Securities & Exchange Organization, Research, Development & Islamic Studies (RDIS)
Journal of Securities and Exchange, Summer 2023, V. 16, No.62, pp. 75-102

Comparing the Accuracy of Selected Machine Learning Models for Price Prediction in the Stock Exchange Market[1]

Hossein Kianizadeh[2], Ali Baghani[3], Mohsen Hamidian[4]

Received: 2023/01/29
Accepted: 2023/06/15

Research Paper

Abstract

The capital market is growing rapidly. This has led to higher demand for information, more effort to predict and invent new models for predicting the future of the market. Predictive models can be classified into three categories. The first group uses technical analysis, the second group uses fundamental analysis, and the third group uses data mining and machine learning. In the present study focusing on data mining method to compare the accuracy of selected machine learning models including neural network, logistic regression, k nearest neighborhood, support vector machine and cross validation to predict stock prices for 12 selected companies of Tehran Stock Exchange that They have been selected through systematic deletion method in the form of machine learning models and the results of this paper showed that among the machine learning algorithms, the support vector machine algorithm has the highest predictive power in stock prices.

Key Words: Stock Exchange, Forecast; Stock Prices, Algorithms, Machine Learning.
JEL Classification: C02, C13, C53, C55, C58, G17.

---

1. DOI: 10.22034/JSE.2021.11636.1733

2. Ph.D. Student, Department of Financial Management, Kish International Branch, Islamic Azad University, Kish Island, Iran. (kianizadeh@gmail.com).
3. Assistant Professor, Department of Accounting, Faculty of Economic and Accounting, South Tehran Branch, Islamic Azad University, Tehran, Iran. (Corresponding Author). (a_baghani@azad.ac.ir).
4. Associate Professor, Department of accounting, Faculty of Economic and Accounting, South Tehran Branch, Islamic Azad University, Tehran, Iran. (m_hamidian@azad.ac.ir).

سازمان بورس و اوراق بهادار، مرکز پژوهش، توسعه و مطالعات اسلامی

فصلنامه بورس اوراق بهادار، سال شانزدهم، شماره ۶۲، تابستان ۱۴۰۲، صص ۱۰۲-۷۵

مقایسه دقت مدل‌های منتخب یادگیری ماشین جهت پیش‌بینی قیمت سهام در بورس اوراق بهادار[۱]

حسین کیانی زاده[۲]، علی باغانی[۳]، محسن حمیدیان[۴]

مقاله پژوهشی

تاریخ دریافت: ۱۴۰۱/۱۱/۰۹
تاریخ پذیرش: ۱۴۰۲/۰۳/۲۵

چکیده

رشد بازار سرمایه با سرعت چشمگیری در حال افزایش است. همین امر باعث تقاضای بالاتر برای اطلاعات، تلاش بیشتر برای پیش‌بینی و ابداع مدل‌های جدید پیش‌بینی آینده بازار شده است. مدل‌های پیش‌بینی در سه دسته قابل طبقه‌بندی هستند. دسته اول از تحلیل تکنیکی، دسته دوم از تحلیل بنیادین و دسته سوم از داده‌کاوی و یادگیری ماشین استفاده می‌کنند. در پژوهش پیش رو با تمرکز بر روش داده کاوی به مقایسه دقت مدل‌های منتخب یادگیری ماشین شامل شبکه عصبی، رگرسیون لجستیک، نزدیک‌ترین همسایه k، ماشین بردار پشتیبان و اعتبارسنجی ضربدری برای پیش‌بینی قیمت سهام برای ۱۲ شرکت منتخب بورس اوراق بهادار تهران که از طریق روش حذف سیستماتیک انتخاب شده‌اند در قالب مدل‌های یادگیری ماشین پرداخته و نتایج این مقاله نشان داد از بین الگوریتمهای یادگیری ماشین، الگوریتم ماشین بردار پشتیبان بیشترین قدرت پیش‌بینی کنندگی در قیمت سهام را به خود اختصاص داده است.

واژه های کلیدی: بورس اوراق بهادار، پیش‌بینی؛ قیمت سهام، الگوریتم، یادگیری ماشین.

طبقه‌بندی موضوعی: C02, C13, C53, C55, C58, G17.

---

۱. DOI: 10.22034/JSE.2021.11636.1733

۲. دانشجوی دکتری، گروه مدیریت مالی، واحد بین الملل کیش، دانشگاه آزاد اسلامی، جزیره کیش، ایران. (kianizadeh@gmail.com).

۳. استادیار، گروه حسابداری، دانشکده اقتصاد و حسابداری، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران. (نویسنده مسئول). (a_baghani@azad.ac.ir).

۴. دانشیار، گروه حسابداری، دانشکده اقتصاد و حسابداری، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران. (m_hamidian@azad.ac.ir).

مقدمه

مفهوم ارزش از گستره اجتماعی در سطح ساده شکل می‌گیرد و به سطح گروه، خانواده، سازمان و جهان توسعه می‌یابد. تعامل ارزش و ارزش آفرینی، حجم افزایی را به ارمغان می‌آورد که بر تمامی فرآیندهای ارزش و مهم از همه بر تفکّر مدیریت مبتنی بر ارزش تأثیر خواهد گذاشت. ارزش مثل ارزش اجتماعی، اقتصادی، مالی و غیره، متغیری با بار معنایی گسترده است. از جمله عوامل مؤثر بر ارزش دارایی‌ها می‌توان به عواملی همچون محیط اقتصادی، زمان، موقعیت دارایی، شرایط فیزیکی، وضعیت بازار و میزان کمیابی اشاره نمود. واژه ارزش در مورد کالاها و خدمات با دو معنای ارزش کمی و ارزش کیفی به کار می‌روند. ارزش کیفی همان چیزی است که باعث ارضاء نیازها و خواسته‌های مصرف کننده می‌شود و ارزش کمی با ارزش مبادله کالاها با واسطه‌ای بنام پول در ارتباط است. مراحلِ فرآیند ارزشیابی شامل محاسبه ارزش شرکت، انتخاب مدل ارزشیابی، پیش‌بینی عملکرد آتی شرکت و شناخت شرکت است. پرسش اصلی در مدیریت مبتنی بر ارزش این است که ما چگونه شرکت خود را مدیریت کنیم تا بیشترین جریان نقدینگی و در نهایت بیشترین ارزش را دریافت کنیم؟ در این راستا، برآورد درست قیمت به ویژه قیمت سهام در کمک به مدیریت درست سرمایه‌گذاری در بازار سرمایه نقش بسزایی دارد (لانگ، سانگ و تیان ۲۰۱۹).

نظریه‌های متفاوتی در خصوص پیش‌بینی قیمت در بازارهای سازمان یافته مطرح شده است. در اوایل قرن بیستم، گروهی از متخصصان صاحب تجربه در ارزیابی اوراق بهادار اعتقاد راسخ بر این امر داشتند که می‌توان از طریق مطالعه و تجزیه و تحلیل روند تاریخی تغییرات قیمت سهام، تصویری را برای پیش‌بینی قیمت آینده سهام ارائه کرد. مطالعات علمی‌تر با تأکید بر شناسایی دقیق رفتار قیمت سهام، گرایش به سمت مدل‌های ارزشیابی قیمت سهام را به وجود آورد. در ابتدا نظریه گام‌های تصادفی به عنوان یک شروع در تعیین رفتار قیمت سهام مطرح شد. سپس به ویژگی‌ها و ساختار بازار سرمایه توجّه شد که نتیجه این مطالعات و بررسی‌ها منجر به فرضیه بازار کارای سرمایه شد. در بازار کارای سرمایه، اعتقاد بر این است که قیمت سهام انعکاسی از اطلاعات جاری مربوط به آن سهم است و تغییرات قیمت سهام دارای الگوی خاص پیش‌بینی نیست. نظریات مطرح شده تا دهه ۱۹۸۰ میلادی به خوبی تعیین کننده رفتار قیمت سهام در بازار بودند تا اینکه تحولات بازار سهام نیویورک در سال ۱۹۸۷، اعتبار فرضیه‌های بازار کارای سرمایه را به شدت زیر سوال برد. در دهه ۱۹۹۰ میلادی و بعد از آن، بیشتر توجّه

متخصصان به یک رفتار آشوبگرایانه همراه با نظم معطوف شد و تلاش در جهت طراحی مدل‌های غیرخطی به منظور پیش‌بینی قیمت سهام اهمیت روز افزونی یافت. بسیاری از مطالعات اخیر نشان داده‌اند بازار سهام، در حقیقت یک سیستم غیرخطی و آشوبناک است که به عوامل سیاسی، اقتصادی و روانی وابسته است. برای غلبه بر محدودیت تکنیک‌های تحلیل سنتی در پیش‌بینی الگوهای غیرخطی، متخصصان طی دو دهه اخیر تکنیک‌های هوشمند و بخصوص شبکه‌های عصبی مصنوعی، الگوریتم ژنتیک، الگوهای رگرسیون، یادگیری ماشین و غیره را برای بهبود پیش‌بینی قیمت سهام به کار برده‌اند (برادران حسن زاده و محرومی، ۱۳۹۶).

روش‌های تجزیه و تحلیل فنی، پیش‌بینی سری‌های زمانی، یادگیری ماشین و داده‌کاوی برای پیش‌بینی رفتار قیمت سهام وجود دارند. دسته نخست از نمودار به عنوان یک ابزار اصلی استفاده می‌کند. تحلیل‌گران از این نمودارها استفاده می‌کنند تا تصمیم به خرید یا فروش بگیرند. هدف دسته دوم پیش‌بینی قیمت‌های آتی سهام با تجزیه و تحلیل بازده قیمت‌های سهام در گذشته است. روش‌های متداول این دسته شامل روش‌های اتورگرسیو یا خودبرگشتی، مدل میانگین متحرک، مدل میانگین متحرک اتورگرسیو و مدل اتورگرسیو آستانه هستند. دسته سوم یا همان داده‌کاوی علم استخراج اطلاعات مفید از مجموعه عظیم داده‌ها یا پایگاه داده است (احمدخان بیگی، عبدالوند ۱۳۹۶). مشکل اصلی پیش‌بینی مسیر قیمت سهام تعداد زیاد داده‌ها است، محبوبیت داده‌کاوی در دنیای مالی رو به افزایش است. مجموعه داده‌ها به قدری عظیم است که نمی‌توان آن را با استفاده از روش‌های غیر داده‌کاوی اداره کرد به طوری که آن‌ها معنای اصلی را مبهم می‌کنند و نمی‌توان اطلاعات مفیدی از آن بدست آورد (تسیناسلانیدیس، کوگیموتزیس، ۲۰۱۴). پیش‌بینی قیمت‌های سهام هدف دارای اهمیتی در دنیای مالی است، زیرا یک پیش‌بینی دقیق معقولانه از این امکان برخوردار است که منافع مالی بسیاری را به همراه داشته باشد و با ریسک بازار مقابله کند (کومار، دنموژی، ۲۰۰۶). یکی از نکات مهم بحث در پژوهش‌های پیشین این است که آیا رویکرد قیمت سهام قابل پیش‌بینی است یا خیر؟ سرمایه‌گذاران مدت‌های مدیدی فرضیه بازار کارا را پذیرفتند. این فرضیه بیان می‌کند که با بررسی سیر تکاملی رفتار گذشته قیمت‌های سهام، هیچ بازده غیرعادی را نمی‌توان بدست آورد (میشل، دیرک، ناتالی و روبن، ۲۰۱۵). در ابتدای قرن بیست و یکم برخی از اقتصاددانان نشان دادند که قیمت‌های سهام در آینده حداقل تا حدی قابل پیش‌بینی هستند. بنابراین، الگوریتم‌های پیش‌بینیِ فراوانی مورد بررسی قرار گرفته و نشان داده شد که رفتار قیمت سهام را به راستی

می‌توان پیش‌بینی کرد. براساس نظریه تصمیم، شخص تصمیم گیرنده با توجّه به شرایط نامطمئن مایل است تا از بین گزینه‌های موجود، بهترین را انتخاب کند. این تصمیم‌گیری‌ها استوار بر انتظارهای عقلایی است. به عبارت دیگر، شخص تصمیم‌گیرنده در تصمیم‌گیری‌های خود بالاترین مطلوبیت را مورد توجّه قرار می‌دهد (بیرد، ردی، یئونگ، ۲۰۱۱).

یادگیری ماشین شاخه‌ای از فناوری هوش مصنوعی است که از شیوه‌ها یا الگوریتم‌های خاص برای ساخت خودکار مدل‌هایی از داده استفاده می‌کند. برخلاف یک سیستم معمولی که با پیروی از قوانین مشخص یک وظیفه را انجام می‌دهد، یک سیستم یادگیری ماشین از طریق تجربه کردن وظایف خود را یاد می‌گیرد. در حالی که یک سیستم مبتنی بر قانون هر بار یک وظیفه را به روشی یکسان (چه خوب، چه بد) انجام می‌دهد، وضعیت عملکرد یک سیستم یادگیری ماشین از طریق یادگیری و تمرین از طریق الگوریتم‌های داده‌کاوی بهبود پیدا می‌کند. الگوریتم‌های یادگیری ماشین با فراگیری از داده‌هایی که دریافت می‌کنند می‌توانند مسائلی را حل کنند که بسیار پیچیده‌تر از آن است که بتوان آن را با برنامه‌نویسی معمولی حل کرد (مینگ، ۲۰۱۹). از الگوریتم‌های یادگیری ماشین می‌توان به الگوریتم‌های طبقه‌بندی، الگوریتم‌های رگرسیون، الگوریتم‌های خوشه‌بندی و الگوریتم‌های کاهش ابعاد اشاره کرد. روش معمول برای یادگیری ماشین تقسیم مجموعه داده‌ها به زیر مجموعه‌هایی برای آموزش، اعتبارسنجی و آزمایش است. یک شبکه عصبی مصنوعی، یک پردازشگر توزیعی موازی و حجیم است که از واحدهای ساده پردازش تشکیل شده است. این شبکه قابلیت یادگیری از دانش تجربی ارائه شده در مقاومت‌های ارتباطی درون واحدی را دارد و می‌تواند چنین دانش موجودی برای استفاده ایجاد کند (ژانگ، ۲۰۰۵). یک شبکه عصبی ساختاری مشابه با یک سلول عصبی مصنوعی دارد. اطلاعات از طریق گره‌های لایه ورودی به شبکه وارد می‌شود. این ورودی‌ها از طریق رابط‌ها به گره‌های لایه‌های پنهان منتقل شده و بعد از پردازش از طریق لایه‌های مختلف از گره‌های لایه خروجی خارج می‌شوند. یک شبکه عصبی مصنوعی، مجموعه‌ای از نرون‌های به هم متصل در لایه‌های مختلف است که اطلاعاتی را برای یکدیگر ارسال می‌کنند. تمام لایه‌های شبکه عصبی به جز لایه ورودی کار پردازش را انجام داده و در نهایت خروجی به دست می‌آید. از انواع این روش‌ها می‌توان به شبکه عصبی پس‌خور و پیشخور اشاره کرد. شبکه‌های عصبی، اطلاعات ورودی را دریافت می‌کنند و پس از پردازش آن خروجی‌ها را نتیجه می‌دهند. برای حل مسائل، شبکه‌های عصبی سه مرحله را طی می‌کنند. آموزش، تعمیم (آزمایش) و اجرا

(ونلانگ، ۲۰۱۹) فرآیندی است که در آن شبکه می‌آموزد تا الگوی موجود در ورودی را که به صورت مجموعه داده‌های آموزشی است، بشناسد. برای این منظور هر شبکه عصبی از مجموعه‌ای از قوانین یادگیری که چگونگی یادگیری را تعریف می‌کنند، استفاده می‌کند. تعمیم یا آموزش، توانایی شبکه را برای ارائه جواب قابل قبول در قبال ورودی‌هایی که جزو مجموعه آموزشی نبوده‌اند، مورد سنجش قرار می‌دهد. استفاده از شبکه برای انجام وظیفه‌ای که به آن منظور طراحی شده است را اجرا می‌گویند. در اثر آموزش دادن شبکه، وزن‌های داخلی که بر روی ورودی‌های هر سلول عمل می‌کند، تغییر کرده و به وضعیت مناسب می‌رسد. یکی از نقاط ضعف شبکه‌های عصبی در این است که نتایج آموزش شبکه یعنی وزن‌های داخلی هیچ گونه تصویر روشنی از اعتبار جواب‌های مسأله بدست نمی‌دهد. این وزن‌ها به درستی قابل درک نیستند، با این وجود جواب‌های تولید شده توسط شبکه بیشتر صحیح هستند و این درستی جواب‌ها در بیشتر اوقات مهم‌تر از توضیح پذیر بودن آن است (به همین خاطر است که به شبکه‌های عصبی مدل‌های جعبه سیاه هم می‌گویند) (بیات و باقری، ۱۳۹۶).

الگوریتم‌های تکاملی از سیر تکاملی شبیه‌سازی شده برای پیدا کردن راه حل‌های مسایل پیچیده استفاده می‌کند. انواع زیادی از این الگوریتم‌ها وجود دارد، بررسی ادبیات این موضوع نشان می‌دهد که الگوریتم‌های ژنتیک و استراتژی‌های تکاملی اساس این الگوریتم‌ها هستند. پردازش تکاملی از شاخه‌های زیر تشکیل شده است. الگوریتم‌های ژنتیک روش‌های بهینه‌سازی– تصادفی بدون مشتق‌پذیری بوده که تا حدودی بر پایه مفاهیم انتخاب طبیعی و فرآیندهای تحولات استوار هستند (رجب پور، تقوا، حسین زاده، بابااحمدی، ۱۳۹۳). روش نزدیک‌ترین همسایه نزدیک یک روش یادگیری موردی است و از جمله ساده‌ترین الگوریتم‌های یادگیری ماشین می‌باشد که به روش K همسایه نزدیک نیز معروف است. در این الگوریتم یک نمونه با رای اکثریت از همسایه‌هایش دسته‌بندی می‌شود و این نمونه در عمومی‌ترین کلاس مابین k همسایه نزدیک تعیین می‌شود. Kیک مقدار مثبت صحیح و عموماً کوچک است. اگر k برابر یک باشد نمونه به سادگی در کلاس همسایگان نزدیکش تعیین می‌شود. فرد بودن مقدارk ، مفید است چون با این کار جلوی آراء برابر گرفته می‌شود. در این روش داده‌های مشابه در نزدیکی یکدیگر قرار گرفته، از این رو فاصله بین داده‌ها بر اساس عدم مشابهت آنها اندازه‌گیری و تفسیر می‌شود. بر این اساس داده‌هایی که در کنار یکدیگر قرار می‌گیرند، همسایه نامیده شده و هر داده جدیدی که به الگوریتم معرفی شود، فاصله آن با دیگر

داده‌ها محاسبه و در دسته‌ای قرار می‌گیرد که در نزدیکترین فاصله قرار دارد (هوانگ، یانگ، چوانگ، ۲۰۰۸). مسأله رگرسیون، نوعی مسأله یادگیری نظارتی است که این مدل را برای پیش‌بینی یک عدد درخواست می‌کند. ساده‌ترین و سریع‌ترین الگوریتم رگرسیون خطی است. لازم به توضیح است که بین الگوریتم‌های یادگیری ماشین رگرسیون و طبقه‌بندی اشتراکاتی وجود دارد (موصلی، جعفر، حسینی، ۲۰۱۲). در الگوریتم‌های اعتبارسنجی ضربدری داده‌ها به Kزیرمجموعه افراز می‌شوند. از این K زیرمجموعه، هر بار یکی برای اعتبارسنجی و ۱ -Kتای دیگر برای آموزش بکار می‌روند. این روال K بار تکرار می‌شود و همه داده‌ها به درستی یکبار برای آموزش و یکبار برای اعتبارسنجی، بکار می‌روند. در نهایت میانگین نتیجه این K بار اعتبارسنجی به عنوان یک تخمین نهایی برگزیده می‌شود. البته می‌توان از روش‌های دیگر برای ترکیب نتایج استفاده کرد. بطور معمول از 10-Fold استفاده می‌شود. در روش K-Fold طبقه‌ای سعی می‌شود نسبت داده‌های هر کلاس در هر زیرمجموعه و در مجموعه اصلی یکسان باشد.

مقاله پیش‌رو با تمرکز بر روش‌های مبتنی بر یادگیری ماشین در صدد بکارگیری این روش در برآورد قیمت سهام و مقایسه آن با سایر روش‌ها است. در صورت تحقق نتایج، امکان دستیابی به یک مدل کاراتر برای برآورد قیمت سهام امکان‌پذیر خواهد بود. از این رو به منظور ارزیابی و مقایسه دقت روش‌های موصوف در پیش‌بینی قیمت سهام، با استفاده از داده‌های روزانه قیمت سهام در بورس اوراق بهادار طی سال‌های ۱۳۸۸-۱۳۹۷ دقت هر یک از الگوریتم‌های موصوف را در پیش‌بینی قیمت سهام محاسبه خواهیم نمود و سپس مدلی کاراتر در پیش‌بینی قیمت سهام از بین مدل‌های مورد بررسی معرفی خواهد شد.

در طول چند سال اخیر رشد و توسعه بازار سرمایه کشور و معرفی ابزارها، سازوکار و پدیده‌های نو در آن، اهمیت بازار سرمایه را در اقتصاد کشور ارتقاء داده است. پیش‌بینی قیمت سهام یکی از مسائل مهم در بازارهای مالی است که توجه بسیاری از پژوهشگران دانشگاهی و کارشناسان این حوزه را در چند دهه گذشته به خود جلب کرده است. اهمیت این موضوع از آنجا ناشی می‌شود که پیش‌بینی قیمت سهام در بازارهای مالی یکی از متغیرهای مهم در زمینه تصمیم‌های سرمایه‌گذاری، قیمت‌گذاری اوراق بهادار (مشتقه‌ها) و مدیریت ریسک است. از آنجا که سرمایه‌گذاران بازارهای بورس همواره علاقه‌مندند از روند بعدی قیمت‌ها مطلع شوند، فعالان این بازار درصدد دستیابی و به‌کارگیری روش‌هایی هستند تا بتوانند با پیش‌بینی آتی قیمت

سهام، سود سرمایه خود را افزایش دهند. بنابراین، ضروری به نظر می‌رسد که روش‌های مناسب، صحیح و متکی به اصول علمی در تعیین قیمت آینده سهام فرآروی افراد سرمایه‌گذار قرار گیرد.

در دهه‌های اخیر روش‌های جدیدی از پیش‌بینی به نام روش‌های داده‌کاوی شامل روش‌های درخت تصمیم‌گیری، ماشین‌های بردار پشتیبان، شبکه‌های بیزین، شبکه‌های عصبی مصنوعی خوشه‌بندی و ... پا به عرصه وجود نهاده‌اند. استفاده موفقیت‌آمیز این روش‌ها در پیش‌بینی‌های اقتصادی، سمت و سویی تازه به پژوهش‌های این‌چنینی داده است. عده‌ای بر این عقیده هستند که این روش‌ها نتایج بهتری را نسبت به روش‌های ریاضی و آماری مانند رگرسیون نشان میدهد، چون بر اساس داده‌های تاریخی که به آنها داده می‌شود، می‌توانند الگوها و روندها را بدون فرمول یا روش خاصی بیاموزند و روابط غیرخطی بین ورودی‌ها و خروجی‌ها را تشخیص و روابط بنیادین بین آنها را شناسایی و از این طریق پیش‌بینی دقیقتری را ارائه کنند (حسینی‌نسب، کریمی تکلو، یوسفی‌نژاد، ۱۳۹۲).

تاکنون در بیشتر پژوهش‌ها از رگرسیون و تکنیک شبکه‌های عصبی مصنوعی و در برخی موارد ماشین‌های بردار پشتیبان برای پیش‌بینی‌های مالی استفاده شده است، بنابراین برای رسیدن به یک پیش‌بینی دقیق‌تر لازم است که روش‌های دیگر داده‌کاوی و یادگیری ماشین نیز مورد بررسی قرار گیرد.

روش شناسی

پژوهش حاضر را می‌توان از جمله پژوهش‌های کاربردی، تحلیلی –ریاضی طبقه‌بندی کرد که قلمرو مکانی آن شرکت های پذیرفته شده در تابلوی اصلی بورس اوراق بهادار تهران و قلمرو زمانی آن از سال ۱۳۸۸ تا ۱۳۹۷ است. جامعه آماری این پژوهش در برگیرنده کلیه شرکت‌های پذیرفته شده در تابلوی اصلی بورس اوراق بهادار تهران است که در سال‌های ۱۳۸۸ تا ۱۳۹۷ در بورس حضور داشته‌اند. به علت گستردگی حجم جامعه آماری و وجود برخی ناهماهنگی‌ها میان اعضاء جامعه، برای گزینش نمونه از روش حذف سیستماتیک استفاده شد و معیارهای اعمال شده برای به‌دست آوردن نمونه به شرح زیر است (امامیان، محمودی راد، ملاعلیزاده زواردهی، نیرومند، ۱۴۰۱):

- سال مالی شرکت منتهی به پایان اسفندماه هر سال باشد (به لحاظ افزایش قابلیت مقایسه).
- شرکت در سالهای ۱۳۸۸ تا ۱۳۹۷ تغییر سال مالی نداده باشد.

- شرکت قبل از سال ۱۳۸۸ در بورس اوراق بهادار تهران پذیرفته شده باشد.
- در طول سال مالی دارای وقفه معاملاتی بیشتر از ۶ ماه نباشند.
- بیش از ششصد هزار سهم آنها در هر سال معامله شده باشد.

با اعمال معیارهای بالا، تعداد ۱۲ شرکت فعال در تابلوی اصلی بورس به عنوان نمونه آماری انتخاب شد.

در این پژوهش مدل پیش‌بینی الگوریتم‌های منتخب به شرح زیر هستند:

الگوریتم شبکه‌های عصبی

در سال‌های اخیر پیش‌بینی و مدل‌سازی به وسیله شبکه‌های عصبی رشد یافته است. یک شبکه‌های عصبی مصنوعی، مجموعه ای از نرونهای به هم متصل در لایه‌های مختلف (چند لایه،MLP) است که اطلاعاتی را برای یکدیگر ارسال می‌کنند. ساده ترین شکل شبکه تنها دو لایه دارد، لایه ورودی و لایه خروجی. ارتباط میان یک ورودی و خروجی به وسیله یک وزن که بیانگر اهمیت نسبی ورودی مذکور در محاسبه ارزش خروجی است، مشخص می‌شود. به این ترتیب، ارزش نرون خروجی Net از ورودی مشاهده tام به وسیله رابطه زیر به دست می‌آید:

$$Net_t = a_0x_0 + a_1x_{1t} + a_2x_{2t} = \sum_{i=0}^{2} a_ix_{it}$$

سپس نرون خروجی، ارزش به دست آمده را با استفاده از یک تابع تبدیل یا فعال سازی (محرک) که با f نشان داده می‌شود- پردازش می کند (مکیان، موسوی، ۱۳۹۰). ساده ترین شکل شبکه‌های عصبی، تابع فعال سازی خطی است. ارزش به دست آمده از رابطه بالا با یک تابع فعال سازی خطی، خروجی نهایی شبکه برای مشاهده t را به صورت زیر می‌سازد.

$$y_t = f(Net_t = a_0x_{0T} + a_1x_{1t} + a_2x_{2t}) = a_0x_{0t} + a_1x_{1t} + a_2x_{2t}$$

تاحدودی تمام شبکه‌های عصبی در بخش هایی از شبکه از توابع فعال سازی غیرخطی استفاده می‌کنند. این مسئله اجازه می‌دهد که شبکه الگوهای غیرخطی مناسبی از مجموعه داده‌های پیچیده تولید کند (زارعی، محمدیان، حاضری نیری، باشکوه اجیرلو، ۱۳۹۷). رایج‌ترین تابع فعال‌سازی مورد استفاده در ادبیات شبکه‌های عصبی تابع سیگموئیدی است:

$$f(x) = \frac{1}{1 + e^{-x}}$$

در این پژوهش، برای پیش‌بینی از طریق شبکه‌های عصبی، تابع فعال‌سازی لایه میانی و لایه خروجی به ترتیب سیگموئید و خطی در نظر گرفته شده است.

از برتری‌های شبکه‌های عصبی می‌توان به قابلیت آن در حل مسائل غیرخطی اشاره کرد. هم چنین شبکه‌های عصبی ابزار مناسبی برای مواردی است که در آنها جواب‌ها مهم تر از درک روابط علت و معلولی است. اما شبکه‌های عصبی قادر به ارائه معادله‌ای با ضرایب معین نیست، ولی رگرسیون از جهت برآورد ضرائب متغیرهای تاثیرگذار بر هدف ما برتری بهتری دارد.

مهم‌ترین محک برای انتخاب یک روش پیش‌بینی، دقت یا به عبارتی نزدیکی مقدار پیش‌بینی به مقدار واقعی است. متداول‌ترین معیار در زمینه پیش‌بینی RMSE (جذر میانگین مجذور خطا) است. در این پژوهش به منظور گزینش بهترین ساختار شبکه‌های عصبی از معیار RMSE و هم چنین برای ارزیابی و مقایسه نتایج شبکه‌های عصبی از معیارهای MAD (میانگین قدرمطلق انحراف) و MAPE (میانگین قدر مطلق درصد خطا) در کنار RMSE استفاده می‌شود (لانگ و همکاران، ۲۰۱۹)

رگرسیون لجستیک

رگرسیون لجستیک نوعی از رگرسیون است که متغیرهای پیش بین (مستقل) می‌تواند هم در مقیاس کمی و هم در مقیاس مقوله‌ای باشد، ولی متغیر وابسته مقوله ای دو سطحی است. این دو مقوله به گونه ای معمول به عضویت یا عدم عضویت در یک گروه اشاره دارد.

این مدل رگرسیون، شبیه رگرسیون معمولی است با این تفاوت که روش تخمین ضرایب یکسان نمی باشد و به جای حداقل کردن مجذور خطاها (که در رگرسیون معمولی انجام می‌شود، احتمالی که یک واقعه رخ می‌دهد را بیشتر می‌کند. در رگرسیون لجستیک از مفهومی به نام بخت برای مقدار متغیر وابسته استفاده می‌شود در اصطلاح آماری بخت به معنی احتمال رخداد یک پیشامد (P_i) بر احتمال عدم رخداد ($I - P_i$) آن است. احتمال بین ۰ و ۱ تغییر می‌کند. در حالی که بخت ممکن است بیش از یک باشد. واژه کلیدی در تحلیل رگرسیون لجستیک سازه ای به نام لوجیت که لگاریتم طبیعی بخت است.

رگرسیون لجستیک به صورت زیر تعریف می‌شود:

$$Z_i = In\left[\frac{P_i}{1-P_i}\right] = \beta o + \sum \beta i Xi$$

در معادله پیشگفته In بیانگر لگاریتم طبیعی است. در مدل رگرسیون لجستیک احتمال رخداد پیشامد بر اساس رابطه زیر محاسبه می‌شود:

$$P_i = IIi\left(x1،x2،\dots،xk\right) = e^{\beta o} + \frac{\sum \beta iXi}{1 + e^{\beta o + \sum \beta iXi}}$$

یکی از منافع رگرسیون لجستیک بی نیازی آن به فرضیه‌های محدود کننده آماری در رابطه با متغیرهاست. مسأله اصلی که در این مدل وجود دارد، این است که تغییرات یکسان وضعیت موجود همیشه تغییرات یکسانی را در احتمال به دست آمده به دنبال ندارد. این موضوع زمانی که احتمال نزدیک مقادیر ۱ یا ۰/۵ باشد، اهمیت بیشتری پیدا می کند (زارعی و همکاران، ۱۳۹۷).

الگوریتم نزدیک ترین همسایه های K

الگوریتم نزدیکترین همسایه های K الگوریتمی است که در آن بر اساس الگوهای تعلیم، داده‌های تست در آن طبقه‌بندی می‌شوند. در این روش داده‌های مشابه در نزدیکی یکدیگر قرار گرفته، از این رو فاصله بین داده‌ها بر اساس عدم مشابهت آنها اندازه گیری و تفسیر می‌شود. بر این اساس داده‌هایی که در کنار یکدیگر قرار می گیرند، همسایه نامیده شده و هر داده جدیدی که به الگوریتم معرفی شود، فاصله آن با دیگر داده‌ها محاسبه و در دسته‌ای قرار می گیرد که در نزدیکترین فاصله قرار دارد (اوبها، نامبی، ۲۰۱۲). K به تعداد همسایگان گفته می‌شود که در فرایند دسته‌بندی نقش مهمی ایفا می کند. تعداد صحیح این متغیر بستگی به اندازه داده‌ها به عنوان یکی از پارامترهای مهم در این فرایند دارد. یکی از معمول‌ترین روش‌ها برای انتخاب صحیح، استفاده از الگوریتم تپه نوردی است (یاکوپ، ملک، مرکان، ۲۰۱۱). در این روش میزان k را مرتبا افزایش داده تا به بهترین انتخاب برسند. معیار بهبود در این روش نیز معیارهای فاصله هستند که عبارتند از یکی از موارد، فاصله اقلیدسی، مجوع قدر مطلق فاصله ها یا فاصله همینگ که مورد آخر برای داده‌های باینری (دودویی) مناسب است. به طور خلاصه روند پیش‌بینی با الگوریتم نزدیک ترین همسایه‌های K شامل مراحل زیر است:

الف: مشخص نمودن تعداد k اولیه در نرم افزار؛

ب: محاسبه فاصله بین نمونه مورد بررسی با تمام داده‌های ورودی؛

ج: مرتب سازی فاصله ها بر اساس کمترین آن ها و مشخص کردن نزدیکترین‌ها بر حسب k؛

د: استفاده از اکثریت نزدیکترین همسایه ها جهت پیش‌بینی مقادیر جدید.

ماشین‌های بردار پشتیبان

فیشر اولین الگوریتم را برای طبقه‌بندی و دسته‌بندی الگوها در سال ۱۹۳۶ ارائه کرد و از طریق کم کردن خطای طبقه‌بندی داده‌های آموزشی، برای بهینه کردن الگوها به کار گرفت. در سال ۱۹۶۵، پژوهشگری روسی به نام ولادیمیر و پنیک، گامی بسیار مهم در طراحی طبقه‌بندی کننده ها برداشت. وی نظریهٔ آماری یادگیری را به صورت مستحکم‌تری بنا نهاد و ماشین‌های بردار پشتیبان را بر این اساس ارائه داد (تسانگ و همکاران، ۲۰۰۷).

در واقع ریشه SVMها در تئوری یادگیری آماری است و کاربردهای فراوانی در رگرسیون، طبقه‌بندی، خوشه بندی و به طور کلی تقریب توابع دارند. رویکرد SVM در بدو ابداع، تنها طبقه‌بندی دو کلاسه را شامل می‌شد. این رویکرد در ادامه با استفاده از انواع تکنیک های ترکیب به طبقه‌بندی چند کلاسه نیز تعمیم یافت (یائو، یو، ژانگ، چن، ۲۰۱۱).

SVMها داده‌ها را به صورت بردار پردازش می‌کنند و همواره در میان تمام ابر صفحاتی که داده‌ها را تفکیک می‌کنند، آن ابر صفحه را بر می گزینند که بیشترین تفکیک پذیری با بیشترین اندازه حاشیه را میان داده‌های کلاس‌های مختلف حاصل کنند. ابر صفحه مطلوب طوری انتخاب می‌شود که فاصله آن از نزدیک ترین داده بیشینه شود. به نزدیک ترین نقاط آموزشی با بیشترین حاشیهٔ ابر صفحه، بردارهای پشتیبان گفته می‌شود. چنین ابر صفحه‌ای در صورت وجود، ابر صفحه با حاشیه بیشینه خوانده می‌شود.

اگر نقاط آموزشی را به صورت $\left[x_i، y_i\right]$، بردار ورودی را $x_i \in Rn$ و ارزش طبقه را $y_i \in \left\{-1،1\right\}$ ، $i = 1، \ldots .، l$ تعریف کنیم، آنگاه در حالتی که داده‌ها به صورت خطی قابل تفکیک هستند، قواعد تصمیم گیری تعریف می‌شود و با یک صفحه بهینه که طبقات تصمیم‌گیری باینری را تفکیک می‌کند، به صورت رابطه زیر خواهد بود.

$$y = sign\left(\sum_{i=1}^{N} y_i a_i (X.X) + b\right)$$

که در آن y: خروجی معادله، y_i: ارزش طبقه نمونه آموزشی و X_i نشان دهنده ضریب داخلی است. بردار $x = \left(x_1، x_2، \ldots، x_n\right)$ نشان دهنده یک داده ورودی و بردارهای $X_i : i = 1، \ldots، N\ Xi$ بردارهای پشتیبان هستند. در رابطه ۱ پارامترهای b، a_iتعیین کننده ابر صفحه هستند.

اگر داده‌ها به صورت خطی قابل تفکیک نباشند، رابطه بالا به صورت رابطه زیر تغییر می‌یابد.

$$Y = sign\left(\sum_{i=1}^{N} y_i a_i K(X.X_i) + b\right)$$

تابع $K(X.X_i)$ تابع کرنلی است که برای ایجاد ماشین‌هایی با انواع مختلفی از سطوح تصمیم‌گیری غیر خطی در فضای داده‌ها، ضربهای داخلی تولید می‌کند (بیات و باقری، ۱۳۹۶). برای مثال سه نوع تابع کرنل که در مدل SVM به کار می‌روند عبارتند از: تابع کرنل چند جمله ای، تابع کرنل پایه شعاعی و تابع کرنل پرسپترون چندلایه. قابلیت به کارگیری کرنلها یکی از نقاط قوت کلیدی SVMها محسوب شده، سبب می‌شود که آنها انعطاف بالایی برای حل انواع مختلف مسائل با دشواری های گوناگون داشته باشند.

برخی ویژگی های ماشین‌های بردار پشتیبان عبارتند از: آموزش ماشین‌های بردار پشتیبان به آسانی صورت می‌گیرد؛ با نمونه های آموزشی کم، نتایج خوبی را ایجاد می‌کند؛ برخلاف شبکه‌های عصبی نیاز به بهینه یابی محلی ندارد و بهترین مدل را پیدا می‌کند. در مقابل این برتری‌ها نیز باید در انتخاب تابع هسته ای مناسب دقت کرد (لانگ و همکاران، ۲۰۱۹).

الگوریتم اعتبار سنجی ضربدری(k-f)

در این نوع اعتبارسنجی، داده‌ها به K زیرمجموعه افراز می‌شوند. از این K زیرمجموعه، هر بار یکی برای اعتبارسنجی و K-1 تای دیگر برای آموزش بکار میروند. این روال K بار تکرار می‌شود و همه داده‌ها دقیقا یکبار برای آموزش و یکبار برای اعتبارسنجی بکار می‌روند. در نهایت میانگین نتیجه این K بار اعتبارسنجی به عنوان یک تخمین نهایی برگزیده می‌شود. البته می‌توان از روش های دیگر برای ترکیب نتایج استفاده کرد. بطور معمول از ۱۰-Fold استفاده می‌شود.

در روش K-Fold طبقه ای سعی می‌شود نسبت داده‌های هر کلاس در هر زیرمجموعه و در مجموعه اصلی یکسان باشد.

مساحت زیر منحنی مشخصه عملیاتی گیرنده AUC به عنوان یک معیار عملکرد طبقه‌بندی‌کننده مناسب و دقیق دیده می‌شود. AUC می‌تواند از ۰٫۵ تا ۱ ارزش داشته باشد. یک مقدار ۰٫۵ به این معنی است که پیش‌بینی‌ها دقیق نیستند، در حالی که مقدار ۱ به این معنی است که پیش‌بینی‌ها عالی هستند.

مزیت AUC بر دیگر معیارهای عملکرد (مانند درصد به طور کامل طبقه‌بندی نشده این است که AUC همه مقادیر قطع‌شده را شامل می‌شود به صورت زیر تعریف می‌شود:

$$AUC = \int_0^1 \frac{TP}{(TP+FN)} d\frac{FP}{(FP+TN)} = \int_0^1 \frac{TP}{P} d\frac{FP}{N}$$

از آنجا که هدف از انجام این طرح، مقایسه روش‌های مختلف است بدین منظور میانگین خطای مطلق، خطای میانگین مربعات و خطای جذر میانگین مربعات مربوط به هر روش را برای انجام مقایسه به ترتیب از طریق روابط زیر به دست می‌آوریم و در نهایت از میانگین ساده و وزن‌دار برای ترکیب نتایج روش‌ها به منظور بهتر شدن نتایج استفاده می‌کنیم که به ترتیب از طریق روابط زیر به دست می‌آید (اوبها، نامبی، ۲۰۱۲).

$$MAE = \sum \frac{|Y_i - \hat{Y}_i|}{n}$$

MAE= میانگین خطای مطلق

Y= مقدار واقعی شاخص روز i ام

Ŷ= مقدار پیش‌بینی شده شاخص برای روز i ام

N= تعداد کل دادگان تست

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

MSE: خطای میانگین مربعات

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

RMSE: خطای جذر میانگین مربعات

fi(ymlp،ysvr،yk-fold،yknn،ylog=

$$\frac{((ymlp.wmlp)+(ysvr.wsvr)+(yk-fold.wk-fold)+(yknn.w.knn)+(ylog.wlog))}{((ymlp)+(ysvr)+(yk-fold)+(yknn.)+(ylog))}$$

fi= خروجی حالص از ترکیب روش‌ها

y= خرجی حاصل از روش x

W= درصد صحت روش x

نتایج

در این پژوهش پس از بررسی های فراوان برای شناسایی متغیرها، ۷ متغیر تاثیرگذار بر قیمت سهام در دو گروه به شرح زیر دسته‌بندی شده‌اند:

متغیرهای مستقل در دو گروه کلی به شرح زیر دسته‌بندی شدند.

گروه اول: متغیرهای فنی شامل بالاترین قیمت سهم، پایین ترین قیمت سهم و حجم معاملات سهم (به میلیون ریال –روزانه) و شاخص قیمت سهام

گروه دوم: متغیرهای اقتصادی شامل نرخ ارز، قیمت هر اونس طلا، قیمت نفت در سبد اوپک

داده‌های این پژوهش از نرم افزار ره آورد نوین به صورت روزانه و برای ۱۲ شرکت از سال ۱۳۸۸ تا ۱۳۹۷ استخراج شده است. ورودیهای هر کدام از شرکتها به طور متوسط شامل ۱۰۰۰ داده هستند که ۸۰ درصد از آنها به عنوان داده‌های آموزش و مابقی داده‌ها به عنوان داده‌های تست در نظر گرفته شده اند (سهرابی، میربرگ کار، چیرانی، خردیار، ۱۴۰۱)

ابتدا پس از انتخاب متغیرها و جمع آوری داده‌ها، به منظور آماده‌سازی داده‌ها برای آموزش وآزمایش ابتدا هر کدام از متغیرها با استفاده از رابطه زیر نرمال شد تا تاثیر اعداد بزرگ کاهش یابد.

$$\tilde{S}_i = \frac{(S_i - S_{\min})}{S_{\max} - S_{\min}}, \quad i = 1,\cdots,9$$

لازم است از بین داده‌ها، توسط روش k-FoldCross-Validation، داده‌های آموزشی و داده‌های آزمایشی تفکیک شدند. بنابراین ۸۰ درصد داده‌ها به عنوان داده‌های آموزشی و ۲۰درصد مابقی به عنوان داده‌های آزمایشی استفاده می‌شود. در این روش مجموعه داده‌ها هر شرکت به K قسمت مساوی، به صورت تصادفی تقسیم می‌شود. در واقع داده‌ها به K مجموعه از متغیرهای مستقل و متغیر وابسته تفکیک می‌شود. در اجرای اول، بخش اول از K قسمت به منظور آزمایش و مابقی داده‌ها برای یادگیری استفاده می‌شود که معمولا ۲۰٪ به عنوان داده‌های آزمایش و ۸۰٪ آن به عنوان داده‌های یادگیری استفاده می‌شود. در اجرای دوم، بخش دوم از

K قسمت به منظور آزمایش و مابقی داده‌ها برای یادگیری استفاده می‌شود. الگوریتم به اندازه K مرتبه تکرار می‌شود. در این روش از تمامی داده‌ها برای یادگیری مدل استفاده می‌شود و هر بار بخشی از داده‌ها به عنوان مجموعه آزمایش توانایی مدل را ارزیابی می‌کند. در پژوهش حاضر مقدار k عدد ۱۲ انتخاب شد.

برای سنجش انطباق یک پیش‌بینی با یک الگوی داده سری زمانی، از معیارهای سنجش خطای پیش‌بینی استفاده خواهد شد، اگر y و $\hat{y}$ نشانگر مقدار واقعی و پیش‌بینی شده متغیر در زمان t باشد خطای پیش‌بینی عبارتست از $e=y-\hat{y}$ بنابراین برای یک دوره زمانی و برای n مقدار پیش‌بینی شده، معیارهای سنجش پیش‌بینی عبارتند از:

جدول ۱. پارامترهای ارزیابی

شماره رابطه	رابطه	معیار
۱	$MSE=\frac{\sum_{i=1}^{n}(e_i)^2}{n}$	MSE (میانگین مجذور خطای پیش‌بینی)
۲	$\sqrt{\frac{1}{N}*\sum_{i=1}^{N}(f_i-y_i)^2}$	RMSE
۳	$R^2=1-\frac{SSE}{SSy}$	R2 (ضریب تعیین)

که در آن $SSE=\sum(y-\hat{y})^2$ و $SSy=\sum(y-\bar{y})^2$ می باشد و y مقدار واقعی، $\hat{y}$ مقدار پیش‌بینی شده توسط مدل و $\bar{y}$ میانگین مقادیر واقعی y است. کلیه موارد بیان شده در بالا برای داده‌های روزانه، هفتگی و ماهیانه محاسبه شده است. بنابراین برای تشخیص بهترین مدل از دیدگاه دقت بین معیارهای MSE،RMSE،R2 ملاک برای مقایسه دقت مدل‌ها معیار MSE است.

آزمایش

در این پژوهش برای ارزیابی روش پیشنهادی از داده‌های سهام ۱۲ شرکت نوشته شده در جدول ۲ ، استفاده شده است.

جدول ۲. سهام‌های بکار گرفته شده در این پژوهش

شرکت	سهام	تعداد داده روزانه
البرز دارو	ولبر	۲۰۳۳
کشتیرانی	حکشتی	۲۰۱۴
بانک ملت	وبملت	۱۸۷۲
گروه صنعتی سدید	وسدید	۲۰۰۸
نفت و گاز و پتروشیمی تامین	تاپیکو	۱۹۴۵
خدمات انفورماتیک	رانفور	۱۸۰۳
پتروشیمی پارس	پارس	۱۸۲۰
فنرسازی خاور	خفنر	۱۵۲۸
پتروشیمی جم	جم	۱۴۹۱
فولاد خوزستان	فخوز	۱۴۸۰
داروسازی کوثر	دکوثر	۱۸۰۳
داروسازی عبید	دعبید	۱۸۲۰

در ادامه برای ارزیابی نتایج روش پیشنهادی از پنج الگوریتم تخمین بهره گرفته شده است. الگوریتم‌های تخمین بکار گرفته شده در این پژوهش به شرح زیر است:

- نزدیک‌ترین همسایه
- لاجستیک
- درخت تصمیم
- شبکه عصبی پرسپترون چند لایه

❖ ماشین بردار پشتیبان

نتایج آزمایش‌ها

در این بخش نتایج پیش‌بینی قیمت سهم هر شرکت بر اساس معیارهای آماری بیان شده است که در ادامه نتایج این آزمایش‌ها آورده شده است.

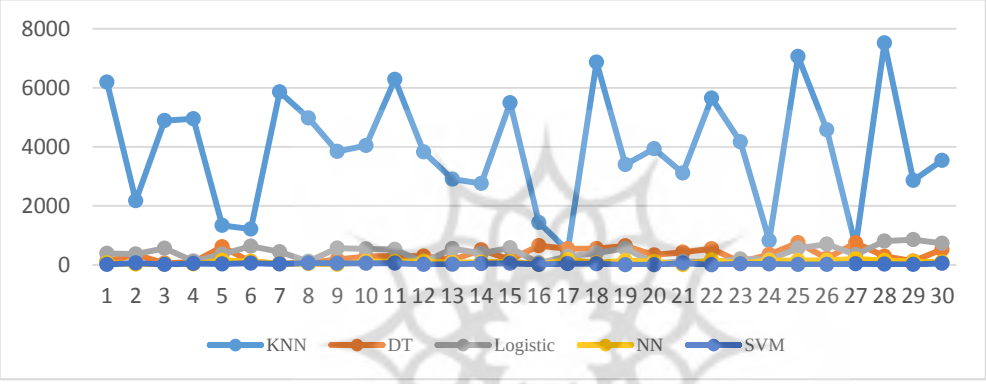
جدول ۳. نتایج آماری پیش‌بینی قیمت سهام

نماد	الگوریتم	MSE	RMSE	R2
والبر	نزدیک‌ترین همسایه	۹۹۰۱۲۵	۹۹۵	۸۴/۰
	درخت تصمیم	۱۵۲۸۹۶	۳۹۱	۹۳/۰
	لاجستیک	۶۵۶۷۸۰	۸۱۰	۸۹/۰
	شبکه عصبی پرسپترون چند لایه	۱۵۰۱۲	۱۲۲	۹۷/۰
	ماشین بردار پشتیبان	۱۹۲۲	۴۳	۹۹/۰
حکشتی	نزدیک‌ترین همسایه	۳۸۳۱۰۷۶	۱۹۵۷	۹۳/۰
	درخت تصمیم	۴۵۸۱۷۵	۶۷۶	۹۵/۰
	لاجستیک	۴۹۰۹۹۳	۷۰۰	۹۴ /۰
	شبکه عصبی پرسپترون چند لایه	۲۶۵۹۴	۱۶۳	۹۸/۰
	ماشین بردار پشتیبان	۴۳۱۱	۶۵	۹۹/۰
وبملت	نزدیک‌ترین همسایه	۷۴۵۶۳۱	۸۶۳	۹۱/۰
	درخت تصمیم	۸۰۰۶۵	۲۸۲	۹۴/۰
	لاجستیک	۶۲۹۸۴	۲۵۰	۹۶/۰
	شبکه عصبی پرسپترون چند لایه	۵۶۸۰	۷۵	۹۸/۰
	ماشین بردار پشتیبان	۶۶۸	۲۵	۹۹/۰
دعبید	نزدیک‌ترین همسایه	۸۲۴۳۲۰۸	۲۸۷۱	۸۹/۰
	درخت تصمیم	۲۴۷۴۵۵۷	۱۵۷۳	۹۳/۰
	لاجستیک	۱۷۶۹۸۳۳	۱۳۳۰	۹۴/۰
	شبکه عصبی پرسپترون چند لایه	۱۶۲۶۶	۱۲۷	۹۹/۰
	ماشین بردار پشتیبان	۱۸۹۹۲۷	۱۳۷	۹۸/۰
دکوثر	نزدیک‌ترین همسایه	۹۳۲۲۲۶	۹۶۵	۸۷/۰
	درخت تصمیم	۱۸۹۲۴۱	۴۳۵	۹۱/۰
	لاجستیک	۱۶۸۰۲۶	۴۰۹	۹۳/۰
	شبکه عصبی پرسپترون چند لایه	۱۴۰۲	۳۷	۹۶/۰
	ماشین بردار پشتیبان	۹۸۵	۳۱	۹۸/۰

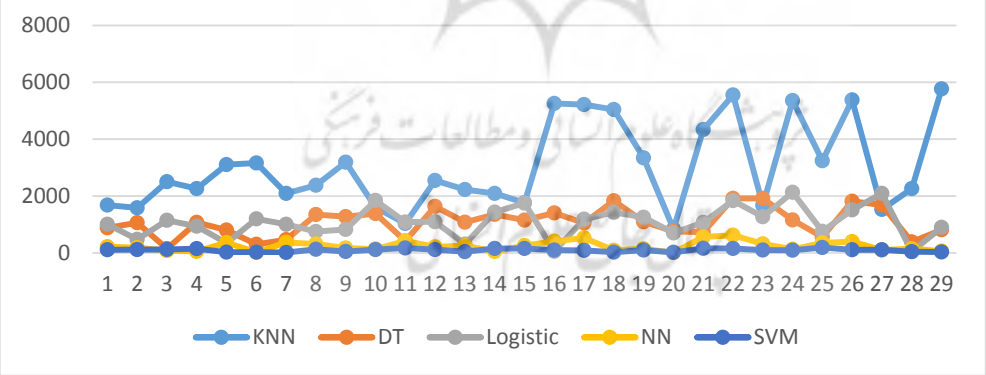
نماد	الگوریتم	MSE	RMSE	R2
فخوز	نزدیک‌ترین همسایه	۳۰۲۶۲۹	۵۵۰	۹۰/۰
	درخت تصمیم	۸۱۰۵۰	۲۸۵	۹۲/۰
	لاجستیک	۴۴۱۳۸	۲۱۰	۹۴/۰
	شبکه عصبی پرسپترون چند لایه	۲۹۰	۱۷	۹۷/۰
	ماشین بردار پشتیبان	۲۵۸	۱۶	۹۸/۰
جم	نزدیک‌ترین همسایه	۳۷۳۷۵۷	۶۱۱	۸۹/۰
	درخت تصمیم	۲۰۰۰۱۷	۴۴۷	۹۳/۰
	لاجستیک	۲۴۳۱۱۱	۴۹۳	۹۴/۰
	شبکه عصبی پرسپترون چند لایه	۲۲۶۸	۴۸	۹۸/۰
	ماشین بردار پشتیبان	۸۹۳	۳۰	۹۹/۰
خفنر	نزدیک‌ترین همسایه	۱۷۹۴۲۵	۴۲۴	۸۵/۰
	درخت تصمیم	۱۰۸۰۳	۱۰۴	۸۹/۰
	لاجستیک	۱۶۵۰۵	۱۲۸	۹۱/۰
	شبکه عصبی پرسپترون چند لایه	۱۴۲۹	۳۸	۹۶/۰
	ماشین بردار پشتیبان	۴۳۷	۲۱	۹۸/۰
رانفور	نزدیک‌ترین همسایه	۸۰۰۸۱۷	۸۹۵	۸۹/۰
	درخت تصمیم	۱۷۳۳۲۰	۴۱۶	۹۰/۰
	لاجستیک	۱۲۱۰۶۲	۳۴۸	۹۲/۰
	شبکه عصبی پرسپترون چند لایه	۷۷۴۷	۸۸	۹۸/۰
	ماشین بردار پشتیبان	۱۰۹۷۸	۱۰۵	۹۶/۰
پارس	نزدیک‌ترین همسایه	۴۶۶۱۷	۲۱۶	۹۴/۰
	درخت تصمیم	۳۶۹۴	۶۱	۹۶/۰
	لاجستیک	۳۵۰۹	۵۹	۹۷/۰
	شبکه عصبی پرسپترون چند لایه	۲۶۹۷	۵۲	۹۸/۰
	ماشین بردار پشتیبان	۳۳۹	۱۸	۹۹/۰
تاپیکو	نزدیک‌ترین همسایه	۱۴۹۲۴۸	۳۸۶	۹۳/۰
	درخت تصمیم	۵۷۰۴۵	۲۳۹	۹۵/۰
	لاجستیک	۴۶۵۲	۶۸	۹۷/۰
	شبکه عصبی پرسپترون چند لایه	۵۴۷	۲۳	۹۹/۰
	ماشین بردار پشتیبان	۲۰۳	۱۴	۹۹/۰
وسدید	نزدیک‌ترین همسایه	۳۳۰۰۷	۱۸۲	۹۲/۰
	درخت تصمیم	۱۵۸۷۸	۱۲۶	۹۴/۰
	لاجستیک	۱۹۵۴۹	۱۴۰	۹۳/۰

نماد	الگوریتم	MSE	RMSE	R2
	شبکه عصبی پرسپترون چند لایه	۱۸۸۷	۴۳	۹۸/۰
	ماشین بردار پشتیبان	۴۳۲	۲۱	۹۹/۰

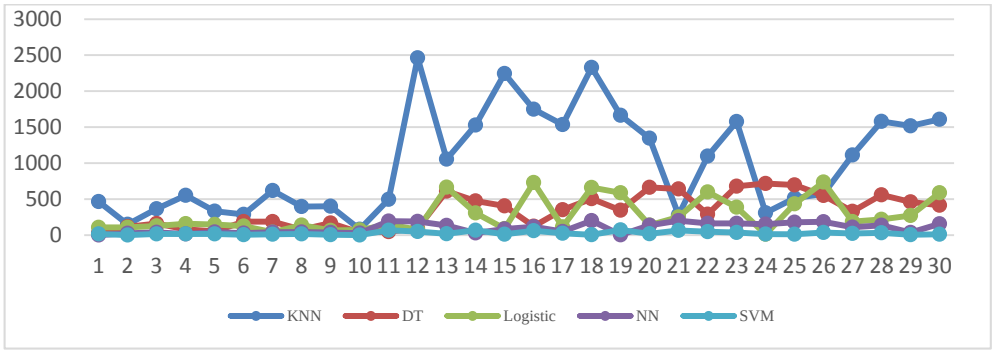
همانطور که در جدول ۳ نشان داده شده است، الگوریتم ماشین بردار پشتیبان از دیگر الگوریتم‌های تخمین از کارایی بالاتری برخوردار است. دلیل این امر ایجاد ابرصفحه مناسب برای پیش‌بینی مقادیر سهام می‌باشد. همچنین الگوریتم ماشین بردار پشتیبان از عملیات ریاضی قوی‌تری نسبت به دیگر الگوریتم‌های تخمین بهره می‌برد.



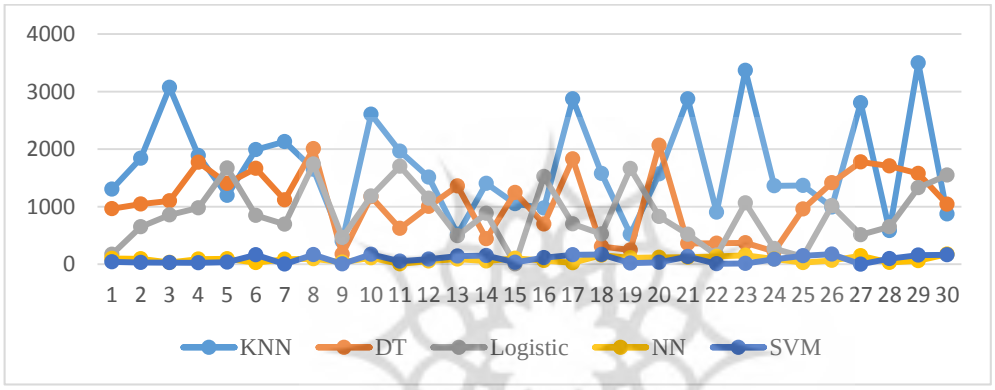
شکل ۱. اختلاف قیمت سهام برای سهام والبر



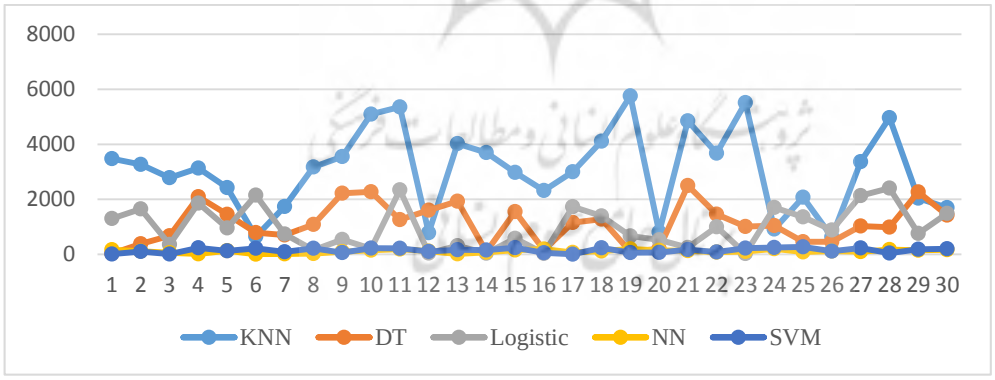
شکل ۲. اختلاف قیمت سهام برای سهام حکشتی



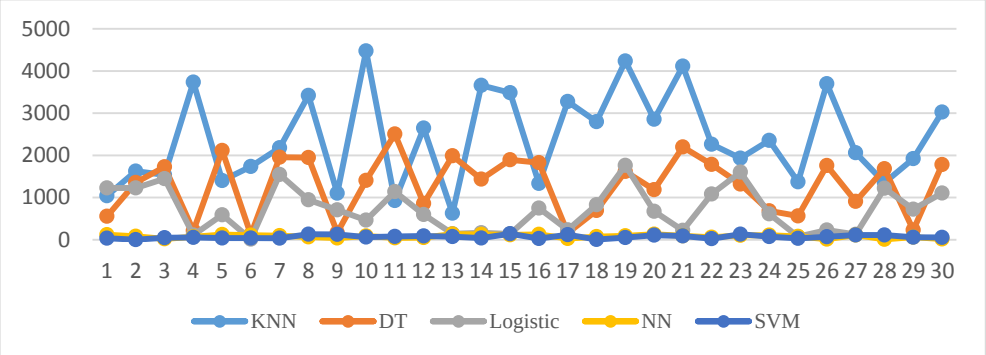
شکل ۳. اختلاف قیمت سهام برای سهام وبملت



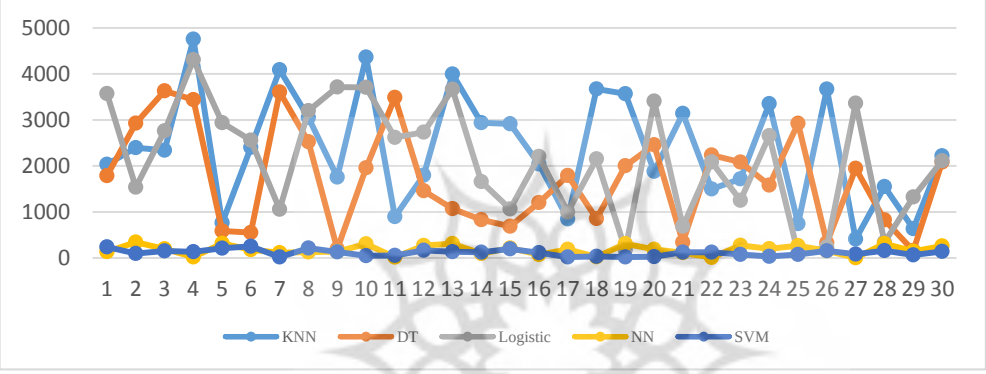
شکل ۴. اختلاف قیمت سهام برای سهام دعبید



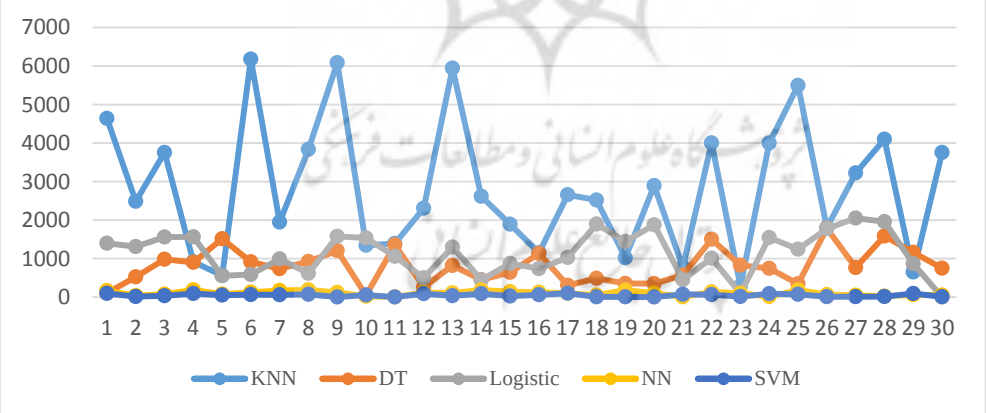
شکل ۵. اختلاف قیمت سهام برای سهام دکوثر



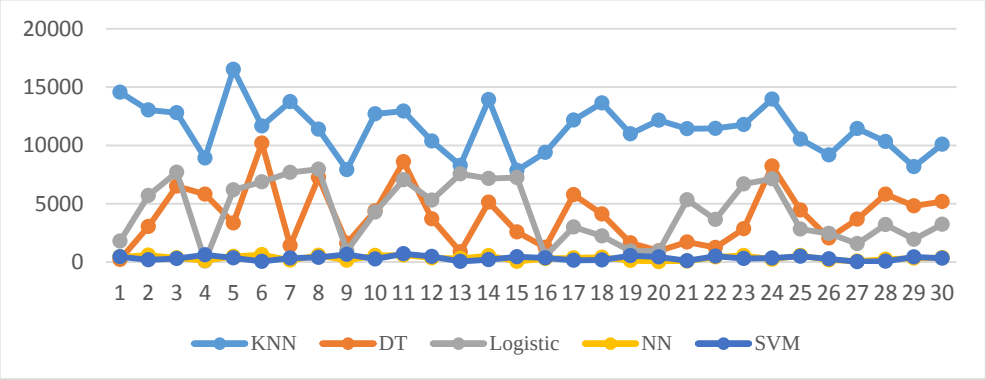
شکل ۶. اختلاف قیمت سهام برای سهام فخوز



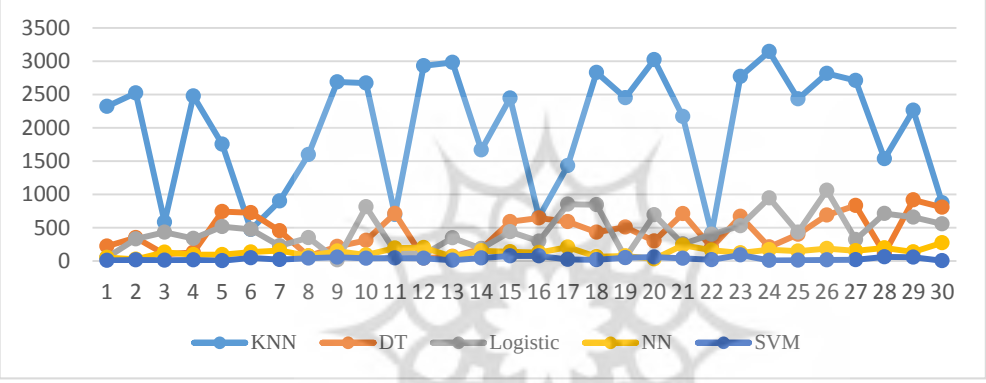
شکل ۷. اختلاف قیمت سهام برای سهام جم



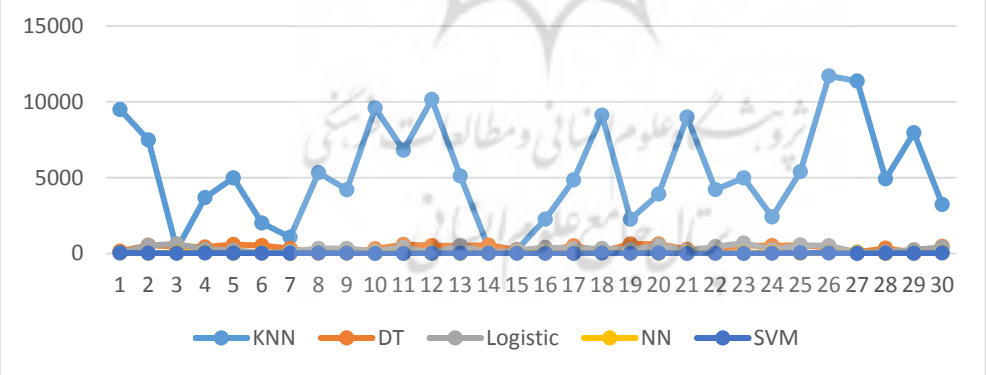
شکل ۸. اختلاف قیمت سهام برای سهام خفنر



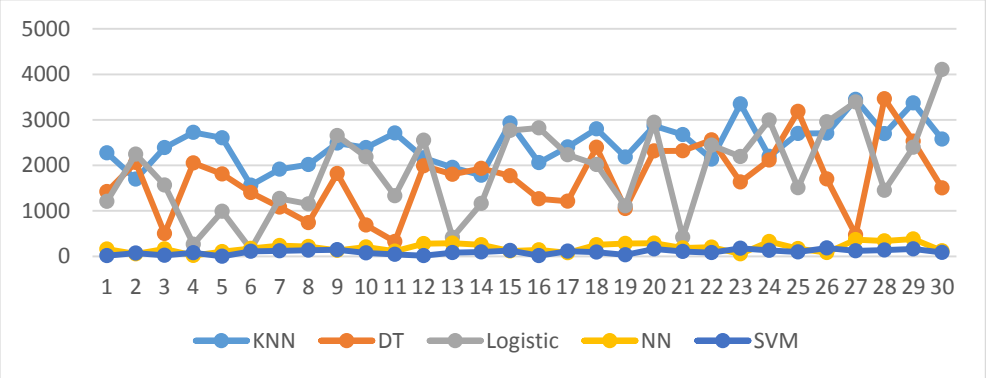
شکل ۹. اختلاف قیمت سهام برای سهام پارس



شکل ۱۰. اختلاف قیمت سهام برای سهام رانفور



شکل ۱۱. اختلاف قیمت سهام برای سهام رانفور



شکل ۱۲. اختلاف قیمت سهام برای سهام وسدید

نتایج حاصل از این آزمایش‌ها نشان‌دهنده‌ی کارایی قابل قبول الگوریتم تخمین ماشین بردار پشتیبان و شبکه عصبی پرسپترون چند لایه است. این امر نشان می‌دهد که با برتری قابل قبول مدل‌های هوشمند نسبت به دیگرروش‌های پیش‌بینی انتخاب صحیح مدل هوشمند برای پیش‌بینی امری مهم است.

بحث و نتیجه‌گیری

امروزه مدیران، سرمایه‌گذاران وسهامداران به دلیل وجود انبوه متغیرهای تاثیرگذار ترجیح می‌دهند، مکانیزمی را در اختیار داشته باشند که بتواند آن‌ها را در امور تصمیم‌گیریشان یاری و مشاوره دهد، به همین دلیل سعی در استفاده از روش‌هایی در پیش‌بینی دارند که به واسطه آن‌ها تخمین‌هایشان به واقعیت نزدیک و خطایشان بسیار کم باشد. رساله حاضر پیش‌بینی قیمت سهام را در سطح داده‌های روزانه توسط پنج الگوریتم تخمین نزدیک‌ترین همسایه، درخت تصمیم، لاجستیک، شبکه عصبی پرسپترون چند لایه و ماشین بردار پشتیبان انجام داده است تا بهترین مدل را از دیدگاه دقت در امر پیش‌بینی، معرفی کند. در واقع هرچه تخمین یک مدل به واقعیت نزدیک‌تر باشد دارای خطای کمتری در پیش‌بینی است از این رو از معیار میانگین مجذور خطا (MSE) استفاده می‌شود که از نظر پژوهشگران معیار قابل قبولی می‌باشد.

با توجه به یافته‌های این پژوهش می‌توان اینگونه نتیجه‌گیری کرد که داده‌های مربوط به اطلاعات درونی شرکت‌ها از جمله نسبت‌های مالی و اطلاعات مربوط به سهام و همچنین داده‌های بیرون از شرکت یا به عبارت دیگر داده‌های کلان اقتصادی به صورت ترکیبی می‌توانند

در پیش‌بینی شاخص‌های تعیین‌کننده قیمت سهام به صورت مؤثری بکار گرفته شوند. گرچه شبکه‌های عصبی قادر به پیش‌بینی بازده سهام با استفاده از متغیرهای کلان اقتصادی است، اما ماشین بردار پشتیبان در این امر موفق‌ترند و می‌توانند خطای پیش‌بینی را به طور معنی‌داری کاهش دهند که در مقایسه با نتایج پژوهش‌های دیگر که اذعان داشته‌اند ماشین بردار پشتیبان عملکرد بهتری نسبت به شبکه عصبی و الگوریتم ژنتیک دارند همخوانی دارد. نتایج شبیه سازی نشان داد که پیش‌بینی‌های مدل ماشین بردار پشتیبان و سپس شبکه‌ی عصبیِ مصنوعی نسبت به دیگر مدل‌ها بهتر است. به عبارتی دیگر، می‌توان گفت که مدل ماشین بردار پشتیبان و شبکه ی عصبی به دلیل انعطاف‌پذیری در مدل بندی اثرات اصلی و متقابل متغیرهای همراه، پیش‌بینی‌های بهتری نسبت به دیگر مدل‌ها فراهم کرده‌اند.

نتایج پژوهش نشان می‌دهد که رفتار سری زمانی قیمت روزانه سهام شرکتها و شاخصها در بورس تهران تصادفی نیست، اما این فرآیند غیر تصادفی دارای پیچیدگی های زیادی است و هنگامی که از شبکه‌های عصبی جهت پیش‌بینی استفاده می‌شود، در طراحی مدل شبکه عصبی نیاز به استفاده از شبکه با تعداد لایه ها و نرونهای میانی متناسب هستیم.

بر اساس معیارهای ارزیابی مدل‌ها می‌توان چنین نتیجه گرفت که مدل‌های استفاده شده در این پژوهش توانایی پیش‌بینی روند قسمت سهام با استفاده از شاخص‌های کل و صنعت را دارا است و این امر تایید دیگری بر توانایی مدل‌های یادگیری ماشین در پیش‌بینی های حوزه مالی، از جمله قیمت سهام در بورس اوراق بهادار تهران است.

ارتباط بین قیمت سهام و متغیرهای کلان اقتصادی همواره ثابت نیست و به علل مختلف، این همبستگی دچار تغییر می‌شود. بنابراین برای برازش مدل‌هایی که هدف آن پیش‌بینی افق‌های کوتاه مدت است بهتر است از سری‌های زمانی نزدیک‌تر استفاده نشود که در مقایسه با پژوهش‌های پیشین که بخش‌های مختلف اقتصاد، همه به یک میزان تحت تاثیر این تغییرات قرار نمی‌گیرند و این خود بستگی به بزرگی آن بخش از اقتصاد دارد. شاید به دلیل آنکه بورس سهام ایران در مقایسه با حجم اقتصاد کشور بسیار کوچک است، نمی‌تواند تغییرات واقعی در اقتصاد را به خوبی نشان دهد. از سوی دیگر از آنجا که اقتصاد ایران یک اقتصاد دولتی است، به دلیل وجود فضای بوروکراتیک و کندی مراحل تصمیم گیری تا اجرا، واکنش صورت گرفته در متغیرهای اقتصادی با تاخیر زمانی صورت می‌گیرد. در کلامی ساده‌تر، تغییر در برخی متغیرها

از قبیل تورم و رشد اشتغال بلافاصله بر هزینه‌ها و درآمدهای شرکت اثر نمی‌گذارد و تاثیرات آن در طول زمان ظاهرمی‌شود و عکس العمل بازار نیز با تاخیر رویت می‌شود.

نتایج به دست آمده برای دوره زمانی مورد بررسی، یعنی بازه زمانی ده ساله ۱۳۸۸ تا ۱۳۹۷ صادق است، بنابراین به منظور بررسی این مساله که آیا مدل یادشده برای همه بازه‌های زمانی کاربرد دارد یا خیر، می‌توان دوره‌های زمانی طولانی‌تر و انجام تقسیم‌بندی‌های جزئی‌تر را به کار برد. برای مثال، می‌توان دوره زمانی ده ساله را به دو دوره پنج‌ساله و یا پنج دوره دوساله تقسیم و در هر دوره اقدام به پیش‌بینی و سپس نتایج را مقایسه کرد.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

منابع

احمدخان بیگی، سهیل، عبدالوند، ندا. (۱۳۹۶). پیش‌بینی قیمت سهام با رویکرد ترکیبی شبکه‌های عصبی مصنوعی و الگوریتم رقابت استعماری مبتنی بر تئوری آشوب. *راهبرد مدیریت مالی*، ۵ (۳)، 27- 73. Doi: 10. 22051/jfm. 2017. 14635. 1319

امامیان، سید محمد طاهر، محمودی راد، علی، ملاعلیزاده زواردهی، صابر، نیرومند، صادق. (۱۴۰۱). الگوریتم سیستم‌های تحلیل ترکیبی در پیش‌بینی مطلوب‌ترین پرتفوی، *فصلنامه بورس اوراق بهادار*، شماره ۵۸، ۱۶۱-۱۸۰، 10.22034/JSE.2021.11619.1725

برادران حسن زاده، رسول، محرومی، رامین. (۱۳۹۶). تأثیر افشای ریسک بر پیش‌بینی قیمت سهام به وسیله سود و ارزش شرکت ها. *راهبرد مدیریت مالی*،۵ (۴)، ۱۸۹-۲۰۹ ، 1274، Doi: 10. 22051/jfm. 2018. 13881

بیات، علی، باقری، زینب. (۱۳۹۶). پیش‌بینی قیمت سهام بااستفاده از الگوریتم کرم شب‌تاب (FA). *دانش مالی تحلیل اوراق بهادار*، ۱۰ (۳۵)، ۱۴۵-۱۳۵.

حسینی‌نسب، حجت، کریمی تکلو، سلیم، یوسفی‌نژاد، مرضیه. (۱۳۹۲). مقایسه دقت رویکردهای ماشین‌های بردار پشتیبان و شبکه‌های عصبی مصنوعی در پیش‌بینی سود هر سهم شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. *دوفصلنامه جستارهای اقتصادی*، ۱۰ (۲۰)، ۱۳۴-۱۰۹.

رجب پور، ابراهیم، تقوا، محمدرضا، حسین زاده یزدی، محمدعلی، بابااحمدی، سارا. (۱۳۹۳). پیش‌بینی قیمت سهام شرکت های بورس اوراق بهادار تهران با استفاده از شبکه‌های عصبی مصنوعی. *فصلنامه پژوهشهای نوین در حسابداری*، ۲ (شماره ۴ (تابستان ۹۳)) ، ۴۵-۵۷.

زارعی، قاسم، محمدیان، رعنا، حاضری نیری، هاتف، باشکوه اجیرلو، محمد. (۱۳۹۷). مقایسه روشهای شبکه‌های عصبی فازی با شبکه‌های عصبی موجک فازی در پیش‌بینی قیمت سهام بانک‌های پذیرفته‌شده در بورس اوراق بهادار تهران. *راهبرد مدیریت مالی*، ۶ (۳)، ۱۳۸-۱۰۹. Doi: 10. 22051/jfm. 2018. 19214. 1606

سهرابی، مریم، میربرگ کار، سید مظفر، چیرانی، ابراهیم، خردیار، سینا، (۱۴۰۱). مدلسازی پیش‌بینی جهش‌های شاخص بازار سهام بر اساس رویکرد شبکه عصبی بازگشتی یادگیری عمیق. *فصلنامه بورس اوراق بهادار*، شماره ۵۹، ۲۴۵-۲۶۸، 10.22034/JSE.2021.11655.1762

مکیان، سید نظام الدین، موسوی، فاطمه السادات. (۱۳۹۰). پیش‌بینی قیمت سهام شرکت فرآورده های نفتی پارس با استفاده از شبکه‌های عصبی و روش رگرسیونی مطالعه موردی: قیمت سهام شرکت فرآورده‌های نفتی پارس. *فصلنامه علمی - پژوهشی مدلسازی اقتصادی، ۶ (۱۸)، ۱۲۱-۱۰۵.*

References

Ahmadkhan Beigi, Sohail, Abdulvand, Neda. (2016). Stock price forecasting with a combined approach of artificial neural networks and colonial competition

algorithm based on chaos theory. *Financial Management Strategy*, 5 (3), 27-73. Doi: 10. 22051/jfm. 2017. 14635. 1319. (In Persian).

Baradaran Hassanzadeh, Rasul, Mahroomi, Ramin. (2016). the effect of risk disclosure on the prediction of stock prices by profit and value of companies. *Financial Management Strategy*, 5 (4), 189-209. Doi: 10. 22051/jfm. 2018. 13881. 1274. (In Persian).

Bayat, Ali, Bagheri, Zainab. (2016). Stock price prediction using firefly algorithm (FA). *Financial knowledge of securities analysis*, 10 (35), 135-145. (In Persian).

Bird, R., Reddy, K., and Yeung, D. (2011). The relationship between uncertainty and the market reaction to information: How Is It Influenced by Market and Stock-Specific Characteristics? Working paper series: The Paul Woolley Centre for Capital Market Dysfunctionality, UTS, 14.

Emamiyan, S.M.T., Mahmoodirad, A., Mola Alizadeh Zavardehi, S., Niroomand, S., (2022). Algorithm of Hybrid Analysis Systems in Predicting Optimal Portfolio, *Journal of Securities Exchange*, 161-180, 10.22034/JSE.2021.11619.1725. (In Persian).

Hosseininasab, Hojjat, Karimi Takallu, Salim, Yousefinejad, Marzieh. (2012). Comparing the accuracy of support vector machines and artificial neural networks approaches in predicting the profit per share of companies listed on the Tehran Stock Exchange. *Bi-quarterly journal of economic essays*, 10 (20), 109-134. (In Persian).

Huang, C. -J., Yang, D. -X., & Chuang, Y. -T. (2008). Application of wrapper approach and composite classifier to the stock trend prediction. Expert Systems with Applications, 34(4), 2870–2878. http://dx. Doi.org/10. 1016/j. eswa. 2007. 05. 035.

Kumar, M., & Thenmozhi, M. (2006). Forecasting Stock index movement: A comparison of support vector machines and random forest. SSRN Scholarly Paper. *Rochester, NY: Social Science Research Network*, January 24, 2006. <http://papers. Ssrn. com/abstract=876544<.

Long Wenabc, Song Linqiuabc, Tian Yingjie (2019), a new graphic kernel method of stock price trend prediction based on financial news semantic and structural similarity, *Expert Systems with Applications*, Volume 118, Pages 411-424.

Makian, Seyyed Nezamuddin, Mousavi, Fatemeh Sadat. (1390). Forecasting the stock price of Pars Oil Products Company using neural networks and regression method. Case study: Pars Oil Products Company stock price. *Economic Modeling Scientific-Research Quarterly*, 6 (18), 105-121. (In Persian).

Michel B., Dirk V., Nathalie H., Ruben G. (2015). Evaluating multiple classifiers for stock price direction prediction, Expert Systems with Applications, *journal homepage: www. Elsevier. com/locate/eswa*.

Ming-Chi Lee, (2009). Using support vector machine with a hybrid feature selection method to thestock trend prediction, Department of Computer Science and Information Engineering, National Pingtung Institute of Commerce, No. 51 Minsheng E. Rd., Pingtung 900, Taiwan, ROC

Mouseli, S., Jaafar, A., and Hussainey, K. (2012). Accruals Quality vis-à-vis Disclosure Quality: Substitutes or Complements. *The British Accounting Review*, 44 (2012), 36–46.

Rajabpour, Ibrahim, Taqwa, Mohammad Reza, Hosseinzadeh Yazdi, Mohammad Ali, Baba Ahmadi, Sara. (2013). Forecasting stock prices of Tehran Stock Exchange companies using artificial neural networks. *Quarterly Journal of New Researches in Accounting*, 2 (Number 4 (Summer 1993)), 45-57. (In Persian).

Sohrabi, M., Mirbargkar, S.M., Chirani, E., Kheradyar, S., (2022), Modeling the Prediction of Stock Market Jumps Based on the Recurrent Neural Network and Deep Learning, *Journal of Securities Exchange*, 245-268, 10.22034/JSE.2021.11655.1762. (In Persian).

Tsang, P. M., Kwok, P., Choy, S. O., Kwan, R., Ng, S. C., Mak, J., Tsang, J., Koong, K., Wong, T, L., 2007. Design and implementation of NN5 for Hong Kong stock price forecasting, E*ngineering Applications of Artificial Intelligence* 20:453–461.

Tsinaslanidis, P. E., & Kugiumtzis, D. (2014). A prediction scheme using perceptually important points and dynamic time warping. Expert Systems with Applications, 41, 6848–6860. http://dx. Doi.org/10. 1016/j. eswa. 2014. 04. 028.

Ubha, M. V., & Nambi, S. T. (2012). Classification of Stock Index movement using K- nearest neighbors (k-NN) algorithm. *WSEAS Transactions on Information Science & Applications*, 9(9), 261–270. P259.

Wen Long Zhichen Lu Lingxiao Cui (2018-2019), deep learning-based feature engineering for stock price movement prediction, *Knowledge-Based Systems* Volume 164, Pages 163-173.

Yao, T., Yu, T., Zhang, T., and Chen, S. (2011). Asset growth and stock returns : Evidence from Asian financial markets. *Pacific-Basin Finance Journal*–115.

Yakup Kara, Melek Acar Boyacioglu, mer Kaan Baykan. (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. 5311–5319.

Zarei, Qasim, Mohammadian, Rana, Hazeri Naieri, Hatef, Bashokh Ajirlu, Mohammad. (2017). Comparison of fuzzy neural network methods with fuzzy wavelet neural networks in predicting stock prices of banks admitted to Tehran Stock Exchange. *Financial Management Strategy*, 6 (3), 109-138. Doi: 10. 22051/jfm. 2018. 19214. 1606. (In Persian).

Zhang, Z. Y., et al. (2005). Stock time series forecasting using support vector machines employeing analyst recommendations, *Springer-Verlag Berlin Heidelberg*.

COPYRIGHTS

This is an open access article under the CC BY-NC 4.0 license.