مجله محاسبات نرم، جلد 14، شماره 2 (پاییز و زمستان 1404)، صفحه 138ـ151
https://doi.org/10.22052/scj.2025.255297.1260

مجله محاسبات نرم
SOFT COMPUTING JOURNAL
تارنمای مجله: scj.kashanu.ac.ir

University of Kashan

مقایسه روش رگرسیون بردار پشتیبان معمولی و گروهی برای پیش‌بینی قیمت سهام✧

سید محمد حسینی[1*]، استادیار، مجید ابتیاع[1]، کارشناس ارشد، محمدرضا آریافر[2]، دانشجوی دکتری

[1] گروه پژوهشی هوش مصنوعی گهر، دانشکده علوم پایه، دانشگاه آیت الله بروجردی (ره)، بروجرد، ایران.

[2] گروه اقتصاد، دانشکده علوم اداری و اقتصادی، دانشگاه فردوسی مشهد، مشهد، ایران.

اطلاعات مقاله

تاریخچه مقاله:
دریافت 28 مرداد ماه 1403
پذیرش 10 اسفند ماه 1403

کلمات کلیدی:
یادگیری ماشین
رگرسیون بردار پشتیبان
روش گروهی
بازار سهام
پیش‌بینی قیمت

چکیده

انباشت سرمایه و نگهداری صحیح از دارایی نقش بسزایی را در رشد اقتصادی کشورها ایفا می‌کند. نقش موثر این عامل و اهمیت آن را به روشنی در سیستم اکثر کشورها به ویژه نظام‌های سرمایه‌داری می‌توان مشاهده نمود. بازار سرمایه یکی از مناسبترین جایگاه‌ها به منظور جذب دارایی‌های اندک و بکارگیری آنها با هدف رشد و ارتقا یک شرکت و همچنین افزایش دارایی شخصی سرمایه‌گذاران است. از اینرو، سرمایه‌گذاری درست، به افراد کمک می‌کند تا با مدیریت صحیح بخش مازاد درآمدهای خود، بتوانند به سرمایه لازم جهت رسیدن به اهداف خود دست پیدا کنند. اساسا قیمت سهام، غیرخطی و آشوبناک است، بنابراین سرمایه‌گذاری در بازار بورس، خطر بالایی به همراه دارد. برای به حداقل رساندن این خطر و کاهش مخاطرات، یک سیستم کارا مورد نیاز است که قادر باشد حرکت قیمت سهام در آینده را با دقت بالایی پیش‌بینی نماید. در این راستا، مدل‌های یادگیری ماشین عملکرد مناسبی در زمینه پیش‌بینی با دقت بالا را دارا هستند. الگوریتم‌های یادگیری ماشین توانایی مدل نمودن سیستم‌های غیرخطی و پیچیده را دارند. در این پژوهش به کمک رویکرد گروهی جدید مبتنی بر رگرسیون بردار پشتیبان یک مدل جمعی یا ترکیبی ساخته می‌شود که از دقت و سرعت بالایی در پیش‌بینی قیمت سهام برخوردار است. سهم‌های مورد استفاده در این پژوهش ده سهام شاخص‌ساز فولاد، شستا، فارس، فملی، شپنا، شتران، شبندر، خودرو، خساپا و پارس است که در بازه زمانی 1400/01/07 تا 1403/04/31 در نظر گرفته شده است. بر اساس معیارهای مختلف، نتایج به دست آمده برتری روش رگرسیون بردار پشتیبان گروهی بر روش معمولی رگرسیون بردار پشتیبان و روش جنگل تصادفی را نشان می‌دهد.

© 1403 نویسندگان. مقاله با دسترسی آزاد تحت مجوز CC-BY

1. مقدمه

بازارهای مالی، به ویژه بازار سرمایه گسترده و کارآمد، به عنوان ابزاری برای تحقق رشد بلندمدت و پایداری اقتصادی، مورد توجه قرار می‌گیرند. تحقق این امر، نیازمند تخصیص بهینه منابع در سطح اقتصاد ملی است. بورس اوراق بهادار یکی از مسیرهای جذب سرمایه است. البته تحقق این امر، مستلزم جلب اعتماد سرمایه‌گذاران است. بازار سرمایه، امکان جذب اعتماد سرمایه‌گذاران را فراهم می‌کند و اوراق بهادار به عنوان ابزارهای

✧ نوع مقاله: پژوهشی
* نویسنده مسئول
پست(های) الکترونیک: sm.hoseini@abru.ac.ir (حسینی)
majid.ebtia@gmail.com (ابتیاع)
m.ariafar@mail.um.ac.ir (آریافر)

نحوه ارجاع به مقاله: سید محمد حسینی، مجید ابتیاع، محمدرضا آریافر، «مقایسه روش رگرسیون بردار پشتیبان معمولی و گروهی برای پیش‌بینی قیمت سهام»، مجله محاسبات نرم، جلد 14، شماره 2، ص138ـ151، پاییز و زمستان 1404.

متنوع جهت مدیریت ریسک و جمع‌آوری سرمایه‌های پراکنده استفاده می‌شوند. این رویکرد نقش مهمی در توسعه اقتصادی و پیشرفت کشورها ایفا می‌کند. گردش سرمایه در حال حاضر بیشترین حجم خود را از طریق بازارهای سرمایه دارد و عملکرد این بازار بر تحولات اقتصادی برخی کشورها با بازار سرمایه دارای شفافیت اطلاعات تاثیرگذار است [1]. با این حال، عوامل ناشناخته و متعددی بر بازارهای اوراق بهادار تاثیر می‌گذارند که سبب عدم اطمینان در تصمیم‌گیری سرمایه‌گذاران شده است. به همین دلیل، تحقیقات و توسعه روش‌های پیش‌بینی دقیق و کاهش خطا در تحلیل بازارها از اهمیت ویژه‌ای برخوردارند [2]. در این راستا، پژوهشگران و سرمایه‌گذاران در تلاشند تا با بهره‌گیری از دانش فنی و تخصص، راهکارهایی برای پیش‌بینی دقیق‌تر و مدیریت بهینه ریسک در بازارهای مالی ایجاد کنند. از این جهت، ترکیب مدل‌های مختلف و بکارگیری ابزارهایی همچون بگینگ[1] و رگرسیون بردار پشتیبان[2] (SVR) می‌تواند بهبودهای مهمی در عملکرد پیش‌بینی و مدیریت سرمایه حاصل کند.

به طور کلی، برای پیش‌بینی قیمت سهام دو نوع روش وجود دارد. یکی از رویکردها، استفاده از روش‌های آماری سنتی مثل میانگین ساده و رگرسیون است. این روش‌ها زمانی کارآمدند که بین متغیرهای مستقل و وابسته ارتباط و همبستگی مشخصی داشته باشند. بازار سهام به شدت نامنظم و آشوبناک است، بنابراین نیاز به یک سیستم کارآمد جهت پیش‌بینی دقیق قیمت سهام وجود دارد. مزیت عمده سیستم‌های هوشمند پیش‌بینی، دقت بالاتر آنها است.

در دهه‌های اخیر، با پیشرفت روزافزون در حوزه فناوری و گسترش بازارهای مالی، پیش‌بینی قیمت سهام به یکی از چالش‌های اساسی در زمینه سرمایه‌گذاری تبدیل شده است. یکی از روش‌های مطرح در این زمینه، استفاده از مدل‌های رگرسیون بردار پشتیبان است. در حال حاضر، با توجه به پیشرفت چشمگیر تکنولوژی و سرعت بالای محاسبات، همچنین در دسترس بودن داده‌های گسترده در حوزه‌های مختلف، دیدگاه‌های علمی متمرکز بر داده، از جمله هوش مصنوعی و یادگیری ماشین، به عنوان رویکردهای اصلی در زمینه تحقیقات، توجه بسیاری از پژوهشگران را به خود جلب کرده است. این دیدگاه‌ها نه‌تنها در زمینه‌های علمی، بلکه در کاربردهای متعددی نیز وارد شده و نتایج قابل‌توجهی را به همراه داشته‌اند. ابزار SVR یکی از ابزارهای پرقدرت در پیش‌بینی قیمت سهام است که بر پایه ایده‌های ماشین بردار پشتیبان[3] (SVM) برای مسائل پیش‌بینی مقادیر عددی پیوسته طراحی شده است. این مدل با توانایی مدیریت داده‌های پیچیده و ارائه پیش‌بینی دقیق، به عنوان یکی از ابزارهای مهم و مورد توجه در زمینه پیش‌بینی قیمت سهام شناخته می‌شود. بگینگ، یک تکنیک موثر در افزایش استحکام و دقت مدل‌های یادگیری ماشین است [3]. این روش شامل نمونه‌گیری تصادفی از داده‌ها، آموزش مدل بر روی هر نمونه و سپس ترکیب پیش‌بینی‌های این مدل‌ها است. این فرآیند منجر به بهبود دقت نهایی مدل می‌شود.

در این مقاله، یک روش جدید ترکیبی با استفاده از مدل‌های بگینگ و SVR ارائه شده است. در این روش پیشنهادی، ابتدا مدل SVR روی نمونه‌های مختلفی از داده‌ها آموزش داده می‌شود. سپس با استفاده از بگینگ، مدل‌های مختلفی ایجاد می‌شوند و هرکدام بر روی یک نمونه تصادفی از داده‌ها آموزش داده می‌شوند. این ترکیب از مدل‌ها منجر به بهبود دقت و مقاومت مدل در برابر نوسانات بازار می‌شود. روش پیشنهادی از قدرت پیش‌بینی SVR بهره می‌برد و با استفاده از بگینگ، دقت و پایداری پیش‌بینی‌ها را افزایش می‌دهد. ارزیابی عملکرد این روش با استفاده از معیارهای دقت پیش‌بینی، خطای میانگین مطلق (MAE) و خطای میانگین مربعات (MSE) صورت گرفته است [4].

نوآوری اصلی این پژوهش در استفاده از ترکیب مدل‌های رگرسیون بردار پشتیبان (SVR) با تکنیک بگینگ برای پیش‌بینی قیمت سهام است. این رویکرد ترکیبی امکانات مدل SVR در مدیریت داده‌های پیچیده و پیش‌بینی دقیق را با افزایش دقت و

[1] Bootstrap Aggregating (Bagging)
[2] Support Vector Regression
[3] Support Vector Machines

پایداری مدل توسط بگینگ ترکیب می‌کند. این نوآوری به سرمایه‌گذاران و تحلیلگران بازار ابزاری قوی ارائه می‌دهد تا در تصمیم‌گیری‌های سرمایه‌گذاری بهترین عملکرد را داشته باشند. همچنین، این رویکرد با مقاومت بیشتری در برابر نوسانات بازار و توانایی تعمیم به شرایط جدید، به عنوان یک ابزار قوی برای مدیریت ریسک در بازار سهام مطرح می‌شود. به طور کلی، استفاده از ترکیب مدل‌های مختلف با روش‌های گروهی یا جمعی برای پیش‌بینی قیمت سهام، به عنوان یک نوآوری مهم و کارآمد در زمینه تحلیل مالی و سرمایه‌گذاری معرفی می‌شود که می‌تواند به بهبود عملکرد مدل‌های پیش‌بینی و افزایش دقت تصمیم‌گیری‌های سرمایه‌گذاران کمک کند.

نتایج مطالعاتی که تا به اینجا بررسی شده‌اند، نشان می‌دهند که روش گروهی باعث بهبود قابل‌توجهی در دقت پیش‌بینی‌ها نسبت به مدل‌های تکی SVR می‌شود. اما با وجود بهبودهای ایجاد شده توسط روش پیشنهادی، هنوز چالش‌هایی نیز وجود دارد که ممکن است تاثیر منفی بر عملکرد مدل داشته باشند. به عنوان مثال، یکی از چالش‌ها مدیریت داده‌های نوسانی بازار است، زیرا بازارهای مالی ممکن است تحت‌تاثیر عوامل ناپیوسته و پیچیده قرار گیرند که مدل را به چالش کشانده و دقت پیش‌بینی‌ها را کاهش دهند. همچنین، حساسیت مدل به تغییرات پارامترها نیز یک چالش مهم است، زیرا نیاز به تنظیم دقیق پارامترها برای حفظ عملکرد بهینه مدل وجود دارد.

در مراحل انجام این پژوهش، ابتدا با مرور ادبیات موجود در زمینه پیش‌بینی قیمت سهام و مدل‌های مورد استفاده، نقاط قوت و ضعف این روش‌ها شناسایی می‌شود. سپس، مدل پیشنهادی بر اساس ترکیب مدل‌های رگرسیون بردار پشتیبان (SVR) و تکنیک بگینگ ارائه شده است. سپس مراحل آموزش مدل و تولید پیش‌بینی‌ها شرح داده شدند. در ادامه، از معیارهای دقت پیش‌بینی، خطای میانگین مطلق (MAE) و خطای میانگین مربعات (MSE) برای ارزیابی عملکرد مدل استفاده شده است. در انتها ارزیابی جامع و دقیق عملکرد مدل انجام می‌شود.

2. پیشینه پژوهش

تاکنون، پیش‌بینی قیمت سهام موضوع بسیاری از پژوهش‌ها بوده است و رویکردها و تکنیک‌های مختلفی برای این منظور بررسی شده‌اند. در سال2022 ایلیاس و همکاران [5]، یک مدل ترکیبی را توسعه دادند که در آن رگرسیون بردار پشتیبان و فیلتر هودریک-پرسکات یکپارچه‌سازی می‌شوند و قابلیت پیش‌بینی قیمت سهام را افزایش می‌دهند. به طور مشابه، در سال 2014، هوآنگ و همکاران [6]، یک چارچوب را نشان دادند که رگرسیون بردار پشتیبان و تبدیل فوریه را برای پیش‌بینی قیمت سهام بر اساس داده‌های تاریخی یکپارچه می‌کند. این مطالعات نشان‌دهنده استفاده از رگرسیون بردار پشتیبان در مدل‌های پیش‌بینی قیمت سهام هستند.

در سال 2019، بشیر و همکاران [7]، بر اهمیت چالش‌هایی که در پیش‌بینی قیمت سهام وجود دارد تاکید کرد و اعلام کرد که در این زمینه موفقیت نسبی به دست آمده است. عدم قطعیت در سیاست‌های اقتصادی و تغییرات در قیمت کالاها می‌تواند تاثیرات غیرمنتظره‌ای بر بازارهای سهام داشته باشد که پیش‌بینی این تغییرات را دشوار می‌کند. روابط غیرخطی و پیچیده بین متغیرهای اقتصادی و قیمت سهام و واکنش‌های متفاوت بازارها به تغییرات مشابه، از جمله چالش‌های اصلی در مدل‌سازی و پیش‌بینی هستند که آنها در تحلیل خود بر آن تاکید کردند. همچنین، در سال 2020، پاسالیس و همکاران [8]، ویژگی‌های مختلفی را شناسایی کرده و از یک روش طبقه‌بندی مبتنی بر رگرسیون بردار پشتیبان (SVM) برای پیش‌بینی حرکات قیمت سهام استفاده کرده‌اند که نمونه‌ای از اعمال تکنیک‌های یادگیری ماشین در این زمینه است. همچنین، ابتونگ و همکاران [9]، در همان سال به تحلیل و بهبود دقت پیش‌بینی قیمت سهام پرداختند. روش‌های مبتنی بر یادگیری ماشین مانند جنگل تصادفی [10] شبکه‌های پرسپترون چندلایه و ماشین‌های رگرسیون بردار پشتیبان [11] و همچنین تکنیک‌هایی مانند تبدیل فوریه [6] و روش جنگل تصادفی[1] و الگوریتم ماشین بردار پشتیبانی[2] [11]

[1] Random Forest
[2] SVM

برای پیش‌بینی قیمت سهام استفاده شده است.

در این راستا، مطالعه لیو و همکاران [13]، بکارگیری یک مدل CAE-LSTM را نشان داده است که با استفاده از اطلاعات فنی، عملکرد بهتری در پیش‌بینی قیمت سهام به دست آورده است. این مطالعه به وضوح نشان می‌دهد که ترکیب شبکه‌های عصبی پیچشی و بازگشتی می‌تواند بهبود معناداری در دقت پیش‌بینی داشته باشد. همچنین، رستم و کینتاندانی [14]، در سال 2019 با تمرکز بر روی بازار سهام اندونزی، از رگرسیون بردار پشتیبان به همراه انتخاب ویژگی، نشان داده‌اند که این روش قابلیت پیش‌بینی بهتر در بازارهای خاصی را ارائه می‌دهد.

در پژوهشی توسط منادی و نجفی [15]، با هدف ارتقای کیفیت انتخاب سبد سرمایه‌گذاری، استفاده از یک مدل پیشرفته رگرسیون بردار پشتیبان با خروجی‌های چندگانه مورد بررسی قرار گرفت. این رویکرد، که در جهت دستیابی به حداکثر بازده با کمترین ریسک است، بر تحلیل دقیق‌تر قیمت سهام تاکید دارد. استفاده از داده‌های مربوط به شرکت‌های برجسته S&P500 طی دوره‌ای معین نشان داد که این مدل، با در نظر گرفتن ارتباطات بین خروجی‌ها، عملکرد بهتری نسبت به روش‌های سنتی پیش‌بینی، به خصوص در زمینه کارایی بر اساس معیار شارپ، از خود نشان می‌دهد.

دهقان و همکاران [16] در پژوهشی خود، مدل‌های طبقه‌بندی مختلف برای دست‌یابی به یک مدل با کارایی و دقت بالا در پیش‌بینی قیمت‌گذاری کمتر از واقع سهام عرضه عمومی اولیه مورد بررسی قرار دادند. نتایج نشان داده‌اند که مدل‌های ماشین بردار پشتیبان، شبکه بیزین دقت بالایی در پیش‌بینی قیمت‌گذاری کمتر از واقع دارند. همچنین، متغیرهای مهم و تاثیرگذار شامل رشد دارایی‌ها، دوره تصدی حسابرس، تخصص حسابرس در صنعت، نسبت فعالیت‌های تامین مالی، نسبت قیمت به سود هر سهم، بازده دارایی‌ها، نسبت فعالیت‌های عملیاتی، اندازه موسسه حسابرسی، فرصت‌های رشد و نوسانات قیمت سهام هستند.

خسروی‌نژاد و همکاران [17] در مطالعه‌ای، ابتدا با استفاده از مدل سری زمانی و شبکه عصبی مصنوعی، متغیر هفتگی شاخص قیمت سهام در بورس اوراق بهادار تهران را برآورد کرده‌اند. این برآوردها برای دوره زمانی 1383 تا 1387 صورت گرفته و سپس قدرت پیش‌بینی دو مدل در دوره زمانی 1387 تا 1389 مورد آزمون و ارزیابی قرار گرفته است.

منجمی و همکاران [18]، یک مدل پیش‌بینی قیمت سهام با استفاده از شبکه‌های عصبی فازی و الگوریتم‌های ژنتیک طراحی کردند. نتایج حاصل این پژوهش مشخص می‌کند که ادغام شبکه‌های عصبی فازی و الگوریتم‌های ژنتیک به عنوان یک مدل ترکیبی، توانمندی بیشتری در پیش‌بینی قیمت سهام از نظر دقت و کارایی از خود نشان می‌دهد. این یافته می‌تواند به توسعه رویکردهای بهبود یافته در حوزه پیش‌بینی بازار سهام کمک کند و به سرمایه‌گذاران ابزاری قدرتمندتر برای تصمیم‌گیری در بازار سرمایه ارائه دهد.

رمضانی و عاملی [19]، در مطالعه‌ای از الگوریتم ژنتیک استفاده کردند. در بخش پیش‌بینی، از شبکه عصبی فازی به همراه دو ساختار مختلف (ممدانی[1] و سوگنو[2]) برای پیش‌بینی قیمت سهام استفاده شده است. الگوریتم پیشنهادی با استفاده از داده‌های 10 شرکت، هر کدام دارای 7 ویژگی، ارزیابی شده است. نتایج نشان داده است که با توجه به نوع شرکت‌ها، ویژگی‌ها و ساختار شبکه عصبی فازی ترکیبی، نتایج متفاوتی به دست می‌دهد. ارزیابی‌ها نشان می‌دهد که الگوریتم پیشنهادی با ساختار سوگنو با الگوریتم ژنتیک، در مقایسه با ساختار ممدانی، دارای عملکرد بهتری است، زیرا تعداد پارامترهای آموزش آن بیشتر است.

زمانی و همکاران [20] در پژوهشی، از مدل‌های ریاضی بهینه‌سازی استفاده کردند تا با توجه به متغیرهایی از قبیل میانگین، واریانس و چولگی، سبدی از سهام را پیش‌بینی کنند. نتایج تحقیق نشان می‌دهد که مدل‌های ارائه شده، در مقایسه با روش‌های سنتی و شاخص‌های بازار، بازدهی بالاتری را برای سرمایه‌گذاران فراهم می‌کنند. این مدل‌ها از ترکیب پتانسیل آتی سهام با اطلاعات گذشته برای تصمیم‌گیری بهینه‌تر استفاده می‌کنند که به سرمایه‌گذاران ابزار قدرتمندتری برای تصمیم‌گیری در بازار بورس ارائه می‌دهد.

[1] Mamdani
[2] Sogno

سادورسکی [10]، یک مطالعه مرتبط در خصوص استفاده از جنگل‌های تصادفی برای پیش‌بینی قیمت سهام ارائه کرده است. یافته‌های این مطالعه نشان می‌دهد که برای پیش‌بینی قیمت سهام، پیش‌بینی‌های جنگل‌های تصادفی دقیق‌تر از مدل‌های لاجیت با توجه به بهره‌مندی از ظرفیت روش‌های گروهی است.

انتی و همکاران [21]، در سال 2020 از یک روش گروهی، یعنی ماشین‌های رگرسیون بردار پشتیبان، برای پیش‌بینی کارایی بازار سهام استفاده کرده است. استفاده از الگوریتم ژنتیک برای بهینه‌سازی ابرپارامترهای SVM، رویکردی برای بهبود عملکرد پیش‌بینی را نشان می‌دهد.

شائو و همکاران [22]، در سال 2019 از تحلیل طیفی تکین[1] و ماشین‌های رگرسیون بردار پشتیبان برای تجزیه و تحلیل و پیش‌بینی قیمت سهام استفاده کرده‌اند. مطالعه مقیاس‌های زمانی مختلف و عوامل موثر بر قیمت‌های سهام، زمینه‌ای را برای چالش‌های پیش‌بینی فراهم می‌کند.

رستام و کینتاندانی [14]، در سال 2019 به طور خاص بر روی رگرسیون بردار پشتیبان برای پیش‌بینی قیمت سهام اندونزی با انتخاب ویژگی تمرکز دارد. این تحقیق نشان‌دهنده کاربرد رگرسیون بردار پشتیبان در بازاری خاص است و همچنین کنترل ویژگی‌های مرتبط در آن بررسی شده است.

کلهری و حسینی [23]، از یک روش مبتنی بر گراف، به کمک الگوریتم کوچک‌ترین درخت پوشا برای خوشه‌بندی سهام شرکت‌های بورس اوراق بهادار تهران استفاده کردند.

تحقیقات موجود از روش‌های گروهی مانند جنگل‌های تصادفی و ماشین‌های رگرسیون بردار پشتیبان و همچنین شبکه‌های عصبی مانند LSTM، برای پیش‌بینی قیمت سهام استفاده کرده‌اند. این مطالعات همچنین چالش‌های مرتبط با داده‌های بازار سهام در مقیاس‌های زمانی مختلف را برجسته می‌کنند و نیاز به انتخاب و بهینه‌سازی مدل را برای بهبود دقت پیش‌بینی نشان می‌دهند. منابع انتخاب شده، پایه‌ای را برای رویکرد ترکیبی پیشنهادی رگرسیون بردار پشتیبان با استفاده از بگینگ فراهم می‌کنند.

در مجموع، تنوع در رویکردها و تکنیک‌های مورد استفاده در حوزه پیش‌بینی قیمت سهام یک چشم‌انداز پویا از پژوهش‌ها را ارائه می‌دهد. این مطالعات نه‌تنها به نقدهای اساسی مرتبط با چالش‌های پیش‌بینی قیمت سهام پرداخته‌اند؛ بلکه همچنین به راهکارهای نوآورانه و بهبود مدل‌های پیش‌بینی اشاره کرده‌اند. این توسعه‌ها و پیشرفت‌ها نشان‌دهنده تعهد محققان به ارتقاء دائمی در زمینه پیش‌بینی قیمت سهام هستند. بررسی پژوهش‌های پیشین در زمینه پیش‌بینی قیمت سهام نشان می‌دهد که این حوزه شامل طیف گسترده‌ای از روش‌ها و تکنیک‌های مختلف است. این مطالعات علاوه بر اشاره به چالش‌ها و پیچیدگی‌های مرتبط با پیش‌بینی قیمت سهام، بر تلاش‌های مداوم برای توسعه و بهبود مدل‌های پیش‌بینی نیز تاکید دارند.

3. پیش‌زمینه

1.3. یادگیری ماشین

امروزه، یادگیری ماشین به عنوان یک زمینه پژوهشی جدید و پرکاربرد در علوم کامپیوتر شناخته می‌شود. این حوزه با مسائل و مفاهیم گوناگونی همچون پیش‌بینی قیمت سهام، هوش مصنوعی، امور پزشکی و مانند آن، ارتباط نزدیکی دارند. الگوریتم‌های یادگیری ماشین به طور کلی به دو دسته باناظر و بدون ناظر تقسیم می‌شوند. هدف در یادگیری ماشین باناظر، یافتن نگاشت بین ورودی‌ها و خروجی‌ها است و برچسب‌ها برای هر زوج ورودی و خروجی مشخص هستند. این دسته از الگوریتم‌ها بر روی داده‌های برچسب خورده قابل استفاده هستند و در پروژه‌هایی که بر اساس تفکیک دقیق دسته‌ها انجام می‌شوند، مورد استفاده قرار می‌گیرند. در مقابل، یادگیری ماشین بدون ناظر، الگوهایی از داده‌ها را بدون نیاز به برچسب یا خروجی مشخص می‌کند. هدف در این دسته از الگوریتم‌ها، کشف الگوهای مخفی و مفهومی در داده‌ها است. این نوع یادگیری در مواردی که برچسب‌ها دسترسی محدود دارند یا دسترس‌ناپذیر هستند، بسیار مورد توجه قرار می‌گیرد. ترکیب این دو نوع یادگیری ماشین، به دلیل قابلیت کاربردی وسیع در زمینه‌های مختلف و توانایی پیش‌بینی دقیق، موضوع مهمی در

[1] Singular spectrum analysis

پژوهش‌های امروزی است.

2.3. جنگل تصادفی

جنگل تصادفی[1] یکی از الگوریتم‌های موثر در حوزه یادگیری ماشین است که برای مسائل دسته‌بندی و پیش‌بینی استفاده می‌شود. این الگوریتم به صورت مجموعه‌ای از درختان تصمیم عمل می‌کند و از رای‌گیری تصمیم‌گیری هر درخت برای تصمیم نهایی استفاده می‌کند. الگوریتم جنگل تصادفی به دلیل ویژگی‌های زیر جذابیت زیادی دارد:

1. تصادفی بودن ویژگی‌ها: در هر مرحله از ساخت هر درخت تصمیم، تعدادی از ویژگی‌ها به صورت تصادفی انتخاب می‌شوند. این اقدام منجر به افزایش تنوع در مدل می‌گردد و از بروز بیش‌فرضی (اطلاعات غلط یا غیرقابل اعتماد) به داده‌ها جلوگیری می‌نماید. این رویکرد به طور معمول در الگوریتم‌های یادگیری ماشین مانند روش جنگل تصادفی مورد استفاده قرار می‌گیرد.
2. استفاده از بوت‌استرپ نمونه‌ها: در هر مرحله از جنگل تصادفی، برای ساخت هر درخت، نمونه‌هایی از مجموعه داده‌ای به صورت بوت‌استرپ (انتخاب با جایگذاری) انتخاب می‌شوند. این باعث متنوع شدن داده‌ها و کاهش اثر تغییرات تصادفی در مجموعه داده‌ای می‌شود.
3. رای‌گیری (میانگین‌گیری) نتایج: نتایج حاصل از تمام درختان تصمیم به صورت میانگین محاسبه می‌شوند. این کار باعث مقاومت بیشتر مدل در مقابل داده‌های نویزی و پرت می‌شود.
4. پیچیدگی کم: درختان تصمیم فرآیندهای قابل فهم و قابل تفسیر هستند. جنگل تصادفی با ترکیب این درختان به عنوان یک مدل مجموعه‌ای به دست می‌آید که همچنان قابل فهم است.

الگوریتم جنگل تصادفی در انواع مختلفی از مسائل کاربرد دارد، از جمله پیش‌بینی قیمت سهام که در متن پیشین به آن اشاره شده است. با توجه به ویژگی‌های یادگیری تصادفی، این الگوریتم به خوبی با داده‌های پیچیده سازگاری دارد و عملکرد بهتری نسبت به روش‌های سنتی مانند رگرسیون یا درخت تصمیم به دست می‌آورد [24].

3.3. الگوریتم رگرسیون بردار پشتیبان

ماشین بردار پشتیبان یکی از الگوریتم‌های باناظر است که در سال 1998 توسط وپنیک کشف شد. به طور کلی مدل‌های SVM به دو گروه رگرسیون ماشین بردار پشتیبان و مدل دسته‌بند ماشین بردار پشتیبان تقسیم می‌شوند. از مدل دسته‌بندی جهت دسته‌بندی داده‌ها و از مدل رگرسیون جهت پیش‌بینی استفاده می‌شود. اگر مجموعه‌ای از نقاط داده به فرم $\{(x_1, y_1), \dots, (x_i, y_i)\}$ نشان داده شود که در آن $x_i \in R$ ورودی و $y_i \in R$ خروجی متناظر آن باشد، فرم کلی رگرسیون بردار پشتیبان بدین شرح است [24]:

$$min \quad \frac{1}{2}\|w\|^2,$$
$$s.t. \quad |y - \hat{y}| \leq \varepsilon,$$

که در آن w بردار وزن ویژگی‌ها است. در ادامه، هسته‌های مختلف قابل استفاده در رگرسیون ماشین بردار پشتیبان ارائه شده است:

1. هسته گاوسی: هسته گاوسی به عنوان یکی از رایج‌ترین هسته‌ها در مدل‌های پیش‌بینی شناخته می‌شود. عملکرد برتر مدل‌های SVR با هسته گاوسی نسبت به دیگر هسته‌ها نشان‌دهنده توانایی بالاتر در دسته‌بندی داده‌های پیچیده و غیرخطی است. توزیع گاوسی این هسته اختلافات بین داده‌ها را به خوبی توزیع می‌کند و از آن برای مدل‌سازی روابط پیچیده استفاده می‌شود.
2. هسته خطی: هسته خطی مختص داده‌هایی است که رابطه خطی بین ویژگی‌ها و پیش‌بینی وجود دارد. این هسته به مدل امکان یادگیری روابط خطی میان ویژگی‌ها را می‌دهد. در شرایطی که داده‌ها دارای روابط خطی هستند،

[1] Random Forest

مدل‌های با هسته خطی به خوبی عمل می‌کنند.

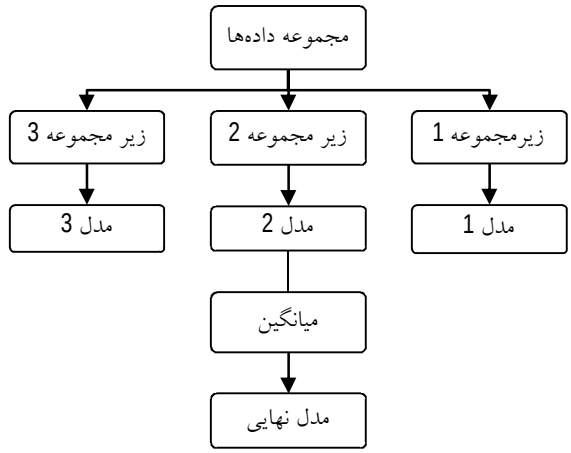
3. هسته چندجمله‌ای: هسته چندجمله‌ای برای مدل‌سازی روابط غیرخطی و پیچیده بین ویژگی‌ها مناسب است. این هسته به مدل امکان می‌دهد تا روابط چندجمله‌ای با ویژگی‌ها را تخمین بزند و برای داده‌های غیرخطی موثر باشد. اهمیت بالایی در مدل‌سازی روابط پیچیده و غیرخطی دارد.

4. هسته سیگموید: هسته‌های سیگموئید در زبان اصطلاحات ریاضی و آماری به مجموعه‌ای از الگوها یا توابع احتمالاتی اشاره دارد که به طور کلی به شکل منحنی‌های «سیگموئید» یا «S-shaped» شناخته می‌شوند. این الگوها به طور معمول در مدل‌های آماری و ریاضی برای نمایش توزیع‌های احتمالاتی، توابع فعال‌سازی در شبکه‌های عصبی و یا در توصیف رفتار و تغییرات در زمان استفاده می‌شوند.

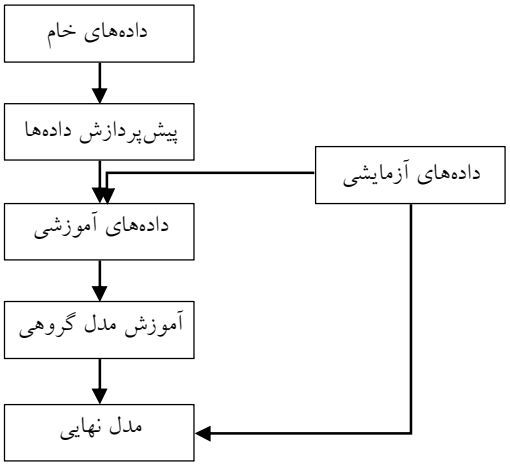
یک هسته سیگموئید به طور معمول به عنوان یک تابع منحنی مانند S از یک متغیر مستقل به یک متغیر وابسته تعریف می‌شود. این توابع اغلب خاصیت‌هایی مانند صعود و نزول یکنواخت در ابتدا و پایان، معکوس‌پذیری و تغییر ناگهانی دارند. از جمله معروف‌ترین موارد استفاده از هسته‌های سیگموئید، مدل‌های رگرسیون لجستیک[1] در آمار و یادگیری ماشین می‌باشند. در این مدل، تابع لجستیک که یک هسته سیگموئید است، برای پیش‌بینی احتمال وقوع یک رویداد یا کلاس بر اساس متغیرهای مستقل استفاده می‌شود.

4.3. الگوریتم رگرسیون بردار پشتیبان گروهی

ایده رگرسیون بردار پشتیبان گروهی از جنگل تصادفی رگرسیون نشات می‌گیرد. در این روش هم ابتدا بر اساس داده‌ها و ایده بوت‌استرپ، چندین رگرسیون بردار پشتیبان تشکیل می‌شود و به هر رگرسیون بردار پشتیبان یک زیرمجموعه تصادفی از داده‌ها با جای‌گذاری و اندازه یکسان ارائه می‌شود. سپس مدل رگرسیون بردار پشتیبان مختلفی آموزش داده می‌شود و در نهایت بین مدل‌های رگرسیون بردار پشتیبان آموزش داده شده، میانگین‌گیری می‌شود و میانگین خروجی‌ها به عنوان مدل نهایی انتخاب می‌شود [25]، [26]. شکل (1) مدل مفهومی رگرسیون بردار پشتیبان گروهی و شکل (2) مدل مفهومی نحوه ورود داده‌ها را نشان می‌دهند.



شکل (1): مدل مفهومی رگرسیون بردار پشتیبان گروهی



شکل (2): مدل مفهومی نحوه ورود داده‌ها

4. روش‌شناسی و شناخت داده‌ها

پیاده‌سازی مدل پیشنهادی با استفاده از برنامه‌نویسی در محیط نرم‌افزار آناکوندا و زبان برنامه‌نویسی پایتون با بهره‌گیری از SciKit-learn انجام شده است. سهم‌های استفاده شده در این پژوهش عبارتند از: شرکت فولاد مبارکه سپاهان (فولاد)، شرکت

[1] Logistic Regression

سرمایه‌گذاری تامین اجتماعی (شستا)، شرکت صنایع پتروشیمی خلیج فارس (فارس)، شرکت ملی صنایع مس ایران (فملی)، شرکت پالایش نفت اصفهان (شپنا)، پالایش نفت تهران (شتران)، پالایش نفت بندرعباس (شبندر)، ایران خودرو (خودرو)، سایپا (خساپا) و پتروشیمی پارس (پارس) که در محدوده زمانی 1400/01/07 تا 1403/04/31است که دلیل این انتخاب تاثیر گذاری 10 سهم (شاخص‌ساز) در 3 سال گذشته بر روی شاخص بورس اوراق بهادار بوده‌اند[1]. این داده‌ها شامل ویژگی‌های بالاترین قیمت، پایین‌ترین قیمت، قیمت باز شدن و قیمت بسته شدن سهام مذکور در بازه زمانی مورد نظر است.

1.4. آماده‌سازی داده‌های ورودی

در این گام به نرمال‌سازی داده‌ها پرداخته می‌شود. نرمال کردن داده‌ها، فرآیندی است که در آن داده‌ها به فرم مناسبی برای الگوریتم‌ها تبدیل می‌شوند. برای نرمال کردن داده‌ها روش‌های مختلفی وجود دارد که به طور کلی پس از استانداردسازی، داده‌ها در بازه $[-1,1]$ قرار می‌گیرند. در این مقاله، از روش زیر برای استانداردسازی استفاده شده است:

$$NRM = \frac{X - \mu}{\sigma},$$

که در فرمول بالا X داده مورد نظر، μ میانگین و σ انحراف معیار را تصریح می‌کند.

2.4. طراحی مدل

ابتدا داده‌ها به دو مجموعه آموزشی و آزمایشی تقسیم می‌شود. اغلب پژوهشگران 80 درصد داده‌ها را به عنوان داده‌های آموزشی و 20 درصد باقیمانده را به عنوان داده‌های آزمایشی در نظر می‌گیرند. تعداد نمونه‌های هر یک از سهم‌ها 700 روز است که 560 روز اول به منظور آموزش و 140 روز دیگر برای آزمایش در نظر گرفته می‌شود. سپس به کمک روش جستجوی شبکه‌ای، پارامترهای مهم مدل‌ها تخمین زده می‌شوند و بعد از اعمال پارامترهای بهینه، مدل آموزش داده می‌شود. در انتها، عملکرد مدل‌ها بر اساس نتایج داده‌های آزمایشی و معیارهای ارزیابی که به شرح زیر است، بررسی می‌شود:

1. میانگین مربعات خطا با رابطه محاسبه می‌شود:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2.$$

2. جذر میانگین مربعات خطا با رابطه تعیین می‌شود و در واقع جذر MSE است:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}.$$

3. میانگین خطای مطلق که با رابطه زیر به دست می‌آید:

$$MAE = \frac{\sum_{i=1}^{N}|y_i - \hat{y}_i|}{N}.$$

4. معیار ارزیابی R^2 که ارتباط بین دو متغیر را اندازه‌گیری می‌کند و به ضریب تعیین شناخته می‌شود همچنین هر چقدر این معیار به یک نزدیک‌تر باشد، مدل عملکرد بهتری خواهد داشت.

در روابط بالا، N، y_i و $\hat{y}_i$ به ترتیب تعداد کل مشاهدات، مقدار واقعی و مقدار پیش‌بینی شده متناظر را نمایش می‌دهند، همچنین هر اندازه مقدار این معیارها به صفر نزدیک باشند، مدل از عملکرد بهتری برخوردار است.

5. برآورد مدل و نتایج

در این بخش، مدل پیشنهادی بر روی داده‌های آموزشی هر یک از ده نماد مورد نظر اعمال و نتیجه‌ی ارزیابی پیش‌بینی مدل بر روی داده‌های آزمایشی در جداول (1) تا (10) گزارش شده است. همچنین به منظور مقایسه، مدل‌های حاصل از روش SVR معمولی، درخت تصمیم و جنگل تصادفی نیز بر روی داده‌های آموزشی به دست آمدند و نتایج ارزیابی پیش‌بینی آنها در جداول (1) تا (10) آورده شده است.

با توجه به نتایج به دست آمده در مورد نماد فولاد که در جدول (1) گزارش شده است، عملکرد SVR معمولی با هسته خطی و

[1] www.tsetmc.com

همچنین روش SVR گروهی با هسته خطی با تفاوت بسیار اندکی دارای بهترین عملکرد نسبت به سایر روش‌ها و هسته‌ها هستند و پس از آنها روش جنگل تصادفی که یک روش گروهی است در رتبه دوم عملکرد قرار می‌گیرد. همچنین در مورد تمام هسته‌ها، SVR گروهی نسبت به SVR معمولی در برخی موارد بهبود ایجاد کرده است.

جدول (1): ارزیابی و مقایسه روش پیشنهادی و سایر روش‌های بر روی داده‌های نماد فولاد.

	روش	R^2	MAE	RMSE	MSE
SVR معمولی	گاوسی	0,9333	0,0512	0,0524	0,0027
	خطی	0,9987	0,0046	0,0071	0,000005
	چندجمله‌ای	0,9356	0,0503	0,0515	0,0026
	سیگموید	0,8781	0,0694	0,0708	0,005
SVR گروهی	گاوسی	0,9289	0,0529	0,0541	0,0029
	خطی	0,9987	0,0046	0,0072	0,000005
	چندجمله‌ای	0,9271	0,0535	0,0547	0,003
	سیگموید	0,8811	0,0683	0,0699	0,0048
درخت تصمیم		0,9929	0,0128	0,017	0,0002
جنگل تصادفی		0,9970	0,0084	0,011	0,0001

در مورد نماد شستا که نتایج ارزیابی در جدول (2) آورده شده است، روش SVR گروهی با هسته خطی بهترین عملکرد را دارد و با اختلاف خوبی، SVR معمولی با هسته‌ی خطی و بعد از آن روش جنگل تصادفی قرار می‌گیرد. هسته‌های گاوسی، چندجمله‌ای و سیگموئید عملکرد متوسطی از خود نشان داده‌اند. در مورد سایر هسته‌ها، اختلاف اندکی بین SVR معمولی و SVR گروهی وجود دارد. جنگل تصادفی نیز به عنوان یک مدل گروهی، عملکرد خوبی در پیش‌بینی سهام شستا دارد. از اینرو، استفاده از مدل‌های ماشین بردار گروهی با هسته خطی و جنگل تصادفی به نظر می‌رسد که دارای عملکرد بهتری برای پیش‌بینی سهام شستا هستند.

در جدول (3) که نتایج ارزیابی روش‌های مورد مطالعه در این مقاله بر روی داده‌های نماد فارس را در خود دارد، SVR معمولی و SVR گروهی با هسته‌های خطی و چندجمله‌ای دارای بهترین عملکرد هستند. همچنین هسته گاوسی و پس از آن جنگل قرار می‌گیرند.

جدول (2): ارزیابی و مقایسه روش پیشنهادی و سایر روش‌های بر روی داده‌های نماد شستا.

	روش	R^2	MAE	RMSE	MSE
SVR معمولی	گاوسی	-2,0644	0,027	0,0301	0,0009
	خطی	0,9954	0,0009	0,0011	0,0000001
	چندجمله‌ای	-27,866	0,0924	0,0924	0,0085
	سیگموید	-0,8622	0,0193	0,0234	0,0005
SVR گروهی	گاوسی	-2,0723	0,0269	0,0301	0,0009
	خطی	0,9968	0,0006	0,0009	0,00000009
	چندجمله‌ای	-27,866	0,0924	0,0924	0,0085
	سیگموید	-1,006	0,02	0,0243	0,0005
درخت تصمیم		0,9823	0,0017	0,0022	0,0000005
جنگل تصادفی		0,9927	0,0011	0,0014	0,0000002

جدول (3): ارزیابی و مقایسه روش پیشنهادی و سایر روش‌های بر روی داده‌های نماد فارس.

	روش	R^2	MAE	RMSE	MSE
SVR معمولی	گاوسی	0,9992	0,0197	0,0247	0,0006
	خطی	0,9995	0,0146	0,0196	0,0003
	چندجمله‌ای	0,9994	0,0172	0,0224	0,0005
	سیگموید	0,9974	0,0373	0,0466	0,0021
SVR گروهی	گاوسی	0,9992	0,0191	0,0243	0,0005
	خطی	0,9995	0,0149	0,020	0,0004
	چندجمله‌ای	0,9995	0,0147	0,0193	0,0003
	سیگموید	0,9973	0,0377	0,0468	0,0021
درخت تصمیم		0,9982	0,0317	0,0388	0,0015
جنگل تصادفی		0,9988	0,0251	0,0311	0,0009

نماد فملی که نتایج ارزیابی روش‌ها بر روی داده‌های آن در جدول (4) آمده است نیز همچنان SVR معمولی و گروهی با

هسته خطی بهترین عملکرد را از خود نشان می‌دهند و بعد از آنها جنگل تصادفی و درخت تصمیم قرار می‌گیرند.

جدول (4): ارزیابی و مقایسه روش پیشنهادی و سایر روش‌های بر روی داده‌های نماد فملی.

	روش	R^2	MAE	RMSE	MSE
SVR معمولی	گاوسی	0,9491	0,0357	0,0384	0,0014
	خطی	0,9962	0,0073	0,0103	0,0001
	چندجمله‌ای	0,9407	0,0394	0,0414	0,0017
	سیگموید	0,868	0,06	0,0618	0,0038
SVR گروهی	گاوسی	0,939	0,0395	0,042	0,0017
	خطی	0,9962	0,0072	0,0103	0,0001
	چندجمله‌ای	0,931	0,0427	0,0447	0,002
	سیگموید	0,8685	0,06	0,0617	0,0038
درخت تصمیم		0,9867	0,0141	0,0189	0,0003
جنگل تصادفی		0,09891	0,0133	0,0177	0,0003

نتایج ارزیابی روش‌ها بر روی داده‌های نماد شپنا که در جدول (5) آمده است نشان می‌دهد که SVR گروهی با هسته خطی بهترین عملکرد را دارد و پس از آن با اختلاف به نسبت خوبی SVR معمولی با هسته خطی قرار می‌گیرد.

جدول (5): ارزیابی و مقایسه روش پیشنهادی و سایر روش‌های بر روی داده‌های نماد شپنا.

	روش	R^2	MAE	RMSE	MSE
SVR معمولی	گاوسی	0,9945	0,0348	0,0422	0,0017
	خطی	0,9993	0,0103	0,0142	0,0002
	چندجمله‌ای	0,9972	0,0283	0,0301	0,0009
	سیگموید	0,9968	0,0268	0,0319	0,001
SVR گروهی	گاوسی	0,9932	0,0395	0,0468	0,0021
	خطی	0,9995	0,0092	0,0124	0,0001
	چندجمله‌ای	0,9983	0,0202	0,0229	0,0005
	سیگموید	0,9972	0,0249	0,0300	0,0009
درخت تصمیم		0,996	0,0216	0,0358	0,0012
جنگل تصادفی		0,9952	0,0184	0,0392	0,0015

در مورد نماد خودرو، براساس نتایج مندرج در جدول (6) به وضوح برتری SVR گروهی با هسته خطی بر سایر روش‌های ارایه شده در این مقاله مشاهده می‌شود و بعد از آن روش‌های SVR گروهی و معمولی با هسته گاوسی و چندجمله‌ای قرار می‌گیرند. در مورد این نماد، جنگل تصادفی و درخت تصمیم در رده‌های آخر قبل از هسته سیگموید قرار می‌گیرند.

جدول (6): ارزیابی و مقایسه روش پیشنهادی و سایر روش‌های بر روی داده‌های نماد خودرو.

	روش	R^2	MAE	RMSE	MSE
SVR معمولی	گاوسی	0,9935	0,0195	0,0269	0,0007
	خطی	0,9929	0,02	0,0283	0,0008
	چندجمله‌ای	0,9911	0,0237	0,0317	0,001
	سیگموید	0,9273	0,0839	0,0909	0,0082
SVR گروهی	گاوسی	0,9932	0,0214	0,0276	0,0007
	خطی	0,9938	0,0189	0,0263	0,0006
	چندجمله‌ای	0,9900	0,0263	0,0336	0,0011
	سیگموید	0,9606	0,0603	0,0668	0,0044
درخت تصمیم		0,9797	0,0368	0,0479	0,0022
جنگل تصادفی		0,9775	0,0397	0,0505	0,0025

نمادهای پارس و شبندر در جدول (7) و (8) به کمک مدل SVR معمولی و گروهی با هسته خطی برترین عملکرد را رقم زده‌اند. همچنین جنگل تصادفی و درخت تصمیم نیز با اختلاف قابل توجهی در رده‌های بعدی قرار می‌گیرند. در مورد نماد شبندر شاهد بهبود عملکرد SVR گروهی نسبت به SVR معمولی هستیم.

نمادهای شتران و خساپا در جدول (9) و (10) نیز همچون اکثر نمادهای فوق بهترین عملکرد را با SVR معمولی و گروهی توسط هسته خطی کسب کرده‌اند. در مورد نماد شتران پس از هسته خطی، SVR گروهی با هسته چندجمله‌ای و نماد خساپا، روش SVR معمولی و گروهی با هسته گاوسی و چندجمله‌ای بهترین عملکرد را از خود نشان داده‌اند.

جدول (7): ارزیابی و مقایسه روش پیشنهادی و سایر روش‌های بر روی داده‌های نماد پارس.

	روش	R^2	MAE	RMSE	MSE
SVR معمولی	گاوسی	-2674,1	0,0793	0,0795	0,0063
	خطی	0,9731	0,0001	0,0002	0,000000006
	چندجمله‌ای	-4106,4	0,0991	0,0991	0,0098
	سیگموید	-1244,5	0,0514	0,0546	0,0029
SVR گروهی	گاوسی	-2692,2	0,08	0,0802	0,0064
	خطی	0,9734	0,0001	0,0002	0,000000006
	چندجمله‌ای	-4106,4	0,0991	0,0991	0,0098
	سیگموید	-1322,4	0,0533	0,0562	0,0031
درخت تصمیم		0,8952	0,0004	0,0005	0,00000002
جنگل تصادفی		0,9547	0,0002	0,0003	0,00000001

جدول (8): ارزیابی و مقایسه روش پیشنهادی و سایر روش‌های بر روی داده‌های نماد شبندر.

	روش	R^2	MAE	RMSE	MSE
SVR معمولی	گاوسی	0,9759	0,0263	0,0305	0,0009
	خطی	0,9944	0,0082	0,0145	0,0002
	چندجمله‌ای	0,938	0,0468	0,0489	0,0023
	سیگموید	0,5918	0,1102	0,1256	0,0157
SVR گروهی	گاوسی	0,9689	0,0302	0,0346	0,0011
	خطی	0,9945	0,0082	0,0145	0,0002
	چندجمله‌ای	0,9677	0,032	0,0353	0,0012
	سیگموید	0,6427	0,1024	0,1175	0,0138
درخت تصمیم		0,9847	0,0172	0,0242	0,0005
جنگل تصادفی		0,9876	0,0155	0,0218	0,0004

نتایج اعمال روش پیشنهادی بر روی ده نماد مورد بررسی در این مقاله نشان می‌دهند که SVR معمولی و گروهی با هسته خطی در اکثر نمادها بهترین عملکرد در پیش‌بینی را به خود اختصاص داده‌اند و بنابراین استفاده از این هسته در روش پیشنهادی توصیه می‌شود. لازم به ذکر است که این توصیه بر مبنای مشاهدات این ده نماد در بازه زمانی مورد نظر در این مقاله است و این نتیجه‌گیری قطعی و حتمی نخواهد بود. از طرف دیگر، نتایج نشان می‌دهند که روش SVR معمولی با SVR گروهی از لحاظ معیارهای ارزیابی مطرح شده در این مقاله با اختلاف خیلی کمی از یکدیگر در اکثر موارد فاصله دارند که قابل چشم‌پوشی است.

جدول (9): ارزیابی و مقایسه روش پیشنهادی و سایر روش‌های بر روی داده‌های نماد شتران.

	روش	R^2	MAE	RMSE	MSE
SVR معمولی	گاوسی	0,8818	0,0978	0,1141	0,013
	خطی	0,9993	0,0052	0,0081	0,000006
	چندجمله‌ای	0,9537	0,0705	0,0714	0,0051
	سیگموید	0,9056	0,0941	0,1020	0,0104
SVR گروهی	گاوسی	0,859	0,0994	0,1246	0,0155
	خطی	0,9993	0,0052	0,0082	0,000006
	چندجمله‌ای	0,9727	0,0537	0,0548	0,003
	سیگموید	0,9278	0,0826	0,0892	0,0079
درخت تصمیم		0,8209	0,0981	0,1405	0,0197
جنگل تصادفی		0,7921	0,1044	0,1514	0,0229

جدول (10): ارزیابی و مقایسه روش پیشنهادی و سایر روش‌های بر روی داده‌های نماد خساپا.

	روش	R^2	MAE	RMSE	MSE
SVR معمولی	گاوسی	0,9885	0,0306	0,0373	0,0013
	خطی	0,9921	0,0211	0,0308	0,0009
	چندجمله‌ای	0,9889	0,0286	0,0367	0,0013
	سیگموید	0,9032	0,1016	0,1085	0,0117
SVR گروهی	گاوسی	0,989	0,0292	0,0365	0,0013
	خطی	0,9922	0,0212	0,0308	0,0009
	چندجمله‌ای	0,9869	0,033	0,0398	0,0015
	سیگموید	0,9262	0,0879	0,0947	0,0089
درخت تصمیم		0,9674	0,0446	0,0629	0,0039
جنگل تصادفی		0,9826	0,0335	0,046	0,0021

از آنجا که به کمک فناوری پردازش موازی، هزینه محاسباتی روش SVR گروهی می‌تواند به طور قابل ملاحظه‌ای از روش SVR معمولی کمتر باشد، لذا استفاده از روش SVR گروهی با معیارهای ارزیابی تقریبا مشابه روش معمولی به‌صرفه‌تر و ارزشمندتر خواهد بود.

6. نتیجه‌گیری و کارهای آتی

روش پیشنهادی با استفاده از ترکیب بگینگ و SVR نشان داده است که چگونه می‌تواند دقت پیش‌بینی قیمت سهام را افزایش دهد. این روش نه‌تنها در مقابل تغییرات بازار مقاومت بیشتری ایجاد می‌کند، بلکه قابلیت تعمیم به شرایط جدید را نیز داراست. این مدل می‌تواند به عنوان یک ابزار قوی در تصمیم‌گیری‌های سرمایه‌گذاری مورد استفاده قرار گیرد و مسیرهای پژوهشی جدیدی را برای توسعه آینده این حوزه ارائه دهد. با اینکه این مقاله سعی در پاسخ به چالش‌های مهم در حوزه پیش‌بینی قیمت سهام داشته است، اما هنوز فرصت‌های بسیاری برای تحقیقات بیشتر و بهبود روش‌های پیشین وجود دارد. در این مقاله، یک روش جدید برای پیش‌بینی قیمت سهام با استفاده از ترکیب مدل‌های SVR با بگینگ معرفی شده است. بر اساس نتایج به دست آمده از تجزیه و تحلیل داده‌های سهام مختلف شامل فولاد، شستا، فارس، فملی، شپنا، شتران، شبندر، خودرو، خساپا و پارس با استفاده از مدل‌های SVR معمولی و گروهی و جنگل تصادفی، می‌توان گفت مدل‌های SVR معمولی و گروهی با هسته خطی در بیشتر موارد، به عنوان مدل‌های با عملکرد برتر هستند.

برای تحقیقات آینده، می‌توان پژوهش بیشتر در زمینه بهینه‌سازی پارامترها و افزودن ویژگی‌های جدید به مدل را مدنظر قرار داد. این اقدامات می‌توانند به بهبود عملکرد مدل و افزایش دقت در پیش‌بینی‌ها کمک کنند. همچنین، تطبیق مدل با محیط‌های پویا و نوسانی بازار یک چالش مهم است که ممکن است تاثیر مستقیم بر قابلیت پیش‌بینی مدل داشته باشد. ترکیب بگینگ و SVR در این مقاله به خوبی نشان داده است که چگونه این چالش می‌تواند با استفاده از چندین مدل و ترکیب پیش‌بینی‌ها کاهش یابد. این روش نه‌تنها دقت بالاتری در پیش‌بینی ارائه می‌دهد؛ بلکه مدل را مقاومت بیشتری در برابر تغییرات سریع بازار به ارمغان می‌آورد [27]. پژوهش‌های آتی می‌توانند به سمت بهینه‌سازی پارامترهای مدل، ارتقای تکنیک‌های بگینگ و استفاده از داده‌های نوآورانه برای آموزش مدل ادامه یابند. همچنین، بررسی تاثیر ترکیب مدل‌های دیگر به‌جای SVR نیز می‌تواند یک جنبه جالب برای پژوهش‌های آینده باشد و به سرمایه‌گذاران امکان می‌دهد تا با در نظر گرفتن تغییرات وضعیت بازار، تصمیمات بهتری اتخاذ کنند. در تحقیقات آتی، می‌توان مدل‌های دیگری را نیز در مقایسه با مدل‌های حاصل از این مطالعه در نظر گرفت. از جمله مدل‌های عمیق یادگیری و شبکه‌های عصبی می‌توانند موضوعات مناسبی برای بررسی باشند. تحلیل تفصیلی‌تر نتایج و تاثیرپذیری از عوامل مختلف مانند حجم معاملات، شاخص‌های اقتصادی و اخبار بازار می‌تواند به بهترین استفاده از این مدل‌ها کمک کند. انجام تحقیقات بر روی دوره‌های آتی بازار سهام به منظور پیش‌بینی عملکرد مدل‌ها در آینده و تعیین احتمال موفقیت آنها می‌تواند مفید باشد. تلاش برای بهبود عملکرد مدل‌ها از طریق بهینه‌سازی پارامترها، افزایش حجم داده‌ها و استفاده از تکنیک‌های پیشرفته یادگیری ماشین می‌تواند تحقیقات آینده را بهبود بخشد. مطالعات آینده می‌توانند معیارهای جدیدی را در تجزیه و تحلیل ارزیابی عملکرد مدل‌ها معرفی کنند تا نتایج بهتر و مطمئن‌تری حاصل شود. تحقیقات بیشتر در این حوزه می‌تواند به بهبود روش‌ها و افزایش دقت پیش‌بینی در بازار سهام کمک کند و تصمیم‌گیری‌های مبتنی بر داده را تسهیل نماید.

تعارض منافع: نویسندگان اعلام می‌کنند که هیچ تعارض منافعی ندارند.

مراجع

[1] R. Oprea and O. Stoica, “Capital Markets Integration and Economic Growth,” Montenegrin J. Econ., vol. 14, no. 3, pp. 23-35, 2018, doi: 10.14254/1800-5845/2018.14-3.2.

[2] M. Nooshyar and A. Ghasemi Marzbali, “Day-Ahead Electricity Price Forecasting by a New Hybrid Algorihtm Based on ELM, Curvelet Transform, Preprocessing System, and Modified VCS Algorithm,” Comput. Intell. Elect. Eng., vol. 10, no. 2, pp. 73-86, 2019, doi: 10.22108/isee.2019.116627.1215 [In Persian].

[3] S. Bajalan, S. Fallahpour, and N. Dana, “The Stock Trend Prediction Using Volume Weighted Support Vector Machine with a Hybrid Feature Selection Method to Predict the Stock Price Trend in Tehran Stock Exchange,” Financ. Manage. Strategy, vol. 4, no. 3, pp. 121-148, 2016, doi: 10.22051/jfm.2016.2575 [In Persian].

[4] H. R. Kordlouie and F. Teimouri, “Comparative Study of the Combination of Gaussian Models and Support Vector Machines in Detecting and Predicting Price Bubbles,” Financ. Eng. Portf. Manage., vol. 6, no. 23, pp. 79-104, 2015, dor: 20.1001.1.22519165.1394.6.23.5.3 [In Persian].

[5] Q. Ilyas, K. Iqbal, S. Ijaz, A. Mehmood, and S. Bhatia, “A Hybrid Model to Predict Stock Closing Price Using Novel Features and a Fully Modified Hodrick–Prescott Filter,” Electronics, vol. 11, no. 21, p. 3588, 2022, doi: 10.3390/electronics11213588.

[6] H. Huang, W. Zhang, G. Deng, and J. Chen, “Predicting Stock Trend Using Fourier Transform and Support Vector Regression,” in Proc. IEEE 17th Int. Conf. Comput. Sci. Eng. (CSE), 2014, pp. 213-216, doi: 10.1109/CSE.2014.70.

[7] S. A. Basher, A. A. Haug, and P. Sadorsky, “The impact of economic policy uncertainty and commodity prices on CARB country stock market volatility,” MPRA Paper 96577, Univ. Library of Munich, Germany, 2019. [Online]. Available: https://ideas.repec.org/p/pra/mprapa/96577.html

[8] N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosifidis, “Temporal Bag-of-Features Learning for Predicting Mid Price Movements Using High Frequency Limit Order Book Data,” IEEE Trans. Emerg. Topics Comput. Intell., vol. 4, no. 6, pp. 774-785, 2020, doi: 10.1109/TETCI.2018.2872598.

[9] M. Obthong, N. Tantisantiwong, W. Jeamwatthanachai, and G. Wills, “A Survey on Machine Learning for Stock Price Prediction: Algorithms and Techniques,” in Proc. 12th Int. Joint Conf. Comput. Intell. (IJCCI), 2020, pp. 63-71, doi: 10.5220/0009340700630071.

[10] P. Sadorsky, “A Random Forests Approach to Predicting Clean Energy Stock Prices,” J. Risk Financ. Manage., vol. 14, no. 2, p. 48, 2021, doi: 10.3390/jrfm14020048.

[11] B. Eguz, F. E. Corbaci, and T. Kaya, “Stock Price Prediction of Turkish Banks Using Machine Learning Methods,” in Intelligent and Fuzzy Techniques for Emerging Conditions and Digital Transformation, C. Kahraman et al., Eds., vol. 308, Cham, Switzerland: Springer, 2022, pp. 222-229, doi: 10.1007/978-3-030-85577-2_26.

[12] P. Samad, S. Mutalib, and S. Rahman, “Analytics of Stock Market Prices Based on Machine Learning Algorithms,” Indones. J. Electr. Eng. Comput. Sci., vol. 16, no. 2, pp. 1050-1058, 2019, doi: 10.11591/ijeecs.v16.i2.pp1050-1058.

[13] H. Liu, L. Qi, and M. Sun, “Short-Term Stock Price Prediction Based on CAE-LSTM Method,” Wireless Commun. Mobile Comput., vol. 2022, p. 4809632, 2022, doi: 10.1155/2022/4809632.

[14] Z. Rustam and P. Kintandani, “Application of Support Vector Regression in Indonesian Stock Price Prediction with Feature Selection Using Particle Swarm Optimisation,” Modelling Simul. Eng., vol. 2019, p. 8962717, 2019, doi: 10.1155/2019/8962717.

[15] M. A. Monadi and A. Najafi, “A Prediction-Based Portfolio Optimization Model Using Support Vector Regression,” Financ. Eng. Portf. Manage., vol. 14, no. 55, pp. 252-270, 2023 [In Persian].

[16] B. Dehghan Khanghahi, J. Bahrisales, S. Jabbarzadeh Kangarlouie, and A. Ashtab, “The Comparative Study of the Accuracy of Prediction of Support Vector Machine, Bayesian Network and C5 Models in Prediction Underpricing for Listed Companies at TSE and OTC,” Financ. Eng. Portf. Manage., vol. 11,

no. 44, pp. 95-113, 2020, dor: 20.1001.1.22519165.1399.11.44.5.0 [In Persian].

[17] A. Khosravinejad and M. Shabani Sadr Pisheh, "Evaluation of Linear and Non-linear Models in Predicting the Stock Price Index in Tehran Stock Exchange," Financ. Econ., vol. 8, no. 27, pp. 51-64, 2014, dor: 20.1001.1.25383833.1393.8.27.4.2 [In Persian].

[18] S. A. Monadjemi, M. Abzari, and A. Rayati Shavazi, "Modeling of Stock Price Forecasting in Stock Exchange Market, using Fuzzy Neural Networks and Genetic Algorithms," J. Quant. Econ. (JQE), vol. 6, no. 3, pp. 1-26, 2009 [In Persian], doi: 10.22055/jqe.2009.10697 [In Persian].

[19] M. Ramazani and A. Ameli, "Forecasting of Stock Price Using Fuzzy Neural Network Based on GA and Compaision with Fuzzy Neural Network," Econ. Modeling Res., vol. 6, no. 22, pp. 61-91, 2015, doi: 10.18869/acadpub.jemr.6.22.61 [In Persian].

[20] M. Zamani, A. Afsar, S. V. Saghafi, and A. Bayat, "An Expert System for Stock Price Prediction and Portfolio Optimization Using Fuzzy Neural Networks, Fuzzy Modeling, and Genetic Algorithms," Financ. Eng. Portf. Manage., vol. 5, no. 21, pp. 107-130, 2014, dor: 20.1001.1.22519165.1393.5.21.7.4 [In Persian].

[21] I. Nti, A. Adekoya, and B. Weyori, "Efficient Stock-Market Prediction Using Ensemble Support Vector Machine," Open Comput. Sci., vol. 10, no. 1, pp. 154-163, 2020, doi: 10.1515/comp-2020-0199.

[22] A. Li, Q. Wei, Y. Shi, and Z. Liu, "Research on Stock Price Prediction from a Data Fusion Perspective," Data Sci. Finance Econ., vol. 3, no. 3, pp. 230-250, 2023, doi: 10.3934/DSFE.2023014.

[23] F. Kalhori and S. M. Hoseini, "Clustering of the Stock Price Using Minimum Spanning Tree," Soft Comput. J., vol. 14, no. 1, pp. 72-87, 2025, doi: 10.22052/scj.2024.253526.1183 [In Persian].

[24] M. Ebtia, S. M. Hoseini, and R. Khochiani, "Credit Rating of Bank Customers Using a New Ensemble Method Based on Support Vector Machine: A Case Study of Pasargad Bank," Soft Comput. J., vol. 10, no. 2, pp. 2-15, 2022, doi: 10.22052/scj.2022.243227.1016 [In Persian].

[25] S. M. Hoseini, M. Ebtia, and R. Khochiani, "An Ensemble Method Based on Bagging SVM for Credit Rating Problem," Soft Comput. J., Accepted, 2024, doi: 10.22052/scj.2024.246330.1063 [In Persian].

[26] A. Smola and B. Scholkopf, "A Tutorial on Support Vector Regression," Statist. Comput., vol. 14, no. 3, pp. 199-222, 2004, doi: 10.1023/B:STCO.0000035301.49549.88.

[27] M. Sedighi, H. Jahangirnia, M. Gharakhani, and S. Farahani Fard, "A Novel Hybrid Model for Stock Price Forecasting Based on Metaheuristics and Support Vector Machine," Data, vol. 4, no. 2, p. 75, 2019, doi: 10.3390/data4020075.