



Sensitivity Analysis of Machine Learning Models in Predicting the Tehran Stock Exchange Index: The Impact of Input Parameters on Performance

Fatemeh Ansari

Ph.D. Candidate, Department of Financial Economics, Faculty of Economics and Management, Urmia University, Urmia, Iran. E-mail: f.ansari@urmia.ac.ir

Shahab Jahangiri *

*Corresponding Author, Associate Prof., Department of Economics, Faculty of Economics and Management, Urmia University, Urmia, Iran. E-mail: kh.jahangiri@urmia.ac.ir

Ali Rezazadeh

Associate Prof., Department of Economics, Faculty of Economics and Management, Urmia University, Urmia, Iran. E-mail: a.rezazade@urmia.ac.ir

Abstract

Objective

This study aims to evaluate the sensitivity of machine learning models to input variables and identify the most significant factors influencing the prediction of the Tehran Stock Exchange index. Additionally, it compares the performance of different models in forecasting the stock index, focusing on the impact of input parameters, and offers strategies for optimizing input data and reducing model complexity. The key input variables considered include open, high, low prices, and trading volume.

Methods

In this research, due to the presence of a target variable (closing price), a supervised learning approach is employed. Furthermore, because the target variable is continuous, the problem is defined as a regression task. Predictions from four powerful machine learning algorithms—

Citation: Ansari, Fatemeh; Jahangiri, Shahab & Rezazadeh, Ali (2026). Sensitivity Analysis of Machine Learning Models in Predicting the Tehran Stock Exchange Index: The Impact of Input Parameters on Performance. *Financial Research Journal*, 28(3), 836 – 859. <https://doi.org/10.22059/frj.2025.387933.1007689> (in Persian)



Linear Model (LM), Support Vector Regression (SVR), Artificial Neural Network (ANN), and Random Forest (RF)—were compared using performance evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and the coefficient of determination (R^2). MAE represents the average magnitude of errors, while MSE indicates the differences between the predicted and actual values. Lower values for both metrics suggest greater accuracy of predictions. The R^2 statistic represents the percentage of variance in the data that is explained by the model, with values close to one indicating a higher level of accuracy. The dataset used in this study consists of six selected indices from various sectors, including pharmaceutical, automotive, financial, food industries, basic metals, and petroleum products, covering the period from 2020 to 2024 daily. Sensitivity analysis was conducted to evaluate the importance of input variables such as open, high, low prices, and trading volume on the predictions of each model.

Results

The sensitivity analysis results revealed that the linear regression, random forest, and artificial neural network models exhibited no response to the trading volume input variable across any of the six indices. However, support vector regression (SVR) demonstrated negative sensitivity to the input parameters in the financial index. This may suggest selling pressure, a weakening bullish trend, or a negative market sentiment favoring safer assets such as bonds and bank deposits. In contrast, other models, including random forest, artificial neural networks (ANN), and linear regression, displayed relatively similar sensitivity to the input variables. These models rely more directly on input parameters and provide a clearer representation of the relative importance of variables in predicting outcomes.

Conclusion

This study identified the key variables influencing stock index fluctuations and demonstrated that more complex models, such as random forest, neural networks, and support vector regression, are more effective in accurately predicting stock indices in the pharmaceutical, petroleum, and food sectors due to their higher sensitivity. These findings can contribute to the enhancement of forecasting and decision-making models in the capital market. Sensitivity analysis aids in identifying critical variables while eliminating redundant features, thereby improving computational efficiency and reducing model complexity. The insights from this research can be valuable for investors and economic policymakers in making more informed decisions.

Keywords: Random forest, Sensitivity analysis, Support vector regression.

تحلیل حساسیت مدل‌های یادگیری ماشین در پیش‌بینی شاخص بورس تهران: تأثیر پارامترهای ورودی بر کارایی

فاطمه انصاری

دانشجوی دکتری، گروه اقتصاد مالی، دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ارومیه، ایران. رایانامه: f.ansari@urmia.ac.ir

شهاب جهانگیری*

نویسنده مسئول، دانشیار، گروه اقتصاد، دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ارومیه، ایران. رایانامه: kh.jahangiri@urmia.ac.ir

علی رضازاده

دانشیار، گروه اقتصاد، دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ارومیه، ایران. رایانامه: a.rezazade@urmia.ac.ir

چکیده

هدف: این پژوهش به ارزیابی حساسیت مدل‌های یادگیری ماشین به متغیرهای ورودی و شناسایی مهم‌ترین متغیرهای تأثیرگذار در پیش‌بینی شاخص بورس تهران پرداخته است. همچنین، مقایسه کارایی مدل‌ها در پیش‌بینی شاخص بورس با تمرکز بر تأثیر پارامترهای ورودی و ارائه راه‌کارهایی برای بهینه‌سازی داده‌های ورودی و کاهش پیچیدگی مدل‌ها، از دیگر اهداف این پژوهش است. متغیرهای ورودی عبارت‌اند از: قیمت‌های باز، بالا، پایین و حجم معاملات.

روش: در پژوهش حاضر به دلیل وجود یک متغیر هدف و ناظر که همان متغیر قیمت پایانی (Close) است، درمی‌یابیم که باید از رویکرد یادگیری نظارت شده استفاده شود. از سوی دیگر، به دلیل کمی یا پیوسته بودن مقیاس متغیر هدف، مسئله به فرم رگرسیون تعریف شده است و در نهایت، پیش‌بینی‌های حاصل از چهار الگوریتم قدرتمند یادگیری ماشین (مدل خطی، رگرسیون بردار پشتیبان، شبکه عصبی مصنوعی و جنگل تصادفی) با استفاده از توابع میانگین خطای مطلق (MAE)، میانگین مربعات خطا (MSE) و ضریب تعیین R^2 با یکدیگر مقایسه و مدل نهایی برای بهترین مدل پیش‌بینی ارائه شده است. معیار خطای مطلق، به نوعی نشان‌دهنده متوسط اندازه خطاها و معیار میانگین مربعات خطا، گویای اختلاف بین مقادیر پیش‌بینی شده و واقعی است؛ بنابراین هرچه مقدار این دو معیار کمتر باشد، دقت بیشتر پیش‌بینی‌ها را نشان می‌دهد. ضریب تعیین R^2 نشان‌دهنده درصدی از واریانس داده‌هاست که توسط مدل توضیح داده می‌شود. مقدار نزدیک به ۱، دقت بسیار زیاد را نشان می‌دهد. داده‌های استفاده‌شده، شامل شش شاخص روزانه منتخب از صنایع دارویی، خودروبی، مالی، صنایع غذایی، فلزات اساسی و فراورده‌های نفتی، طی سال‌های ۱۳۹۹ تا ۱۴۰۳ است. برای تحلیل حساسیت، اهمیت متغیرهای ورودی مانند قیمت‌های باز، بالا، پایین و حجم معاملات بر پیش‌بینی خروجی هر مدل بررسی شد.

استناد: انصاری، فاطمه؛ جهانگیری، شهاب و رضازاده، علی (۱۴۰۵). تحلیل حساسیت مدل‌های یادگیری ماشین در پیش‌بینی شاخص بورس تهران: تأثیر پارامترهای ورودی بر کارایی. *تحقیقات مالی*، ۲۸ (۳)، ۸۳۶ - ۸۵۹.

تاریخ دریافت: ۱۴۰۳/۱۰/۱۱

تاریخ ویرایش: ۱۴۰۳/۱۱/۱۷

تاریخ پذیرش: ۱۴۰۴/۰۷/۳۰

تاریخ انتشار: ۱۴۰۵/۰۷/۰۹

doi: <https://doi.org/10.22059/frj.2025.387933.1007689>

تحقیقات مالی، ۱۴۰۵، دوره ۲۸، شماره ۳، صص. ۸۳۶ - ۸۵۹

ناشر: دانشکدان مدیریت دانشگاه تهران

نوع مقاله: علمی پژوهشی

© نویسندگان

یافته‌ها: نتایج تحلیل حساسیت نشان داد که مدل‌های خطی، جنگل تصادفی و شبکه عصبی مصنوعی، در هیچ‌یک از شش شاخص نسبت به متغیر ورودی حجم معاملات واکنشی نشان ندادند. رگرسیون بردار پشتیبان (SVR) به پارامترهای ورودی در شاخص مالی حساسیت منفی نشان داد. این موضوع می‌تواند نشان‌دهنده فشار فروش، ضعف روند صعودی یا جو روانی منفی در بازار مالی، به سمت دارایی‌های امن‌تر مثل اوراق قرضه و سپرده‌های بانکی باشد. در مقابل، مدل‌های دیگر شامل جنگل تصادفی، شبکه عصبی مصنوعی (ANN) و رگرسیون خطی، حساسیت نسبتاً مشابهی به متغیرهای ورودی نشان دادند. این مدل‌ها به پارامترهای ورودی وابستگی مستقیم‌تری دارند و اهمیت نسبی متغیرها در پیش‌بینی خروجی را بهتر نمایان می‌کنند.

نتیجه‌گیری: این پژوهش به شناسایی مهم‌ترین متغیرهای تأثیرگذار بر تغییرات شاخص بورس کمک می‌کند و نشان می‌دهد که مدل‌های پیچیده‌تر، مانند جنگل تصادفی، شبکه عصبی و رگرسیون بردار پشتیبان، در شاخص دارویی و فراورده‌های نفتی و غذایی به دلیل حساسیت بیشتر، می‌توانند در پیش‌بینی دقیق‌تر شاخص‌های بورس مؤثرتر عمل کنند. این نتایج می‌تواند به بهبود طراحی مدل‌های پیش‌بینی و تصمیم‌گیری در بازار سرمایه کمک کند. تحلیل حساسیت به شناسایی متغیرهای مهم و حذف ویژگی‌های غیرضروری کمک می‌کند که به افزایش سرعت و کاهش پیچیدگی مدل‌ها منجر می‌شود. نتایج این پژوهش می‌تواند برای سرمایه‌گذاران و سیاست‌گذاران اقتصادی، به منظور تصمیم‌گیری‌های آگاهانه‌تر مفید باشد.

کلیدواژه‌ها: تحلیل حساسیت، جنگل تصادفی، رگرسیون بردار پشتیبان.

مقدمه

بازار بورس به‌عنوان یکی از شاخص‌های اقتصادی بسیار مهم، در تخصیص منابع مالی و ارزیابی سلامت اقتصادی کشورها نقش کلیدی ایفا می‌کند (فرمان‌آرا، کمیجانی، فرزین‌شو و غفاری، ۱۳۹۸). همه می‌خواهند در کوتاه‌ترین زمان ممکن، درآمد کسب کنند. بازار سهام می‌تواند یکی از ابزارهای تحقق چنین رؤیاهایی باشد. بخش عمده‌ای از جامعه تصور می‌کنند که کسب درآمد از بازار سهام، بازی کودکانه است؛ اما واقعیت نشان می‌دهد که اگر بتوان اطلاعات مربوط به قیمت‌های سهام قبلی را به‌خوبی پیش‌پردازش و الگوریتم‌های مناسبی را اعمال کرد، می‌توان سود شایان توجهی به‌دست آورد (خانا، کولشرستا، سینگ، تاوکار و شریواستاوا^۱، ۲۰۲۲). مالکیل و فاما^۲ (۱۹۷۰) نشان دادند که پیش‌بینی داده‌های قیمت سهام بر اساس معاملات، امکان‌پذیر است؛ زیرا قیمت‌ها منعکس‌کننده تمام اطلاعات و نااطمینانی‌ها هستند. برای پیش‌بینی بازار سهام به‌کمک داده‌های تاریخی، روش‌های مختلفی وجود دارد. در یک طبقه‌بندی، این روش‌ها در دو گروه خطی و غیرخطی قرار می‌گیرند و در نوع دیگر، به روش‌های یادگیری ماشین و آماری دسته‌بندی می‌شوند؛ اما دسته‌بندی مناسب، گروه‌بندی آن‌ها به روش‌های هوشمند و کلاسیک است (پاتل، شاه، تاکار و کوتچا^۳، ۲۰۱۵). پژوهش‌ها در زمینه یادگیری ماشین، در حال پیشرفت است و بررسی و توانایی دقت آن‌ها برای پیش‌بینی سری‌های زمانی، بیش از پیش اهمیت دارد (سهرابی، میربرگ‌کار، چیرانی و خردیار، ۱۴۰۲).

پیش‌بینی شاخص‌های بورس، به‌دلیل پیچیدگی و رفتار غیرخطی بازارهای مالی، همواره مورد توجه پژوهشگران و سرمایه‌گذاران بوده است. در این راستا، مدل‌های یادگیری ماشین به‌عنوان ابزارهای قدرتمند برای تحلیل داده‌های مالی و شناسایی الگوهای پیچیده معرفی شده‌اند (فیشر و کراوس^۴، ۲۰۱۸). این مدل‌ها از متغیرهای ورودی متعددی برای پیش‌بینی استفاده می‌کنند و تأثیر هر یک از این متغیرها بر عملکرد مدل، به‌درستی مشخص نیست. درک میزان حساسیت مدل‌ها به متغیرهای ورودی نظیر قیمت باز، بالا، پایین و حجم معاملات، می‌تواند نقش مهمی در بهینه‌سازی مدل‌ها و کاهش پیچیدگی آن‌ها داشته باشد.

در پژوهش‌های پیشین، مدل‌های مختلف یادگیری ماشین مانند رگرسیون خطی، جنگل تصادفی، شبکه عصبی مصنوعی و رگرسیون بردار پشتیبان در پیش‌بینی شاخص‌های مالی به‌کار گرفته شده‌اند (آیالا، گارسیا-تورس، نوگوئرا، گومز ولا و دیوینا^۵، ۲۰۲۱، پاتل و همکاران، ۲۰۱۵، قلی‌زاده کپورچالی، فلاح و مینویی، ۱۴۰۳). با این حال، تحلیل حساسیت این مدل‌ها به متغیرهای ورودی و تأثیر دقیق این ویژگی‌ها بر عملکرد مدل‌ها، در کمتر مطالعه‌ای بررسی شده است. بیشتر پژوهش‌ها تمرکز خود را روی دقت پیش‌بینی معطوف کرده‌اند و از بررسی چگونگی تأثیر پارامترهای ورودی بر پیش‌بینی مدل‌ها غافل مانده‌اند. این مسئله، به‌ویژه در بازار بورس تهران که تحت تأثیر عوامل اقتصادی، سیاسی و نوسان‌های بازار قرار دارد، اهمیت بیشتری می‌یابد. تحلیل حساسیت مدل‌های یادگیری ماشین، می‌تواند راهنمای

1. Khanna, Kulshrestha, Singh, Thawkar & Shrivastava
2. Malkiel & Fama
3. Patel, Shah, Thakkar & Kotecha
4. Fischer & Krauss
5. Ayala, García-Torres, Noguera, Gómez-Vela & Divina

ارزشمندی برای شناسایی متغیرهای کلیدی و بهبود پیش‌بینی شاخص‌ها ارائه دهد. اهدافی که این پژوهش دنبال می‌کند، عبارت‌اند از:

۱. ارزیابی حساسیت مدل‌های یادگیری ماشین، شامل رگرسیون خطی، جنگل تصادفی، شبکه عصبی مصنوعی و رگرسیون بردار پشتیبان به متغیرهای ورودی؛

۲. شناسایی مهم‌ترین متغیرهای ورودی (مانند قیمت باز، بالا، پایین و حجم معاملات) در پیش‌بینی شاخص بورس تهران؛

۳. مقایسه کارایی مدل‌ها در پیش‌بینی شاخص بورس با تمرکز بر تأثیر پارامترهای ورودی؛

۴. ارائه راه‌کارهایی برای بهینه‌سازی داده‌های ورودی و کاهش پیچیدگی مدل‌ها.

در این پژوهش سؤال‌های زیر مطرح شده است:

۱. کدام‌یک از متغیرهای ورودی، بر عملکرد مدل‌های یادگیری ماشین در پیش‌بینی شاخص بورس بیشترین تأثیر را دارد؟

۲. آیا حساسیت مدل‌های پیچیده‌تر مانند شبکه عصبی و جنگل تصادفی در مقایسه با مدل‌های ساده‌تر مانند رگرسیون خطی و SVR متفاوت است؟

۳. چگونه تحلیل حساسیت می‌تواند به بهبود پیش‌بینی و کاهش پیچیدگی مدل‌های یادگیری ماشین کمک کند؟

۴. کدام مدل یادگیری ماشین دقت و کارایی بیشتری در پیش‌بینی شاخص بورس تهران دارد؟

بدین ترتیب، فرضیه‌های پژوهش به‌صورت زیر مطرح شده‌اند:

فرضیه ۱. مدل‌های پیچیده‌تر مانند جنگل تصادفی و شبکه عصبی حساسیت بیشتری به متغیرهای ورودی دارند.

فرضیه ۲. برخی از متغیرهای ورودی (مانند قیمت باز و حجم معاملات) روی پیش‌بینی شاخص بورس تأثیر بیشتری دارند.

فرضیه ۳. تحلیل حساسیت می‌تواند متغیرهای غیرمؤثر را شناسایی و به کاهش پیچیدگی مدل‌ها کمک کند.

فرضیه ۴. حساسیت مدل‌ها به متغیرهای ورودی، می‌تواند با توجه به ساختار آن‌ها متفاوت باشد و عملکرد کلی مدل‌ها را تحت تأثیر قرار دهد.

ضرورت‌هایی که باعث شده است این پژوهش انجام شود، عبارت‌اند از:

۱. پیشرفت در پیش‌بینی شاخص بورس: این پژوهش می‌تواند به بهبود پیش‌بینی‌های مالی و کاهش خطاهای پیش‌بینی کمک کند.

۲. بهینه‌سازی مدل‌ها: تحلیل حساسیت به شناسایی متغیرهای مهم و حذف ویژگی‌های غیرضروری کمک می‌کند و به افزایش سرعت و کاهش پیچیدگی مدل‌ها منجر می‌شود.

۳. کاربرد در سیاست‌گذاری و سرمایه‌گذاری: نتایج این پژوهش می‌تواند برای سرمایه‌گذاران و سیاست‌گذاران اقتصادی به‌منظور تصمیم‌گیری‌های آگاهانه‌تر مفید باشد.

با تمرکز بر تحلیل حساسیت مدل‌های یادگیری ماشین، این پژوهش گامی نوین در ادبیات پیش‌بینی بازارهای مالی و به‌ویژه بورس تهران خواهد بود. این پژوهش با تمرکز بر تحلیل حساسیت، نه‌تنها دقت مدل‌های یادگیری ماشین را بررسی کرده است، بلکه به درک عمیق‌تری از تأثیر متغیرهای ورودی بر پیش‌بینی شاخص‌ها پرداخته است که در ادبیات موجود کمتر به آن پرداخته شده است و نوآوری این پژوهش محسوب می‌شود.

پیشینه نظری پژوهش

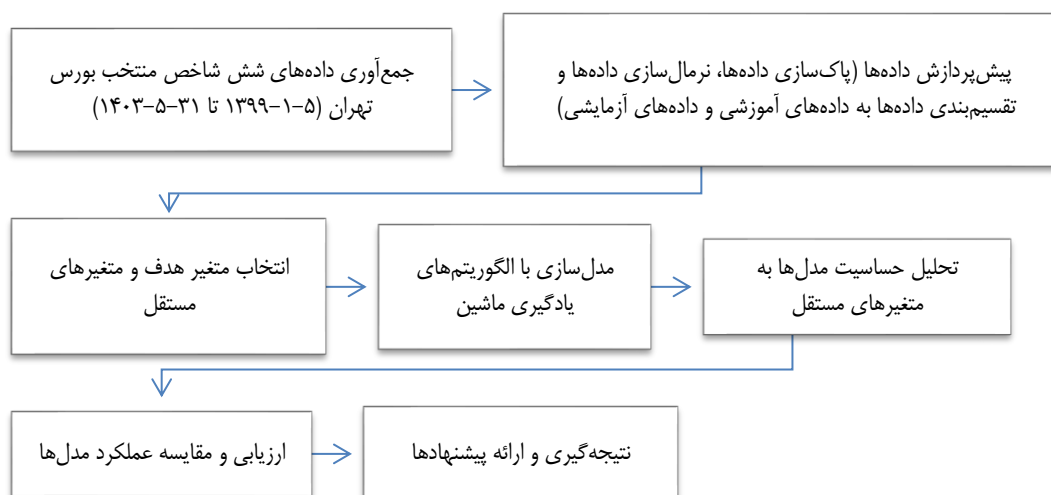
فلسفه پشت یادگیری ماشینی، استخراج دانش از داده‌هاست (کوبات^۱، ۲۰۱۷: ۱). مشکل اصلی پیش‌بینی مسیر قیمت سهام، تعداد زیاد داده‌هاست. محبوبیت داده‌کاوی در دنیای مالی رو به افزایش است. مجموعه داده‌ها به‌قدری عظیم است که نمی‌توان آن را با استفاده از روش‌های غیر داده‌کاوی اداره کرد؛ به‌طوری که آن‌ها معنای اصلی را مبهم می‌کنند و نمی‌توان اطلاعات مفیدی از آن‌ها به‌دست آورد (تسیناسلانی‌دیس و کوگیموتزیس^۲، ۲۰۱۴).

یادگیری ماشین، شاخه‌ای از فناوری هوش مصنوعی است که از شیوه‌ها یا الگوریتم‌های خاص برای ساخت خودکار مدل‌هایی از داده استفاده می‌کند. برخلاف یک سیستم معمولی که با پیروی از قوانین مشخص، یک وظیفه را انجام می‌دهد، یک سیستم یادگیری ماشین، از طریق تجربه‌کردن وظایف خود را یاد می‌گیرد. در حالی که یک سیستم مبتنی بر قانون، هر بار یک وظیفه را به‌روشی یکسان (چه خوب، چه بد) انجام می‌دهد، وضعیت عملکرد یک سیستم یادگیری ماشین از طریق یادگیری و تمرین، از طریق الگوریتم‌های داده‌کاوی بهبود پیدا می‌کند. الگوریتم‌های یادگیری ماشین با فراگیری از داده‌هایی که دریافت می‌کنند، می‌توانند مسائلی را حل کنند که بسیار پیچیده‌تر از آن است که بتوان آن را با برنامه‌نویسی معمولی حل کرد (لی^۳، ۲۰۱۹).

به‌طور کلی، مدل‌های یادگیری ماشین از دو پارامتر تشکیل شده‌اند: پارامترهای مدل و ابرپارامترها. پارامترهای مدل توسط مدل در مرحله آموزش آموخته می‌شوند. در مقابل، ابرپارامترهای مدل یادگیری ماشین، باید در ابتدای آموزش پیکربندی شوند و در طول آموزش، ممکن است تغییر کند (توقف زود هنگام و کاهش سرعت یادگیری) یا ثابت بماند (شراتز، مونچو و ایتوریتسا^۴، ۲۰۱۹؛ هاتر، کوتوف و ونشورن^۵، ۲۰۱۹).

در این آزمایش، ابرپارامترهای مدل ثابت فرض شده است و پارامترهای مدل که همان ورودی ما محسوب می‌شوند (ویژگی‌ها X و برچسب Y) به‌بوته آزمایش گذاشته شده‌اند. این مطالعه تجربی به‌طور کامل با استفاده از پایتون اجرا شده است.

1. Kubat
2. Tsinaslanidis & Kugiumtzis
3. Lee
4. Schratz, Muenchow & Iturritxa
5. Hutter, Kotthoff & Vanschoren



شکل ۱. چارچوب کلی پژوهش

پیشینه تجربی پژوهش

پژوهش‌های زیادی در داخل و خارج، از اندیکاتورهای مختلف تکنیکی و الگوریتم‌های یادگیری ماشین، برای پیش‌بینی روند بازار سهام بهره گرفته‌اند و با مرور پیشینه پژوهش‌های انجام شده، به نظر می‌رسد که این اولین پژوهشی است که به ارزیابی حساسیت مدل‌های یادگیری ماشین به متغیرهای ورودی و شناسایی مهم‌ترین متغیرهای تأثیرگذار در پیش‌بینی شاخص بورس تهران پرداخته است. در نتیجه در این بخش، کارهای مرتبط انجام شده در پیش‌بینی قیمت سهام با استفاده از مدل‌های یادگیری ماشین بررسی خواهد شد.

هوک و الجمان^۱ (۲۰۲۱) یک ارزیابی تجربی از هشت مدل یادگیری ماشین مرسوم برای پیش‌بینی قیمت سهام یازده شرکت متعلق به بورس اوراق بهادار عربستان انجام دادند. عملکرد پیش‌بینی ارزیابی شده توسط دو معیار خطای معروف ریشه میانگین مربعات خطا (RMSE) و میانگین درصد مطلق خطا (MAPE) انجام شده است. نتایج نشان داد که پس از تنظیم ابرپارامترها، رگرسیون بردار پشتیبان با اختلاف آماری چشمگیری از سایر مدل‌های پیش‌بینی بهتر عمل کرده است.

چاجر، شاه و کشیرساگر^۲ (۲۰۲۱) در مطالعه‌ای با عنوان «کاربردهای شبکه‌های عصبی مصنوعی، ماشین‌های بردار پشتیبان و حافظه بلندمدت کوتاه‌مدت (LSTM) برای پیش‌بینی بازار سهام»، بیان کردند که بازار سهام آشفته است؛ اما استفاده از هوش مصنوعی برای انجام پیش‌بینی‌ها، قبل از سرمایه‌گذاری امکان‌پذیر و توصیه می‌شود.

خان و همکاران^۳ (۲۰۲۲) از الگوریتم‌های هوشمند یادگیری ماشین برای پیش‌بینی روند شاخص‌های بورس

1. Hoque & Aljamaan

2. Chaajer, Shah & Kshirsagar

3. Khan et al

نیویورک و لندن بهره بردند و از اطلاعات موجود در رسانه‌های اجتماعی و اخبار مالی استفاده کردند. نتایج نشان داد که این الگوریتم‌ها برای پیش‌بینی روند شاخص بسیار توانمندند.

سنگیتا و الیفّا^۱ (۲۰۲۴) در مقاله‌ای با عنوان «پیش‌بینی بازار سهام مالی با استفاده از تکنیک یادگیری ماشینی مبتنی بر رگرسیون خطی ارزیابی شده»، از یادگیری ماشین نظارت شده ELR پیشنهادی، روی مجموعه داده‌ای که از یاهوفايننس به دست آمده، استفاده کردند. این مجموعه داده، پنج متغیر دارد که عبارت است از: باز، بسته، کم، زیاد و حجم. نتایج پژوهش نشان داد که با انتخاب یک مدل پیش‌بینی کارآمد، معامله‌گران می‌توانند تصمیم بگیرند که آیا یک شیء را بخرند یا بفروشند.

آجیگا و همکاران^۲ (۲۰۲۴) به بررسی نقش محوری انتخاب ویژگی و مهندسی در افزایش قدرت پیش‌بینی مدل‌های ML پرداختند. مجموعه ویژگی‌ها عبارت بودند از: شاخص‌های مالی، متغیرهای کلان اقتصادی، تحلیل احساسات از مقاله‌های خبری و داده‌های رسانه‌های اجتماعی برای تأثیر آن‌ها بر دقت پیش‌بینی. این بررسی درک دقیقی از چشم‌انداز مدل‌های ML برای پیش‌بینی بازار سهام ارائه داد و رویکردها، چالش‌ها و فرصت‌های متنوع را در تلاش برای بهبود دقت برجسته کرد.

فلاح‌پور، گل‌ارضی و فتوره‌چیان (۱۳۹۲)، در پژوهشی روند حرکتی قیمت سهام را با استفاده از ماشین بردار پشتیبان بر پایه الگوریتم ژنتیک در بورس اوراق بهادار تهران پیش‌بینی کردند. این پژوهش نشان داد که ماشین بردار پشتیبان بر پایه الگوریتم ژنتیک، در مقایسه با ماشین بردار پشتیبان ساده، برای پیش‌بینی حرکت قیمت سهام از دقت بسیار بیشتری برخوردار است.

مرادزاده‌فر، دارابی و شاهعلی‌زاده (۱۳۹۳) به یکپارچه‌سازی تکنیک‌های هوش مصنوعی جهت ارائه مدل پیش‌بینی قیمت سهام پرداختند. نتایج این پژوهش برتری مدل ترکیبی نسبت به سایر مدل‌های مورد بررسی را نشان داد. افشاری‌راد، علوی و سینایی (۱۳۹۷) به پیش‌بینی روند با روش‌های تحلیل تکنیکال و روش‌های هوشمند یادگیری ماشین پرداختند و برای پیش‌بینی از شاخص کل استفاده کردند. در نهایت این نتیجه حاصل شد که روش پیشنهادی به‌طور متوسط نرخ پیش‌بینی صحیح ۹۷ درصد دارد.

مرادی، بحری ثالث، جبارزاده کنگرلویی و آشتاب (۱۴۰۱) بیست‌وسه عامل مشخص شده در مطالعات کتابخانه‌ای را بر اساس داده‌های ۱۵۴ شرکت فعال در فاصله زمانی ۱۳۸۸ تا ۱۳۹۸ استخراج و در قالب دو خوشه افراز کردند. ارزیابی ارتباط متغیرها در مدل‌های یادگیری ماشینی نشان داد که جریان نقدی غیرعادی، هزینه اختیاری غیرعادی، خطای تخمین اقلام تعهدی، تفاوت بین سرمایه در گردش تحقق یافته و موردانتظار، سهام شناور آزاد، دوره تصدی حسابرس، حق الزحمه حسابرسی، سهم بازار حسابرس، محافظه‌کاری و تغییر حسابرس، در خوشه‌بندی بیشترین تأثیر را دارند.

سهرابی و همکاران (۱۴۰۲) توانایی مدل‌های مختلف جنگل تصادفی، شبکه عصبی مصنوعی، شبکه عصبی

1. Sangeetha & Alfia

2. Ajiga et al

بازگشتی و ماشین بردار پشتیبان را بررسی کردند. نتایج پژوهش آن‌ها نشان داد که شبکه عصبی بازگشتی حافظه طولانی کوتاه مدت (LSTM) در مقایسه با سایر مدل‌ها نتیجه بهتری داشته است.

ایزدی، آشتاب و زواری رضایی (۱۴۰۳) با استفاده از اطلاعات مالی ۱۳۹ شرکت تولیدی پذیرفته شده در بورس اوراق بهادار تهران، در دوره زمانی ۱۵ ساله، طی سال‌های ۱۳۸۷ تا ۱۴۰۱ و با به کارگیری ۲۵ مدل یادگیری ماشین و ۱۰ مدل آماری، به بررسی مقایسه کارایی مدل‌های یادگیری ماشین و مدل‌های آماری در پیش‌بینی جهت تغییرات اجزای سود از جمله سود (زیان) خالص، سود (زیان) ناخالص و سود (زیان) عملیاتی پرداختند. نتایج آزمون فرضیه‌های پژوهش، نویددهنده کارایی بالای مدل‌های یادگیری ماشین در پیش‌بینی جهت تغییرات سود (زیان) خالص، سود (زیان) ناخالص و سود (زیان) عملیاتی نسبت به مدل‌های آماری است.

این پژوهش با تمرکز بر تحلیل حساسیت مدل‌های یادگیری ماشین، گامی نوین در ادبیات پیش‌بینی بازارهای مالی و به‌ویژه بورس تهران برداشته است. این پژوهش، نه تنها دقت مدل‌های یادگیری ماشین را بررسی کرده، بلکه به درک عمیق‌تری از تأثیر متغیرهای ورودی بر پیش‌بینی شاخص‌ها پرداخته است که در ادبیات موجود کمتر به آن پرداخته شده است.

روش‌شناسی پژوهش

یادگیری ماشین معمولاً به سه دسته پیش‌بینی‌کننده (یادگیری نظارت شده)، توصیف‌کننده (یادگیری بدون ناظر) و یادگیری تقویتی تقسیم می‌شود. در یادگیری نظارت‌شده، هدف پیدا کردن الگویی مناسب برای پیش‌بینی متغیر هدف بر اساس مجموعه‌ای از متغیرهای پیش‌گو یا خصیصه یا ویژگی است. معمولاً متغیر هدف را با y به‌عنوان خروجی و متغیرهای پیش‌گو یا ویژگی‌ها را با x نشان داده و از مدل زیر استفاده می‌شود.

$$D = \{(x_i, y_i)\}_{i=1}^N \quad \text{رابطه ۱}$$

اما در حالت دوم که یادگیری بدون ناظر است، هدف پیدا کردن الگویی مناسب برای توصیف هرچه بهتر روابط بین داده‌هاست. در حالت یادگیری فاقد نظارت، مجموعه داده‌های آموزشی متغیر هدف ندارد و ما تنها با مجموعه‌ای از ورودی‌ها سروکار داریم (قلی‌زاده کیورچالی و همکاران، ۱۴۰۳).

$$D = \{x_i\}_{i=1}^N \quad \text{رابطه ۲}$$

متغیرهای وابسته و مستقل

قیمت پایانی (Close)، به‌عنوان متغیر هدف و متغیرهای قیمت باز (Open)، بیشترین قیمت (High)، کمترین قیمت (Low) و حجم معاملات (Volume) به‌عنوان متغیرهای مستقل انتخاب شده‌اند. دلایل انتخاب متغیر Close به‌عنوان متغیر وابسته (هدف) عبارت‌اند از:

۱. نمایشگر عملکرد روزانه بازار: قیمت پایانی، به عنوان یکی از مهم‌ترین شاخص‌های بورس، نمایانگر ارزش نهایی سهم در پایان روز معاملاتی است و معمولاً در تحلیل‌های مالی به عنوان معیار اصلی ارزیابی وضعیت بازار در نظر گرفته می‌شود.

۲. مبنای تصمیم‌گیری سرمایه‌گذاران: سرمایه‌گذاران اغلب بر اساس قیمت پایانی روز قبل، استراتژی‌های معاملاتی خود را برای روز بعد برنامه‌ریزی می‌کنند. بنابراین پیش‌بینی قیمت پایانی اهمیت زیادی برای تصمیم‌گیری‌های مالی دارد.

۳. معیار اصلی در شاخص‌های مالی: بسیاری از شاخص‌های بازار بورس و ارزش‌گذاری‌ها بر اساس قیمت پایانی محاسبه می‌شوند. این موضوع باعث می‌شود که قیمت پایانی متغیر وابسته مناسبی برای مدل‌سازی و پیش‌بینی باشد.

متغیرهای Low, High, Open و Volume به دلایل زیر به عنوان متغیرهای مستقل انتخاب شده‌اند:

۱. ارتباط مستقیم با رفتار قیمت پایانی: قیمت باز (Open) نشان‌دهنده نقطه شروع معاملات است و اغلب اطلاعات مهمی درباره روند اولیه بازار ارائه می‌دهد که می‌تواند تأثیر مستقیمی بر قیمت پایانی داشته باشد. بیشترین قیمت (High) در طول روز، نشان‌دهنده سطوح مقاومت بازار و نوسان‌های روزانه است که بر روند و سطح قیمت پایانی تأثیرگذار است. کمترین قیمت (Low) می‌تواند بیانگر سطح حمایت بازار باشد و اطلاعات ارزشمندی درباره رفتار بازار در طول روز ارائه دهد. حجم معاملات (Volume) نشان‌دهنده فشار خرید و فروش حجم معاملات است. حجم معاملات به عنوان یکی از مهم‌ترین متغیرهای بازار، منعکس‌کننده میزان علاقه و فعالیت سرمایه‌گذاران در بازار است. تغییرات حجم معاملات می‌تواند تأثیر مستقیمی بر قیمت‌ها، به ویژه قیمت پایانی، داشته باشد؛

۲. کاهش پیچیدگی و تمرکز بر ویژگی‌های کلیدی: انتخاب این متغیرها از میان سایر ویژگی‌های ممکن (مانند داده‌های خارجی یا متغیرهای کلان اقتصادی)، به دلیل ارتباط مستقیم آن‌ها با قیمت پایانی و نقش کلیدی آن‌ها در نوسان‌های روزانه است. این متغیرها به طور مستقیم بر رفتار قیمت تأثیر می‌گذارند و در عین حال باعث افزایش کارایی و کاهش پیچیدگی تحلیل می‌شوند؛

۳. کاربرد در تحلیل حساسیت: این متغیرها ویژگی‌های مستقیم و کلیدی معاملات روزانه هستند و تحلیل حساسیت آن‌ها می‌تواند به درک دقیق‌تری از چگونگی تأثیر هر یک از این متغیرها بر کارایی مدل‌های پیش‌بینی کمک کند؛

۴. مطابقت با ادبیات پژوهش: در بسیاری از مطالعات پیشین مرتبط با پیش‌بینی شاخص‌های بورس، این متغیرها به عنوان ورودی‌های اصلی مدل‌سازی مورد استفاده قرار گرفته‌اند (فیشر و کراوس، ۲۰۱۸؛ آیالا و همکاران، ۲۰۲۱؛ سنگیتا و ایفا، ۲۰۲۴) که نشان‌دهنده اهمیت و اعتبار آن‌ها در پیش‌بینی قیمت پایانی است.

در پژوهش حاضر، به دلیل وجود یک متغیر هدف و ناظر که همان متغیر قیمت پایانی (Close) است، درمی یابیم که باید از رویکرد یادگیری نظارت شده استفاده شود و از سوی دیگر، به دلیل کمی یا پیوسته بودن مقیاس متغیر هدف، مسئله به فرم رگرسیون تعریف شده است و در نهایت پیش بینی های حاصل از چهار الگوریتم قدرتمند یادگیری ماشین (مدل خطی (LM)، رگرسیون بردار پشتیبان (SVR)، شبکه عصبی مصنوعی (ANN) و جنگل تصادفی (RF)) با استفاده از توابع میانگین خطای مطلق (MAE)، میانگین مربعات خطا (MSE) و ضریب تعیین R^2 با یکدیگر مقایسه و مدل نهایی به جهت بهترین مدل پیش بینی ارائه شده است. معیار MAE به نوعی نشان دهنده متوسط اندازه خطاها و معیار MSE بیانگر اختلاف بین مقادیر پیش بینی شده و واقعی است؛ بنابراین هرچه مقدار این دو معیار کمتر باشد، دقت بیشتر پیش بینی ها را نشان می دهد. ضریب تعیین R^2 نشان دهنده درصدی از واریانس داده ها است که توسط مدل توضیح داده می شود. مقدار نزدیک به یک نشان دهنده دقت بالا است.

مدل رگرسیون خطی

مدل های خطی از جمله ساده ترین و پرکاربردترین روش ها در یادگیری ماشین هستند که برای پیش بینی مقادیر پیوسته (رگرسیون) و دسته بندی داده ها (طبقه بندی) استفاده می شوند. در مدل های خطی، فرض بر این است که متغیر هدف یک ترکیب خطی از متغیرهای مستقل است. انواع مدل های خطی در یادگیری ماشین، شامل رگرسیون خطی (برای پیش بینی مقادیر پیوسته استفاده می شود و مناسب برای مسائلی است که رابطه خطی قوی بین متغیرهای مستقل و وابسته دارند)، رگرسیون لجستیک (برای مسائل طبقه بندی مثلاً باینری استفاده می شود و از توابع احتمال (مانند سیگموئید) برای پیش بینی دسته بندی استفاده می کند) و رگرسیون چندگانه (شامل چندین متغیر مستقل برای پیش بینی متغیر وابسته) است.

مدل جنگل تصادفی

روش جنگل تصادفی جزئی از روش های طبقه بندی تحت نظارت گروهی در حوزه داده کاوی است. جنگل تصادفی یا به اختصار RF، پتانسیل بسیار زیادی برای تبدیل به یک روش محبوب برای طبقه بندی های آینده دارد زیرا کارایی آن قابل مقایسه با دیگر روش های گروهی می باشد. به عنوان یک الگوریتم گروهی (رأی گیری) RF چندین درخت تصمیم مختلف به عنوان طبقه بندی های پایه تولید می کند و رأی اکثریت را برای ترکیب با نتایج درختان اصلی اعمال می کند (سعادت و زمانی بروجنی، ۱۳۹۵).

مدل رگرسیون بردار پشتیبان

برای حل مسائل طبقه بندی باینری، ولادیمیر^۱ از آزمایشگاه های AT&T Bell، پس از معرفی ماشین بردار پشتیبانی (SVM) آن را به عنوان یک مسئله بهینه سازی محدب فرموله کرد (واپنیک^۲، ۱۹۹۹). برای حل مشکلات رگرسیون،

1. Vladimir
2. Vapnik

رگرسیون بردار پشتیبان (SVR) با مشخص کردن ناحیه فشرده ε در اطراف تابع که می‌تواند مقادیر ادامه را با پیچیدگی مدل معقول تقریبی کند، بازسازی شد. علاوه بر این، SVR مسائل رگرسیون را به عنوان یک مسئله بهینه‌سازی فرموله می‌کند که با کاهش خطای پیش‌بینی، باریک‌ترین مسیر ممکن را در اطراف سطح جست‌وجو می‌کند (آواد و خانا، ۲۰۱۵). این شرط را می‌توان به صورت معادله زیر نشان داد (اسمولا و شولکوف، ۲۰۰۴):

$$\text{رابطه ۳} \quad \text{minimize } \frac{1}{2} ||w||^2$$

استفاده از SVR برای پیش‌بینی‌های سری‌های زمانی چالش‌هایی دارد؛ چون این مدل به‌طور مستقیم نمی‌تواند وابستگی‌های زمانی را مدل کند. برای حل این مشکل، باید داده‌ها را به شکلی پیش‌پردازش کرد که اطلاعات زمان‌بندی به مدل منتقل شود؛ از جمله ایستایی (مانایی) متغیرها، استفاده از پنجره‌های زمانی، مقیاس‌بندی داده‌ها و تقسیم داده‌ها به مجموعه آموزش و آزمون.

مدل شبکه عصبی مصنوعی

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

$$\text{رابطه ۴} \quad Net_t = a_0x_0 + a_1x_{1t} + a_2x_{2t} = \sum_{i=0}^2 a_ix_{it}$$

سپس نرون خروجی، ارزش به دست آمده را با استفاده از یک تابع تبدیل یا فعال‌سازی (محرک) پردازش می‌کند که با f نشان داده می‌شود (مکیان و موسوی، ۱۳۹۰).

روش‌های جلوگیری از بیش‌برازشی

کنترل بیش‌برازشی^۳ یکی از مسائل اساسی در مدل‌های یادگیری ماشین است و زمانی به وجود می‌آید که مدل بیش از حد به داده‌های آموزش وابسته شده و در داده‌های جدید عملکرد ضعیفی دارد. در نتیجه برای رفع این مشکل با تنظیم مدل یا استفاده از روش‌های کاهش پیچیدگی که در جدول زیر آمده، مشکل بیش‌برازشی قبل از تخمین مدل برطرف شده است.

1. Awad and Khanna
2. Smola & Schölkopf
3. overfitting

جدول ۱. روش‌های جلوگیری از بیش‌برازشی برای هر مدل

روش مناسب کنترل بیش‌برازشی	مدل
کاهش تعداد درخت‌ها (n_estimators)، افزایش حداقل تعداد نمونه‌ها در هر برگ (min_samples_leaf)، افزایش عمق درخت (max_depth)	جنگل تصادفی
تنظیم دقیق c، استفاده از epsilon برای کاهش حساسیت به نویز، تست هسته‌های مختلف (kernel=linear, rbf)	رگرسیون بردار پشتیبان
Dropout، L1/L2 Regularization (kernel_regularizer)، افزایش داده‌ها (Augmentation Data)	شبکه عصبی مصنوعی
Lasso Ridge Regression (L2 Regularization) (Linear Regression, Ridge, Lasso) Regression (L1 Regularization)	مدل خطی

معرفی داده‌های پژوهش

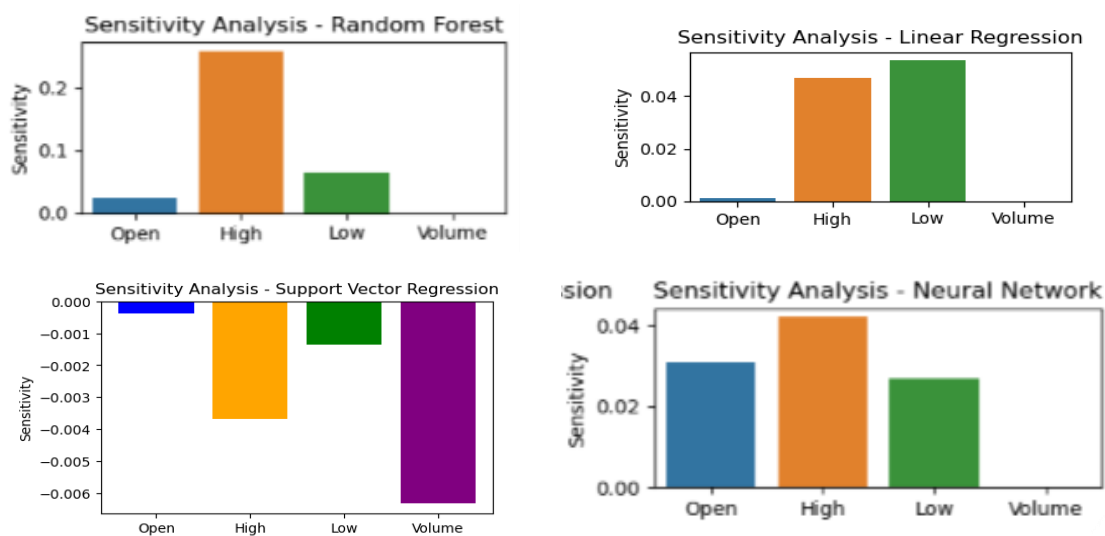
نمونه مورد مطالعه شامل شش شاخص منتخب از گروه‌های فلزات، فراورده‌های نفتی، خودرویی، دارویی، غذایی و واسطه‌گری مالی خواهد بود. گروه‌های مورد نظر بخش عمده ارزش بازار سهام تهران را به خود اختصاص می‌دهد و بنابراین می‌تواند به‌عنوان نماینده از کل بازار در نظر گرفته شوند. داده‌ها از سایت بورس اوراق بهادار تهران جمع‌آوری شده است.

جدول ۲. خلاصه داده‌های شاخص‌ها

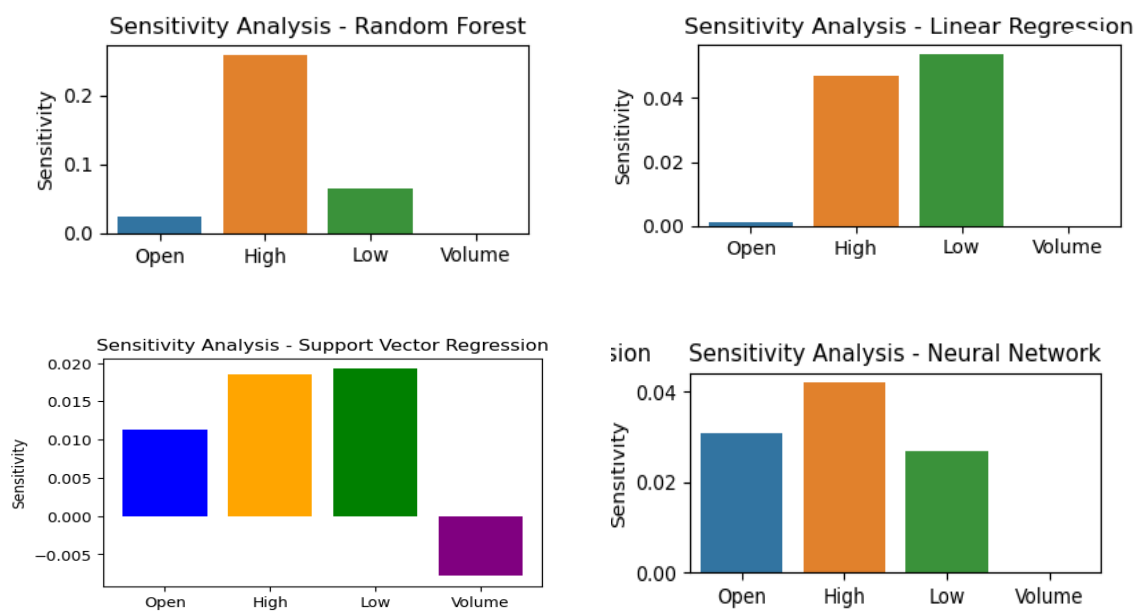
Observations	End	Start	Abbreviation	Name
۱۰۵۹	۱۴۰۳/۵/۳۱	۱۳۹۹/۱/۵	AUT	Automotive Index
۱۰۵۹	۱۴۰۳/۵/۳۱	۱۳۹۹/۱/۵	FIN	Financial Index
۱۰۵۹	۱۴۰۳/۵/۳۱	۱۳۹۹/۱/۵	PHAR	Pharmaceutical Index
۱۰۵۹	۱۴۰۳/۵/۳۱	۱۳۹۹/۱/۵	FOOD	Food Industry Index
۱۰۵۹	۱۴۰۳/۵/۳۱	۱۳۹۹/۱/۵	BME	Basic Metals Index
۱۰۵۹	۱۴۰۳/۵/۳۱	۱۳۹۹/۱/۵	OIL	Petroleum Products Index

یافته‌های پژوهش

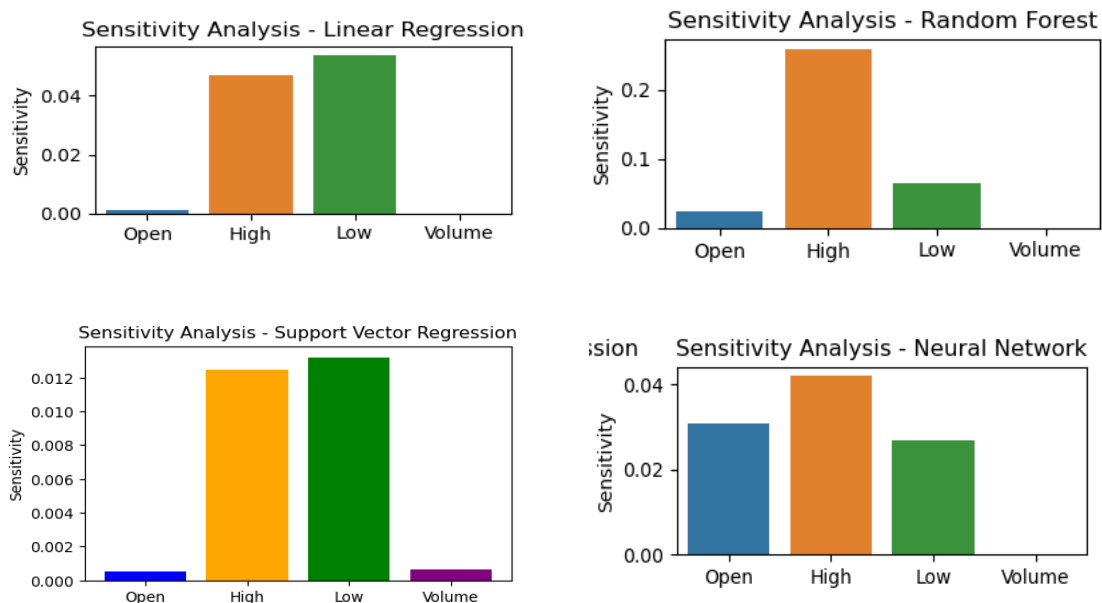
تحلیل حساسیت در مدل‌های یادگیری ماشین ابزاری قدرتمند برای درک تأثیر متغیرهای ورودی بر نتایج پیش‌بینی است. این روش به پژوهشگران و تحلیلگران کمک می‌کند تا مشخص کنند کدام متغیرهای ورودی بیشترین تأثیر را بر عملکرد مدل دارند و چگونه تغییر در این متغیرها می‌تواند نتایج پیش‌بینی را تغییر دهد. در این بخش، به بررسی حساسیت مدل‌های یادگیری ماشین به متغیرهای ورودی شامل Low، High، Open و Volume در پیش‌بینی شاخص‌های مختلف بورس تهران پرداخته شده است. این شاخص‌ها شامل شش صنعت منتخب دارویی، خودرویی، مالی، صنایع غذایی، فلزات اساسی و فراورده‌های نفتی هستند که داده‌های مربوط به آن‌ها در بازه زمانی ۱۳۹۹ تا ۱۴۰۳ به‌صورت روزانه گردآوری شده است. نمودارهای ارائه‌شده در این بخش نشان‌دهنده تفاوت حساسیت مدل‌ها به این متغیرها و بیانگر الگوهای پنهان در تأثیر متغیرهای ورودی بر شاخص‌های مختلف هستند. این تحلیل می‌تواند راهنمایی برای انتخاب مؤثرترین متغیرها در بهبود عملکرد مدل‌ها و بهینه‌سازی پیش‌بینی‌ها باشد.



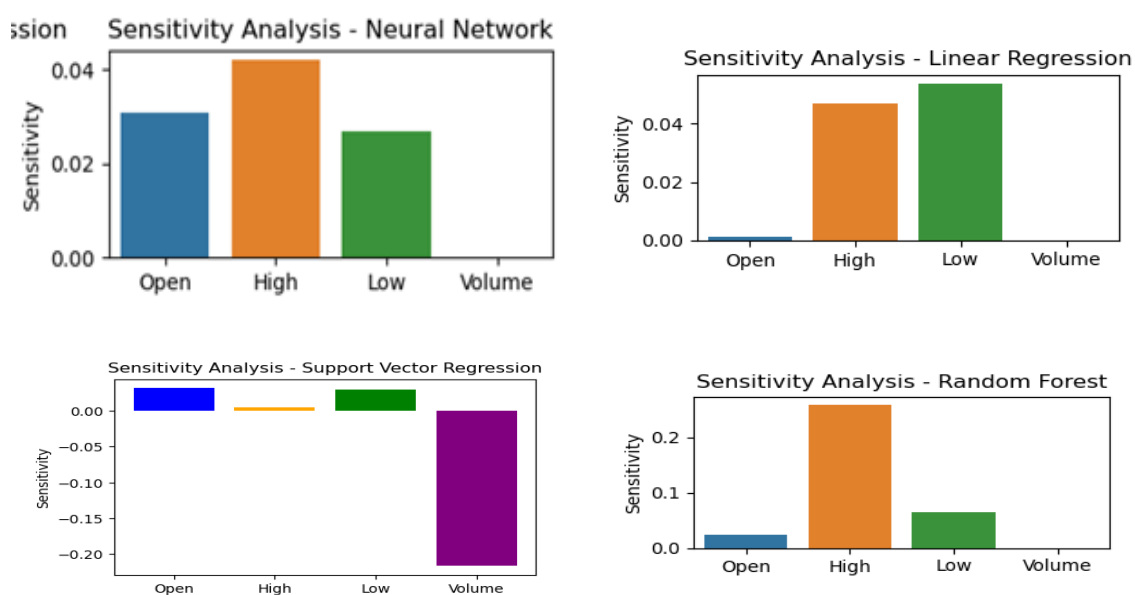
شکل ۲. نمودارهای تحلیل حساسیت مدل‌های یادگیری ماشین در شاخص مالی



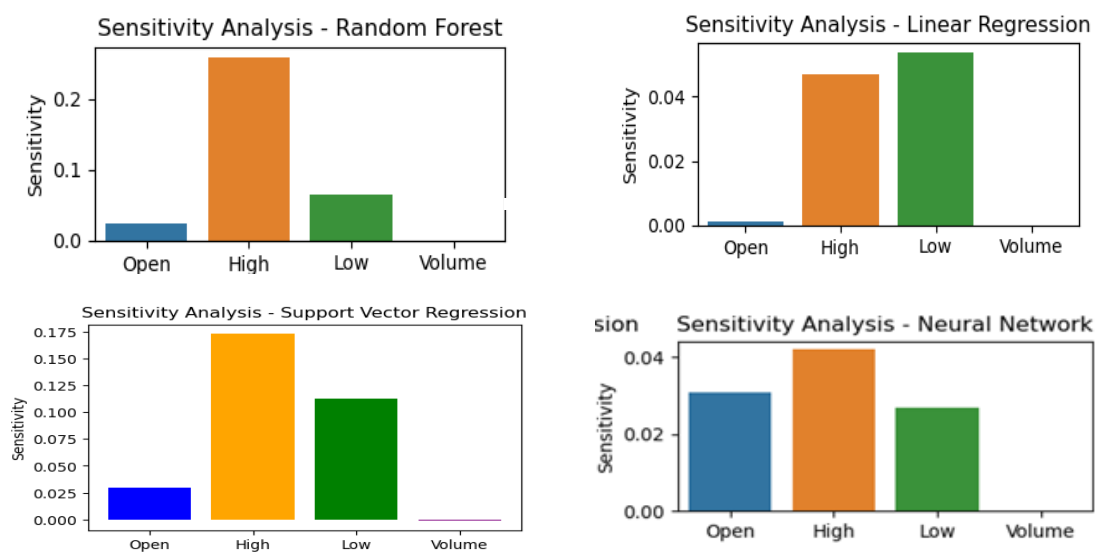
شکل ۳. نمودارهای تحلیل حساسیت مدل‌های یادگیری ماشین در شاخص خودرویی



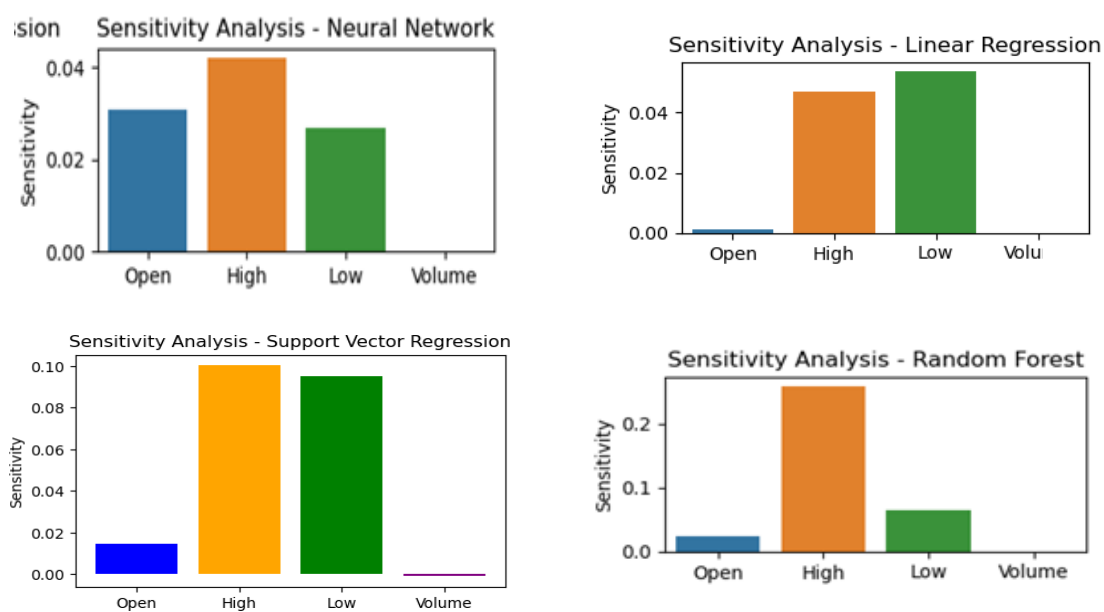
شکل ۴. نمودارهای تحلیل حساسیت مدل‌های یادگیری ماشین در شاخص صنایع غذایی



شکل ۵. نمودارهای تحلیل حساسیت مدل‌های یادگیری ماشین در شاخص فلزات اساسی



شکل ۶. نمودارهای تحلیل حساسیت مدل‌های یادگیری ماشین در شاخص فرآورده‌های نفتی



شکل ۷. نمودارهای تحلیل حساسیت مدل‌های یادگیری ماشین در شاخص دارویی

این نمودارها به تحلیل حساسیت مدل‌ها به متغیرهای ورودی (مانند Open، High، Low و Volume) اشاره دارند. تحلیل حساسیت معمولاً برای فهم این است که کدام متغیرهای ورودی روی خروجی مدل تأثیر بیشتری دارند. اگر

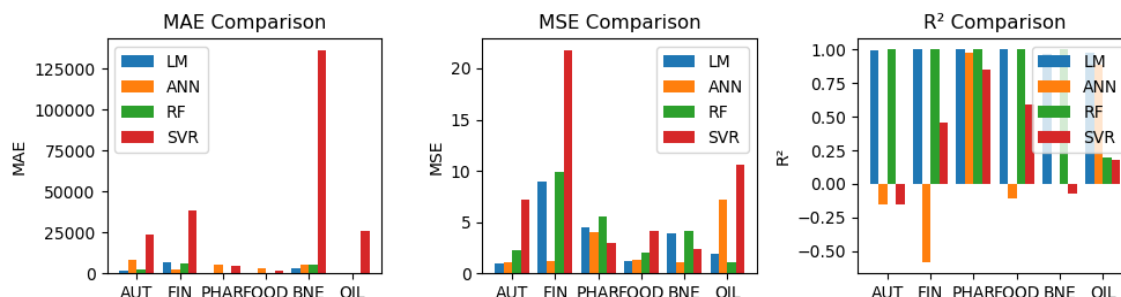
حساسیت یک متغیر به مدل مثبت باشد (میل به بالای محور صفر است)، به این معناست که افزایش مقدار آن متغیر باعث افزایش پیش‌بینی مدل می‌شود. اگر حساسیت منفی باشد (میل به زیر محور صفر است)، به این معناست که افزایش مقدار آن متغیر باعث کاهش پیش‌بینی مدل می‌شود. هرچه ارتفاع میله‌ها (قدر مطلق حساسیت) بزرگ‌تر باشد، آن ویژگی (ورودی) تأثیر بیشتری روی مدل دارد. این نشان‌دهنده اهمیت آن ویژگی در پیش‌بینی مدل است. ویژگی‌هایی که حساسیت بسیار کمی دارند، ممکن است برای مدل کم اهمیت باشند و در برخی موارد قابل حذف باشند (در فرایند کاهش ابعاد یا انتخاب ویژگی).

در جدول ۳ عملکرد هر الگوریتم یادگیری برای شش شاخص منتخب ارائه شده است. در اینجا، سه معیار مختلف ارزیابی (میانگین مطلق خطا (MAE)، میانگین مربعات خطا (MSE) و ضریب تعیین R^2) برای هر یک از الگوریتم‌ها در هر بخش ارائه شده است.

جدول ۳. میانگین عملکرد هر الگوریتم یادگیری در پنجره تاریخی w

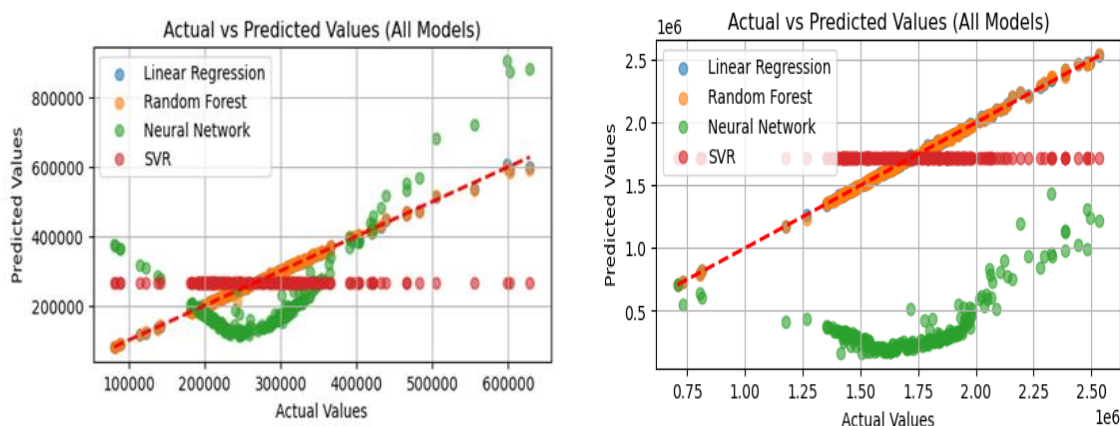
SVR	RF	ANN	LM	Metric	w
۲۳۵۰۹/۷۰۳۸۹۱۶۲	۲۲۲۵/۲۴۱۸۰۴	۸۳۶۷۰۲/۹۳۷۹۱۷	۱۹۲۴/۳۸۵۷۴۵	MAE	AUT
۷/۲۲۲۵۵۱	۲/۲۲۸۰۷۰	۱/۱۳۳۴۳۶	۱/۰۸۷۲۶۸	MSE	
-۰/۱۵۰۰	-۰/۹۹۷۰۱۷	-۰/۱۵۷۳۰۲۳۵	-۰/۹۹۸۵۴۵	R^2	
۳۸۲۵۷/۰۱۳۰۶۱	۶۲۷۱/۳۲۰۳۶۴	۲۴۹۲۶۷/۹۱۰۰۲۲	۶۴۷۳/۲۸۸۹۷۴	MAE	FIN
۲۱/۸۰۲۱۱۸۱۴	۹/۸۹۲۱۵۵	۱/۳۹۴۸۳۸	۹/۰۶۲۷۹۰	MSE	
-۰/۴۵۴۴	-۰/۹۹۸۸۷۷	-۰/۵۸۴۱۷۹	-۰/۹۹۸۹۷۱	R^2	
۴۴۴۹/۳۴۴۸۲۴	۴۸۵/۱۰۹۱۳۹	۵۲۴/۰۸۷۲۱۶۴	۴۶۵/۵۴۰۷۴۸	MAE	PHAR
۲/۹۱۹۰۰	۵/۵۸۴۰۷۸	۴/۰۰۳۴۱۹	۴/۴۴۴۰۰۲	MSE	
-۰/۸۵۳۸۱۰۳۲	-۰/۹۹۹۶۵۴	-۰/۹۷۵۱۸۳	-۰/۹۹۹۷۲۵	R^2	
۱۶۷۹/۱۹۴۴۴۷	۲۴۷/۹۵۶۱۵۵	۳۲۶۵۶/۵۵۱۶۷۶	۲۵۷/۳۲۲۲۵۳	MAE	FOOD
۴/۱۱۸۲۰۰	۱/۳۲۱۶۹۹	۱/۲۴۵۷۳۷	۱/۲۴۹۸۶۹	MSE	
-۰/۵۹۲۸۸۱	-۰/۹۹۸۷۹۵	-۰/۱۰۳۶۱۱۷۱	-۰/۹۹۸۸۶۰	R^2	
۱۳۶۲۹۴/۰۰۳۶۴۸۹۲۷۴	۵۰۵۹/۹۷۶۶۹۸	۵۰۳۸۹۰/۶۲۷۴۱۸	۲۸۵۹۰/۸۵۵۱۰۹	MAE	BME
۲/۳۹۹۴۱	۴/۹۸۱۱۹۲	۱/۱۲۵۷۰۶	۳/۹۲۷۱۵۷	MSE	
-۰/۰۷۳۴۷۱۴	-۰/۹۹۹۵۵۷	-۰/۰۰۳۷۰۸	-۰/۹۶۵۱۰۱	R^2	
۲۶۰۸۵۹۶/۸۶۵۰۶۰	۴/۵۴۵۹۴۰	۳/۵۳۲۳۸۳	۸/۶۹۹۲۰۶	MAE	OIL
۱۰/۶۴۶۷۲۲	۱/۱۳۷۵۷۲	۷/۸۴۶۷۶۹	۱/۹۹۶۳۵۶	MSE	
-۰/۱۷۶۰	-۰/۹۹۸۳۷۷	-۰/۸۸۸۰۸۲	-۰/۹۷۱۵۲۶	R^2	

Performance Comparison of Machine Learning Models

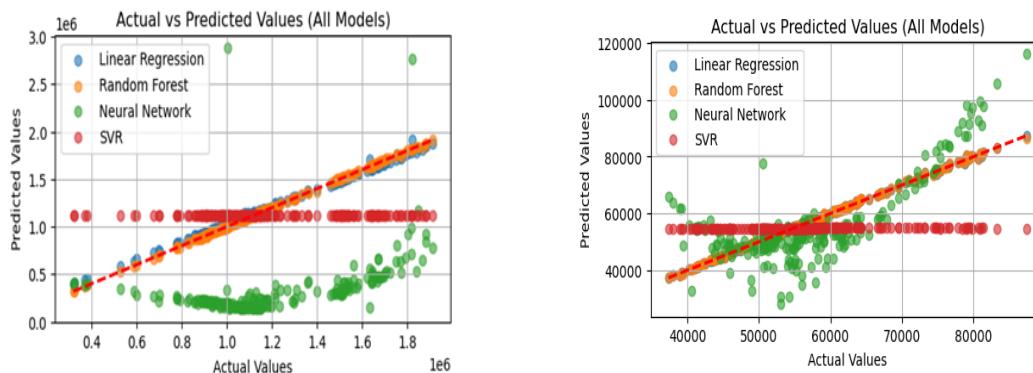


شکل ۸. نمودارهای مقایسه عملکرد مدل‌های یادگیری ماشین

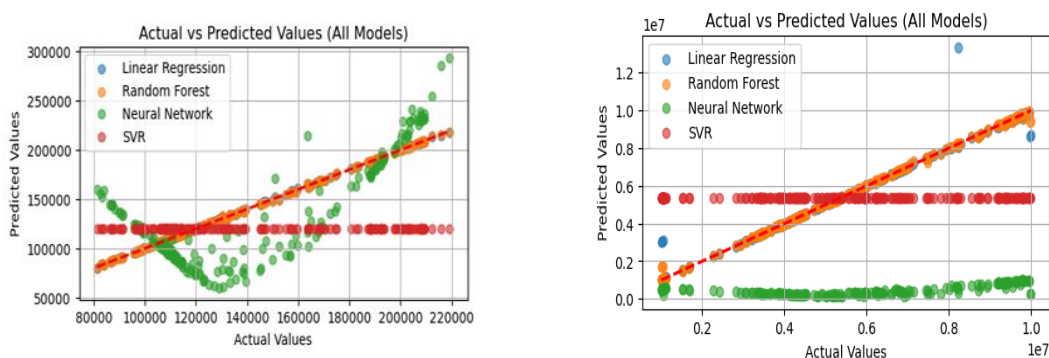
این نمودارها عملکرد مدل‌های مختلف یادگیری ماشین (LM, ANN, RF و SVR) را با استفاده از سه معیار MAE, MSE و R^2 در پیش‌بینی داده‌های مربوط به شاخص‌های مختلف (AUT, FIN, PHAR, FOOD, BNE و OIL) مقایسه می‌کند. نتایج را می‌توان بدین صورت تفسیر کرد: به لحاظ MAE مدل رگرسیون بردار پشتیبان بدترین عملکرد در شاخص فلزات اساسی، مالی، خودرویی و نفتی دارد. به لحاظ MSE همچنان مدل رگرسیون بردار پشتیبان بدترین عملکرد در شاخص مالی، نفت و خودرویی دارد. R^2 نشان‌دهنده میزان تطابق مدل با داده‌های واقعی است (یک بهترین تطابق و صفر بدون تطابق). شبکه عصبی مصنوعی با داشتن R^2 منفی در شاخص مالی، خودرویی، غذایی و فلزات اساسی و رگرسیون بردار پشتیبان در شاخص خودرویی و فلزات اساسی بدترین عملکرد را دارند و مدل‌های LM و RF در همه شاخص‌ها عملکرد بهتری دارند.



شکل ۹. مقایسه مقادیر واقعی و پیش‌بینی شده در همه مدل‌ها نسبت به شاخص مالی و خودرویی



شکل ۱۰. مقایسه مقادیر واقعی و پیش‌بینی شده در همه مدل‌ها نسبت به شاخص غذایی و فلزات اساسی



شکل ۱۱. مقایسه مقادیر واقعی و پیش‌بینی شده در همه مدل‌ها نسبت به شاخص فراورده‌های نفتی و دارویی

در نمودارهای فوق، مقادیر واقعی و پیش‌بینی شده در شش شاخص منتخب براساس چهار تکنیک یادگیری ماشین مطرح شده در پژوهش نشان داده شده است. خط قرمز رنگ، مقادیر واقعی را نشان می‌دهد. نمودارها قدرت بسیار زیاد مدل پیش‌بینی جنگل تصادفی نسبت به سایر مدل‌ها را نشان می‌دهند.

براساس نتایج به‌دست‌آمده پاسخ سؤال‌های پژوهش به شرح زیر است:

۱. کدامیک از متغیرهای ورودی بیشترین تأثیر را دارند؟ براساس نتایج، متغیرهای قیمت باز (Open)، بالا (High) و پایین (Low) بیشترین تأثیر را داشته‌اند.

۲. آیا حساسیت مدل‌های پیچیده‌تر نسبت به مدل‌های ساده‌تر متفاوت است؟ بله، مدل‌های پیچیده‌تر مانند جنگل تصادفی و شبکه عصبی حساسیت بیشتری به متغیرهای ورودی نشان دادند، در حالی که مدل‌های ساده‌تر مانند رگرسیون خطی و SVR حساسیت کمتری داشتند.

۳. چگونه تحلیل حساسیت می‌تواند به بهبود مدل‌ها کمک کند؟ تحلیل حساسیت به شناسایی متغیرهای مؤثر کمک می‌کند و می‌تواند به حذف متغیرهای غیرضروری و کاهش پیچیدگی مدل‌ها منجر شود.
۴. کدام مدل کارایی بیشتری دارد؟ مدل جنگل تصادفی، به دلیل دقت بسیار زیاد و تطابق مناسب با داده‌های واقعی، بهترین عملکرد را در پیش‌بینی شاخص‌های بورس ارائه داد.

نتیجه‌گیری و پیشنهادها

تحلیل حساسیت مدل‌های یادگیری ماشین در پیش‌بینی شاخص بورس تهران نشان داد که تأثیر متغیرهای ورودی در مدل‌های مختلف، به‌طور چشمگیری متفاوت است. براساس یافته‌های این پژوهش، مدل جنگل تصادفی (RF) به دلیل ساختار گروهی و استفاده از تعداد زیادی درخت تصمیم، بیشترین دقت و کارایی را در پیش‌بینی شاخص‌های مختلف بورس تهران ارائه کرد و به متغیرهای ورودی حساسیت زیادی نشان داد. مدل شبکه عصبی مصنوعی (ANN) توانست روابط پیچیده و غیرخطی بین متغیرهای ورودی را به‌خوبی مدل‌سازی کند؛ اما در مقایسه با جنگل تصادفی، عملکرد آن اندکی ضعیف‌تر بود. مدل رگرسیون خطی (LM) به دلیل فرضیه‌های ساده‌تر و خطی بودن، در پیش‌بینی شاخص‌ها دقت کمتری داشت؛ اما همچنان توانست نقش متغیرهای ورودی کلیدی را نمایان سازد. مدل رگرسیون بردار پشتیبان (SVR) در شاخص دارویی، فراورده‌های نفتی و غذایی، به دلیل حساسیت بیشتر، توانستند در پیش‌بینی دقیق‌تر شاخص‌های بورس مؤثرتر عمل کنند. یافته‌ها همچنین نشان داد که متغیرهای قیمت باز (Open)، بالا (High) و پایین (Low) بیشترین تأثیر را در مدل‌ها دارند و می‌توانند به‌عنوان متغیرهای کلیدی در بهینه‌سازی پیش‌بینی‌ها در نظر گرفته شوند.

پیشنهادها

بهینه‌سازی مدل‌ها با استفاده از روش‌های ترکیبی (مانند ترکیب شبکه عصبی و جنگل تصادفی) می‌تواند دقت پیش‌بینی را بهبود بخشد. توجه به متغیرهای کلیدی، مانند قیمت باز و بالا، باید به‌عنوان ورودی‌های اصلی در مدل‌سازی آینده در کانون توجه قرار گیرد. استفاده از داده‌های خارجی مانند افزودن متغیرهای کلان اقتصادی یا داده‌های مرتبط با اخبار و تحلیل احساسات، می‌تواند به بهبود دقت مدل‌ها کمک کند. توسعه روش‌های تحلیل حساسیت با استفاده از روش‌های پیچیده‌تر مانند تحلیل حساسیت جهانی، برای درک بهتر تأثیر متغیرها و تحلیل حساسیت براساس ابرپارامترها پیشنهاد می‌شود. نتایج این پژوهش می‌تواند به دیگر بازارهای مالی یا حوزه‌های پیش‌بینی غیرمالی تعمیم داده شود. این نتایج و پیشنهادها می‌تواند در کانون توجه سیاست‌گذاران و سرمایه‌گذاران در بازار بورس قرار گیرد و به تصمیم‌گیری آگاهانه‌تر در این حوزه کمک کند.

منابع

افشاری راد، الهام؛ علوی، سیدعنایت‌اله و سنایی، حسنعلی (۱۳۹۷). مدلی هوشمند برای پیش‌بینی روند سهام با استفاده از روش‌های تحلیل تکنیکال. *تحقیقات مالی*، ۲۰(۲)، ۲۴۹-۲۶۴.

- ایزدی، مریم؛ آشتاب، علی و زواری رضایی، اکبر (۱۴۰۴). بررسی و تطبیق قدرت تخمین مدل‌های یادگیری ماشین و مدل‌های آماری در پیش‌بینی جهت تغییرات اجزای سود و انتخاب مدل بهینه. *تحقیقات مالی*، ۲۷(۱)، ۳۱-۵۷.
- سعادت، محمد و زمانی بروجنی، فرساد (۱۳۹۵). مروری بر روش‌های بهبود کارایی تکنیک جنگل تصادفی. *اولین همایش ملی فناوری اطلاعات، ارتباطات و محاسبات نرم، اصفهان دانشگاه آزاد اسلامی واحد اصفهان (خوراسگان)*.
- سهرابی، مریم؛ میربرگ کار، سید مظفر؛ چیرانی، ابراهیم و خردیار، سینا (۱۴۰۲). مقایسه مدل‌های مختلف یادگیری ماشین در پیش‌بینی شاخص بازار سهام. *مهندسی مالی و مدیریت اوراق بهادار*، ۵۶(۱۴)، ۱۹-۳۶.
- فرمان آرا، وحید؛ کمیجانی، اکبر؛ فرزین‌وش، اسداله و غفاری، فرهاد (۱۳۹۸). نقش بازار سرمایه در تأمین مالی و رشد اقتصادی (مطالعه موردی ایران و منتخبی از کشورهای در حال توسعه). *اقتصاد مالی*، ۱۳(۴۷)، ۱۹-۳۸.
- فلاح‌پور، سعید؛ گل ارضی، غلامحسین و فتوره‌چیان، ناصر (۱۳۹۲). پیش‌بینی روند حرکتی قیمت سهام با استفاده از ماشین بردار پشتیبان برپایه الگوریتم ژنتیک در بورس اوراق بهادار تهران. *تحقیقات مالی*، ۱۵(۲)، ۲۶۹-۲۸۸.
- قلی‌زاده کپورچالی، لیلا؛ فلاح، میرفیض و مینویی، مهرزاد (۱۴۰۳). طراحی مدل تبیین‌کننده اثر عوامل مؤثر بر پیش‌بینی سود شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از مدل یادگیری ماشین، شبکه عصبی و رگرسیون خطی. *دانش حسابداری و حسابرسی مدیریت*، ۱۳(۵۰)، ۲۴۵-۲۵۷.
- مرادزاده‌فرد، مهدی؛ دارابی، رویا و شاهعلی‌زاده، رامین (۱۳۹۳). یکپارچه‌سازی تکنیک‌های هوش مصنوعی جهت ارائه مدل پیش‌بینی قیمت سهام. *پژوهش‌های حسابداری مالی و حسابرسی (پژوهشنامه حسابداری مالی و حسابرسی)*، ۶(۲۴)، ۸۹-۱۰۱.
- مرادی، بابک؛ بحری ثالث، جمال؛ جبارزاده‌کنگرلویی، سعید و آشتاب، علی (۱۴۰۱). تبیین و ارائه مدلی برای پیش‌بینی نقدشوندگی سهام در بورس اوراق بهادار تهران. *تحقیقات مالی*، ۲۴(۱)، ۱۳۴-۱۵۶.
- مکیان، سید نظام‌الدین و موسوی، فاطمه سادات (۱۳۹۰). پیش‌بینی قیمت سهام شرکت فراورده‌های نفتی پارس با استفاده از شبکه‌های عصبی و روش رگرسیون. *مطالعه موردی: قیمت سهام شرکت فراورده‌های نفتی پارس. فصلنامه علمی - پژوهشی مدل‌سازی اقتصادی*، ۶(۱۸)، ۱۰۵-۱۲۱.

References

- Afsharirad, E., Alavi, S. E. & Sinaei, H. (2018). Developing an Intelligent Model to Predict Stock Trend Using the Technical Analysis. *Financial Research Journal*, 20(2), 249-264. <https://doi.org/10.22059/frj.2018.233763.1006452> (in Persian)
- Ajiga, D. I., Adeleye, R. A., Tubokirifuruar, T. S., Bello, B. G., Ndubuisi, N. L., Asuzu, O. F. & Owolabi, O. R. (2024). Machine learning for stock market forecasting: a review of models and accuracy. *Finance & Accounting Research Journal*, 6(2), 112-124.
- Awad, M. & Khanna, R. (2015). *Efficient learning machines: theories, concepts, and applications for engineers and system designers* (p. 268). Springer nature.

- Ayala, J., García-Torres, M., Noguera, J. L. V., Gómez-Vela, F. & Divina, F. (2021). Technical analysis strategy optimization using a machine learning approach in stock market indices. *Knowledge-Based Systems*, 225, 107-119.
- Chaaajer, P., Shah, M. & Kshirsagar, A. (2021). The applications of artificial neural networks, support vector machines, and long-short term memory for stock market prediction. *Decision Analytics Journal*, 100015.
- Fallahpour, S., Golarzi, G. & Fatourechian, N. (2013). Predicting Stock Price Movement Using Support Vector Machine Based on Genetic Algorithm in Tehran Stock Exchange Market. *Financial Research Journal*, 15(2), 269–288. <https://doi.org/10.22059/jfr.2013.51081> (in Persian)
- Farman Ara, V., Komijani, A., Farzinvas, A. & Ghaffari, F. (2019). The role of capital markets in financing economic growth A Case Study of Iran and some selected developing countries. *Journal of Financial Economics (Financial Economics and Development)*, 13(47), 19-37. SID. <https://sid.ir/paper/229015/en> (in Persian)
- Fischer, T. & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market prediction. *European Journal of Operational Research*, 270(2), 654–669.
- Gholizadeh Kapourchali, L., Fallah, M. F. & Minooei, M. (2024). Presenting an explanatory model of factors which explains the effect of the company's factors on the prediction of accepted profits in the Tehran Stock Exchange using machine learning, neural network and linear regression. *Journal of Management Accounting and Auditing Knowledge*, 13(50), 245-257. (in Persian)
- Hoque, K. E. & Aljamaan, H. (2021). Impact of hyperparameter tuning on machine learning models in stock price forecasting. *IEEE Access*, 9, 163815-163830.
- Hutter, F., Kotthoff, L. & Vanschoren, J. (2019). *Automated machine learning: methods, systems, challenges* (p. 219). Springer Nature.
- Izadi, M., Ashtab, A. & Zavari Rezaei, A. (2025). Comparing the Estimation Power of Machine Learning Models and Statistical Models in Predicting Profit Component Changes and Selecting the Optimal Model. *Financial Research Journal*, 27(1), 31–57. <https://doi.org/10.22059/frj.2024.373472.1007580> (in Persian)
- Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H. & Alfakeeh, A. S. (2022). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, 1-24.
- Khanna, M., Kulshrestha, M., Singh, L. K., Thawkar, S. & Shrivastava, K. (2022). Performance evaluation of machine learning algorithms for stock price and stock index movement prediction using trend deterministic data prediction. *International Journal of Applied Metaheuristic Computing (IJAMC)*, 13(1), 1-30.
- Kubat, M. (2017). *An introduction to machine learning*. Springer, Cham.

- Lee, M. C. (2009). Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert systems with applications*, 36(8), 10896-10904.
- Makian, S.N. & Mousavi, F.S. (2011). Stock Price Forecasting Through Using Ann and Arima Techniques: A Case Study of Pars Petroleum Company. *Economic Modeling Scientific-Research Quarterly*, 6 (18), 105-121. SID. <https://sid.ir/paper/176324/en> (in Persian)
- Malkiel, B. G. & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417.
- Moradi, B., Bahri Sales, J., Jabbarzadeh Kangarloo, S. & Ashtab, A. (2022). Explaining and Proposing a Market Liquidity Prediction Model in Tehran Stock Exchange. *Financial Research Journal*, 24(1), 134-156. <https://doi.org/10.22059/frj.2021.323611.1007184> (in Persian)
- Moradzadehfard, M., Darabi, R. & Shahalizadeh, R. (2015). Combination of Artificial Intelligence Methods for Presenting Stock Price Forecasting Models. *The Financial Accounting and Auditing Research*, 6(24), 89-101. SID. <https://sid.ir/paper/198056/en> (in Persian)
- Patel, J., Shah, S., Thakkar, P. & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert systems with applications*, 42(4), 2162-2172.
- Saadat, M. & Zamani Boroujeni, F. (2016). A review of methods for improving the efficiency of the random forest technique. *Proceedings of the First National Conference on Information Technology, Communications, and Soft Computing*, Isfahan, Iran: Islamic Azad University, Isfahan (Khorasgan) Branch. (in Persian).
- Sangeetha, J. M. & Alfia, K. J. (2024). Financial stock market forecast using evaluated linear regression based machine learning technique. *Measurement: Sensors*, 31, 100-950.
- Schratz, P., Muenchow, J., Iturritxa, E., Richter, J. & Brenning, A. (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecological Modelling*, 406, 109-120.
- Smola, A. J. & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14, 199-222.
- Sohrabi, M., Mirbargkar, S.M., Chirani, E. & Kheradiar, S. (2023). Comparison of different machine learning models in stock market index prediction. *Financial engineering and securities management*. 56 (14), 19-36. (in Persian)
- Tsinaslanidis, P. E. & Kugiumtzis, D. (2014). A prediction scheme using perceptually important points and dynamic time warping. *Expert Systems with Applications*, 41, 6848-6860.
- Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5), 988-999.