فصلنامه علمی پژوهشی دانش سرمایه‌گذاری
دوره ۱۶ / شماره ۳ (پیاپی ۶۳) / پائیز ۱۴۰۶
صفحه ۳۸۳ تا ۴۱۶

# مدل‌های یادگیری ماشین برای پیش‌بینی میان‌مدت مبتنی‌بر داده‌های شاخص بورس برای مدیریت سبد سرمایه‌گذاری

**بهناز بدائی**
گروه حسابداری و مالی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران.
b.badaei@iau.ac.ir
**حسین بدیعی**
گروه حسابداری و مالی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران. (نویسنده مسئول)
badiei@iau.ac.ir
**محسن حمیدیان**
گروه حسابداری و مالی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران.
m_hamidian@azad.ac.ir
**علی امیری**
گروه حسابداری و مالی، واحد بندرعباس، دانشگاه آزاد اسلامی، بندرعباس، ایران.
a_amiri@iau.ac.ir

تاریخ دریافت: ۱۴۰۳/۱۰/۱۰ تاریخ پذیرش: ۱۴۰۳/۱۱/۳۰

## چکیده

این پژوهش به بررسی پیش‌بینی رفتارهای مالی با استفاده از مدل‌های پیشرفته یادگیری ماشین می‌پردازد. هدف اصلی این تحقیق، ارائه الگوی پرتفوی مبتنی بر مدل‌های پیشرفته یادگیری ماشین، از جمله NARX، RF و SVR و بهینه‌سازی چندهدفه است. با توجه به اهمیت پیش‌بینی دقیق رفتارهای مالی در تصمیم‌گیری‌های سرمایه‌گذاری و مدیریت ریسک، این پژوهش سعی دارد به روشن شدن نقاط قوت و ضعف هر یک از مدل‌های یادگیری ماشین مورد بررسی و ارائه الگوی بهینه‌سازی پورتفولیو بر این اساس کمک کند.

در این پژوهش، از داده‌های تاریخی بازار سهام تهران (از سال ۱۳۹۰ تا ۱۴۰۲)، شامل اطلاعات مربوط به قیمت و متغیرهای کلان اقتصادی استفاده شده است. داده‌ها از منابع معتبر و قابل اعتماد گردآوری شده و به‌منظور تحلیل بهتر، پردازش و آماده‌سازی شدند. همچنین، برای ارزیابی دقیق عملکرد مدل‌ها، از روش‌های یادگیری ماشین استفاده شده و پارامترهای مدل‌ها تنظیم گردیدند.

نتایج این تحقیق نشان می‌دهد که مدل‌های NARX، RF و SVR دارای قابلیت‌های متفاوتی در پیش‌بینی قیمت‌ها هستند و عملکرد آن‌ها تحت تأثیر ویژگی‌های داده‌ها قرار دارد. به‌ویژه، مدل NARX، عملکرد باثبات‌تری از خود نشان می‌دهد. به علاوه، پورتفولیوهای مبتنی بر الگوریتم‌های یادگیری ماشین، در همه حالات توانسته‌اند بهتر از پورتفولیو با اوزان برابر عمل کنند. همچنین، این مطالعه به شناسایی چالش‌های مربوط به تنظیم پارامترها و پیچیدگی‌های محاسباتی پرداخته و به اهمیت به‌روزرسانی مدل‌ها با توجه به تغییرات بازار تأکید می‌کند. در نهایت، پیشنهاداتی برای بهبود مدل‌ها و راهکارهایی برای پژوهش‌های آتی ارائه می‌شود تا دقت پیش‌بینی‌ها افزایش یابد و نتایج بهتری در تحلیل رفتارهای مالی به دست آید.

**واژه‌های کلیدی:** مدل NARX، بهینه‌سازی پورتفولیو، یادگیری ماشین.

## ۱- مقدمه

در چشم‌انداز مالی معاصر، که با افزایش نوسانات و پیچیدگی مشخص می‌شود، سرمایه‌گذاران دائماً به دنبال استراتژی‌هایی هستند که نه تنها عملکرد پرتفوی را بهبود می‌بخشد، بلکه ریسک‌ها و هزینه‌های مرتبط را نیز به طور موثر مدیریت می‌کند (چنگ[1] و همکاران، ۲۰۲۴). چالش در ایجاد یک مدل سبد مستحکم نهفته است که می‌تواند عملکرد بهینه را در حین پایبندی به محدودیت‌ها و دستیابی به اهداف متعدد به دست آورد. این چالش در هنگام تلاش برای متعادل کردن این اهداف در چارچوبی که به طور دقیق رفتار دارایی را در طول زمان پیش‌بینی می‌کند، آشکارتر می‌شود (اودیمی[2] و همکاران، ۲۰۲۴).

محور این تحقیق توسعه یک مدل پورتفولیو است که از قابلیت‌های مدل شبکه عصبی خودرگرسیون غیرخطی با ورودی‌های برون زا (NARX[3]) در ارتباط با تکنیک‌های بهینه‌سازی چند هدفه استفاده می‌کند. مدل NARX که به دلیل قدرت خود در مدیریت داده‌های سری زمانی غیرخطی و درنظرگرفتن روابط پیچیده بین متغیرها مشهور است، پایه محکمی برای پیش‌بینی بازده دارایی و روند بازار فراهم می‌کند. با این حال، ادغام این قدرت پیش‌بینی با بهینه‌سازی چند هدفه برای بهینه‌سازی همزمان چندین هدف متضاد، یک حوزه تحقیق پیچیده و ناشناخته باقی مانده است (گوارینو[4] و همکاران، ۲۰۲۴).

هدف این مطالعه بررسی شکاف قابل توجه در ادبیات بهینه‌سازی پورتفولیو موجود با ترکیب دقت پیش‌بینی NARX با مزایای استراتژیک بهینه‌سازی چند هدفه است. هدف اصلی ایجاد مدلی است که نه تنها بازده دارایی‌ها را با دقت بالا پیش‌بینی کند، بلکه سبد را در ابعاد مختلف، از جمله به حداکثر رساندن بازده و به حداقل رساندن ریسک‌ها بهینه کند. این کار نیاز به یک رویکرد پیچیده برای توسعه مدل، ترکیب تکنیک‌های پیش‌بینی پیشرفته و الگوریتم‌های بهینه‌سازی دارد.

## بیان مسئله

یکی از جنبه‌های محوری این تحقیق، بهره‌گیری از یادگیری ماشین برای پیش‌بینی دقیق بازده دارایی‌ها و استفاده از این پیش‌بینی‌ها در بهینه‌سازی پورتفولیو است. در این مطالعه، مسئله بهینه‌سازی پورتفولیو به‌عنوان یک مسئله چند هدفه تعریف شده که نیازمند تعادل میان اهداف متعارضی مانند حداکثرسازی بازده و حداقل‌سازی ریسک است. با این حال، نقطه کلیدی این تحقیق، تمرکز بر پیش‌بینی دقیق بازده دارایی‌ها از طریق الگوریتم‌های پیشرفته یادگیری ماشین است. این پیش‌بینی‌ها نقشی حیاتی در هدایت تصمیم‌گیری‌های بهینه ایفا می‌کنند و چارچوبی فراهم می‌کنند که قادر به شناسایی الگوهای پیچیده و روابط پنهان در داده‌ها باشد.

این تحقیق، کارایی مدل پیشنهادی را با تاکید ویژه بر توانایی آن در پیش‌بینی‌های دقیق ارزیابی می‌کند. دقت پیش‌بینی‌ها به‌طور مستقیم به موفقیت مدل در دستیابی به نتایج بهینه پورتفولیو مرتبط است. علاوه بر این، مدل

[1] Cheng
[2] Odeyemi
[3] Nonlinear Autoregressive with Exogenous Inputs
[4] Guarino

پیشنهادی از لحاظ قابلیت پیش‌بینی در سناریوهای واقعی سرمایه‌گذاری و مقایسه با روش‌های موجود بررسی می‌شود. تمرکز اصلی این مطالعه بر این است که چگونه پیش‌بینی‌های مبتنی بر یادگیری ماشین می‌توانند مبنای تصمیم‌گیری‌های بهینه باشند و کارایی فرآیند بهینه‌سازی پورتفولیو را به‌طور قابل‌توجهی ارتقا دهند.

بازارهای سرمایه پویا و قابل اعتماد در هر کشور از عوامل تسریع رشد اقتصادی و جذب منابع برای سرمایه گذاران است که در نتیجه کشورها با داشتن بازارهای بورس فعال می‌تواند سرمایه را به جریان انداخته و رساندن منابع مالی به بخش‌های مولد را تسریع نمایند. انتخاب رویکرد مناسب برای مدیریت سبد سهام یکی از عوامل موفقیت سرمایه گذاران در بازار سرمایه می‌باشد (مالاندرونیتس[1] و همکاران، ۲۰۲۴). شیوه‌های مدیریت سبد سهام به دو رویکرد مدیریت فعال و غیرفعال تقسیم بندی می‌شود که در مدیریت فعال، هدف مدیر پرتفوی کسب بازدهی بالاتر از بازده بازار (شاخص کل) است. با توجه به اینکه مدیریت فعال به دنبال کسب بازدهی بالاتر از بازار می‌باشد می‌توان گفت مدیرانی که از استراتژی فعال استفاده می‌کنند اعتقاد دارند بازار از کارایی کامل برخوردار نمی‌باشد. در واقع پژوهشگران در با بررسی داده‌های تاریخی، الگوهایی را مشاهده کرده‌اند که با کارایی بازار در تعارض است و نشان دهنده امکان کسب بازدهی بالاتر از شاخص با اتخاذ استراتژی سرمایه‌گذاری فعال می‌باشد. از این رو همواره پژوهش‌ها به منظور ارائه رویکرد یا استراتژی مدیریت فعال پرتفوی جهت کسب بازدهی بالاتر از شاخص مورد توجه فعالان بازارهای مالی خواهد بود. از طرفی در سالهای اخیر رشد و پیشرفت تکنولوژی و فناوری اطلاعات منجر به تحول دنیای مالی و مدیریت سرمایه‌گذاری گردیده و روش و ابزارهایی نوین را به فعالان بازارهای مالی معرفی کرده و امکان دسترسی و تحلیل داده‌ها و استفاده از الگوریتم‌های مالی را فراهم ساخته است (گارسیا[2] و همکاران، ۲۰۲۴).

در نهایت، این تحقیق با ارائه یک مدل جدید که پیش‌بینی مبتنی بر مدل‌های پیشرفته یادگیری ماشین، از جمله NARX، RF و SVR را با تکنیک‌های بهینه‌سازی چندهدفه ادغام می‌کند، به دنبال پیشرفت زمینه بهینه‌سازی پورتفولیو است. هدف ارائه راه حلی جامع است که به چالش‌های چند وجهی مدیریت پرتفولیو می‌پردازد و ابزاری را به سرمایه گذاران ارائه می‌دهد که عملکرد را افزایش می‌دهد و ریسک‌ها را مدیریت می‌کند. با پر کردن شکاف بین مدل‌سازی پیش‌بینی پیشرفته و بهینه‌سازی چند هدفه، این مطالعه با هدف کمک به بینش‌های ارزشمند و راه‌حل‌های عملی در زمینه سرمایه‌گذاری مالی است.

## اهداف تحقیق

### هدف اصلی

- ارائه الگوی پرتفوی مبتنی بر مدل‌های پیشرفته یادگیری ماشین، از جمله NARX، RF و SVR و بهینه‌سازی چند هدفه

[1] Malandreniotis
[2] García

**اهداف فرعی**

۱) ارزیابی مدل‌های یادگیری ماشین در پیش‌بینی شاخص کل

۲) مقایسه بازده استراتژی معاملاتی با استفاده از الگوی پرتفوی مبتنی بر مدل‌های پیشرفته یادگیری ماشین، از جمله NARX، RF و SVR و بهینه‌سازی چندهدفه در بورس تهران با بازدهی پورتفولیو با اوزان برابر

## مبانی نظری و مرور ادبیات

### بهینه‌سازی میانگین-واریانس

بهینه‌سازی میانگین-واریانس (MVO[1])، که توسط مارکوویتز در سال ۱۹۵۲ معرفی شد، یک نظریه اساسی در امور مالی مدرن برای ساخت پرتفوی سرمایه‌گذاری بهینه است. ایده اصلی این است که با انتخاب ترکیبی از دارایی‌هایی که بازده مورد انتظار را برای سطح معینی از ریسک به حداکثر می‌رساند یا ریسک را برای سطح معینی از بازده مورد انتظار به حداقل می‌رساند، تعادل بین ریسک و بازده ایجاد شود (لوین[2] و همکاران، ۲۰۱۷). در اینجا توضیح مفصلی از MVO آورده شده است:

۱. بازده مورد انتظار:

بازده مورد انتظار یک پرتفوی مجموع وزنی از بازده مورد انتظار دارایی‌های منفرد در پرتفوی است و از نظر ریاضی به صورت زیر بیان می‌شود:

$$E(R_p) = \sum_{i=1}^{m} w_i * E(R_i) \quad (۱)$$

جایی که:

- $E(R_p)$ بازده مورد انتظار سبد است.
- $w_i$ وزن دارایی $i$ در پورتفولیو است.
- $E(R_i)$ بازده مورد انتظار دارایی i است.

۲. واریانس پرتفوی:

واریانس (به عنوان یک معیار ریسک) یک پرتفوی نه تنها به واریانس دارایی‌های فردی بستگی دارد، بلکه به کوواریانس بین آنها نیز بستگی دارد. واریانس پورتفولیو توسط عبارت زیر اندازه‌گیری می‌شود:

$$Var(R_p) = \sum_{i=1}^{m} \sum_{j \geq i}^{m} w_i * w_j * Cov(R_i, R_j) \quad (۲)$$

جایی که:

- $Var(R_p)$ واریانس پورتفولیو است.

[1] Mean-Variance Optimization
[2] Lwin

- $w_i$ و $w_j$ به ترتیب وزن دارایی‌های $i$ و $j$ در پورتفولیو هستند.
- $\mathrm{Cov}(\mathrm{R_i}, \mathrm{R_j})$ کوواریانس بین بازده دارایی‌های $i$ و $j$ است.

۳. کوواریانس:
کوواریانس نحوه حرکت دو دارایی با هم را اندازه گیری می‌کند. به صورت زیر محاسبه می‌شود:

$$Cov(R_i, R_j) = \frac{\sum_{t=1}^{N}\left[\left(R_{i,t} - E(R_i)\right) * \left(R_{j,t} - E(R_j)\right)\right]}{(N-1)} \quad (۳)$$

جایی که:
- $R_{i,t}$ و $R_{j,t}$ بازده دارایی‌های $i$ و $j$ در زمان $t$ هستند.
- $E(R_i)$ و $E(R_j)$ بازده مورد انتظار دارایی‌های $i$ و $j$ هستند.
- $N$ تعداد مشاهدات است.

## روش‌های حل مسائل چندهدفه

روش جمع وزنی

روش جمع وزنی یکی از ساده ترین رویکردها برای بهینه‌سازی چند هدفه است. با ترکیب چند هدف در یک تابع هدف، مسئله چند هدفه را می‌توان به یک مسئله تک هدفه تبدیل می‌کند. این کار با تخصیص وزن به هر هدف با توجه به اهمیت یا ترجیح آن و سپس تشکیل یک مجموع وزنی از اهداف انجام می‌شود. از نظر ریاضی، اگر m هدف $\mathrm{f_{i(x)}}$ وجود داشته باشد، تابع هدف ترکیبی $\mathrm{F(x)}$ را می‌توان به صورت زیر بیان کرد:

$$\mathrm{F(x)} = \sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{w_i f_{i(x)}} \quad (۴)$$

که در آن $\mathrm{w_i}$ وزن مرتبط با هدف $i$ است. با تغییر وزن‌ها می‌توان مبادلات مختلف بین اهداف را بررسی کرد. محدودیت اصلی این روش این است که نیاز به دانش یا فرضیات قبلی در مورد وزن دارد که ممکن است همیشه در دسترس نباشد.

روش اپسیلون-محدودیت

روش اپسیلون-محدودیت یک رویکرد رایج در بهینه‌سازی چند هدفه است که با تبدیل یک مسئله چند هدفه به یک مسئله تک هدفه به تعادل اهداف متعارض کمک می‌کند. این روش شامل اضافه کردن توابع هدف به عنوان قیود (با کران اپسیلون مخصوص آن هدف) و نگهداری تنها یک هدف به عنوان تابع هدف مسئله جدید است. با تغییر این مقادیر اپسیلون، تصمیم گیرندگان می‌توانند مبادلات مختلف را بررسی کنند و راه حل‌هایی بیابند که به بهترین وجه بین اهداف متعارض تعادل برقرار کند. روش اپسیلون-محدودیت با تعریف توابع هدف چندگانه که

نیاز به بهینه‌سازی دارند شروع می‌شود. این اهداف ممکن است معیارهای مختلفی مانند هزینه، کیفیت، زمان یا پایداری را نشان دهند. یکی از اهداف به عنوان هدف اصلی انتخاب می‌شود، در حالی که بقیه به عنوان محدودیت در نظر گرفته می‌شوند. هدف اولیه بهینه شده، و اهداف ثانویه در مقادیر اپسیلون خاصی برای یافتن بهترین سازش بین اهداف متعارض محدود می‌شوند. از نظر ریاضی، روش محدودیت اپسیلون را می‌توان به صورت زیر فرموله کرد:

$$Max/Min\ f(x) \qquad (۵)$$
$$g_{i(x)} \leq epsilon_i, for\ i = 1,2,\dots,k$$

در اینجا، $f(x)$ تابع هدف اولیه است که باید بهینه شود، $g_{i(x)}$ توابع هدف ثانویه هستند که باید محدود شوند، و $epsilon_i$ مقادیر اپسیلون از پیش تعریف‌شده هستند که حداکثر انحرافات مجاز را از راه‌حل مورد نظر برای اهداف ثانویه نشان می‌دهند.

## مدل‌های یادگیری ماشین

### مدل NARX

شبکه‌های عصبی مصنوعی (ANN[1]) با موفقیت در تعدادی از کارهای پیش‌بینی و مدل‌سازی سری‌های زمانی، از جمله پیش‌بینی سری‌های زمانی مالی (دابلمونت[2] و همکاران، ۲۰۰۳)، پیش‌بینی جریان رودخانه (عطیه[3] و همکاران، ۱۹۹۹)، مدل‌سازی سری‌های زمانی زیست‌پزشکی (کویل[4] و همکاران، ۲۰۰۵)، پیش‌بینی ترافیک شبکه‌های ارتباطی (کویل و همکاران، ۲۰۰۵) استفاده شده‌اند. به طور خاص، زمانی که سری‌های زمانی نویز دارند و سیستم دینامیکی زیربنایی غیرخطی است، مدل‌های ANN غالباً از تکنیک‌های خطی استاندارد مانند مدل‌های معروف باکس-جنکینز[5] (باکس و همکاران[6]، ۱۹۹۴) بهتر عمل می‌کنند. در چنین مواردی، غیرخطی بودن ذاتی مدل‌های ANN و استواری بالاتر نسبت به نویز، عملکرد پیش‌بینی بهتر آنها را توضیح می‌دهد.

شبکه‌های بازگشتی ساده (SRN[7]) شامل یک کلاس از مدل‌های عصبی بازگشتی پیش‌خور حاوی تعداد کمی حلقه‌های بازخورد محلی و/یا سراسری در معماری‌های خود هستند. اگرچه شبکه‌های شبه MLP[8] پیشخور به راحتی می‌توانند برای پردازش سری‌های زمانی از طریق وقفه‌های متغیر هدف به عنوان ورودی تطبیق داده شوند، و شبکه عصبی با تأخیر زمانی (TDNN[9]) را به وجود آورد، اما می‌توان آن‌ها را نیز به راحتی به SRN تبدیل کرد. شایان ذکر است که وقتی یک مدل TDNN پیش‌خور برای پیش‌بینی بلندمدت اعمال می‌شود، در نهایت به‌عنوان

[1] Artifitial Neural Networks
[2] Dablemont
[3] Atiya
[4] Coyle
[5] Box-Jenkins
[6] Box
[7] Simple Recurrent Network
[8] Multi-Layer Perceptron
[9] Time delay neural network

نوعی معماری SRN رفتار می‌کند، زیرا یک حلقه سراسری برای بازخورد مقدار تخمینی فعلی به بردار ورودی مورد نیاز است.

معماری‌های بازگشتی بالا معمولاً با استفاده از انواع الگوریتم‌های پس‌انتشار آموزش داده می‌شوند (پیرلموتر[1]، ۱۹۹۵). با این حال، یادگیری انجام وظایفی که در آن وابستگی‌های زمانی موجود در سیگنال‌های ورودی/خروجی در فواصل زمانی طولانی است، می‌تواند با استفاده از الگوریتم‌های یادگیری مبتنی بر گرادیان بسیار دشوار باشد (بنگیو[2] و همکاران، ۱۹۹۴). لین[3] و همکاران (۱۹۹۷) گزارش می‌دهند که یادگیری چنین وابستگی‌های زمانی طولانی مدت با تکنیک‌های گرادیان-نزولی در کلاسی از مدل SRN به نام مدل خودرگرسیون غیرخطی با ورودی برون زا (NARX) موثرتر از مدل‌های بازگشتی مبتنی بر MLP است.

علیرغم مزایای ذکر شده شبکه NARX، امکان سنجی اجرای آن به عنوان یک ابزار غیر خطی برای مدل سازی و پیش‌بینی سری‌های زمانی تک متغیره هنوز به طور کامل بررسی نشده است. به عنوان مثال، در لین و همکاران (۱۹۹۷)، مدل NARX در واقع به مدل TDNN تقلیل می‌یابد تا در پیش‌بینی سری‌های زمانی اعمال شود.

زمینه‌های بالقوه کاربرد این نوع شبکه، مشخص‌سازی ترافیک شبکه ارتباطی و پیش‌بینی سری‌های زمانی است (کانتز[4] و همکاران، ۲۰۰۶)، زیرا نشان داده شده است که این نوع داده‌ها به دلیل ماهیت خود متشابه، وابستگی بلندمدت دارند.

مدل خودرگرسیون غیرخطی با ورودی‌های برون‌زا (NARX) یک کلاس مهم از سیستم‌های غیرخطی زمان-گسسته است که می‌تواند به‌صورت ریاضی به صورت زیر نمایش داده شود.

$$y(n+1) = f[y(n), \dots, y(n-d_y+1); u(n-k), u(n-k+1), \dots, u(n-d_u-k+1)] \qquad (۶)$$

که در آن $u(n) \in \mathbb{R}$ و $y(n) \in \mathbb{R}$ به ترتیب ورودی و خروجی مدل را در مقطع زمانی گسسته n نشان می‌دهند، در حالی که $d_u \geq 1$ و $d_y \geq 1$، $d_u \geq d_y$ به ترتیب درجه حافظه ورودی و خروجی حافظه هستند. پارامتر $k$ عبارت تاخیر است که به عنوان زمان مرگ فرآیند شناخته می‌شود.

با فرض $k = 0$ مدل NARX زیر را به دست می‌آوریم:

$$y(n+1) = f[y(n), \dots, y(n-d_y+1); u(n), u(n-1), \dots, u(n-d_u+1)] \qquad (۷)$$

که به صورت برداری به صورت زیر نوشته می‌شود:

$$y(n+1) = f[\boldsymbol{y}(n); \boldsymbol{u}(n)] \qquad (۸)$$

[1] Pearlmutter
[2] Bengio
[3] Lin
[4] Kantz

که در آن بردارهای $\boldsymbol{y}(n); \boldsymbol{u}(n)$ نشان دهنده رگرسورهای خروجی و ورودی هستند. نگاشت غیرخطی $f$ به طور کلی ناشناخته است و می‌توان آن را به عنوان مثال، توسط یک شبکه پرسپترون چند لایه استاندارد (MLP) تقریب زد.

آموزش شبکه NARX را می‌توان در یکی از دو حالت زیر انجام داد:

- **حالت سری موازی (SP[1])** - در این حالت، رگرسیون خروجی تنها با مقادیر واقعی خروجی سیستم تشکیل می‌شود:

$$\hat{y}(n+1) = \hat{f}[\boldsymbol{y_{sp}}(n); \boldsymbol{u}(n)] = \hat{f}[y(n), \dots, y(n-d_y+1); u(n), u(n-1), \dots, u(n-d_u+1)] \quad (9)$$

که در آن نماد هت (^) برای نشان دادن مقادیر (یا توابع) برآورد شده استفاده می‌شود.

- **حالت موازی (P[2])** - در این مورد، خروجی‌های تخمینی بازخورد داده می‌شوند و در رگرسیون خروجی گنجانده می‌شوند :

$$\begin{aligned}\hat{y}(n+1) &= \hat{f}[\boldsymbol{y_{sp}}(n); \boldsymbol{u}(n)] \\ &= \hat{f}[\hat{y}(n), \dots, \hat{y}(n-d_y+1); u(n), u(n-1), \dots, u(n-d_u+1)]\end{aligned} \quad (10)$$

برای پیش‌بینی سری‌های زمانی تک متغیره غیرخطی با شبکه NARX، درجه خروجی حافظه معمولاً $d_y = 0$ درنظر گرفته می‌شود، بنابراین شبکه NARX به معماری TDNN کاهش می‌یابد یعنی:

$$y(n+1) = f[\boldsymbol{u}(n)] = f[u(n), u(n-1), \dots, u(n-d_u+1)] \quad (11)$$

## رگرسیون بردار پشتیبان

رگرسیون بردار پشتیبان (SVR[3]) بر اساس همان اصولی است که ماشین‌های بردار پشتیبان (SVM[4]) برای طبقه‌بندی دارند. در SVM، ما سعی می‌کنیم یک ابر صفحه پیدا کنیم که نقاط داده را با حداکثر حاشیه به کلاس‌های مختلف تفکیک کند. از سوی دیگر، هدف SVR یافتن ابرصفحه ای است که بهترین تناسب با نقاط داده را در یک رگرسیون دارد (آواد[5] و همکاران، ۲۰۱۵).

ایده کلیدی پشت SVR یافتن یک ابر صفحه در فضای ویژگی است که نزدیک به نقاط داده باشد اما همچنان یک حاشیه تلورانس ε را حفظ کند. ابر صفحه با معادله $f(x) = w^T * x + b$ تعریف می‌شود که $w$ بردار وزن، $x$ بردار ویژگی ورودی و $b$ عبارت بایاس است. در SVR، می‌خواهیم انحرافات خروجی پیش‌بینی‌شده $f(x)$ را از

[1] Series-Parallel
[2] Parallel
[3] Support Vector Regression
[4] Support Vector Machine
[5] Awad

خروجی واقعی $y$ به حداقل برسانیم، در حالی که همچنان آنها را در حاشیه تحمل $\varepsilon$ نگه داریم. این با یافتن بردار وزن بهینه $w$ و عبارت بایاس $b$ که تابع هدف را به حداقل می‌رساند به دست می‌آید:

$$\left(\frac{1}{2}\right) * ||w||^2 + C * \sum_{i=1}^{m}(\max(0, |y_i - f(x_i)| - \varepsilon)) \quad (۱۲)$$

که در آن $||w||^2$ نرم بردار وزن $w$ را نشان می‌دهد، $C$ یک ابرپارامتر است که مبادله بین پیچیدگی مدل و خطاها را کنترل می‌کند و انحراف خروجی‌های پیش‌بینی‌شده از خروجی‌های واقعی را جریمه می‌کند. عبارت $\sum_{i=1}^{m}(\max(0, |y_i - f(x_i)| - \varepsilon))$ تضمین می‌کند که خروجی پیش‌بینی شده در حاشیه تلورانس $\varepsilon$ برای همه نقاط داده قرار دارد و نقاط داده خارج از این حاشیه به تابع زیان کمک می‌کنند و جریمه می‌شود. برای حل مسئله بهینه‌سازی، SVR از روش لاگرانژ برای تبدیل مسئله به شکل دوگان[1] آن استفاده می‌کند، که امکان بهینه‌سازی کارآمد را با استفاده از یک الگوریتم برنامه‌ریزی درجه دوم فراهم می‌کند. هنگامی که بردار وزن بهینه $w$ و عبارت بایاس $b$ تعیین شد، می‌توانیم از تابع $f(x)$ برای پیش‌بینی داده‌های ورودی جدید استفاده کنیم.

### جنگل تصادفی

جنگل تصادفی[2] یک الگوریتم یادگیری گروهی است که می‌تواند برای کارهای طبقه بندی و رگرسیون استفاده شود. در زمینه رگرسیون، رگرسیون جنگل تصادفی درخت‌های تصمیم گیری متعددی را در طول آموزش ایجاد می‌کند و پیش‌بینی‌های آنها را برای پیش‌بینی نهایی تجمیع می‌کند. هر درخت در جنگل تصادفی بر روی یک زیرمجموعه نمونه‌برداری شده تصادفی متفاوت از داده‌ها و ویژگی‌های آموزشی آموزش داده می‌شود و عنصر تصادفی را به مدل اضافه می‌کند (سدورسکی[3]، ۲۰۲۱).

هر درخت تصمیم در جنگل تصادفی با استفاده از زیرمجموعه ای از داده‌ها و ویژگی‌های آموزشی ساخته می‌شود. این فرآیند شامل تقسیم بازگشتی داده‌ها به زیر مجموعه‌ها بر اساس ویژگی‌ها برای ایجاد یک ساختار درختی است که متغیر هدف را پیش‌بینی می‌کند. نقاط تقسیم بر اساس معیارهایی مانند به حداقل رساندن واریانس متغیر هدف در هر زیر مجموعه انتخاب می‌شوند.

جنگل تصادفی با نمونه‌برداری از داده‌های آموزشی و ویژگی‌های هر درخت، تصادفی بودن را لحاظ می‌کند. این فرآیند به کاهش بیش برازش[4] و افزایش توانایی تعمیم مدل کمک می‌کند. به طور معمول، هر درخت بر روی یک نمونه بوت استرپ (نمونه گیری با جایگزینی) از داده‌های آموزشی و یک زیر مجموعه تصادفی از ویژگی‌ها برای تقسیم در هر گره آموزش داده می‌شود.

[1] Dual
[2] Random Forest
[3] Sadorsky
[4] Overfitting

فرض کنید یک جنگل تصادفی با N درخت تصمیم داریم که با $f_{i(x)}$ برای $i = 1$ تا N نشان داده شده است. در این حالت برای تجمیع داریم (که در آن $f_{i(x)}$ پیش‌بینی درخت تصمیم i–ام است):

$$y_{hat} = \left(\frac{1}{N}\right) * \sum_{i=1}^{N} f_{i(x)} \quad (۱۳)$$

## مرور ادبیات

### پیشینه خارجی

مدل NARX بخاطر دقتی که در پیش‌بینی قیمت سهام و شاخص بازار سهام دارد در مقایسه با روش‌های معمولی پیش‌بینی سری‌های زمانی محبوب‌تر است. گریگورین[1] (۲۰۱۵) تجزیه و تحلیل مولفه‌های اصلی را بر روی ۳۰ اندیکاتور تحلیل تکنیکال و پنج متغیر قیمتی از ۱۲ مارس ۲۰۱۲ تا ۳۰ دسامبر ۲۰۱۴ انجام داد تا متغیرهای برون‌زا را برای NARX به دست آورد و قیمت پایانی شاخص NASDAQ را برای دو دوره پیش‌بینی کرد. قیمت‌های پیش‌بینی شده، اثربخشی پیش‌بینی NARX را نشان می‌دهد. کومار و گاندمال[2] (۲۰۲۱) با استفاده از ۱۲ اندیکاتور تحلیل تکنیکال که از داده‌های تاریخی از ژانویه ۲۰۰۰ تا مارس ۲۰۱۹ استخراج شده ، قیمت دو سهم را با استفاده از NARX با موفقیت پیش‌بینی کردند. آنها از تحلیل تکنیکال به میزان زیادی استفاده کردند که بسیار وقت گیر است. ژو و چن[3] (۲۰۱۷) از NARX برای پیش‌بینی شاخص اوراق بهادار شانگهای در ۱۰۰ روز استفاده کردند تا ثابت کنند که پیش‌بینی بلند مدت امکان پذیر است. با این حال ، اثربخشی NARX برای پیش‌بینی طولانی مدت در شکل قوی بازار کارا مورد شک و تریدد است زیرا بورس اوراق بهادار شانگهای شکل ضعیف کارایی بازار است. لبده[4] و همکاران (۲۰۱۷) در پژوهش خود NARX را برای پیش‌بینی قیمت سهام در بازار سهام هند معتبر یافتند. با این حال ، سه متغیر برون زا برای NARX همه شاخص‌های اقتصاد خرد هستند. مونتنگرو و مولینا[5] (۲۰۱۹) از NARX با داده‌های تاریخی ۱۲۵۹ روز معاملاتی برای پیش‌بینی شاخص S&P 500 استفاده کردند. نتایج اعتبار پیش‌بینی کوتاه مدت را تأیید می‌کند، اما ۱۰ متغیر برون‌زا برای NARX همه شاخص‌های اقتصاد خرد هستند. جیا[6] و همکاران (۲۰۲۰) با استفاده از NARX به پیش‌بینی بازده شاخص S&P 500 با استفاده از ۱۸ متغیر برون زا برای NARX پرداخته و سپس سهام براساس بازدهی پیش‌بینی شده از NARX انتخاب می‌شود. سپس با استفاده از بهینه‌سازی چندهدفه وزن سهام در پرتفو به‌نحوی تخصیص داده می‌شود که بازدهی حداکثر و ریسک حداقل شود.

مدل بهینه‌سازی پورتفولیو در عمل به دلیل مسائل تخمینی زمانی که برای داده‌های واقعی اعمال می‌شود، تأثیر محدودی دارد. برای پرداختن به این موضوع، بن[7] و همکاران (۲۰۱۸) دو روش منظم‌سازی و اعتبارسنجی

[1] Grigoryan
[2] Gandhmal & Kumar
[3] Xiu & Chen
[4] Labde
[5] Montenegro & Molina
[6] Jia
[7] Ban

متقابل را برای بهینه‌سازی پورتفولیو استفاده کرده‌اند. ابتدا، آن‌ها منظم‌سازی مبتنی بر عملکرد (PBR) را معرفی می‌کنند، که در آن ایده محدود کردن واریانس‌های نمونه ریسک و بازده پرتفوی برآورد شده است، که راه‌حل را به سمت راه‌حل مرتبط با خطای برآورد کمتر هدایت می‌کند. آن‌ها PBR را برای مسائل میانگین-واریانس و میانگین- ارزش در معرض خطر شرطی (CVaR) در نظر می‌گیرند. برای مسئله میانگین-واریانس، PBR یک محدودیت چند جمله‌ای را معرفی می‌کند، که برای آن دو تقریب محدب ایجاد می‌شود: یکی بر اساس تقریب درجه اول و دیگری بر اساس تقریب درجه دوم محدب. تقریب درجه اول PBR یک سوگیری به تخصیص بهینه اضافه می‌کند و تقریب درجه دوم محدب PBR ماتریس کوواریانس نمونه را کوچک می‌کند. برای مسئله میانگین-CVaR، مدل PBR یک مسئله بهینه‌سازی ترکیبی است. آن‌ها نشان می‌دهند که مدل‌های PBR را می‌توان به‌عنوان مسائل بهینه‌سازی استوار با مجموعه‌های عدم قطعیت جدید مطرح کرد و بهینه‌سازی مجانبی را از راه‌حل‌های تقریب میانگین نمونه (SAA) و PBR و مرزهای کارآمد مربوطه ایجاد کرد. آنها در نهایت نشان می‌دهند که PBR بر تمام معیارهای دیگر برای دو مورد از سه مجموعه داده فاما-فرنچ غالب است.

ادغام پیش‌بینی بازده مدل‌های سری زمانی سنتی در تشکیل پورتفولیو می‌تواند عملکرد مدل بهینه‌سازی سبد اصلی را بهبود بخشد. از آنجایی که مدل‌های یادگیری ماشین و یادگیری عمیق برتری قابل توجهی نسبت به مدل‌های سری زمانی نشان داده‌اند، مقاله ما[1] و همکاران (۲۰۲۱) پیش‌بینی بازده در تشکیل پورتفولیو را با دو مدل یادگیری ماشین، یعنی جنگل تصادفی (RF) و رگرسیون برداری پشتیبان (SVR) و سه مدل یادگیری عمیق شبکه عصبی LSTM، پرسپترون چند لایه عمیق (DMLP) و شبکه عصبی کانولوشن ترکیب می‌کند. به طور خاص، این مقاله ابتدا این مدل‌های پیش‌بینی را برای پیش‌انتخاب سهام قبل از تشکیل پورتفولیو اعمال می‌کند. سپس، نتایج پیش‌بینی آن‌ها را در مدل‌های بهینه‌سازی پورتفولیو امگا و میانگین واریانس ترکیب می‌کند. به منظور ارائه برتری این مدل‌ها، از مدل‌های پورتفولیو با پیش‌بینی بازده میانگین متحرک یکپارچه اتورگرسیو به عنوان معیار استفاده می‌شود. ارزیابی بر اساس داده‌های تاریخی ۹ ساله از سال ۲۰۰۷ تا ۲۰۱۵ از سهام جزء شاخص ۱۰۰ اوراق بهادار چین است. نتایج تجربی نشان می‌دهد که مدل‌های MV و امگا با پیش‌بینی بازده RF، یعنی RF+MVF و RF+OF، از مدل‌های دیگر بهتر عمل می‌کنند. علاوه بر این، RF+MVF نسبت به RF+OF برتری دارد. با توجه به اهمیت بالای این دو مدل، این مقاله به بررسی عملکرد آنها پس از کسر کارمزد تراکنش می‌پردازد. آزمایش‌ها نشان می‌دهد که RF+MVF همچنان بهترین عملکرد را در بین مدل‌های MVF دارد و مدل امگا با پیش‌بینی SVR (SVR+OF) بهترین عملکرد را در بین مدل‌های OF دارد. بنابراین، این مقاله به سرمایه گذاران توصیه می‌کند که پورتفو MVF را با پیش‌بینی بازده RF برای سرمایه‌گذاری معاملات روزانه بسازند.

در مقاله چن[2] و همکاران (۲۰۲۱)، یک رویکرد جدید ساخت پرتفوی با استفاده از یک مدل ترکیبی مبتنی بر یادگیری ماشین برای پیش‌بینی سهام و مدل میانگین واریانس (MV) برای انتخاب سبد توسعه داده شده است. به طور خاص، دو مرحله در این مدل درگیر است: پیش‌بینی سهام و انتخاب پرتفوی. در مرحله اول، یک مدل

[1] Ma
[2] Chen

ترکیبی از تقویت گرادیان پیشرفته (XGBoost) با یک الگوریتم کرم شب تاب بهبود یافته (IFA[1]) برای پیش‌بینی قیمت سهام برای دوره بعدی پیشنهاد شده است. IFA برای بهینه‌سازی‌هایپرپارامترهای XGBoost توسعه یافته است. در مرحله دوم، سهام با پتانسیل بازده بالاتر انتخاب شده و از مدل MV برای انتخاب پرتفوی استفاده می‌شود. با استفاده از بورس شانگهای به عنوان نمونه مطالعه، نتایج به‌دست‌آمده نشان می‌دهد که روش پیشنهادی بر روش‌های سنتی (بدون پیش‌بینی سهام) از نظر بازده و ریسک برتری دارد.

پیش‌بینی بازده سهام یک چشم انداز مهم در زمینه سری‌های زمانی مالی است. مطالعه پایوا و همکاران (۲۰۱۹) یک مدل تصمیم گیری منحصر به فرد را برای سرمایه‌گذاری‌های معاملاتی روزانه در بازار سهام پیشنهاد می‌کند. در این راستا، مدل با استفاده از رویکرد تلفیقی طبقه‌بندی‌کننده مبتنی بر یادگیری ماشین، با روش ماشین بردار پشتیبان (SVM) و میانگین-واریانس (MV) برای انتخاب پورتفولیو توسعه داده شد. ارزیابی تجربی مدل بر اساس دارایی‌های شاخص بورس اوراق بهادار سائوپائولو بود. پنجره‌های ماهانه برای انتخاب مجموعه‌های پارامتر با بهترین عملکرد (مرحله درون نمونه) و آزمایش (مرحله خارج از نمونه) استفاده شد. پنجره‌های ماهانه از پنجره‌های روزانه، با آموزش جدید الگوریتم طبقه بندی و بهینه‌سازی پورتفولیو تشکیل شده است. در مجموع ۸۱ دسته پارامتر فرموله شد. برای مقایسه عملکرد مدل پیشنهادی، دو مدل دیگر پیشنهاد شد: (۱) SVM+1/N که روند طبقه‌بندی دارایی‌هایی را که به هدف معینی از سود رسیده‌اند حفظ می‌کند و به طور مساوی در همه دارایی‌هایی که مثبت بودند سرمایه‌گذاری می‌کندو (ii) تصادفی+MV، که همچنین انتخاب آن دارایی‌ها را با تمایل به دستیابی به هدف معینی از سود حفظ می‌کند، اما انتخاب به‌طور تصادفی تعریف شده است. سپس ترکیب پورتفولیو با استفاده از روش MV تعیین می‌شود. آزمایش‌ها با استفاده از داده‌های تاریخی برای ۳۷۱۶ روز معاملاتی برای تجزیه و تحلیل خارج از نمونه فرموله شدند. شبیه‌سازی‌ها بدون احتساب هزینه‌های مبادله و همچنین با درج بخشی از این هزینه‌ها انجام شد. این مطالعه همچنین عملکرد طبقه‌بندی‌کننده، اندازه پورتفولیوها و بازده و ریسک مدل‌ها را ارزیابی کرد. مدل اصلی پیشنهادی نتایج قابل‌توجهی نشان داد، اگرچه تقاضا برای ارزش معاملاتی می‌تواند یک عامل محدودکننده برای اجرای آن باشد. با این وجود، این مطالعه کاربرد نظری یادگیری ماشین را گسترش داده و یک رویکرد بالقوه عملی برای پیش‌بینی قیمت سهام ارائه می‌دهد.

در مطالعه وو[2] و همکاران (۲۰۲۲) ساده‌سازی یک پورتفولیو با استفاده از روش‌ها و عناصر یادگیری ماشین، و همچنین یک استراتژی برای ایجاد پورتفولیو که بر کاهش ریسک یک پورتفولیو به تعهدات عامل ریسک متکی است، مورد بررسی قرار می‌گیرد. نتایج نشان میدهد روش‌های یادگیری ماشین ساختارهای کوواریانس و وزن پورتفولیو را ایجاد می‌کنند که ارزیابی آنها دشوار است. پرتفوی با کمترین تغییر در به حداقل رساندن ریسک از معیارهای ساده تر بهتر عمل می‌کند. در طول دوره‌های بی‌ثباتی بالا، بازدهی تعدیل‌شده بر اساس ریسک وجود دارد و این اثرات برای سرمایه‌گذارانی که حساسیت بیشتری نسبت به تغییرات شانسی در بازده دارند، تقویت می‌شود.

[1] Improved Firefly Algorithm
[2] Wu

مقاله سن و دوتا[1] (۲۰۲۳) یک رویکرد سیستماتیک برای طراحی سبد با استفاده از دو رویکرد بهینه‌سازی ریسک، الگوریتم برابری ریسک سلسله مراتبی و الگوریتم سهم ریسک مساوی سلسله مراتبی در هفت بخش مهم و سهام NIFTY 50 فهرست‌شده در بازار سهام هند ارائه می‌کند. پرتفوی‌ها بر اساس قیمت‌های تاریخی سهام از ۱ ژانویه ۲۰۱۶ تا ۳۱ دسامبر ۲۰۲۰ ساخته شده اند. عملکرد پرتفوی بر اساس داده‌های آزمایشی از ۱ ژانویه ۲۰۲۱ تا ۱ نوامبر ۲۰۲۱ ارزیابی می‌شود. نتایج بک تست پرتفوی‌ها نشان می‌دهد که عملکرد پورتفولیو الگوریتم برابری ریسک سلسله مراتبی نسبت به پورتفولیو الگوریتم سهم ریسک مساوی سلسله مراتبی در آموزش و داده‌های آزمون برای اکثر بخش‌های مورد مطالعه در این کار برتر است.

در مقاله بهارا[2] و همکاران (۲۰۲۳) به عنوان یک تکنیک منحصربه‌فرد ساخت پورتفولیو ارائه شده است. مدل‌های رگرسیون یادگیری ماشین مانند جنگل تصادفی، تقویت گرادیان پیشرفته (XGBoost)، تقویت تطبیقی (AdaBoost)، رگرسیون ماشین بردار پشتیبان (SVR)، k نزدیک‌ترین همسایه‌های (KNN) و شبکه عصبی مصنوعی (ANN) برای پیش‌بینی سهام استفاده می‌شوند. ارزش‌های دوره بعدی سهام با بازدهی بیشتر در مرحله اول انتخاب می‌شوند. علاوه بر این، مدل بهینه‌سازی پرتفوی میانگین-VaR برای انتخاب پورتفولیو در مرحله دوم استفاده می‌شود. مجموعه داده‌های ماهانه بورس اوراق بهادار بمبئی هند، بورس اوراق بهادار توکیو ژاپن و بورس شانگهای چین، به عنوان نمونه تحقیق استفاده می‌شود و یافته‌ها نشان می‌دهد که مدل میانگین-VaR با پیش‌بینی AdaBoost از سایر مدل‌ها بهتر عمل می‌کند.

هباب و کامپوریدیس[3] (۲۰۲۴) ارزیابی کاملی از ۵ الگوریتم یادگیری ماشین (ML) از جمله رگرسیون خطی حداقل مربعات معمولی (LR)، رگرسیون بردار پشتیبانی (SVR)، k-نزدیکترین همسایگان (KNN)، تقویت گرادیان پیشرفته (XGBoost). و شبکه‌های عصبی حافظه کوتاه‌مدت بلند (LSTM)، و همچنین سایر معیارهای مالی مانند هموارسازی نمایی هولت (HES)، مثلثاتی، تبدیل باکس-کاکس، خطاهای ARMA، روند، و مؤلفه‌های فصلی و میانگین متحرک یکپارچه رگرسیون خودکار (ARIMA) انجام داده اند. آنها این الگوریتم‌ها را برای پیش‌بینی قیمت‌های آینده برای ۳۰ REIT از ایالات متحده، بریتانیا و استرالیا و همچنین ۳۰ سهام و ۳۰ اوراق قرضه به کار برده اند. سپس دارایی‌ها به عنوان بخشی از پورتفولیو مورد استفاده قرار گرفتند که با استفاده از الگوریتم ژنتیک آن را بهینه کردند. نتایج نشان داد که استفاده از الگوریتم‌های ML برای پیش‌بینی قیمت حداقل سه برابر بازدهی نسبت به مدل‌های معیار ارائه می‌کند و ریسک را تقریباً دو برابر کاهش می‌دهد.

## پیشینه داخلی

تهرانی و همکاران (۱۳۹۴) رویکردی جدید برای مدیریت فعال پرتفوی ارائه دادند که در ان ابتدا با استفاده از الگوریتم ژنتیک، با ادغام روش‌های پیش‌بینی شبکه عصبی و نزدیک‌ترین همسایگی اقدام به پیش‌بینی قیمت سهام می‌نمایند. با توجه به قیمت پیش‌بینی شده برای هر سهم، با توجه به قیمت پیش‌بینی شده برای هرسهم مطابق با روش‌های پیش‌بینی شبکه عصبی مبتنی بر الگوریتم ژنتیک و روش مجاورت در نزدیکترین همسایه بر

[1] Sen & Dutta
[2] Behera
[3] Habbab & Kampouridis

پایه الگوریتم ژنتیک و استراتژی معاملاتی تعریف شده، معاملات هوشمند سهام صورت می‌پذیرد. استراتژی معاملاتی تعریف شده بر اساس قیمت‌های پیش‌بینی شده دارای‌های خروجی (۱و۰و۱ -) بوده که به معنای خرید سهم، فروش سهم و یا انجام هیچ گونه عملی تفسیر می‌گردند. بازده حاصل از پرتفوهای روش‌های شبکه عصبی مبتنی بر الگوریتم ژنتیک و روش مجاورت در نزدیکترین همسایه بر پایه الگوریتم ژنتیک و پرتفوی خرید و نگهداری به عنوان نماینده رویکرد مدیریت غیر فعال پرتفوی در هر یک از چهار سال محاسبه گردید. پرتفوی حاصل از روش شبکه عصبی مبتنی بر الگوریتم ژنتیک دارای بالاترین بازده در دوره چهار ساله را دارا بوده که این خود نشان از برتری رویکرد مدیریت فعال پرتفوی نسبت به رویکرد مدیریت غیر فعال پرتفوی می‌باشد.

در ایران نیز پیش‌بینی قیمت سهام و شاخص کل همواره مورد توجه محققان بوده و تحقیقات بسیاری در این زمینه صورت گرفته است. میرعلوی و همکاران (۱۳۹۸) با استفاده از روش‌های فراابتکاری و شبکه‌های عصبی مصنوعی مدلی جهت پیش‌بینی قیمت سهام ارائه دادند و نتیجه گرفتند که شبکه عصبی پرسپترون نسبت به مدل‌های دیگر در پیش‌بینی شاخص سهام خطای کمتری دارد. شبان و همکاران (۱۳۹۹) به بررسی سرایت‌پذیری شاخص کل بورس اوراق بهادار تهران از قیمت دارایی‌های موازی با بکارگیری شبکه عصبی پویا پرداخته‌اند که در آن قیمت سکه تمام بهار آزادی، قیمت هر متر مربع ساختمان مسکونی، قیمت هر بشکه نفت خام ایران و نرخ دلار آمریکا و نوسانات شرطی آنها به عنوان متغیرهای ورودی در نظر گرفته شده است واز مدل خودرگرسیون با متغیرهای برونزا NARX استفاده شده است. نتایج تحقیق نشان داده که با استفاده از متغیرهای مذکور و الگوهای طراحی شده با حداکثر دو وقفه زمانی شاخص کل بورس از قیمت دارایی‌ها سرایت پذیری دارد که نشان‌دهنده کارایی ضعیف بازار اوراق بهادار تهران می‌باشد. این پژوهش یکی از اولین موارد بررسی در بازار سرمایه ایران است. همچنین شبکه عصبی طراحی شده قادر به پیش‌بینی شاخص کل بورس اوراق بهادار تهران و نوسانات آتی شاخص می‌باشد.

هدف مقاله کرامتی و همکاران (۱۴۰۱) کاربرد معاملات الگوریتمی با تمرکز بر رویکرد یادگیری تقویتی برای بهینه‌سازی پرتفوی سهام‌های منتخب است. این پژوهش از حیث هدف، کاربردی و از نظر نوع داده، کمّی و از لحاظ روش، توصیفی - اکتشافی و از منظر طرح تحقیق، پس‌رویدادی است. جامعه آماری این پژوهش، ۶۷۲ شرکت بورسی است که از این تعداد، داده‌های پنج شرکت (نمونه آماری) طی دوره زمانی ۱۳۹۶-۱۴۰۰ بررسی شده است. یافته‌های تحقیق در دوره‌های صعودی و نزولی بازار نشان داد که رویکرد یادگیری تقویتی در بازارهای صعودی و نزولی به صورت معناداری بر رویکرد خرید و نگهداری برتری دارد و عملکرد بهتری ارائه داده است و نتایج با عملکرد الگوریتم‌ها در بازارهای بورس سازگار است. نتایج آشکار کرد که از دیدگاه سودآوری، رویکرد یادگیری تقویتی نسبت به رهیافت خرید و نگه‌داری، عملکرد بهتر و موثرتری داشته است؛ بنابراین، به‌کارگیری روش یادگیری تقویتی پیشنهاد می‌شود.

هدف پژوهش اقتصاد و محمدی (۱۴۰۲) کمینه‌سازی خطای پیش‌بینی به عنوان تابع هدف است. این نوآوری به سرمایه‌گذاران توصیه می‌کند که در تشکیل سبد سرمایه‌گذاری علاوه بر سود و ریسک، بر معیار مهم قابل پیش‌بینی بودن گزینه‌های سرمایه‌گذاری نیز تاکید گردد. ادغام پیش‌بینی بازده مدل‌های سری زمانی سنتی در

تشکیل پورتفولیو می‌تواند عملکرد مدل بهینه‌سازی سبد اصلی را بهبود بخشد. از آنجایی که مدل‌های یادگیری ماشین و یادگیری عمیق برتری قابل توجهی نسبت به مدل‌های سری زمانی نشان داده‌اند، این مقاله پیش‌بینی بازده در تشکیل پورتفولیو را با مدل یادگیری ماشین، یعنی جنگل تصادفی و مدل یادگیری عمیق حافظه‌ی کوتاه‌مدت طولانی ترکیب می‌کند. به منظور ارزیابی عملکرد مدل پیشنهادی، داده‌های تاریخی ۵ ساله از سال ۱۳۹۶ تا ۱۴۰۱ از شاخص ۵ صنعت بانکی، خودرویی، دارویی، فلزی و نفتی است. نتایج تجربی نشان می‌دهد که مدل‌های بهینه‌سازی میانگین واریانس با پیش‌بینی بازدهی به‌وسیله جنگل تصادفی، بهتر عمل می‌کنند.

هدف مقاله علیزاده و کیانفر (۱۴۰۲) توسعهٔ مدل مارکوویتز به‌گونه‌ای است که با شرایط دنیای واقعی تطابق بیشتری داشته باشد و اضافه‌کردن انواع عوامل تحلیل بنیادی و محدودیت‌های بازار سرمایه در این مدل است. در این پژوهش که روی بورس اوراق بهادار تهران انجام شده، از دو معیار میانگین نیمه‌واریانس و میانگین قدر مطلق انحرافات در کنار معیار واریانس در مدل مارکوویتز برای تخمین بهتر میزان ریسک استفاده شده است. به‌علاوه، از چندین محدودیت مانند محدودیت کاردینالیتی، آستانه و بخش‌بندی برای نزدیک‌شدن نتایج مدل مبتنی بر مارکوویتز به واقعیت استفاده می‌شود. برای اینکه معیار بازدهٔ سهام فقط براساس تغییرات قیمتی سهام نباشد، از ۹ معیار مهم تحلیل بنیادی در فیلترسازی سهام شرکت‌ها و معیار بازدهٔ مدل مارکوویتز استفاده شده است. به دلیل پیچیدگی محاسباتی زیاد مدل برنامه‌ریزی ریاضی، نمونهٔ مسائل با الگوریتم فراابتکاری جستجوی‌هارمونی نیز حل شد. نتایج نشان‌دهندهٔ آن بود که مدل ریاضی در معیار فاصله از نقطهٔ ایدئال کارایی بهتری دارد. در صورتی که الگوریتم جستجوی‌هارمونی در معیارهای یکنواختی و گسترش جواب‌های پارتو و زمان حل برتری دارد. با افزایش بازهٔ تغییرات محدودیت‌های کاردینالیتی و آستانه، تعداد سهام بیشتری در سبد انتخاب می‌شود و توابع هدف ریسک و بازده به‌صورت هم‌زمان بهبود خواهند یافت.

هدف از مطالعه رستمی و سهیلی (۱۴۰۲) پیش‌بینی شاخص بورس با استفاده از مدل پویای میانگین‌گیری و نیز روش مدل پویای انتخابی و استفاده از داده‌های فصلی سال‌های ۱۳۸۰-۱۳۹۹ و به‌کارگیری نرم‌افزار متلب می‌باشد. مزیت اصلی مدل مورداستفاده در مطالعه حاضر ورود تعداد زیادی متغیر مستقل به جهت پویایی آن است بدون اینکه مشکل معمول برازش بیش‌ازحد در مدل ظاهر شود. در این مقاله اثر برخی متغیرهای کلان اقتصادی بر فرآیندِ مدل‌سازی و تخمین بازده سهام بورس اوراق بهادار بررسی شد. نتایج مقاله نشان داد که احتمال ورود متغیرهای رشد حجم پول، رشد شبه پول، تورم، رشد شاخص قیمت زمین در شهرهای بزرگ بیشتر از سایر متغیرهای ورودی است.

## روش پژوهش

از لحاظ روش‌شناختی، تحقیق حاضر را می‌توان تحت دسته‌بندی‌های مختلفی بررسی نمود. این تحقیق از نوع تحقیقات توسعه‌ای و کاربردی است. هدف اصلی این نوع تحقیقات، توسعه مدل‌ها و روش‌های جدیدی است که در عمل قابل استفاده باشند.

منابع داده برای این تحقیق شامل پایگاه‌های اطلاعاتی مالی مانند سایت بورس اوراق بهادار تهران، بانک مرکزی و مرکز آمار ایران است که داده‌های تاریخی جامعی را در مورد قیمت سهام، شاخص‌های بازار، شاخص‌های اقتصادی و سایر معیارهای مالی مرتبط ارائه می‌دهند. معیارهای انتخاب شامل انتخاب داده هایی است که مربوط به ابزارهای مالی و بازارهای مورد مطالعه است. معیارهای انتخاب سهام شامل موارد زیر است :

- داده‌های سهم یک دوره تاریخی (۱۳۹۰ تا ۱۴۰۱) را پوشش می‌دهد تا چرخه‌ها و رویدادهای مختلف بازار را به تصویر بکشد .
- سهم انتخابی برای حداقل ۸۵٪ روزهای معاملاتی مورد بررسی اطلاعات قیمتی در دسترس داشته باشد.

پس از اینکه پیش‌بینی بازده دوره بعد شاخص سهام از طریق مدل‌های یادگیری ماشین انجام شد، انتخاب بین سرمایه گذاری در بازار سهام یا سرمایه گذاری در اوراق بهادار با درآمد ثابت بر اساس قائده زیر انجام می‌شود:

- چنانچه بازدهی پیش بینی شده برای شاخص بازار بیشتر یا مساوی بازدهی مورد انتظار برای اوراق بهادار با درآمد ثابت باشد، رویکرد سرمایه گذاری در بازار سهام پیشنهاد می‌گردد.
- چنانچه بازدهی پیش بینی شده برای شاخص بازار کمتر از بازدهی مورد انتظار برای اوراق بهادار با درآمد ثابت باشد، رویکرد سرمایه گذاری در اوراق بهادار با درآمد ثابت انتخاب می‌شود.

در این بخش‌مدل بهینه سازی چند هدفه به منظور تخصیص دارایی بین سهام مختلف بررسی می‌شود. چنانچه از بررسی صورت گرفته در بخش قبل تصمیم بر سرمایه گذاری در بازار سهام حاصل شود، موضوع انتخاب بهترین سهام برای تخصیص دارایی موضوعیت پیدا خواهد کرد و رویکرد مطرح شده در این بخش برای تخصیص دارایی استفاده می‌شود.

مدل بهینه سازی چند هدفه به صورت زیر تعریف می‌شود:

$$Max\left(R_P = \sum_{i=1}^{n} x_i r_i\right)$$

$$Min\left(MpSn = \sum_{i=1}^{n} x_i p_i\right)$$

$subject$ to:

$$\sum_{i=1}^{i} x_i = 1$$

$$0 \le x_i \le \frac{1}{\text{mininnumber}}$$

در این فرمول $n$ نمایانگر تعداد سهام مورد بررسی، $R_p$ نشانگر بازدهی پرتفوی، $MsPn$ توزیع احتمال شرطی معکوس پرتفوی $p$، $x_i$ میزان سرمایه گذاری شده در سهم شماره $i$، $r_i$ بازدهی سهم $i$، $p_i$ توزیع احتمال معکوس سهم $i$ و $mininnumber$ حداقل تعداد سهامی است که می‌خواهیم در آن سرمایه گذاری کنیم.

توزیع احتمال شرطی معکوس یک صندوق یا یک پرتفوی به عنوان عامل ریسک مشخص شده و مورد استفاده قرار می‌گیرد. این متغیر میزان احتمال آنکه یک صندوق سرمایه گذاری در زمانی که بازدهی شاخص بازار بیشتر

از اوراق بهادار با درآمد ثابت بوده است و بازدهی صندوق کمتر از میزان شاخص کل بوده است را مورد ارزیابی قرار می‌دهد.

در این فرمول، $P_i$ به صورت زیر محاسبه می‌شود:

$$p_i = \frac{p(MpSn)i}{p(Mp)}$$

در این فرمول صورت کسر توزیع احتمال مشترک آن است که شاخص، بازدهی بیشتر از اوراق بهادار با درآمد ثابت داشته باشد و سهم $i$ موفق به کسب بازدهی حداقل به اندازه شاخص نشده باشد و مخرج کسر احتمال کسب بازدهی بیشتر از اوراق بهادار با درآمد ثابت توسط شاخص است. توجه داشته باشید، از آنجایی که حل مدل فوق باید به مجموعه مشخصی از متغیرهای تصمیم برای تخصیص دارایی رسیده و بر اساس آن پورتفولیو ساخته شود، نمی‌توان از رویکردهای مبتنی بر پارتو برای حل آن استفاده شود. به همین دلیل، مدل دو هدفه فوق با استفاده از رویکرد مجموع وزنی ساده یا اپسیلون محدودیت به یک مدل تک‌هدفه تبدیل شده و حل خواهد شد. مزیت این روش آن است که یک مجموعه وزن مشخص به عنوان خروجی مدل حاصل می‌شود که می‌توان بر اساس آن به تخصیص دارایی پرداخت.

به طور خلاصه، فرایند گام به گام اجرای تحقیق به صورت زیر است:

- جمع‌آوری داده: در این گام داده‌های قیمت سهام بورسی برای بازه ۱۳۹۰ تا ۱۴۰۱ به همراه داده‌های شاخص کل بورس اوراق بهادار (TEDPIX) و داده‌های شاخص‌های کلان اقتصادی تهیه می‌شود. پس از جمع‌آوری داده‌های مذکور، گام‌های پاک‌سازی و پیش پردازش، شامل حذف شرکت‌های خارج از معیارهای انتخاب، اصلاح داده‌های دورافتاده و ازدست‌رفته طی می‌شود.
- محاسبه بازده و مهندسی ویژگی
- نرمال‌سازی و بخش‌بندی داده‌ها برای ورود به مدل‌های یادگیری ماشین.
- آموزش و اعتبارسنجی مدل‌های یادگیری ماشین

مدلسازی و حل مسئله برنامه‌ریزی دوهدفه طراحی شده، به منظور تخصیص دارایی: برای این کار، بازه‌های زمانی یک‌ماهه برای متعادل‌سازی پورتفولیو در نظر گرفته می‌شود. به همین دلیل، به صورت پنجره غلطان با طول پنجره ۳۰ روزه پیش رفته و در هر پنجره، پارامترهای مدل برنامه‌ریزی دوهدفه تخصیص دارایی با استفاده از اطلاعات همان دوره محاسبه شده و مدل بر اساس آن‌ها مجددا حل می‌شود. به این صورت، اوزان پورتفولیو به صورت پویا در هر دوره به‌روزرسانی می‌شوند.

## یافته‌های پژوهش

در این فصل، نتایج محاسباتی حاصل از اجرای الگوریتم‌های پیشرفته یادگیری ماشین و تکنیک‌های بهینه‌سازی بر روی مدل پرتفوی پیشنهادی ارائه شده است. جدول ۱ آمار مربوط به بازده ماهانه ۲۰ شرکت منتخب و شاخص کل (TEDPIX) را نشان می‌دهد و شامل پنج ستون اصلی است: میانگین، انحراف معیار، چارک اول، میانه و چارک

سوم. در این تحلیل، هدف بررسی و مقایسه عملکرد مالی این شرکت‌ها بر اساس بازده ماهانه آنهاست. با توجه به این که میانگین بازده مثبت است، به نظر می‌رسد که شرکت‌ها به طور کلی عملکرد مناسبی دارند، اگرچه پراکندگی در بازده‌ها متفاوت است.

**جدول ۱: آماره‌های توصیفی بازده ماهانه**

| | میانگین | انحراف معیار | چارک اول | میانه | چارک سوم |
|---|---|---|---|---|---|
| ومعادن | 3.71 | 14.11 | -5.11 | 0.5 | 9.81 |
| وتجارت | 2.82 | 14.83 | -3.9 | 0 | 7.22 |
| وبملت | 2.5 | 13.14 | -4.21 | 0 | 8.28 |
| وبصادر | 2.33 | 14.11 | -3.95 | 0 | 7.84 |
| کگل | 3.89 | 12.73 | -4.81 | 0.61 | 10.54 |
| کچاد | 2.97 | 13.76 | -4.83 | 0 | 9.01 |
| کاما | 3.69 | 20.09 | -8.75 | 0.22 | 10.25 |
| فملی | 4.01 | 15.16 | -3.51 | 0.53 | 9.3 |
| فخوز | 3.88 | 12.45 | -3.95 | 1.62 | 10.94 |
| غگل | 4.14 | 19.29 | -3.43 | 0.35 | 7.08 |
| شلرد | 2.88 | 16.82 | -5.78 | 0 | 9.08 |
| شپنا | 5.4 | 31.23 | -5.42 | 0 | 9.29 |
| سمگا | 5.56 | 21.9 | -5.78 | 0 | 11.71 |
| خودرو | 3.76 | 21.79 | -8.65 | -0.89 | 9.17 |
| خساپا | 3.61 | 21.99 | -7.57 | 0 | 10.87 |
| خزامیا | 3.32 | 21.53 | -8.07 | -1.16 | 9.57 |
| تایرا | 4.01 | 14.46 | -4.93 | -0.36 | 10.4 |
| پارسان | 5.94 | 36.42 | -4.18 | 1.36 | 10.13 |
| بترانس | 3.49 | 16.36 | -7.49 | 0.07 | 12.59 |
| اخابر | 2.13 | 14.28 | -4.31 | 0.17 | 5.09 |
| شاخص کل | 0.14 | 1.12 | -0.33 | 0.05 | 0.55 |
| تغییرات نرخ ارز | 8.72 | 16.43 | -0.90 | 4.10 | 13.32 |
| تغییرات نرخ تورم | 3.99 | 17.02 | -6.16 | -0.11 | 13.67 |
| نرخ بهره سالانه | 17.02 | 2.27 | 15.00 | 17.00 | 18.00 |

در این پژوهش، از الگوریتم‌های یادگیری ماشین به دو منظور استفاده شده است: پیش‌بینی شاخص و پیش‌بینی بازده شرکت‌ها. در پیش‌بینی شاخص، متغیرهای ورودی مدل‌ها در هر ماه شامل تغییرات نرخ ارز، تغییرات نرخ

تورم و نرخ بهره در ماه فعلی به همراه مقادیر بازده شاخص در ماه‌های گذشته است. متغیر خروجی نیز پیش‌بینی بازده شاخص در دوره بعد است. به منظور انتخاب تعداد ماه‌های گذشته که به عنوان ورودی انتخاب می‌شوند از تابع خودهمبستگی (ACF[1]) استفاده شده است. این تابع، مقدار همبستگی شاخص با مقادیر گذشته خود را نشان می‌دهد. بر این اساس، دوره‌های پیشین که دارای قدر مطلق خودهمبستگی بیشتر از ۰.۲ هستند به عنوان ورودی مدل در نظر گرفته می‌شوند. در پیش‌بینی بازده شرکت‌ها، متغیرهای ورودی مدل‌ها در هر ماه شامل مقادیر بازده شرکت در ماه‌های گذشته و مقادیر بازده شاخص در ماه‌های گذشته است. متغیر خروجی نیز پیش‌بینی بازده شرکت در دوره بعد است. مجددا به منظور انتخاب تعداد ماه‌های گذشته که به عنوان ورودی انتخاب می‌شوند از تابع خودهمبستگی استفاده شده است. بر این اساس، دوره‌های پیشین که دارای قدر مطلق خودهمبستگی بیشتر از ۰.۲ هستند به عنوان ورودی مدل در نظر گرفته می‌شوند. پس از ایجاد ورودی‌ها و خروجی‌های مدل‌ها، داده‌ها به سه بخش تقسیم می‌شوند: ۷۰ درصد برای آموزش، ۱۰ درصد برای اعتبارسنجی و ۲۰ درصد برای آزمون مدل‌ها. مدل‌ها با استفاده از داده‌های آموزش، آموزش داده‌شده و توسط داده‌های اعتبارسنجی، اعتبارسنجی شده‌اند. درنهایت، بهترین ترکیبات از پارامترهای مدل بر اساس معیار میانگین مربعات خطا (MSE[2]) انتخاب شده‌اند. از یک شبکه NARX با تابع هسته چند جمله‌ای استفاده شده است. بنابراین یکی از پارامترهای تنظیمی درجه این تابع چندجمله‌ای است که می‌تواند دارای مقادیر ۱ تا ۴ باشد. به علاوه، از الگوریتم Adam برای پیاده‌سازی گرادیان نزولی به عنوان تابع بهینه‌ساز استفاده شده است. در الگوریتم Adam، که یکی از روش‌های محبوب بهینه‌سازی در یادگیری عمیق است، از دو پارامتر به نام‌های بتا-۱ ($\beta_1$) و بتا-۲ ($\beta_2$) استفاده می‌شود. این پارامترها به‌طور خاص برای کنترل میانگین‌ها و واریانس‌های متحرک به کار می‌روند. دو پارامتر اصلی الگوریتم جنگل تصادفی مورد بررسی قرار گرفته‌اند که شامل تعداد تخمین‌گرها (درختان تصمیم) و حداکثر عمق درختان است. پارامترهای اصلی مورد بررسی در رگرسیون بردار پشتیبان نیز شامل C، Kernel و Epsilon هستند که هر یک نقش مهمی در تنظیم عملکرد مدل دارند. جدول ۲ پارامترهای منتخب الگوریتم‌های یادگیری ماشین را نشان می‌دهد.

**جدول ۲: پارامترهای منتخب الگوریتم‌های یادگیری ماشین**

| نماد | NARX | | | | RF | | SVR | | |
|---|---|---|---|---|---|---|---|---|---|
| | درجه تابع چندجمله‌ای | بتا-۱ | بتا-۲ | تعداد نورون لایه پنهان | تعداد تخمین گر | حداکثر عمق | C | Kernel | Epsilon |
| ومعادن | 4 | 0.5 | 0.85 | 2 | 100 | 2 | 1 | sigmoid | 0.2 |
| وتجارت | 1 | 0.99 | 0.99 | 2 | 10 | 2 | 0.01 | linear | 1 |
| وبملت | 4 | 0.5 | 0.75 | 2 | 20 | 2 | 1 | rbf | 1 |

[1] Auto-Correlation Function
[2] Mean Squared Error

| نماد | NARX | | | | RF | | SVR | | |
|---|---|---|---|---|---|---|---|---|---|
| | درجه تابع جندجمله‌ای | بتا-۱ | بتا-۲ | تعداد نورون لایه پنهان | تعداد تخمین گر | حداکثر عمق | C | Kernel | Epsilon |
| وبصادر | 4 | 0.99 | 0.99 | 2 | 10 | 10 | 5 | rbf | 1 |
| کگل | 3 | 0.9 | 0.85 | 2 | 100 | 2 | 5 | rbf | 0.5 |
| کچاد | 1 | 0.9 | 0.99 | 2 | 100 | 2 | 5 | rbf | 0.01 |
| کاما | 1 | 0.1 | 0.85 | 2 | 100 | 2 | 5 | rbf | 1 |
| فملی | 2 | 0.99 | 0.75 | 2 | 20 | 2 | 5 | rbf | 0.01 |
| فخوز | 2 | 0.9 | 0.85 | 2 | 20 | 2 | 5 | rbf | 0.2 |
| غگل | 2 | 0.5 | 0.85 | 2 | 2 | 2 | 0.5 | sigmoid | 1 |
| شلرد | 3 | 0.1 | 0.85 | 2 | 20 | 2 | 1 | sigmoid | 1 |
| شپنا | 3 | 0.99 | 0.75 | 2 | 10 | 2 | 5 | linear | 1 |
| سمگا | 2 | 0.1 | 0.75 | 2 | 5 | 2 | 1 | rbf | 1 |
| خودرو | 4 | 0.9 | 0.85 | 2 | 100 | 100 | 5 | rbf | 0.05 |
| خساپا | 1 | 0.9 | 0.75 | 2 | 2 | 5 | 5 | rbf | 1 |
| خزامیا | 2 | 0.9 | 0.75 | 2 | 20 | 2 | 5 | rbf | 0.5 |
| تایرا | 2 | 0.5 | 0.75 | 2 | 5 | 2 | 0.5 | rbf | 1 |
| پارسان | 4 | 0.99 | 0.85 | 2 | 100 | 2 | 5 | rbf | 0.2 |
| بترانس | 1 | 0.5 | 0.85 | 2 | 5 | 2 | 1 | rbf | 1 |
| اخابر | 1 | 0.1 | 0.99 | 2 | 10 | 2 | 5 | rbf | 0.5 |

جدول ۳ نتایج مدل‌های پیش‌بینی بازده را نشان می‌دهد. در این مقایسه، سه مدل SVR (رگرسیون بردار پشتیبانی)، RF (جنگل تصادفی)، و NARX (شبکه عصبی خودبازگشتی) برای پیش‌بینی داده‌های آزمون با چهار شاخص میانگین مربعات خطا (MSE)، جذر میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE1)، و

[1] Mean Absolute Error

میانگین خطای پیش‌بینی (MFE[1]) بررسی شده‌اند. مدل‌ها در چندین نماد بورسی ارزیابی شده‌اند. به‌طور کلی، هر سه مدل SVR، RF و NARX در پیش‌بینی نمادهای مختلف بازار سرمایه عملکرد متفاوتی را نشان می‌دهند و نمی‌توان به طور قاطع یکی از آن‌ها را به عنوان بهترین مدل در همه نمادها انتخاب کرد. این تفاوت عملکرد به میزان پیچیدگی داده‌ها، میزان نوسانات، و حساسیت مدل‌ها نسبت به شاخص‌های خطا بستگی دارد. مدل NARX در برخی نمادها در شاخص‌های MSE و RMSE عملکرد قابل‌قبولی ارائه داده است و به دلیل ساختار شبکه عصبی خودبازگشتی، قادر به دنبال کردن روندهای پیچیده‌تر و شناسایی الگوهای غیرخطی است. این مدل در نمادهایی مانند «کچاد» و «کاما» عملکرد مناسبی از خود نشان داده که نشان می‌دهد توانایی تشخیص الگوهای دنباله‌دار را بهتر از سایر مدل‌ها دارد. این امر ممکن است به دلیل خصوصیت بازگشتی مدل NARX باشد که اجازه می‌دهد تا مدل از اطلاعات گذشته برای پیش‌بینی‌های دقیق‌تر استفاده کند.

**جدول ۳: مقایسه عملکرد مدل‌های پیش‌بینی در داده‌های آزمون**

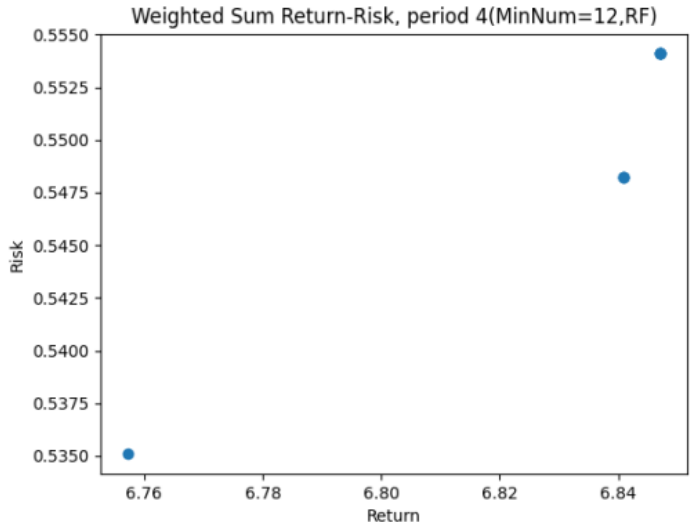
| نماد | SVR | | | | RF | | | | NARX | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MSE | RMSE | MAE | MFE | MSE | RMSE | MAE | MFE | MSE | RMSE | MAE | MFE |
| ومعادن | 141.03 | 11.88 | 20.9 | 0.18 | 143.64 | 11.99 | 16.43 | 1.84 | 152.95 | 12.37 | 7.89 | 3.84 |
| وتجارت | 177.24 | 13.31 | 2.2 | 1.68 | 182.2 | 13.5 | 2.86 | 3.61 | 216.56 | 14.72 | 5.57 | 6.79 |
| وبملت | 126.65 | 11.25 | 6.09 | -0.8 | 118.2 | 10.87 | 6.54 | 0.46 | 124.41 | 11.15 | 7.38 | -1.12 |
| وبصادر | 173.22 | 13.16 | 11.57 | 1.76 | 213.86 | 14.62 | 4.48 | 2.77 | 312.03 | 17.66 | 13.47 | 9.3 |
| کگل | 110.73 | 10.52 | 23.78 | 0.95 | 108.55 | 10.42 | 13.58 | 3.35 | 118.13 | 10.87 | 26.47 | 3.69 |
| کچاد | 121.95 | 11.04 | 12.39 | 1.25 | 120.61 | 10.98 | 4.18 | 1.75 | 94.16 | 9.7 | 8.28 | -0.21 |
| فملی | 116.85 | 10.81 | 16.3 | -3.02 | 448.3 | 21.17 | 21.83 | 4.91 | 352.35 | 18.77 | 18.83 | 2.82 |
| کاما | 344.09 | 18.55 | 6.85 | 2.31 | 107.45 | 10.37 | 13.76 | 1.22 | 91.23 | 9.55 | 5.23 | 2.37 |
| فخوز | 126.19 | 11.23 | 10.1 | 2.17 | 100.26 | 10.01 | 8.01 | 2.94 | 98.36 | 9.92 | 10.13 | 1.9 |
| غگل | 198.72 | 14.1 | 0.89 | -0.87 | 1059.93 | 32.56 | 3.17 | 6.8 | 837.22 | 28.93 | 5.47 | 23.91 |
| شلرد | 181.38 | 13.47 | 22.2 | -0.78 | 180.71 | 13.44 | 20.15 | 1.5 | 181.93 | 13.49 | 22.83 | -0.9 |
| شپنا | 158.17 | 12.58 | 16.79 | -1.54 | 172.32 | 13.13 | 14.33 | 0.17 | 140.35 | 11.85 | 12.53 | -2.7 |
| سمگا | 188.18 | 13.72 | 24.56 | 3.05 | 257.21 | 16.04 | 24.96 | 6.13 | 281.64 | 16.78 | 26.89 | 7.31 |
| خودرو | 188.28 | 13.72 | 11.96 | -0.85 | 144.31 | 12.01 | 12.18 | 0.98 | 171.72 | 13.1 | 9.65 | 3.47 |
| خساپا | 160.48 | 12.67 | 18.12 | -1.45 | 372.9 | 19.31 | 22.44 | 4.33 | 164.05 | 12.81 | 26.42 | 0.58 |
| خزامیا | 246.23 | 15.69 | 6.2 | -3.93 | 195.32 | 13.98 | 6.15 | -2.19 | 248.02 | 15.75 | 13 | -5.66 |
| تایرا | 154.01 | 12.41 | 8.36 | -2.05 | 149.76 | 12.24 | 6.96 | 2.49 | 152.86 | 12.36 | 4.01 | 3.68 |
| پارسان | 125.79 | 11.22 | 14.05 | -0.68 | 163 | 12.77 | 20.94 | 0.61 | 153.82 | 12.4 | 17.28 | 6.42 |
| بترانس | 218.2 | 14.77 | 4.03 | 0.14 | 249.39 | 15.79 | 1.75 | 1.43 | 234.28 | 15.31 | 7.2 | -0.35 |
| اخابر | 284.75 | 16.87 | 7.12 | 0.06 | 285.54 | 16.9 | 2.62 | 3.12 | 311.24 | 17.64 | 3.97 | -0.49 |

[1] Mean Forecast Error

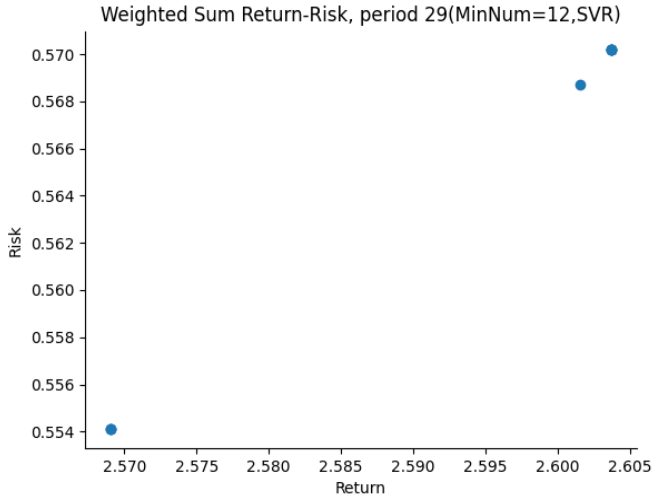
به منظور حل مسئله دو هدفه بهینه‌سازی پورتفولیو از دو رویکرد مجموع وزنی استفاده شده است. فرض شده است که ترکیب پورتفولیو هر ماه به‌روزرسانی می‌شود. بنابراین، در هر رویکرد، ۳۳ مسئله برنامه‌ریزی خطی (برای ۳۳ دوره آزمون) حل و تحلیل می‌شود. در هر دوره، پیش‌بینی بازده به دست آمده توسط الگوریتم‌های مختلف به عنوان بازده مورد انتظار هر سهم در تابع هدف استفاده می‌شود. در این رویکرد، مسئله دو هدفه اصلی به یک مسئله یک هدفه با تابع هدف جدید تبدیل می‌شود. تابع هدف مسئله جدید به صورت زیر است:

$$Max\ Z = w_1 * \left(\sum_{i=1}^{n} x_i r_i\right) - w_2 * \left(\sum_{i=1}^{n} x_i p_i\right)$$

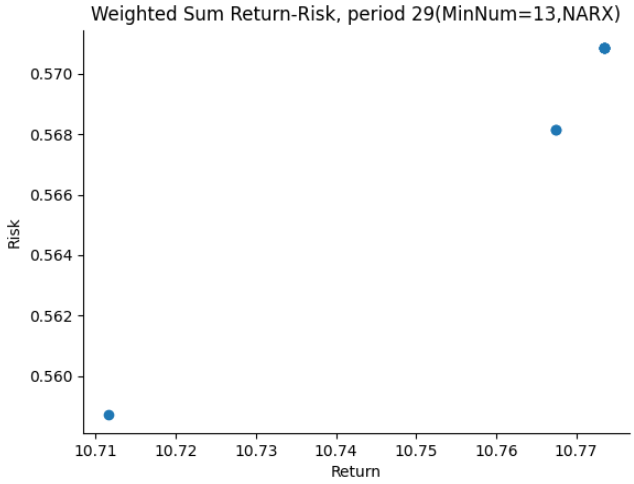
که در آن $w_1$ و $w_2$ به ترتیب اوزان تابع هدف اول (بازده) و تابع هدف دوم (ریسک) هستند. انتخاب این دو پارامتر به همراه پارامتر حداقل تعداد سهام در پورتفولیو ($mininnumber$) بستگی به ترجیحات سرمایه‌گذار دارد و نمی‌توان مقدار مشخصی برای آن در نظر گرفت. به همین منظور، مسئله با درنظر گرفتن مقادیر مختلف برای این پارامترها حل شده است (مقادیر ۰.۱ تا ۰.۹ برای اوزان و ۱ تا ۲۰ برای حداقل تعداد سهام در پورتفولیو). شکل‌های ۱ تا ۳ جبهه‌های پارتو را برای برخی از حالات نشان می‌دهد. از این شکل‌ها می‌توان دریافت که (همانگونه که انتظار می‌رود) ارتباط بین ریسک و بازده پوتفولیو مستقیم است. به عبارتی، برای ایجاد بازد مورد انتظار بیشر، ریسک بیشتری باید تحمل کرد.




**شکل ۱: جبهه پارتو برای پورتفولیو دوره ۴ با حداقل تعداد سهام در پورتفولیو برابر ۱۲ و استفاده از الگوریتم RF برای پیش‌بینی بازده**




**شکل ۲: جبهه پارتو برای پورتفولیو دوره ۲۹ با حداقل تعداد سهام در پورتفولیو برابر ۱۲ و استفاده از الگوریتم SVR برای پیش‌بینی بازده**



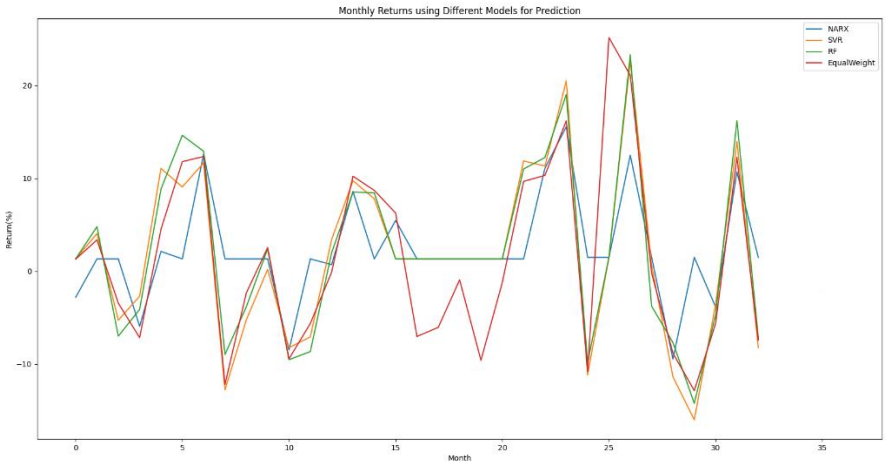

**شکل ۳: جبهه پارتو برای پورتفولیو دوره ۲۹ با حداقل تعداد سهام در پورتفولیو برابر ۱۳ و استفاده از الگوریتم NARX برای پیش‌بینی بازده**

در جدول ۴ بازده محقق شده پورتفولیو بهینه‌سازی‌شده بر اساس پیش‌بینی بازده با استفاده از الگوریتم‌های مختلف در ۳۳ ماه به‌صورت مقایسه‌ای ارائه شده است. الگوریتم‌های مورد استفاده شامل NARX، SVR و RF هستند و

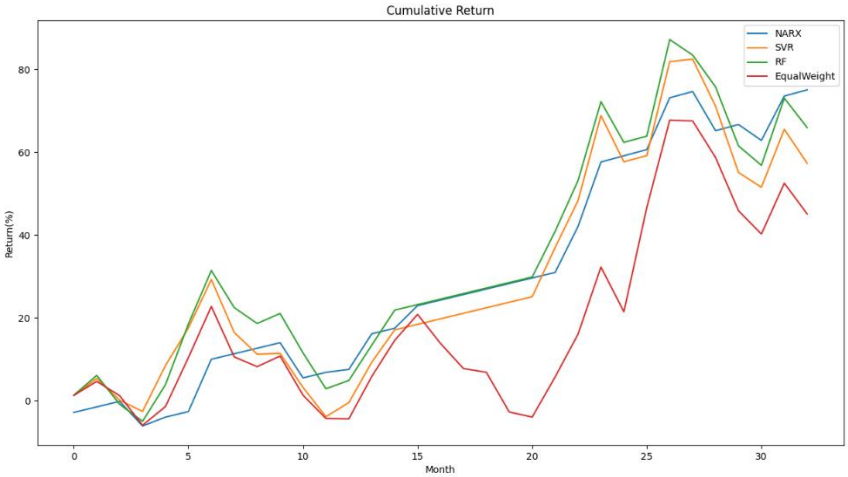
نتایج این الگوریتم‌ها با یک پورتفولیو با وزن مساوی (EqualWeight) نیز مقایسه شده‌اند. در این جدول، میزان بازده برای هر ماه با حداقل تعداد سهام در پورتفولیو برابر با ۱۰ و اوزان ۰.۲ و ۰.۸ برای بازده و ریسک (به طور نمونه) نشان داده شده است. شکل‌های ۴ و ۵ مقادیر بازده ماهانه و تجمعی را به صورت بصری نشان داده و می‌توان دید که پورتفولیوهای مبتنی بر پیش‌بینی همگی بهتر از پورتفولیو با اوزان برابر عمل کرده‌اند.

**جدول ۴: بازده محقق شده پورتفولیو بهینه‌سازی‌شده بر اساس بازده پیش‌بینی‌شده با الگوریتم‌های مختلف با حداقل تعداد سهام در پورتفولیو برابر ۱۰ و اوزان ۰.۲ و ۰.۸ برای اهداف بازده و ریسک**

| EqualWeight | RF | SVR | NARX | ماه |
|---|---|---|---|---|
| 1.30 | 1.33 | 1.33 | -2.81 | 1 |
| 3.37 | 4.80 | 4.06 | 1.33 | 2 |
| -3.43 | -6.99 | -5.27 | 1.33 | 3 |
| -7.15 | -4.11 | -2.68 | -5.94 | 4 |
| 4.52 | 8.85 | 11.08 | 2.13 | 5 |
| 11.81 | 14.64 | 9.08 | 1.33 | 6 |
| 12.36 | 12.93 | 11.68 | 12.63 | 7 |
| -12.19 | -8.96 | -12.78 | 1.33 | 8 |
| -2.35 | -3.83 | -5.24 | 1.33 | 9 |
| 2.55 | 2.42 | 0.20 | 1.33 | 10 |
| -9.44 | -9.52 | -8.23 | -8.49 | 11 |
| -5.63 | -8.65 | -7.07 | 1.33 | 12 |
| -0.10 | 2.00 | 3.41 | 0.71 | 13 |
| 10.24 | 8.53 | 9.76 | 8.60 | 14 |
| 8.71 | 8.45 | 7.80 | 1.33 | 15 |
| 6.27 | 1.33 | 1.33 | 5.47 | 16 |
| -7.02 | 1.33 | 1.33 | 1.33 | 17 |
| -6.04 | 1.33 | 1.33 | 1.33 | 18 |
| -0.92 | 1.33 | 1.33 | 1.33 | 19 |
| -9.58 | 1.33 | 1.33 | 1.33 | 20 |
| -1.21 | 1.33 | 1.33 | 1.33 | 21 |
| 9.68 | 11.02 | 11.88 | 1.33 | 22 |
| 10.33 | 12.26 | 11.34 | 11.10 | 23 |
| 16.21 | 19.05 | 20.51 | 15.56 | 24 |
| -10.79 | -9.82 | -11.16 | 1.50 | 25 |
| 25.16 | 1.50 | 1.50 | 1.50 | 26 |
| 21.10 | 23.32 | 22.67 | 12.50 | 27 |
| -0.19 | -3.75 | 0.63 | 1.52 | 28 |
| -8.83 | -7.68 | -11.37 | -9.46 | 29 |
| -12.85 | -14.22 | -16.00 | 1.50 | 30 |
| -5.63 | -4.71 | -3.57 | -3.84 | 31 |
| 12.28 | 16.21 | 13.98 | 10.71 | 32 |
| -7.44 | -7.12 | -8.23 | 1.50 | 33 |




**شکل ۴: بازده ماهانه پورتفولیو بهینه‌سازی‌شده بر اساس بازده پیش‌بینی‌شده با الگوریتم‌های مختلف با حداقل تعداد سهام در پورتفولیو برابر ۱۰ و اوزان ۰.۲ و ۰.۸ برای اهداف بازده و ریسک**




**شکل ۵: بازده تجمعی پورتفولیو بهینه‌سازی‌شده بر اساس بازده پیش‌بینی‌شده با الگوریتم‌های مختلف با حداقل تعداد سهام در پورتفولیو برابر ۱۰ و اوزان ۰.۲ و ۰.۸ برای اهداف بازده و ریسک**

آمار توصیفی مقادیر بازده با ترکیب اوزان و تعداد سهم مورد نظر در جدول ۵ نشان داده شده است. بر اساس این جدول، مشاهده می‌شود که پورتفولیو ساخته شده بر اساس همه الگوریتم‌های NARX، SVR و RF به طور میانگین بازدهی بهتری نسبت به اوزان برابر دارند که نشان دهنده عملکرد بهتر آن‌ها است. همچنین پورتفولیو ساخته شده بر اساس هر سه الگوریتم دارای انحراف معیار بازده کمتری نسبت به پورتفولیو با اوزان برابر هستند. به خصوص، پورتفولیو ساخته شده بر اساس NARX کمترین انحراف معیار بازده را دارد که نشان دهنده باثبات‌ترین عملکرد است. به علاوه، کمترین زیان نیز مربوط به NARX است.

**جدول ۵: آمار توصیفی مقادیر بازده با ترکیب اوزان و تعداد سهم منتخب**

| | NARX | SVR | RF | EqualWeight |
|---|---|---|---|---|
| میانگین | 2.27 | 1.74 | 2.00 | 1.37 |
| انحراف معیار | 5.56 | 9.31 | 9.24 | 9.92 |
| حداقل | -9.46 | -16.00 | -14.22 | -12.85 |
| میانه | 1.33 | 1.33 | 1.33 | -0.19 |
| حداکثر | 15.56 | 22.67 | 23.32 | 25.16 |

نتایج جداول ۴ و ۵ برای یک مجموعه پارامتر خاص و به صورت نمونه ارائه شده است. درصورتی که در هر ماه بهترین مجموعه پارامترها انتخاب شود این پارامترها در جدول ۶ نمایش داده شده است. سلول‌هایی که با عبارت "سرمایه‌گذاری بدون ریسک" پر شده است نشان می‌دهد که در دوره مربوطه، الگوریتم متناظر با آن سلول پیش‌بینی بازده شاخص کمتر از نرخ بازده بدون ریسک را داشته و بنابراین نیاز به حل مدل بهینه‌سازی نبوده است. جدول ۷، بازده محقق شده پورتفولیو بهینه‌سازی‌شده بر اساس پیش‌بینی بازده با استفاده از الگوریتم‌های مختلف در ۳۳ ماه را نشان می‌دهد. شکل‌های ۶ و ۷ مقادیر بازده ماهانه و تجمعی را نشان می‌دهد. از این شکل‌ها می‌توان مشاهده کرد که پورتفولیوهای مبتنی بر پیش‌بینی همگی بهتر از پورتفولیو با اوزان برابر عمل کرده‌اند.

**جدول ۶: بهترین پارامترهای مدل بهینه‌سازی پورتفولیو در هر دوره**

| | NARX | | | RF | | | SVR | | |
|---|---|---|---|---|---|---|---|---|---|
| **ماه** | **حداقل تعداد سهم** | **وزن بازده** | **وزن ریسک** | **حداقل تعداد سهم** | **وزن بازده** | **وزن ریسک** | **حداقل تعداد سهم** | **وزن بازده** | **وزن ریسک** |
| **1** | **15** | **0.1** | **0.9** | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک |
| **2** | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | **12** | **0.1** | **0.9** | **12** | **0.1** | **0.9** |
| **3** | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | **20** | **0.1** | **0.9** | **20** | **0.1** | **0.9** |
| **4** | **1** | **0.1** | **0.9** | **5** | **0.1** | **0.9** | **5** | **0.1** | **0.9** |

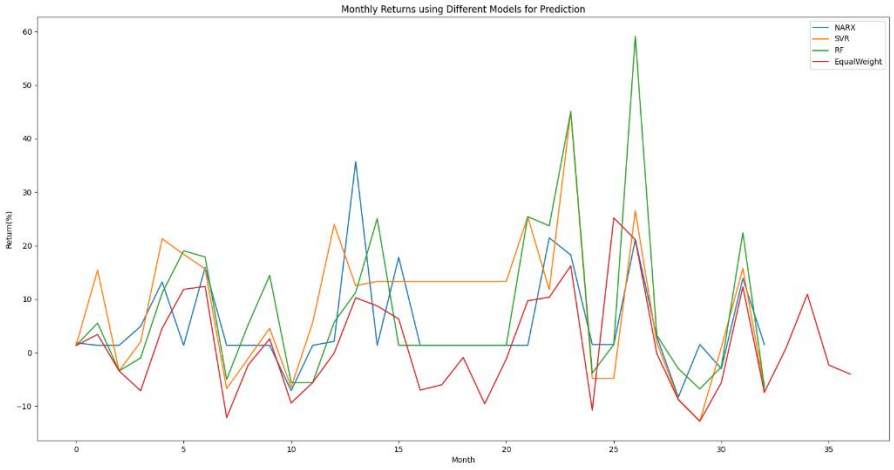
| | NARX | | | RF | | | SVR | | |
|---|---|---|---|---|---|---|---|---|---|
| ماه | حداقل تعداد سهم | وزن بازده | وزن ریسک | حداقل تعداد سهم | وزن بازده | وزن ریسک | حداقل تعداد سهم | وزن بازده | وزن ریسک |
| 5 | 2 | 0.3 | 0.7 | 4 | 0.1 | 0.9 | 4 | 0.1 | 0.9 |
| 6 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | 1 | 0.1 | 0.9 | 1 | 0.1 | 0.9 |
| 7 | 3 | 0.1 | 0.9 | 1 | 0.1 | 0.9 | 1 | 0.1 | 0.9 |
| 8 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | 1 | 0.1 | 0.9 | 1 | 0.1 | 0.9 |
| 9 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | 2 | 0.1 | 0.9 | 2 | 0.1 | 0.9 |
| 10 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | 2 | 0.1 | 0.9 | 2 | 0.1 | 0.9 |
| 11 | 9 | 0.1 | 0.9 | 2 | 0.1 | 0.9 | 2 | 0.1 | 0.9 |
| 12 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | 20 | 0.1 | 0.9 | 20 | 0.1 | 0.9 |
| 13 | 13 | 0.1 | 0.9 | 1 | 0.2 | 0.8 | 1 | 0.2 | 0.8 |
| 14 | 1 | 0.1 | 0.9 | 7 | 0.1 | 0.9 | 7 | 0.1 | 0.9 |
| 15 | 0 | 0 | 0 | 1 | 0.1 | 0.9 | 1 | 0.1 | 0.9 |
| 16 | 1 | 0.1 | 0.9 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک |
| 17 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک |
| 18 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک |
| 19 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک |
| 20 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک |
| 21 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک |
| 22 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | 1 | 0.1 | 0.9 | 1 | 0.1 | 0.9 |
| 23 | 1 | 0.1 | 0.9 | 2 | 0.1 | 0.9 | 2 | 0.1 | 0.9 |
| 24 | 11 | 0.1 | 0.9 | 1 | 0.1 | 0.9 | 1 | 0.1 | 0.9 |
| 25 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | 1 | 0.1 | 0.9 | 1 | 0.1 | 0.9 |

| | NARX | | | RF | | | SVR | | |
|---|---|---|---|---|---|---|---|---|---|
| ماه | حداقل تعداد سهم | وزن بازده | وزن ریسک | حداقل تعداد سهم | وزن بازده | وزن ریسک | حداقل تعداد سهم | وزن بازده | وزن ریسک |
| 26 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک |
| 27 | **20** | **0.1** | **0.9** | **1** | **0.1** | **0.9** | **1** | **0.1** | **0.9** |
| 28 | **14** | **0.1** | **0.9** | **2** | **0.2** | **0.8** | **2** | **0.2** | **0.8** |
| 29 | **11** | **0.3** | **0.7** | **2** | **0.1** | **0.9** | **2** | **0.1** | **0.9** |
| 30 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | **3** | **0.1** | **0.9** | **3** | **0.1** | **0.9** |
| 31 | **9** | **0.1** | **0.9** | **8** | **0.1** | **0.9** | **8** | **0.1** | **0.9** |
| 32 | **4** | **0.2** | **0.8** | **1** | **0.1** | **0.9** | **1** | **0.1** | **0.9** |
| 33 | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | سرمایه‌گذاری بدون ریسک | **1** | **0.1** | **0.9** | **1** | **0.1** | **0.9** |

**جدول ۷: بازده محقق شده پورتفولیو بهینه‌سازی‌شده بر اساس بازده پیش‌بینی‌شده با الگوریتم‌های مختلف و بهترین پارامترهای هر دوره**

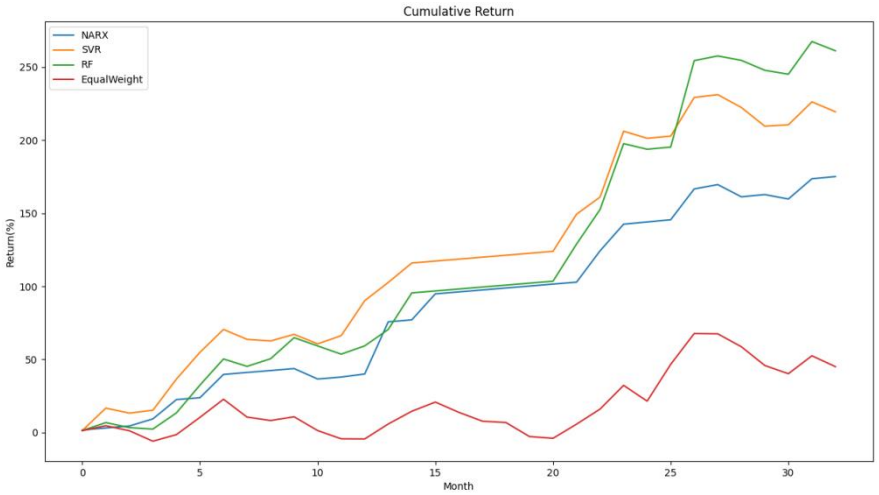
| ماه | NARX | SVR | RF | EqualWeight |
|---|---|---|---|---|
| 1 | 1.81 | 1.33 | 1.33 | 1.30 |
| 2 | 1.33 | 15.40 | 5.50 | 3.37 |
| 3 | 1.33 | -3.43 | -3.43 | -3.43 |
| 4 | 4.88 | 1.99 | -1.02 | -7.15 |
| 5 | 13.18 | 21.28 | 11.09 | 4.52 |
| 6 | 1.33 | 18.36 | 19.02 | 11.81 |
| 7 | 15.91 | 15.62 | 17.86 | 12.36 |
| 8 | 1.33 | -6.78 | -5.06 | -12.19 |
| 9 | 1.33 | -1.11 | 5.21 | -2.35 |
| 10 | 1.33 | 4.53 | 14.41 | 2.55 |
| 11 | -7.13 | -6.50 | -5.64 | -9.44 |
| 12 | 1.33 | 5.60 | -5.63 | -5.63 |
| 13 | 2.09 | 23.92 | 5.67 | -0.10 |
| 14 | 35.70 | 12.48 | 11.24 | 10.24 |
| 15 | 1.33 | 13.29 | 25.01 | 8.71 |
| 16 | 17.80 | 1.33 | 1.33 | 6.27 |
| 17 | 1.33 | 1.33 | 1.33 | -7.02 |
| 18 | 1.33 | 1.33 | 1.33 | -6.04 |
| 19 | 1.33 | 1.33 | 1.33 | -0.92 |
| 20 | 1.33 | 1.33 | 1.33 | -9.58 |
| 21 | 1.33 | 1.33 | 1.33 | -1.21 |
| 22 | 1.33 | 25.39 | 25.39 | 9.68 |

| EqualWeight | RF | SVR | NARX | ماه |
|---|---|---|---|---|
| 10.33 | 23.69 | 11.75 | 21.43 | 23 |
| 16.21 | 45.02 | 45.02 | 18.24 | 24 |
| -10.79 | -3.85 | -4.85 | 1.50 | 25 |
| 25.16 | 1.50 | 1.50 | 1.50 | 26 |
| 21.10 | 59.12 | 26.47 | 21.10 | 27 |
| -0.19 | 3.21 | 1.89 | 2.97 | 28 |
| -8.83 | -3.01 | -8.74 | -8.34 | 29 |
| -12.85 | -6.81 | -12.85 | 1.50 | 30 |
| -5.63 | -2.68 | 0.96 | -3.02 | 31 |
| 12.28 | 22.38 | 15.70 | 13.85 | 32 |
| -7.44 | -6.36 | -6.79 | 1.50 | 33 |




**شکل ۶: بازده ماهانه پورتفولیو بهینه‌سازی‌شده بر اساس بازده پیش‌بینی‌شده با الگوریتم‌های مختلف با بهترین پارامترها**

آمار توصیفی مقادیر بازده بر اساس بهترین ترکیب در جدول ۸ نشان داده شده است. بر اساس این جدول، مشاهده می‌شود که پورتفولیو ساخته شده بر اساس همه الگوریتم‌های NARX، SVR و RF به طور میانگین بازدهی بهتری نسبت به پورتفولیو با اوزان برابر دارند که نشان دهنده عملکرد بهتر آن‌ها است. همچنین پورتفولیو ساخته شده بر اساس NARX دارای انحراف معیار بازده کمتری نسبت به پورتفولیوهای دیگر است که نشان دهنده باثبات‌ترین عملکرد است.




**شکل ۷: بازده تجمعی پورتفولیو بهینه‌سازی‌شده بر اساس بازده پیش‌بینی‌شده با الگوریتم‌های مختلف با بهترین پارامترها**

**جدول ۸: آمار توصیفی مقادیر بازده با بهترین پارامترها**

| | NARX | SVR | RF | EqualWeight |
|---|---|---|---|---|
| میانگین | 5.31 | 6.65 | 7.91 | 1.37 |
| انحراف معیار | 9.22 | 12.45 | 15.06 | 9.92 |
| حداقل | -8.34 | -12.85 | -6.81 | -12.85 |
| میانه | 1.33 | 1.50 | 1.33 | -0.19 |
| حداکثر | 35.70 | 45.02 | 59.12 | 25.16 |

## بحث و نتیجه‌گیری

این بخش به بررسی نتایج به‌کارگیری مدل‌های یادگیری ماشین از جمله NARX، RF، و SVR در بهبود پیش‌بینی بازدهی و کنترل ریسک پورتفولیو پرداخته است. مدل NARX به دلیل ساختار شبکه عصبی غیرخطی و توانایی استفاده از ورودی‌های برون‌زا، در پیش‌بینی بازده‌های باثبات‌تر و بلندمدت عملکرد مطلوبی داشته است. از سوی دیگر، RF با ترکیب چندین درخت تصمیم توانسته در شناسایی الگوهای پیچیده و کاهش نویز در پیش‌بینی‌های کوتاه‌مدت مؤثر باشد. مدل SVR نیز با ساختار خطی و کنترل خطا، در پیش‌بینی‌های کوتاه‌مدت و داده‌های ساده‌تر کارایی بالایی داشته است.

نتایج نشان می‌دهند که ترکیب مدل‌های پیش‌بینی با بهینه‌سازی چندهدفه موجب بهبود عملکرد پورتفولیو شده و باعث کاهش ریسک و افزایش بازدهی می‌شود. این روش‌ها، به‌ویژه در شرایط بی‌ثبات بازار، ثبات نسبی بیشتری را ارائه کرده و در بلندمدت مزیت‌های قابل توجهی دارند. همچنین، مدل دوهدفه به‌عنوان ابزاری برای تعادل بین بازده و ریسک پیشنهاد شده که اثربخشی بیشتری در کنار پیش‌بینی‌های یادگیری ماشین دارد. این یافته‌ها مدیران سرمایه‌گذاری را ترغیب می‌کند که از ترکیب مدل‌های پیش‌بینی و بهینه‌سازی برای دستیابی به اهداف مالی خود استفاده کنند.

برای ارتقای دقت و قابلیت پیش‌بینی در داده‌های مالی، تاکید اصلی باید بر استفاده از یادگیری ماشین و توسعه مدل‌های پیشرفته پیش‌بینی باشد. پژوهشگران باید با همکاری‌های بین‌سازمانی میان نهادهای مالی، دانشگاه‌ها و مراکز تحقیقاتی، کیفیت و دسترسی به داده‌ها را بهبود بخشند و از تکنیک‌های داده‌کاوی برای استخراج و پاکسازی داده‌ها استفاده کنند. تمرکز اصلی در پژوهش‌های آتی باید بر توسعه الگوریتم‌های هوشمند پیش‌بینی، مانند مدل‌های یادگیری ماشین از جمله NARX، RF و SVR باشد. این مدل‌ها باید بهینه‌سازی شوند تا دقت پیش‌بینی را افزایش داده و روابط پیچیده در داده‌ها را شناسایی کنند.

علاوه بر این، طراحی مدل‌هایی که به‌طور پویا قادر به تنظیم خودکار پارامترها و تطبیق با شرایط متغیر بازار باشند، از اهمیت بالایی برخوردار است. این امر می‌تواند با استفاده از روش‌های پیشرفته مانند بهینه‌سازی ژنتیک، یادگیری تقویتی و تکنیک‌های جدید یادگیری ماشین محقق شود. بازبینی و به‌روزرسانی دوره‌ای این مدل‌ها برای حفظ دقت و تعمیم‌پذیری آنها ضروری است، به‌ویژه در مواجهه با تغییرات ناگهانی در بازارهای مالی. یادگیری ماشین با قدرت پیش‌بینی بالا و توانایی پردازش سریع داده‌های پیچیده می‌تواند به‌عنوان ابزاری کلیدی در استخراج ارزش از داده‌های مالی و بهبود تصمیم‌گیری در این حوزه عمل کند.

## فهرست منابع


اقتصاد, امیرعلی؛ محمدی,عمران. (1402). بهینه سازی سبد سرمایه گذاری به کمک پیش بینی بازده مورد انتظار با استفاده از روش های شبکه عصبی LSTM، جنگل تصادفی و ARIMA. *چشم انداز مدیریت مالی*.

تهرانی رضا؛ محمد هندیجانی زاده، و عیسی نوروزیان لکوان ، (۱۳۹۴). ارائه رویکردی جدید برای مدیریت فعال پرتفوی وانجام معاملات هوشمند سهام با تاکید بر نگرش انتخاب ویژگی. دانش سرمایه‌گذاری، ۴(۱۳): ۱۰۷۱۲۶.

رستمی, ژیلا, فتاحی, سهیلی. (۱۴۰۲). مدل‌سازی و تخمین بازده بورس اوراق بهادار تهران با استفاده از مدل‌های پویا. اقتصاد مالی, ۱۷(۶۲), ۱۸۵۲۱۶.

شبان مهدی ؛ اله نخعی حبیب؛ الله طالب نیاقدرت و نازنین بشیری منش. (۱۳۹۹). طراحی الگوی غیرخطی سرایت‌پذیری شاخص کل بورس اوراق بهادار تهران از بازار دارایی‌های فیزیکی (کاربردی از مدل شبکه عصبی مصنوعی .(NARXمهندسی مالی و مدیریت اوراق بهادار، ۱۲(۴۶): ۴۷۶۴۹۴.

علیزاده, حسین؛ کیانفر, کامران. (۱۴۰۲). توسعهٔ مدل مارکوویتز در بهینه‌سازی سبد سهام با در نظر گرفتن محدودیت‌های واقع‌گرایانه. مدیریت دارایی و تامین مالی, ۱۱(۴), ۶۵۹۲.

کرامتی, اسفندیار؛ غلامی جمکرانی, و کاشفی نیشابوری. (۱۴۰۱). بهینه‌سازی پرتفوی سهام در بورس اوراق بهادار تهران (کاربرد رهیافت یادگیری تقویتی). مدلسازی اقتصادی, ۱۶(۵۸), ۵۱۶۶.

میرعلوی سید حسن؛ زهرا پورزمانی و آزیتا جهانشاد. (۱۳۹۸). ارائه مدلی مبتنی بر رفتار مالی سرمایه گذاران جهت پیش بینی قیمت سهام با استفاده از روش های فرا ابتکاری شبکه های عصبی. فصلنامه بورس اوراق بهادار، ۴۷(۱۲): ۱۰۹۱۷۶.

Atiya, A. F., El-Shoura, S. M., Shaheen, S. I., & El-Sherif, M. S. (1999). A comparison between neural-network forecasting techniques - case study: river flow forecasting. IEEE Transactions on neural networks, 10(2), 402-409.

Awad, M., Khanna, R., Awad, M., & Khanna, R. (2015). Support vector regression. Efficient learning machines: Theories, concepts, and applications for engineers and system designers, 67-80.

Ban, G. Y., El Karoui, N., & Lim, A. E. (2018). Machine learning and portfolio optimization. Management Science, 64(3), 1136-1154.

Behera, J., Pasayat, A. K., Behera, H., & Kumar, P. (2023). Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. Engineering Applications of Artificial Intelligence, 120, 105843.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks, 5(2), 157-166.

Box, G. E., Jenkins, G. M., & Reinsel, G. C. (1994). Time series analysis: Forecasting and control. 3rd Prentice Hall. Englewood Cliffs NJ, 614.

Chen, W., Zhang, H., Mehlawat, M. K., & Jia, L. (2021). Mean–variance portfolio optimization using machine learning-based stock price prediction. Applied Soft Computing, 100, 106943.

Cheng, Q., Yang, L., Zheng, J., Tian, M., & Xin, D. (2024). Optimizing Portfolio Management and Risk Assessment in Digital Assets Using Deep Learning for Predictive Analysis. arXiv preprint arXiv:2402.15994.

Coyle, D., Prasad, G., & McGinnity, T. M. (2005). A time-series prediction approach for feature extraction in a brain-computer interface. IEEE transactions on neural systems and rehabilitation engineering, 13(4), 461-467.

Dablemont, S., Simon, G., Lendasse, A., Ruttiens, A., Blayo, F., & Verleysen, M. (2003). Time series forecasting with SOM and local non-linear models-Application to the DAX30 index prediction. In WSOM 2003, Workshop on Self-Organizing Maps.

Gandhmal Dattatray P و Kumar K. (2021). Wrapper-Enabled Feature Selection and CPLM-Based NARX Model for Stock Market Prediction. The Computer Journal٦٤، (2): 169-184.

García-Gonzalo, E., García-Nieto, P. J., Fidalgo Valverde, G., Riesgo Fernández, P., Sánchez Lasheras, F., & Suárez Gómez, S. L. (2024). Hybrid DE-Optimized GPR and NARX/SVR Models for Forecasting Gold Spot Prices: A Case Study of the Global Commodities Market. Mathematics, 12(7), 1039.

Grigoryan Hakob. (2015). Stock market prediction using artificial neural networks. Case Study of TAL1T, Nasdaq OMX Baltic Stock. Database Systems Journal٦، (2): 14-23.

Guarino, A., Santoro, D., Grilli, L., Zaccagnino, R., & Balbi, M. (2024). EvoFolio: a portfolio optimization method based on multi-objective evolutionary algorithms. Neural Computing and Applications, 36(13), 7221-7243.

Habbab, F. Z., & Kampouridis, M. (2024). An in-depth investigation of five machine learning algorithms for optimizing mixed-asset portfolios including REITs. Expert Systems with Applications, 235, 121102.

Jia Lu؛ Shazemeen Noor Muhammad و Martinkute-Kauliene Raimonda. (2020). Portfolio Decision Using Time Series Prediction and Multi-objective Optimization. Romanian Journal of Economic Forecasting۲۳ ،(4): 118.

Kantz, H., & Schreiber, T. (2003). Nonlinear time series analysis. Cambridge university press.

Labde Saurabh؛ Patel Stuti و Shukla Megh. (2017). Time series regression model for prediction of closing values of the stock using an adaptive narx neural network. International Journal of Computer Applications۱۵۸ ،(10): 0975-8887.

Lin, T., Horne, B. G., & Giles, C. L. (1998). How embedded memory in recurrent neural network architectures helps learning long-term temporal dependencies. Neural Networks, 11(5), 861-868.

Lwin, K. T., Qu, R., & MacCarthy, B. L. (2017). Mean-VaR portfolio optimization: A nonparametric approach. European Journal of Operational Research, 260(2), 751-766.

Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. Expert Systems with Applications, 165, 113973.

Malandreniotis, D. (2024). Probabilistic Forecasting Models for Multidimensional Financial Time-series With Applications to Systematic Portfolio Management (Doctoral dissertation, UCL (University College London)).

Montenegro Carlos و Molina Marco. (2019). Using Deep Neural Networks for Stock Market Data Forecasting: An Effectiveness Comparative Study. the The International Conference on Advances in Emerging Trends and Technologies.

Odeyemi, O., Mhlongo, N. Z., Nwankwo, E. E., Scholatica, U. C., & Okoye, C. C. (2024). Big data applications in portfolio management: A review of techniques and strategies. International Journal of Science and Research Archive, 11(1), 1984-1999.

Pearlmutter, B. A. (1995). Gradient calculations for dynamic recurrent neural networks: A survey. IEEE Transactions on Neural networks, 6(5), 1212-1228.

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

Sen, J., & Dutta, A. (2023). Portfolio optimization for the Indian stock market. In Encyclopedia of Data Science and Machine Learning (pp. 1904-1951). IGI Global.

Wu, L., Ahmad, M., Qureshi, S. A., Raza, K., & Khan, Y. A. (2022). An analysis of machine learning risk factors and risk parity portfolio optimization. Plos one, 17(9), e0272521.

Xiu Yan و Chen Xinye. (2017). Study on prediction of the shanghai composite index based on EMD and NARX neural network Information Technology and Intelligent Transportation Systems (pp. 590-596): IOS Press.

*Financial Economics*
*Vol. (16) Issue (62) Summer 2027*

# Multi-objective optimization of the stock portfolio based on machine learning models and comparison with the stock market index

**Behnaz Badaei**
Department of Accounting and Finance, South Tehran Branch, Islamic Azad University, Tehran, Iran
b.badaei@iau.ac.ir

**Hossein Badiei**
Department of Accounting and Finance, South Tehran Branch, Islamic Azad University, Tehran, Iran
(Corresponding Author)
badiei@iau.ac.ir

**Mohsen Hamidian**
Department of Accounting and Finance, South Tehran Branch, Islamic Azad University, Tehran, Iran
m_hamidian@azad.ac.ir

**Ali Amiri**
Department of Accounting and Finance, Bandar Abbas Branch, Islamic Azad University, Bandar Abbas, Iran.
a_amiri@iau.ac.ir

## Abstract

This thesis examines the prediction of financial behaviors using advanced machine learning models, including NARX, RF and SVR. The main goal of this research is to present a portfolio model based on this machine learning models and multi-objective optimization. Considering the importance of accurately predicting financial behaviors in investment decisions and risk management, this research tries to help clarify the strengths and weaknesses of each of the examined models and provide a portfolio optimization model based on it.

In this research, the historical data of Tehran stock market (from 2011 to 2023), including price information and macroeconomic variables, has been used. Data were collected from reliable and reliable sources and processed and prepared for better analysis. Also, to accurately evaluate the performance of the models, machine learning methods were used and the parameters of the models were fine-tuned.

The results of this research show that NARX, RF and SVR models have different capabilities in price prediction and their performance is affected by data characteristics. In particular, the NARX model shows a more stable performance. In addition, portfolios based on machine learning algorithms have been able to perform better than portfolios with equal weights in all cases. Also, this study identifies challenges related to setting parameters and computational complexity and emphasizes the importance of updating models according to market changes. Finally, suggestions for improving the models and solutions for future research are provided to increase the accuracy of predictions and obtain better results in the analysis of financial behaviors.

**Keywords:** NARX model, portfolio optimization, machine learning