Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
DOI: 10.22034/jscit.2025.506154.2107
Volume 14, Number 2, Summer 2025, pp. 45-59
Received:02/14/2025, Revised:06/13/2025, Accepted: 07/14/2025

Using Machine Learning Methods to Predict Stock Returns Considering the Impact of Price Fluctuation Range Changes

Fateme Hosseini[1], Narges Norouzi[2*]
1- Department of Industrial Engineering, NT.C., Islamic Azad University, Tehran, Iran.
2*- Department of Industrial Engineering, NT.C., Islamic Azad University, Tehran, Iran.
[1]f.hosseini1000@gmail.com, [2*] nargesnorouzi@iau.ac.ir

Corresponding author's address: Narges Norouzi, Department of Industrial Engineering, NT.C., Islamic Azad University, Tehran, Iran.

Abstract- This study investigates stock price prediction in the Iran Farabourse using machine learning methods, specifically Random Forest and Support Vector Regression. The primary objective is to improve prediction accuracy and examine the effect of changes in the price fluctuation range on model performance. To this end, data from four selected industrial stocks listed on the first tier of the Iran Farabourse were collected over two distinct time periods: 2015–2019 (for model evaluation under a stable fluctuation regime) and 2020–2022 (to assess the impact of fluctuation range changes). The input features include open price, Simple Moving Average (SMA), and Relative Strength Index (RSI), while the output is the closing price. After preprocessing and normalization, models were trained using the specified algorithms. The models' performance was evaluated using standard metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). In the first period, the Random Forest model demonstrated superior performance with an MSE of 162,217.18, MAE of 156.01, and R^2 of 1.00, compared to the Support Vector Regression model with an MSE of 355,674,554.14 and R^2 of 0.62. In the second period, although prediction errors increased in both models, Random Forest maintained better performance (MSE: 901,875.77; MAPE: 1.98%) than Support Vector Regression (MSE: 82,930,332.00; MAPE: 16.60%). A slight decrease in the R^2 value of Random Forest from 1.00 to 0.99 was also observed. These findings indicate that Random Forest consistently outperforms Support Vector Regression in both time frames and that changes in the price fluctuation range negatively affect prediction accuracy in both models. Overall, machine learning approaches can serve as effective tools for analyzing capital market dynamics, reducing investment risk, and supporting investors in making more informed financial decisions.

Keywords- Stock Returns, Machine Learning, Price Fluctuation Limits, Support Vector Regression, Random Forest

مجله علمی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
شناسه دیجیتال: 10.22034/jscit.2025.506154.2107
جلد ۱۴، شماره ۲، تابستان ۱۴۰۴، صفحه ۴۵-۵۹
دریافت: ۱۴۰۳/۱۱/۲۶، بازنگری: ۱۴۰۴/۰۳/۲۳، پذیرش: ۱۴۰۴/۰۴/۲۳

استفاده از روش‌های یادگیری ماشین جهت پیش بینی بازده سهام با در نظر گرفتن تاثیر تغییر دامنه نوسان قیمت

فاطمه حسینی [۱]، نرگس نوروزی [۲*]
۱- گروه مهندسی صنایع، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران.
*۲- گروه مهندسی صنایع، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران.
[1]f.hosseini1000@gmail.com, [2*] nargesnorouzi@iau.ac.ir

* نشانی نویسنده مسئول: نرگس نوروزی، گروه مهندسی صنایع، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران

چکیده- این پژوهش به بررسی پیش‌بینی قیمت سهام در فرابورس ایران با بهره‌گیری از روش‌های یادگیری ماشین، شامل جنگل تصادفی و رگرسیون بردار پشتیبان، می‌پردازد. هدف مطالعه، پیش‌بینی دقیق‌تر قیمت سهام و تحلیل تأثیر تغییرات دامنه نوسان قیمت بر عملکرد مدل‌هاست. برای این منظور، داده‌های چهار نماد منتخب از سهام‌های صنعتی بازار پایه اول فرابورس ایران در دو بازه زمانی سال‌های ۱۳۹۴ تا ۱۳۹۹ (جهت ارزیابی دقت مدل‌ها در شرایط بدون تغییر دامنه) و سال‌های ۱۴۰۰ تا ۱۴۰۲ (جهت بررسی اثر تغییر دامنه نوسان) گردآوری شده‌اند. متغیرهای ورودی شامل قیمت باز، میانگین متحرک ساده و شاخص قدرت نسبی و متغیر خروجی، قیمت بسته بوده است. داده‌ها پس از اعمال روش‌های پیش‌پردازش و نرمال‌سازی، با استفاده از الگوریتم‌های یادشده آموزش داده شدند. عملکرد مدل‌ها با معیارهایی چون میانگین مربعات خطا، جذر آن، میانگین قدر مطلق خطا، میانگین درصد خطای مطلق و ضریب تعیین ارزیابی شد. در بازه اول، مدل جنگل تصادفی با میانگین مربعات خطای ۱۶۲٬۲۱۷٫۱۸، میانگین قدر مطلق خطای ۱۵۶٫۰۱ و ضریب تعیین ۱، عملکرد بهتری نسبت به مدل رگرسیون بردار پشتیبان با خطای ۳۵۵٬۶۷۴٬۵۵۴٫۱۴ و ضریب تعیین ۰٫۶۲ داشت. در بازه دوم، اگرچه میزان خطا در هر دو مدل افزایش یافت، جنگل تصادفی همچنان برتری خود را حفظ کرد (خطای ۹۰۱٬۸۷۵٫۷۷ و میانگین درصد خطای ۱٫۹۸٪) در مقابل رگرسیون بردار پشتیبان (با خطای ۸۲٬۹۳۰٬۳۳۲٫۰۰ و درصد خطای ۱۶٫۶۰٪). کاهش اندک ضریب تعیین مدل جنگل تصادفی از ۱ به ۰٫۹۹ نیز مشاهده شد. نتایج نشان می‌دهد جنگل تصادفی در هر دو دوره عملکرد بهتری دارد و تغییر دامنه نوسان قیمت موجب کاهش دقت پیش‌بینی در هر دو مدل شده است. در مجموع، روش‌های یادگیری ماشین می‌توانند ابزاری کارآمد برای تحلیل بازار سرمایه، کاهش ریسک و بهینه‌سازی تصمیم‌گیری سرمایه‌گذاران باشند.

واژه‌های کلیدی: بازده سهام، یادگیری ماشین، دامنه نوسان قیمت، رگرسیون بردار پشتیبان، جنگل تصادفی.

۱- مقدمه

در عصر حاضر، پیش‌بینی قیمت سهام به‌عنوان یکی از موضوعات محوری در حوزه‌های اقتصاد، مدیریت مالی و علوم داده، جایگاه ویژه‌ای یافته و توجه طیف گسترده‌ای از پژوهشگران، سرمایه‌گذاران و تصمیم‌گیران بازارهای مالی را به خود جلب کرده است. بازارهای مالی نقش حیاتی در تخصیص بهینه منابع، تأمین مالی بنگاه‌ها و هدایت سرمایه در اقتصاد ایفا می‌کنند. با توجه به ماهیت پویا، غیرخطی و پیچیده این بازارها، بهره‌گیری از ابزارهای تحلیلی نوین مبتنی بر فناوری‌های هوش مصنوعی و یادگیری ماشین می‌تواند به کاهش ریسک سرمایه‌گذاری، افزایش کارایی بازار و ارتقای کیفیت تصمیم‌گیری‌های اقتصادی منجر شود. یکی از عوامل کلیدی که بر رفتار بازار و بازدهی سرمایه‌گذاری تأثیرگذار

است، دامنه نوسان قیمت سهام است. اعمال محدودیت در نوسان روزانه قیمت، که از سوی نهادهای ناظر با هدف کاهش هیجانات و کنترل نوسانات شدید صورت می‌گیرد، می‌تواند الگوهای معاملاتی، رفتار سرمایه‌گذاران و نقدشوندگی بازار را دستخوش تغییر کند. بررسی اثر این محدودیت‌ها بر عملکرد مدل‌های پیش‌بینی، به‌ویژه در بستر بازارهای مالی نوظهور مانند ایران، اهمیت زیادی دارد. با وجود مطالعات متعدد در زمینه پیش‌بینی قیمت سهام با استفاده از الگوریتم‌های یادگیری ماشین، بیشتر آن‌ها شرایط پایدار و بدون مداخله‌های سیاستی را بررسی کرده‌اند و به تأثیر مستقیم سیاست‌های محدودکننده نظیر تغییر دامنه نوسان کمتر پرداخته‌اند. در برخی مطالعات اخیر، از الگوریتم‌های پیشرفته‌تری مانند شبکه‌های عصبی بازگشتی، مدل‌های ترکیبی هیبریدی، و الگوریتم‌های مبتنی بر یادگیری تقویتی برای پیش‌بینی قیمت سهام استفاده شده است. برای نمونه، ژانگ در سال ۲۰۲۲ از مدل ترکیبی CNN-LSTM برای پیش‌بینی قیمت سهام در بازار چین استفاده کرده‌اند [۱]. همچنین در پژوهش دیگری در سال ۲۰۲۴ با بهره‌گیری از مدل LSTM و شاخص‌های تحلیل تکنیکال در بازار بورس ویتنام، به دقت پیش‌بینی ۹۳٪ دست یافته‌اند [۲]. در ایران، مطالعه‌ای با تمرکز بر شرکت‌های مپنا و توکاریل در بازار بورس تهران، از مدل‌های یادگیری عمیق برای پیش‌بینی قیمت سهام استفاده کرده و نتایج قابل قبولی ارائه داده است [۳]، مدل‌های ترکیبی هیبریدی، و الگوریتم‌های مبتنی بر یادگیری تقویتی برای پیش‌بینی قیمت سهام استفاده شده است. با این حال، بسیاری از این مدل‌ها یا بر داده‌های بین‌المللی متمرکز بوده‌اند و یا شرایط واقعی بازار ایران را مدنظر قرار نداده‌اند. مسئله اصلی این پژوهش آن است که آیا تغییر در دامنه نوسان می‌تواند بر عملکرد مدل‌های یادگیری ماشین در پیش‌بینی قیمت سهام تأثیرگذار باشد یا خیر. نوآوری پژوهش حاضر در ترکیب الگوریتم‌های یادگیری ماشین با تحلیل اثرپذیری آن‌ها از سیاست‌های کنترلی بازار، در بستر واقعی فرابورس ایران و با استفاده از داده‌های مربوط به دو دوره زمانی پیش و پس از تغییر دامنه نوسان نهفته است. همچنین، مزیت این پژوهش نسبت به مطالعات گذشته در تمرکز هم‌زمان بر تحلیل سیاست‌های بازار، استفاده از داده‌های واقعی داخلی، و مقایسه عملکرد دو مدل معتبر در شرایط متفاوت بازار است. در سال‌های اخیر، توسعه الگوریتم‌های یادگیری ماشین، به‌ویژه مدل‌هایی چون رگرسیون بردار پشتیبان و جنگل تصادفی، فرصت‌های جدیدی را برای تحلیل و پیش‌بینی دقیق‌تر رفتار بازار فراهم کرده‌اند. این پژوهش با تمرکز بر بازار فرابورس ایران، تلاش می‌کند با استفاده از داده‌های تاریخی و شاخص‌های تحلیل تکنیکال، بازده سهام را در دو بازه زمانی مختلف بررسی و تأثیر تغییرات دامنه نوسان قیمت را بر دقت پیش‌بینی تحلیل نماید. جامعه آماری تحقیق شامل چهار نماد صنعتی منتخب از بازار پایه اول فرابورس بوده و متغیرهایی چون قیمت باز، میانگین متحرک ساده و شاخص قدرت نسبی به‌عنوان ورودی و قیمت بسته به‌عنوان خروجی مدل در نظر گرفته شده است.

داده‌های گردآوری‌شده، پس از پیش‌پردازش و نرمال‌سازی، در قالب دو الگوریتم یادگیری ماشین آموزش داده شده‌اند و عملکرد مدل‌ها با استفاده از معیارهای ارزیابی استاندارد از جمله میانگین مربعات خطا، جذر میانگین مربعات خطا، میانگین قدر مطلق خطا، میانگین قدر مطلق درصد خطا و ضریب تعیین سنجیده شده است. نتایج حاصل از تحلیل مدل‌ها نشان می‌دهد که جنگل تصادفی در هر دو بازه زمانی عملکرد بهتری نسبت به رگرسیون بردار پشتیبان داشته و همچنین تغییر دامنه نوسان قیمت منجر به کاهش دقت پیش‌بینی در هر دو مدل شده است. این یافته‌ها نشان می‌دهد که تصمیمات نظارتی در خصوص محدودیت‌های معاملاتی می‌توانند بر دقت تحلیل‌های پیش‌بینی اثرگذار باشند و باید در طراحی استراتژی‌های معاملاتی مورد توجه قرار گیرند.

از میان روش‌های متعدد یادگیری ماشین، در این پژوهش الگوریتم‌های جنگل تصادفی و رگرسیون بردار پشتیبان به دلایل فنی و عملیاتی خاص انتخاب شده‌اند. جنگل تصادفی با ساختار گروهی خود، توانایی بالایی در مقابله با داده‌های پرت، کاهش بیش‌برازش و افزایش پایداری مدل دارد و برای داده‌های مالی با نوسانات شدید بسیار مناسب است. از سوی دیگر، رگرسیون بردار پشتیبان با بهره‌گیری از هسته‌های غیرخطی همچون RBF، می‌تواند روابط پیچیده بین متغیرهای ورودی و خروجی را مدل‌سازی کند و عملکرد خوبی در مجموعه‌داده‌های با حجم محدود ارائه دهد. با توجه به محدودیت‌های داده‌ای، وجود نویز، و ساختار غیرخطی داده‌های بازار فرابورس ایران، این دو الگوریتم به عنوان گزینه‌های بهینه برای این پژوهش انتخاب شده‌اند.

ساختار مقاله به‌گونه‌ای طراحی شده است که مراحل پژوهش را به‌صورت نظام‌مند و گام‌به‌گام تبیین می‌کند. در بخش دوم، مروری بر ادبیات نظری و مطالعات پیشین مرتبط با پیش‌بینی قیمت سهام و تأثیر دامنه نوسان ارائه شده است. در بخش سوم، روش تحقیق شامل معرفی جامعه آماری، متغیرهای پژوهش، روش‌های پیش‌پردازش داده و الگوریتم‌های مورد استفاده تشریح می‌شود. بخش چهارم به آموزش مدل‌ها و تحلیل نتایج اختصاص دارد، و در

بخش پنجم، تأثیر تغییر دامنه نوسان بر عملکرد مدل‌ها بررسی می‌شود. در نهایت، بخش ششم به نتیجه‌گیری کلی، تحلیل یافته‌ها و ارائه پیشنهادهایی برای تحقیقات آینده می‌پردازد.

۲- مبانی نظری و پیشینه تحقیق

در تحقیقات اخیر، استفاده از الگوریتم‌های یادگیری ماشین برای پیش‌بینی بازده سهام نتایج امیدوارکننده‌ای به‌دست داده است. به‌عنوان نمونه، شاشیکالا و هارشیتا در پژوهشی با عنوان "تحلیل و پیش‌بینی بازار سهام با استفاده از الگوریتم‌های یادگیری ماشین" دریافتند که ماشین‌های بردار پشتیبان در پیش‌بینی حرکات اقتصادی و کمک به سرمایه‌گذاران در بهینه‌سازی سود بسیار مؤثر هستند. این الگوریتم‌ها توانسته‌اند دقت پیش‌بینی را به طور قابل توجهی افزایش دهند همچنین، در مطالعه‌ای با عنوان "ارزیابی پیش‌بینی بازار سهام با استفاده از تکنیک‌های یادگیری ماشین نظارت شده" به این نتیجه رسیدند که ماشین‌های بردار پشتیبان با دقت ۹۱٪ عملکرد بهتری نسبت به سایر الگوریتم‌ها از جمله درختان رگرسیون خطی و درخت تصمیم دارند [۴].

پژوهشی با عنوان "تأثیر انتخاب نامطلوب بر استراتژی قیمت متخصصان بورس اوراق بهادار با استفاده از روش‌های چارچوب مدل‌سازی انتخاب نامطلوب" با تکیه بر تکنیک شبیه‌سازی مونت کارلو به بررسی تاثیر قیمت نامطلوب، میزان تاثیر عدم قطعیت قیمت و نرخ‌های سهام انجام دادند که یافته‌های پژوهش حاکی از این است که تعداد معاملات مطابق با کاهش عدم اطمینان رشد می‌کند، زیرا متخصصان احتمال بیشتری برای پیش‌بینی قیمت بالقوه محصول مالی دارند [۵].

در پژوهشی با عنوان "مقایسه مدل‌های مختلف یادگیری ماشین در پیش‌بینی شاخص بازار سهام" دریافتند که مدل‌های شبکه عصبی بازگشتی LSTM عملکرد بهتری نسبت به سایر مدل‌ها برای پیش‌بینی شاخص‌های بازار بورس تهران دارند [۶].

در تحقیقی دیگر خود از مدل‌های شبکه عصبی حافظه کوتاه‌مدت طولانی، جنگل تصادفی و روش سری زمانی ARIMA برای پیش‌بینی بازدهی استفاده کرده‌اند. با مقایسه نتایج حاصل از این مدل‌ها مشخص شد که مدل شبکه عصبی حافظه کوتاه‌مدت طولانی عملکرد بهتری نسبت به سایر روش‌ها در پیش‌بینی بازدهی داشته است [۷].

امیری نیز در تحقیقی به بررسی قدرت مدل‌های مبتنی بر هوش مصنوعی در پیش‌بینی روند قیمت سهام بورس اوراق بهادار تهران پرداخت. نتایج این پژوهش نشان داد که مدل‌های یادگیری عمیق نسبت به مدل‌های دیگر عملکرد بهتری دارند، هرچند که نتایج حاصل، آن‌طور که باید، برای سرمایه‌گذاران حائز اهمیت نیست.

پژوهشی با عنوان "بررسی اثر تغییر دامنه نوسان قیمت بر تلاطم تحقق یافته قیمت در بازار بورس تهران" با استفاده از روش‌های تلاطم، نقدشوندگی، حجم معاملات، کارایی انجام دادند که یافته‌های پژوهش حاکی از این است که کاهش تلاطم پس از اعمال دامنه نوسانی، رابطه‌ی معناداری بین تغییر دامنه نوسان و کارایی قیمت یافت نشد، تاثیر بر کاهش نقدشوندگی و نیز کاهش حجم معاملات پس از تغییر دامنه نوسان در بازار می‌گذارد[۸].

این پژوهش به بررسی کارایی الگوریتم‌های یادگیری ماشین در پیش‌بینی سری‌های زمانی بازارهای مالی می‌پردازد. چالش اصلی، افزایش دقت مدل‌های پیش‌بینی برای بهبود تصمیم‌گیری اقتصادی است. در این مطالعه، از الگوریتم‌های رگرسیون مبتنی بر یادگیری ماشین با تأکید بر انتخاب ویژگی‌های فنی مؤثر استفاده شده است. داده‌های مورد بررسی شامل اطلاعات دو شرکت در بورس تهران طی سال‌های ۱۳۸۷ تا ۱۳۹۷ است. نتایج نشان می‌دهد که انتخاب بهینه ویژگی‌های فنی، دقت پیش‌بینی را افزایش داده و الگوریتم‌های ماشین بردار پشتیبان و پرسپترون چندلایه عملکرد بهتری در کاهش خطا دارند [۹].

در پژوهشی با عنوان "یادگیری ماشین مبتنی بر رویکرد سلسله مراتبی برابری ریسک"، نشان دادند که رویکرد برابری ریسک سلسله مراتبی، عملکرد بهتری نسبت به رویکرد حداقل واریانس در مدیریت پرتفوهای سهام بورس تهران داشته است. این پژوهش‌ها نشان‌دهنده اهمیت روزافزون استفاده از روش‌های یادگیری ماشین در پیش‌بینی قیمت سهام و تحلیل بازار بورس ایران است. همچنین، تأثیر تغییرات دامنه نوسان قیمت سهام بر نوسانات و پیش‌بینی بازده سهام مورد توجه بسیاری از محققان قرار گرفته است [۱۰].

این پژوهش به پیش‌بینی بازده سهام بورس اوراق بهادار با توجه به تأثیر تغییرات دامنه نوسان قیمت سهام می‌پردازد. بورس اوراق بهادار به‌عنوان یک بازار مالی کلیدی، نقش مهمی در تأمین مالی اقتصاد ایفا می‌کند. قیمت سهام در این بازار تحت تأثیر عواملی مانند سیاست‌های اقتصادی، نوسانات نرخ ارز و قیمت نفت قرار دارد. تغییر دامنه نوسان قیمت سهام یکی از این عوامل مؤثر است که بر الگوی معاملات و بازده سهام تأثیر می‌گذارد. بنابراین، پیش‌بینی دقیق بازده سهام با در نظر گرفتن این تغییرات، اهمیت زیادی در تصمیم‌گیری‌های سرمایه‌گذاری دارد. برای این منظور، از روش‌های یادگیری ماشین استفاده می‌شود که به دلیل توانایی

مدل‌سازی پیچیدگی‌های بازار سهام، ابزار مناسبی برای پیش‌بینی محسوب می‌شوند. یکی از اهداف این پژوهش مقایسه کاربردی مبتنی بر الگوریتم‌های یادگیری ماشین (جنگل تصادفی و ماشین بردار پشتیبان) جهت پیش‌بینی قیمت سهام فرابورس ایران است که در آن تصمیمات سرمایه‌گذاری توسط شاخص‌های تکنیک‌های یادگیری ماشین است.

در این پژوهش مروری بر مطالعات اخیر در حوزه پیش‌بینی قیمت سهام با استفاده از الگوریتم‌های یادگیری ماشین صورت گرفته است. در جدول ۱، مدل‌ها و روش‌های به‌کاررفته در هر پژوهش به همراه مزایا، معایب و نوع داده‌های مورد استفاده ارائه شده است. این مرور نه‌تنها تنوع رویکردهای مدل‌سازی را نشان می‌دهد، بلکه جایگاه و نوآوری پژوهش حاضر را نیز در مقایسه با روش‌های پیشین روشن‌تر می‌سازد.

جدول۱: تجمیع پژوهش‌های مشابه

عنوان مقاله	روش‌ها / مدل‌ها	نوع داده‌های مورد استفاده	معایب	مزایا
الگوریتم‌های یادگیری ماشین با استفاده از داده‌های معاملات داخلی[۱۱]	درخت تصمیم، جنگل تصادفی، الگوریتم بردار پشتیبانی، خوشه‌بندی میانگین‌ها	داده‌های معاملات داخلی شرکت‌ها	حساسیت الگوریتم بردار پشتیبانی به پارامتر کرنل، محدودیت داده‌ها	ارزیابی تطبیقی الگوریتم‌ها، دقت خوب الگوریتم بردار پشتیبانی
الگوریتم جدیدی برای استخراج داده‌های سهام در اینترنت چیزهای چندرسانه‌ای [۱۲]	پردازش تصویر با تبدیل موجک و تجزیه مقادیر منفرد	داده‌های تصویری و تاریخی قیمت سهام	پیچیدگی در پیاده‌سازی الگوریتم، نیاز به داده‌های تصویری با کیفیت بالا	ارائه رویکرد نوین برای تحلیل نوسانات بازار
تحلیل و پیش‌بینی بازار سهام با استفاده از الگوریتم‌های یادگیری ماشین[۱۳]	الگوریتم بردار پشتیبانی	داده‌های تاریخی بازار سهام	حساسیت به پارامترها، خطر بیش‌برازش	دقت بالا در پیش‌بینی جهت حرکت بازار، مناسب برای پیش‌بینی‌های کوتاه‌مدت
پیش‌بینی قیمت سهام با استفاده از تکنیک‌های یادگیری ماشین [۱۴]	رگرسیون خطی چندگانه، الگوریتم بردار پشتیبانی، شبکه پرسپترون چندلایه، الگوریتم تقویت‌شده X	داده‌های تاریخی ۱۵ ساله سهامTCS	خطای بیشتر مدل‌های پیچیده‌تر مانند شبکه پرسپترون چندلایه، الگوریتم‌های گرادیان افزایشی	ارزیابی الگوریتم‌ها؛ عملکرد بهتر رگرسیون خطی چندگانه با خطای کمتر
رویکرد ترکیبی در پیش‌بینی قیمت سهام [۱۵]	ترکیب شبکه عصبی عمیق، حافظه بلندمدت، شبکه بازگشتی دروازه‌دار، مدل گرادیان افزایشی سبک، رگرسیون خطی	داده‌های سری زمانی و ویژگی‌های استخراج‌شده	ساختار بسیار پیچیده، زمان‌بر بودن آموزش مدل‌ها	دقت بالا، کاهش خطا با ترکیب مدل‌ها
ارزیابی پیش‌بینی بازار سهام با استفاده از الگوریتم‌های نظارت‌شده [۱۶]	رگرسیون خطی، رگرسیون بردار پشتیبانی، رگرسیون چندجمله‌ای، رگرسیونLASSO	داده‌های تاریخی قیمت سهام.AppleInc	عملکرد ضعیف رگرسیون بردار پشتیبانی؛ نیاز به تنظیم دقیق	رگرسیون خطی و LASSO با دقت بالا در پیش‌بینی قیمت بسته روز بعد
مقایسه مدل‌های مختلف یادگیری ماشین در پیش‌بینی شاخص بازار سهام [۴]	جنگل تصادفی، الگوریتم بردار پشتیبانی، شبکه عصبی مصنوعی، شبکه بازگشتی با حافظه بلندمدت	شاخص کل بورس اوراق بهادار تهران (۲۰۱۳ تا ۲۰۲۰)	نیاز به داده‌های زیاد، عملکرد ضعیف مدل‌های سنتی	عملکرد بهتر مدل حافظه بلندمدت در بازه‌های زمانی مختلف
بهینه‌سازی پرتفوی با پیش‌بینی بازده با استفاده از حافظه بلندمدت، جنگل تصادفی و مدل ARIMA [۵]	مدل حافظه بلندمدت، جنگل تصادفی، مدل سری زمانی خودرگرسیون میانگین متحرک	داده‌های ۵ صنعت از سال ۲۰۱۷ تا ۲۰۲۱	نیاز به تنظیم دقیق و پیچیدگی بالا	پیش‌بینی بازده و بهینه‌سازی پرتفوی هم‌زمان؛ عملکرد بهتر نسبت به مدل سنتی
مقایسه مدل‌های مختلف یادگیری ماشین برای پیش‌بینی قیمت سهام [۱۷]	رگرسیون خطی، الگوریتم نزدیک‌ترین همسایه، شبکه عصبی با حافظه بلندمدت	داده‌های تاریخی سهام شرکت تسلا	عملکرد ضعیف رگرسیون خطی و نزدیک‌ترین همسایه	مدل حافظه بلندمدت عملکرد بهتری در پیش‌بینی قیمت سهام تسلا نشان داد
پیش‌بینی قیمت سهام با استفاده از مدل ترکیبی عصبی و الگوریتم بهینه‌سازی [۱۸]	شبکه عصبی ترکیبی حافظه بلندمدت - پیچشی، بهینه‌سازی با الگوریتم ازدحام ذرات	داده‌های ۱۰ نماد از ۲۰ شهریور ۱۳۹۲ تا ۲۰ شهریور ۱۴۰۲ شامل اندیکاتورها	پیچیدگی در تنظیم پارامترها، نیاز به توان پردازشی بالا	عملکرد بهتر مدل ترکیبی حافظه بلندمدت و شبکه عصبی پیچشی، بهینه‌سازی با الگوریتم ازدحام ذرات
شبکه عصبی به سبک تبدیل‌کننده در پیش‌بینی قیمت سهام[۱۹]	مدل شبکه عصبی مبتنی بر تبدیل‌کننده (ترنسفورمر) سفارشی‌شده	داده‌های تکنیکال بازار سهام	نیاز به داده زیاد، پیچیدگی و زمان‌بر بودن آموزش	دقت بالا در پیش‌بینی، یادگیری روابط بلندمدت زمانی

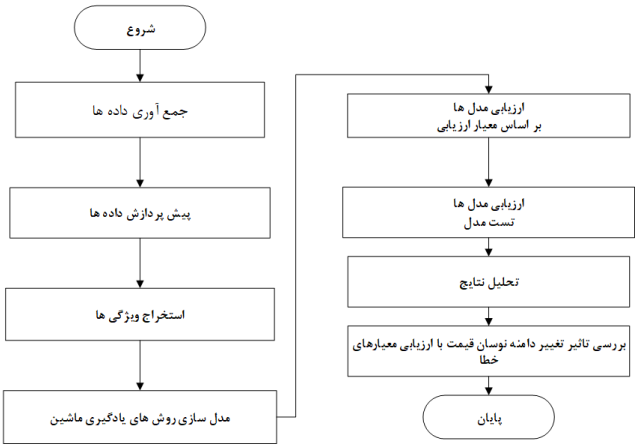
همان‌گونه که در جدول مشاهده می‌شود، الگوریتم‌های ترکیبی (مانند حافظه بلندمدت و شبکه‌های توجه‌محور) و الگوریتم‌های کلاسیک (مانند رگرسیون بردار پشتیبان و جنگل تصادفی) در سال‌های اخیر مورد توجه پژوهشگران قرار گرفته‌اند. پژوهش حاضر با تمرکز بر تأثیر تغییرات دامنه نوسان قیمت و استفاده از دو الگوریتم معتبر در بستر واقعی بازار ایران، دارای تمایز روش‌شناختی و محتوایی نسبت به مطالعات جدول ۱ است.

۳- روش‌شناسی پژوهش

یادگیری ماشین شاخه‌ای از هوش مصنوعی است که به سیستم‌ها امکان می‌دهد از داده‌ها یاد بگیرند و عملکرد خود را بهبود بخشند. در این پژوهش، هدف اصلی مقایسه کارآیی دو الگوریتم جنگل تصادفی رگرسیون بردار پشتیبان برای پیش‌بینی بازده سهام در بورس ایران است. همچنین، تأثیر تغییرات دامنه نوسان بر پیش‌بینی قیمت سهام مورد بررسی قرار می‌گیرد. داده‌های ورودی شامل قیمت باز سهام، میانگین متحرک و شاخص قدرت نسبی است و داده خروجی شامل قیمت بسته سهام می باشد. لازم به ذکر است که کلیه مراحل این پژوهش در نرم افزار پایتون و مطابق با شکل ۱ پیاده سازی می‌شود.

یکی از مهم‌ترین مراحل در طراحی مدل‌های یادگیری ماشین، انتخاب روش‌هایی متناسب با ماهیت داده‌ها و ویژگی‌های مسئله مورد بررسی است. در این پژوهش، داده‌ها مربوط به بازار فرابورس ایران بوده که دارای ویژگی‌هایی چون نوسانات شدید قیمتی، حجم متوسط داده‌ها، روابط غیرخطی بین متغیرها، و ساختار تکنیکال پیچیده هستند. در چنین شرایطی، استفاده از الگوریتم‌هایی مانند جنگل تصادفی و رگرسیون بردار پشتیبان توجیه‌پذیر است.

الگوریتم جنگل تصادفی به عنوان یک مدل قدرتمند گروهی، توانایی بالایی در مقابله با نویز و جلوگیری از بیش‌برازش دارد و از طریق نمونه‌گیری تصادفی و ترکیب نتایج درخت‌های تصمیم‌گیری، عملکرد باثباتی در مسائل غیرخطی و پیچیده از خود نشان می‌دهد. از سوی دیگر، الگوریتم رگرسیون بردار پشتیبان، با استفاده از هسته‌های غیرخطی نظیر RBF، قادر است ساختارهای پنهان در داده را شناسایی کرده و روابط غیرخطی میان متغیرها را مدل‌سازی کند. در مقابل، مدل‌هایی نظیر شبکه‌های عصبی بازگشتی یا یادگیری عمیق اگرچه در برخی موارد عملکرد بالایی دارند، اما نیازمند حجم بسیار بالای داده، زمان آموزش بیشتر، و تنظیمات حساس‌تر هستند که با توجه به ماهیت داده‌های در دسترس و نیاز پژوهش حاضر، کارایی عملی لازم را نداشتند. بنابراین، انتخاب دو الگوریتم مذکور با در نظر گرفتن تناسب با حجم داده، پایداری مدل، مقاومت در برابر نویز و توانایی مدل‌سازی روابط غیرخطی انجام شده است.



شکل۱. فرآیند انجام کار

۳-۱- جمع آوری داده‌ها

داده‌های این پژوهش از منابع معتبر مانند وب‌سایت رسمی بورس اوراق بهادار www.tsetmc.com جمع‌آوری شده است. جامعه آماری پژوهش شامل سهام‌های فعال بورس ایران است و برای تحلیل، نمونه‌ای از سهام‌های حوزه صنعتی در فرابورس ایران انتخاب شده‌اند. دوره زمانی ۱۳۹۴ تا ۱۳۹۹ برای مقایسه عملکرد الگوریتم‌ها انتخاب شده است، در حالی که تغییرات دامنه نوسان در دوره ۱۴۰۰ تا ۱۴۰۲ مورد بررسی قرار میگیرد. با توجه به تغییرات متعدد دامنه نوسانات مجاز قیمت در سهام فرابورس اوراق بهادار ایران طی سال‌های ۱۴۰۰ الی ۱۴۰۲، این بازه زمانی به‌عنوان بازه‌ای جهت بررسی تأثیر تغییر دامنه نوسان قیمت سهام بر عملکرد پیش‌بینی قیمت سهام با استفاده از روش‌های یادگیری ماشین در نظر گرفته شده است. پس از بررسی و مقایسه عملکرد الگوریتم‌های یادگیری ماشین طی سال‌های ۱۳۹۴ الی ۱۳۹۹، به پیش‌بینی قیمت سهام برای سال‌های ۱۴۰۰ الی ۱۴۰۲ با همان داده‌های آموزش پرداخته می‌شود. بررسی تأثیر تغییر دامنه نوسان قیمت بر پیش‌بینی سهام با بهره‌گیری از مقایسه عملکرد الگوریتم‌های یادگیری ماشین در پیش‌بینی قیمت سهام طی دو بازه زمانی متفاوت انجام می‌گیرد.

۳-۲- انتخاب جامعه آماری

در این پژوهش، جامعه آماری شامل سهام‌های صنعتی فعال در بازار پایه اول فرابورس ایران است. انتخاب این بخش به دلیل نقش کلیدی آن در بازار سرمایه ایران و تأثیرگذاری آن بر روندهای

اقتصادی کشور انجام شده است. از آنجا که صنایع فلزی و معدنی از مهم‌ترین بخش‌های تولیدی محسوب می‌شوند و نقش استراتژیکی در جایگزینی درآمدهای غیرنفتی دارند، بررسی روند قیمتی سهام این شرکت‌ها می‌تواند به تحلیل دقیق‌تر پویایی‌های بازار کمک کند.

به‌منظور انجام این مطالعه، چهار نماد از شرکت‌های فعال در این حوزه که سوابق اطلاعاتی آنها از سال ۱۳۹۴ موجود است شامل زرین معدن آسیا (فزرین)، صنعت روی زنگان (زنگان)، تولید و توسعه سرب و روی ایرانیان (فتوسا) و صنایع فولاد آلیاژی یزد (فولای)، به‌عنوان نمونه انتخاب شدند. معیار انتخاب این نمادها بر اساس میزان نقدشوندگی، حجم معاملات بوده است.

داده‌های مورد بررسی طی دو دوره زمانی ۱۳۹۴ تا ۱۳۹۹ و ۱۴۰۰ تا ۱۴۰۲ گردآوری شده‌اند. هدف از این تفکیک زمانی، ارزیابی عملکرد الگوریتم‌های یادگیری ماشین در پیش‌بینی قیمت سهام در یک دوره پایدار و سپس بررسی تأثیر تغییرات دامنه نوسان قیمت بر دقت این پیش‌بینی‌ها در سال‌های اخیر است. در نتیجه، این پژوهش امکان مقایسه نتایج به‌دست‌آمده در شرایط متفاوت بازار را فراهم می‌کند و می‌تواند به درک بهتر از نحوه تأثیرگذاری تغییرات دامنه نوسان بر روندهای قیمتی سهام کمک کند.

۳-۳- پیش‌پردازش داده‌ها

در این پژوهش، به‌منظور بهبود دقت مدل‌های پیش‌بینی و افزایش کیفیت داده‌ها، فرآیند پیش‌پردازش داده‌ها انجام شده است. این فرآیند شامل دو مرحله اصلی پاک‌سازی داده‌ها و نرمال‌سازی مقادیر عددی است. در مرحله پاک‌سازی، داده‌های دارای مقادیر گمشده یا نامعتبر شناسایی و اصلاح شده‌اند تا از تأثیر منفی آن‌ها بر عملکرد مدل جلوگیری شود. سپس، برای یکسان‌سازی مقیاس ویژگی‌ها و جلوگیری از تأثیرگذاری بیش از حد برخی متغیرها بر مدل، داده‌ها با استفاده از روش StandardScaler نرمال‌سازی شده‌اند.

ویژگی‌های منتخب برای مدل‌سازی شامل قیمت باز، قیمت بسته، میانگین متحرک ۱۴ روزه و شاخص قدرت نسبی هستند. این ویژگی‌ها به دلیل اهمیت آن‌ها در تحلیل تکنیکال بازار سهام و تأثیرگذاری بر روند قیمت انتخاب شده‌اند. نرمال‌سازی داده‌ها باعث می‌شود که مدل بتواند الگوهای قیمتی را بهتر تشخیص دهد و عملکرد بهتری در پیش‌بینی داشته باشد.

در فرایند پیش‌پردازش داده‌ها، ابتدا شاخص‌های تحلیل تکنیکال از جمله میانگین متحرک ساده ۱۴ روزه و شاخص قدرت نسبی ۱۴ روزه محاسبه گردید. سپس، مقادیر گمشده حاصل از محاسبه این شاخص‌ها حذف شدند تا از ورود نویز به مدل جلوگیری شود. همچنین ستون تاریخ که دارای فرمت‌های ناسازگار یا تاریخ‌های نامعتبر بود، به کمک تبدیل نوع و فیلترگذاری تصحیح گردید.

همچنین، به‌منظور افزایش دقت مدل و جلوگیری از پراکندگی داده‌ها، کلیه داده‌های مربوط به سهام منتخب به‌صورت یکجا مورد استفاده قرار گرفته‌اند. این رویکرد باعث ایجاد مجموعه داده‌ای جامع‌تر شده و مدل را قادر می‌سازد تا از الگوهای مشترک میان سهام مختلف برای بهبود دقت پیش‌بینی بهره ببرد. از آنجا که نوسانات قیمتی و روندهای حاکم بر بازار ممکن است در نمادهای مختلف شباهت‌هایی داشته باشند، این ترکیب داده‌ها به مدل کمک می‌کند تا رفتار کلی بازار را بهتر درک کرده و عملکرد بهتری در پیش‌بینی قیمت‌ها داشته باشد.

۳-۴- استخراج ویژگی‌ها

در این پژوهش، فرآیند استخراج ویژگی‌ها به‌عنوان یکی از مراحل کلیدی در مدل‌سازی یادگیری ماشین انجام شده است تا اطلاعات مؤثری برای پیش‌بینی قیمت سهام فراهم شود. انتخاب ویژگی‌های مناسب تأثیر بسزایی در بهبود عملکرد مدل دارد، زیرا این متغیرها اطلاعات ارزشمندی درباره روندهای قیمتی و رفتار بازار ارائه می‌دهند. برای این منظور، چهار ویژگی اصلی از داده‌های بازار سهام استخراج شده است که شامل قیمت باز، قیمت بسته، میانگین متحرک ساده و شاخص قدرت نسبی هستند. قیمت باز نشان‌دهنده اولین قیمتی است که سهم در ابتدای روز معاملاتی با آن معامله می‌شود و می‌تواند جهت‌گیری ابتدایی بازار را بازتاب دهد. قیمت بسته شدن سهام نیز که در این پژوهش به عنوان متغیر هدف در نظر گرفته شده است نمایانگر آخرین معامله سهم در پایان روز معاملاتی است و به‌طور گسترده‌ای به عنوان متغیر هدف در پیش‌بینی قیمت روز بعد به کار می‌رود[۲۰]. میانگین متحرک ساده، که معمولاً در بازه‌های زمانی ۱۰، ۱۴ یا ۲۰ روزه محاسبه می‌شود که در این پژوهش میانگین متحرک ساده با دوره ۱۴ روزه مورد استفاده قرار گرفته است. به‌عنوان شاخصی تکنیکال برای تشخیص روند کلی بازار استفاده می‌گردد و با کاهش نوسانات قیمت، سیگنال‌های مناسبی برای خرید و فروش ارائه می‌دهد [۲۱]. همچنین، از شاخص قدرت نسبی که یک شاخص نوسانی برای اندازه‌گیری قدرت و ضعف حرکت قیمت در دوره‌های اخیر است، بهره گرفته شده است. این شاخص معمولاً برای شناسایی شرایط اشباع خرید و فروش مورد استفاده قرار می‌گیرد و می‌تواند در تحلیل رفتار قیمت نقش مهمی ایفا کند. این شاخص اغلب در

ترکیب با سایر ویژگی‌ها در مدل‌سازی قیمت سهام مورد استفاده قرار گرفته است [۲۲].در این پژوهش، برای محاسبه این شاخص نیز بازه ۱۴ روزه در نظر گرفته شده است. برای کاهش تأثیر تفاوت مقیاس ویژگی‌ها، از نرمال‌سازی با استفاده از استانداردسازی (StandardScaler) استفاده شد. نمودارهای مقایسه‌ای داده‌های قبل و بعد از نرمال‌سازی نشان‌دهنده هم‌مقیاس شدن متغیرها و توزیع متعادل‌تر آن‌ها در مدل نهایی است. با توجه به نقش کلیدی این ویژگی‌ها در تحلیل و پیش‌بینی قیمت سهام، انتخاب آن‌ها بر اساس مطالعات پیشین و تجربه عملی در بازارهای مالی انجام شده است. این متغیرها اطلاعات لازم را برای شناسایی الگوهای قیمتی و بررسی نوسانات بازار در اختیار مدل یادگیری ماشین قرار می‌دهند و موجب بهبود دقت پیش‌بینی و تحلیل رفتار آتی سهام می‌شوند.

۳-۵- ارزیابی مدل‌ها

برای ارزیابی عملکرد مدل‌های یادگیری ماشین در پیش‌بینی قیمت سهام، داده‌ها به دو مجموعه آموزشی و آزمایشی تقسیم می‌شوند. این تقسیم‌بندی به‌صورت ۸۰-۲۰ انجام می‌شود، به این معنا که درصد بیشتری از داده‌ها برای آموزش مدل و بخش کوچکتری برای ارزیابی آن در نظر گرفته می‌شود. این مرحله اهمیت زیادی دارد، زیرا هدف، آموزش مدل به گونه‌ای است که بتواند بر اساس الگوهای موجود در داده‌های گذشته، پیش‌بینی‌های دقیقی برای داده‌های جدید ارائه دهد.

پس از آموزش مدل، عملکرد آن با استفاده از معیارهای مختلف ارزیابی می‌شود تا دقت پیش‌بینی‌ها سنجیده شده و بهترین مدل انتخاب گردد. معیارهای مورد استفاده در این پژوهش شامل میانگین مربعات خطا، جذر میانگین مربعات خطا، میانگین قدر مطلق خطا، ضریب تعیین و میانگین درصد خطای مطلق هستند که هر یک از آن‌ها ویژگی‌های خاصی در تحلیل خطای مدل دارند.

میانگین مربعات خطا به‌عنوان یکی از پرکاربردترین معیارهای ارزیابی مدل، میزان خطای مطلق مدل را اندازه‌گیری می‌کند، اما به دلیل آنکه مجذور اختلاف‌ها را در نظر می‌گیرد، نسبت به خطاهای بزرگ حساس‌تر است. جذر میانگین مربعات خطا که از جذر مقدار به‌دست می‌آید، خطا را در همان مقیاس اصلی داده‌ها نشان می‌دهد و تفسیر آن ساده‌تر است. میانگین قدر مطلق خطا نیز معیاری برای سنجش خطای مطلق مدل بوده و تأثیر خطاهای بزرگ را در ارزیابی کاهش می‌دهد. علاوه بر این، ضریب تعیین بیانگر میزان واریانس توضیح داده‌شده توسط مدل است و هرچه مقدار آن به عدد ۱ نزدیک‌تر باشد، نشان‌دهنده دقت بالاتر مدل در پیش‌بینی قیمت سهام است. در نهایت، میانگین درصد خطای مطلق برای ارزیابی میزان دقت پیش‌بینی به‌صورت درصدی به‌کار گرفته می‌شود و در مواردی که اهمیت مقیاس داده‌ها مطرح باشد، معیار مفیدی محسوب می‌شود.

با استفاده از این معیارها، عملکرد مدل‌ها مقایسه شده و مدلی که کمترین خطا و بالاترین دقت را داشته باشد، به‌عنوان مدل برتر برای پیش‌بینی قیمت سهام انتخاب شده‌اند.

$$MSE = \frac{1}{N} + \sum_{N}^{i=1} (\hat{Y}_i + Y_i) \quad (۱)$$

$$RSME = \sqrt{\frac{1}{N} \sum_{N}^{i=1} (\hat{Y}_i + Y_i)} \quad (۲)$$

$$MAE = \frac{1}{N} \sum_{N}^{i=1} |\hat{Y}_i + Y_i| \quad (۳)$$

$$MAPE = \frac{1}{N} \sum_{N}^{i=1} \left| \frac{\hat{Y}_i + Y_i}{\hat{Y}_i} \right| \quad (۴)$$

$$R^2 = 1 - \frac{\sum_{N}^{i=1} (Y_i - \overline{Y_i})^2}{\sum_{N}^{i=1} (Y_i - \overline{Y})^2} \quad (۵)$$

Y_i = مقدار واقعی شاخص در روز i ام

$\hat{Y}_i$ = مقدار پیشبینی شده ی شاخص برای روز i ام

N = تعداد کل داده‌های تست [۲۳]

۳-۶- پیاده‌سازی مدل‌ها

رگرسیون بردار پشتیبان یکی از الگوریتم‌های یادگیری ماشین است که برای پیش‌بینی داده‌های پیوسته استفاده می‌شود. این مدل به دنبال یافتن یک تابع است که بتواند داده‌ها را با کمترین انحراف پیش‌بینی کند. خروجی مدل رگرسیون بردار پشتیبان تابعی است که رابطه بین ورودی‌ها و خروجی‌ها را به خوبی مدل‌سازی می‌کند پس انتخاب پارامترهای ارزیابی از اهمیت ویژه ای برخوردار است. تنظیمات پارامترهای اصلی در پیاده سازی این روش شامل موارد زیر می باشد:

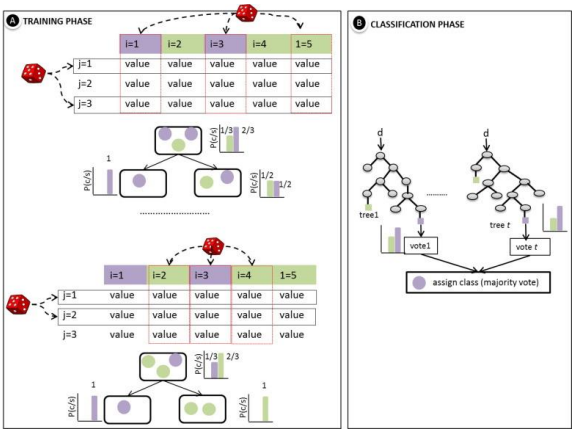
- kernel=rbf: نوع هسته که در اینجا رادیکال بیزی (RBF) انتخاب شده است.
- C=100 : پارامتر تنظیم که نشان‌دهنده تعادل بین حفظ خطاها و ایجاد یک مدل پیچیده‌تر است.
- gamma=0.1: پارامتر تنظیم دیگری که تأثیر ویژگی‌ها

را بر پیش‌بینی مدل کنترل می‌کند.

- epsilon=0.01: تعیین می‌کند که مدل تا چه میزان خطا را مجاز می‌داند.

جنگل تصادفی ترکیبی از چندین درخت تصمیم را برای ایجاد یک مدل قوی‌تر استفاده می‌کند. این الگوریتم مقاوم به اورفیتینگ و نوسانات داده‌های مالی است و می‌تواند الگوهای پیچیده را کشف کند. جنگل تصادفی می‌تواند از ویژگی‌های پیچیده و نویزی داده‌ها به‌خوبی استفاده کرده و پیش‌بینی‌های دقیق‌تری ارائه دهد. مطابق با شکل ۲ پیش‌بینی نهایی از طریق رای‌گیری یا میانگین‌گیری نتایج درخت‌های مختلف حاصل می‌شود. این مدل به‌ویژه برای پیش‌بینی روندهای کوتاه‌مدت و بلندمدت قیمت سهام مفید است. برای تولید درختان جنگل، تعداد درختان تصمیم گیری برای ۵۰۰ قرار داده شده است زیرا خطاها بعد از دستیابی به این تعداد درخت دسته بندی تثبیت میشوند .هر درخت تصمیم به طور مستقل و بدون هرس تولید می‌شود و هر گره با استفاده از تعداد متغیرهای تعریف شده توسط کاربر تقسیم می‌شود، که به طور تصادفی انتخاب شده است. جنگل تصادفی ابتدا داده‌ها به زیرمجموعه‌هایی تقسیم می‌شوند و از نمونه‌برداری تصادفی استفاده می‌شود. ویژگی‌ها به‌طور تصادفی انتخاب می‌شوند تا به کاهش همبستگی بین درخت‌ها کمک کنند. تعداد درختان تصمیم گیری برای تولید در این پژوهش ۵۰۰ در نظر گرفته شده است. در جنگل تصادفی اگر S یک مجموعه داده با n مشاهده (نمونه) است و G مجموعه داده آموزشی با m مشاهده می باشد. در فرایند تشکیل جنگل تصادفی از مجموعه آموزشی G در هر مرحله، باز نمونه گیری خودگردان گرفته می‌شود. احتمال انتخاب نشدن داده x در هر مرحله به‌صورت زیر است:

$$P(X \text{ انتخاب نشدن}) = \left(\frac{m-1}{m}\right)^m \quad (۶)$$



شکل۲. مراحل آموزش (الف) و طبقه بندی (ب) جنگل تصادفی

هر درخت تصمیم با ویژگی‌های انتخابی ساخته می‌شود و سپس پیش‌بینی‌ها با استفاده از میانگین یا رای‌گیری از نتایج درخت‌ها انجام می‌شود. در نهایت، این الگوریتم به‌طور گسترده در مسائل رگرسیون و طبقه‌بندی مورد استفاده قرار می‌گیرد و عملکرد بسیار خوبی در پردازش داده‌های پیچیده دارد. به صورت کلی پارامترهای پیاده سازی هر دو مدل به صورت خلاصه در جدول ۲ ذکر شده است.

جدول ۲: پارامترهای پیاده سازی مدل

مدل	پارامتر	مقدار
رگرسیون بردار پشتیبان	C	100
	gamma	0.1
	epsilon	0.01
جنگل تصادفی	n_estimators (Ntree)	500
	max_features (Mtree)	sqrt(n)

۴- تحلیل داده‌ها و یافته‌ها

۴-۱- بررسی داده‌های آموزش و تست

۴-۱-۱- بررسی داده‌های آموزش و تست در بازه زمانی ۱۳۹۴ تا ۱۳۹۹

با توجه به نتایج جدول شماره ۳، تفاوت عملکرد دو مدل در مرحله آموزش و آزمون قابل توجه است. مدل رگرسیون بردار پشتیبان با داشتن ضریب تعیین برابر با ۰/۶۲ در هر دو بخش آموزش و آزمون، و همچنین افزایش محسوس در میزان خطا در داده‌های آزمون (مثلاً میانگین قدر مطلق خطا از ۴٬۳۵۶/۶۴ در آموزش به ۴٬۷۵۱/۵۷ در آزمون رسیده است)، نشانه‌هایی از بیش‌برازش نسبی[1] را نشان می‌دهد. این موضوع به‌ویژه از آن‌جا اهمیت دارد که دقت مدل در داده‌های آموزشی نسبتاً بالا بوده، اما در داده‌های آزمون کاهش دقت و افزایش خطا مشاهده شده است. در مقابل، مدل جنگل تصادفی با ضریب تعیین ۱ در هر سه بخش (آموزش، آزمون و کل داده‌ها) و همچنین اختلاف کم بین خطاهای آموزش و آزمون، عملکردی بسیار پایدار و تعمیم‌پذیر ارائه داده است. این موضوع نشان می‌دهد که مدل جنگل تصادفی با وجود پیچیدگی کمتر، از قابلیت تعمیم بهتر به داده‌های جدید برخوردار بوده و در معرض بیش‌برازش قرار نگرفته است. بنابراین، می‌توان نتیجه گرفت که مدل رگرسیون بردار پشتیبان به‌احتمال زیاد در مرحله

آموزش بیش از حد به داده‌ها وابسته شده و در پیش‌بینی روی داده‌های جدید دچار ضعف نسبی بوده، در حالی‌که مدل جنگل تصادفی توانسته تعادل بهتری میان یادگیری و تعمیم برقرار کند.

لازم به ذکر است که معیارهایی مانند میانگین مربعات خطا و میانگین قدر مطلق خطا تحت تأثیر مستقیم مقیاس متغیر هدف هستند. از آن‌جا که در این پژوهش، متغیر هدف قیمت پایانی سهام است که در مقیاس واقعی ریال قرار دارد، بزرگی عددی این شاخص‌ها امری طبیعی است. همچنین این مقادیر در این معیارها الزاماً نشان‌دهنده‌ی عملکرد ضعیف مدل نیستند؛ بلکه باید در چارچوب مقیاس داده‌ها تفسیر شوند. همچنین نرمال‌سازی فقط روی ویژگی‌های ورودی (x) انجام شده است تا سرعت همگرایی و پایداری آموزش دو مدل رگرسیون بردار پشتیبان و جنگل تصادفی بهبود یابد. در واقع، مدل‌ها بر روی داده‌هایی آموزش دیده‌اند که خروجی آن‌ها در مقیاس واقعی باقی مانده و همین امر امکان تحلیل دقیق‌تر و قابل درک‌تر عملکرد را فراهم کرده است. از این رو به جهت شفافیت بیشتر، از معیارهای نسبی مانند میانگین درصد خطای مطلق و ضریب تعیین نیز استفاده شده است که دقت مناسب مدل‌ها را تأیید می‌کنند.

جدول ۳: پیاده سازی روش‌های یادگیری ماشین در بازه ۱۳۹۴ تا ۱۳۹۹

مدل / مجموعه داده‌ها	ضریب تعیین	میانگین درصد خطای مطلق	میانگین قدر مطلق خطا	جذر میانگین مربعات خطا	میانگین مربعات خطا
رگرسیون بردار پشتیبان – آموزش	۰٫۶۲	۱۶٫۸۳٪	۴٬۳۵۶٫۶۴	۱۸٬۰۰۰٫۹۵	۳۲۴٬۳۴۵٬۰۵۲٫۲۳
رگرسیون بردار پشتیبان – آزمون	۰٫۶۲	۱۷٫۰۸٪	۴٬۷۵۱٫۵۷	۱۸٬۸۵۹٫۳۳	۳۵۵٬۶۷۴٬۵۵۴٫۱۴
رگرسیون بردار پشتیبان – کل داده‌ها	۰٫۶۲	۱۶٫۸۸٪	۴۴۳۵٫۶۶	۱۸٬۱۸۰٫۲۷	۳۳۰٬۶۱۳٬۳۲۹٫۶۶
جنگل تصادفی – آموزش	۱٫۰۰	۰٫۳۵٪	۵۰٫۹۰۱	۱۵۷٫۵۴۸	۲۴٬۸۲۱٫۵۱
جنگل تصادفی – آزمون	۱٫۰۰	۰٫۹۳٪	۱۵۶٫۰۱۴	۴۰۲٫۷۶۲	۱۶۲٬۲۱۷٫۱۸
جنگل تصادفی – کل داده‌ها	۱٫۰۰	۰٫۴۶٪	۷۱٫۹۴۴	۲۲۸٫۷۵۱	۵۲٬۳۲۶٫۸۸

لازم به ذکر است که معیارهایی مانند میانگین مربعات خطا و میانگین قدر مطلق خطا تحت تأثیر مستقیم مقیاس متغیر هدف هستند. از آن‌جا که در این پژوهش، متغیر هدف قیمت پایانی سهام است که در مقیاس واقعی ریال قرار دارد، بزرگی عددی این شاخص‌ها امری طبیعی است. همچنین این مقادیر در این معیارها الزاماً نشان‌دهنده‌ی عملکرد ضعیف مدل نیستند؛ بلکه باید در چارچوب مقیاس داده‌ها تفسیر شوند. همچنین نرمال‌سازی فقط روی ویژگی‌های ورودی (x) انجام شده است تا سرعت همگرایی و پایداری آموزش دو مدل رگرسیون بردار پشتیبان و جنگل تصادفی بهبود یابد. در واقع، مدل‌ها بر روی داده‌هایی آموزش دیده‌اند که خروجی آن‌ها در مقیاس واقعی باقی مانده و همین امر امکان تحلیل دقیق‌تر و قابل درک‌تر عملکرد را فراهم کرده است. از این رو به جهت شفافیت بیشتر، از معیارهای نسبی مانند میانگین درصد خطای مطلق و ضریب تعیین نیز استفاده شده است که دقت مناسب مدل‌ها را تأیید می‌کنند.

۴-۱-۲- بررسی داده‌های آموزش و تست در بازه زمانی ۱۴۰۰ تا ۱۴۰۲

طبق جدول شماره ۴ نتایج حاصل از آموزش و آزمون دو مدل در بازه زمانی ۱۴۰۰ تا ۱۴۰۲ نشان می‌دهد که مدل رگرسیون بردار پشتیبان در داده‌های آموزشی عملکرد نسبتاً مناسبی داشته (ضریب تعیین برابر با ۹۰ صدم)، اما در داده‌های آزمون، افت قابل توجهی در دقت مدل مشاهده می‌شود (ضریب تعیین ۸۷ صدم و افزایش مقدار خطاها مانند میانگین مربعات خطا و میانگین قدر مطلق خطا). این اختلاف بین آموزش و آزمون نشان‌دهنده کاهش تعمیم‌پذیری مدل و احتمال بروز بیش‌برازش است؛ به‌ویژه آن‌که میانگین درصد خطای مطلق در آموزش و آزمون تقریباً برابر ولی نسبتاً بالا باقی مانده است (حدود ۱۶ درصد)، که می‌تواند نشان‌دهنده حساسیت مدل به نویز یا رفتارهای غیرخطی داده‌ها باشد.

جدول ۴: پیاده سازی روش‌های یادگیری ماشین در بازه ۱۳۹۴ تا ۱۳۹۹

مدل/ مجموعه داده‌ها	ضریب تعیین	میانگین درصد خطای مطلق	میانگین قدر مطلق خطا	جذر میانگین مربعات خطا	میانگین مربعات خطا
رگرسیون بردار پشتیبان – آموزش	۰٫۹۰	۱۶٫۳۵٪	۳٬۲۲۴٫۰۶	۷٬۹۴۰٫۷۰	۶۳٬۰۵۴٬۷۶۰٫۰۰
رگرسیون بردار پشتیبان – آزمون	۰٫۸۷	۱۶٫۶۴٪	۳٬۴۵۵٫۴۲	۹٬۱۰۶٫۶۱	۸۲٬۹۳۰٬۳۲۰٫۰۰
رگرسیون بردار پشتیبان – کل داده‌ها	۰٫۸۹	۱۶٫۴۱٪	۳٬۲۷۰٫۳۴	۸٬۱۸۷٫۱۸	۶۷٬۰۲۹٬۸۷۰٫۰۰
جنگل تصادفی – آموزش	۱٫۰۰	۷٫۴۷٪	۱۶۳٫۲۷	۳۳۰٫۴۰۵	۱۰۹٬۱۶۷٫۱۵
جنگل تصادفی – آزمون	۰٫۹۹	۱٫۹۸٪	۴۳۳٫۴۴	۹۴۹٫۶۷۱	۹۰۱٬۸۷۵٫۷۷
جنگل تصادفی – کل داده‌ها	۱٫۰۰	۹٫۹۳٪	۲۱۷٫۳۰	۵۱۷٫۴۰۶	۲۶۷٬۷۰۸٬۸۸

در مقابل، مدل جنگل تصادفی در همین بازه عملکردی به‌مراتب بهتر از خود نشان داده است. با وجود افزایش نسبی در مقادیر خطا در داده‌های آزمون، ضریب تعیین این مدل همچنان بسیار بالا باقی مانده است (۹۹ صدم در آزمون و عدد کامل یک در آموزش)، و همچنین میانگین درصد خطای مطلق در آزمون تنها حدود ۲ درصد بوده است. این نتایج بیانگر آن است که مدل جنگل تصادفی علاوه بر دقت بالا در یادگیری، توانایی تعمیم مناسب به داده‌های جدید را نیز دارا بوده و از بروز بیش‌برازش جلوگیری کرده است.

۴-۲- مقایسه روش‌های یادگیری ماشین در بازه زمانی ۱۳۹۹-۱۳۹۴

در این پژوهش، دو روش یادگیری ماشین رگرسیون بردار پشتیبان و جنگل تصادفی برای پیش‌بینی بازده سهام در بازه زمانی ۱۳۹۴ تا ۱۳۹۹ مورد بررسی قرار گرفتند. معیارهای ارزیابی عملکرد شامل میانگین مربعات خطا، جذر میانگین مربعات خطا، میانگین قدر مطلق خطا، ضریب تعیین، میانگین قدر مطلق درصد خطا است. در ادامه، نتایج هر معیار بررسی می‌شود، جدول ۵ نتایج پیش بینی هر یک از روش‌های یادگیری ماشین در بازه زمانی ۱۳۹۴ تا ۱۳۹۹ را نشان می دهد.

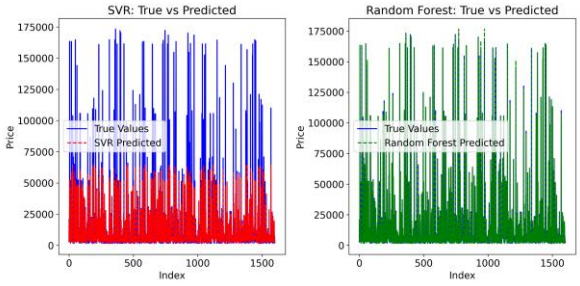
مطابق با جدول ۵ مدل جنگل تصادفی خطای بسیار کمتری نسبت به رگرسیون بردار پشتیبان دارد. این اختلاف بزرگ نشان

جدول۵. مقایسه روش‌های یادگیری ماشین در بازه زمانی ۱۳۹۴-۱۳۹۹

معیار ارزیابی	رگرسیون بردار پشتیبان	جنگل تصادفی	اختلاف
میانگین مربعات خطا	۳۵۵،۶۷۴،۵۵۴،۱۴	۱۶۲،۲۱۷،۱۸	۳۵۵،۵۱۲،۳۳۶،۹۶
جذر میانگین مربعات خطا	۱۸،۸۵۹،۳۴	۴۰۲،۷۶	۱۸،۴۵۶،۵۷
میانگین قدر مطلق خطا	۴،۷۵۱،۵۸	۱۵۶،۰۱	۴،۵۹۵،۵۶
میانگین قدر مطلق درصد خطا	۱۷،۱۰٪	۰،۹۳٪	۱۶،۲۰٪
ضریب تعیین	۰،۶۲	۱،۰۰	۰،۳۸

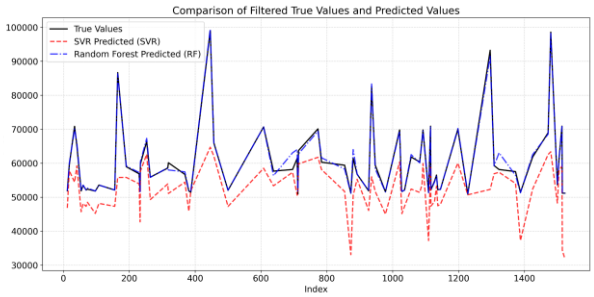
می‌دهد که مدل جنگل تصادفی در پیش‌بینی دقیق‌تر عمل کرده و توانسته نوسانات قیمت را بهتر مدل‌سازی کند. میانگین خطای مطلق برای جنگل تصادفی بسیار کمتر از رگرسیون بردار پشتیبان است (۱۵۶،۰۱ در مقابل ۴،۷۵۱،۵۸)، که نشان‌دهنده دقت بالاتر جنگل تصادفی در مقایسه با رگرسیون بردار پشتیبان است. درصد

خطای مطلق مدل جنگل تصادفی فقط ۰،۹۳٪ است که نشان‌دهنده پیش‌بینی‌هایی بسیار نزدیک به مقادیر واقعی است، در حالی که میانگین قدر مطلق درصد خطا مدل رگرسیون بردار پشتیبان برابر با ۱۷،۱۰٪ است و ضعف آن در پیش‌بینی را آشکار می‌کند. مقدار ۱،۰۰ برای جنگل تصادفی نشان‌دهنده این است که مدل توانسته تقریباً تمام واریانس داده‌های واقعی را توضیح دهد. در مقابل، مقدار ۰،۶۲ برای رگرسیون بردار پشتیبان نشان‌دهنده محدودیت این مدل در درک تغییرات داده‌ها است. مدل جنگل تصادفی به‌طور قابل‌توجهی عملکرد بهتری نسبت به رگرسیون بردار پشتیبان در بازه ۱۳۹۴-۱۳۹۹ داشته است. با خطاهای کمتر و دقت بالاتر در معیارهای میانگین مربعات خطا، Rمیانگین مربعات خطا، میانگین قدر مطلق خطا و میانگین قدر مطلق درصد خطا ، جنگل تصادفی توانسته نوسانات قیمت را بهتر پیش‌بینی کند و برای این بازه زمانی به‌عنوان روش برتر شناخته می‌شود. مطابق با نمودار (۱) نمودار مقایسه پیش‌بینی قیمت سهام با مدل‌های رگرسیون بردار پشتیبان و جنگل تصادفی نشان می‌دهد که مدل جنگل تصادفی در تطابق با مقادیر واقعی و پیش‌بینی نوسانات عملکرد بهتری داشته است. مدل رگرسیون بردار پشتیبان با پیش‌بینی‌های ثابت و محدود نتوانسته نوسانات پیچیده را شبیه‌سازی کند، در حالی که جنگل تصادفی به دلیل دقت بالا و توانایی بهتر در دنبال کردن تغییرات، به‌عنوان روش برتر برای بازه زمانی ۱۳۹۴-۱۳۹۹ شناخته می‌شود. همچنین نتایج نشان داده شده در نمودار (۲) به وضوح نشان از برتری قابل توجه جنگل تصادفی در این بازه زمانی است.



نمودار۱. مقایسه عملکرد پیش بینی رگرسیون بردار پشتیبان با جنگل تصادفی (۱۳۹۹-۱۳۹۴)

مطابق با نمودار (۲) نمودار مقایسه پیش‌بینی قیمت سهام با مدل‌های رگرسیون بردار پشتیبان و جنگل تصادفی نشان می‌دهد که مدل جنگل تصادفی در تطابق با مقادیر واقعی و پیش‌بینی نوسانات عملکرد بهتری داشته است. مدل رگرسیون بردار پشتیبان با پیش‌بینی‌های ثابت و محدود نتوانسته نوسانات پیچیده را شبیه‌سازی کند، در حالی که جنگل تصادفی به دلیل دقت بالا و توانایی بهتر در دنبال کردن تغییرات، به‌عنوان روش برتر برای بازه زمانی ۱۳۹۴-۱۳۹۹ شناخته می‌شود. همچنین نتایج نشان داده شده در نمودار (۲) به وضوح نشان از برتری قابل توجه جنگل تصادفی در این بازه زمانی است.



نمودار۲. مقایسه عملکرد پیش بینی رگرسیون بردار پشتیبان با جنگل تصادفی (۱۳۹۹-۱۳۹۴)

۴-۳- مقایسه روش‌های یادگیری ماشین در بازه زمانی ۱۴۰۰-۱۴۰۲:

در این پژوهش، دو روش یادگیری ماشین رگرسیون بردار پشتیبان و جنگل تصادفی برای پیش‌بینی بازده سهام در بازه زمانی ۱۴۰۰ تا ۱۴۰۲ مورد بررسی قرار گرفتند. معیارهای ارزیابی عملکرد شامل میانگین مربعات خطا، میانگین مربعات خطا، میانگین قدر مطلق خطا، میانگین قدر مطلق درصد خطا و ضریب تعیین است. در ادامه، نتایج هر معیار بررسی می‌شود، جدول ۶ نتایج پیش بینی هر یک از روش‌های یادگیری ماشین در بازه زمانی ۱۴۰۰ تا ۱۴۰۲ را نشان می دهد. مطابق با جدول ۶ در بازه زمانی ۱۴۰۰-۱۴۰۲، عملکرد دو مدل رگرسیون بردار پشتیبان و جنگل تصادفی در پیش‌بینی قیمت سهام بر اساس معیارهای مختلف مورد بررسی قرار گرفت. نتایج نشان‌دهنده برتری واضح جنگل تصادفی نسبت به رگرسیون بردار پشتیبان بود:

میانگین مربعات خطا معیار خطای پیش‌بینی مدل را نمایش می‌دهد و هرچه کمتر باشد، عملکرد بهتر است. مقدار میانگین مربعات خطا برای جنگل تصادفی برابر با ۹۰۱،۸۷۵,۷۷ بود که نشان‌دهنده خطای بسیار کم نسبت به مقدار ۸۲،۹۳۰،۳۲۰,۰۰ برای رگرسیون بردار پشتیبان است. این اختلاف نشان‌دهنده توانایی بالای جنگل تصادفی در کاهش خطای پیش‌بینی است. Rمیانگین مربعات خطا معیار خطای پیش‌بینی را در همان واحد داده نشان می‌دهد و مقدار کمتر به معنای دقت بالاتر است. Rمیانگین مربعات خطا برای جنگل تصادفی برابر با ۹۴۹,۶۷۱ عملکرد بسیار بهتری نسبت به میانگین مربعات خطا برای رگرسیون بردار پشتیبان برابر با ۹،۱۰۶,۶۱ دارد.

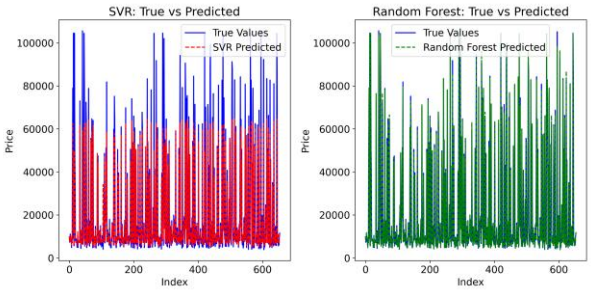
جدول ۶ مقایسه روش‌های یادگیری ماشین در بازه زمانی ۱۴۰۲-۱۴۰۰

معیار ارزیابی	رگرسیون بردار پشتیبان	جنگل تصادفی	اختلاف
میانگین مربعات خطا	۸۲،۹۳۰،۳۳۲,۰۰	۹۰۱،۸۷۵,۷۷	۸۲،۰۲۸،۴۴۴,۲۳
میانگین مربعات خطا	۹،۱۰۶,۶۱	۹۴۹,۶۷	۸،۱۵۶,۹۳۹
میانگین قدر مطلق خطا	۳،۴۵۵,۴۲	۴۳۳,۴۴	۳،۰۲۱,۹۷۸
میانگین قدر مطلق درصد خطا	۱۶,۶۰٪	۱,۹۸٪	۱۴,۶۲٪
ضریب تعیین	۰,۸۷	۰,۹۹	-۰,۱۲

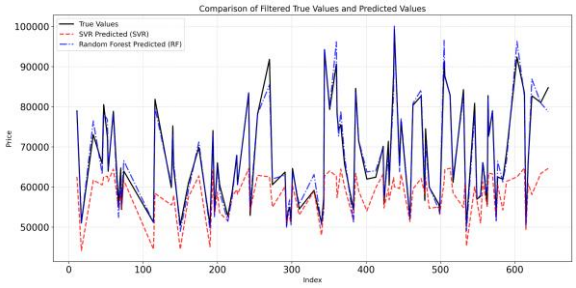
میانگین قدر مطلق خطا معیار میانگین اختلافات مطلق پیش‌بینی‌ها با مقادیر واقعی را محاسبه می‌کند. مقدار کمتر بهتر است و جنگل تصادفی با میانگین قدر مطلق خطا برابر با ۴۳۳,۴۴۲ نسبت به مقدار ۳،۴۵۵,۴۲ برای رگرسیون بردار پشتیبان، پیش‌بینی‌های دقیق‌تری ارائه داد. میانگین قدر مطلق درصد خطا معیار درصد خطای مطلق را نسبت به مقادیر واقعی نشان می‌دهد و مقدار کمتر، دقت بالاتر مدل را نشان می‌دهد. جنگل تصادفی با میانگین قدر مطلق درصد خطا برابر با ۱,۹۸٪ در مقایسه با مقدار ۱۶,۶۰٪ برای رگرسیون بردار پشتیبان، خطای نسبی بسیار کمتری داشت. ضریب تعیین معیار نشان‌دهنده درصد واریانس توضیح داده شده توسط مدل است و هرچه به ۱ نزدیک‌تر باشد، بهتر است. مقدار ضریب تعیین برای جنگل تصادفی برابر با ۰,۹۹۹ بود که نشان‌دهنده قدرت بسیار بالای مدل در توضیح واریانس داده‌ها است، در حالی که مقدار ۰,۸۷ برای رگرسیون بردار پشتیبان بیانگر محدودیت در توانایی پیش‌بینی دقیق این مدل بود. در مجموع، جنگل تصادفی در تمام معیارها عملکرد بهتری از خود نشان داد و با خطای کمتر و دقت بالاتر،

مدل مناسبی برای پیش‌بینی قیمت سهام در این بازه زمانی بود.

در نمودار (۳) مقایسه‌ی پیش‌بینی‌های رگرسیون بردار پشتیبان و جنگل تصادفی در برابر مقادیر واقعی قیمت سهام (۱۴۰۰-۱۴۰۲)، خط آبی (جنگل تصادفی) نسبت به خط قرمز (رگرسیون بردار پشتیبان) تطابق بیشتری با خط سیاه (مقادیر واقعی) دارد. جنگل تصادفی توانسته تغییرات قیمت را بهتر دنبال کند، در حالی که رگرسیون بردار پشتیبان در پیش‌بینی نوسانات شدید ضعیف عمل کرده است. همچنین نتایج نشان داده شده در نمودار (۴) به وضوح نشان از برتری قابل توجه جنگل تصادفی در این بازه زمانی است.



نمودار۳. مقایسه عملکرد پیش بینی رگرسیون بردار پشتیبان با جنگل تصادفی (۱۴۰۲-۱۴۰۰)



نمودار۴. مقایسه عملکرد پیش بینی رگرسیون بردار پشتیبان با جنگل تصادفی (۱۴۰۲-۱۴۰۰)

۴-۴- بررسی تاثیر تغییر دامنه نوسان قیمت سهام بر پیش‌بینی قیمت سهام:

این بخش به بررسی تأثیر تغییر دامنه نوسان قیمت سهام بر دقت پیش‌بینی قیمت سهام با استفاده از مدل جنگل تصادفی می‌پردازد. دو بازه زمانی مورد مطالعه قرار گرفته‌اند: پیش از تغییر دامنه نوسان و دوره‌ای شامل تغییرات مکرر دامنه نوسان. نتایج نشان می‌دهند که جنگل تصادفی به دلیل انعطاف‌پذیری و توانایی بالا در مدل‌سازی روابط غیرخطی، عملکرد بهتری نسبت به رگرسیون بردار پشتیبان داشته است.

برای تحلیل، معیارهای میانگین مربعات خطا، Rمیانگین مربعات خطا، میانگین قدر مطلق خطا، میانگین قدر مطلق درصد خطا و ضریب تعیین به‌کار گرفته شده‌اند تا دقت پیش‌بینی در هر بازه و تفاوت عملکرد مدل در شرایط مختلف بررسی شود. این تحلیل، تأثیر سیاست تغییر دامنه نوسان بر پیش‌بینی قیمت سهام را مشخص کرده و اطلاعات مفیدی برای سیاست‌گذاران و تحلیل‌گران بازار سرمایه ارائه می‌دهد. جدول ۷ به مقایسه خروجی جنگل تصادفی در دو بازه زمانی این پژوهش پرداخته است.

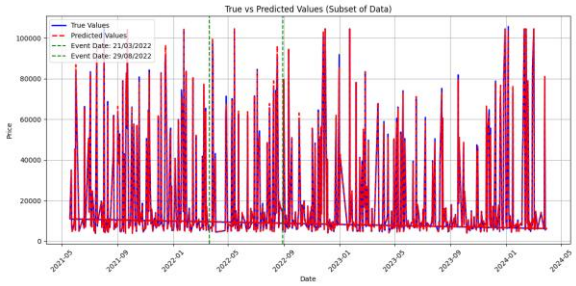
جدول ۷. مقایسه خروجی جنگل تصادفی در دو بازه زمانی

معیار ارزیابی	۱۳۹۹-۱۳۹۴	۱۴۰۲-۱۴۰۰	اختلاف
میانگین مربعات خطا	۱۶۲,۲۱۷,۱۸	۹۰۱,۸۷۵,۷۷	۷۳۹,۶۵۸,۵۹-
میانگین مربعات خطا	۴۰۲,۷۶۲	۹۴۹,۶۷	-۵۴۶,۹۰۹
میانگین قدر مطلق خطا	۱۵۶,۰۱۴	۴۳۳,۴۴	-۲۷۷,۴۲۸
میانگین قدر مطلق درصد خطا	۰,۹۳٪	۱,۹۸٪	-۱,۰۵
ضریب تعیین	۱,۰۰	۰,۹۹	۰,۰۱

مطابق با جدول ۷ نتایج معیارهای ارزیابی مدل جنگل تصادفی در دو بازه زمانی (۱۳۹۹-۱۳۹۴ و ۱۴۰۲-۱۴۰۰) نشان می‌دهد که میانگین مربعات خطا و میانگین مربعات خطا در بازه دوم (۱۴۰۲-۱۴۰۰)، افزایش قابل‌توجهی داشته‌اند که نشان‌دهنده کاهش دقت مدل در این بازه است.میانگین قدر مطلق خطا میانگین خطای مطلق پیش‌بینی‌ها در بازه دوم بیشتر است، که به نوسانات بیشتر بازار و تغییرات سریع قیمت‌ها مرتبط است. میانگین قدر مطلق درصد خطا درصد خطای مطلق از ۰,۹۶٪ به ۱,۹۸٪ افزایش یافته است. اگرچه این افزایش نشان‌دهنده کاهش دقت نسبی مدل است، اما همچنان مقدار میانگین قدر مطلق درصد خطا زیر ۲٪ است که عملکرد کلی مدل را قابل‌قبول می‌کند. ضریب تعیین در هر دو بازه نزدیک به ۱ باقی مانده است، که بیانگر قدرت بالای مدل در توضیح تغییرات داده‌ها است. در مجموع، معیارهای خطا نشان‌دهنده کاهش دقت مدل در بازه دوم هستند که احتمالاً به دلیل نوسانات بیشتر بازار و کمبود داده‌ها بوده است. با این حال، مقدار بالای ضریب تعیین عملکرد کلی مدل را تأیید می‌کند.

طبق نمودار (۵) با توجه به تغییرات دامنه نوسان قیمت سهام در تاریخ‌های ۲۱ مارس ۲۰۲۲ و ۲۹ آگوست ۲۰۲۲، پیش‌بینی‌های مدل با مقادیر واقعی تفاوت بیشتری پیدا کرده‌اند. نقاط پیش‌بینی

شده (قرمز) از مقادیر واقعی (آبی) جدا شده و انحرافات بیشتری را نشان می‌دهند. این تغییرات ممکن است ناشی از نوسانات بیشتر بازار و تغییرات غیرقابل پیش‌بینی در داده‌ها باشد. باوجود مشابه بودن برخی پیش‌بینی‌ها و مقادیر واقعی، خطاها (میانگین مربعات خطا) افزایش یافته‌اند و این کاهش دقت مدل را در شرایط نوسانی نشان می‌دهد.



نمودار۵. مقایسه داده‌های واقعی و پیش بینی روش جنگل تصادفی با بررسی تاریخ‌ها تغییر دامنه نوسان قیمت

۵- بحث

یافته‌های این پژوهش نشان می‌دهد که الگوریتم جنگل تصادفی در پیش‌بینی قیمت سهام در هر دو بازه زمانی بررسی‌شده، عملکرد بهتری نسبت به الگوریتم رگرسیون بردار پشتیبان داشته است. این تفاوت را می‌توان به ماهیت الگوریتم جنگل تصادفی نسبت داد که با استفاده از تجمیع نتایج چندین درخت تصمیم و تکنیک تصادفی‌سازی، توانایی بالاتری در مدل‌سازی روابط پیچیده و غیرخطی بازار سهام دارد. در مقابل، رگرسیون بردار پشتیبان به‌عنوان یک مدل حساس به تنظیم پارامترهای کرنل (مانند C و γ)، در صورت انتخاب نامناسب پارامترها ممکن است دچار بیش‌برازش شود. این مسئله در نتایج مشاهده‌شده نیز مشهود است؛ به‌ویژه زمانی که خطای مدل در داده‌های آموزش نسبت به تست به‌طور غیرمعمول پایین‌تر است. از سوی دیگر، بررسی نتایج نشان می‌دهد که در بازه زمانی دوم (۱۴۰۰ تا ۱۴۰۲) که دامنه نوسان قیمت تغییر یافته، دقت هر دو مدل کاهش یافته است. علت اصلی این کاهش را می‌توان در برهم خوردن ساختار رفتاری و الگویی بازار دانست که ناشی از سیاست‌گذاری‌های جدید و واکنش‌های هیجانی سرمایه‌گذاران بوده است. به بیان دیگر، مدل‌هایی که بر پایه داده‌های تاریخی با الگوهای ثابت آموزش دیده‌اند، در مواجهه با ساختار جدید بازار کارایی پیشین خود را از دست داده‌اند. همچنین، با تحلیل مقادیر ضریب تعیین میزان خطاهای مدل‌ها در داده‌های آموزش و آزمون، می‌توان بیان کرد که در برخی موارد مدل رگرسیون بردار پشتیبان احتمالاً دچار بیش‌برازش شده است، در حالی که الگوریتم جنگل تصادفی با ساختار مقاوم‌تری که دارد، توانسته تعادل بهتری میان دقت و تعمیم‌پذیری ایجاد کند. در مجموع، با در نظر گرفتن ماهیت پویای بازار سرمایه ایران و تغییرات سیاست‌گذاری مانند تغییر دامنه نوسان، به نظر می‌رسد استفاده از الگوریتم‌های ترکیبی و مقاوم‌سازی مدل‌ها در برابر تغییرات ساختاری، چشم‌اندازی مناسب برای تحقیقات آینده در این حوزه باشد.

۶- نتیجه‌گیری

نتایج این پژوهش که به ارزیابی و مقایسه عملکرد مدل‌های یادگیری ماشین رگرسیون بردار پشتیبان و جنگل تصادفی در پیش‌بینی بازده سهام فرابورس ایران و تأثیر تغییر دامنه نوسان قیمت در دو بازه زمانی (۱۳۹۴-۱۳۹۹ و ۱۴۰۰-۱۴۰۲) پرداخته است، نشان داد که مدل جنگل تصادفی عملکرد بهتری نسبت به مدل رگرسیون بردار پشتیبان داشت. معیارهای خطا مانند میانگین مربعات خطا، میانگین مربعات خطا، میانگین قدر مطلق خطا و میانگین قدر مطلق درصد خطا برای مدل جنگل تصادفی کمتر بودند که نشان از دقت بالاتر آن است، همچنین ضریب تعیین برای مدل جنگل تصادفی در هر دو بازه زمانی نزدیک به ۱ سبود. این نشان می‌دهد که مدل جنگل تصادفی قادر است تغییرات قیمت سهام را به‌دقت پیش‌بینی کند و عملکرد آن در مقایسه با مدل رگرسیون بردار پشتیبان بهتر است. مدل رگرسیون بردار پشتیبان به دلیل حساسیت به نوسانات کوچک قیمت و احتمال overfiting، عملکرد ضعیف‌تری داشت، به‌ویژه در بازه ۱۴۰۲-۱۴۰۰ که نوسانات بیشتر بود. همچنین، تغییر دامنه نوسان قیمت تأثیر منفی بر دقت پیش‌بینی‌ها داشت و سبب افزایش معیارهای خطا شد. این امر نشان می‌دهد که دامنه نوسان وسیع‌تر می‌تواند منجر به نوسانات بیشتر در پیش‌بینی‌ها شود که مدل‌ها را با چالش روبرو می‌کند. با این حال، مدل جنگل تصادفی برتری خود را در پیش‌بینی قیمت سهام حفظ کرده است و توانست به‌خوبی بر نوسانات بازار غلبه کند.

۷- پیشنهادها و محدودیت‌ها

برای بهبود دقت پیش‌بینی مدل‌ها، پیشنهاد می‌شود از بازه‌های زمانی هم‌طول با مدل‌های ترکیبی برای داده‌های نوسانی استفاده گردد. این مدل‌ها می‌توانند به‌طور بهینه‌تری از اطلاعات موجود بهره ببرند و نتایج دقیق‌تری ارائه دهند. همچنین، افزایش ویژگی‌های جدید به‌ویژه عواملی که تأثیرات اقتصادی و اخبار بازار را در نظر می‌گیرند، می‌تواند به بهبود پیش‌بینی‌ها کمک کند.

بهینه‌سازی پارامترهای مدل‌ها و گنجاندن ویژگی‌های اقتصادی و اطلاعات کلان‌اقتصادی می‌تواند به‌طور چشمگیری دقت پیش‌بینی‌ها را افزایش دهد.

یکی از محدودیت‌های این پژوهش، تفاوت در طول بازه‌های زمانی بود. بازه اول (۱۳۹۴-۱۳۹۹) برای تحلیل دقیق‌تر طولانی‌تر انتخاب شد، اما به دلیل محدودیت دسترسی به داده‌ها، طول بازه دوم (۱۴۰۰-۱۴۰۲) محدودتر بود. این تفاوت در طول بازه‌ها ممکن است بر معیارهای حساس به تعداد داده‌ها تأثیر بگذارد و نتایج مدل را تحت تأثیر قرار دهد. در صورت دسترسی به داده‌های بیشتر در بازه دوم، می‌توان نتایج دقیق‌تر و بهتری ارائه داد.

مراجع

[1] R. D. Singh, H. Chouhan, and M. Neema, "Predicting stock prices using machine learning: A comparative study," *International Journal of Advances in Engineering and Management*, vol. 5, no. 4, pp. 1812-1821, 2023.

[2] P. L. Sagar, C. Maheshbabu, G. V. S. Teja, and C. Vishnu, "An evaluation of stock market prediction using supervised machine learning techniques," in *Proc. 2023 Int. Conf. Inventive Computation Technol. (ICICT)*, 2023, pp. 255-263, doi: 10.1109/ICICT57646.2023.10133970.

[3] A. Muratov, R. Szabo, and R. Varadi, "The impact of adverse selection on the quotation strategy of stock market specialists: A Monte Carlo simulation approach," *Journal of Economic Dynamics & Control*, vol. 106, pp. 249-261, 2019, doi: 10.1016/j.jedc.2019.05.002.

[4] M. Sohrabi, S. M. Mir Barzegar, and E. Chirani, "Comparison of different machine learning models in stock market index prediction," *Financial Engineering and Securities Management Quarterly*, vol. 14, no. 56, pp. 19-36, 2023.

[5] A. Eghtesad and M. Mohammadi, "Portfolio optimization by return prediction using LSTM, random forest, and ARIMA," *Financial Management Perspective*, vol. 13, no. 43, pp. 9-28, 2023, doi: 10.48308/jfmp.2024.104191.

[6] D. Riahi and A. Mazighi, "The effect of fluctuation limit changes on actual price volatility in the Tehran Stock Exchange," M.S. thesis, University of Tehran, Tehran, Iran, 2022.

[7] M. Dehghani, M. Ghasemzadeh, and H. Ansari-Samani, "Machine learning algorithms for time series in financial markets," *Soft Computing and Information Technology Journal*, vol. 8, no. 3, pp. 60-67, 2019.

[8] N. Kuhestani, I. Dadashi, H. Gholamnia Roshan, and K. Azinfar, "Day features and stock market performance," *Financial Management Perspective*, vol. 13, no. 44, pp. 33-58, 2024, doi: 10.48308/jfmp.2024.104307.

[9] N. Gholami and N. Shams Gharni, "An optimized CNN-LSTM model for stock price prediction in the Tehran Stock Exchange," *Financial Management Perspective*, vol. 14, no. 45, pp. 123-147, 2024, doi: 10.48308/jfmp.2024.104892.

[10] M. Nourahmadi and H. Sadeghi, "Machine learning based on the hierarchical risk parity approach (Case study: Stock portfolio of the top 30 companies listed on the Tehran Stock Exchange)," *Financial Research Journal*, vol. 24, no. 2, pp. 236–256, 2022.

[11] A. Chakravorty and N. Elsayed, "A Comparative Study of Machine Learning Algorithms for Stock Price Prediction Using Insider Trading Data," *arXiv*, Feb. 12, 2025.

[12] A. Dezhkam, M. T. Manzuri, A. Aghapour, A. Karimi, A. Rabiee, and S. M. Shalmani, "A Bayesian-based classification framework for financial time series trend prediction," *The Journal of Supercomputing*, vol. 79, pp. 4622–4659, 2023, doi: 10.1007/s11227-022-04746-8.

[13] S. V. Shashikala and H. S. Harshitha, "Stock Market Analysis and Prediction using Machine Learning Algorithm," *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, vol. 11, no. 7, pp. 1092–1097, 2023, doi: 10.22214/ijraset.2023.54812

[14] C. L. de Souza, "Previsão de Preços de Ações com Aprendizado de Máquina," ResearchGate, May 2024.

[15] M. Sui, C. Zhang, L. Zhou, S. Liao, and C. Wei, "An Ensemble Approach to Stock Price Prediction Using Deep Learning and Time Series Models," Preprints, vol. 2024, no. 09, p. 2077, 2024, doi: 10.20944/preprints202409.2077.v1

[16] R. He, Z. Zhang, S. Zhang, and Q. Song, "Evaluating Regression and Ensemble Models for Financial Forecasting: The Case of Apple Stock," ResearchGate, Nov. 2023.

[17] A. Chakravorty and N. Elsayed, "A Comparative Study of Machine Learning Algorithms for Stock Price Prediction Using Insider Trading Data," arXiv, Feb. 12, 2025

[18] C. L. de Souza, "Previsão de Preços de Ações com Aprendizado de Máquina," ResearchGate, May 2024.

[19] K. Ł. Szydłowski and J. A. Chudziak, "Hidformer: Transformer-Style Neural Network in Stock Price Forecasting," in USB Proceedings of the 21st International Conference on Modeling Decisions for Artificial Intelligence (MDAI 2024), Tokyo, Japan, Aug. 2024.

[20] R. Kumar et al., "The Effectiveness of Technical Indicators in Stock Price Prediction," International Journal of Financial Analysis, vol. 8, pp. 120–135, 2021.

[21] A. Muratov, R. Szabo, and R. Varadi, "The impact of adverse selection on the quotation strategy of stock market specialists: A Monte Carlo simulation approach," J. Econ. Dyn. Control, vol. 106, pp. 249–261, 2019.

[22] P. L. Sagar, C. Maheshbabu, G. V. S. Teja, and C. Vishnu, "An Evaluation of Stock Market Prediction using Supervised Machine Learning Techniques," in Proc. 2023 Int. Conf. Inventive Comput. Technol. (ICICT), Apr. 2023, doi: 10.1109/ICICT57646.2023.10133970.

[23] N. Kuhestani, I. Dadashi, H. Gholamnia Roshan, and K. Azinfar, "Daily features and stock market performance," Financial Management Perspective, vol. 13, no. 44, pp. 33-58, 2024, doi: 10.48308/jfmp.2024.104307.

[24] M. Hassannejad, M. Davalu, and F. Shabani, "The effects of price limit changes on intraday volatility in the Iranian stock market using realized variance (RV) and discrete Fourier transform (DFT)," Asset Management and Financial Engineering, vol. 11, no. 2, pp. 19-34, 2023, doi: 10.22108/amf.2023.134778.1755.

[25] S. Behrampour and K. Kiani, "Stock market trend prediction using data mining in big data," M.S. thesis, Semnan University, Electrical Engineering, 2015.

[26] I. Davoudi Kasbi, I. Dadashi, and K. Azinfar, "Stock price analysis using machine learning methods," Financial Advances and Mathematical Applications, vol. 6, no. 2, pp. 285-301, 2021.

[27] Sh. Bagheri and H. Izadbaksh, "Stock market prediction using machine learning, deep learning, and econometric methods: A comparative study," M.S. thesis, Shahid Beheshti University, 2021.

[28] A. Li, Q. Wei, Y. Shi, and Z. Liu, "Research on stock price prediction from a data fusion perspective," Data Science in

Finance and Economics, vol. 3, no. 3, pp. 230-250, 2023, doi: 10.3934/DSFE.2023014.

[29] Md. S. Ansary, "Machine learning for predicting the stock price direction with trading indicators," ENP Engineering Science Journal, vol. 3, no. 2, pp. 34-40, 2023, doi: 10.53907/enpesj.v3i2.178.

[30] W. Bao, Y. Cao, Y. Yang, H. Che, J. Huang, and S. Wen, "Data-driven stock forecasting models based on neural networks: A review," Information Fusion, vol. 102, art. 102616, 2024, doi: 10.1016/j.inffus.2024.102616.

[31] S. Pashankar, J. Shendage, and J. Pawar, "Machine learning techniques for stock price prediction: A comparative analysis of linear regression, random forest, and support vector regression," Journal of Advanced Zoology, vol. 45, no. S4, pp. 118-127, 2024, doi: 10.53555/jaz.v45iS4.4164.

[32] T. Li, F. Ma, X. Zhang, and Y. Zhang, "Economic policy uncertainty and the Chinese stock market volatility: Novel evidence," Economic Modelling, vol. 87, pp. 24-33, 2020, doi: 10.1016/j.econmod.2019.07.002.

[33] P. Gao, R. Zhang, and X. Yang, "The application of stock index price prediction with neural network," Mathematics and Computers in Applications, vol. 25, no. 3, art. 53, 2020, doi: 10.3390/mca25030053.

[34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[35] M. Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms*, 2nd ed. Hoboken, NJ, USA: Wiley, 2011, doi: 10.1002/9781118029145.

[36] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock market index using fusion of machine learning techniques," Expert Systems with Applications, vol. 42, no. 4, pp. 1-9, 2014, doi: 10.1016/j.eswa.2014.10.031.

پاورقی‌ها:

[1] Overfitting