲۶۲۳	۲۱/۰۱/۱۴۰۱
۵۶۶۷	۱۱۵۹۴	۱۵۸۸	۲۳/۰۳/۱۴۰۱	۷۶۳۰	۱۳۶۰۵	۳۴۷۰	۲۲/۰۱/۱۴۰۱
۵۳۸۰	۱۱۵۲۴	۱۷۳۴	۲۴/۰۳/۱۴۰۱	۷۲۰۸	۱۱۸۰۳	۲۵۲۷	۲۳/۰۱/۱۴۰۱
۶۲۰۵	۱۶۹۱۹	۱۸۱۱	۲۵/۰۳/۱۴۰۱	۸۹۷۷	۱۸۷۱۵	۲۵۹۸	۲۴/۰۱/۱۴۰۱
۵۴۲۶	۱۳۶۰۸	۱۷۶۵	۲۸/۰۳/۱۴۰۱	۷۴۴۷	۱۳۸۰۳	۲۷۴۶	۲۷/۰۱/۱۴۰۱
۵۵۶۴	۱۲۵۰۰	۱۷۳۱	۲۹/۰۳/۱۴۰۱	۸۴۷۲	۱۵۶۷۰	۲۸۷۸	۲۸/۰۱/۱۴۰۱
۵۰۸۵	۱۱۳۶۲	۱۵۴۱	۳۰/۰۳/۱۴۰۱	۷۴۰۳	۱۵۰۵۹	۲۹۷۶	۲۹/۰۱/۱۴۰۱
۴۶۶۴	۱۰۷۶۴	۱۴۰۰	۳۱/۰۳/۱۴۰۱	۷۲۱۵	۱۵۱۰۰	۲۴۳۵	۳۰/۰۱/۱۴۰۱
۵۶۲۵	۱۶۳۳۳	۱۳۱۶	۰۱/۰۴/۱۴۰۱	۱۰۳۴۴	۲۲۵۳۱	۲۷۴۳	۳۱/۰۱/۱۴۰۱
۳۷۲۵	۱۱۵۰۵	۱۲۴۱	۰۴/۰۴/۱۴۰۱	۷۰۸۱	۱۳۶۲۷	۲۱۳۶	۰۴/۰۲/۱۴۰۱
۴۳۶۵	۱۲۸۰۲	۱۰۹۹	۰۵/۰۴/۱۴۰۱	۹۱۱۲	۱۶۴۶۵	۲۵۹۸	۰۵/۰۲/۱۴۰۱
۴۰۵۵	۱۲۱۲۵	۱۱۴۰	۰۶/۰۴/۱۴۰۱	۷۹۳۳	۱۶۲۰۵	۲۸۵۷	۰۶/۰۲/۱۴۰۱
۳۹۴۸	۱۲۱۰۴	۱۱۶۶	۰۷/۰۴/۱۴۰۱	۹۲۰۸	۲۲۶۲۶	۲۴۴۷	۰۷/۰۲/۱۴۰۱
۴۹۷۷	۱۵۷۷۱	۱۴۸۹	۰۸/۰۴/۱۴۰۱	۷۹۴۱	۱۶۲۲۴	۲۴۰۱	۱۰/۰۲/۱۴۰۱
۴۰۸۳	۱۰۸۴۹	۱۲۰۶	۱۱/۰۴/۱۴۰۱	۶۳۲۳	۱۴۷۲۳	۱۹۳۵	۱۱/۰۲/۱۴۰۱
۳۹۵۳	۱۰۲۹۳	۱۰۵۱	۱۲/۰۴/۱۴۰۱	۱۱۲۶۱	۲۱۵۱۱	۲۶۱۵	۱۲/۰۲/۱۴۰۱
۴۱۶۳	۱۱۰۳۹	۱۲۰۸	۱۳/۰۴/۱۴۰۱	۷۲۶۶	۱۳۸۱۷	۲۰۶۳	۱۷/۰۲/۱۴۰۱
۴۲۷۳	۱۰۵۷۰	۱۳۰۰	۱۴/۰۴/۱۴۰۱	۷۵۶۵	۱۴۷۱۵	۲۳۶۷	۱۸/۰۲/۱۴۰۱

جدول ۳ نشان می‌دهد که مدل متن‌کاوی قادر بوده است پیام‌های استخراج‌شده را به‌صورت ساختاریافته در سه طبقه اصلی تفکیک کند. در شکل ۱، نمودار توزیع پیام‌های خرید، فروش و خنثی استخراج‌شده توسط مدل نشان داده شده است. این نمودار، توزیع سه کلاس پیام را در کل دوره زمانی نشان می‌دهد. مقایسه ارتفاع ستون‌ها بیانگر تفاوت نسبی میان فراوانی سیگنال‌های خرید، فروش و پیام‌های خنثی است. این الگوی توزیع، در ادامه به‌عنوان مبنای تحلیل رابطه میان فعالیت سیگنال‌دهی و ارزش معاملات بازار مورد استفاده قرار گرفت.



شکل ۱. نمودار توزیع پیام‌های خرید، فروش و خنثی استخراج‌شده توسط مدل

### نتایج ارزیابی مدل

به‌منظور ارزیابی عملکرد مدل طبقه‌بندی، داده‌های برچسب‌گذاری‌شده دستی به سه بخش آموزش (۷۰ درصد)، اعتبارسنجی (۱۵ درصد) و آزمون (۱۵ درصد) تقسیم شدند. مجموعه آزمون که پیش‌تر در فرایند آموزش استفاده نشده بود، برای سنجش نهایی مدل به کار گرفته شد. عملکرد مدل با استفاده از ماتریس درهم‌ریختگی ارزیابی گردید. با توجه به اینکه پیام‌ها در سه کلاس خرید، فروش و خنثی دسته‌بندی شده بودند، در مرحله نخست ماتریس ارزیابی به‌صورت  $3 \times 3$  محاسبه شد. با این حال، نظر به تمرکز پژوهش بر سیگنال‌های معاملاتی، تحلیل نهایی بر اساس ماتریس  $2 \times 2$  شامل دو کلاس «خرید» و «فروش» انجام شد که نتایج به شرح جدول ۴ ارائه شده است. نتایج نشان می‌دهد که مدل در این حالت عملکرد بسیار مطلوبی دارد؛ به‌گونه‌ای که میزان صحت<sup>۱</sup> کلی برابر با  $96/60$  درصد به دست آمده است. همچنین مقادیر دقت<sup>۲</sup> و فراخوانی<sup>۳</sup>، به‌ترتیب  $89/15$  و  $90/82$  درصد بوده است که بیانگر تعادل مناسب میان دقت

1. Accuracy  
2. Precision  
3. Recall

پیش‌بینی و توان بازیابی نمونه‌های واقعی است. افزایش چشمگیر صحت مدل در حالت دوکلاسه نسبت به حالت سه‌کلاسه نشان می‌دهد که حذف کلاس خنثی، موجب تمرکز بیشتر مدل بر سیگنال‌های معاملاتی شده است.

جدول ۴. ماتریس درهم‌ریختگی

predict class				
	B	N	S	
۸۳/۲۸%	۳۸۶	۲۸۱	۳۸۲۰	S
۹۰/۶۴%	۱۸۱۸	۲۲۳۳۰	۴۸۸	N
۹۰/۱۷%	۲۰۳۵۳	۱۷۵۵	۴۶۵	B
۸۹/۷۸%	۹۰/۲۳%	۹۱/۲۷%	۸۰/۰۳%	
actual class				
predict class				
	B	S		
	۹۰/۸۲%	۳۸۶	۳۸۲۰	S
	۹۷/۷۷%	۲۰۳۵۳	۴۶۵	B
	۹۶/۶۰%	۹۸/۱۴%	۸۹/۱۵%	
actual class				

### تحلیل همبستگی سیگنال‌ها و ارزش معاملات

پس از اجرای مدل طبقه‌بندی بر داده‌های مدنظر و استخراج تعداد روزانه سیگنال‌های خرید، فروش و پیام‌های خنثی، به‌منظور بررسی رابطه بین خروجی‌های مدل و متغیر بازار، از ضریب همبستگی پیرسون<sup>۱</sup> استفاده شده است. در این مرحله، رابطه میان تعداد سیگنال‌های خرید، تعداد سیگنال‌های فروش، پیام‌های خنثی، تفاضل سیگنال‌های خرید و فروش و نیز ارزش معاملات روزانه محاسبه شد. نتایج در قالب ماتریس همبستگی در جدول ۵ ارائه شده است.

جدول ۵. ماتریس ضرایب همبستگی

ارزش معاملات	تفاضل تعداد سیگنال خرید و فروش	تعداد سیگنال خرید	تعداد سیگنال خنثی	تعداد سیگنال فروش	
				۱/۰۰	تعداد سیگنال فروش
			۱/۰۰	۰/۶۹	تعداد سیگنال خنثی
		۱/۰۰	۰/۷۶	۰/۹۰	تعداد سیگنال خرید
	۱/۰۰	۰/۹۷	۰/۷۴	۰/۷۶	تفاضل خرید و فروش
۱/۰۰	۰/۶۱	۰/۷۲	۰/۵۸	۰/۸۰	ارزش معاملات

همان‌طور که از ماتریس ضرایب همبستگی برداشت می‌شود، بیشترین همبستگی با ۸۰ درصد همبستگی، بین تعداد سیگنال‌های فروش و ارزش معاملات است و در مقابل، پیام‌های خنثی یا خبری با ارزش معاملات همبستگی ضعیف‌تری (۵۸ درصد) دارد که می‌تواند از ماهیت غیرسیگنال‌محور این دسته از پیام‌ها نشئت گرفته باشد. به بیان

ساده‌تر، هرگاه تعداد پیام فروش افزایش یابد، ۸۰ درصد احتمال دارد که ارزش معاملات نیز افزایش پیدا کند؛ اما در رابطه با سیگنال خرید، این عدد ۷۲ درصد است. تفاضل تعداد سیگنال‌ها نیز ۶۱ درصد و نسبت به هر یک از سیگنال‌های خرید و فروش به تنهایی کمتر است؛ بنابراین این گونه استنباط می‌شود که سیگنال‌های صادر شده از سوی کاربران در کانال‌های پیام‌رسان تلگرام با ارزش معاملات (شاخصی از پویایی بازار) با درصد قابل قبولی هم‌بستگی دارد و بر آن اثر گذار است.

### جمع‌بندی و نتیجه‌گیری

پژوهش حاضر با هدف بررسی تأثیر سیگنال‌های منتشرشده در شبکه‌های اجتماعی بر ارزش معاملات روزانه بورس اوراق بهادار تهران انجام شد. با توجه به متداول بودن استفاده از پیام‌رسان تلگرام در بین فعالان بازار سرمایه، این شبکه اجتماعی برای نمونه پژوهش انتخاب شد. در این راستا با بهره‌گیری از روش‌های داده‌کاوی، متن‌کاوی و الگوریتم‌های یادگیری ماشین در علم هوش مصنوعی از جمله الگوریتم نایو بیس، ماشین بردار پشتیبان و برت در زبان برنامه‌نویسی پایتون، مدلی ایجاد شد که بتواند پیام‌ها و محتواهای ورودی را بگیرد و آن‌ها را در سه دسته سیگنال خرید، سیگنال فروش و خنثی دسته‌بندی و طبقه‌بندی کند. ارزش معاملات روزانه کل بازار، به‌عنوان متغیر وابسته از معاملات سهام در نظر گرفته شد که خود مجموعه‌ای از سهام بازار بورس و بازار فرابورس و حق تقدم‌های آن‌هاست.

نتایج ارزیابی مدل نشان داد که الگوریتم طراحی شده در حالت دوکلاسه (خرید/فروش) دقت زیادی دارد (صحت = ۹۶/۶ درصد) و شاخص‌های دقت و فراخوانی نیز بیانگر عملکرد متوازن و قابل اتکای آن در تشخیص سیگنال‌های معاملاتی است. به‌ویژه عملکرد بسیار خوب مدل در شناسایی سیگنال‌های فروش (دقت حدود ۹۸ درصد و فراخوانی حدود ۹۷ درصد) نشان‌دهنده وضوح الگوی زبانی این دسته از پیام‌ها و کارایی مناسب فرایند پیش‌پردازش و مدل‌سازی است. این یافته تأیید می‌کند که داده‌های غیرساختاریافته شبکه‌های اجتماعی، قابلیت تبدیل به داده‌های کمی و تحلیلی را دارند.

در مرحله بعد، تحلیل هم‌بستگی پیرسون نشان داد که میان تعداد سیگنال‌های فروش و ارزش معاملات روزانه رابطه مثبت و نسبتاً قوی وجود دارد (۸۰ درصد هم‌بستگی). همچنین تعداد سیگنال‌های خرید و تفاضل سیگنال‌های خرید و فروش نیز با ارزش معاملات هم‌بستگی مثبت و معنادار نشان دادند. این نتایج بیانگر آن است که افزایش فعالیت سیگنال‌دهی در کانال‌های تلگرامی با افزایش سطح معاملات بازار همراه است و میان پویایی فضای رسانه‌ای و رفتار معاملاتی سرمایه‌گذاران رابطه‌ای آماری وجود دارد.

یافته‌های پژوهش حاضر با نتایج مطالعات پیشین در حوزه تأثیر رسانه‌های اجتماعی بر بازارهای مالی هم‌سو است. پژوهش‌های داخلی و خارجی نشان داده‌اند که محتوای تولیدشده در شبکه‌های اجتماعی می‌تواند بر انتظارات سرمایه‌گذاران، احساسات بازار و در نهایت، رفتار معاملاتی اثرگذار باشد. با این حال، نوآوری پژوهش حاضر در چند بُعد درخور توجه است: نخست، استفاده از مدل‌های یادگیری ماشین برای استخراج سیگنال‌های معاملاتی از داده‌های فارسی؛ دوم، تمرکز بر ارزش معاملات به‌عنوان شاخص پویایی بازار (به‌جای صرفاً بازده یا قیمت) و سوم، تحلیل داده‌های یک سال کامل معاملاتی در بستر بازار سرمایه ایران.

از جنبه نظری، نتایج این پژوهش در چارچوب نظریه‌های مالی رفتاری و نقش اطلاعات غیررسمی در شکل‌گیری تصمیم‌های سرمایه‌گذاری قابل تبیین است. شبکه‌های اجتماعی می‌توانند به‌عنوان بستر انتقال سریع اطلاعات و شکل‌گیری انتظارات جمعی عمل کنند و از این طریق بر حجم معاملات و تحرک بازار تأثیر بگذارند. با این حال، باید تأکید کرد که یافته‌های پژوهش، گویای وجود رابطه هم‌بستگی است و نمی‌توان بر اساس آن، رابطه علی‌قطعی‌ای میان سیگنال‌های تلگرامی و تغییرات ارزش معاملات استنتاج کرد.

از جنبه کاربردهای مدیریتی و نظارتی، یافته‌های این پژوهش می‌تواند برای سازمان بورس و نهادهای ناظر در جهت پایش فعالیت‌های سیگنال‌دهی در شبکه‌های اجتماعی مورد استفاده قرار گیرد. با توجه به نقش چشمگیر این سیگنال‌ها در افزایش فعالیت معاملاتی، توسعه ابزارهای تحلیلی مبتنی بر داده‌کاوی می‌تواند به شناسایی رفتارهای غیرعادی، کاهش احتمال دست‌کاری بازار و بهبود شفافیت اطلاعاتی کمک کند. همچنین، برای سرمایه‌گذاران، توجه به ماهیت و اعتبار سیگنال‌های منتشرشده در شبکه‌های اجتماعی، می‌تواند در اتخاذ تصمیم‌های آگاهانه‌تر مؤثر باشد. در سطح سیاست‌گذاری نیز، نتایج این پژوهش می‌تواند مبنایی برای تدوین چارچوب‌های نظارتی در حوزه انتشار سیگنال‌های معاملاتی و مدیریت جریان اطلاعات در بازار سرمایه قرار گیرد.

### محدودیت‌های پژوهش

اولین محدودیت پژوهش، به‌طور قطع فیلتر بودن پیام‌رسان تلگرام است که در گرفتن محتوا موانعی ایجاد می‌کند. محدودیت بعدی، ادبیات متفاوت افراد در تولید محتواست که یادگیری ماشین و ایجاد مدل کارا و دقیق و صحیح را با مشکل مواجه می‌سازد؛ زیرا ابزارهای موجود برای یادگیری ماشین، روی زبان فارسی به‌خوبی زبان انگلیسی کار نمی‌کنند و به پیش‌پردازش‌های فراوان و دقیق نیاز دارند. همچنین تمرکز پژوهش بر یک شبکه اجتماعی و یک شاخص معاملاتی، ممکن است تعمیم نتایج را به سایر بسترها و متغیرها محدود سازد.

### پیشنادهایی برای پژوهش‌های آتی

۱. استفاده از الگوریتم‌های پیشرفته‌تر یادگیری ماشین و مدل‌های عمیق‌تر مانند الگوریتم SVC<sup>۱</sup> در حوزه پردازش زبان طبیعی برای بهبود دقت طبقه‌بندی پیام‌ها.
۲. بررسی تأثیر سیگنال‌های شبکه‌های اجتماعی بر سایر شاخص‌های بازار سرمایه نظیر بازده سهام، نوسان‌های قیمتی، شاخص کل یا عملکرد صنایع مختلف.
۳. تحلیل تأثیر اخبار کلان اقتصادی (مانند تغییر نرخ ارز، سیاست‌های پولی یا تصمیمات دولتی) بر روند معاملات و مقایسه آن با اثر سیگنال‌های شبکه‌های اجتماعی.
۴. طبقه‌بندی پیام‌ها بر اساس منبع تولید یا انتشار (اشخاص حقیقی، کارگزاری‌ها، نهادهای مالی و غیره) و بررسی میزان اثرگذاری هر گروه بر رفتار بازار.

## منابع

- ابراهیمیان، کامل؛ عباسی، ابراهیم؛ عالم‌تبریز، اکبر و محمدزاده، امیر (۱۴۰۰). پیش‌بینی روند روزانه قیمت سهام با استفاده از متن‌کاوی احساسات کاربران شبکه اجتماعی و داده‌کاوی نماگرهای تکنیکال. *فصلنامه دانش سرمایه‌گذاری*، ۱۰(۴۰)، ۱-۲۶.
- الوندی، پدرام؛ خانیکی، هادی و اکبرزاده جهرمی، سید جمال (۱۴۰۱). پلتفرمی‌شدن و بازتعریف اکوسیستم رسانه‌های خبری. *جامعه فرهنگ رسانه*، ۱۱(۴۴)، ۱۱-۴۱.
- پیرایی، خسرو و شهسوار، محمدرضا (۱۳۸۸). بررسی تأثیر متغیرهای کلان اقتصادی بر شاخص قیمت سهام در ایران. *پژوهش‌های اقتصادی ایران*، ۱۱۵-۱۳۶.
- راعی، رضا؛ حسینی، سید فرهنگ و کیانی هرچگانی، مانده (۱۳۹۵). بررسی توانایی نظرات کاربران شبکه‌های اجتماعی بر پیش‌بینی جهت و قیمت سهام در بورس اوراق بهادار تهران. *فصلنامه دانش سرمایه‌گذاری*، ۵(۱۹)، ۱۰۷-۱۲۸.
- عباسی، مهدی (۱۳۹۵). *طبقه‌بندی توئیت‌های فارسی شبکه اجتماعی توئیتر با استفاده از روش متن‌کاوی*. پایان‌نامه دوره کارشناسی ارشد. دانشگاه علامه طباطبائی.
- عزیزی، زهرا؛ عبدالوند، ندا؛ قالیباف‌اصل، حسن و رجائی هرندی، سعیده (۱۳۹۹). بررسی تأثیر اخبار فارسی بر بازدهی سهام با استفاده از تکنیک‌های متن‌کاوی. *مجله ایرانی مطالعات مدیریت*، ۱۴(۴)، ۷۹۹-۸۱۶.
- فروغی مجد، فرزاد (۱۳۹۸). *ارزیابی تأثیر نظرات تحلیلگران مالی رسانه‌های اجتماعی در پیش‌بینی روند حرکت سهام*. پایان‌نامه دوره کارشناسی ارشد. دانشگاه خوارزمی.
- کشاوری‌حداد، غلامرضا و حیدری، هادی (۱۳۹۰). بررسی تأثیر اخبار سیاسی بر تلاطم بازار سهام تهران (مقایسه مدل‌های عمومی MSM و FAGARCH). *فصلنامه تحقیقات اقتصادی*، ۴۶(۱)، ۱۱۱-۱۳۵.
- هاتفی قهفرخی و شمس‌فرد (۱۳۹۹). پیش‌بینی بورس اوراق بهادار تهران با استفاده از تحلیل احساسات نظرات متنی آنلاین. *سیستم‌های هوشمند در حسابداری، مالی و مدیریت*، ۲۷(۱)، ۳۷-۲۲.
- هیبتی، فرشاد و زندیه، وحید (۱۳۹۰). *بیش‌واکنش سرمایه‌گذاران بازار سهام ایران به اخبار بحران مالی جهانی*. *دانش مالی تحلیل اوراق بهادار (مطالعات مالی)*، ۴(۹)، ۷۵-۱۰۰.

## References

- Abbasi, M. (2016). Classification of Persian social network tweets using text mining method. *End quote Senior Undergraduates Allameh Tabataba'i University*. <https://ganj.irandoc.ac.ir/#/articles/7f4f8e0bcd0255a404e6427f4d14fe2/fulltext> (in Persian)
- Alvandi, P., Khaniki, H. & Akbarzadeh Jahromi, S. J. (2023). Platformization and Redefining of news media ecosystem. *Society Culture Media*, 11(44), 11-41. <https://dor.isc.ac/dor/20.1001.1.38552322.1401.11.44.1.9> (in Persian)

- Azizi, Z., Abdolvand, N., Ghalibaf Asl, H. & Rajaei Harandi, S. (2021). The Impact of Persian News on Stock Returns Through Text Mining Techniques. *Interdisciplinary Journal of Management Studies*, 14(4), 799-816. [https://ijms.ut.ac.ir/article\\_80598\\_600bf9e0bd2f46554d6d6487cf0d4473.pdf](https://ijms.ut.ac.ir/article_80598_600bf9e0bd2f46554d6d6487cf0d4473.pdf) (in Persian)
- Benthaus, J. & Beck, R. (2015). *It's more about the content than the users! The influence of social broadcasting on stock markets.*
- Bentolila, A. & Thompkins, T. (2022). A Comparative Analysis Between the Influence of Social Media and Mass Media on the Stock Market. *Journal of Student Research*, 11(3). <https://doi.org/10.47611/jsrhs.v11i3.3430>
- Boyd, D. M. & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- Davis, A. K., Piger, J. M. & Sedor, L. M. (2012). Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research*, 29(3), 845–868. <https://doi.org/10.1111/j.1911-3846.2011.01130.x>
- Derakhshan, A. & Beigy, H. (2019). Sentiment analysis on stock social media for stock price movement prediction. *Engineering Applications of Artificial Intelligence*, 85, 569–578. <https://doi.org/10.1016/j.engappai.2019.07.002>
- Diaz, A. & Jareno, F. (2009). Explanatory factors of the inflation news impact on stock returns by sector: the Spanish case. *Research in International Business and Finance*, 23(3), 349–368. <https://doi.org/10.1016/j.ribaf.2008.10.006>
- Ebrahimian, K., Abbasi, E., Alam Tabriz, A. & Mohammadzadeh, A. (2021). Daily Stock Price Movement Prediction Using Sentiment text mining of social network and data mining of Technical indicators. *Journal of Investment Knowledge*, 10, (40), 451-469. <https://sanad.iau.ir/fa/Journal/jik/DownloadFile/843047> (in Persian)
- Fruoghimajd, F. (2020). *Evaluation of the impact of social media financial analysts' comments on the trend of stock movement.* Thesis of senior undergraduate study. Kharazmi University. (in Persian)
- Han, J., Kamber, M. & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). *Morgan Kaufmann*. <https://doi.org/10.1016/C2009-0-61819-5>
- Hatefi Ghahfarrokhi, A. & Shamsfard, M. (2020). Tehran stock exchange prediction using sentiment analysis of online textual opinions. *Intelligent Systems in Accounting, Finance and Management*, 27(1), 22-37. <https://doi.org/10.1002/isaf.1465>
- Heybati, f., Zandiyeh, V. (2011). Investors' Overreaction in IRAN Stock Market to Global Financial Crisis News.” *Financial Knowledge of Securities Analysis*, 4, (9), 75-100. <https://sanad.iau.ir/fa/Journal/jfksa/Article/803276> (in Persian)
- Jiang, T. & Zeng, A. (2023). Financial sentiment analysis using FinBERT with application in predicting stock movement. *arXiv*. <https://doi.org/10.48550/arXiv.2306.02136>
- Kaplan, A. M. & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59–68. <https://doi.org/10.1016/j.bushor.2009.09.003>

- Keshavarz, G. & Heidari, H. (2011). The Impact of Political News on Tehran Stock Exchange (AFIGARCA and MSM) Approach. *Journal of Economic Research (Tahghighat- E-Eghtesadi)*, 46(1), 111-135. [https://jte.ut.ac.ir/article\\_22448.html](https://jte.ut.ac.ir/article_22448.html) (in Persian)
- Kietzmann, J. H., Hermkens, K., McCarthy, I. P. & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, 54(3), 241–251. <https://doi.org/10.1016/j.bushor.2011.01.005>
- Loughran, T. & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187–1230. <https://doi.org/10.1111/1475-679X.12123>
- Maqbool, J., Aggarwal, P., Kaur, R., Mittal, A. & Ganaie, I. A. (2023). Stock prediction by integrating sentiment scores of financial news and MLP-regressor: A machine learning approach. *Procedia Computer Science*, 218, 1067–1078. <https://doi.org/10.1016/j.procs.2023.01.086>
- Mehta, P., Pandya, S. & Doshi, N. (2020). Sentiment analysis on social media and stock price prediction using machine learning and deep learning models. *International Journal of Information Management Data Insights*, 1(2), 100017. <https://doi.org/10.1016/j.jjime.2020.100017>
- Nti, I. K., Adekoya, A. F. & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, 53(4), 3007–3057. <https://doi.org/10.1007/s10462-019-09754-z>
- Önder, Z. & Şimşak-Mugan, C. (2006). How do political and economic news affect emerging markets? Evidence from Argentina and Turkey. *Emerging Markets Finance and Trade*, 42(4), 50–77. <https://doi.org/10.2753/REE1540-496X420403>
- Padmanayana, V. & Bhavya, K. (2021). Stock market prediction using Twitter sentiment analysis. *International Journal of Scientific Research in Science and Technology*, 7(4), 265–270. <https://doi.org/10.32628/CSEIT217475>
- Pang, X., Zhou, Y., Wang, P., Lin, W. & Chang, V. (2020). An innovative neural network approach for stock market prediction. *Journal of Supercomputing*, 76(3), 2098–2118. <https://doi.org/10.1007/s11227-017-2228-y>
- Pearce, D. K. & Roley, V. V. (1985). Stock prices and economic news. *The Journal of Business*, 58(1), 49–67. <https://doi.org/10.1086/296282>
- Pirae, K., Shahsavari, M.R. (2010). The impacts of macroeconomic variables on Iranian stock market. *Iranian Journal of Economic Research*. 115-136. <https://ensani.ir/fa/article/83697/> (in Persian)
- Raie, R., Farhang Hoseini, S., Kiani Harchegani, M. (2016). Evaluate the Ability of Social Networks to Predict the Direction and Stock Prices in Tehran Stock Exchange. *Journal of Investment Knowledge*, 5, (19), 107-128. <https://www.magiran.com/paper/1592175> (in Persian)
- Safko, L., Brake, D. K. (2009). The Social Media Bible: Tactics, Tools, and Strategies for Business Success. *Ukraine: Wiley*.

- Smith, S. & O'Hare, A. (2022). Comparing traditional news and social media with stock price movements; which comes first, the news or the price change? *Journal of Big Data*, 9(1), 1-20. DOI: <https://doi.org/10.1186/s40537-022-00591-6>
- Sprenger, T. O. & Welpe, I. M. (2011). News or noise? The Stock Market Reaction to Different Types of Company-Specific News Events (January 4, 2011).
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62(3), 1139–1168. <https://doi.org/10.1111/j.1540-6261.2007.01232.x>
- Tetlock, P. C., Saar-Tsechansky, M. & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance*, 63(3), 1437–1467. <https://doi.org/10.1111/j.1540-6261.2008.01362.x>
- Wang, J.L. & Chan, S.H. (2007). Stock market trading rule discovery using pattern recognition and technical analysis. *Expert Systems with Applications*, 33(2), 304–315. <https://doi.org/10.1016/j.eswa.2006.05.002>
- Wang, Z., Hu, Z., Li, F., Ho, S.-B. & Cambria, E. (2021). Learning-based stock trending prediction by incorporating technical indicators and social media sentiment. *Cognitive Computation*, 15(3), 1092–1102.
- Wu, X., Wang, X., Ma, S. & Ye, Q. (2017). The influence of social media on stock volatility. *Frontiers of Engineering Management*, 4(2), 201–211. <https://doi.org/10.15302/J-FEM-2017018>
- Zhang, H., Chen, Y., Rong, W., Wang, J. & Tan, J. (2022). Effect of social media rumors on stock market volatility: A case of data mining in China. *Frontiers in Physics*, 10, 987799. DOI: <https://doi.org/10.3389/fphy.2022.987799>