Securities & Exchange Organization, Research, Development & Islamic Studies (RDIS)
Journal of Securities and Exchange, Fall 2025, V. 18, No.71, pp. 309-358

Stock Price Forecasting with Sentimental Analysis of Social Media Data[1]

Soroush Khajeh Hagh Verdi[2], Shapour Mohammadi[3], Reza Raei[4]

Receive Date: 08 January 2022 Revise Date: 07 March 2022
Accept Date: 12 February 2022 Publish Date: 23 October 2025

Research Paper

Highlights

- The study analyzes 15,736,204 Telegram messages and matches them with 19,312 symbol-day stock return records to evaluate whether social-media sentiment can predict stock performance.
- Unlike previous research, this work separates messages by content type, by sender importance, and by message importance (views) to improve sentiment-based forecasting.
- Six neural-network modeling approaches were trained on 36 companies in the Tehran Stock Exchange 30-Company Index, using 6–10 hidden neurons and 1–5 lags.
- Results show that models distinguishing influential senders and high-visibility messages significantly outperform all other models in predicting stock-price direction.
- Directional forecasting accuracy of the enhanced model reaches about 69%, roughly 9% higher than models using only message counts or unspecialized sentiment.
- In return-value forecasting, the enhanced model also achieves significantly lower RMSE, confirming superior quantitative prediction.
- The authors recommend using real-time sentiment systems and dynamic sender/message importance ranking to support algorithmic and high-frequency trading.

Abstract

Social media has an important rule in desion making. In recent years, Telegram groups and channels become considerable parameter in user's opinion. Which make people pay more attention to the content of the media. Previous researches show the importance of sentiments in forecasting models and researches find significant relation between sentimental analysis and stock price prediction. They consider buy, sell and hold contents as independent variable. But in reality, there is a difference between people who publish their opinion in social media. Means people consider the speaker's background and also for the viral messages there are same situation. In this research we consider importnace of the publisher and message in social media. Then we separate content into different variables. We analyse 15.736.204 Telegram messages and according to our sentimental dictionary, we tag them then put merge them with 19.312 records related to symbols in different days. Selected symbols are the stocks in 30 company index portfolio and the duration of survey is from march 2019 to august 2021. Six variable selection approach was considered in this survey, for each symbol (among 36 selected companies) we train different artificial neural network which the number of neurons in hidden layer varies from 6 to 10 and the lag of data for input variable differs from 1 to 5. In order to remove initial values effect in neural network, each network runs 10 times. Results showed binary prediction in a model which separates messages by importance of people and amount of view has better performance than other models and RMSE error in that model for return forecasting significantly lower than other models.

Key Words: Social Media, Price Forecasting, Sentimental Analysis, Artificial Neural Network.
JEL Classification: G12.

---

1. doi: 10.22034/JSE.2022.11900.1933
2. Ph.D. Student, Faculty of Management, Tehran University, Tehran, Iran. (Corresponding Author). (haghverdi.314@ut.ac.ir).
3. Professor, Faculty of Management, Tehran University, Tehran, Iran. (shmohmad@ut.ac.ir).
4. Professor, Faculty of Management, Tehran University, Tehran, Iran. (raei@ut.ac.ir).

Copyright © 2025 The Authors. Published by Securities and Exchange Organization. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (https://creativecommons.org/licenses/by-nc/4.0/). Non-commercial uses of the work are permitted, provided the original work is properly cited.

Cite This Article: Khajeh Hagh Verdi, Soroush; Mohammadi, Shapour; Raei, Reza. (2025). Stock Price Forecasting with Sentimental Analysis of Social Media Data. ***Journal of Securities Exchange***, 18 (71), 309-358. http://10.22034/JSE.2022.11900.1933.

Introduction

Social networks have become deeply embedded in everyday communication, information exchange, and collective decision-making, reshaping how individuals interact and form opinions. The attached article emphasizes that networks such as Telegram play an increasingly influential role in Iran's capital market, where investors—both novice and experienced—regularly consult social-media channels for insights, interpretations, and decision signals related to stock trading.

Prior studies showed that social-media sentiment holds predictive power for stock returns, but these works typically relied on a simplified treatment of message counts classified merely into "buy," "sell," or "hold." The article argues that such an approach fails to reflect real market behavior because, in practice, the identity of the message sender and the visibility of the message strongly influence how investors react. Certain senders—due to reputation, expertise, or follower count—exert disproportionate influence. Likewise, messages that receive high numbers of views or reposts shape collective sentiment more strongly than low-visibility messages. Thus, sentiment cannot be treated merely as a uniform stream of opinions; instead, it must be weighted and segmented to reflect realistic dynamics of information diffusion and investor response.

The article therefore positions its contribution in refining sentiment analysis for financial forecasting. It uses a massive dataset of more than 15.7 million Telegram messages, processed through linguistic normalization, tokenization, morphological analysis, and lexicon-based sentiment scoring. Messages are categorized into four content types-fundamental analysis, technical analysis, buy signals, and sell signals-while also being assigned importance indicators based on the sender's influence level and message view counts. The goal is to determine whether these additional layers of information meaningfully improve the predictive power of artificial neural networks in forecasting both the direction (up or down) and magnitude (return value) of stock-price movements.

The study seeks to answer core research questions:

1. What type of content dominates Telegram discussions about the market?
2. Which types of messages attract the most user attention?
3. Does incorporating sender importance and message visibility significantly enhance predictive models relative to traditional sentiment measures?

4. Do different content types (analysis vs. signals) yield different predictive strengths?

The research builds on the premise that markets are shaped not only by fundamentals and macroeconomic news but also by the behavioral patterns and emotions of investor communities, which social media reveals in real time. It situates itself within behavioral finance, sentiment analysis, and machine learning, drawing from theoretical frameworks on investor sentiment and empirical studies linking collective mood to market fluctuations.

Methodology

The dataset spans from March 2019 to August 2021 and contains messages collected through the Telegram API. The researchers began by filtering only messages mentioning any of the 30-Index companies or general market terms, ensuring relevance to the stock universe under investigation. Messages were further filtered to exclude irrelevant or conversational noise.

Text Processing Pipeline

To prepare the messages for automated analysis, the study applies a multi-level natural-language-processing pipeline using the HAZM toolkit:

1. Text segmentation: Messages are broken into sentences for finer sentiment scoring.
2. Normalization: Persian textual inconsistencies—letter variations, compound forms, affixes—are standardized.
3. Morphological tagging: Each word is labeled with its part of speech, enabling accurate detection of negations affecting sentiment polarity.
4. Stemming / root extraction: Verbs are reduced to their roots to distinguish inherently negative forms (e. g; verbs beginning with "ن").
5. Sentiment scoring: A custom lexicon assigns positive or negative scores to words; negation rules modify sentence-level polarity. Sentence scores aggregate into message-level sentiment values.

Content Classification

Each message is categorized into four types using keyword-based scoring:

- Fundamental analysis
- Technical analysis
- Buy signal
- Sell signal

A message may receive partial credit for multiple categories, but its highest-scoring category determines its classification. Messages lacking sufficient relevant keywords are excluded from later modeling.

Sender and Message Importance Segmentation

A central innovation of the study is segmenting messages based on:

1. Sender importance: Senders whose messages rank in the top decile of average monthly views are considered influential.
2. Message importance: Messages that individually fall in the top decile of view counts each month are treated as high-impact messages.

This produces a refined sentiment dataset where each observation is weighted by both content type and perceived market impact through sender/message prominence.

Merged Financial Dataset

The sentiment datasets are merged with daily financial variables for each symbol:

- Daily log returns
- Opening auction returns
- Tehran exchange market index returns
- Volatility of currency, oil, gold, copper, urea, aluminum, zinc, iron ore, methanol
- Sentiment variables (as constructed above)

This results in 19,312 symbol-day observations.

Neural Network Modeling

The researchers use multilayer perceptron (MLP) neural networks with:

- Hidden layer neurons: 6, 7, 8, 9, 10
- Lags: 1 to 5 for all input variables
- Training/testing/validation split: 70% training, 15% testing, 15% validation

- Activation function: hyperbolic tangent
- Training algorithm: Levenberg–Marquardt
- Repetition: each model configuration is run 10 times to avoid initialization bias

Overall, 108,000 neural networks are trained across six major model groups:

1. Full sentiment + sender importance + message importance
2. Sentiment with content categories but excluding "other" categories
3. Sentiment classified only into four content types
4. Sentiment simplified to positive/negative/neutral
5. Raw message counts only
6. No sentiment variables (historical/market variables only)

The first two models uniquely incorporate sender/message importance.

Evaluation Metrics

- Directional accuracy (binary classification of up/down)
- RMSE for return forecasts (continuous predictions)
- Paired t-tests to test statistical significance of model performance differences

Results

Content Distribution and User Attention

Analyses reveal that approximately 60–66% of Telegram market messages are signals, not analyses—making them the dominant communication form. This pattern holds across channels, super-groups, and all time segments of the trading day.

Messages containing buy signals receive significantly more attention (views) than sell signals, while fundamental analyses receive more attention than technical ones. These patterns are statistically significant and consistent across multiple paired-mean tests. Investors appear to favor actionable content and fundamental reasoning over chart-based commentary.

Market Regime Behavior

The study tests whether attention to message types varies under bullish, bearish, or neutral market regimes. Surprisingly, there is no statistically significant difference: users do not systematically shift their attention

distribution across regimes, implying stable attention patterns independent of market direction.

Directional Forecasting Results

Model 1—incorporating full sentiment segmentation plus sender and message importance—achieves:

- Mean directional accuracy: ~0.693
- ~9% higher accuracy than models lacking these refinements
- Significantly better quartile distribution of prediction values
- Lower variance, indicating more stable forecasting performance

Paired t-tests confirm significance at the 1% level for all comparisons between Model 1 and Models 3–6. Similarly, Model 2 also outperforms all baseline models, though slightly below Model 1.

Return Forecasting Results

Model 1 yields the lowest RMSE (~0.0246) among all models, outperforming others by margins of 0.0024–0.0031, all statistically significant. Model 2 again performs similarly but slightly worse. Kernel density plots show Model 1's error distribution is shifted left (lower error) with thicker concentrations around minimal-error regions.

Discussion

The findings strongly support the article's central hypothesis: message sender importance and message visibility materially enhance predictive power. This aligns with real-world investment behavior, where markets respond more strongly to messages from trusted sources or viral posts that rapidly influence sentiment.

The dominance of signal messages and the higher attention toward buy signals reflect behavioral biases such as optimism, herd behavior, and promotional activity, all of which the sentiment models implicitly capture. The fact that users maintain consistent attention patterns across market regimes suggests that social-media behavior is relatively independent of macro-market sentiment and may act as an orthogonal information stream.

The superior performance of the enhanced models demonstrates that incorporating information hierarchy within social-media data is essential. Treating all messages equally ignores critical variations in influence and visibility, leading to diluted models and weaker predictive performance.

Furthermore, the successful integration of lexicon-based sentiment with machine-learning forecasting confirms that structured behavioral signals can significantly complement traditional financial variables.

Conclusions

The study provides compelling evidence that refined sentiment analysis, which distinguishes between message types, sender influence, and message visibility, significantly improves stock-return forecasting for companies in the Tehran Stock Exchange 30-Company Index.

Key conclusions include:

1. Signal messages dominate Telegram's market-related content.
2. Investors pay more attention to fundamental analyses and buy signals.
3. Market-regime shifts do not significantly alter attention behavior.
4. Incorporating message/sender importance produces statistically superior forecasting models for both direction and magnitude.
5. Neural networks exploiting refined sentiment variables outperform all baselines, including models with only historical/market data.
6. These methods hold practical value for algorithmic trading, real-time sentiment monitoring, and potentially high-frequency trading (HFT) given adequate infrastructure.

The article recommends developing real-time sentiment-analysis systems and dynamic decile-based classification of sender/message importance to further enhance prediction accuracy.

Author Contributions

Conceptualization, Soroush Khajeh Hagh Verdi; methodology, Soroush Khajeh Hagh Verdi; software, Soroush Khajeh Hagh Verdi and Shapour Mohammadi; validation, Soroush Khajeh Hagh Verdi; formal analysis, Soroush Khajeh Hagh Verdi and Shapour Mohammadi; investigation, Soroush Khajeh Hagh Verdi; resources, Soroush Khajeh Hagh Verdi; data curation, Soroush Khajeh Hagh Verdi; writing—original draft preparation, Soroush Khajeh Hagh Verdi; writing—review and editing, Soroush Khajeh Hagh Verdi, Shapour Mohammadi and Reza Raei; visualization, Soroush Khajeh Hagh Verdi; supervision, Soroush Khajeh Hagh Verdi, Shapour Mohammadi and Reza Raei; project administration, Soroush Khajeh Hagh Verdi.

All authors contributed equally to the conceptualization of the article and writing of the original and subsequent drafts.

Data Availability Statement

Data available on request from the authors.

Acknowledgements

The authors would like to thank all participants in the present study.

Ethical Considerations

The authors avoided data fabrication, falsification, and plagiarism, and any form of misconduct.

Conflict of Interest

The authors declare no conflict of interest.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

سازمان بورس و اوراق بهادار، مرکز پژوهش، توسعه و مطالعات اسلامی
فصلنامه بورس اوراق بهادار، سال هجدهم، شماره ۷۱، پاییز ۱۴۰۴، صص ۳۵۸-۳۰۹

پیش‌بینی قیمت سهام با استفاده از تحلیل محتوای احساسات شبکه‌های اجتماعی[۱]

سروش خواجه حق وردی[۲]، شاپور محمدی[۳]، رضا راعی[۴]

مقاله پژوهشی

تاریخ دریافت:۱۴۰۰/۱۰/۱۸ تاریخ بازنگری:۱۴۰۰/۱۰/۱۶
تاریخ پذیرش:۱۴۰۰/۱۱/۲۳ تاریخ انتشار:۱۴۰۴/۰۸/۰۱

نکات برجسته

- این پژوهش با تحلیل ۱۵٬۷۳۶٬۲۰۴ پیام تلگرامی و تطبیق آن با ۱۹٬۳۱۲ رکورد بازده روزانه نمادها بررسی می‌کند که آیا احساسات موجود در شبکه‌های اجتماعی قابلیت پیش‌بینی بازدهی سهام را دارند یا خیر.
- شش رویکرد مدل‌سازی با استفاده از شبکه عصبی برای ۳۶ شرکت موجود در پرتفوی شاخص ۳۰ شرکت اجرا شد و تعداد نورون‌های لایه پنهان بین ۶ تا ۱۰ و تعداد وقفه‌ها بین ۱ تا ۵ متغیر بود.
- نتایج نشان می‌دهد مدلی که در آن اهمیت اشخاص تأثیرگذار و پیام‌های پربازدید منظور شده است، در پیش‌بینی برای حرکت قیمت سهام به‌طور معناداری عملکرد بهتری نسبت به سایر مدل‌ها دارد. دقت پیش‌بینی برای قیمت این مدل حدود ۶۹ درصد گزارش شده که کمابیش ۹ درصد بیشتر از مدل‌هایی است که تنها بر تعداد پیام‌ها یا تحلیل احساسات ساده متکی‌اند.
- در پیش‌بینی بازده مقداری نیز مدل تقویت‌شده دارای کمترین خطای RMSE و عملکرد به‌طور معناداری برتر از مدل‌های جایگزین است.

چکیده

شبکه‌های اجتماعی نقش بسیار مهمی در شکل‌گیری ارتباطات، بحث، تبادل نظر و در نتیجه تصمیم‌گیری مبتنی بر خردجمعی دارند. در سال‌های اخیر گروه‌ها و کانال‌های تلگرامی مربوط به بازار سرمایه در ایران توانسته‌اند جایگاه ویژه‌ای نزد کاربران حوزه بازار سرمایه کسب کنند و الگوریتم‌های مختلفی برای استخراج احساسات موجود در فضای مجازی ایجاد شده تا با استفاده از نتایج آن‌ها بتوان اقدام به گرفتن تصمیمات سرمایه‌گذاری و نیز انجام معامله کرد؛ بر اساس آنچه در سایر پژوهش‌ها انجام شده است، احساسات موجود در شبکه‌های اجتماعی در پیش‌بینی بازده سهام موثر بوده و این عامل می‌تواند به عنوان یکی از عناصر قابل استفاده در پیش‌بینی‌ها مورد استفاده قرار گیرد. پژوهش‌های پیشین تنها از تعداد پیام‌های منتشرشده در جهت خرید، فروش یا نگهداری به عنوان عامل تحلیل احساسات استفاده می‌کرد. در دنیای امروز، گوینده پیام و همچنین اهمیت خود پیام نیز قابل تفکیک و دسته‌بندی است، بر اساس آنچه که در محیط واقعی کسب و کار بازار سرمایه رخ می‌دهد، فرستنده پیام، جایگاه ویژه‌ای پیدا می‌کند و با توجه به امکان بازنشر پیام‌ها، آنچه که دارای اهمیت باشد به سرعت به سایر افراد نیز اطلاع رسانی می‌شود پس تفکیک کردن پیام‌های منتشرشده از سوی اشخاص مهم و نیز پیام‌هایی که مورد توجه کاربران قرار می‌گیرند، به نظر می‌رسد بتواند در پیش‌بینی بازدهی سهام کمک کننده‌تر باشد. در خصوص محتوای احساسی، تعداد ۱۵,۷۳۶,۲۰۴ پیام کانال‌ها و گروه‌های تلگرامی که دارای محتوای قابل بررسی بودند، استخراج شد و پس از برچسب‌گذاری بر اساس لغتنامه احساسی، موارد در تعداد ۱۹,۳۱۲ داده مربوط به نماد-روز حاوی بازدهی و محتوای احساسی خلاصه شد. بدین منظور ۶ رویکرد داده‌ای مختلف برای ۳۶ نماد معاملاتی (نمادهای موجود در شاخص ۳۰ شرکت) در بازه زمانی ابتدای سال ۱۳۹۸ تا انتهای مرداد سال ۱۴۰۰ به صورت جهتی و پیش‌بینی مقداری با ۱ الی ۵ وقفه و ۶ الی ۱۰ نرون در لایه پنهان و ۱۰ مرتبه اجرا برای هر حالت مورد بررسی قرار گرفت. نتایج پیش‌بینی جهتی و بازدهی قیمت با استفاده از شبکه عصبی مصنوعی نشان داد مدلی که داده‌های مربوط به شخص ارسال کننده پیام (اشخاصی که متوسط مشاهده پیام‌های ارسالی ایشان در دهک اول قرار داشته باشد) و نیز نوع پیام (پیام‌هایی که در دهک اول میزان مشاهدات قرار داشتند) را تفکیک کرده در پیش‌بینی جهت به صورت معنی دار عملکرد بهتری نسبت به مدل‌های دیگر (سایر مدل‌ها: استفاده از محتوای ارسالی تنها به صورت جهتی، بر اساس نوع محتوا و همچنین نیز بدون در نظر گرفتن محتوای احساسی) داشته و نیز پیش‌بینی مقداری بازدهی در روش مورد نظر این پژوهش، به صورت معنی داری دارای خطای RMSE کمتری است.

واژه‌های کلیدی: تسلط اجتماعی، ماکیاولیسم، دستکاری هیجانی و بیش نمایی سود.
طبقه‌بندی موضوعی: G12.

---

۱. doi: 10.22034/JSE.2022.11900.1933
۲. دانشجوی دکتری، گروه مالی، دانشکده مدیریت، دانشگاه تهران، تهران، ایران. (نویسنده مسئول). (haghverdi.314@ut.ac.ir).
۳. استاد، گروه مالی و بیمه، دانشکده مدیریت، دانشگاه تهران، تهران، ایران. (shmohmad@ut.ac.ir).
۴. استاد، گروه مالی و بیمه، دانشکده مدیریت، دانشگاه تهران، تهران، ایران. (raei@ut.ac.ir).

حق انتشار این مستند، متعلق به نویسندگان آن است. ۱۴۰۴ ©. ناشر این مقاله، سازمان بورس و اوراق بهادار است.
این مقاله تحت گواهی زیر منتشرشده و هر نوع استفاده غیرتجاری از آن مشروط بر استناد صحیح به مقاله و با رعایت شرایط مندرج در آدرس زیر مجاز است.
Creative Commons Attribution-NonCommercial 4.0 International license
(https://creativecommons.org/licenses/by-nc/4.0/)

استناد: خواجه حق وردی، سروش؛ محمدی، شاپور؛ راعی، رضا. (۱۴۰۴). پیش‌بینی قیمت سهام با استفاده از تحلیل محتوای احساسات شبکه‌های اجتماعی. *فصلنامه بورس اوراق بهادار*، ۱۸ (۷۱)، ۳۵۸-۳۰۹.
https://10.22034/JSE.2022.11900.1933

مقدمه

شبکه اجتماعی مجموعه‌ای از ارتباطات است که در آن افراد بدون واسطه به یکدیگر متصل می‌شوند و با این روش تعداد چشمگیری از افراد به یکدیگر متصل می‌شوند. شبکه‌های اجتماعی می‌توانند وظایفی مانند تبادل اخبار و انجام گفت و گو را داشته باشند. این شبکه‌ها در بستر واسطه‌های ارتباطی صورت می‌گیرد(نیوبک و گلاسبرگ[1]، ۲۰۰۴). شبکه‌های اجتماعی تسهیل‌کننده مناسب برای افرادی است که به تازگی وارد بازار سرمایه شده‌اند که با این طریق اطلاعات مفیدی به دست آورند(شپارد[2]، ۲۰۱۲).

یک شبکه اجتماعی، سایت یا مجموعه سایتی است که به کاربران اجازه می‌دهد تا علاقه‌مندی‌ها، افکار و فعالیت‌های خود را با دیگران به اشتراک بگذارند و فضایی برای تبادل نظر داشته باشند که در آن‌ها افراد عقاید و نظرات خود را بیان می‌کنند.

مزایایی که شبکه‌های اجتماعی دارند باعث استقبال بسیار زیاد از آن‌ها شده است. شامل این موارد است: انتشار سریع و آزادانه اخبار، اطلاعات و شایعات که امکان انتشار اطلاعات نادرست بسیار زیاد است اما مخاطب می‌تواند مطالب را با هم مقایسه کند و حتی می‌تواند از این طریق قدرت نقد و تحلیل خود را افزایش دهد. البته شبکه‌های اجتماعی پیامدهای منفی هم دارند از قبیل شکل‌گیری و ترویج سریع شایعات و اخبار نادرست.

توسعه شبکه‌های اجتماعی به حدی رسیده است که شبکه اجتماعی بازار سرمایه با نام توییت سهام در اکتبر سال ۲۰۰۸ راه‌اندازی شده است. این شبکه اجتماعی زیر ساختی جداگانه از توییتر دارد و تنها مخصوص پست‌های مرتبط با بازار بورس است. در این محیط افراد بر اساس نماد شرکت‌ها، نظرات، تحلیل‌ها و خبرهای خود را منتشر می‌کنند. این شبکه اجتماعی با پلفرم سایر شبکه‌های اجتماعی نیز هماهنگ است و کاربران سایر شبکه‌های اجتماعی به محتوای آن دسترسی دارند. شبکه‌های اجتماعی بازار بورس، فضای اطلاع رسانی یک طرفه از سوی رسانه‌ها را به فضای دوطرفه مبدل نموده است و سرمایه‌گذاران می‌توانند نظرات و تحلیل‌ها و اخبار خود را به صورت لحظه‌ای منتشر و در اختیار سایر تحلیل‌گران و سرمایه‌گذاران قرار دهند.

---

1. Neubeck & Glasberg
2. Shepard

پیش‌بینی و برآورد آینده یکی از ضروریات تصمیم‌گیری است. اما مساله مهمتر این است که چگونه در دنیای عدم اطمینان، بتوان رویدادهای آتی را پیش‌بینی کرد (راعی و حسینی، ۱۳۹۵). بنابراین با توجه به اهمیت پیش‌بینی در کسب و کارهای مالی و یا حتی کسب و کارهایی که به صورت غیرمستقیم با مالی سروکار دارند، یافتن روشی با قابلیت اتکای بیشتر برای بهره‌برداری، بسیار سودمند خواهد بود.

دو روش قابل اعتماد برای پیش‌بینی شامل استفاده از سیستم‌های خبره و استفاده از خرد جمعی است (راعی و حسینی، ۱۳۹۵).

از آنجایی که معاملات در بازار برآیند نظرات خریداران و فروشندگان است و از طرفی با توجه به اینکه در شبکه‌های اجتماعی امروزه، هر فردی می‌تواند نظرات خود را ثبت کند، به نظر می‌رسد پیش از معاملات در جلسات معاملات آتی می‌توان به نظرات افراد در شبکه‌های اجتماعی دست یافت. در نتیجه ایجاد مدلی که بتواند نظرات افراد را در محاسبات قیمتی منظور کند، بسیار مفید خواهد بود. در این خصوص اقتصاد رفتاری بیان می‌دارد که احساسات عمیقاً می‌توانند بر رفتار فردی و تصمیم‌گیری تاثیرگذار باشد. از این دیدگاه، بررسی احساس و حال تعداد زیادی از افراد در شبکه‌های اجتماعی و شاخص داو جونز بررسی شده است. بر این اساس وضعیت هر شخص به یکی از ۶ حال آرام، نگران، مطمئن، حیاتی، مهربان و خوشحال تقسیم‌بندی شده که نتایج بیانگر دقت ۸۷,۶ درصدی در پیش‌بینی روند روزانه شاخص داوجونز است (بولن و مائو[1]، ۲۰۱۱).

یکی از مهمترین مسائل موجود در متن کاوی، استخراج عقیده (نظر) از داده‌های متنی است. بخش شایان توجهی از متن‌های منتشر شده در وب به تولید عقیده‌های متفاوت کاربران تعلق دارد. برای مثال به نظرهایی که افراد مختلف در خصوص یک محصول جدید از شرکت خاصی منتشر کرده‌اند و می‌تواند مورد استفاده این شرکت خاص قرار گیرد، اشاره کرد. کاوش در متن‌های عقیده محور می‌تواند به منظور خلاصه‌سازی و آشکار شدن نظر کاربران در موضوع مربوطه استفاده شود. برای مثال می‌تواند انتظاراتی که مشتریان یک شرکا از

1. Bollen & Mao

محصولات آن دارند را مشخص کرده یا احساس رضایت یا نارضایتی کاربران یک محصول مشخص را بررسی کند. این مساله در پیشینه موضوع با نام تحلیل احساس شناخته می‌شود(آگاروال و ژای[1]، ۲۰۱۲).

آنالیز احساس که با عنوان عقیده‌کاوی نیز شناخته می‌شود، شامل سیستمی برای جمع‌آوری و تحلیل نظر درباره پست‌ها در بلاگ‌ها، کامنت‌ها، نقدها یا توییت‌هاست. آنالیز احساس کاربردهای متنوعی دارد؛ برای مثال در بازاریابی می‌تواند به موفقیت یک کمپین تبلیغاتی یا راه‌اندازی یک محصول جدید کمک کند، محصولات و سرویس‌های خاص و محبوب بین مشتریان را نمایان کرده و حتی مشخص کند کدام جمعیت چه ویژگی‌هایی را پسندیده و چه ویژگی‌هایی را نمی‌پسندند. پژوهش‌های بسیاری درباره تحلیل داده‌هایی که از عقیده کاربران سرچشمه گرفته‌اند، وجود دارد. بیشتر این پژوهش‌ها به بررسی احساسات و قضاوت درباره قطبیت نظرهای (مثبت یا منفی بودن نظرها) کاربران می‌پردازند. آنالیز احساسات بیشتر در سه سطح سند، جمله و منظر (مشخصه) بررسی می‌شود(وینودینی و چاندراسکاران[2]، ۲۰۱۲). در سطح سند، تمام سند به عنوان ورودی در نظر گرفته می‌شود. به بیان دیگر تمام تحلیل‌های مربوط به آنالیز احساس روی کل سند انجام می‌شود. در سطح جمله هر جمله از یک سند به عنوان ورودی تجزیه و تحلیل می‌شود. در مفهوم دیگر، تحلیل‌های مربوط به آنالیز احساس روی تک تک جملات موجود در یک سند انجام شده و در پایان با توجه به خروجی ایجاد شده، درباره کل سند قضاوت می‌شود(پرادهان، والا و بالانی[3]، ۲۰۱۶).

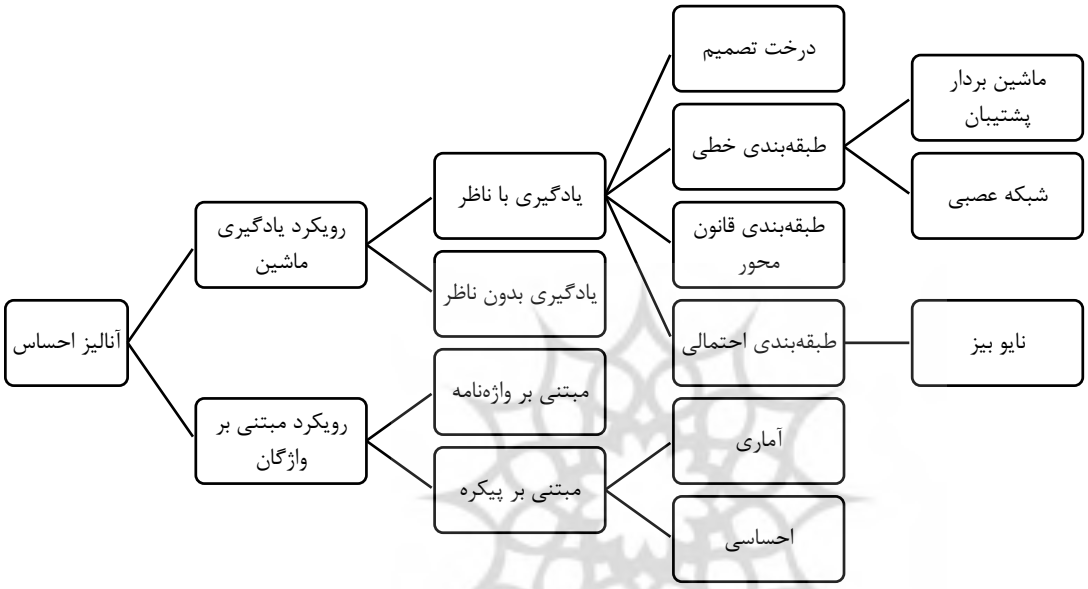
تحلیل احساسات موجود در پیام‌های منتشرشده می‌تواند در پیش‌بینی قیمت کمک کننده باشد. در حال حاضر استفاده از سیستم‌هایی که اقدام به تحلیل احساسات می‌کند برای انجام معاملات به صورت الگوریتمی و نیز به صورت معاملات پربسامد (HFT) در سایر کشورها رو به گسترش است. در این خصوص تبیین هرچه دقیق‌تر و بهتر جهت و شدت احساسات در شبکه‌های اجتماعی در خصوص یک یا چند نماد معاملاتی و نیز آگاهی از میزان تاثیرگذاری آن ضروری است.

---

1. Aggarwal & Zhai
2. Vinodhini & Chandrasekaran
3. Pradhan, Vala & Balani

مبنای نظری

در خصوص تحلیل احساسات، دو رویکرد کلی برای آنالیز احساس وجود دارد، رویکرد یادگیری ماشین و رویکرد مبتنی بر واژگان. همچنین روش‌های پردازش زبان طبیعی (NLP) در این زمینه به خصوص در تشخیص احساسات سند، استفاده می‌شود. در شکل زیر به طبقه‌بندی روش‌های آنالیز احساس پرداخته می‌شود.



بر این اساس، تکنیک‌های طبقه‌بندی احساس را می‌توان به سه رویکرد زیر دسته‌بندی کرد:

- رویکرد مبتنی بر واژگان
- رویکرد یادگیری ماشین
- رویکرد ترکیبی

رویکرد یادگیری ماشین از الگوریتم‌های معروف یادگیری و مبتنی بر ویژگی‌های زبانی استفاده می‌کند. رویکرد مبتنی بر واژگان بر اساس واژه‌نامه احساسی بنا شده که این واژه نامه مجموعه‌ای از لغت‌ها و واژه‌های از پیش تعیین شده است و خود به یر بخش‌های واژه نامه‌محور و پیکره محور طبقه‌بندی می‌شود و از روش‌های آماری و احساسی برای یافتن قطب احساسی جمله‌ها (مثبت یا منفی بودن جمله) استفاده می‌کند. رویکردهای

ترکیبی نیز از هر دو روش بالا بهره می‌برند تا بتوانند بار احساسی جمله‌ها را بررسی کنند(مدهات، حسن و کراشای[1]، ۲۰۱۴).

آن دسته از روش‌های طبقه‌بندی احساس که از رویکرد یادگیری ماشین استفاده می‌کنند را می‌توان به دو گروه یادگیری با ناظر و یادگیری بدون ناظر دسته‌بندی کرد. رویکردهای یادگیری با ناظر زمانی به کار می‌روند که موارد بسیار زیادی مستند (داده‌های دارای برچسب) به منظور یادگیری توسط ماشین وجود داشته باشد. در غیر این صورت از روش‌های یادگیری بدون ناظر استفاده می‌شود.

رویکردهای مبتنی بر واژگان به پیدا کردن واژه‌هایی وابسته‌اند که با خود بار احساسی به همراه دارند. روش مبتنی بر واژه‌نامه، به دنبال یافتن برخی واژه‌های خاص در عبارت است و با استفاده از واژه‌نامه مدنظر به یافتن برخی واژه‌های هم معنا یا مخالف با واژه یادشده می‌پردازند. روش پیکره محور با فهرستی تک کلمه‌هایی که بار عقیده‌ای دارند، کار را آغاز کرده و سعی می‌کند واژه‌های دیگری را بیابد که می‌توانند جهت‌گیری مفهومی جمله‌ها را مشخص کنند(وینودهینی و چاندراسکاران[2]، ۲۰۱۲).

در پژوهشی در سال ۲۰۱۴ مشخص شد که کمابیش ۷۵ درصد انسان‌های بالغ در حداقل یکی از شبکه‌های اجتماعی فعالیت دارند(دوگان، الیسون، لمپ، لنهارت و مادن[3]، ۲۰۱۵). شبکه‌های اجتماعی با اهداف متنوعی از اشتراک زندگی روزمره تا بیان ایده‌ها و نظرات علمی و گرفتن تصمیمات سرمایه‌گذاری به کار می‌روند(اُه و شنگ[4]، ۲۰۱۱).

تعداد بیشماری از سرمایه‌گذاران حرفه‌ای و مبتدی و تحلیلگران از شبکه‌های مجازی برای بیان نظرات و تحلیل‌های خود استفاده می‌کنند(اسپرینگر، تاماسجان، ساندنر و ولپ[5]، ۲۰۱۴).

بازده سهام تحت تاثیر عوامل بسیاری از جمله عوامل بنیادی، هزینه معاملات و ... قرار دارد، در این بین احساسات سرمایه‌گذاران نیز بازده سهام را تحت تاثیر قرار می‌دهند(بیکر و وورگلر، ۲۰۰۷)[6].

---

1. Medhat, Hassan & Korashy
2. Vinodhini & Chandrasekaran
3. Duggan, Ellison, Lampe, Lenhart & Madden
4. Oh & Sheng
5. Sprenger, Tumasjan, Sandner & Welpe
6. Baker & Wurgler

مبنای تجربی

پیش‌بینی و برآورد آینده یکی از ضروریات تصمیم‌سازی در سرمایه‌گذاری است. موضوع پیش‌بینی‌پذیر بودن بازده سهام در بازارهای مالی، همواره محل بحث و مناقشه بوده است. برپایه فرضیه بازار کارا[۱]، بازده سهام به صورت تصادفی است و قابلیت پیش بینی ندارد(اِپن و فاما[۲]، ۱۹۶۹). همچنین می‌دانیم که اطلاعات و اخبار جدید نقش مهمی در پیش‌بینی‌پذیری بازار سهام داشته و به سرعت سبب نوسان در بازدهی و قیمت سهام می‌شود، از آنجایی‌که اطلاعات به سرعت در بازار پخش می‌شود و سرمایه‌گذاران رفتار عقلایی دارند، بنابراین فرصت‌های سرمایه‌گذاری سودآور چندانی باقی نمی‌ماند(هانگ و استین[۳]، ۱۹۹۹).

روش‌شناسی

شبکه‌های عصبی مصنوعی، سیستم‌هایی بر مبنای هوش مصنوعی است که می‌کوشد کارکرد مغز انسان را به عنوان شبکه‌ای از نرون‌های متصل به هم در فرآیند تصمیم‌گیری تقلید کند. نرون‌ها کوچکترین واحدهای محاسبه و تصمیم‌گیری در شبکه‌های عصبی هستند. در هر یک از نرون‌ها یک معادله تبدیل تعریف شده است. معادله تعریف شده در هر یک از نرون‌ها می‌تواند یک مدل اقتصادسنجی یا هر مدل ریاضی دیگر مثل توابع سیگمویدی باشند. با استفاده از این معادله سعی می‌شود که وزن هر یک از متغیرها به گونه‌ای تعیین شود که ارتباط معنی‌داری بین بردار داده‌ها و بردار ستاده‌ها برقرار شود. معمولاً تعیین ضرایب در هر یک از نرون‌ها به صورت آزمون و خطا است. بدین ترتیب که ابتدا وزن‌های کوچک به هر یک از متغیرها ارائه می‌شود و سپس با استفاده از الگوریتم بازخورد، خطاهای ضرایب تعدیل می‌شوند. اینکار تا زمانی ادامه می‌یابد که خطاها به حداقل ممکن تعیین شده از سوی پژوهشگر برسد (کوپر[۴]، ۱۹۹۹).

مدل‌های شبکه عصبی ابزارهای مدل‌سازی ناپارامتری، غیرخطی و انعطاف‌پذیری هستند که هر تابع پیچیده و بزرگی را با دقت مطلوب طرح‌ریزی می‌کنند. هر مدل از تعدادی لایه مختلف تشکیل شده است و هر لایه، خود شامل تعدادی اجزای حسابگر است که در اصطلاح گره (نورون) نامیده می‌شود. هر گره، یک سیگنال ورودی را از یک گره دیگر یا از داده‌هایی که

---

۱. Efficient Market Hypothesis (EMH)
2. Eppen & Fama
3. Hong, & Stein
4. Cooper

در لایه ورودی سیستم قرار دارند، دریافت می‌کند و بعد از پردازش آن، نتیجه را به وسیله یک تابع انتقال به صورت یک سیگنال تغییر یافته به گره‌های دیگر یا لایه خروجی ارسال می‌کند. در شکل کلی، این مدل‌ها به صورت یک پرسپترون چندلایه‌ای[1] طراحی می‌شوند. در این نوع مدل‌ها پایین‌ترین لایه که اطلاعات خارجی را دریافت می‌کند، به عنوان لایه ورودی است و آخرین لایه که مخفی است، وظیفه پردازش اطلاعات و طراحی تابع را بر عهده دارد. پردازش‌هایی که در لایه مخفی صورت می‌گیرد، از دید مدل‌ساز پوشیده است و تنها جواب نهایی از این مدل‌ها استخراج می‌شود. چنین مدل‌هایی در اصطلاح جعبه سیاه نام دارند. تعیین پارامترهای این‌گونه مدل‌ها از طریق فرآیند آموزش سیستم به دست می‌آید. به این منظور ابتدا تعدادی از نمونه‌ها را برای سنجش برازندگی مدل نهایی به صورت تصادفی انتخاب می‌کنند و کنار می‌گذارند، سپس با استفاده از نمونه‌های باقیمانده که در مورد مبحث رتبه‌بندی اعتباری نسبت‌های مالی و یا سایر متغیرهای مرتبط است، به عنوان اطلاعات به لایه ورودی ارسال می‌شود. در مرحله بعد، سیستم این داده‌ها را به لایه‌های مخفی منتقل می‌کند و ضمن این کار به هر کدام از متغیرها ضرایب وزنی داده می‌شود. این عمل بارها و بارها انجام می‌شود تا نتایج مدل به اندازه مطلوب درآید و با واقعیت همخوانی کافی داشته باشد.

خصوصیات شبکه‌های عصبی

- درک این مدل‌ها بسیار دشوار است.
- رابطه بین وزن‌ها و متغیرها بسیار پیچیده است.
- هرچقدر قدرت یادگیری بالا می‌رود خطاهای مدل کمتر می‌شود.
- در شبکه‌های عصبی مصنوعی این مدل‌ها سعی دارند که ارتباط بین ورودی‌ها و خروجی‌های نهایی را تقلید کنند و شکاف بین مقادیر پیش‌بینی شده و اندازه واقعی را به حداقل برسانند.
- این مدل‌ها از فرایندهای یادگیری استفاده می‌کنند و فرایند یادگیری نیز از حلقه بازخورد استفاده می‌کند (شیرزادی، ۱۳۸۵).

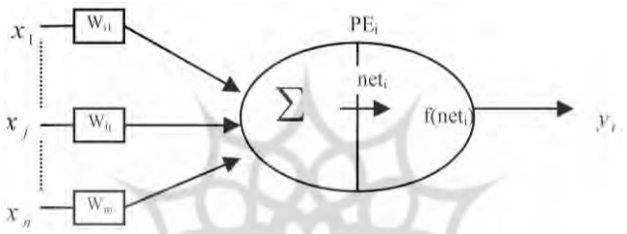
---

1. Multi-Layer Perceptron

شبکه‌های پرسپترون چندلایه

هدف شبکه‌های عصبی، تلاش برای ساخت الگوهایی است که همانند مغز انسان عمل می‌کنند. کار شبکه عصبی ایجاد یک الگوی خروجی بر اساس الگوی ورودی ارائه شده به شبکه است. شبکه‌های عصبی متشکل از تعدادی عناصر پردازشی (نرون‌های مصنوعی) هستند که این نرون‌ها درون داده‌ها را دریافت و پردازش می‌کند و در نهایت یک برون‌داد از آن ارائه می‌دهد. درون‌داد می‌تواند داده‌های خام یا برون‌داد دیگر عناصر پردازشی باشد. برون‌داد می‌تواند محصول نهایی یا درون‌دادی برای یک نرون دیگر باشد.

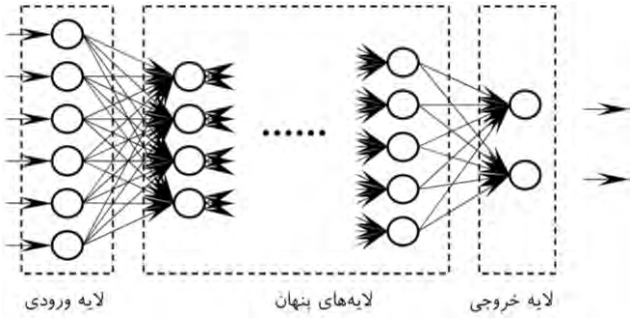
یک شبکه عصبی مصنوعی متشکل از نرون‌های مصنوعی است که در واقع همان عناصر پردازشی هستند. در شکل زیر به صورت ساده یک عنصر پردازشی توضیح داده شده است.



Wها وزن‌های اختصاص یافته به هر ورودی، net تابع مجموع، f تابع تبدیل، xها ورودی‌های نرون و yها خروجی‌های نرون هستند.

تابع مجموع، مجموع موزون ورودی‌ها را محاسبه می‌کند و فرمول آن $net = \sum w_{ij}x_j$ است، رابطه بین سطح فعال شدن و برون‌داد با استفاده از تابع تبدیل توصیف می‌شود که دارای انواع مختلفی از جمله تانژانت هیپوربولیک، سیگموئید و ... است.

شبکه‌های پیشخور شبکه‌هایی هستند که ارتباط تنها یک طرفه است و از هر نرون داده‌ها تنها به نرون بعدی منتقل می‌شود اما در شبکه‌های پسخور ارتباط دوطرفه است. پرسپترون چندلایه و یا MLP معماری از شبکه‌های عصبی مصنوعی است که پیشخور بوده و پردازنده‌های شبکه به چند لایه مختلف تقسیم می‌شوند. در این شبکه‌ها لایه اول، ورودی، لایه آخر، خروی و لایه‌های میانی، لایه‌های پنهان نامیده می‌شوند. این معماری را پرکاربردترین معماری شبکه‌های عصبی می‌توان نامید، شکل زیر نمای عمومی پرسپترون چندلایه را نشان می‌دهد:



یک شبکه عصبی مصنوعی اطلاعات را از لایه اول با استفاده از تابع انتقال به لایه منتقل می‌کند. در یک شبکه پیش-خور با یک لایه پنهان، نورون jام در لایه پنهان به صورت زیر تعریف می‌شود:

$$h_j = f_j(\alpha_{0j} + \sum_{i \to j} w_{ij} x_i)$$

α_{0j} خطا نامیده می‌شود. جمع $i \to j$ نشان‌دهنده جمع تمامی نورون‌های ورودی به نورون jبوده که در آن وزن‌ها به صورت w_{ij} تعریف شده‌اند. همچنین x_i مقدار iامین نورون ورودی بوده و $f_j(0)$ یک تابع انتقال است. برای لایه خروجی، نورون o به صورت زیر تعریف می‌شود:

$$o = f_0(\alpha_{0o} + \sum_{j \to o} w_{jo} h_j)$$

(سای، ۲۰۰۵).

در این پژوهش در مدل‌های جهتی متغیر وابسته با مقادیر ۱ برای صعود و نیز مقدار ۰ برای نزول در نظر گرفته شده و در مدل‌های مقداری نیز بازدهی روزانه هر سهم که به صورت لگاریتمی حساب شده است، به عنوان متغیر ورودی مدل (متغیر وابسته) در نظر گرفته شده است. متغیرهای مستقل پژوهش نیز شامل بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول و محتوای احساسی بوده که به تفصیل در بخش مراحل انجام کار توضیح داده شده است. در مدل مورد استفاده در پژوهش حاضر ۷۰ درصد

داده‌ها برای آموزش استفاده شده و از ۳۰ درصد باقیمانده، نیمی برای آزمون و نیمی برای اعتبارسنجی مورد استفاده قرار گرفته است. پس از انجام فرایند یادگیری، داده‌ها با یک پنجره متحرک روزانه رو به جلو برای پیش‌بینی بر اساس شبکه آموزش دیده، مورد استفاده قرار گرفت. همچنین در این پژوهش تابع تبدیل (انتقال) تانژانت هیپوربولیک بوده و از الگوریتم Levenberg-Marquardt برای آموزش شبکه عصبی استفاده شده است.

تحلیل احساسات

قطعه‌بندی متن[1]

قطعه‌بندی متن عملی است که در آن دنباله‌ای از رشته‌ها به قطعاتی مانند جملات، کلمات، کلمات کلیدی، عبارات، نمادها و سایر عناصری که به آن‌ها قطعه[2] گویند، شکسته می‌شوند. درواقع وظیفه قطعه‌بند[3]، تکه‌تکه کردن سند به قسمت‌های کوچک به نام قطعه است. قطعه‌بندی می‌تواند در سطح جملات، کلمات و سایر موارد بیان شده انجام شود و واحدهای استخراج شده می‌توانند به عنوان ورودی ماژول‌های دیگر مورد استفاده قرار گیرند. تشخیص مرز بین این واحدها در متن و تبدیل آن به دنباله‌ای از آن‌ها برعهده‌ی این ابزار است. این مرز بیشتر می‌تواند نقطه باشد.

یکسان‌سازی متن[4]

در پردازش‌های متنی، به‌منظور این‌که واحدهای متنی برای ماشین قابل پردازش باشند، عملیات یکسان‌سازی در آن‌ها انجام می‌شود. دلیل انجام این عملیات این است که بعضی از کاراکترها در متون مانند «ی» و «ک» دارای یونی‌کدهای متفاوتی هستند. همچنین شیوه اتصال وندها به کلمات اصلی به شیوه‌های مختلفی انجام می‌شود. مورد بعدی این است که اجزای کلمات مرکب به روش‌های مختلف به هم متصل می‌شوند و سرانجام کلماتی که املای آن‌ها به چند صورت متفاوت نوشته می‌شود. همه این موارد، دلایلی هستند که موجب می‌شود تا به منظور پردازش بهتر از ابزارهای یکسان‌سازی واحدهای متنی در آن‌ها استفاده شود.

---

1. tokenizing
2. Token
3. Tokenizer
4. Normalization

تحلیل صرفی جملات

فرآیند انتساب مقوله نحوی به هر کلمه در پیکره متنی را تحلیل صرفی کلمات هر جمله می‌گویند. مقوله‌ی نحوی که هر کلمه به آن تعلق دارد را جزءکلام[1] می‌نامند.

یافتن ریشه کلمات[2]

منظور از یافتن ریشه کلمات در این پژوهش، عمل حذف وندهای تصریفی از کلمه است. با وندهای تصریفی کلمه‌ی جدیدی ساخته نمی‌شود بلکه صورت تازه‌ای از کلمه به دست می‌آید که در موقعیت‌های مختلف برای سازگاری با نحو، استفاده می‌شوند به طور مثال اسم مفرد به جمع تبدیل می‌شود. با به‌کارگیری وندهای تصریفی می‌توان ابعاد دستوریِ شمار، شخص، زمان دستوری، نمود، حالت و جز آن‌ها را بازنمود.

پرسش‌ها و فرضیه‌ها

در پژوهش‌های پیشین از جمله (راعی و حسینی، ۱۳۹۵)، پیام‌ها و نظرات موجود در شبکه‌های اجتماعی تنها از نظر جهتی مورد توجه پژوهش‌گران واقع شده بود. در این پژوهش سه عامل محتوا، فرستنده و میزان توجه به پیام مورد توجه و مد نظر برای تفکیک پیام‌ها است. بنابراین به منظور مدلسازی بهتر ابتدا لازم است بدانیم آیا محتوای منتشرشده از نظر نوع محتوا (تحلیل یا سیگنال) به صورت معناداری متفاوت منتشر می‌شود یا خیر. پاسخ پرسش‌ها می‌تواند در مدلسازی کمک کننده باشد و در صورت معنی دار نبودن موارد، می‌تواند منجر به تجمیع متغیرها شود.

پرسش اول: غالب محتوای منتشرشده در شبکه‌های اجتماعی مرتبط با بازار سرمایه حاوی چیست؟

پرسش دوم: چه نوع محتوایی در شبکه‌های اجتماعی بازار سرمایه بیشتر مورد توجه مخاطبان قرار می‌گیرد؟

فرضیه اول: قدرت پیش‌بینی جهتی قیمت سهام با استفاده از داده‌های تاریخی تفاوت معنی‌داری با قدرت پیش‌بینی جهت قیمت سهام با استفاده از داده‌های تاریخی و نظرات غالب افراد (نظرات اشخاص صاحب شهرت) در شبکه‌های اجتماعی دارد.

---

1. Part Of Speech (POS)
2. Stemming

فرضیه دوم: دقت پیش‌بینی بازده قیمت سهام با استفاده از داده‌های تاریخی تفاوت معنی‌داری با دقت پیش‌بینی بازده قیمت سهام با استفاده از داده‌های تاریخی و نظرات غالب افراد (نظرات اشخاص صاحب شهرت) در شبکه‌های اجتماعی دارد.

فرضیه سوم: قدرت پیش‌بینی جهت قیمت در انواع مختلف محتوای اطلاعاتی منتشر شده در شبکه‌های اجتماعی مانند تحلیل یا سیگنال تفاوت معنی داری با استفاده از داده‌های تاریخی و نظرات غالب افراد (نظرات اشخاص صاحب شهرت) در شبکه‌های اجتماعی دارد.

فرضیه چهارم: دقت پیش‌بینی بازده قیمت در انواع مختلف محتوای اطلاعاتی منتشر شده در شبکه‌های اجتماعی مانند تحلیل یا سیگنال تفاوت معنی‌داری با استفاده از داده‌های تاریخی و نظرات غالب افراد (نظرات اشخاص صاحب شهرت) در شبکه‌های اجتماعی دارد.

مراحل انجام کار

پیام‌های تلگرامی مورد استفاده در این پژوهش در بازه زمانی ابتدای سال ۱۳۹۸ تا پایان مرداد ۱۴۰۰ است. پیام‌های یادشده از طریق API تلگرام دریافت شده است. در تمامی موارد این پژوهش از ابزار هضم[1] استفاده شده است. به منظور دستیابی به پیام‌های مورد نظر برای این پژوهش می‌بایست فیلتری را بر روی این پیام‌ها منظور کرد تا هم سرعت پردازش بالا رود و هم از پردازش داده‌های هرز جلوگیری شود. بدین منظور با منظور کردن نمادهای موجود در شاخص ۳۰ شرکت، پیام‌هایی که نام نماد یا شرکت یکی از شرکت‌های یادشده یا کلیات بازار سرمایه در آن بود انتخاب شد. در مرحله‌ی بعد از بین این پیام‌های منتخب پیام‌هایی که در حداقل یکی از دسته‌های بنیادی، تکنیکال، سیگنال خرید و سیگنال فروش قرار داشت جمع‌آوری و در مراحل آتی از آن‌ها استفاده شد. لازم به بیان است برای دسته‌بندی پیام‌ها در هر یک از ۴ دسته مذکور کلماتی انتخاب و براساس این که چه تعداد از کلمات هر دسته در هر پیام وجود داشته باشد به هر دسته امتیازی داده می‌شود که میزان تعلق هر پیام به هر دسته را نشان می‌دهد. در این زمان پیام‌های قابل پردازش معلوم و برای استفاده در سایر مراحل استفاده می‌شوند. به‌منظور این‌که پیام‌های تلگرام برای ماشین قابل پردازش باشند، عملیات یکسان‌سازی در آن‌ها انجام می‌شود. برای امتیازدهی به هر پیام براساس مثبت یا منفی بودن آن می‌بایست هر پیام با استفاده

1. http://www.sobhe.ir/hazm./

از ابزار قطعه‌بندی متن به جملات تبدیل شوند و امتیاز هر جمله محاسبه شود تا امتیاز نهایی هر پیام مشخص شود. هر جمله می‌تواند براساس دارا بودن تعداد کلمات مثبت یا منفی امتیازی در هر دسته داشته باشد و جمع این دو امتیاز، امتیاز نهایی جمله را تعیین می‌کند. همچنین امتیاز هر پیام نیز از جمع امتیازهای جملات آن پیام مشخص می‌شود.

همانطور که پیش گفته شد امتیاز هر جمله براساس دارا بودن تعداد کلمات مثبت یا منفی مشخص می‌شود و جمع این دو امتیاز، امتیاز نهایی جمله را تعیین می‌کند. لازم به بیان است ممکن است در یک جمله کلمه مثبت به کار رفته باشد اما فعل جمله منفی باشد مانند «شرایط بازار مناسب نیست». در این جمله کلمه «مناسب» یک کلمه مثبت است اما جمله، جمله منفی است. برای حل این مشکل می‌بایست مثبت یا منفی بودن فعل هر جمله به صورت +۱ یا −۱ مشخص و در امتیاز آن جمله ضرب شود. برای تشخیص افعال هر جمله از ابزار تحلیل صرفی جملات استفاده می‌شود. منفی یا مثبت بودن هر فعل را می‌توان از به همراه داشتن حرف «ن» در ابتدای آن فعل تشخیص داد. البته ممکن است حرف «ن» جزء حروف خود فعل باشد به همین خاطر ابتدا باید ریشه فعل را با استفاده از ابزار یافتن ریشه کلمات پیدا کرد، در صورتی که حرف اول ریشه فعل حرف «ن» باشد و در ابتدای فعل موجود در جمله نیز تنها یک حرف «ن» به کار رفته باشد آن فعل مثبت و در صورت وجود دو حرف «ن» متوالی در ابتدای آن، فعل منفی می‌شود مانند «ننوشت».

به منظور بررسی اهمیت تفکیک محتوا، ابتدا با استفاده از اطلاعات تعداد پیام منتشرشده به تفکیک محتوا به بررسی محتواهای غالب در شبکه اجتماعی تلگرام می‌پردازیم. در نتیجه پس از آن با استفاده از اطلاعات مربوط به تعداد مشاهده هر پیام، به بررسی میزان مورد مطالعه قرار گرفتن انواع پیام‌ها می‌پردازیم. با توجه به اینکه ممکن است رفتار افراد در رژیم‌های مختلف صعودی، نزولی یا خنثی بازار از حیث انتشار و نیز مطالعه پیام‌های ارسالی تفاوت داشته باشد، به بررسی این موضوع در رژیم‌های محتلف بازار نیز می‌پردازیم. چنانچه رفتار افراد در شرایط بازار تفاوت داشته باشد، موضوع را در مدلسازی منظور خواهیم کرد.

در این پژوهش افزون بر محتوای منتشرشده اعم از سیگنال خرید، سیگنال فروش، تحلیل بنیادی و نیز تحلیل تکنیکال و جهت پیام (خرید یا فروش)، به نوع فرستنده و نیز میزان توجه به پیام هم توجه شده است.

در حالی که سایر مقالات و پژوهش‌ها، تنها تعداد پیام در خصوص یک نماد معاملاتی مد نظر نویسندگان قرار می‌گرفت و نیز در برخی پژوهش‌ها جهت محتوای منتشرشده (مثبت یا منفی) مورد نظر پژوهشگر بود، با توجه به اینکه هدف از این پژوهش بررسی موثر بودن یا نبودن سرچشمه پیام و کیفیت پیام منتشرشده در پیش‌بینی بازده سهام بود، از بین محتوای دسته‌بندی شده، کانال‌هایی که متوسط تعداد مشاهده پیام‌های آن کانال در هر ماه در دهک اول قرار داشت (کانال‌هایی با بیشترین میزان مشاهده) به عنوان گروهی مجزا قرار گرفتند و محتوای منتشر شده از آن‌ها در هر طبقه تفکیک شد. از سوی دیگر با توجه به اینکه هر یک از پیام‌ها نیز دارای اطلاعات میزان مشاهده هستند، پیام‌هایی که بیشترین مشاهده را داشتند به ترتیب در هر ماه مرتب شدند و پیام‌های با مشاهده بیشتر که در دهک اول جای می‌گرفتند به صورت مجزا به عنوان متغیر تعریف شدند. پس از آن بازدهی سهام شرکت‌ها و نیز تعداد پیام‌های منتشرشده در هر دسته، در داده‌ها با یکدیگر ادغام شد.

بدین منظور ۶ مدل با استفاده از شبکه عصبی مصنوعی با متغیرهای مختلف برای پیش‌بینی بازده سهام موجود در شاخص ۳۰ شرکت اجرا شد. تعداد سهام موجود در شاخص یادشده با توجه به تغییر در پرتفوی شاخص در دوره زمانی ابتدای سال ۱۳۹۸ تا پایان مرداد ماه ۱۴۰۰، ۳۶ شرکت بود. به ازای هر مدل، ۵ حالت برای تعداد نورون‌های موجود در لایه پنهان (از ۶ الی ۱۰ نورون) و نیز به ازای هر یک از موارد مورد اشاره از ۱ الی ۵ وقفه زمانی برای داده‌های یاد شده در نظر گرفته شد. بدین معنی که به ازای هر مدل-نماد ۲۵ ریز مدل برای یادگیری جهت پیش‌بینی جهتی و ۲۵ مدل برای یادگیری برای پیش‌بینی مقداری وجود دارد. در نتیجه به ازای هر ریز مدل هم ۱۰ مرتبه اجرا صورت گرفت و میانگین نتایج هر اجرا مورد استفاده قرار گرفت. با توضیحات بالا با توجه به تعداد نمادهای معاملاتی، ۱۰۸,۰۰۰ مدل اجرا شد. در این خصوص تعداد ۱۵,۷۳۶,۲۰۴ پیام کانال تلگرامی که دارای محتوای قابل بررسی بودند استخراج شده (موارد غیرقابل بررسی شامل صحبت‌های متعارف و موارد غیرمرتبط با نمادهای معاملاتی است) و پس از برچسب‌گذاری بر اساس لغتنامه احساسی، موارد در تعداد ۱۹,۳۱۲ داده مربوط به نماد-روز حاوی بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول و محتوای احساسی خلاصه شد. مدل‌های اجرا شده به شرح زیر هستند:

مدل ۱: این مدل افزون بر سایر داده‌های مورد اشاره در خصوص بازده نماد یا سایر عوامل موثر بر بازار، شامل محتوای احساسی به شرح داده‌های تحلیل بنیادی مثبت، تحلیل بنیادی منفی، تحلیل بنیادی سایر، تحلیل تکنیکال مثبت، تحلیل تکنیکال منفی، تحلیل تکنیکال سایر، سیگنال خرید و سیگنال فروش به عنوان نوع محتوای منتشرشده بوده و درآن تفکیک موثر بودن یا نبودن فرستنده نیز منظور شده و همچنین مهم بودن پیام از حیث تعداد مشاهده نیز از جمله موارد تفکیک کننده بوده است.

مدل ۲: این مدل افزون بر سایر داده‌های مورد اشاره در خصوص بازده نماد یا سایر عوامل موثر بر بازار، شامل محتوای احساسی به شرح داده‌های تحلیل بنیادی مثبت، تحلیل بنیادی منفی، تحلیل تکنیکال مثبت، تحلیل تکنیکال منفی، سیگنال خرید و سیگنال فروش به عنوان نوع محتوای منتشرشده بود و درآن تفکیک موثر بودن یا نبودن فرستنده نیز منظور شده و همچنین مهم بودن پیام از حیث تعداد مشاهده نیز از جمله موارد تفکیک کننده بوده است. تفاوت این مدل با مدل قبلی حذف متغیرهای تحلیل بنیادی سایر و تحلیل تکنیکال سایر است.

مدل ۳: این مدل افزون بر سایر داده‌های مورد اشاره در خصوص بازده نماد یا سایر عوامل موثر بر بازار، شامل محتوای احساسی به شرح داده‌های تحلیل بنیادی، تحلیل تکنیکال، سیگنال خرید و سیگنال فروش است.

مدل ۴: این مدل افزون بر سایر داده‌های مورد اشاره در خصوص بازده نماد یا سایر عوامل موثر بر بازار، شامل محتوای احساسی به شرح شامل محتوای مثبت، منفی و خنثی است.

مدل ۵: این مدل افزون بر سایر داده‌های مورد اشاره در خصوص بازده نماد یا سایر عوامل موثر بر بازار، شامل محتوای احساسی تنها بر اساس داده‌ی تعداد پیام منتشرشده است.

مدل ۶: این مدل فاقد محتوای داده‌های تلگرامی بوده و تنها از موارد مورد اشاره در خصوص بازدهی و عوامل موثر بر بازار استفاده می کند.

به منظور پردازش اولیه اطلاعات از نرم افزار Microsoft SQL Server Management Studio 18 استفاده شد. پس از آن از نرم‌افزارهای Matlab 2021 برای انجام محاسبات مربوط به شبکه‌های عصبی مصنوعی استفاده شد و در نهایت نیز از نرم افزار IBM SPSS Statistics 26 برای انجام آزمون‌های مقایسه میانگین زوجی استفاده شد.

یافته‌ها

در این بررسی از تعداد ۱۹۳۱۲ روز-نماد معاملاتی استفاده شده است. با توجه به اینکه برخی محتوای منتشره در روزهای تعطیل منتشر می‌شود، پیام‌هایی که در روزهای تعطیل ارسال می‌شود در روز معاملاتی قبلی لحاظ می‌شود. از سوی دیگر با توجه به اینکه برخی پیام‌ها از ساعت ۰۰:۰۰ بامداد لغایت شروع جلسه معاملاتی ارسال می‌شوند، با توجه به زمان ارسال، پیام‌های مذکور در روز قلب منظور می‌شوند. شایان بیان است هر روز به سه بازه زمانی ۰۰:۰۰ لغایت ۸:۳۰ (قبل از بازار)، ۸:۳۰ لغایت ۱۲:۳۰ (حین بازار) و نیز ۱۲:۳۰ لغایت ۲۳:۵۹ (پس از بازار) تقسیم شده است. در پاسخ به پرسش اول مبنی بر غالب محتوای منتشرشده در شبکه‌های اجتماعی مرتبط با بازار سرمایه حاوی چیست؟ آزمون مقایسه میانگین (جدول ۱) به ازای درصد محتوای منتشرشده در هر ماه از بازه مورد بررسی نشان می‌دهد حدوداً ۳۶ درصد از محتوای منتشرشده تحلیل بوده و سایر محتوای منتشره (۶۴ درصد) سیگنال شمرده می‌شود. همچنین جداول ۳ و ۴ نیز آزمون میانگین زوجی برای نوع محتوای ارسالی در کانال‌ها را نشان می‌دهد. ملاحظه می‌شود حدود ۴۳ درصد از محتوای منتشرشده در کانال‌ها تحلیل بوده و ۵۷ درصد پیام‌های منتشرشده سیگنال است. بنابراین با توجه به میانگین موجود در جداول ۱ و ۲، می‌توان نتیجه گرفت تعداد سیگنال‌های منتشرشده در گروه‌های تلگرامی دارای تعداد و در نتیجه وزن بیشتری است. در جدول‌های ۵ و ۶ مشاهده می‌شود که نسبت انتشار سیگنال در سوپر گروه‌ها ۶۶ درصد است. جداول ۷ و ۸ نیز نشان دهنده نتایج آزمون‌های میانگین زوجی برای سیگنال و تحلیل منتشره به تفکیک کانال/سوپر گروه و نیز زمان انتشار است. در نتیجه در پاسخ به پرسش اول، می‌توان نتیجه گرفت بیش از حدوداً ۶۰ درصد از پیام‌های موجود در بازار سرمایه در قالب سیگنال طبقه‌بندی می‌شوند.

جدول ۱. آزمون مقایسه میانگین زوجی نوع پیام‌های ارسالی

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	tahlil	.۳۶۱۴	۲۹	.۰۲۴۳۱	.۰۰۴۵۱
	signal	.۶۳۸۶	۲۹	.۰۲۴۳۱	.۰۰۴۵۱

جدول ۲. آماره آزمون میانگین زوجی

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	tahlil - signal	-.۲۷۷۲۷	.۰۴۸۶۱	.۰۰۹۰۳	-.۲۹۵۷۶	-.۲۵۸۷۸	-۳۰.۷۱۵	۲۸	.۰۰۰

جدول ۳. آزمون مقایسه میانگین زوجی پیام‌های ارسالی در کانال‌ها

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	tahlil	.۴۳۱۰	۲۹	.۰۲۵۳۳	.۰۰۴۷۰
	Signal	.۵۶۹۰	۲۹	.۰۲۵۳۳	.۰۰۴۷۰

جدول ٤. آزمون مقایسه میانگین زوجی پیام‌های ارسالی در کانال‌ها

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	tahlil - signal	-.۱۳۸۱۰	.۰۵۰۶۵	.۰۰۹۴۱	-.۱۵۷۳۷	-.۱۱۸۸۳	-۱۴.۶۸۱	۲۸	.۰۰۰

جدول ۵. آزمون مقایسه میانگین زوجی پیام‌های ارسالی در سوپرگروه‌ها

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	tahlil	.۳۴۳۷	۲۹	.۰۱۹۰۳	.۰۰۳۵۳
	signal	.۶۵۶۳	۲۹	.۰۱۹۰۳	.۰۰۳۵۳

جدول ٦. آزمون مقایسه میانگین زوجی پیام‌های ارسالی در سوپرگروه‌ها

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	tahlil - signal	-.۳۱۲۶۰	.۰۳۸۰۷	.۰۰۷۰۷	-.۳۲۷۰۹	-.۲۹۸۱۲	-۴۴.۲۲۰	۲۸	.۰۰۰

جدول ۷. آزمون مقایسه میانگین زوجی پیام‌های ارسالی در کانال‌ها و گروه‌ها به تفکیک زمان انتشار

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	supAfter-tahlil	.۳۶۸۵	۲۹	.۰۱۵۳۵	.۰۰۲۸۵
	supAfter-signal	.۶۳۱۵	۲۹	.۰۱۵۳۵	.۰۰۲۸۵
Pair 2	supBefore-tahlil	.۳۴۴۳	۲۹	.۰۳۶۱۲	.۰۰۶۷۱
	supBefore-signal	.۶۵۵۷	۲۹	.۰۳۶۱۲	.۰۰۶۷۱
Pair 3	supMiddle-tahlil	.۳۱۰۶	۲۹	.۰۲۵۳۲	.۰۰۴۷۰
	supMiddle-signal	.۶۸۹۴	۲۹	.۰۲۵۳۲	.۰۰۴۷۰
Pair 4	channelAfter-tahlil	.۴۵۴۰	۲۹	.۰۲۰۸۰	.۰۰۳۸۶
	channelAfter-signal	.۵۴۶۰	۲۹	.۰۲۰۸۰	.۰۰۳۸۶
Pair 5	channelBefore-tahlil	.۴۲۲۱	۲۹	.۰۳۵۶۲	.۰۰۶۶۲
	channelBefore-signal	.۵۷۷۹	۲۹	.۰۳۵۶۲	.۰۰۶۶۲
Pair 6	channelMiddle-tahlil	.۴۱۱۵	۲۹	.۰۳۴۹۴	.۰۰۶۴۹
	channelMiddle-signal	.۵۸۸۵	۲۹	.۰۳۴۹۴	.۰۰۶۴۹

جدول ۸. آزمون مقایسه میانگین زوجی پیام‌های ارسالی در کانال‌ها و گروه‌ها به تفکیک زمان انتشار

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	supAfter	-.۲۶۳۰۶	.۰۳۰۷۰	.۰۰۵۷۰	-.۲۷۴۷۴	-.۲۵۱۳۹	-۴۶.۱۴۶	۲۸	.۰۰۰
Pair 2	supBefore	-.۳۱۱۳۳	.۰۷۲۲۵	.۰۱۳۴۲	-.۳۳۸۸۱	-.۲۸۳۸۵	-۲۳.۲۰۶	۲۸	.۰۰۰
Pair 3	supMiddle	-.۳۷۸۷۳	.۰۵۰۶۵	.۰۰۹۴۱	-.۳۹۷۹۹	-.۳۵۹۴۶	-۴۰.۲۶۹	۲۸	.۰۰۰
Pair 4	channelAfter	-.۰۹۱۹۹	.۰۴۱۶۰	.۰۰۷۷۲	-.۱۰۷۸۱	-.۰۷۶۱۷	-۱۱.۹۰۹	۲۸	.۰۰۰
Pair 5	channelBefore	-.۱۵۵۸۰	.۰۷۱۲۵	.۰۱۳۲۳	-.۱۸۲۹۱	-.۱۲۸۷۰	-۱۱.۷۷۶	۲۸	.۰۰۰
Pair 6	channelMiddle	-.۱۷۷۰۹	.۰۶۹۸۷	.۰۱۲۹۷	-.۲۰۳۶۷	-.۱۵۰۵۲	-۱۳.۶۴۹	۲۸	.۰۰۰

در خصوص سوال دوم مبنی بر اینکه چه نوع محتوایی در شبکه‌های اجتماعی بازار سرمایه بیشتر مورد توجه مخاطبان قرار می‌گیرد؟ از آنجایی که تعداد دفعات ارسال پیام و میزان محتوای موجود نمی‌تواند به تنهایی نشان دهنده توجه کاربران باشد، تعداد مشاهده پیام‌های ارسال شده نیز باید مورد بررسی قرار گیرد. بدین معنی که ممکن است محتوای مورد نظر ارسال شده لکن مورد توجه اشخاص قرار نگیرد. بدین منظور متوسط مشاهده پیام‌های منتشرشده را با استفاده از آزمون مقایسه میانگین زوجی مورد بررسی قرار دادیم. نتایج بررسی به شرح جداول ۹ و ۱۰ نشان می‌دهد، متوسط مشاهده پیام‌هایی که سیگنال در نظر گرفته می‌شوند به صورت معنی‌داری در سطح معنی داری ۹۵ درصد بیش از پیام‌هایی با محتوای تحلیلی است. در این خصوص بررسی دقیق‌تر نشان می‌دهد: بر اساس جداول ۱۱ و ۱۲ افراد در خصوص پیام‌هایی با محتوای تحلیلی به صورت معنی داری به پیام‌های با محتوای بنیادی توجه می‌کنند و در خصوص سیگنال‌های ارسالی توجه بیشتری به سیگنال‌های خرید دارند.

جدول ۹. آزمون مقایسه میانگین زوجی مشاهده پیام‌ها

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Signal	۱۵۶۲۷.۱۴۰۶	۸۸۶	۱۲۲۵۹.۲۰۸۳۷	۴۱۱.۸۵۶۱۶
	tahlil	۱۴۲۹۲.۵۶۶۳	۸۸۶	۶۱۷۳.۲۰۹۲۸	۲۰۷.۳۹۳۰۲

جدول ۱۰. آزمون مقایسه میانگین زوجی مشاهده پیام‌ها

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	total	۱۳۳۴.۵۷۴۳۰	۱۱۵۲۲.۴۵۲۰۲	۳۸۷.۱۰۴۳۵	۵۷۴.۸۲۴۶۸	۲۰۹۴.۳۲۳۹۲	۳.۴۴۸	۸۸۵	.۰۰۱

جدول ۱۱. آزمون مقایسه میانگین زوجی مشاهده پیام‌ها به تفکیک نوع تحلیل و جهت سیگنال

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Total-bonyadi	۱۴۷۶۶.۸۷۹۴	۸۸۶	۶۹۰۹.۳۶۲۵۹	۲۳۲.۱۲۴۵۸
	Total-technical	۱۳۸۰۰.۴۹۳۱	۸۸۶	۷۶۹۱.۶۱۱۹۸	۲۵۸.۴۰۴۷۶
Pair 2	total-SignalB	۱۷۲۷۱.۴۰۲۵	۸۸۶	۲۱۷۸۱.۷۰۷۱۰	۷۳۱.۷۷۰۷۶
	total-SignalS	۱۳۹۹۸.۵۷۶۱	۸۸۶	۷۱۶۵.۵۳۳۶۹	۲۴۰.۷۳۰۸۱

جدول ۱۲. آزمون مقایسه میانگین زوجی مشاهده پیام‌ها به تفکیک نوع تحلیل و جهت سیگنال

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Bonyadi-technical	۹۶۶.۳۸۶۳۰	۷۸۵۸.۰۴۹۱۳	۲۶۳.۹۹۶۳۲	۴۴۸.۲۵۴۴۱	۱۴۸۴.۵۱۸۱۸	۳.۶۶۱	۸۸۵	.۰۰۰
Pair 2	Signal B&S	۳۲۷۲.۸۲۶۴۴	۲۱۱۸۵.۵۸۷۴۸	۷۱۱.۷۴۳۷۳	۱۸۷۵.۹۲۳۹۴	۴۶۶۹.۷۲۸۹۴	۴.۵۹۸	۸۸۵	.۰۰۰

همانگونه که عنوان شد، در این پژوهش داده‌های شبکه اجتماعی در سه سطح نوع محتوا، نوع فرستنده و نیز میزان توجه به پیام طبقه‌بندی شد. در خصوص سطح اول نیاز بود نخست محتوای غالب در شبکه اجتماعی تلگرام در حوزه بازار سرمایه از حیث انتشار بررسی شود که در این خصوص پرسش اول، مبنی بر غالب محتوای منتشرشده در شبکه‌های اجتماعی مرتبط با بازار سرمایه حاوی چیست؟ طرح شد. در پاسخ به این پرسش، آزمون مقایسه میانگین زوجی نشان داد به صورت کلی حدوداً ۶۴ درصد از محتوای منتشرشده در تلگرام مربوط به بازار سرمایه مربوط به محتوای سیگنال (اظهار نظر صریح در خصوص خرید، فروش یا نگهداری یک ورقه بهادار) است. هرچند این مقدار بسته به زمان انتشار و نیز محل انتشار (کانال یا گروه) متفاوت است اما در تمامی حالات سیگنال محتوای غالب بوده و بیشترین میزان آن در خصوص گروه های تلگرامی و در حین ساعات معاملاتی است. پس از آن پرسش و در ادامه پرسش اول

برای تبیین بهتر ضرورت تفکیک طبقه اول، لازم بود به سوال دیگری نیز پاسخ داده شود. چه نوع محتوایی در شبکه‌های اجتماعی بازار سرمایه بیشتر مورد توجه مخاطبان قرار می‌گیرد؟ در پاسخ به پرسش اینکه افراد چه نوع محتوایی را مورد توجه قرار می‌دهند، نتایج بدست آمده نشان می‌دهد متوسط توجه به محتوای تحلیل بنیادی بیش از تحلیل تکنیکال بوده و در خصوص سیگنال نیز متوسط توجه به سیگنال خرید بیش از سیگنال فروش است. نتایج به دست آمده از آزمون مقایسه میانگین زوجی برای متوسط مشاهده پیام‌ها نشان داد پیام‌های سیگنالی به صورت کلی بیشتر از پیام‌های تحلیلی مورد توجه افراد قرار می‌گیرند. در این خصوص لازم به توجه است، پیام‌های بازنشر شده از کانال‌ها به گروه‌ها تنها امکان شمارش میزان مشاهده را دارند که نتایج به دست آمده نشان می‌دهد اشخاص با انتخاب موارد مد نظر خود از کانال‌ها و نیز بازنشر کردن آن در گروه‌ها توجه بیشتری به موارد سیگنالی کرده و در خصوص آن با یکدیگر تبادل نظر می‌کنند. هرچند این رفتار با توجه به ذی‌نفع بودن ایشان در معاملات دور از انتظار نیست. بدین منظور که اگر فردی سیگنالی را زودتر متوجه شود، به دلیل اینکه ممکن است ناشی از افشای آن قیمت افزایش یا کاهش داشته باشد، فرض عقلایی ایجاب می‌نماید که اول خود از آن استفاده نماید. سپس آن را به اشتراک بگذارد. رفتاری که از میزان توجه مشاهده می‌شود این موضوع را به ذهن می‌آورد که اشخاص نمادهایی را تبلیغ می‌کنند که خود در آن نفع معینی دارند. آنچه در پاسخ به این سوال به دست آمد تفاوت میان توجه افراد به محتوای ارسالی است. نتایج بدست آمده از پرسش یادشده ، تاکید کننده بر لزوم تفکیک نمودن محتوای منتشرشده در سطح اول مبنی بر نوع محتوای ارسالی است زیرا نخست میزان انتشار موارد متفاوت است دوم اشخاص رفتارهای متفاوتی در توجه به آن‌ها دارند. به منظور مدلسازی موارد نیز ضروری است تفاوت رفتار افراد در توجه به محتوای منتشرشده در رژیم‌های مختلف بازار مورد بررسی قرار گیرد.

به منظور در نظر گرفتن رژیم‌های بازار در مدلسازی، ابتدا باید به پرسش «آیا تفاوت معنی‌داری میان رفتار افراد در رژیم‌های صعودی، نزولی یا خنثی در شبکه‌های اجتماعی وجود دارد؟» پاسخ داد. چنانچه رفتار افراد از حیث توجه به پیام‌های ارسالی در رژیم‌های مختلف بازار متفاوت باشد، این پارامتر نیز می‌بایست در مدلسازی مد نظر قرار گیرد. نتایج بدست آمده از بررسی رفتار اشخاص در مشاهده انواع پیام‌ها در روندهای صعودی، نزولی و خنثی بازار به شرح جداول ۱۳ الی ۱۸ نشان می‌دهد رفتار افراد در روندهای مختلف بازار از نظر توجه به پیام‌های

مختلف و نیز تغییر توجه به آن‌ها تفاوت معنی‌داری با یکدیگر ندارد. لذا د ر مدلسازی نیز از روندهای بازار به عنوان یک متغیر استفاده نخواهیم کرد.

جدول ۱۳. آزمون مقایسه میانگین زوجی مشاهده پیام‌ها در روند خنثی و صعودی بازار

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	sig0	۱۳۱۸۵.۹۹۲۰	۲۸	۴۴۱۷.۴۵۸۸۴	۸۳۴.۸۲۱۲۵
	sig1	۱۳۰۰۶.۹۹۹۹	۲۸	۳۹۶۶.۷۸۳۱۹	۷۴۹.۶۵۱۵۶
Pair 2	tah0	۱۴۶۸۱.۰۹۱۷	۲۸	۵۵۶۵.۸۳۵۱۰	۱۰۵۱.۸۴۳۹۶
	tah1	۱۵۰۸۵.۱۷۱۰	۲۸	۵۲۶۷.۸۷۰۶۷	۹۹۵.۵۳۳۹۸

جدول ۱٤. آزمون مقایسه میانگین زوجی مشاهده پیام‌ها در روند خنثی و صعودی بازار

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	sig0 - sig1	۱۷۸.۹۹۲۱۲	۲۸۰۹.۷۹۸۶۲	۵۳۱.۰۰۲۰۳	-۹۱۰.۵۳۴۰۵	۱۲۶۸.۵۱۸۲۸	.۳۳۷	۲۷	.۷۳۹
Pair 2	tah0 - tah1	-۴۰۴.۰۷۹۲۹	۳۹۸۹.۵۰۶۰۸	۷۵۳.۹۴۵۷۸	-۱۹۵۱.۰۴۸۲۶	۱۱۴۲.۸۸۹۶۷	-.۵۳۶	۲۷	.۵۹۶

جدول ۱۵. آزمون مقایسه میانگین زوجی مشاهده پیام‌ها در روند خنثی و نزولی بازار

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	sign1	۱۳۱۳۲.۹۰۳۳	۱۸	۵۳۲۸.۸۲۵۴۰	۱۲۵۶.۰۱۶۱۹
	sig0	۱۴۱۰۷.۱۰۲۱	۱۸	۴۸۹۲.۵۶۱۲۳	۱۱۵۳.۱۸۷۷۴
Pair 2	tahn1	۱۵۱۶۲.۵۵۶۴	۱۸	۸۸۹۰.۴۵۷۹۲	۲۰۹۵.۵۰۱۰۳
	tah0	۱۴۴۵۲.۲۱۰۳	۱۸	۵۸۶۶.۹۵۴۳۸	۱۳۸۲.۸۵۴۴۱

جدول ۱٦. آزمون مقایسه میانگین زوجی مشاهده پیام‌ها در روند خنثی و نزولی بازار

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	sign1 - sig0	-۹۷۴.۱۹۸۸۸	۳۸۷۶.۶۹۱۷۲	۹۱۳.۷۴۵۰۰	-۲۹۰۲.۰۳۲۳۲	۹۵۳.۶۳۴۵۶	-۱.۰۶۶	۱۷	.۳۰۱
Pair 2	tahn1 - tah0	۷۱۰.۳۴۶۱۳	۵۴۶۹.۲۳۳۰۲	۱۲۸۹.۱۱۰۵۸	-۲۰۰۹.۴۳۹۴۶	۳۴۳۰.۱۳۱۷۳	.۵۵۱	۱۷	.۵۸۹

جدول ۱۷. آزمون مقایسه میانگین زوجی مشاهده پیام‌ها در روند نزولی و صعودی بازار

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	sign1	۱۳۱۳۲.۹۰۳۳	۱۸	۵۳۲۸.۸۲۵۴۰	۱۲۵۶.۰۱۶۱۹
	sig1	۱۳۲۸۵.۳۶۶۲	۱۸	۴۳۳۲.۴۲۶۰۳	۱۰۲۱.۱۶۲۶۱
Pair 2	tahn1	۱۵۱۶۲.۵۵۶۴	۱۸	۸۸۹۰.۴۵۷۹۲	۲۰۹۵.۵۰۱۰۳
	tah1	۱۴۲۶۲.۴۱۲۱	۱۸	۴۹۳۲.۵۷۶۲۸	۱۱۶۲.۶۱۹۳۸

جدول ۱۸. آزمون مقایسه میانگین زوجی مشاهده پیام‌ها در روند نزولی و صعودی بازار

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	sign1 - sig1	-۱۵۲.۴۶۲۹۳	۴۸۶۰.۹۹۴۶۵	۱۱۴۵.۷۴۷۴۳	-۲۵۶۹.۷۷۸۷۰	۲۲۶۴.۸۵۲۸۴	-.۱۳۳	۱۷	.۸۹۶
Pair 2	tahn1 - tah1	۹۰۰.۱۴۴۳۱	۷۸۶۵.۱۰۳۸۷	۱۸۵۳.۸۲۲۷۶	-۳۰۱۱.۰۷۹۸۳	۴۸۱۱.۳۶۸۴۴	.۴۸۶	۱۷	.۶۳۳

فرضیه‌ها

نتایج به دست آمده از اجرای ۶ رویکرد داده‌ای در شبکه عصبی (در نظر گرفتن متغیرهای مختلف) با استفاده از ۶ الی ۱۰ نورون در لایه پنهان و نیز در نظر گرفتن ۱ الی ۵ وقفه زمانی در تمامی متغیرهای مستقل و ۱۰ مرتبه اجرا در هر حالت، با توجه به اینکه ۳۶ نماد معاملاتی مورد بررسی بود، تعداد ۹,۰۰۰ (۱۰*۳۶*۵*۵) مرتبه اجرای مدل را به همراه داشت. بر اساس جدول ۱۹ مدل‌های ۱ و ۲ که دارای محتوای پیام‌های مهم و نیز پیام‌های ارسالی توسط اشخاص با اهمیت است، قدرت پیش‌بینی جهتی بالاتری نسبت به سایر مدل‌ها را نشان می‌دهد بدین معنی که مدل ۱ و ۲ حدوداً ۶۹ درصد قدرت پیش‌بینی داشته و سایر مدل‌ها قدرت پیش‌بینی حدود ۶۰ درصد را دارا می‌باشند. از سوی دیگر چارک سوم مدل اول ۶۱ درصد بوده در حالی که این آماره برای سایر مدل‌ها شامل اعداد بسیار کمتری بوده و در محدوده ۶۲ درصد است. به منظور بررسی معنی‌داری نتایج و مقایسه آن‌ها با یکدیگر، مدل اول با سایر مدل‌ها به شرح جدول ۲۰ و ۲۱ مقایسه شد. از سوی دیگر با توجه به اینکه مدل دوم شباهت زیادی با مدل اول دارد و تنها دو متغیر از متغیرهای مدل اول را در خود ندارد، مدل دوم نیز با سایر مدل‌ها به شرح جداول ۲۲ و ۲۳ مقایسه شد. مقایسه‌های یادشده تماماً نشان‌دهنده بازدهی بالاتر مدل‌هایی است که در

آن پیام‌ها افزون بر تفکیک محتوایی بر اساس میزان مشاهده پیام که نشان دهنده با اهمیت بودن پیام است و نیز میزان اهمیت فرستنده تفکیک شده‌اند. نتایج بررسی‌های به عمل آمده نشان داد مدل‌های ۱ و ۲ تماماً به صورت معنی داری دارای قدرت پیش‌بینی جهتی بالاتر بوده و به طور متوسط قدرت پیش‌بینی مدل‌های یادشده حدوداً ۹ درصد از قدرت پیش بینی سایر مدل‌های ارائه شده در این پژوهش بیشتر است. از سوی دیگر با توجه به شباهت مدل‌های ۱ و ۲، مقایسه معنی داری فاصله ۰,۳ درصدی میان قدرت پیش‌بینی هر دو مدل نشان می‌دهد افزودن متغیرهای تحلیل بنیادی سایر و تحلیل تکنیکال سایر در بهبود قدرت پیش‌بینی مدل موثر است. شکل ۱ توزیع کرنل مقادیر پیش‌بینی جهتی مدل‌ها را نشان می‌دهد. همانگونه که مشاهده می‌شود، مدل‌های ۱ و ۲ دارای میانگین بیشتری از سایر مدل‌ها بوده ولی انحراف معیار کمتری نسبت به آن‌ها دارند و۲ دارای ارتفاع کمتری نسبت به سایرین هستند. مدل شماره ۱ نسبت به مدل ۲ دارای فراوانی کمتری از مشاهدات در بخش میانی داده‌ها بوده ولی در سمت راست توزیع دارای فراوانی بیشتری (مطلوبتری) نسبت به مدل شماره ۲ است.در نتیجه، با توجه به آزمون مقایسه میانگین زوجی انجام شده و مقادیر p-value موجود در جداول ۲۶ و ۲۷ ، در سطح معنی داری ۱ درصد، فرضیه اول مبنی بر تفاوت در قدرت پیش‌بینی جهتی قیمت در مدل‌هایی که در آن افزون بر داده‌های تاریخی سهم از نظرات اشخاص صاحب شهرت نیز استفاده شده است، رد نمی‌شود. همچنین با توجه به تفاوت ۸,۹ درصدی در قدرت پیش‌بینی مدل ۱ نسبت به مدل ۳ که در سطح معنی‌داری ۱ درصد تفاوت مذکور معنی‌دار است، فرضیه سوم مبنی بر متفاوت بودن قدرت پیش‌بینی مدلی که در آن از نظرات اشخاص دارای شهرت افزون بر نوع محتوای استفاده شده است نسبت به مدلی که بر محتوای ارسالی بر اساس نظرات اشخاص دارای شهرت تغییر نکرده است، رد نمی‌شود.

جدول ۱۹. آماره های مقادیر پیش بینی جهتی مدل‌ها

آماره/مدل	مدل ۱	مدل ۲	مدل ۳	مدل ۴	مدل ۵	مدل ۶
میانگین	۰.۶۹۲۸	۰.۶۸۹۷	۰.۶۰۳	۰.۶۰۱۲	۰.۵۹۶۸	۰.۵۹۶۱
انحراف معیار	۰.۰۳۷۵	۰.۰۳۴۸	۰.۰۲۸۷	۰.۰۲۹۳	۰.۰۲۸۲	۰.۰۲۸۱
چارک اول	۰.۶۶۶۲	۰.۶۶۶	۰.۵۸۳۶	۰.۵۸۱۷	۰.۵۷۷۳	۰.۵۷۵۷
میانه	۰.۶۹۰۷	۰.۶۸۷۵	۰.۶۰۱۸	۰.۶۰۲	۰.۵۹۶۴	۰.۵۹۶
چارک سوم	۰.۷۱۸	۰.۷۱۱۷	۰.۶۲۳۱	۰.۶۲۰۳	۰.۶۱۶۴	۰.۶۱۷۲

جدول ۲۰. آزمون مقایسه میانگین زوجی قدرت پیش بینی جهتی مدل‌ها

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	مدل ۱	.۶۹۲۸	۹۰۰	.۰۳۷۵۴	.۰۰۱۲۵
	مدل ۲	.۶۸۹۷	۹۰۰	.۰۳۴۸۰	.۰۰۱۱۶
Pair 2	مدل ۱	.۶۹۲۸	۹۰۰	.۰۳۷۵۴	.۰۰۱۲۵
	مدل ۳	.۶۰۳۰	۹۰۰	.۰۲۸۷۳	.۰۰۰۹۶
Pair 3	مدل ۱	.۶۹۲۸	۹۰۰	.۰۳۷۵۴	.۰۰۱۲۵
	مدل ۴	.۶۰۱۲	۹۰۰	.۰۲۹۳۳	.۰۰۰۹۸
Pair 4	مدل ۱	.۶۹۲۸	۹۰۰	.۰۳۷۵۴	.۰۰۱۲۵
	مدل ۵	.۵۹۶۸	۹۰۰	.۰۲۸۲۰	.۰۰۰۹۴
Pair 5	مدل ۱	.۶۹۲۸	۹۰۰	.۰۳۷۵۴	.۰۰۱۲۵
	مدل ۶	.۵۹۶۱	۹۰۰	.۰۲۸۱۰	.۰۰۰۹۴

جدول ۲۱. آزمون مقایسه میانگین زوجی قدرت پیش بینی جهتی مدل‌ها

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	مدل ۱ -مدل ۲	.۰۰۳۱۴	.۰۲۸۰۶	.۰۰۰۹۴	.۰۰۱۳۱	.۰۰۴۹۸	۳.۳۶۰	۸۹۹	.۰۰۱
Pair 2	مدل ۱ -مدل ۳	.۰۸۹۷۸	.۰۳۳۸۶	.۰۰۱۱۳	.۰۸۷۵۷	.۰۹۲۰۰	۷۹.۵۳۶	۸۹۹	.۰۰۰
Pair 3	مدل ۱ -مدل ۴	.۰۹۱۶۰	.۰۳۳۹۴	.۰۰۱۱۳	.۰۸۹۳۸	.۰۹۳۸۲	۸۰.۹۶۸	۸۹۹	.۰۰۰
Pair 4	مدل ۱ -مدل ۵	.۰۹۵۹۷	.۰۳۲۶۹	.۰۰۱۰۹	.۰۹۳۸۳	.۰۹۸۱۱	۸۸.۰۷۴	۸۹۹	.۰۰۰
Pair 5	مدل ۱ -مدل ۶	.۰۹۶۶۷	.۰۳۲۵۱	.۰۰۱۰۸	.۰۹۴۵۵	.۰۹۸۸۰	۸۹.۱۹۶	۸۹۹	.۰۰۰

جدول ۲۲. آزمون مقایسه میانگین زوجی قدرت پیش بینی جهتی مدل‌ها حذف پیام‌های خنثی

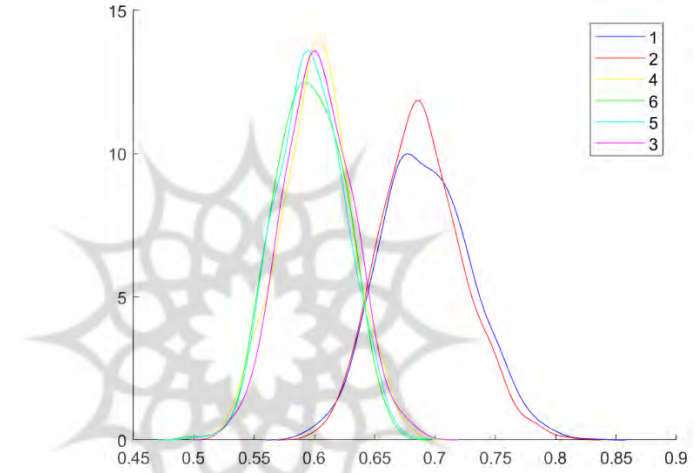
Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	مدل ۲	.۶۸۹۷	۹۰۰	.۰۳۴۸۰	.۰۰۱۱۶
	مدل ۳	.۶۰۳۰	۹۰۰	.۰۲۸۷۳	.۰۰۰۹۶
Pair 2	مدل ۲	.۶۸۹۷	۹۰۰	.۰۳۴۸۰	.۰۰۱۱۶
	مدل ۴	.۶۰۱۲	۹۰۰	.۰۲۹۳۳	.۰۰۰۹۸
Pair 3	مدل ۲	.۶۸۹۷	۹۰۰	.۰۳۴۸۰	.۰۰۱۱۶
	مدل ۵	.۵۹۶۸	۹۰۰	.۰۲۸۲۰	.۰۰۰۹۴
Pair 4	مدل ۲	.۶۸۹۷	۹۰۰	.۰۳۴۸۰	.۰۰۱۱۶
	مدل ۶	.۵۹۶۱	۹۰۰	.۰۲۸۱۰	.۰۰۰۹۴

جدول ۲۳. آزمون مقایسه میانگین زوجی قدرت پیش بینی جهتی مدل‌ها حذف پیام‌های خنثی

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	مدل ۲ -مدل ۳	۰۸۶۶۴.	۰۳۲۹۸.	۰۰۱۱۰.	۰۸۴۴۸.	۰۸۸۸۰.	۷۸.۷۹۹	۸۹۹	۰۰۰.
Pair 2	مدل ۲ -مدل ۴	۰۸۸۴۶.	۰۳۲۶۹.	۰۰۱۰۹.	۰۸۶۳۲.	۰۹۰۶۰.	۸۱.۱۷۳	۸۹۹	۰۰۰.
Pair 3	مدل ۲ -مدل ۵	۰۹۲۸۲.	۰۳۰۹۴.	۰۰۱۰۳.	۰۹۰۸۰.	۰۹۴۸۵.	۸۹.۹۹۲	۸۹۹	۰۰۰.
Pair 4	مدل ۲ -مدل ۶	۰۹۳۵۳.	۰۳۱۹۵.	۰۰۱۰۶.	۰۹۱۴۴.	۰۹۵۶۲.	۸۷.۸۲۶	۸۹۹	۰۰۰.



شکل ۱. توزیع کرنل مقادیر پیش بینی جهتی مدل‌ها

به منظور بررسی فرضیه‌ها ۲ و ۴، شبکه‌های عصبی مشابه با شرایط شبکه‌هایی که به منظور پیش‌بینی باینری (جهتی) مورد استفاده قرار گرفته بود، اجرا شد. جدول ۲۴ نشان دهنده خطای RMSE مدل شماره ۱ کمتر از سایر مدل‌ها بوده و این تفاوت بر اساس آزمون مقایسه میانگین زوجی مندرج در جدول ۲۶ معنی‌دار است. هرچند خطای RMSE مدل شماره ۱ تفاوت بسیار کمی با خطای RMSE مدل ۲ دارد اما تفاوت یادشده نیز معنی دار است. بر اساس مقادیر مندرج در جداول ۲۷ و ۲۸ نیز خطای پیش‌بینی مقادیر بازدهی در مدل شماره ۲ با سایر مدل‌ها دارای تفاوت معنی دار است. بنابراین می‌توان نتیجه گرفت افزون بر تفکیک نمودن پیام‌ها بر اساس

نوع و جهت پیام، ایجاد تفکیک بین فرستنده پیام و میزان مشاهده پیام می‌تواند در بهبود پیش‌بینی مدل موثر باشد. تفاوت خطای RMSE مدل‌های ۱ و ۲ با سایر مدل‌ها حدوداً ۰,۰۰۲۸ است. شکل ۲ نشان‌دهنده توزیع کرنل مقادیر پیش‌بینی مدل‌های مختلف است. با توجه به اینکه توزیع‌ها دارای مقادیر متفاوتی در دُم سمت راست هستند، در شکل ۳ مقادیر پیش‌بینی مدل‌ها در ۹ دهک ابتدایی ارائه شده است. همانگونه که در هر دو شکل مشاهده می‌شود، توزیع خطای RMSE در مدل‌های ۱ و ۲ دارای میانگین کمتری از سایر روش‌ها است. از سوی دیگر روش ۱ دارای فله (بیشترین فراوانی) در سمت چپ روش ۲ است که نشان از خطای کمتر مدل شماره ۱ از مدل ۲ دارد. از سوی دیگر با توجه به مقادیر موجود در جدول ۲۵ و ۲۸، میانگین خطای مدل اول به صورت معنی‌داری کمتر از میانگین خطای مدل سوم است.

جدول ۲٤. آماره های مقادیر پیش بینی بازدهی مدل‌ها

آماره/مدل	مدل ۱	مدل ۲	مدل ۳	مدل ۴	مدل ۵	مدل ۶
میانگین	۰.۰۲۴۶	۰.۰۲۴۷	۰.۰۲۷۰	۰.۰۲۷۱	۰.۰۲۷۵	۰.۰۲۷۷
انحراف معیار	۰.۰۰۲۳	۰.۰۰۲۳	۰.۰۰۳۱	۰.۰۰۳۲	۰.۰۰۳۵	۰.۰۰۳۶
چارک اول	۰.۰۲۳۰	۰.۰۲۳۱	۰.۰۲۴۹	۰.۰۲۵۱	۰.۰۲۵۴	۰.۰۲۵۶
میانه	۰.۰۲۴۵	۰.۰۲۴۷	۰.۰۲۶۹	۰.۰۲۶۹	۰.۰۲۷۲	۰.۰۲۷۳
چارک سوم	۰.۰۲۶۲	۰.۰۲۶۳	۰.۰۲۸۵	۰.۰۲۸۶	۰.۰۲۸۹	۰.۰۲۹۱

جدول ۲۵. آزمون مقایسه میانگین زوجی قدرت پیش بینی مقداری بازده مدل‌ها

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	مدل ۱	.۰۲۴۶	۹۰۰	.۰۰۲۲۶	.۰۰۰۰۸
	مدل ۲	.۰۲۴۷	۹۰۰	.۰۰۲۳۴	.۰۰۰۰۸
Pair 2	مدل ۱	.۰۲۴۶	۹۰۰	.۰۰۲۲۶	.۰۰۰۰۸
	مدل ۳	.۰۲۷۰	۹۰۰	.۰۰۳۱۴	.۰۰۰۱۰
Pair 3	مدل ۱	.۰۲۴۶	۹۰۰	.۰۰۲۲۶	.۰۰۰۰۸
	مدل ۴	.۰۲۷۱	۹۰۰	.۰۰۳۱۵	.۰۰۰۱۱
Pair 4	مدل ۱	.۰۲۴۶	۹۰۰	.۰۰۲۲۶	.۰۰۰۰۸
	مدل ۵	.۰۲۷۵	۹۰۰	.۰۰۳۵۴	.۰۰۰۱۲
Pair 5	مدل ۱	.۰۲۴۶	۹۰۰	.۰۰۲۲۶	.۰۰۰۰۸
	مدل ۶	.۰۲۷۷	۹۰۰	.۰۰۳۵۵	.۰۰۰۱۲

جدول ۲۶. آزمون مقایسه میانگین زوجی قدرت پیش بینی مقداری بازده مدل‌ها

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	مدل ۱ -مدل ۲	-.۰۰۰۱۶	.۰۰۱۰۹	.۰۰۰۰۴	-.۰۰۰۲۳	-.۰۰۰۰۹	-۴.۴۲۴	۸۹۹	.۰۰۰
Pair 2	مدل ۱ -مدل ۳	-.۰۰۲۴۲	.۰۰۲۳۳	.۰۰۰۰۸	-.۰۰۲۵۸	-.۰۰۲۲۷	-۳۱.۲۴۴	۸۹۹	.۰۰۰
Pair 3	مدل ۱ -مدل ۴	-.۰۰۲۵۱	.۰۰۲۲۴	.۰۰۰۰۷	-.۰۰۲۶۶	-.۰۰۲۳۶	-۳۳.۵۵۷	۸۹۹	.۰۰۰
Pair 4	مدل ۱ -مدل ۵	-.۰۰۲۸۸	.۰۰۲۷۶	.۰۰۰۰۹	-.۰۰۳۰۶	-.۰۰۲۷۰	-۳۱.۲۳۷	۸۹۹	.۰۰۰
Pair 5	مدل ۱ -مدل ۶	-.۰۰۳۰۹	.۰۰۲۷۶	.۰۰۰۰۹	-.۰۰۳۲۷	-.۰۰۲۹۱	-۳۳.۵۴۷	۸۹۹	.۰۰۰

جدول ۲۷. آزمون مقایسه میانگین زوجی قدرت پیش بینی مقداری بازده مدل‌ها حذف پیام‌های خنثی

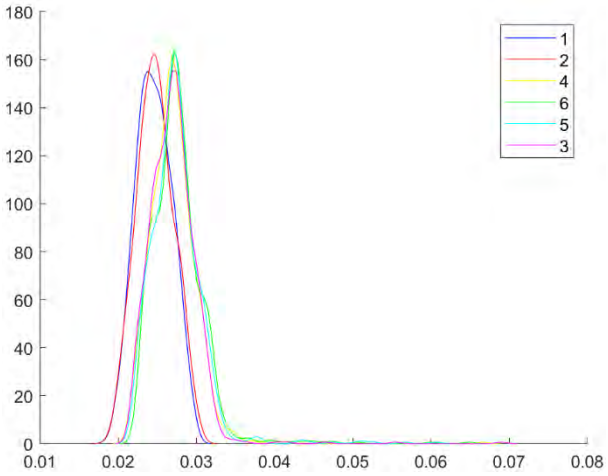
Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	مدل ۲	.۰۲۴۷	۹۰۰	.۰۰۲۳۴	.۰۰۰۰۸
	مدل ۳	.۰۲۷۰	۹۰۰	.۰۰۳۱۴	.۰۰۰۱۰
Pair 2	مدل ۲	.۰۲۴۷	۹۰۰	.۰۰۲۳۴	.۰۰۰۰۸
	مدل ۴	.۰۲۷۱	۹۰۰	.۰۰۳۱۵	.۰۰۰۱۱
Pair 3	مدل ۲	.۰۲۴۷	۹۰۰	.۰۰۲۳۴	.۰۰۰۰۸
	مدل ۵	.۰۲۷۵	۹۰۰	.۰۰۳۵۴	.۰۰۰۱۲
Pair 4	مدل ۲	.۰۲۴۷	۹۰۰	.۰۰۲۳۴	.۰۰۰۰۸
	مدل ۶	.۰۲۷۷	۹۰۰	.۰۰۳۵۵	.۰۰۰۱۲

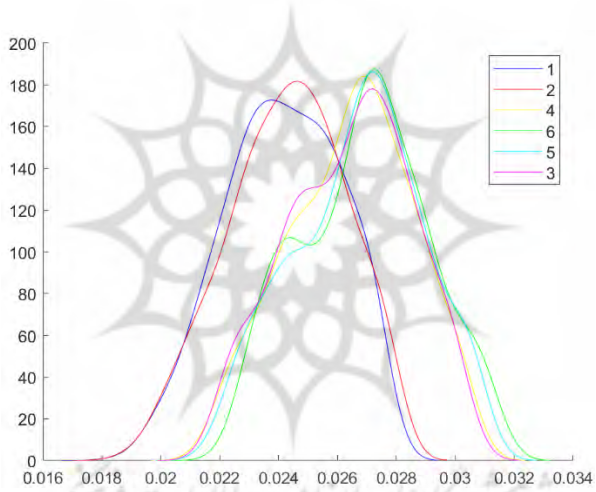
جدول ۲۸. آزمون مقایسه میانگین زوجی قدرت پیش بینی مقداری بازده مدل‌ها حذف پیام‌های خنثی

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	مدل ۲ -مدل ۳	-.۰۰۲۲۶	.۰۰۲۳۰	.۰۰۰۰۸	-.۰۰۲۴۱	-.۰۰۲۱۱	-۲۹.۴۸۶	۸۹۹	.۰۰۰
Pair 2	مدل ۲ -مدل ۴	-.۰۰۲۳۵	.۰۰۲۲۱	.۰۰۰۰۷	-.۰۰۲۴۹	-.۰۰۲۲۰	-۳۱.۸۷۱	۸۹۹	.۰۰۰
Pair 3	مدل ۲ -مدل ۵	-.۰۰۲۷۲	.۰۰۲۷۴	.۰۰۰۰۹	-.۰۰۲۸۹	-.۰۰۲۵۴	-۲۹.۷۷۶	۸۹۹	.۰۰۰
Pair 4	مدل ۲ -مدل ۶	-.۰۰۲۹۳	.۰۰۲۷۴	.۰۰۰۰۹	-.۰۰۳۱۱	-.۰۰۲۷۵	-۳۲.۰۰۹	۸۹۹	.۰۰۰



شکل ۲. توزیع کرنل مقادیر پیش بینی بازدهی مدل‌ها



شکل ۳. توزیع کرنل مقادیر پیش بینی بازدهی مدل‌ها (داده‌های ۹ دهک اول)

بحث، نتیجه‌گیری و ارائه پیشنهادها

در سال‌های اخیر، شبکه‌های اجتماعی جایگاه ویژه‌ای در زندگی روزمره افراد پیدا کرده‌اند. بدین صورت که بخش عمده‌ی جمعیت کشور حداقل در یکی از شبکه‌های اجتماعی عضویت دارند. بنابراین تبادل نظر در زمینه‌های مختلف با محدودیت‌های مربوط به موقعیت جغرافیایی و حتی زمانی کنار رفته است. به گونه‌ای که افراد می‌توانند در هر ساعت از شبانه روز محتوای مد

نظر خود را منتشر کند و مخاطبین نیز در مشاهده پیام یادشده دارای اختیار هستند. در این شرایط با توجه به تمایل افراد به بهره‌مندی از نظرات سایرین و نیز با توجه به اینکه برخی افراد دارای تخصص لازم در زمینه‌های مختلف نیستند، استفاده از نظرات سایر اشخاص دارای اهمیت است. بدین صورت که افراد یا به سخنان یک یا چند شخص خاص توجه بیشتری می‌کنند. یا تحت تاثیر تکرار سخنان مشابه قرار می‌گیرند. در این حالت بیشتر تصمیمات افراد، مستقل شمرده نمی‌شود و می‌تواند تصمیمات بیان شده را مرتبط به سایر عوامل مربوط به شبکه‌های اجتماعی دانست. در سال‌های اخیر تحلیل احساسات موجود در شبکه‌های اجتماعی نقش بسیار مهمی در پژوهش‌های علمی یافته است. پژوهشگران با رویکردهای مختلف سعی در طبقه‌بندی و برچسب زنی پیام‌ها از نظر احساسات می‌کنند. در این پژوهش نیز با توجه به موارد یاد شده، تمامی پیام‌های قابل بررسی در شبکه اجتماعی تلگرام از بیش از ۱۳۰۰ کانال و گروه تلگرامی موجود در بازار سرمایه جمع‌آوری شد. سپس پیام‌هایی که مربوط به نمادهای موجود در شاخص ۳۰ شرکت بود (۳۶ شرکت در بازه زمانی ابتدای سال ۱۳۹۸ الی پایان مرداد ۱۴۰۰)، با استفاده از فهرست واژگان ساخته شده در حداقل یکی از دسته‌های بنیادی، تکنیکال، سیگنال خرید و سیگنال فروش قرار گرفت. برای دسته‌بندی پیام‌ها در هر یک از ۴ دسته یادشده کلماتی انتخاب و براساس این که چه تعداد از کلمات هر دسته در هر پیام وجود داشته باشد به هر دسته امتیازی داده شد که میزان تعلق هر پیام به هر دسته را نشان دهد. سپس برای اعطای امتیاز مثبت یا منفی، هر پیام به جملات تبدیل شد و امتیاز جملات بر اساس تعداد کلمات مثبت و منفی در هر دسته مشخص شد و امتیاز نهایی برابر با جمع امتیاز جملات در نظر گرفته شد. به منظور تفکیک اهمیت پیام و نیز اهمیت فرستنده پیام‌ها یکبار از نظر میزان مشاهده دهک بندی شده و پیام‌هایی که میزان مشاهده آن‌ها در دهک بالایی قرار داشت، به عنوان پیام مهم طبقه‌بندی شد. در خصوص فرستنده نیز متوسط مشاهده پیام‌های افراد دهک‌بندی شد و افرادی که در دهک بالایی قرار داشتند به عنوان اشخاص با اهمیت تلقی شدند. در نهایت به توجه به موارد یادشده، داده‌های مربوط به تحلیل احساسات (حدودا ۱۶ میلیون داده) اشخاص همراه با داده‌های بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول به ازای هر سهم در مدلسازی مورد استفاده قرار گرفت. به منظور بررسی تاثیرگذاری تفکیک

نمودن داده‌ها از یکدیگر، ۶ رویکرد داده‌ای مد نظر قرار گرفت. در مدل ۱ پیام‌های شبکه‌های اجتماعی بر اساس نوع محتوا، نوع منشا و نیز میزان مشاهده پیام تفکیک شده و افزون بر آن از سایر داده‌های مورد اشاره نیز استفاده شد. مدل شماره ۲ مشابه با مدل شماره ۱ بوده ولی دو متغیر مربوط به تحلیل بنیادی و تحلیل تکنیکالی که به لحاظ احساسی خنثی در نظر گرفته شده حذف شدند. در مدل شماره ۳ از نوع محتوا نظیر تحلیل بنیادی، تحلیل تکنیکال و ... به عنوان محتوای احساسی استفاده شد. محتوای احساسی در مدل شماره ۴ تنها شامل جهت پیام بود و در مدل شماره ۵ محتوای احساسی تنها نشان دهنده تعداد پیام ارسالی در خصوص نماد بود. مدل شماره ۶ نیز فاقد محتوای احساسی مدلسازی شد. به ازای هر یک از رویکردهای داده‌ای یادشده (۶ رویکرد) برای هر نماد معاملاتی (۳۶ نماد) دو نوع پیش‌بینی جهتی و مقداری با استفاده از شبکه عصبی پیش خور صورت گرفت. بدین صورت که شبکه عصبی با تعداد بین ۵ الی ۱۰ نرون در لایه پنهان مدلسازی شد و در هر حالت از وقفه‌های ۱ الی ۵ روزه برای داده‌ها استفاده شد. در نتیجه به منظور خنثی کردن اثر مقادیر اولیه وزن‌ها در شبکه عصبی، هر مدل ۱۰ مرتبه اجرا شد. بدین صورت ۱۰۸,۰۰۰ مدل اجرا شد.

در پاسخ به پرسش مربوط به نوع غالب محتوای منتشرشده در شبکه‌های اجتماعی و پیام‌های توجه مخاطبان در بازار سرمایه، نتایج حاصل از آزمون مقایسه میانگین زوجی نشان داد حدود ۳۶ درصد از پیام‌های ارسالی به صورت معنی‌داری در قالب تحلیل طبقه‌بندی می‌شوند و سایر پیام‌ها از انواع سیگنال‌ها هستند. بررسی میزان توجه به پیام‌ها نیز نشان دهنده بالاتر بودن میانگین مشاهده پیام‌های سیگنال نسبت به پیام‌های تحلیلی به صورت معنی‌دار بود. بررسی تغییر در توجه افراد به پیام‌ها در رژیم‌های صعودی، خنثی و نزولی بازار نیز نشان دهنده معنی دار نبودن تفاوت در میزان مشاهده پیام‌ها بود بنابراین رژیم‌های بازار در مدلسازی مدنظر قرار نگرفت.

در خصوص فرضیه‌ها نیز نتایج حاصل از مدل‌های یادشده در پیش‌بینی جهتی نشان دهنده معنی داری برتری قدرت پیش‌بینی مدل شماره ۱ (حاوی تفکیک پیام‌ها از لحاظ محتوا، نوع فرستنده و نوع اهمیت پیام) نسبت به سایر مدل‌ها بود. به صورت مشابه برای پیش‌بینی مقداری نیز خطای RMSE برای مدل شماره ۱ به صورت معنی داری از سایر مدل‌ها کمتر بود.

بنابراین به شرکت‌های سرمایه‌گذاری و نیز سایر اشخاصی که امکان پیاده‌سازی بسترهای خزش و واکشی داده‌های موجود در شبکه‌های اجتماعی مانند تلگرام را دارند، پیشنهاد می‌شود با دریافت داده‌های مورد اشاره و نیز ایجاد واژه‌نامه احساسی از کلمات فارسی، متغیرهای موجود

در پژوهش حاضر را مورد استفاده قرار داده و از آن در پیش‌بینی بازده و نیز انجام معامله به صورت الگوریتمی استفاده کنند. در این خصوص ایجاد بسترهای تشخیص برخط احساسات نیز می‌تواند در پیش‌بینی بازده به صورت لحظه‌ای و انجام معامله درون روز و نیز با تایم فریم‌های مختلف کمک‌کننده باشد. از سوی دیگر امکان انجام معامله به صورت HFT نیز با متغیرهای موصوف ساده‌تر می‌شود. همچنین متغیرهای موجود در این پژوهش به ارکان نظارتی بازار نیز کمک می‌کند تا از علل واکنش قیمت سهم در شرایط خاص مطلع شده و بتواند در صورتی که رفتارغیرنرمال در متغیرهای بیان شده مشاهده کند موضوع را بررسی کرده و در صورت مشاهده بدرفتاری تعمدی، موارد را پیگیری کند.

به پژوهشگران نیز پیشنهاد می‌شود دهک بندی صورت گرفته در این پژوهش را به صورت پویا انجام داده یا از دهک‌های دیگر نیز در تشخیص با اهمیت بودن فرد یا پیام استفاده کنند. هرچند مطلوب آن است که بازه دهک بندی نیز محدودتر شود ولی با توجه به فراوانی داده‌ها نیز محتمل است نتیجه مطلوب حاصل نشود. از سوی دیگر با توجه به اینکه برخی از محتوای ارسالی، متاثر از اخبار موجود هستند، تفکیک اخبار و استفاده از آن به عنوان یک رویداد می‌تواند در بهبود مدل موثر واقع شود.

ملاحظات اخلاقی

پیروی از اصول اخلاق پژوهش

نویسندگان اصول اخلاقی را در انجام و انتشار این پژوهش علمی رعایت کرده‌اند و این موضوع مورد تأیید همه آنهاست.

مشارکت نویسندگان

جمع‌آوری داده‌ها: سروش خواجه حق وردی تهیه گزارش پژوهش: سروش خواجه حق وردی تحلیل داده‌ها: سروش خواجه حق وردی، شاپور محمدی، رضا راعی.

مشارکت نویسندگان در مقاله مستخرج از پایان‌نامه تقریباً به شکل زیر باشد:

نویسنده اول: تهیه و آماده‌سازی نمونه‌ها، انجام آزمایش و گردآوری داده‌ها، انجام محاسبات، تجزیه و تحلیل آماری داده‌ها، تحلیل و تفسیر اطلاعات و نتایج، تهیه پیشنویس مقاله؛

نویسنده دوم: استاد راهنمای پایان‌نامه، طراحی پژوهش، نظارت بر مراحل انجام پژوهش، بررسی و کنترل نتایج، اصلاح، بازبینی و نهایی‌سازی مقاله؛

نویسنده سوم: استاد مشاور پایان‌نامه، مشارکت در طراحی پژوهش، نظارت بر پژوهش، مطالعه و بازبینی مقاله.

تعارض منافع

بنا بر اظهار نویسندگان این مقاله تعارض منافع ندارد.

سپاسگزاری

از معاونت محترم پژوهشی دانشکده مدیریت دانشگاه تهران به خاطر حمایت معنوی در اجرای پژوهش حاضر سپاسگزاری می‌شود.

از آقای دکتر شاپور محمدی و رضا راعی به خاطر بازبینی متن مقاله و ارائه نظرهای ساختاری تشکر و قدردانی می‌شود.

از داوران محترم به خاطر ارائه نظرهای ساختاری و علمی سپاسگزاری می‌شود.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

References

Aggarwal, C. C; & Zhai, C. (Eds.). (2012). Mining text data. Springer Science & Business Media. https://doi.org/10.1007/978-1-4614-3223-4_1

Bollen, J; Mao, H; & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2(1), 1-8. https://doi.org/10.1016/j.jocs.2010.12.007

Baker, M; & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of economic perspectives*, 21(2), 129-152. DOI: 10.1257/jep.21.2.129

Oh, C; & Sheng, O. (2011). Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. https://aisel.aisnet.org/icis2011/proceedings/knowledge/17/

Duggan, M; Ellison, N. B; Lampe, C; Lenhart, A; & Madden, M. (2015). Social media update 2014. *Pew research center*, 19, 1-2. https://media.myworshiptimes22.com/wp-content/uploads/sites/6/2015/07/20140539/Social-Media-Site-Usage-2014-_-Pew-Research-Centers-Internet-American-Life-Project.pdf

Hong, H; & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of finance*, 54(6), 2143-2184. https://doi.org/10.1111/0022-1082.00184

Eppen, G. D; & Fama, E. F. (1969). Cash balance and simple dynamic portfolio problems with proportional costs. *International Economic Review*, 10(2), 119-133. https://doi.org/10.2307/2525547

Cooper. John C.B. (1999) ”Artificial neural network versus multivariate statistics: an application from economics”, *Journal of applied statistics*,vol.26”:8: 909-921. https://doi.org/10.1080/02664769921927

Medhat, W; Hassan, A; & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), 1093-1113. https://doi.org/10.1016/j.asej.2014.04.011

Neubeck, K. J; & Glasberg, D. S. (2004). Sociology: Diversity, conflict, and change. Recording for the Blind & Dyslexic. https://cir.nii.ac.jp/crid/1970304959948828299

Pradhan, V. M; Vala, J; & Balani, P. (2016). A survey on Sentiment Analysis Algorithms for opinion mining. *International Journal of Computer Applications*, 133(9), 7-11. DOI:10.5120/ijca2016907977

Sprenger, T. O; Tumasjan, A; Sandner, P. G; & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial*

Management, 20(5), 926-957. https://doi.org/10.1111/j.1468-036X.2013.12007.x

Shepard, J. M. (2012). Cengage advantage books: Sociology. Nelson Education. book link

Tsay, R. S. (2005). Analysis of financial time series. John wiley & sons. book link

Vinodhini, G; & Chandrasekaran, R. M. (2012). *Sentiment analysis and opinion mining: a survey*. *International Journal*, 2(6), 282-292. https://www.researchgate.net/profile/Vinodhini-G-2/publication/265163299_Sentiment_Analysis_and_Opinion_Mining_A_Survey/links/54018f330cf2bba34c1af133/Sentiment-Analysis-and-Opinion-Mining-A-Survey.pdf

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

پیوست

مراحل انجام پژوهش:

۱. در مرحله اول ابتدا تمامی پیام‌های حاوی نام نماد یا نام شرکت در شبکه اجتماعی تلگرام مورد تفکیک قرار گرفته است و پس از آن با استفاده از ابزار هضم و بر أساس لغتنامه‌ی تهیه شده برای طبقه‌بندی نوع محتوا، پیام‌ها به دسته‌های بنیادی، تکنیکال، سیگنال خرید و سیگنال فروش تفکیک شدند.

۲. برای انجام آزمون مقایسه میانگین زوجی در پاسخ به سوال اول، نسبت انتشار پیام‌ها در دسته‌های مختلف به تفکیک سیگنال و تحلیل به صورت ماهانه مورد استفاده قرار گرفت و بر أساس آن آزمون در نرم افزار SPSS اجرا شد.

۳. در پاسخ به سوال دوم، متوسط میزان مشاهده پیام‌های سیگنالی و تحلیلی به صورت روزانه به عنوان داده‌های ورودی برای انجام آزمون مقایسه میانگین زوجی بوده و آزمون در نرم افزار SPSS اجرا شد.

۴. برای اجرای مدل‌های شبکه عصبی و نیز ماشین بردار پشتیبان (مربوط به فرضیه‌ها) ابتدا لازم بود تا محتوای منتشره بر اساس نوع فرستنده و نیز میزان توجه به پیام در طبقات مختلف تفکیک شود. بدین منظور متوسط مشاهده پیام‌های هر کانال به صورت ماهانه استخراج شد و کانال‌هایی که از نظر متوسط میزان مشاهده در دهک اول قرار گرفتند (متوسط میزان مشاهده پیام‌های آن‌ها از سایر کانال‌ها بیشتر بود)، به صورت مجزا علامت‌گذاری شدند. پیام‌ها نیز بر اساس متوسط مشاهده دهک بندی شده و موارد موجود در دهک اول علامت‌گذاری شد. سپس مدل‌ها بر اساس متغیرهای جدول زیر اجرا شد. در نهایت نیز آزمون مقایسه میانگین زوجی برای بررسی دقت یا قدرت پیش‌بینی مدل انجام شد.

روش	نوع پیش بینی	متغیر ورودی	متغیر خروجی	شماره مدل
شبکه عصبی مصنوعی	جهتی	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول،محتوای	بازده روزانه سهم (باینری)	۱

روش	نوع پیش بینی	متغیر ورودی	متغیر خروجی	شماره مدل
		احساسی شامل تحلیل بنیادی مثبت، تحلیل بنیادی منفی، تحلیل بنیادی سایر، تحلیل تکنیکال مثبت، تحلیل تکنیکال منفی، تحلیل تکنیکال سایر، سیگنال خرید و سیگنال فروش (در محتوای احساسی تفکیک موثر بودن یا نبودن فرستنده نیز لحاظ شده و همچنین مهم بودن پیام از حیث تعداد مشاهده نیز از جمله موارد تفکیک کننده بوده است)		
شبکه عصبی مصنوعی	جهتی	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، محتوای احساسی شامل؛ تحلیل بنیادی مثبت، تحلیل بنیادی منفی، تحلیل تکنیکال مثبت، تحلیل تکنیکال منفی، سیگنال خرید و سیگنال فروش (در محتوای احساسی تفکیک موثر بودن یا نبودن فرستنده نیز لحاظ شده و همچنین مهم بودن پیام از حیث تعداد مشاهده نیز از جمله موارد تفکیک کننده بوده است)	بازده روزانه سهم (باینری)	۲
شبکه عصبی مصنوعی	جهتی	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، محتوای احساسی به تفکیک تحلیل بنیادی، تحلیل تکنیکال، سیگنال خرید و سیگنال فروش	بازده روزانه سهم (باینری)	۳
شبکه عصبی مصنوعی	جهتی	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، محتوای احساسی به تفکیک مثبت، منفی و خنثی	بازده روزانه سهم (باینری)	۴
شبکه عصبی مصنوعی	جهتی	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت	بازده روزانه سهم (باینری)	۵

روش	نوع پیش بینی	متغیر ورودی	متغیر خروجی	شماره مدل
		روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، تعداد پیام منتشره		
شبکه عصبی مصنوعی	جهتی	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول	بازده روزانه سهم (باینری)	۶
ماشین بردار پشتیبان	جهتی	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول،محتوای احساسی شامل تحلیل بنیادی مثبت، تحلیل بنیادی منفی، تحلیل بنیادی سایر، تحلیل تکنیکال مثبت، تحلیل تکنیکال منفی، تحلیل تکنیکال سایر، سیگنال خرید و سیگنال فروش (در محتوای احساسی تفکیک موثر بودن یا نبودن فرستنده نیز لحاظ شده و همچنین مهم بودن پیام از حیث تعداد مشاهده نیز از جمله موارد تفکیک کننده بوده است)	بازده روزانه سهم (باینری)	۱
ماشین بردار پشتیبان	جهتی	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، محتوای احساسی شامل؛ تحلیل بنیادی مثبت، تحلیل بنیادی منفی، تحلیل تکنیکال مثبت، تحلیل تکنیکال منفی، سیگنال خرید و سیگنال فروش (در محتوای احساسی تفکیک موثر بودن یا نبودن فرستنده نیز لحاظ شده و همچنین مهم بودن پیام از حیث تعداد مشاهده نیز از جمله موارد تفکیک کننده بوده است)	بازده روزانه سهم (باینری)	۲
ماشین بردار پشتیبان	جهتی	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، محتوای	بازده روزانه سهم (باینری)	۳

روش	نوع پیش بینی	متغیر ورودی	متغیر خروجی	شماره مدل
		احساسی به تفکیک تحلیل بنیادی، تحلیل تکنیکال، سیگنال خرید و سیگنال فروش		
ماشین بردار پشتیبان	جهتی	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، محتوای احساسی به تفکیک مثبت، منفی و خنثی	بازده روزانه سهم (باینری)	۴
ماشین بردار پشتیبان	جهتی	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، تعداد پیام منتشره	بازده روزانه سهم (باینری)	۵
ماشین بردار پشتیبان	جهتی	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول	بازده روزانه سهم (باینری)	۶
شبکه عصبی مصنوعی	مقداری	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول،محتوای احساسی شامل تحلیل بنیادی مثبت، تحلیل بنیادی منفی، تحلیل بنیادی سایر، تحلیل تکنیکال مثبت، تحلیل تکنیکال منفی، تحلیل تکنیکال سایر، سیگنال خرید و سیگنال فروش (در محتوای احساسی تفکیک موثر بودن یا نبودن فرستنده نیز لحاظ شده و همچنین مهم بودن پیام از حیث تعداد مشاهده نیز از جمله موارد تفکیک کننده بوده است)	بازده روزانه سهم	۱
شبکه عصبی مصنوعی	مقداری	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت	بازده روزانه سهم	۲

روش	نوع پیش بینی	متغیر ورودی	متغیر خروجی	شماره مدل
		روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، محتوای احساسی شامل؛ تحلیل بنیادی مثبت، تحلیل بنیادی منفی، تحلیل تکنیکال مثبت، تحلیل تکنیکال منفی، سیگنال خرید و سیگنال فروش (در محتوای احساسی تفکیک موثر بودن یا نبودن فرستنده نیز لحاظ شده و همچنین مهم بودن پیام از حیث تعداد مشاهده نیز از جمله موارد تفکیک کننده بوده است)		
شبکه عصبی مصنوعی	مقداری	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، محتوای احساسی به تفکیک تحلیل بنیادی، تحلیل تکنیکال، سیگنال خرید و سیگنال فروش	بازده روزانه سهم	۳
شبکه عصبی مصنوعی	مقداری	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، محتوای احساسی به تفکیک مثبت، منفی و خنثی	بازده روزانه سهم	۴
شبکه عصبی مصنوعی	مقداری	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، تعداد پیام منتشره	بازده روزانه سهم	۵
شبکه عصبی مصنوعی	مقداری	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول	بازده روزانه سهم	۶
ماشین بردار پشتیبان	مقداری	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت	بازده روزانه سهم	۱

روش	نوع پیش بینی	متغیر ورودی	متغیر خروجی	شماره مدل
		مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول،محتوای احساسی شامل تحلیل بنیادی مثبت، تحلیل بنیادی منفی، تحلیل بنیادی سایر، تحلیل تکنیکال مثبت، تحلیل تکنیکال منفی، تحلیل تکنیکال سایر، سیگنال خرید و سیگنال فروش (در محتوای احساسی تفکیک موثر بودن یا نبودن فرستنده نیز لحاظ شده و همچنین مهم بودن پیام از حیث تعداد مشاهده نیز از جمله موارد تفکیک کننده بوده است)		
ماشین بردار پشتیبان	مقداری	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، محتوای احساسی شامل؛ تحلیل بنیادی مثبت، تحلیل بنیادی منفی، تحلیل تکنیکال مثبت، تحلیل تکنیکال منفی، سیگنال خرید و سیگنال فروش (در محتوای احساسی تفکیک موثر بودن یا نبودن فرستنده نیز لحاظ شده و همچنین مهم بودن پیام از حیث تعداد مشاهده نیز از جمله موارد تفکیک کننده بوده است)	بازده روزانه سهم	۲
ماشین بردار پشتیبان	مقداری	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، محتوای احساسی به تفکیک تحلیل بنیادی، تحلیل تکنیکال، سیگنال خرید و سیگنال فروش	بازده روزانه سهم	۳
ماشین بردار پشتیبان	مقداری	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، محتوای احساسی به تفکیک مثبت، منفی و خنثی	بازده روزانه سهم	۴

روش	نوع پیش بینی	متغیر ورودی	متغیر خروجی	شماره مدل
ماشین بردار پشتیبان	مقداری	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول، تعداد پیام منتشره	بازده روزانه سهم	۵
ماشین بردار پشتیبان	مقداری	بازدهی قیمتی روزانه، بازده قیمت حراج ناپیوسته آغازین بازار، بازده شاخص کل بورس اوراق بهادار تهران، نوسان قیمت ارز، نوسان قیمت نفت، نوسان قیمت طلا، نوسان قیمت مس، نوسان قیمت اوره، نوسان قیمت آلومینیوم، نوسان قیمت روی، نوسان قیمت سنگ آهن، نوسان قیمت متانول	بازده روزانه سهم	۶

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی