Journal of Advances in Finance and Investment
Volume 4, Issue 2 , 2023
pp. 1-30.
Paper type: Research paper

Ability of machine learning algorithms and artificial neural networks in predicting accounting profit information content before announcing it

Hossein Alizadeh[1], Majid Zanjirdar[2], Gholam Ali Haji[3]

Received: 24/10/2022 Accepted: 01/03/2023

Extended Abstract

Introduction

Research in recent years shows that investors' response to quarterly profit announcements of companies has increased significantly. Of course, most price changes happen before the official announcement of profits, which is caused by the transactions of the holders of confidential information. Beaver *et al.* (2020) approaches and theories about the commerce of the holders of confidential information are directed to two opposing views, and the main issue is the lack of understanding among researchers, the prohibition of the transactions of the holders of confidential information, or the lack of monitoring. The first point of view considers transactions based on confidential information to increase market efficiency, and the second is to decrease market efficiency. The current research is based on experts' opinions that trading based on secret information does not increase market efficiency. Machine learning and artificial intelligence for stock market forecasting have significantly been developed in recent years, and while the price behavior in the stock market seems chaotic, the use of artificial intelligence to make calculated predictions before investing is highly recommended. Chaajer *et al.* (2021), and therefore, in this research, considering the mentioned contradictions, artificial neural networks and machine learning algorithms have been used to predict the information content of profit before its announcement.

1. Department of Financial Management, Arak Branch, Islamic Azad University, Arak, Iran.
2. Department of Financial Management, Arak Branch, Islamic Azad University, Arak, Iran. (Corresponding Author). zanjirdar08@gmail.com
3. Department of Economics, Arak Branch, Islamic Azad University, Arak, Iran.

How to cite this paper: Alizadeh, H., Zanjirdar, M., Haji, G.A. (2023). Ability of machine learning algorithms and artificial neural networks in predicting accounting profit information content before announcing it. *Advances in Finance and Investment*, *4*(2), 1-30. [In Persian]

https://doi.org/10.30495/afi.2023.1962857.1137

Literature Review

Company controllers and informed investors benefit from confidential information before the company's profits are announced, and information may be published in a biased manner by the company's controllers. Therefore external investors are looking for alternative ways to detect the behavior of informed investors and holders of confidential information. Check before earnings announcements; from 2001 to 2016, there has been a significant increase in the market reaction to quarterly earnings announcements. Management guidance, analyst forecasts, and financial statement items are each increasingly disclosed with earnings announcements during this period. These simultaneous information items have significant explanatory power to increase the market response (Beaver *et al.*, 2020).

Chaajer *et al.* (2021), in a study titled Applications of Artificial Neural Networks, Support Vector Machines and Long-Short-Term Memory (LSTM) for stock market forecasting, stated that the stock market is chaotic, but using artificial intelligence to make predictions, It is possible and recommended before investing. This study provides an overview of artificial intelligence and machine learning as predictive analysis tools in the stock market. They studied the applications of three machine learning technologies in stock market forecasting, including artificial neural networks, support vector machines, and long-short-term memory. Yin *et al.* (2023), in research titled "Stock trend prediction based on optimal random forest, "showed that the average accuracy of the random forest after optimizing the above process increases by 17%, which is 18% more than the average accuracy of the light gradient amplification device model. The combination of ROC curve performance and the Recision-Recall curve also guarantees the stability of the model, which shows the advantages of random forest in predicting the medium and long-term trends of the stock market. Madeeh and Abdullah (2021), in research as an efficient forecasting model based on machine learning techniques for stock market forecasting, expressed the use of effective Machine Learning Techniques to build a robust model for stock market forecasting. The tests of the results showed that the two proposed models obtained a high accuracy rate, and the RF model had the best prediction accuracy, which reached 93.23, 93.12, and 93.17 percent, respectively, according to the measurements of precision, recall, and F-measure.

Research Methodology

The statistical population of the current research is the number of companies present in the Iranian capital market at the end of March 2019, which was equal to 458 companies. As a result of applying the conditions in the systematic elimination sampling, 88 companies from the statistical population were selected to perform the tests. For testing the first hypothesis, the research variables were defined and calculated first. The dependent variable of the research is the abnormal return of the company's stock $AR_{i.t}$ calculated from the difference

between the company's actual return and the market return Piotroski *et al.* (2005) For testing the second hypothesis, the abnormal return resulting from the output of the artificial neural network model and machine learning algorithms is indicated by the symbol $NNAR_{i.t}$ that the output of the artificial neural network model and algorithms causes the abnormal return of the i-th share at time t. It is machine learning. All data were normalized in the pre-processing stage by MATLAB software.

The recurrent neural network has three layers: the input, the inner (hidden), and the output layer. The number of memory cells is equal to 200, the number of complete connection layers is equal to 1, the number of neurons in the whole connection layer is equal to 40, the number of neurons in the first layer (input variables) is equal to 10, the number of IPAC is equal to 5, and the minimum batch size is considered equal to 8. The hidden layer's transfer function (activation) is the sigmoid tangent.

In the random forest method, the maximum number of trees in the forest, which determines the depth of the trees, is equal to the number of experimental data. In using the Support Vector Machine method, the input variables of the Support Vector Machine method are the same as the above two methods. First, the variables are normalized and then used as input to the model. The function used in the kernel polynomial is the 4th degree. The output of the above models includes the three states of maintenance, sale, and purchase for the next day, which are classified based on the performance evaluation criteria compared to the actual data for the training and test period.

In testing the second to fourth hypotheses, the relevance test is used to measure the informational content of profit. The information content of profit is measured in experimental accounting research by testing the relationship between profit and efficiency. In other words, the more the profit can explain the stock price or return, the more information content it has, or in other words, the more relevant it is.

Results

The results of the first hypothesis test (the ability to predict stock prices before the quarterly profit announcement of companies by support vector machine, random forest algorithm, and artificial neural networks) were obtained as follows: in the Support Vector Machine Method, the maximum accuracy achieved in the training period is equal to 97/51%, and the average accuracy obtained was 72/9%; for the test period, the ultimate accuracy was 95/45%, and the average accuracy was 43/45%. The maximum accuracy achieved in the training period was 100% in the random forest method. The average accuracy was 99/99%; for the test period, the ultimate accuracy was 91/53%, and the average accuracy was 49/01%. In RNN neural networks, the LSTM technique was used to increase prediction accuracy. The average accuracy obtained was 39/9%, and F_score was equal to 25/6%, which is less than the accuracy of the random method (33/33%). For the test period, the average accuracy obtained was 40/82%, and providing a F_score

higher than the random model was impossible. Therefore it was concluded that the model could not predict the stock price at the time of quarterly profit announcement of companies and create a suitable strategy for buying, selling, and holding. According to the higher average and significant accuracy of RF and SVM methods in the test period compared to other methods and acceptable accuracy compared to the Random Method, it is concluded that the support Vector Machine, Random Forest algorithm, can predict the future price of stocks before the announcement of quarterly profits of companies. And artificial neural networks cannot provide a F_measure higher than the random model. Therefore it is concluded that the model cannot predict the stock price at the time of quarterly profit announcement of companies and create a suitable strategy for buying, selling, and holding.

In the second hypothesis, based on the daily buy/sell and maintenance strategy predicted by the SVM method, the coefficient of determination is equal to 0/22, which means that 22% of the changes in the abnormal return variable are explained by using this regression model, the analysis of variance of this regression model It is meaningful. The percentage change of the profit variable was observed with a statistic of 0/029, a coefficient of -27/2, and a significance level of less than 5% (0/032). Therefore, the null hypothesis, which means that there is no relationship between the two variables, is rejected, and it can be said that the percentage change of the profit variable has a direct and significant relationship with abnormal returns. Since an increase (decrease) of one unit in the profit change percentage causes an increase (decrease) of 2/9% in the abnormal return, the support vector machine algorithm can predict the information content of the profit before announcing it.

In the third hypothesis, the coefficient of determination is equal to 0/021, which means that 2% of the changes in the abnormal return variable are explained using this regression model. This regression model's variance analysis is insignificant at 0/05, considering the significance level of 0/054. The significance level has been more than 5% (0/054). Therefore, the null hypothesis, the lack of relationship between the two variables, has been confirmed. The variable of profit percentage change does not have a direct and significant relationship with abnormal returns. Therefore, based on RF data, the third hypothesis was rejected, which means that the RF machine learning algorithm could not predict the information content of the benefit before it was announced.

The fourth hypothesis, taking into account the results of the first hypothesis, confirms the inability of artificial neural networks to predict the future price of stocks before the quarterly profit announcement of companies. As a result, the failure to create a holding, buying, and selling strategy, and therefore, this model can predict the information content of profits. He did not have it before the announcement.

Discussions and Conclusions

The high ability of machine learning algorithms and artificial neural networks in the field of predicting market behavior, compared to linear methods by researchers such as Chaajer *et al.* (2021), Yin *et al.* (2023), Srivastava *et al.* (2021), Sadorsky (2021), Sharma *et al.* (2022), Gholamian and Davoodi (2018) It has been proven, but so far in Iran, there has been no research on the ability of artificial neural networks and machine learning algorithms to predict the information content of companies' quarterly profit before its announcement. This research sought to find solutions that can predict the information content of profit before the announcement to increase the speed of information transmission to the market and its absorption, and this factor leads to reducing the effect of information asymmetry and confidential information and increasing market efficiency. Is done, which has reached its goal in this field and has answered part of the research vacuum related to the research problem, which is the lack of tools and techniques for predicting the information content of profit before announcing it. The results of the research showed that the Support Vector Machine and Random Forest algorithm could predict the future price of stocks before the announcement of quarterly profits of companies, which leads to help increase in the efficiency of the market by reducing the effects of hidden information, which is by the views of Sadorsky (2021), also Based on the results obtained from the support vector machine method, the acceptable accuracy of the model in predicting the purchase, sale and maintenance strategy, as well as the confirmation in the relevance test, which confirms the ability to explain the changes in returns caused by the seasonal profit changes of the companies, and therefore the vector machine algorithm The supporter could predict the information content of companies' quarterly profit before announcing it. Based on the results of the random forest method, despite the acceptable accuracy of the model in predicting the buying, selling, and holding strategy, in the test of relevance, it could not explain the changes in profit. Therefore it is concluded that while the random forest algorithm can predict the future stock price has had, but cannot predict the information content of the profit before announcing it. Based on the results of artificial neural networks in predicting the strategy of buying, selling, and holding stocks, compared to two Random Forest algorithms and Support Vector Machine, it provided lower accuracy. Therefore, artificial neural networks couldn't predict the future price of stocks and information content before announcing profit. The results obtained are different from the research results of Srivastava *et al.* (2021) from the point of view of comparing the ability of the artificial neural network method with machine learning algorithms.

Conflict of Interest

The authors of this article declared no conflict of interest regarding the authorship or publication of this article.

Keywords: Artificial Neural Network, Profit Information Content, Random Forest, Support Vector Machine.

JEL Classification: G11, G14.

پیشرفت‌های مالی و سرمایه‌گذاری
سال چهارم، تابستان ۱۴۰۲- شماره ۲
صفحات ۳۰-۱

نوع مقاله: پژوهشی

توانایی الگوریتم‌های یادگیری ماشین و شبکه‌های عصبی مصنوعی در پیش‌بینی محتوای اطلاعاتی سود حسابداری قبل از اعلان آن

حسین علیزاده[۱]، مجید زنجیردار[۲]، غلامعلی حاجی[۳]

تاریخ دریافت: ۱۴۰۱/۰۸/۰۲ تاریخ پذیرش: ۱۴۰۱/۱۲/۱۰

چکیده

هدف: هدف از پژوهش حاضر بررسی توانایی شبکه‌های عصبی مصنوعی و الگوریتم‌های یادگیری ماشین شامل ماشین بردار پشتیبان و جنگل تصادفی در پیش‌بینی محتوای اطلاعاتی سود حسابداری قبل از اعلان آن در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران در بازه زمانی سال‌های ۱۳۹۴ الی ۱۳۹۹ می‌باشد.

روش‌شناسی پژوهش: برای جمع‌آوری داده‌های روزانه موردنیاز پژوهش از نرم‌افزار ره‌آورد نوین استفاده شده است و با استفاده از روش نمونه‌گیری حذفی سیستماتیک ۸۸ شرکت انتخاب شدند. از نرم‌افزار متلب به‌منظور مدل‌سازی شبکه‌های عصبی و الگوریتم‌های یادگیری ماشین استفاده شده است و به‌منظور محاسبه بازده غیرعادی در شبکه‌های عصبی و الگوریتم‌های یادگیری ماشین از کد نوشته شده در نرم‌افزار پایتون استفاده شده است. محتوای اطلاعاتی سود از طریق آزمون ارتباط بین سود و بازده غیرعادی و بر اساس مدل پورتی و همکاران (Poretti *et al.*, 2018) سنجیده شده است. متغیرهای ورودی شبکه‌های عصبی و الگوریتم‌های یادگیری ماشین اندیکاتورها تحلیل تکنیکال می‌باشند. به‌منظور ارزیابی عملکرد طبقه‌بندی، از معیارهای صحت، دقت، فراخوانی و اندازه F استفاده شده است.

یافته‌ها: نتایج حاصل از پیش‌بینی سه مدل شبکه‌های عصبی مصنوعی، ماشین بردار پشتیبان و جنگل تصادفی نشان داد که ماشین بردار پشتیبان و جنگل تصادفی از دقت بالاتری نسبت به شبکه‌های عصبی مصنوعی، در پیش‌بینی استراتژی خرید، فروش و نگهداری برخوردار بوده‌اند و تنها ماشین بردار پشتیبان از سه مدل مذکور توان پیش‌بینی محتوای اطلاعاتی سود را داشته است.

اصالت / ارزش‌افزوده علمی: طراحی مدل پیش‌بینی جهت حرکت قیمت سهام در روز آتی معاملاتی، با سه روش شبکه‌های عصبی مصنوعی، ماشین بردار پشتیبان و جنگل تصادفی به‌عنوان اصلی‌ترین نوآوری پژوهش می‌باشد و همچنین یافته‌های پژوهش می‌تواند سرعت انتقال اطلاعات به بازار و جذب آن را افزایش دهد که این عامل منجر به کاهش اثر عدم تقارن اطلاعاتی و معاملات مبتنی بر اطلاعات نهانی و در نهایت افزایش کارایی بازار خواهد شد.

واژه‌های کلیدی: جنگل تصادفی، شبکه‌های عصبی مصنوعی، ماشین بردار پشتیبان، محتوای اطلاعاتی سود.

طبقه‌بندی موضوعی: G11, G14.

۱. گروه مدیریت مالی، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران.
۲. گروه مدیریت مالی، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران. (نویسنده مسئول). zanjirdar08@gmail.com
۳. گروه اقتصاد، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران.

استناد: علیزاده، حسین؛ زنجیردار، مجید؛ حاجی، غلامعلی. (۱۴۰۲). توانایی الگوریتم‌های یادگیری ماشین و شبکه‌های عصبی مصنوعی در پیش‌بینی محتوای اطلاعاتی سود حسابداری قبل از اعلان آن. *پیشرفت‌های مالی و سرمایه‌گذاری*، ۴(۲). ۳۰-۱.

https://doi.org/10.30495/afi.2023.1962857.1137

۱- مقدمه

پژوهش‌های سال‌های اخیر نشان‌می‌دهد که پاسخ سرمایه‌گذاران به اعلامیه‌های سود فصلی شرکت‌ها افزایش چشمگیری داشته است و البته بیشتر تغییر قیمت قبل از اعلام رسمی سود اتفاق می‌افتد که ناشی از معاملات دارندگان اطلاعات نهانی می‌باشد. بیور و همکاران (Beaver *et al.*, 2020) رویکردها و نظریه‌ها در مورد معاملات دارندگان اطلاعات نهانی به دو دیدگاه متضاد معطوف شده است و اصلی‌ترین موضوع عدم تفاهم محققین، منع معاملات دارندگان اطلاعات نهانی و یا عدم نظارت بر آن می‌باشد. دیدگاه اول معاملات متکی بر اطلاعات نهانی را منجر به افزایش کارایی بازار و دیدگاه دوم معتقد به کاهش کارایی بازار می‌داند. محققین در این ارتباط تاکنون بر یک چارچوب نظری یکسان به توافق نرسیده‌اند و این مسئله منجر به ارائه پژوهش‌های متنوع و مدل‌های علمی متفاوتی گردیده است. از جمله صاحب‌نظرانی که موافق نقش معاملات متکی بر اطلاعات نهانی بر افزایش کارایی بازار هستند می‌توان به موارد ذیل اشاره نمود: اشتون و ترین (Ashton and Trinh, 2018) در پژوهشی با عنوان"معامله با اطلاعات نهانی؛ دو دیدگاه" معتقد است اگر یکی از وظایف اولیه بازارها کشف قیمت دقیق باشد، استثنا کردن معاملات با اطلاعات نهانی معنی ندارد. با حذف معامله‌گران دارای اطلاعات نهانی، آگاهی بازارها کمتر شده و قیمت‌های بازار انعکاس‌دهنده دانش تجمعی همه فعالان بازار نمی‌باشد. همچنین انگلن و ون لیدکرکه (Engelen and Van Liedekerke, 2007) در پژوهشی با عنوان "اخلاق معاملات با اطلاعات نهانی" بیان می‌دارد که با وارد شدن اطلاعات نهانی در قیمت، قیمت‌های اوراق‌بهادار بهتر و سریع‌تر انعکاس‌دهنده ارزش بنیادی واقعی شرکت خواهند بود و به معامله‌گران با اطلاعات نهانی این اجازه را خواهد داد که کارایی تخصیصی بازار اوراق‌بهادار را افزایش دهند.

از صاحب‌نظرانی که مخالف نقش معاملات نهانی در افزایش کارایی بازار می‌باشند می‌توان به سادورسکی (Sadorsky, 2021) اشاره کرد که در پژوهشی با عنوان "عایدی سرمایه" باور دارد معامله با اطلاعات نهانی یک روش ناکارای دستیابی به کارایی بازار است، زیرا تمام سود معامله‌گر نهانی ناشی از وقفه میان زمانی است که وی شروع به فروش می‌کند و زمانی که بازار متوجه آنچه اتفاق افتاده می‌شود. آنچه به نفع بازار سرمایه است، شفافیت است، در حالی که معامله‌گران نهانی از عدم‌شفافیت و ابهام منتفع خواهند شد. پژوهش حاضر مبتنی بر نظر آن دسته از صاحب‌نظرانی است که اعتقاد دارند معاملات مبتنی بر اطلاعات نهانی به افزایش کارایی بازار منجر نمی‌گردد. یادگیری ماشین و هوش مصنوعی برای پیش‌بینی بازار سهام در سال‌های اخیر به‌شدت توسعه یافته است و در عین اینکه رفتار قیمت در بازار سهام آشفته به نظر می‌رسد، لیکن استفاده از هوش مصنوعی برای انجام پیش‌بینی‌های

حساب شده قبل از سرمایه‌گذاری به‌شدت توصیه می‌شود. چاجر و همکاران (Chaajer *et al.*, 2021) پیش‌بینی بازار سهام یک مشکل کلاسیک و در عین حال چالش‌برانگیز هم برای اقتصاددانان و هم دانشمندان کامپیوتر بوده است. اخیراً، مدل‌های یادگیری عمیق به‌عنوان مرزهای جدیدی برای این موضوع معرفی‌شده‌اند و توسعه سریع یافته‌اند. جیانگ (Jiang, 2021) در این پژوهش با در نظر داشتن تضادهای اشاره‌شده، از شبکه‌های عصبی مصنوعی و الگوریتم‌های یادگیری ماشین به‌عنوان راهکارهایی برای پیش‌بینی محتوای اطلاعاتی سود قبل از اعلان آن استفاده شده است. تحقیقات زیادی در زمینه محتوای اطلاعاتی سود حسابداری در ایران صورت گرفته است به‌عنوان مثال بررسی نقش عدم اطمینان سیاست‌های اقتصادی در ارتباط بین افشای داوطلبانه اطلاعات و عدم تقارن اطلاعاتی در شرکت‌های پذیرفته شده در بورس اوراق‌بهادار تهران نشان داد که رابطه بین افشای داوطلبانه اطلاعات و عدم تقارن اطلاعاتی از نظر آماری منفی و معنادار است. یافته‌ها همچنین نشان داد که عدم اطمینان سیاست‌ها موجب تضعیف رابطه بین افشای داوطلبانه اطلاعات و عدم تقارن اطلاعاتی می‌شود. این موضوع به معنی آن است که در یک شرایط عدم اطمینان سیاست‌های اقتصادی با افزایش سیاست‌های مالی و پولی ناکارآمد و همچنین کاهش بی‌اعتمادی مردم نسبت به تصمیم‌های اقتصادی، افشای داوطلبانه اطلاعات کاهش یافته و در نتیجه عدم تقارن اطلاعاتی افزایش می‌یابد (Ghasemi *et al.*, 2022). لیکن در عین وجود پژوهش‌های مشابه، کمتر به موضوع پیش‌بینی محتوای اطلاعاتی سود، روش‌ها، ابزارها و تکنیک‌های پیش‌بینی محتوای اطلاعاتی سود حسابداری قبل از اعلان آن پرداخته شده است و تاکنون پژوهشی در مورد توانایی شبکه‌های عصبی مصنوعی و الگوریتم‌های یادگیری ماشین در پیش‌بینی محتوای اطلاعاتی سود حسابداری در بازار سرمایه ایران صورت نگرفته‌است و لذا در این پژوهش قصد پرداختن به این خلأ تحقیقاتی وجود دارد و البته اصلی‌ترین نوآوری پژوهش طراحی مدل پیش‌بینی جهت حرکت قیمت سهام در روز آتی معاملاتی می‌باشد که با سه روش شبکه‌های عصبی مصنوعی، ماشین بردار پشتیبان و جنگل تصادفی انجام شده است. نتایج این پژوهش ابزاری برای تصمیم‌گیری بهتر و به‌موقع‌تر در اختیار سرمایه‌گذاران قرار می‌دهد و همچنین به افزایش کارایی بازار کمک شایانی خواهد نمود و در عین حال امکان درک بخشی از اثرات اطلاعات نهانی را مهیا می‌نماید. سرمایه‌گذاران خرد که قادر به دستیابی به‌موقع به اطلاعات بنیادی و محرمانه شرکت‌ها نیستند، همچنین تحلیل‌گران بازار سرمایه و سرمایه‌گذاران کوتاه‌مدتی و سفته‌بازان، مشاوران صندوق‌های سرمایه‌گذاری و بازارگردانان سهام، که با در نظر گرفتن حجم وسیع سرمایه‌گذاران، اعتباردهندگان و فعالان اقتصادی و توجه روز افزون عموم مردم به بازار سرمایه، که بدون شک آنها به‌دنبال کسب بازده بیشتر نسبت به سایر بازارها می‌باشند، ذی‌نفعان یافته‌های این پژوهش خواهند بود.

۲- مبانی نظری و پیشینه پژوهش

محتوای اطلاعاتی بدین معناست که وقوع یک رویداد مانند اعلام سود بتواند بر قیمت تأثیر بگذارد اگر محتوای اطلاعات شناخته شده نباشد و به‌عبارتی جدید باشد، نوسان قیمت را باعث می‌گردد لیکن به‌واسطه تئوری عدم تقارن اطلاعاتی که فرض اساسی آن عدم کارایی کامل بازار است، تمامی سرمایه‌گذاران به‌صورت همزمان از اطلاعات مطلع نمی‌گردند و لذا اعلام سود منجر به واکنش قیمت‌ها قبل و بعد از اعلام رسمی آن می‌گردد. کنترل‌کنندگان شرکت و سرمایه‌گذاران آگاه قبل از اعلام سود شرکت‌ها از اطلاعات نهانی بهره‌مند هستند و همچنین ممکن است اطلاعات به‌صورت مغرضانه از سوی کنترل‌کنندگان شرکت منتشر گردد و لذا سرمایه‌گذاران بیرونی به‌دنبال یافتن راه‌های جایگزینی هستند تا از طریق آن، رفتار سرمایه‌گذاران آگاه و دارندگان اطلاعات نهانی را قبل از اعلام سود بررسی کنند، از سال ۲۰۰۱ تا ۲۰۱۶ افزایش قابل توجهی در واکنش بازار به اعلامیه‌های سود فصلی وجود داشته است. راهنمایی‌های مدیریت، پیش‌بینی‌های تحلیلگر و اقلام صورت‌های مالی هر کدام به‌طور فزاینده‌ای با اطلاعیه‌های سود در این مدت افشا می‌شوند. هر یک از این اقلام اطلاعاتی همزمان دارای قدرت توضیحی قابل توجهی برای افزایش پاسخ بازار هستند. بیور و همکاران (Beaver *et al.*, 2020) داده‌هایی که بر مشارکت کنندگان بازار اثری معنادار بگذارد حاوی اطلاعات است. به باور آنها اطلاعات باید قادر به کاهش عدم اطمینان باشد و به تصمیم‌گیرنده پیامی دهد که ارزش آن از هزینه کسب اطلاعات بیشتر باشد و از سوی دیگر بتواند به‌صورت بالقوه بر تصمیماتی مؤثر واقع شود که اشخاص می‌گیرند. پیام تنها زمانی می‌تواند حاوی اطلاعات نامیده شود که بتوان آن را علت برخی تصمیم‌گیری‌های دریافت‌کننده پیام دانست. به‌صورت خاص محتوای اطلاعات سود خالص گزارش شده می‌تواند با میزان تغییر قیمت سهام و یا با میزان بازده غیرعادی در حوالی زمانی تعریف شود که بازار از سود خالص دوره جاری آگاه می‌شود. ازاین‌رو استدلال بر این است که سرمایه‌گذاران در زمان آگاه شدن از سود حاوی اطلاعات نظریات خود در مورد بازده آینده سهام را تعدیل و بر این اساس به خریدوفروش اقدام می‌کنند که نتیجه آن تغییر قیمت سهام خواهد بود (Patel *et al.*, 2015). در یک معامله، چنانچه یکی از طرفین از شرایط واقعی کالاهای موردنظر آگاهی نداشته‌باشد، طرف دوم می‌تواند سود بیشتری حاصل نماید. در چنین شرایطی، گونه‌ای از نبود اطمینان و عدم قطعیت بر فضای معامله حاکم خواهد بود که در نهایت به تضعیف جریان بازار یا به‌عبارت‌دیگر به شکست بازار می‌انجامد (Levin, 2001). فرضیه بازار کارا اساساً این ادعا را مطرح کرد که اگر بازار سرمایه از کارایی نسبی برخوردار باشد. در فرایند جذب اطلاعات به‌سرعت و قوت عمل نموده و قیمت‌های سهام در بازار همواره انعکاس مطلوبی از وقایع مرتبط به شرکت خواهد بود. بر اساس این فرضیه هیچ‌کس نمی‌تواند در بلندمدت به‌طور

سیستماتیک بیشتر از میزان ریسکی که قبول کرده است بازده کسب کند. یکی از مشخصه‌های بازار کارا این است که قیمت سهام در هر لحظه از زمان منعکس‌کننده ارزش ذاتی سهام باتوجه‌به اطلاعات موجود و مربوط در بازار می‌باشد و بازده غیرعادی در این بازار وجود ندارد؛ بنابراین در صورتی که بتوان این بازده غیرعادی را در بازار کاهش داد میزان کارایی بازار بیشتر می‌شود (Poretti *et al.*, 2018). این پژوهش درصدد است، راهکارهایی را بررسی نماید که قدرت پیش‌بینی محتوای اطلاعاتی سود را قبل از اعلان داشته‌باشد تا از این طریق سرعت انتقال اطلاعات به بازار و جذب آن را افزایش دهد که این عامل منجر به افزایش کارایی بازار خواهد شد. توسعه روش‌های پیش‌بینی مبتنی بر یادگیری ماشینی[1] در سال‌های اخیر موجبات برخورد با داده‌های بزرگ[2] را در مسائل پیش‌بینی فراهم آورده‌اند. جمع‌آوری و نگهداری داده‌های بزرگ در حوزه‌های مختلف این امکان را به صاحبان کسب‌وکارها داده است که بتوانند به منظورهای خاص از پایگاه‌های داده استفاده نموده و الگوهای معنی‌دار را استخراج نمایند. مدل‌های پیش‌بینی باتوجه‌به نقش کلیدی‌شان در سودآوری و تضمین بقای شرکت در حوزه‌ها و صنایع گوناگونی مورد استفاده واقع‌شده‌اند یادگیری عمیق[3] نقش مهمی در برخورد با داده‌های بزرگ دارد، زیرا می‌تواند اطلاعات ارزشمندی را از سیستم‌های پیچیده و داده‌های بزرگ به‌دست‌آورد. روش یادگیری در شبکه‌های عصبی بازگشتی[4] تحت‌عنوان الگوریتم بازگشت به عقب از طریق زمان[5] شناخته می‌شود، در عین حال بخشی از اطلاعات در حین یادگیری ممکن است که از بین برود. لذا از روش حافظه کوتاه–بلندمدت[6] برای حل این نقصان تکنیک گرادیان محور استفاده می‌گردد. این روش یکی از روش‌های پرکاربرد و توانمند در RNN است که شامل یک سلول حافظه برای به یادسپاری در طول مقاطع زمانی می‌باشد. ماشین بردار پشتیبان[7] یکی از روش‌های یادگیری با ناظر است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. این روش از جمله روش‌های نسبتاً جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر برای طبقه‌بندی، از جمله نسبت به شبکه عصبی نشان داده است (Liu *et al.*, 2017). ماشین بردار پشتیبان یک مدل ویژه از روش‌های دسته‌بندی‌کننده است که با استفاده از توابع کرنل مختلف، حاشیه‌ی ایمن و بردارهای پشتیبان به دسته‌بندی میپردازد، ریسک را کاهش می‌دهد و از قابلیت تعمیم بالایی برخوردار است (Chen and Ho, 2005). ماشین‌بردار پشتیبان به‌خوبی با داده‌های غیرخطی رگرسیون و سری‌های زمانی کار می‌کند. روش جنگل تصادفی

1. Machine Learning
2. Big Data
3. Deep Learning
4. Recurrent Neural Network (RNN)
5. Backpropagation Through Time (BPTT)
6. Long Short-Term Memory (LSTM)
7. Support Vector Machines (SVM)

جزئی از روش‌های طبقه‌بندی تحت نظارت گروهی در حوزه داده‌کاوی است. جنگل تصادفی، پتانسیل بسیار زیادی برای تبدیل به یک روش محبوب برای طبقه‌بندی‌های آینده دارد زیرا کارایی آن قابل مقایسه با دیگر روش‌های گروهی می‌باشد. به‌عنوان یک الگوریتم گروهی (رأی‌گیری) چندین درخت تصمیم مختلف به‌عنوان طبقه‌بندی‌های پایه تولید می‌کند و رأی اکثریت را برای ترکیب با نتایج درختان اصلی اعمال می‌کند. جنگل تصادفی هم توانایی رگرسیون، و هم توانایی کلاس‌بندی را با هم دارد. یین و همکاران (Yin *et al.*, 2023) بر اساس چارچوب نظری و سوابق تجربی فوق فرضیه‌های پژوهش به‌صورت زیر تدوین می‌گردد:

فرضیه اول: ماشین بردار پشتیبان، الگوریتم جنگل تصادفی و شبکه‌های عصبی مصنوعی توانایی پیش‌بینی قیمت آتی سهام قبل از اعلان سود فصلی شرکت‌ها را دارند.

فرضیه دوم: ماشین بردار پشتیبان توانایی پیش‌بینی محتوای اطلاعاتی سود را قبل از اعلان آن دارند.

فرضیه سوم: الگوریتم جنگل تصادفی توانایی پیش‌بینی محتوای اطلاعاتی سود را قبل از اعلان آن دارند.

فرضیه چهارم: شبکه‌های عصبی مصنوعی توانایی پیش‌بینی محتوای اطلاعاتی سود را قبل از اعلان آن دارند.

چاجر و همکاران (Chaajer *et al.*, 2021)، در مطالعه‌ای با عنوان کاربردهای شبکه‌های عصبی مصنوعی، ماشین‌های بردار پشتیبان و حافظه بلندمدت – کوتاه‌مدت (LSTM) برای پیش‌بینی بازار سهام، بیان کردند که بازار سهام آشفته‌است، اما استفاده از هوش مصنوعی برای انجام پیش‌بینی‌ها، قبل از سرمایه‌گذاری امکان‌پذیر و توصیه می‌شود. این مطالعه مروری بر هوش مصنوعی و یادگیری ماشین به‌عنوان ابزارهای تحلیل پیش‌بینی در بازار سهام ارائه می‌کند. آنها کاربردهای سه فناوری یادگیری ماشین را در پیش‌بینی بازار سهام، از جمله شبکه‌های عصبی مصنوعی، ماشین‌های بردار پشتیبان و حافظه بلندمدت – کوتاه‌مدت را مطالعه کردند. یین و همکاران (Yin *et al.*, 2023) در تحقیقی با عنوان پیش‌بینی روند سهام بر اساس جنگل تصادفی بهینه، نشان دادند که میانگین دقت جنگل تصادفی پس از بهینه‌سازی فرایند فوق به میزان ۱۷ درصد افزایش می‌یابد که ۱۸ درصد بیشتر از دقت متوسط مدل دستگاه تقویت گرادیان سبک بوده است. ترکیب عملکرد منحنی[1] و منحنی دقت – بازخوانی[2]، پایداری مدل را نیز تضمین می‌نماید که مزیت‌های جنگل تصادفی را در پیش‌بینی روند میان‌مدت و

1. Receiver operating characteristic (ROC)
2 .Precision-Recall

بلندمدت بازار سهام نشان‌می‌دهد. سریواستاوا و همکاران (Srivastava *et al.*, 2021) در تحقیقی به‌عنوان رویکرد شبکه عصبی عمیق و سری زمانی برای سیستم‌های مالی: پیش‌بینی حرکت بازار سهام هند، با استفاده از شاخص‌های فنی مانند میانگین متحرک[1]، میانگین متحرک همگرایی – واگرایی[2]، نوسانگر استوکاستیک[3] و میانگین متحرک سه‌روزه مربوطه، مقدار نشانگر قدرت نسبی[4] و نشانگر ویلیامز[5]، قدرت پیش‌بینی شبکه‌های عصبی عمیق نسبت به سایر تکنیک‌های یادگیری ماشین را نشان‌می‌دهد و دامنه آینده یادگیری عمیق را در پیش‌بینی سری‌های زمانی چند پارامتری معرفی می‌نماید. امیوما و ادکی (Emioma and Edeki, 2021) در تحقیقی با عنوان پیش‌بینی قیمت سهام با استفاده از یادگیری ماشین بر اساس رگرسیون خطی حداقل مربعات، با هدف استفاده از یک الگوریتم یادگیری ماشینی برای تخمین قیمت پایانی سهام به تدوین مدل با یک مجموعه‌داده برای کمک به پیش‌بینی قیمت سهام پرداخت. هدف این مدل این بود که از آن به‌عنوان راهنمای معاملات روزانه استفاده شود. الگوریتم مورد استفاده مدل رگرسیون خطی حداقل مربعات نامیده می‌شود. یک متغیر وابسته مدنظر است که می‌تواند در این مورد، قیمت بسته‌شدن سهام باشد و یک متغیر مستقل است که روز ثبت هر قیمت سهام است. سادورسکی (Sadorsky, 2021) در تحقیقی با عنوان رویکرد جنگل‌های تصادفی برای پیش‌بینی قیمت سهام انرژی پاک، از روش یادگیری ماشینی جنگل‌های تصادفی برای پیش‌بینی جهت قیمت سهام صندوق‌های قابل معامله در بورس انرژی پاک استفاده‌کرد. برخی از شاخص‌های فنی شناخته شده به‌عنوان ویژگی مورد استفاده قرار گرفت. پیش‌بینی‌های درخت تصمیم و جنگل‌های تصادفی از جهت قیمت سهام دقیق‌تر از مدل‌های لاجیت بود. برای افق پیش‌بینی ۲۰ روزه، روش‌های بسته‌بندی درخت و جنگل‌های تصادفی نرخ‌های دقت بین ۸۵ تا ۹۰ درصد را تولید می‌کنند درحالی‌که مدل‌های لاجیت نرخ‌های دقت بین ۵۵ تا ۶۰ درصد را تولید می‌کنند. مادی و عبدالله (Madeeh and Abdullah, 2021) در تحقیقی به‌عنوان یک مدل پیش‌بینی کارآمد بر اساس تکنیک‌های یادگیری ماشین برای پیش‌بینی بازار سهام، استفاده از تکنیک‌های یادگیری ماشینی مؤثر را برای ساخت یک مدل قوی برای پیش‌بینی بازار سهام بیان کرد. آزمایش‌های نتایج نشان داد که دو مدل پیشنهادی نسبت دقت بالایی به‌دست آوردند و مدل RF بهترین دقت پیش‌بینی را داشت که باتوجه‌به اندازه‌گیری‌های ارزیابی دقت، یادآوری و F-measure به‌ترتیب به ۹۳/۲۳، ۹۳/۱۲ و ۹۳/۱۷ درصد رسید. شارما و همکاران (Sharma *et al.*, 2022) در تحقیقی با عنوان ادغام الگوریتم ژنتیک با

1. Moving Average (MA)
2. Moving Average Convergence Divergence (MACD)
3. stochastic oscillator (%K)
4. Relative Strength Index (RSI)
5. Larry Williams, Williams (%R)

شبکه عصبی مصنوعی برای پیش‌بینی بازار سهام، یک روش پیش‌بینی هوشمند مبتنی بر ترکیبی از یک شبکه عصبی مصنوعی و یک الگوریتم ژنتیک پیشنهاد و از دو شاخص بورس ایالات متحده، میانگین صنعتی داو جونز[1] و نزدک-۱۰۰[2] برای پیش‌بینی استفاده کردند. یافته‌های تجربی به‌دست‌آمده با استفاده از میانگین صنعتی داو جونز و و نزدک - ۱۰۰ نشان‌می‌دهد که دقت مدل ترکیبی GA[3] و ANN[4] برای میانگین صنعتی داو جونز و و نزدک - ۱۰۰ هم در کوتاه‌مدت و هم در بلندمدت بیشتر از تکنیک ANN منفرد می‌باشد. بیور و همکاران (Beaver *et al.*, 2020) طی پژوهشی تحت‌عنوان "افزایش پاسخ بازار به اعلامیه‌های سود در قرن بیست و یکم: پژوهش تجربی" به بررسی نقش اطلاعات هم‌زمان در افزایش چشمگیر پاسخ سرمایه‌گذاران به اعلامیه‌های سود از سال ۲۰۰۱ تا ۲۰۱۶ پرداختند که یافته‌ها نشان‌می‌دهد که طی دو دهه گذشته تاریخ ورود اطلاعات به درآمد به‌طور قابل‌توجهی افزایش‌یافته است و مؤلفه‌های اصلی این افزایش افشای اطلاعات توسط راهنمایی‌های مدیریت و اقلام صورت‌های مالی و پیش‌بینی تحلیلگران است. شائو و همکاران (Shao *et al.*, 2020) در پژوهشی با عنوان قدرت اطلاعات حسابداری در توضیح بازده، نشان دادند که اخبار مربوط به درآمد، توضیح‌دهنده حرکت بیشتر قیمت در طول زمان است. در سال‌های پس از ۲۰۰۳، بازده اعلامیه سود تقریباً ۲۰ درصد از بازده سالانه را توضیح می‌دهد که دو برابر بیشتر از قبل می‌باشد. این موضوع نشان‌می‌دهد که در سال‌های اخیر اطلاعات اساسی در توضیح بازده سهام مهم‌تر شده است. این الگو برای دیگر اطلاعات بنیادی شرکت نیز وجود دارد. بیور و همکاران (Beaver *et al.*, 2018) طی پژوهشی تحت‌عنوان "محتوای اطلاعاتی اعلامیه‌های سود: بینش جدیدی از رفتار بین دوره‌ای و مقطعی" به مطالعه محتوای اطلاعاتی اعلامیه‌های سود سه‌ماهه پرداختند که در آن میزان بازنگری قیمت سهام ناشی از اطلاعیه‌های سود نسبت به تغییر قیمت‌ها ناشی از دیگر عوامل و در دیگر زمان‌ها مورد توجه قرار دارد. چهار یافته به شرح ادامه حاصل شد، اول، افزایش چشمگیری در محتوای اطلاعات در تاریخ درآمد از سال ۲۰۰۱ به‌بعد وجود دارد. دوم، واکنش بازار به شرکت‌های زیان‌ده نسبت به شرکت‌های سودآور کمتر است. سوم، واکنش چشمگیری نسبت به شرکت‌های بزرگ‌تر وجود دارد. چهارم، واکنش در تاریخ درآمد به‌طور قابل‌توجهی در پوشش تحلیلگر افزایش می‌یابد و هنگامی که پوشش تحلیلگر کنترل می‌شود، ارتباط با اندازه کمتر قابل‌توجه است. پورتی و همکاران (Poretti *et al.*, 2018) در پژوهش خود به بررسی استقلال کمیته حسابرسی و محتوای اطلاعاتی سود در اروپای

1. DOW30
2. NASDAQ100
3. Genetic Algorithm (GA)
4. Artificial Neural Network (ANN)

غربی پرداختند. نتایج به‌طورکلی پس از کنترل بسیاری از موضوعات روش‌شناختی بررسی شد و نشان داد که اگر کمیته‌های حسابرسی مستقل جایگزین کمیته حسابرسی ضعیف شوند، افزایش اعتبار اعلان‌های سود مشاهده می‌شود. کارا و همکاران (Kara *et al.*, 2011) در پژوهشی به پیش‌بینی جهت حرکت شاخص قیمت سهام با استفاده از شبکه‌های عصبی مصنوعی و ماشین بردار پشتیبان پرداختند و به‌عنوان نمونه شاخص بورس اوراق‌بهادار استانبول را پیش‌بینی کردند. نتایج تجربی نشان داد که میانگین عملکرد مدل ANN (۷۵/۷۴ درصد) به‌طور قابل‌توجهی بهتر از مدل SVM (۷۱/۵۲ درصد) بود. پورقدیمی و همکاران (Pourgadimi *et al.*, 2022) در پژوهشی با عنوان ارائه الگوی توسعه یافته مدل بنیش با تأکید بر ویژگی‌های کیفیت حسابرسی با استفاده از شبکه عصبی، ماشین بردار و جنگل تصادفی به ارائه مدلی به‌منظور کشف تقلب با استفاده از مدل ماشین بردار و جنگل تصادفی و شبکه عصبی به‌منظور برازش مدل بسط یافته بهره گرفته شده است. نتایج نشان‌دهنده دقت ضرایب به‌دست‌آمده از مدل جنگل تصادفی ۹۸/۴ درصد و بیشتر از دو مدل شبکه عصبی و ماشین بردار ۹۳ درصد است. همچنین مدل بسط یافته نسبت به مدل پایه دقت بیشتری دارد. ویژگی‌های حسابرسی در پیش‌بینی تقلب در صورت‌های مالی تأثیرگذار است و بهتر است توسط مشارکت‌کنندگان بازار سرمایه در نظر گرفته شود. غلامیان و داوودی (Gholamian and Davoodi, 2018) در پژوهشی با عنوان "پیش‌بینی روند قیمت در بازار سهام با استفاده از الگوریتم جنگل تصادفی" با استفاده از رویکرد جنگل تصادفی به‌همراه اندیکاتورها تکنیکال: شاخص قدرت نسبی قیمت، استوکاستیک، ویلیامز، بازدهی روزانه و شاخص میانگین متحرک همگرایی – واگرایی به پیش‌بینی روند قیمت در بازار سهام و مقایسه آن با روش‌های موجود پرداختند که نتیجه پژوهش نشان‌می‌دهد که دقت روش پیشنهادی در برآورد روند بازار ۶۴ درصد می‌باشد و نسبت به دو روش مقایسه شده رگرسیون لجستیک و روش کاملاً تصادفی از دقت بالاتری برخوردار است.

۳- روش‌شناسی پژوهش

پژوهش حاضر از نظر نتیجه اجرا از نوع پژوهش‌های کاربردی است؛ زیرا می‌تواند در تصمیم‌گیری سرمایه‌گذاران درباره استفاده از شبکه‌های عصبی مصنوعی و الگوریتم‌های یادگیری ماشین مؤثر واقع شود برای جمع‌آوری داده‌های موردنیاز پژوهش از سایت اطلاع‌رسانی ناشران بازار سرمایه و نرم‌افزار ره‌آورد نوین استفاده شده است. جامعه آماری پژوهش حاضر تعداد شرکت‌های حاضر در بازار سرمایه ایران در پایان اسفند ۱۳۹۹ می‌باشد که برابر ۴۵۸ شرکت بوده است. در نتیجه اعمال شرایط در نمونه‌گیری حذفی سیستماتیک ۸۸ شرکت از جامعه آماری جهت انجام آزمون‌ها انتخاب شدند.

به‌منظور آزمون فرضیه اول در ابتدا متغیرهای پژوهش تعریف و محاسبه گردید. متغیر وابسته پژوهش بازده غیرعادی سهام شرکت $AR_{i.t}$ از تفاوت بازده واقعی شرکت و بازده بازار محاسبه می‌شود (Piotroski *et al.*, 2005). متغیر فوق لازم است بر اساس برآوردهای حاصل از سیگنال‌های خرید و فروش شبکه‌های عصبی مصنوعی و الگوریتم‌های یادگیری ماشین برآورد گردد که در رابطه (۱) ارائه شده است:

$$NNAR_{i.t} = R_{i.t} - R_{m.t}$$ رابطه (۱)

$NNAR_{i.t}$: بازده غیرعادی سهم i ام در زمان t ناشی از شبکه‌های عصبی و الگوریتم‌های یادگیری ماشین.

$NNR_{i.t}$: بازده عادی سهم i ام در زمان t ناشی از شبکه‌های عصبی و الگوریتم‌های یادگیری ماشین

R_{mt}: بازده بازار در زمان t.

بازده غیرعادی ناشی از خروجی مدل شبکه‌های عصبی مصنوعی و الگوریتم‌های یادگیری ماشین با درنظرداشتن رابطه (۱) با نماد $NNAR_{i.t}$ نشان داده‌می‌شود که در آن:

$NNAR_{i.t}$: بازده غیرعادی سهم i ام در زمان t ناشی از خروجی مدل شبکه‌های عصبی مصنوعی و الگوریتم‌های یادگیری ماشین.

$NNR_{i.t}$: بازده عادی سهم i ام در زمان t ناشی از خروجی مدل شبکه‌های عصبی مصنوعی و الگوریتم‌های یادگیری ماشین.

R_{mt}: بازده بازار در زمان t.

تمامی داده‌ها در مرحله پیش‌پردازش توسط نرم‌افزار متلب و بر اساس رابطه (۲) نرمال‌سازی می‌گردد:

$$X_{norm} = \frac{X_i - X_{min}}{X_{Max} - X_{Min}}$$ رابطه (۲)

در رابطه (۲)، X_{Min} حداقل مقدارها و X_{Max} نیز حداکثر مقادیر را نشان‌می‌دهد (Esmaeeli *et al.*, 2017).

به‌منظور محاسبه بازده عادی نیاز به سیگنال‌های خریدوفروش می‌باشد که پیش‌بینی با کمک سه روش جنگل تصادفی، ماشین بردار پشتیبان و شبکه‌های عصبی مصنوعی استفاده شده است. باتوجه‌به ویژگی‌های منحصربه‌فرد شبکه‌های عصبی بازگشتی RNN در پیش‌بینی سری‌های زمانی، در این پژوهش به‌منظور پیش‌بینی قیمت سهم در محدوده زمانی اعلان سود بر اساس شبکه عصبی طراحی

شده برچسب‌های خرید، فروش و نگهداری به‌عنوان خروجی طبقه‌بندی شبکه ایجاد می‌گردد و بازده عادی مطابق رابطه بازده جامع و بر اساس فاصله بین خرید و فروش‌های اعلامی شبکه عصبی محاسبه می‌گردد. شبکه عصبی بازگشتی به کار گرفته شده دارای سه لایه شامل لایه ورودی، لایه داخلی (پنهان) و لایه خروجی می‌باشد. تعداد سلول حافظه برابر ۲۰۰، تعداد لایه اتصال کامل برابر ۱، تعداد نورون درون لایه اتصال کامل برابر با ۴۰ تعداد نرون‌های لایه اول (متغیرهای ورودی) برابر ۱۰ و تعداد ایپاک برابر ۵ و سایز حداقل بچ نیز برابر ۸ در نظر گرفته شده است. تابع انتقال (فعال‌سازی) لایه پنهان تانژانت سیگموئیدی می‌باشد. در هر یک از ۸۸ سهم انتخابی معادل ۹۰ درصد از داده‌های روزانه به‌عنوان داده‌های مورد نیاز دوره آموزش مدل شبکه عصبی بازگشتی استفاده شده‌اند و ۱۰ درصد از داده‌ها به‌عنوان خروجی مدل در دوره آزمایش مدل پیش‌بینی شده است که برای آزمون مدل و اعتبارسنجی از تمامی داده‌های دوره آزمایشی استفاده شده است. برای طراحی مدل از نرم‌افزار متلب[1] ۲۰۱۸ استفاده شده و ساختار داده‌های ورودی و خروجی در نرم‌افزار متلب کدنویسی گردیده و مدل در آن پیاده‌سازی شده است که تعداد نرون‌های ورودی‌های مدل ۱۰ نرون که شامل نه اندیکاتورهای مهم تحلیل تکنیکال به‌علاوه متغیر قیمت پایانی به شرح مندرج در جدول (۱) می‌باشند:

ازآنجاکه در فریم قیمت سهم روزانه می‌باشد و هدف آن پیش‌بینی جهت سهم در روز آتی می‌باشد، تمام اندیکاتورها نیز در فریم روزانه به‌صورت متغیر ورودی استفاده می‌گردند که در جدول متغیرها درج‌شده‌اند. در استفاده از روش جنگل تصادفی در هر یک از ۸۸ سهم انتخابی معادل ۹۰ درصد از داده‌های روزانه به‌عنوان داده‌های موردنیاز دوره آموزش جنگل تصادفی استفاده شده‌اند و ۱۰ درصد از داده‌ها به‌عنوان خروجی مدل در دوره آزمایش مدل پیش‌بینی شده است که برای آزمون مدل و اعتبارسنجی از تمامی داده‌های دوره آزمایشی استفاده شده است. بیشینه تعداد درخت‌های جنگل که تعیین‌کننده عمق درخت‌ها می‌باشد، برابر تعداد داده‌های آزمایشی در نظر گرفته شده است. در استفاده از روش ماشین بردار پشتیبان متغیرهای ورودی روش ماشین بردار پشتیبان نیز همانند دو روش فوق بوده است و ابتدا متغیرها نرمال گردیده و بعد به‌عنوان ورودی مدل مورداستفاده قرار گرفته‌اند. تابع به‌کاررفته در مدل کرنل پلی نومیال درجه ۴ می‌باشد.

متغیرهای پژوهش در جدول (۱) ارائه شده است:

1. Matlab R2018b

جدول (۱) متغیرهای پژوهش

Table (1) Research variables

ردیف	نوع	نام متغیر	نماد متغیر	تعریف عملیاتی
۱	وابسته	بازده غیرعادی سهم	$NNAR_{i.t}$	$NNR_{i.t} - R_{mt}$
۲	وابسته	بازده عادی سهم	$NNR_{i.t}$	$\frac{NNP_t(1+\alpha+\beta) - (NNP_{t-1} + C\alpha) + D_t}{NNP_{t-1} + C\alpha}$
۳	مستقل	بازده بازار	Rm	$\frac{TEDPIX2 - TEDPIX1}{TEDPIX1}$
۴	مستقل	تغییرات سود	$\Delta NI_{i.t}$	$NI_{i.t} - NI_{i.t-1}$
۵	مستقل	همگرایی / واگرایی میانگین متحرک قیمت پایانی سهم i در دوره $t-1$	$\text{MACD}_{i.t-1}$	$MACD = EMA(5) - EMA(20)$
۶	مستقل	شاخص قدرت نسبی سهم i در دوره $t-1$	$RSI_{i.t-1}$	$RSI = 100 - \frac{100}{1-RS}$، $RS = \frac{SMA(U)}{SMA(D)}$
۷	مستقل	نرخ تغییر قیمت سهم i در دوره $t-1$	$ROC_{i.t-1}$	$ROC = \frac{CLOSE_{today} - CLOSE_{\text{قبل روز } N}}{CLOSE_{\text{قبل روز } N}}$
۸	مستقل	شاخص ویلیامز سهم i در دوره $t-1$	$Williams\ R\%_{i.t-1}$	$Williams\ R\% = \frac{HighestHigh - Close}{HighestHigh - LowestLow}$
۹	مستقل	استوکاستیک سریع سهم i در دوره $t-1$	$K\%\ .D\%_{i.t-1}$	$Fast\ K\% = 100 \times \frac{Close - Low}{High - Low}$
۱۰	مستقل	شاخص جریان پول سهم i در دوره $t-1$	$MFI_{i.t-1}$	$MFI = 100 - \frac{100}{1+MFR}$، $MFR = \frac{5period\ positive\ money\ flow}{5period\ negative\ money\ flow}$
۱۱	مستقل	حجم معاملات متوازن سهم i در دوره $t-1$	$OBV_{i.t-1}$	$OVB = OVB_{prev} \begin{cases} volume & \text{if } close > close_{prev} \\ 0 & \text{if } close = close_{prev} \\ -volume & \text{if } close < close_{prev} \end{cases}$
۱۲	مستقل	شاخص جهت میانگین	$ADX_{i.t-1}$	خروجی از ره‌آورد نوین
۱۳	مستقل	شاخص کانال قیمت سهم i در دوره $t-1$	$CCI_{i.t-1}$	$CCI = \frac{(M-A)}{0.015 \times D}$
۱۴	مستقل	قیمت بسته‌شدن (پایانی) سهام i ام در دوره $t-1$	$Pc_{i.t-1}$	

خروجی مدل‌های فوق شامل سه وضعیت نگهداری، فروش و خرید برای روز آتی می‌باشد که در مقایسه با داده‌های واقعی برای دوره آموزش و آزمون بر اساس معیارهای ارزیابی عملکرد طبقه‌بندی می‌شوند. یکی از نکات مهم در استفاده از طبقه‌بندی‌کننده‌های مختلف، ارزیابی کیفیت روش‌های طبقه‌بندی است. روش‌های ارزیابی عملکرد مختلفی برای طبقه‌بندی‌کننده‌های گسسته و احتمالی وجود دارد. در طبقه‌بندی‌کننده‌های گسسته، یک کلاس به‌عنوان نتیجه طبقه‌بندی در نظر گرفته‌می‌شود. برای طبقه‌بندی‌کننده گسسته، ارزیابی عملکرد از قبیل F-measure, recall, precision, accuracy

استفاده می‌شود. به‌منظور ارزیابی عملکرد طبقه‌بندی‌کننده دو مجموعه‌داده که مجموعه آموزش و آزمون نامیده می‌شود موردنیاز است. از مجموعه‌داده آموزش برای ساخت مدل استفاده می‌شود و از مجموعه‌داده‌های آزمون برای ارزیابی عملکرد طبقه‌بندی‌کننده استفاده می‌شود. مجموعه آزمون و آموزش می‌تواند شامل کل یا بخشی از مجموعه‌داده‌ها باشد. استفاده از کل مجموعه‌داده به‌عنوان هم مجموعه آزمون و هم آموزش دارای این مزیت است که این داده‌ها کاملاً کاوش می‌شوند اگرچه این نتایج گاهی اوقات بیش از حد خوش‌بینانه و غیرواقعی است زیرا کلاس‌های این تاپل‌ها از قبل شناخته شده هستند. رایج‌ترین روش تقسیم مجموعه‌داده به دو قسمت است: مجموعه‌داده‌های آزمون و آموزش و معمولاً مجموعه‌داده آموزش بزرگ‌تر از مجموعه‌داده آزمون است (Wagner and Rondinelli, 2016). صحت، دقت، فراخوانی و F-measure معیارهای معمول برای ارزیابی نتایج طبقه‌بندی‌ها می‌باشد. پیش‌بینی‌های یک طبقه‌بندی‌کننده در شکل مثبت و منفی جدول احتمالات چهار سلولی خلاصه می‌شود. طبقه‌بندی‌کننده به اعداد منفی و مثبت که نشان‌دهنده شرایط آزمون است برمی‌گردد (Tabar *et al.*, 2021). وقتی توزیع کلاس یکنواخت نیست، از ماتریس سردرگمی برای ارزیابی دقت مدل استفاده می‌شود. خان و همکاران (Khan *et al.*, 2020) یکی از نکات مهم در استفاده از طبقه‌بندی‌کننده‌های مختلف، ارزیابی کیفیت روش‌های طبقه‌بندی است که تعیین می‌کند چه زمانی از تکنیک‌های خاص برای دستیابی به بهترین نتایج باید استفاده‌کرد که در این مطالعه ماتریس سردرگمی ۳×۳ است. ماتریسی که در آن مقادیر ستون‌ها مقدار روند واقعی را نشان می‌دهند، مقادیر نشان‌داده‌شده در ردیف‌ها پیش‌بینی شده روندها است، درحالی‌که مقادیر مورب روندهای پیش‌بینی شده درست را نشان می‌دهند. مقادیر موجود در ماتریس توسط برخی اصطلاحات خاص نشان داده می‌شوند. به‌عنوان‌مثال، مثبت درست (TP) برای روند مثبت صحیح، خنثی درست (TNT) برای روند خنثی صحیح، منفی درست (TNG) برای روند منفی درست استفاده می‌شود. به‌طور مشابه، روند مثبت طبقه‌بندی شده در صورت خنثی یا منفی بودن (مثبت کاذب) به‌ترتیب توسط FPnt و FPng و روند خنثی طبقه‌بندی‌شده در صورت مثبت یا منفی (خنثی کاذب) به‌ترتیب، توسط FNTP و FNTng نشان داده‌می‌شود، درحالی‌که روند منفی طبقه‌بندی شده در صورت مثبت یا خنثی (منفی کاذب) به‌ترتیب توسط FNGp و FNGnt نشان داده می‌شود. دقت کلی در رابطه (۳) محاسبه می‌شود:

$$Accuracy = \frac{TP + TNT + TNG}{TP + TNT + TNG + FPnt + FPng + FNTP + FNTng + FNGp + FNGnt} \quad \text{رابطه (۳)}$$

معیار صحت، بیانگر نسبت موارد مثبت پیش‌بینی شده است که واقعاً مثبت بوده است. در رابطه (۴) برای کلاس روند مثبت بیان می‌شود.

$$Precision = TP/TP + FPnt + FPng$$ رابطه (۴)

فراخوانی، بیانگر نسبت مثبت‌های واقعی است که به‌درستی مثبت پیش‌بینی شده است. رابطه (۵) برای کلاس روند مثبت ارائه شده است.

$$Recall = TP/TP + FNTp + FNGp$$ رابطه (۵)

نمره F ترکیبی از معیارهای صحت و فراخوانی است و می‌تواند با رابطه (۶) برای کلاس روند مثبت محاسبه شود (Khan *et al.*, 2020).

$$F_{-Measure} = \frac{2 \times (Precision \times Recall)}{Precision + Recall}$$ رابطه (۶)

ازآنجاکه مسئله فعلی یک طبقه‌بندی چندطبقه است و توزیع کلاس یکنواخت نیست، ماتریس سردرگمی انتخاب گردید، معیارهای دقت، صحت و اندازه‌گیری F برای ارزیابی الگوریتم‌های یادگیری ماشین انتخاب شده است. به‌منظور ایجاد قابلیت مقایسه روش‌های پیش‌بینی شبکه‌های عصبی و الگوریتم‌های یادگیری ماشین و روش تصادفی به‌عنوان مبنای مقایسه قرار می‌گیرد که باتوجه‌به اینکه خروجی مدل‌ها سه وضعیت خرید فروش و نگهداری می‌باشد و لذا انتظار می‌رود دقت روش تصادفی در تعداد تکرارهای زیاد با ۳۳/۳ درصد برابر گردد؛ یعنی هر یک از وضعیت‌های فوق‌الذکر شانس برابری در انتخاب‌شدن به‌عنوان وضعیت خروجی دارد و لذا درصورتی‌که F-measure دوره آزمون حاصل از روش‌ها بالاتر از روش تصادفی باشد آن روش توانا محسوب می‌گردد.

در آزمون فرضیه‌های دوم الی چهارم از آزمون مربوط بودن برای به‌منظور سنجش محتوای اطلاعاتی سود بهره برده می‌شود. محتوای اطلاعاتی سود در پژوهش‌های تجربی حسابداری از طریق آزمون ارتباط بین سود و بازده سنجیده می‌شود. به‌بیان‌دیگر هر چه سود از توانایی بیشتری در توضیح قیمت یا بازده سهام برخوردار باشد، محتوای اطلاعاتی بیشتری داشته، یا به‌بیان‌دیگر، مربوطتر است (Nasiri Pour *et al.*, 2016). به‌صورت خاص محتوای اطلاعات سود خالص گزارش‌شده می‌تواند با میزان تغییر قیمت سهام و یا با میزان بازده غیرعادی در حوالی زمانی تعریف شود که بازار از سود خالص دوره جاری آگاه می‌شود. ازاین‌رو استدلال بر این است که سرمایه‌گذاران در زمان آگاه شدن از سود حاوی اطلاعات نظریات خود در مورد بازده آینده تعدیل و بر این اساس به خریدوفروش سهام اقدام می‌کنند که نتیجه آن تغییر قیمت سهام خواهد بود (Poretti *et al.*, 2018). آزمون مربوط بودن محتوای اطلاعاتی سود با مدل‌های مختلفی از سوی محققین سنجیده می‌شود که یکی از مدل‌های

متداول اندازه‌گیری محتوای اطلاعاتی سود حسابداری مدل پورتتی و همکاران (Poretti *et al.*, 2018) استفاده از رابطه (۷) ارائه می‌گردد:

$$AR_{i.t} = \alpha_0 + \beta_1 \Delta NI_{i.t} + \varepsilon_{it} \quad \text{رابطه (۷)}$$

$AR_{i.t}$: بازده غیرعادی سهام شرکت که متغیر وابسته این پژوهش است.

α_0: جزء ثابت.

β_1: ضریب رگرسیون.

ε: جزء خطا.

$\Delta NI_{i.t}$: تغییرات دوره‌ای سود خالص شرکت i در دوره مالی t.

ضریب واکنش سود[1] (ضریب آماره t متغیر $\Delta NI_{i.t}$) برای اندازه‌گیری محتوای اطلاعاتی سود حسابداری استفاده می‌شود. ضرایب واکنش سود، ضرایب معادله خطی تغییرات سود هر سهم با بازده دوره‌ای سهام است. به‌عبارت دیگر معنی‌داری مدل رگرسیون رابطه (۸) نشان‌دهنده محتوای اطلاعاتی سود حسابداری است. اگر تغییرات سود هم زمان با بازده سهام تغییر کند نشان‌دهنده محتوای اطلاعاتی است. به‌عبارت دیگر تغییر در بازده سهام واکنش به تغییرات سود در یک دوره مشخص است. بنابراین بازدهی به سود مربوط بوده و ضریب تغییر سود به‌عنوان معیاری از محتوای اطلاعاتی سود در نظر گرفته‌می‌شود و به زبان آماری در صورتی که ضریب متغیر، β_1 معنادار باشد نشان دهنده این است که تغییرات سودآوری با بازده سهام رابطه معنادار دارد و می‌توان نتیجه گرفت سود حسابداری دارای محتوای اطلاعاتی می‌باشد.

از بین روش‌های سنجش محتوای اطلاعات سود حسابداری، مدل پورتی و همکاران (Poretti *et al.*, 2018) باتوجه‌به هدف پژوهش و اینکه بیشترین رجوع را توسط محققین داشته است، به‌عنوان مدل مرجع در این پژوهش نیز مورداستفاده قرار می‌گیرد.

و لذا مدل آماری فرضیه‌های دوم الی چهارم به شرح زیر می‌باشد:

به‌منظور آزمون فرضیه‌های دوم الی چهارم مبنی بر توانایی پیش‌بینی محتوای اطلاعاتی سود قبل از اعلان توسط ماشین بردار پشتیبان، الگوریتم جنگل تصادفی و شبکه‌های عصبی مصنوعی از مدل آماری رابطه (۸) بهره برده می‌شود.

$$NNAR_{i.t} = \alpha_0 + \beta_1 \Delta NI_{i.t} + \varepsilon_{i.t}$$
$$H_0: \beta_1 \leq 0 \quad \text{رابطه (۸)}$$
$$H_1: \beta_1 > 0$$

1 Earning Response Coefficient (ERC)

$NNAR_{i.t}$ بازده غیرعادی سهم i ام در دوره مالی t برآورد شده توسط مدل ماشین بردار پشتیبان، جنگل تصادفی و شبکه‌های عصبی.

$\Delta NI_{i.t}$ درصد تغییر سود سهم i ام در دوره مالی t.

٤- تجزیه‌وتحلیل داده‌ها

برای بررسی نرمال بودن مشاهدات از آزمون شپیرو–ویلکاکسون استفاده می‌گردد و باتوجه‌به نتایج، ازآنجاکه نوع داده‌های پژوهش پنل دیتا است، فرضیه نرمال بودن مشاهدات رد می‌شود. به‌منظور آزمون مانایی داده‌ها از آزمون ریشه واحد دیکی فولر تعمیم‌یافته استفاده شده است که در هر دو متغیر آن مقدار احتمال (p-value) کوچک‌تر از سطح معنی‌داری ۹۵ درصد ($\alpha = 0/05$) بوده که نشان از رد فرضیه صفر و مانایی متغیرها دارد.

فرضیه اول: ماشین بردار پشتیبان، الگوریتم جنگل تصادفی و شبکه‌های عصبی مصنوعی توانایی پیش‌بینی قیمت آتی سهام قبل از اعلان سود فصلی شرکت‌ها را دارند.

متغیرهای موردنیاز ورودی در مدل‌های شبکه عصبی و یادگیری ماشین از طریق نرم‌افزار ره‌آورد نوین استخراج گردیده است و در نرم‌افزار اکسل ۲۰۱۷ به‌منظور ورود به نرم‌افزار متلب آماده‌سازی شده است.

پارامترهای مدل در الگوریتم ماشین بردار پشتیبان[1] به شرح زیر می‌باشد:

- به‌منظور آموزش از تابع fitcecoc استفاده شده است.
- تابع کرنل پلی‌نومیال درجه ۴ به‌عنوان تابع پیش‌بینی (با سعی و خطا) انتخاب گردیده است.
- ۹۰ درصد داده‌ها به‌عنوان دوره آموزش و ۱۰ درصد داده‌ها به‌عنوان دوره آزمون لحاظ شده است.
- به‌واسطه گسسته بودن متغیر خروجی، دقت مدل بر اساس ماتریس سردرگمی تعیین می‌گردد.

متغیرهای ورودی روش ماشین بردار پشتیبان نیز همانند دو روش فوق بوده است و متغیرها نرمال گردیده و بعد به‌عنوان ورودی مدل مورداستفاده قرار گرفته‌اند. تابع به‌کاررفته در مدل کرنل پلی‌نومیال درجه ۴ می‌باشد.

جدول (۲) شاخص‌های ارزیابی عملکرد طبقه‌بندی دوره آموزش و آزمون

Table (2) Performance evaluation indicators for classification of training and test

شرح روش	Accuracy	Precision	Recall	F_score
SVM-Training	٪ ۷۲/۹۰	٪ ۷۱/۳۰	٪ ۷۳/۵۰	٪ ۷۲/۳۰
SVM-Test	٪ ۴۳/۴۵	٪ ۳۶/۳۰	٪ ۳۶/۸۰	٪ ۵۹/۵۴

1. SVM

در روش ماشین بردار پشتیبان ماکزیمم دقت حاصل شده در دوره آموزش معادل ۹۷/۵۱ درصد و میانگین دقت کسب شده ۷۲/۹۰ درصد و برای دوره آزمون ماکزیمم دقت حاصل شده معادل ۹۵/۴۵ درصد و میانگین دقت کسب شده ۴۳/۴۵ درصد می‌باشد.

پارامترهای الگوریتم جنگل تصادفی به شرح زیر می‌باشد:

- از تابع treebager به‌منظور آموزش استفاده می‌گردد و باتوجه‌به اینکه خروجی پژوهش حاضر گسسته می‌باشد خروجی این تابع طبقه‌بندی تعریف می‌گردد.
- خروجی‌ها در این مدل متنی هستند و لذا در انتهای پیش‌بینی لازم است به عدد تبدیل گردند.
- پارامتر اصلی تعداد درختان می‌باشد که بیشینه تعداد درخت‌های جنگل که تعیین‌کننده عمق درخت‌ها می‌باشد، برابر تعداد داده‌های آزمایشی در نظر گرفته شده است.
- ۹۰ درصد داده‌ها به‌عنوان دوره آموزش و ۱۰ درصد داده‌ها به‌عنوان دوره آزمون لحاظ شده است.
- به‌واسطه گسسته بودن متغیر خروجی دقت مدل بر اساس ماتریس سردرگمی تعیین می‌گردد.

جدول (۳) شاخص‌های ارزیابی عملکرد طبقه‌بندی دوره آموزش و آزمون

Table (3) Performance evaluation indicators for classification of training and test

شرح روش	Accuracy	Precision	Recall	F_score
RF -Training	٪ ۹۹/۹۹	٪ ۹۹/۹۹	٪ ۹۹/۹۹	٪ ۹۹/۹۹
RF -Test	٪ ۴۹/۰۱	٪ ۳۵/۶۳	٪ ۳۷/۱۵	٪ ۶۳/۸۱

در روش جنگل تصادفی ماکزیمم دقت حاصل شده در دوره آموزش معادل ۱۰۰ درصد و میانگین دقت کسب شده ۹۹/۹۹ درصد و برای دوره آزمون ماکزیمم دقت حاصل شده معادل ۹۱/۵۳ درصد و میانگین دقت کسب شده ۴۹/۰۱ درصد می‌باشد.

در شبکه‌های عصبی RNN تکنیک LSTM به‌منظور افزایش دقت پیش‌بینی به کار گرفته شده است که پارامترهای مدل به شرح زیر است:

- تعداد نرون‌های لایه ورودی به تعداد متغیرها تعریف شده است که معادل ۱۰ نرون می‌باشد.
- نوع لایه پنهان LSTM دوطرفه bilstmLayer می‌باشد و لایه‌های چسبیده به لایه bilstmLayer به‌صورت fullyConnectedLayer و تعداد لایه اتصال کامل برابر ۱، تعداد نورون درون لایه اتصال کامل برابر با ۳۰.
- تابع انتقال (فعال‌سازی) لایه پنهان تانژانت سیگموئیدی می‌باشد.
- نوع داده‌ها لازم است به‌صورت سلولی تعریف گردیده و تعداد سلول حافظه برابر ۲۰۰ در نظر گرفته شده است.

- دسته‌بندی داده‌ها به‌منظور آموزش به دسته‌های ۸تایی تقسیم گردیده است (miniBatchSize) به‌منظور اینکه حجم آموزش کاهش یابد.
- تعداد ایپاک برابر ۵ در نظر گرفته شده است.
- روش آموزش SGDM می‌باشد.
- تواتر validation به ازاء هر سه بار اجرا یک‌بار صورت می‌گیرد و در صورت مشاهده ۴۰ بار عدم بهبود فرایند آموزش متوقف می‌گردد.
- GradientThreshold برابر یک قرار داده شده است و نرخ یادگیری اولیه برابر ۰/۰۱ قرار داده شده است.
- ۹۰ درصد داده‌ها به‌منظور دوره آموزش و ۱۰ درصد داده‌ها به‌عنوان دوره آزمون در نظر گرفته شده است.
- به‌واسطه گسسته بودن متغیر خروجی دقت مدل بر اساس ماتریس سردرگمی تعیین می‌گردد.
- برای تمامی سهم‌ها مدل فوق اجرا گردید و شاخص‌های ارزیابی عملکرد طبقه‌بندی به شرح جدول (۴) می‌باشد.

جدول (۴) شاخص‌های ارزیابی عملکرد طبقه‌بندی دوره آموزش و آزمون

Table (4) Performance evaluation indicators for classification of training and test

شرح روش	Accuracy	Precision	Recall	F_score
LSTM -Training	٪ ۳۹/۹۰	٪ ۳۳/۰۰	٪ ۲۳/۳۰	٪ ۲۵/۶۰
LSTM -Test	٪ ۴۰/۸۲	٪ ۳۳/۳۳		

در روش شبکه‌های عصبی میانگین دقت کسب شده ۳۹/۹ درصد و F_score معادل ۲۵/۶۰ درصد که کمتر از دقت روش تصادفی (۳۳/۳۳ درصد) می‌باشد برای دوره آزمون میانگین دقت حاصل شده ۴۰/۸۲ درصد و امکان ارائه F_score بالاتر از مدل تصادفی را نداشته است و لذا نتیجه می‌گردد مدل توان پیش‌بینی قیمت سهام در زمان اعلان سود فصلی شرکت‌ها و ایجاد استراتژی مناسب خرید، فروش و نگهداری را ندارد.

در ادامه خلاصه دقت حاصل شده از سه روش در دوره‌های آموزش و آزمون در مقایسه با روش تصادفی در جدول (۵) ارائه شده است.

جدول (۵) نتایج شاخص‌های ارزیابی طبقه‌بندی روش‌ها در دوره آموزش

Table (5) Results of Methods for Evaluating the Classification of Methods in the Training

شرح روش	Accuracy	Precision	Recall	F_score
SVM-Training	٪ ۷۲/۹۰	٪ ۷۱/۳۰	٪ ۷۳/۵۰	٪ ۷۲/۳۰
RF-Training	٪ ۹۹/۹۹	٪ ۹۹/۹۹	٪ ۹۹/۹۹	٪ ۹۹/۹۹
LSTM-Training	٪ ۳۹/۹۰	٪ ۳۳/۰۰	٪ ۲۳/۳۰	٪ ۲۵/۶۰

جدول (۶) نتایج شاخص‌های ارزیابی طبقه‌بندی روش‌ها در دوره آزمون

Table (6) Results of evaluation classification indicators of methods in the test period

شرح روش	Accuracy	Precision	Recall	F_score
SVM-test	٪ ۴۳/۴۰	٪ ۳۶/۳۰	٪ ۳۶/۸۰	٪ ۳۶/۴۰
RF-test	٪ ۴۹/۱۰	٪ ۳۵/۶۰	٪ ۷۱/۱۰	٪ ۳۶/۳۰
LSTM-test	٪ ۴۰/۸۰	٪ ۳۳/۳۰		

جدول (۷) خلاصه دقت (Accuracy) حاصل از سه روش در مقایسه با روش تصادفی

Table (7) Summary of accuracy obtained from three methods compared to the random method

روش پیش‌بینی	دوره آموزش		دوره آزمون	
	میانگین دقت	ماکزیمم دقت	میانگین دقت	ماکزیمم دقت
RF	٪ ۹۹/۹۹	٪ ۱۰۰	٪ ۴۴/۰۸	٪ ۹۱/۵۳
RNN-LSTM	٪ ۲۳/۳۰	٪ ۲۵/۶۰		
SVM	٪ ۷۲/۹۰	٪ ۹۷/۵۱	٪ ۴۳/۳۰	٪ ۹۵/۴۵
روش تصادفی	٪ ۳۳/۳۰	٪ ۳۳/۳۰	٪ ۳۳/۳۰	٪ ۳۳/۳۰

با توجه به میانگین بالاتر و معنادار دقت دو روشRF و SVM در دوره آزمون در مقایسه با دو روش دیگر و همچنین دقت قابل‌قبول در مقایسه با روش تصادفی نتیجه می‌گردد ماشین بردار پشتیبان، الگوریتم جنگل تصادفی توانایی پیش‌بینی قیمت آتی سهام قبل از اعلان سود فصلی شرکت‌ها را دارند و شبکه‌های عصبی مصنوعی امکان ارائه F_measure بالاتر از مدل تصادفی را نداشته است و لذا نتیجه می‌گردد مدل توان پیش‌بینی قیمت سهام در زمان اعلان سود فصلی شرکت‌ها و ایجاد استراتژی مناسب خرید، فروش و نگهداری را ندارد.

فرضیه دوم: ماشین بردار پشتیبان توانایی پیش‌بینی محتوای اطلاعاتی سود را قبل از اعلان آن دارند.

بر اساس استراتژی خرید/فروش و نگهداری روزانه پیش‌بینی‌شده توسط روش SVM و با لحاظ قیمت واقعی سهم در همان تاریخ‌ها و در بازه زمانی ۳۰ روز قبل تا ۱۰ روز بعد از تاریخ اعلان سود اقدام به محاسبه بازده غیرعادی بر اساس روش معرفی شده می‌گردد به‌منظور انجام محاسبات در محیط پایتون کدنویسی شده است و در ادامه به‌منظور آزمون مربوط بودن بازده غیرعادی ایجاد شده به تغییرات سود، از مدل پورتی و همکاران (Poretti *et al.*, 2018) بهره برده می‌شود که آزمون آن در رابطه (۷) صورت‌بندی گردیده است.

بر اساس مندرجات جدول (۸) ضریب تعیین برابر است با ۰/۲۲، به این معنی که ۲۲ درصد از تغییرات متغیر بازده غیرعادی با استفاده از این مدل رگرسیونی تبیین می‌شود، بر اساس نتایج جدول (۹) آنالیز واریانس این مدل رگرسیونی معنی‌دار است.

جدول (۸) خلاصه مدل

Table (8) Model Summary

مدل	ضریب همبستگی	ضریب تعیین	ضریب تعیین تعدیل‌شده
۱	۰/۱۴۹[a]	۰/۲۲	۰/۱۸

جدول (۹) نتایج آنالیز واریانس

Table (9) Results of analysis of variance

	مدل	مجموع مربعات	درجه آزادی	میانگین مربعات	F	سطح معنی‌داری
۱	رگرسیون	۸۱۲۲/۳۸۹	۱	۸۱۲۲/۳۸۹	۴/۶۶۰	۰/۰۳۲[b]
	باقیمانده‌ها	۳۵۵۴۱/۲۲۸	۲۰۴	۱۷۴۲/۸۴۹		
	کل	۳۶۳۶۳/۶۱۶	۲۰۵			

جدول (۱۰) ضرایب مدل رگرسیونی

Table (10) Coefficients of regression model

	مدل	ضرایب استاندارد		t	سطح معنی‌داری
		B	خطای استاندارد	Beta		
۱	(Constant)	-۲۷/۲۰۰	۲/۹۸۱		-۹/۱۲۴	۰/۰۰۰
	درصد تغییر سود	۰/۰۲۹	۰/۰۱۳	۰/۱۴۹	۲/۱۵۹	۰/۰۳۲

باتوجه‌به نتایج جدول (۱۰)، مشاهده می‌شود که متغیر درصد تغییر سود دارای آماره ۰۲۹/ ۰، ضریب ۲۷/۲- و سطح معناداری کمتر از ۵ درصد (۰/۰۳۲) است. ازاین‌رو فرضیه صفر یعنی عدم ارتباط میان دو متغیر رد شده و می‌توان گفت متغیر درصد تغییر سود دارای ارتباط مستقیم و معنادار با بازده غیرعادی است. به این مفهوم که افزایش (کاهش) یک واحد در درصد تغییر سود باعث می‌شود که ۲/۹ درصد بر بازده غیرعادی افزوده (کاسته) شود و فرضیه دوم مورد تأیید است و لذا الگوریتم ماشین بردار پشتیبان توانایی پیش‌بینی محتوای اطلاعاتی سود را قبل از اعلان آن دارد.

فرضیه سوم: الگوریتم جنگل تصادفی توانایی پیش‌بینی محتوای اطلاعاتی سود را قبل از اعلان آن دارند.

بر اساس استراتژی خرید/فروش و نگهداری روزانه پیش‌بینی‌شده توسط روش RF و با لحاظ قیمت واقعی سهم در همان تاریخ‌ها و در بازه زمانی ۳۰ روز قبل تا ۱۰ روز بعد از تاریخ اعلان سود اقدام به محاسبه بازده غیرعادی بر اساس روش معرفی شده می‌گردد به‌منظور انجام محاسبات در محیط پایتون

کدنویسی شده است و در ادامه به‌منظور آزمون مربوط بودن بازده غیرعادی ایجاد شده به تغییر سود همانند فرضیه دوم، از روش پورتی و همکاران (Poretti *et al.*, 2018) بهره برده می‌شود که آزمون آن در رابطه (۷) صورت‌بندی گردیده است.

بر اساس مندرجات جدول (۱۱) ضریب تعیین برابر است با ۰/۰۲۱ به این معنی که ۲ درصد از تغییرات متغیر بازده غیرعادی با استفاده از این مدل رگرسیونی تبیین می‌شود، بر اساس نتایج جدول (۱۲) آنالیز واریانس این مدل رگرسیونی با توجه سطح معنی‌داری ۰/۰۵۴ در سطح ۰/۰۵ معنی‌دار نمی‌باشد.

جدول (۱۱) خلاصه مدل

Table (11) Model Summary

مدل	ضریب همبستگی	ضریب تعیین	ضریب تعیین تعدیل‌شده
۱	۰/۱۴۷[a]	۰/۰۲۱	۰/۰۱۶

جدول (۱۲) نتایج آنالیز واریانس

Table (12) Results of analysis of variance

	مدل	مجموع مربعات	درجه آزادی	میانگین مربعات	F	سطح معنی‌داری
۱	رگرسیون	۴۹۳۲/۵۶۲	۱	۴۹۳۲/۵۶۲	۳/۷۵۶	۰/۰۵۴[b]
	باقیمانده‌ها	۲۲۴۵۹۲/۰۱۵	۱۷۱	۱۳۱۳/۴۰۴		
	کل	۲۲۹۵۲۴/۵۷۷	۱۷۲			

جدول (۱۳) ضرایب مدل رگرسیونی

Table (13) Coefficients of regression model

	مدل	ضرایب استاندارد		t	سطح معنی‌داری
		B	خطای استاندارد	Beta		
۱	(Constant)	-۲۵/۶۹۹	۲/۸۴۲		-۹/۰۴۲	۰/۰۰۰
	درصد تغییر سود	۰/۰۲۸	۰/۰۱۴	۰/۱۴۷	۱/۹۳۸	۰/۰۵۴

باتوجه‌به نتایج جدول (۱۳)، مشاهده می‌شود که سطح معناداری بیش از ۵ درصد (۰/۰۵۴) است. ازاین‌رو فرضیه صفر یعنی عدم ارتباط میان دو متغیر تأیید شده و می‌توان گفت متغیر درصد تغییر سود دارای ارتباط مستقیم و معنادار با بازده غیرعادی نمی‌باشد و لذا بر اساس داده‌های RF فرضیه سوم رد می‌گردد و به معنی این است که الگوریتم یادگیری ماشین RF توانایی پیش‌بینی محتوای اطلاعاتی سود را قبل از اعلان آن را ندارد.

فرضیه چهارم: شبکه‌های عصبی مصنوعی توانایی پیش‌بینی محتوای اطلاعاتی سود را قبل از اعلان آن دارند.

با مدنظر داشتن نتایج حاصل از فرضیه اول که مؤید عدم توانایی شبکه‌های عصبی مصنوعی در پیش‌بینی قیمت آتی سهام قبل از اعلان سود فصلی شرکت‌ها و در نتیجه عدم توانایی در ایجاد استراتژی نگهداری، خریدوفروش و لذا این مدل توانایی پیش‌بینی محتوای اطلاعاتی سود را قبل از اعلان آن ندارد.

۵- بحث و نتیجه‌گیری

توانایی بالای الگوریتم های یادگیری ماشین و شبکه‌های عصبی مصنوعی در زمینه پیش‌بینی رفتار بازده (توده واری)، در مقایسه با روش‌های خطی مبتنی بر رگرسیون توسط پژوهشگرانی نظیر چاجر و همکاران (Chaajer *et al.*, 2021)، یین و همکاران (Yin *et al.*, 2023)، سریواستاوا و همکاران (Srivastava *et al.*, 2021)، امیوما و ادکی (Emioma and Edeki, 2021)، سادورسکی (Sadorsky, 2021)، شارما و همکاران (Sharma *et al.*, 2022)، غلامیان و داوودی (Gholamian and Davoodi, 2018) به اثبات رسیده است لیکن تاکنون در ایران پژوهشی مبنی بر توان پیش‌بینی محتوای اطلاعاتی سود توسط شبکه‌های عصبی مصنوعی و الگوریتم‌های یادگیری ماشین صورت نگرفته است. در پژوهش حاضر، بر اساس نتایج حاصل از جنگل تصادفی، علیرغم دقت قابل قبول مدل در پیش‌بینی استراتژی خرید، فروش و نگهداری، لیکن در آزمون مربوط بودن توان توضیح تغییرات سود را نداشته و لذا نتیجه می‌گردد که الگوریتم جنگل تصادفی توانایی پیش‌بینی محتوای اطلاعاتی سود قبل از اعلان آن را ندارد. بر اساس نتایج حاصل از ماشین بردار پشتیبان، دقت قابل قبول مدل در پیش‌بینی استراتژی خرید، فروش و نگهداری و همچنین تأیید در آزمون مربوط بودن موید توان توضیح تغییرات سود می باشد و لذا الگوریتم ماشین بردار پشتیبان توانایی پیش‌بینی محتوای اطلاعاتی سود قبل از اعلان آن را دارد. شبکه‌های عصبی مصنوعی در پیش‌بینی استراتژی خرید، فروش و نگهداری سهام، نسبت به دو الگوریتم جنگل تصادفی و ماشین بردار پشتیبان دقت پایین‌تری را ارائه نمود و لذا مشخص گردید، شبکه‌های عصبی مصنوعی قادر به پیش‌بینی محتوای اطلاعاتی سود قبل از اعلان آن نبوده است. لیکن با نتایج پژوهش سریواستاوا و همکاران (Srivastava *et al.*, 2021) از منظر مقایسه توانایی روش شبکه‌های عصبی مصنوعی با الگوریتم‌های یادگیری ماشین متفاوت می‌باشد. نتایج به‌دست‌آمده در این پژوهش با مستندات اشاره‌شده در چارچوب نظری پژوهش و ادبیات مالی مطابقت دارد که در ادامه اشاره می گردد. پژوهش درصدد یافتن راهکارهایی بود که قدرت پیش‌بینی محتوای اطلاعاتی سود را قبل از اعلان داشته باشد تا از این طریق سرعت انتقال اطلاعات به بازار و جذب آن را افزایش دهد که این عامل منجر به کاهش اثر عدم تقارن اطلاعاتی و اطلاعات نهانی و

افزایش کارایی بازار خواهد شد، که در این زمینه به هدف خود رسیده است و بخشی از خلاء تحقیقاتی مربوط به مسئله پژوهش که نبود ابزارها و تکنیک‌های پیش‌بینی محتوای اطلاعاتی سود قبل از اعلان آن بوده است را پاسخ داده است. برای سرمایه‌گذاران خرد که قادر به دستیابی به موقع به اطلاعات بنیادی و محرمانه شرکت‌ها نیستند، همچنین تحلیل‌گران بازار سرمایه و سرمایه‌گذاران کوتاه‌مدتی و سفته‌بازان، مشاوران صندوق‌های سرمایه‌گذاری و بازارگردانان سهام، ذینفعان یافته‌های این پژوهش مفید خواهد بود. مبتنی بر نتایج پژوهش مشخص گردید که سود اعلام شده فصلی شرکت‌ها دارای محتوای اطلاعاتی می‌باشد و لذا اعلام سود منجر به تغییر قیمت سهام در حوالی هفته‌های اعلام سود می‌گردد که بخشی از تغییرات قیمت قبل از اعلام سود روی می‌دهد. این موضوع ناشی از نشر و یا افشاء اطلاعات نهانی قبل از اعلام رسمی تغییرات سود می‌باشد که این نتایج با پژوهش بیور و همکاران (Beaver *et al.*, 2020) مطابقت دارد و همچنین نتایج بر توانایی الگوریتم ماشین بردار پشتیبان جهت پیش‌بینی محتوای اطلاعاتی سود و شناسایی اثر تغییرات سود قبل از اعلان آن صحه گذاشت که این نتیجه می‌تواند به سرمایه‌گذاران و تحلیل‌گران در تصمیم‌گیری کمک نماید و همچنین منجر به کمک به افزایش کارایی بازار با کاهش اثرات اطلاعات نهانی گردد که مطابق دیدگاه‌های سادورسکی (Sadorsky, 2021) می‌باشد، زیرا تمام سود معامله‌گر نهانی ناشی از وقفه میان زمانی است که وی شروع به فروش میکند و زمانی که بازار متوجه آنچه اتفاق افتاده میشود. آنچه به نفع بازار سرمایه است، شفافیت است، در حالیکه معامله‌گران نهانی از عدم‌شفافیت و ابهام منتفع خواهند شد. با توجه به نتایج پژوهش پیشنهاد می‌گردد بازارگردانها، سبدگردانها، سرمایه‌گذاران و تحلیل‌گران از الگوریتم ماشین‌بردار پشتیبان، به منظور انتخاب استراتژی خرید، نگهداری و فروش سهام در مقاطع زمانی اعلان سود فصلی شرکتها استفاده کنند که این موضوع علاوه بر ایجاد بازده غیرعادی، منجر به کاهش اثرات اطلاعات نهانی و افزایش کارایی بازار نیز می‌گردد. به قانون‌گذاران پیشنهاد می‌گردد به منظور کاهش اثر معاملات بر اساس اطلاعات نهانی نسبت به تصریح مصادیق تخلفات معاملات بر اساس اطلاعات نهانی و تدوین روش‌های کشف به موقع این نوع معاملات اقدام نمایند تا با افزایش شفافیت معاملات و کاهش عدم تقارن اطلاعات به افزایش کارایی بازار کمک گردد. به قانونگذاران پیشنهاد می‌گردد با توجه به اثبات محتوای اطلاعاتی سود در گزارش‌های فصلی، نسبت به اصلاح روش اعلام سود فصلی شرکتها به بورس و به حداقل رساندن زمان بین آماده شدن گزارشات فصلی در شرکت‌های ناشر تا اعلام عمومی اقدام نمایند.

٦- تعارض منافع

هیچ‌گونه تعارض منافع در این پژوهش وجود ندارد.

۷- منابع

Ashton, D., & Trinh, C. (2018). Evaluating the information content of earnings forecasts. *Accounting and Business Research, 48*(6), 674-699.

Beaver, W. H., McNichols, M. F., & Wang, Z. Z. (2018). The information content of earnings announcements: new insights from intertemporal and cross-sectional behavior. *Review of Accounting Studies, 23*, 95-135.

Beaver, W. H., McNichols, M. F., & Wang, Z. Z. (2020). Increased market response to earnings announcements in the 21st century: An empirical investigation. *Journal of Accounting and Economics, 69*(1), 101244.

Chaajer, P., Shah, M., & Kshirsagar, A. (2021). The applications of artificial neural networks, support vector machines, and long-short term memory for stock market prediction. *Decision Analytics Journal*, 100015.

Chen, K. Y., & Ho, C. H. (2005). An improved support vector regression modeling for Taiwan Stock Exchange market weighted index forecasting. In *2005 International conference on neural networks and brain* (Vol. 3, pp. nil15-1638). IEEE.

Emioma, C. C., & Edeki, S. O. (2021). Stock price prediction using machine learning on least-squares linear regression basis. In *Journal of Physics: Conference Series* (Vol. 1734, No. 1, p. 012058). IOP Publishing.

Engelen, P. J., & Van Liedekerke, L. (2007). The ethics of insider trading revisited. *Journal of Business Ethics, 74*, 497-507.

Esmaeeli, H., Rahmani, S., Kazemi, A., & Aliahmadi, M. (2017). Evaluation of E-Learning of the virtual learning program from the student's point of view. *Public Management Researches, 9*(34), 221-241. [In Persian]

Ghasemi, G., Khodamipour, A., & Shamsadini, K. (2022). The role of economic policy uncertainty in the relationship between voluntary disclosure of information and information asymmetry. *Advances in Finance and Investment, 3*(6), 31-52. [In Persian]

Gholamian, E., & Davoodi, S. M. R. (2018). Predicting the Direction of Stock Market Prices Using Random Forest. *Financial Engineering and Portfolio Management, 9*(35), 301-322. [In Persian]

Jiang, W. (2021). Applications of deep learning in stock market prediction: recent progress. *Expert Systems with Applications, 184*, 115537.

Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert systems with Applications, 38*(5), 5311-5319.

Khan, W., Malik, U., Ghazanfar, M. A., Azam, M. A., Alyoubi, K. H., & Alfakeeh, A. S. (2020). Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis. *Soft Computing, 24*, 11019-11043.

Levin, J. (2001). Information and the Market for Lemons. *RAND Journal of Economics*, 657-666.

Liu, C., Wang, W., Wang, M., Lv, F., & Konan, M. (2017). An efficient instance selection algorithm to reconstruct training set for support vector machine. *Knowledge-Based Systems, 116*, 58-73.

NasiriPour Dori, S., Banimahd, B., & Ahmadzadeh, H. (2016). Managerial Overconfidence and Earnings Overstatement. *Management Accounting, 9*(30), 55-65. [In Persian]

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert systems with applications, 42*(1), 259-268.

Piotroski, J. D., & Roulstone, D. T. (2005). Do insider trades reflect both contrarian beliefs and superior knowledge about future cash flow realizations?. *Journal of Accounting and Economics, 39*(1), 55-81.

Poretti, C., Schatt, A., & Bruynseels, L. (2018). Audit committees' independence and the information content of earnings announcements in Western Europe. *Journal of Accounting Literature, 40*(1), 29-53.

Pourgadimi, K., Bahri Sales, J., Jabbarzadeh Kangarluei, S., & Zavar Rezaee, A. (2022). Presenting the developed model of benish model with emphasis on audit quality fea-tures using neural network, vector machine and random forest. *Advances in Finance and Investment, 3*(6), 1-30. [In Persian]

Sadorsky, P. (2021). A random forests approach to predicting clean energy stock prices. *Journal of Risk and Financial Management, 14*(2), 48.

Srivastava, P. R., Zhang, Z. J., & Eachempati, P. (2021). Deep neural network and time series approach for finance systems: predicting the movement of the Indian stock market. *Journal of Organizational and End User Computing (JOEUC), 33*(5), 204-226.

Shao, S., Stoumbos, R. C., & Zhang, F. (2020). The Power of Firm Fundamental Information in Explaining Stock Returns. *SSRN Electronic Journal.*

Tabar, S., Sharma, S., & Volkman, D. (2021). Stock Market Prediction Using Elliot Wave Theory and Classification. *International Journal of Business Analytics (IJBAN), 8*(1), 1-20.

Wagner, N., & Rondinelli, J. M. (2016). Theory-guided machine learning in materials science. *Frontiers in Materials, 3*, 28.

Yin, L., Li, B., Li, P., & Zhang, R. (2023). Research on stock trend prediction method based on optimized random forest. *CAAI Transactions on Intelligence Technology, 8*(1), 274-284.

COPYRIGHTS

© 2023 by the authors. Published by Islamic Azad University, Esfarayen Branch. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) (http://creativecommons.org/licenses/by/4.0/).