Journal of Defense Economics & Sustainable Development, Vol. 10, No. 37, September 2025, pp. 121-147

Defense Economics & Sustainable Development

Journal homepage: https://eghtesad.sndu.ac.ir/

# Analyzing the Application of Artificial Intelligence and Machine Learning Algorithms for Evaluating the Financial Performance of Publicly Listed Investment Companies with an Emphasis on the Case of Ghadir Investment Group

Hossein Shirani[1*], Sajjad Mirzaei[2]

1. Ph.D., Department of Chemistry, Molecular Simulation Research Laboratory, University of Science and Technology, Tehran. Corresponding Author. Email:shiranihossein@gmail.com
2. M.A., Department of Financial Management, Faculty of Economics and Management, Urmia University, Urmia, Iran. Email: st_s.mirzaei@urmia.ac.ir

**Article Info**

**Article type:**
Research Article

**Article history:**
Received: 22-12-2024
Accepted: 03-05-2025

**Keywords**:
Machine Learning, Feature Engineering, Ensemble Learning, Performance Prediction.

**Abstract**

The objective of the present study was to identify the optimal combination of machine learning algorithms, feature selection methods, and clustering techniques for predicting company performance on the Tehran Stock Exchange. The statistical population includes all companies listed on the Tehran Stock Exchange between the years 1389 and 1402. A systematic elimination method was employed to select the sample, resulting in a final dataset of 171 companies. The collected data were analyzed using RapidMiner for machine learning processes and EViews for econometric analysis.

The results obtained from conducting comparative analysis revealed that advanced models such as support vector machines, random forests, and neural networks generally outperform traditional methods in financial forecasting tasks. However, the competitive performance of simpler models, such as logistic regression in specific scenarios, highlights the ongoing relevance of model interpretability and suggests caution against unnecessary algorithmic complexity. Our findings indicate that dimensionality reduction techniques—namely principal component analysis (PCA) and self-organizing maps (SOMs)—often result in reduced predictive accuracy in this context, suggesting that their use should be guided by a clear understanding of the data characteristics and the predictive objectives. Moreover, ensemble approaches, particularly bagging, significantly enhanced model robustness, demonstrating the value of combining multiple learners to navigate the inherent volatility of financial markets. Notably, our top-performing models generalized well when applied to the portfolio of Ghadir Investment Holding, underscoring their practical applicability. The findings of the study demonstrate that models trained on broader market data can provide valuable insights for specialized investment portfolios. Therefore, the findings offer meaningful implications for investment firms and financial analysts, highlighting the potential of machine learning for improving decision-making processes.

**Cite this article:** Shirani, H., & Mirzaei, S. (2025). Analyzing the Application of Artificial Intelligence and Machine Learning Algorithms for Evaluating the Financial Performance of Publicly Listed Investment Companies with an Emphasis on the Case of Ghadir Investment Group. *Journal of Defense Economics & Sustainable Development*, 10 (37), 121-147.

فصلنامه اقتصاد دفاع و توسعه پایدار، دوره ۱۰، شماره ۳۷، مهر ۱۴۰۴، صفحه ۱۲۱-۱۴۷

اقتصاد دفاع و توسعه پایدار

شاپا چاپی: ۳۰۶۰-۱۵۳۱ شاپا الکترونیکی: ۲۹۸۱-۰۶۱X

# هوش مصنوعی و الگوریتم های مختلف یادگیری ماشین در ارزیابی عملکرد مالی شرکت های سرمایه‌گذاری بورسی: با تأکید بر گروه سرمایه گذاری غدیر

حسین شیرانی[1*]، سجاد میرزایی[2]

۱. دکترا، گروه شیمی، آزمایشگاه تحقیقاتی شبیه سازی مولکولی، دانشگاه علم وصنعت، تهران، ایران. نویسنده مسئول.
رایانامه: shiranihossein@gmail.com

۲. کارشناسی ارشد، گروه مدیریت مالی، دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ارومیه، ایران.
رایانامه: st_s.mirzaei@urmia.ac.ir

**اطلاعات مقاله**

**نوع مقاله:**
مقاله علمی

**تاریخچه مقاله:**
تاریخ ارسال: ۱۴۰۳/۱۰/۰۲
تاریخ پذیرش: ۱۴۰۴/۰۲/۱۳

**واژگان کلیدی:**
هوش مصنوعی، یادگیری ماشین، یادگیری گروهی، مهندسی ویژگی، پیش‌بینی عملکرد.

## چکیده

پژوهش حاضر به دنبال پاسخ به این سوال می باشد که چه ترکیبی از الگوریتم‌های یادگیری ماشین، روش‌های انتخاب ویژگی و تکنیک‌های گروهی، قوی‌ترین پیش‌بینی عملکرد شرکت ها را در بورس اوراق بهادار تهران ارائه می‌کند؟ در این مسیر تمامی شرکت‌های پذیرفته‌شده در بورس تهران طی سال‌های ۱۳۸۹ تا ۱۴۰۲ به عنوان جامعه آماری در نظر گرفته شدند و برای انتخاب نمونه، از روش حذف سیستماتیک استفاده شد که در نهایت، ۱۷۱ شرکت به عنوان نمونه نهایی انتخاب گردید. داده‌های گردآوری‌شده با استفاده از نرم‌افزار یادگیری ماشین RapidMiner و نرم‌افزار اقتصادسنجی EViews مورد تجزیه‌وتحلیل قرار گرفته شد.

نتایج تحلیل مقایسه‌ای مدل‌های مختلف یادگیری ماشین نشان داد که ماشین‌های بردار پشتیبان، جنگل تصادفی و شبکه‌های عصبی عموماً از سایر الگوریتم‌ها در وظایف پیش‌بینی مالی بهتر عمل می‌کنند. عملکرد رقابتی مدل‌های ساده‌تر مانند رگرسیون لجستیک در سناریوهای خاص، ارزش تفسیرپذیری مدل و خطر پیچیدگی غیرضروری را به ما یادآوری کرد، بنابراین یک رویکرد متعادل با در نظر گرفتن عملکرد مدل و تفسیرپذیری، در وظایف پیش‌بینی مالی بسیار ضروری به نظر می¬رسد. مطالعه ما نشان داد که تکنیک‌های کاهش ابعاد مانند تجزیه و تحلیل مؤلفه اصلی و نقشه‌های خودسازمان دهنده در این زمینه مفید نیستند و اغلب منجر به کاهش عملکرد می‌شوند. این یافته مهم نسبت به کاربرد بی¬رویه چنین تکنیک‌هایی هشدار می‌دهد و بر نیاز به در نظر گرفتن دقیق ماهیت داده‌ها و وظیفه پیش‌بینی خاص تاکید می‌کند. استفاده از روش‌های گروهی، به‌ویژه بگینگ، عملکرد مدل‌های قوی را بیشتر افزایش داد، بنابراین ترکیب چندین مدل یا نسخه از یک مدل می‌تواند به پیش‌بینی‌های قوی‌تر در محیط پرنوسان بازار سهام منجر شود. شاید مهمترین نکته این باشد که بهترین مدل‌های پژوهش تعمیم بسیار خوبی به پرتفوی خاص مانند هلدینگ سرمایه گذاری غدیر داشتند. این نشان‌دهنده کاربرد عملی روش پژوهش حاضر است و نشان می‌دهد که مدل‌های آموزش داده شده بر روی داده‌های گسترده‌تر بازار می‌توانند بینش‌های ارزشمندی برای پرتفوی‌های سرمایه‌گذاری خاص ارائه دهند.

استناد به مقاله: شیرانی، حسین و میرزایی، سجاد. (۱۴۰۴). هوش مصنوعی و الگوریتم های مختلف یادگیری ماشین در ارزیابی عملکرد مالی شرکت های سرمایه‌گذاری بورسی: با تأکید بر گروه سرمایه گذاری غدیر، *فصلنامه اقتصاددفاع و توسعه پایدار*، ۱۰(۳۷)، ۱۲۱-۱۴۷.

ناشر: انجمن علمی اقتصاد دفاع ایران

## ۱. مقدمه

در سال‌های اخیر، نوسانات در بازارهای مالی بین‌المللی به دلیل شوک‌های ناشی از بیماری‌های همه‌گیر و خطرات ژئوپلیتیک افزایش یافته است و به دنبال آن رشد اقتصادی اقتصادهای بزرگ جهان و کشورهای نوظهور روند نزولی داشته است. شوک‌های خارجی جریان منابع تولیدی مانند نیروی کار و سرمایه را مختل نموده و باعث کاهش سرمایه‌گذاری، ظرفیت تولید، مصرف و صادرات و همچنین درآمد شرکت‌ها شده است (ونگ و همکاران[۱]، ۲۰۲۳)

بازار سهام ایران در مقایسه با بازار سهام کشورهای توسعه یافته به دلیل نوسانات مکرر و چشمگیر قیمت از بلوغ کافی برخوردار نیست؛ بنابراین، پیش‌بینی عملکرد مالی شرکت‌های پذیرفته شده در بورس اهمیت عملی و ارزش مرجع تصمیم‌گیری زیادی برای توسعه سیستم مالی دارد. در سطح شرکت، پیش‌بینی مؤثر عملکرد مالی شرکت به مدیران کمک می‌کند تا مشکلات در عملیات را به‌موقع تشخیص دهند و در نتیجه کنترل داخلی و مدیریت عملیات را تقویت کنند. از دیدگاه سرمایه‌گذاران، درک روند آینده عملکرد مالی شرکت‌ها برای ارزیابی به‌موقع ریسک‌های مالی شرکت‌ها و اتخاذ تصمیمات سرمایه‌گذاری صحیح مفید است. علاوه بر این، برای تنظیم‌گر بازار، درک روند عملکرد مالی شرکت‌ها برای نظارت بر ریسک‌های مالی بازار و جلوگیری از وقوع ریسک‌های سیستماتیک در سیستم مالی مفید است (لیانگ و همکاران[۲]، ۲۰۲۴).

صورت‌های مالی اسناد اساسی هستند که وضعیت مالی شرکت، نتایج عملیاتی و جریان‌های نقدی آن را در طول یک دوره حسابداری خاص منعکس می‌کنند (جان[۳]، ۲۰۱۸). نسبت‌های مالی به‌عنوان ابزار سنتی و در عین حال برجسته در پیش‌بینی سلامت مالی شرکت ظاهر شده‌اند و درک بهتری نسبت به مقادیر مطلق شناسایی‌شده در صورت‌های مالی ارائه می‌دهند. این نسبت‌ها به مقایسه عملکرد نسبی شرکت‌ها بین صنایع، درون گروه‌ها و بخش‌های مختلف در خود شرکت کمک می‌کنند. (مانوگنا و همکاران[۴]، ۲۰۲۲).

از نظر فنی، پیش‌بینی عملکرد مالی شرکت‌ها، اظهارنظر در مورد عملکرد آینده شرکت‌ها است (فریمن و همکاران[۵]، ۱۹۸۲). مدل‌های آماری و یادگیری ماشین مانند ماشین بردار پشتیبان و شبکه عصبی مصنوعی به طور مؤثری در پیش‌بینی عملکرد مالی شرکت‌ها در مطالعات مرتبط استفاده شده‌اند (لیانگ و همکاران، ۲۰۲۴). مدل‌های پیش‌بینی باید تا حد ممکن، صرف نظر از تکنیک‌های استفاده شده برای توسعه آن‌ها، مؤثر باشند. بنابراین، مطالعات عمدتاً بر تعیین بهترین مدلی که می‌تواند بالاترین دقت پیش‌بینی یا کمترین خطای پیش‌بینی را ارائه دهد، تمرکز کرده‌اند. چندین عامل کلیدی بر عملکرد نهایی مدل‌های پیش‌بینی تأثیر می‌گذارند. یکی از عوامل مهم، انتخاب ویژگی و استخراج ویژگی است. از آنجا که انواع ویژگی‌های مرتبط که می‌توانند بالاترین قدرت تمایز را برای تشخیص بین کلاس‌های مختلف ارائه دهند، نامشخص است، انجام انتخاب ویژگی برای

---

[1] Wong et al
[2] Liang et al
[3] Jan
[4] Manogna et al
[5] Freeman et al

تجزیه و تحلیل سطح نمایندگی ویژگی از مجموعه داده‌های جمع‌آوری شده معمولاً برای بهبود عملکرد پیش‌بینی مدل‌ها مفید است. به عنوان مثال، لیانگ و همکاران[1] (۲۰۱۵) اثر اعمال روش‌های انتخاب ویژگی مبتنی بر فیلتر را بر عملکرد پیش‌بینی وضعیت مالی نامناسب را بررسی کردند. میرزایی و همکاران (۱۴۰۲) تاثیر انتخاب ویژگی رگرسیون لاسو را بر بهبود پیش‌بینی نگهداشت وجه نقد بررسی کردند. کاهش ابعاد نیز به روش‌های مختلف قبلاً توسط چندین محقق مانند تاسی و همکاران (۲۰۲۱) و جابر و همکاران[2] (۲۰۲۲) آزمایش شده است. برخلاف دو عامل قبلی یعنی انتخاب ویژگی و استخراج ویژگی که مربوط به مسائل پیش‌پردازش داده‌ها هستند، تکنیک‌های یادگیری گروهی که با ترکیب چندین طبقه‌بندی‌کننده کار می‌کنند، نشان داده‌اند که در پیش‌بینی ورشکستگی و امتیازدهی اعتباری بهتر از طبقه‌بندی‌کننده‌های منفرد عمل می‌کنند (سان و همکاران[3]، ۲۰۲۰).

مطالب فوق اهمیت درنظرگرفتن تکنیک‌های انتخاب ویژگی، استخراج ویژگی و یادگیری گروهی را برای بهبود پیش‌بینی نشان می‌دهد. بااین‌حال، بیشتر پژوهش‌ها فقط بر یکی از این سه رویکرد تمرکز کرده‌اند؛ بنابراین، هیچ مطالعه‌ای هر سه عامل را با هم برای پیش‌بینی عملکرد شرکت به کار نگرفته است. این مقاله قصد دارد دو شکاف موجود در مرور ادبیات یعنی تأثیر روش‌های استخراج ویژگی و انتخاب ویژگی بر عملکرد مدل‌های پیش‌بینی طبقه‌بندی و تأثیر ترکیب روش‌های یادگیری گروهی با انتخاب ویژگی را بر پیش‌بینی عملکرد شرکت ها پر کند. به‌عبارت‌دیگر، انگیزه این مقاله پاسخ‌دادن به این سؤال است که چه ترکیبی از الگوریتم‌های یادگیری ماشین، روش‌های انتخاب ویژگی و تکنیک‌های گروهی قوی‌ترین پیش‌بینی عملکرد شرکت را در بورس اوراق بهادار تهران ارائه می‌کند؟ بر اساس بررسی های انجام شده، این سؤال تحقیقاتی قبلاً در حوزه ادبیات مرتبط با پیش‌بینی عملکرد شرکت پاسخ داده نشده است. البته هدف این مقاله تنها توسعه یک مدل پیش‌بینی واحد نیست، بلکه ارائه یک نمای کلی از نحوه پردازش داده‌های واقعی توسط روش‌های طبقه‌بندی، انتخاب ویژگی، استخراج ویژگی و یادگیری گروهی است. علاوه بر این، این مقاله روشی نظام‌مند را برای خوانندگان ارائه می‌دهد که تأثیری را که این روش‌ها ممکن است بر تشخیص عملکرد شرکت داشته باشند، توصیف می‌کند.

## ۲. مبانی نظری و پیشینه پژوهش

### ۲-۱. مبانی نظری پژوهش

#### ۲-۱-۱. انتخاب ویژگی و متغیر هدف

ویژگی‌های مورد استفاده برای پیش‌بینی عملکرد مالی شرکت‌ها را می‌توان به ویژگی‌های مالی و غیرمالی تقسیم کرد. مانوگنا و همکاران (۲۰۲۲) از ویژگی‌های مالی مانند: نسبت‌های نقدینگی، نسبت‌های سودآوری، نسبت‌های رشد، نسبت‌های ساختار دارایی و نسبت‌های توان پرداخت بدهی به عنوان ویژگی‌های مالی برای

[1] Liang et al
[2] Jabeur et al
[3] Sun et al

پیش‌بینی عملکرد شرکت استفاده کردند. لو و همکاران[1] (۲۰۰۱) از ۱۱ ویژگی مالی برای پیش‌بینی بازده مالی آتی شرکت‌های بورس چین استفاده کرد و نشان داد که اطلاعات موجود در ویژگی‌های مالی برای پیش‌بینی بازده مالی آتی مفید است. علاوه بر ویژگی‌های مالی، تأثیر ویژگی‌های غیرمالی نیز در بسیاری از مطالعات در نظر گرفته شده است. آلبوکرکه و همکاران[2] (۲۰۱۹) دریافتند که عملکرد خوب مسئولیت اجتماعی شرکت‌ها می‌تواند ریسک سیستماتیک را کاهش داده و ارزش شرکت را افزایش دهد.

برای متغیر هدف، شاخص‌های مالی مانند بازده دارایی‌ها، نسبت کیو[3] توبین و درآمد هر سهم اغلب برای اندازه‌گیری عملکرد مالی شرکت‌ها انتخاب می‌شوند. از این میان، کیو توبین اندازه‌گیری می‌کند که چگونه بازار کارایی عملیاتی و توانایی ایجاد عملکرد مالی خوب شرکت‌ها را ارزش‌گذاری می‌کند (وانگ و سارکیس[4]، ۲۰۱۷). سود هر سهم معمولاً برای نشان‌دادن نتایج عملیاتی شرکت‌ها استفاده می‌شود که یکی از ویژگی‌های مهم برای ارزیابی ثروت شرکت‌ها است (اویونو و همکاران[5]، ۲۰۱۱). بازده دارایی‌ها نسبت سود خالص به کل دارایی‌ها است که اصول اساسی عملکرد مالی و توانایی‌های عملیاتی شرکت را به طور جامع نشان می‌دهد. عواملی مانند بدهی و دست‌کاری تأثیر کمی بر آن دارند و بازده دارایی را به مؤثرترین و پرکاربردترین معیار عملکرد مالی شرکت‌ها تبدیل می‌کنند (بنر و ولوسو[6]، ۲۰۰۸). بسیاری از محققان بازده دارایی را به‌عنوان متغیر هدف هنگام بررسی تأثیر ویژگی‌های مالی و غیرمالی بر عملکرد مالی شرکت‌ها به کار گرفته‌اند (انجلیا و سورینینگسیه[7]، ۲۰۱۵؛ کییر و آوسلوس[8]، ۲۰۲۱).

### ۲-۱-۲. مدل‌های آماری

در حوزه تحلیل داده‌ها، رگرسیون خطی به عنوان ابزاری کارآمد برای پیش‌بینی مقادیر ناشناخته با بهره‌گیری از داده‌های مرتبط و در دسترس شناخته می‌شود. این تکنیک با مدل‌سازی ریاضی رابطه بین متغیر وابسته (ناشناخته) و متغیر مستقل (شناخته شده) در قالب یک معادله خطی، امکان برآورد دقیق‌تری از مقادیر مجهول را فراهم می‌سازد. در مواردی که بیش از یک متغیر مستقل بر متغیر وابسته تأثیرگذارند، از تحلیل رگرسیون چندگانه استفاده می‌شود. این روش با تعمیم مفهوم رگرسیون خطی ساده، امکان بررسی و تعیین میزان تأثیر هر یک از متغیرهای مستقل بر متغیر وابسته را فراهم می‌آورد. به عبارت دیگر، رگرسیون چندگانه علاوه بر پیش‌بینی مقدار متغیر وابسته، به تحلیل چگونگی و میزان تأثیرگذاری متغیرهای مستقل مختلف بر آن می‌پردازد. شایان ذکر است که برای برآورد پارامترهای مدل در رگرسیون چندگانه، از روش کمترین مربعات

---

[1] Lu et al
[2] Albuquerque et al
[3] Q
[4] Wang & Sarkis
[5] Oeyono et al
[6] Benner & Veloso
[7] Angelia & Suryaningsih
[8] Kyere & Ausloos

استفاده می‌شود. این روش با کمینه‌سازی مجموع مربعات خطا، بهترین برازش را برای داده‌ها فراهم کرده و دقت پیش‌بینی مدل را افزایش می‌دهد (توتکو و همکاران[1]، ۲۰۲۴).

پژوهش‌های گذشته در حوزه پیش‌بینی عملکرد مالی شرکت‌ها، غالباً از رویکردهای رگرسیون خطی یا لجستیک بهره می‌برند. برای نمونه، او و پنمن[2] (۱۹۸۹) با استفاده از ۶۸ شاخص مالی و اعمال مدل رگرسیون لجستیک، به پیش‌بینی روند تغییرات عملکرد شرکت‌ها پرداختند. عیسی و آنتوی[3] (۲۰۱۷) نیز از مدل رگرسیون خطی چندگانه برای پیش‌بینی بازده دارایی شرکت‌ها استفاده کردند و دریافتند که متغیرهای اقتصاد کلان در پیش‌بینی عملکرد مالی شرکت‌ها مؤثر هستند. مدل‌های آماری مذکور، اگرچه قادر به تأیید روابط خطی بین شاخص‌های مالی و عملکرد شرکت‌ها هستند، اما در مواجهه با روابط غیرخطی، کارایی آنها به شدت کاهش می‌یابد. عملکرد پیش‌بینی مدل‌های رگرسیونی به عوامل مختلفی مانند شکل تابع، انتخاب ویژگی‌ها، انتخاب برآوردگر و رفتار جمله خطا بستگی دارد. شاید برخی از این عوامل محدودکننده باعث عملکرد ضعیف‌تر مدل‌های رگرسیونی پیش‌بینی‌های برون‌نمونه شوند (آناند و همکاران[4]، ۲۰۱۹). با این حال، این مدل‌ها به دلیل نیاز به حجم داده‌های کمتر، سادگی در اجرا، بررسی روابط علّی، تفسیرپذیری بالا و قابلیت ترکیب با دانش نظری پیشین، همچنان در حوزه‌های مالی با ساختارهای خطی و پایدار کاربرد دارند (میرزایی و همکاران، ۱۴۰۲). در مقابل، با پیشرفت روزافزون هوش مصنوعی، الگوریتم‌های یادگیری ماشین با قابلیت استخراج روابط غیرخطی بین شاخص‌های مالی و متغیرهای هدف، دقت و پایداری فرآیندهای پیش‌بینی را به طور چشمگیری افزایش داده‌اند.

### ۲-۱-۳. مدل‌های یادگیری ماشین

هوش مصنوعی را می‌توان به چهار شاخه اصلی کاربردی تقسیم نمود: طراحی سیستم‌های خبره، حل مسائل اکتشافی، پردازش زبان طبیعی و بینایی ماشین. هسته مرکزی این کاربردها، یادگیری ماشین است که به سیستم‌های هوش مصنوعی امکان می‌دهد تا از داده‌ها بیاموزند و بدون نیاز به برنامه‌نویسی مجدد، عملکرد خود را ارتقا دهند. بهرام میرزایی[5] (۲۰۱۰) در پژوهشی جامع، کاربردهای هوش مصنوعی در حوزه مالی را بررسی کرده است.

تحلیل و بررسی شاخه‌های مختلف یادگیری ماشین، گامی ضروری در راستای درک عمیق از این حوزه پرکاربرد محسوب می‌شود. روش‌های یادگیری تحت نظارت، بدون نظارت، تقویتی و عمیق به عنوان رویکردهای اصلی در این زمینه مطرح بوده و انتخاب هر یک از آن‌ها مستلزم شناخت دقیق الگوریتم‌ها و هدف مشخص از تحلیل داده‌ها می‌باشد. هدف غایی در یادگیری ماشین، یادگیری تابع نگاشتی است که بتواند به دقت رابطه میان متغیرهای ورودی و خروجی مورد انتظار را مدل‌سازی نماید. به عبارت دیگر، هدف پیش‌بینی

---

[1] Tutcu et al.
[2] Ou and Penman
[3] Issah & Antwi
[4] Anand et al.
[5] Bahrammirzaee

دقیق خروجی‌های جدید بر اساس داده‌های ورودی جدید است. بر اساس میزان دانش قبلی موجود در مورد داده‌ها، روش‌های یادگیری ماشین به دو دسته کلی یادگیری تحت نظارت و بدون نظارت تقسیم‌بندی می‌شوند. در یادگیری تحت نظارت، الگوریتم با مجموعه‌ای از داده‌های آموزشی برچسب‌گذاری شده کار می‌کند که در آن هر نمونه داده با یک پاسخ صحیح مرتبط است. وجود این برچسب‌ها، امکان ارزیابی کمی دقت مدل را فراهم آورده و به الگوریتم اجازه می‌دهد تا ارتباط مستقیمی میان ویژگی‌های ورودی و خروجی مورد نظر برقرار نماید. مسائل طبقه‌بندی و رگرسیون از جمله مسائل متداولی هستند که با استفاده از روش‌های یادگیری تحت نظارت همچون رگرسیون لجستیک، ماشین‌های بردار پشتیبان، شبکه‌های عصبی مصنوعی و جنگل‌های تصادفی قابل حل می‌باشند (به عنوان مثال، ژو و همکاران[۱]، ۲۰۱۹؛ سمیل[۲]، ۲۰۲۰، سرمپینیس و همکاران[۳]، ۲۰۱۷؛ کاتوال و همکاران[۴]، ۲۰۲۰). اگرچه یادگیری تحت نظارت در بسیاری از حوزه‌های کاربردی از اهمیت بالایی برخوردار است، اما دسترسی به مجموعه داده‌های کاملاً برچسب‌گذاری شده همواره با محدودیت‌هایی همراه است. به ویژه در عصر کلان داده، تهیه و برچسب‌گذاری دستی مجموعه داده‌ها، به دلیل هزینه‌های بالا و پیچیدگی‌های عملیاتی، به چالشی جدی تبدیل شده است. در چنین شرایطی، یادگیری بدون نظارت به عنوان یک راهکار جایگزین مطرح می‌شود. در این روش، الگوریتم‌ها با هدف کشف الگوها و ساختارهای نهفته در داده‌های بدون برچسب، به تحلیل داده می‌پردازند. بدون نیاز به اطلاعات قبلی در مورد پاسخ صحیح، این روش قادر است روابط پنهان در داده‌ها را آشکار سازد. از این رو، تکنیک‌های یادگیری بدون نظارت به طور گسترده‌ای در مسائلی همچون خوشه‌بندی داده‌ها، تشخیص ناهنجاری‌ها و تخمین چگالی به کار گرفته می‌شوند. برخی از پرکاربردترین روش‌های یادگیری بدون نظارت شامل الگوریتم‌های همسایگی نزدیک، k–میانگین و تجزیه مقادیر منفرد می‌باشند (چن و هائو[۵]، ۲۰۱۷؛ هستی و همکاران[۶]، ۲۰۰۹؛ شن و همکاران[۷]، ۲۰۱۱). در ادامه، برخی از تکنیک‌های پرکاربرد در یادگیری ماشین تشریح شده است.

شبکه‌های عصبی به عنوان ابزارهای پیشرفته در حوزه یادگیری ماشین، با بهره‌گیری از لایه‌های پنهان، قادر به پردازش و تبدیل اطلاعات ورودی به خروجی هستند. این لایه‌های پنهان با استفاده از وزن‌های تصادفی اولیه که به صورت تکراری برای بهبود دقت پیش‌بینی تعدیل می‌شوند، اطلاعات را از گره‌های ورودی (متغیرهای توضیحی) دریافت و پردازش می‌کنند. فرآیند یادگیری در این شبکه‌ها از طریق تصحیح خطاهای پیش‌بینی در هر مشاهده صورت می‌گیرد. اگرچه شبکه‌های عصبی در مقایسه با روش‌های رگرسیون، برازش بهتری برای داده‌های غیرخطی ارائه می‌دهند، اما پیچیدگی آنها، به ویژه در تفسیر رابطه بین متغیرهای توضیحی و نتیجه پیش‌بینی، چالش‌برانگیز است. تعداد گره‌های پنهان که وظیفه انجام رگرسیون‌های غیرخطی

[1] Zhou et al.
[2] Smyl
[3] Sermpinis et al.
[4] Katuwal et al.
[5] Chen & Hao
[6] Hastie et al.
[7] Shen et al.

را بر عهده دارند، تعیین‌کننده میزان این پیچیدگی است. به همین دلیل، شبکه‌های عصبی اغلب به عنوان جعبه سیاه در نظر گرفته می‌شوند، زیرا نحوه دقیق محاسبات پیش‌بینی در آنها به طور کامل قابل مشاهده نیست. با این وجود، دقت بالای این شبکه‌ها در پیش‌بینی مزیت قابل توجهی است که می‌تواند این محدودیت را تا حدودی جبران کند (اوز و همکاران[1]، ۲۰۲۱).

تکنیک‌های یادگیری عمیق با استفاده از لایه‌های متعدد، به‌تدریج سطوح بالاتری از ویژگی‌ها را از مجموعه‌داده‌های خام ورودی استخراج می‌کنند. رایج‌ترین مدل‌ها، شبکه‌های عصبی پیچشی هستند (جو و همکاران[2]، ۲۰۱۸) که از مکانیسم طبیعی درک بصری موجودات زنده الهام گرفته‌اند. آن‌ها قادر به یادگیری بدون نظارت از داده‌های بدون ساختار یا برچسب‌گذاری شده هستند، در حالی که به ویژه در تشخیص تصویر نیز توانمند هستند.

به گفته کورتس و واپینک[3] (۱۹۹۵)، ماشین بردار پشتیبان رویکردی نوین در حوزه طبقه‌بندی داده‌ها ارائه می‌دهد. ماشین بردار پشتیبان امکان تحلیل و تفکیک داده‌های ورودی را بر اساس نگاشت غیرخطی آنها به فضایی با ابعاد بالاتر فراهم می‌کند. در این روش، ابتدا بردارهای ورودی از طریق یک تبدیل غیرخطی به فضای ویژگی با ابعاد بالا منتقل می‌شوند. سپس، با استفاده از ابرصفحه‌های بهینه، عمل طبقه‌بندی بین گروه‌های مختلف انجام می‌شود. هدف اصلی در محاسبات ماشین بردار پشتیبان، دستیابی به بیشترین جدایی بین دو گروه است. با توجه به خطی یا غیرخطی بودن داده‌ها، جداکننده بین دو گروه می‌تواند تغییر کند. از آنجا که فرض خطی بودن داده‌ها در عمل نادر است، ماشین بردار پشتیبان از توابع کرنل برای تحلیل داده‌های غیرخطی استفاده می‌کند (آتول و موناگان[4]، ۲۰۱۵).

درخت تصمیم، روشی ساده و در عین حال قدرتمند در دسته‌بندی داده‌ها به شمار می‌رود. این روش، با ساختاری شبیه به نمودار درختی، داده‌ها را بر اساس ویژگی که دارند، دسته‌بندی می‌کند. هر گره در این درخت، نمایانگر یک پرسش یا تصمیم درباره یکی از ویژگی‌های داده است. شاخه‌های منشعب از هر گره، پاسخ‌های گوناگون به آن پرسش را نشان می‌دهند و در نهایت، برگ‌های درخت، دسته‌ی نهایی داده‌ها را مشخص می‌کنند (دیلین و همکاران[5]، ۲۰۱۳).

جنگل تصادفی یا تکنیک جنگل بوت‌استرپ می‌تواند برای طبقه‌بندی و رگرسیون استفاده شود. این روش، درخت‌های تصمیم جداگانه را از طریق بوت‌استرپینگ و تعداد نامحدودی از مجموعه‌های داده تصادفی تولید می‌کند. علاوه بر این، متغیرهای توضیحی مشتق‌شده به صورت تصادفی، درخت‌های تصمیم مختلفی را ایجاد می‌کنند که مشکل بیش‌برازش را کاهش می‌دهند و قابلیت تعمیم نتایج پیش‌بینی را افزایش می‌دهند (آتول و موناگان، ۲۰۱۵).

---

[1] Oz et al.
[2] Gu et al.
[3] Cortes & Vapnik
[4] Attewell & Monaghan
[5] Delen et al.

تقویت گرادیان با هدف تبدیل مجموعه‌ای از یادگیرنده‌های ضعیف به یک یادگیرنده قوی، به دنبال بهبود نتایج پیش‌بینی است. این روش با بهینه‌سازی تابع زیان و آموزش یادگیرنده‌های ضعیف بر اساس خطاهای بین مشاهدات پیش‌بینی‌شده و واقعی، عمل می‌کند. در این فرآیند، یادگیرنده‌ها به صورت افزایشی و سریالی ساخته می‌شوند، به گونه‌ای که هر یادگیرنده جدید، خطای یادگیرنده قبلی را اصلاح می‌کند. این سری تا زمانی که به پیش‌بینی رضایت‌بخش دست یابد، ادامه دارد (اوز و همکاران، ۲۰۲۱؛ فریدمن[1]، ۲۰۰۱).

### ۲-۱-۴. تکنیک‌های انتخاب ویژگی

انتخاب ویژگی یا کاهش ابعاد به دنبال انتخاب تعدادی ویژگی نماینده از یک مجموعه‌داده آموزشی داده شده است، جایی که مجموعه‌داده آموزشی با کاهش ابعاد می‌تواند قدرت تمایز بیشتری برای تمایز بین کلاس‌ها نسبت به مجموعه اصلی ارائه دهد. در نتیجه، مدلی که با استفاده از یک مجموعه داده آموزشی با کاهش ابعاد ساخته می‌شود، احتمالاً از مدلی که با استفاده از مجموعه داده آموزشی اصلی به دست می‌آید، عملکرد بهتری خواهد داشت (تاسی و همکاران[2]، ۲۰۲۱).

به‌طورکلی، فرایند انتخاب ویژگی شامل چهار مرحله: تولید زیرمجموعه، ارزیابی زیرمجموعه، معیار توقف و اعتبارسنجی نتیجه است. تولید زیرمجموعه بر اساس استفاده از برخی استراتژی‌های جستجو برای تولید زیرمجموعه‌های ویژگی نامزد برای مراحل ارزیابی بعدی انجام می‌شود. در ارزیابی زیرمجموعه، هر زیرمجموعه نامزد با بهترین زیرمجموعه قبلی بر اساس برخی معیارهای ارزیابی مقایسه می‌شود. مراحل تولید زیرمجموعه و ارزیابی زیرمجموعه تا زمانی که یک معیار توقف معین برآورده شود، تکرار می‌شوند. مرحله نهایی اعتبارسنجی نتیجه بر اساس استفاده از مجموعه‌داده‌های مصنوعی یا واقعی برای اعتبارسنجی بهترین زیرمجموعه انتخاب شده است. الگوریتم‌های انتخاب ویژگی را می‌توان به سه دسته تقسیم کرد: الگوریتم‌هایی که از روش‌های فیلتر، پوششی و تعبیه شده استفاده می‌کنند (بولون کاندو[3] و همکاران، ۲۰۱۳؛ لی و همکاران[4]، ۲۰۱۷). روش‌های فیلتر از برخی معیارهای رتبه‌بندی ویژگی برای شناسایی ویژگی‌هایی که قدرت تمایز بالاتری دارند، استفاده می‌کنند. در مقابل، روش‌های پوششی از یک پیش‌بینی‌کننده یا طبقه‌بندی‌کننده انتخابی استفاده می‌کنند که عملکرد آن به‌عنوان تابع هدف برای ارزیابی زیرمجموعه ویژگی استفاده می‌شود. روش‌های تعبیه شده فرایند انتخاب ویژگی را به‌عنوان بخشی از روش آموزش مدل ادغام می‌کنند، به این معنی که انتخاب ویژگی در مرحله ساخت مدل پیش‌بینی جاسازی شده است. در نتیجه، ویژگی‌های نماینده هنگام توسعه مدل پیش‌بینی انتخاب می‌شوند.

---

[1] Friedman et al.
[2] Tsai et al
[3] Bolón-Canedo et al.
[4] li et al

### ۲-۱-۵. تکنیک استخراج ویژگی

تکنیک‌های کاهش بعد در تحلیل داده و یادگیری ماشین ضروری هستند و هدف آن‌ها ساده‌سازی مجموعه‌داده‌ها با کاهش تعداد ویژگی‌ها درحالی‌که ویژگی‌های اساسی آن‌ها حفظ می‌شود. این تکنیک‌ها به کاهش مشکلاتی مانند نفرین ابعاد کمک می‌کنند، جایی که داده‌های با بعد بالا می‌تواند منجر به بیش‌برازش، افزایش هزینه‌های محاسباتی و مشکلات در تجسم شود. با تبدیل داده‌ها به یک فضای با بعد پایین‌تر، کاهش بعد عملکرد مدل را بهبود می‌بخشد، قابلیت تفسیر را افزایش می‌دهد و شناسایی الگوها در داده‌ها را تسهیل می‌کند.

تحلیل مولفه‌های اصلی یک تکنیک آماری چندمتغیره و ناپارامتریک است که به منظور تحلیل ماتریس داده‌هایی با مشاهدات کمی پیوسته مورد استفاده قرار می‌گیرد. هدف اصلی این روش، کاهش ابعاد داده‌ها از طریق شناسایی و حفظ بیشترین میزان واریانس موجود در آنها است. در تحلیل مولفه‌های اصلی، داده‌های ورودی به متغیرهای جدید و غیرقابل مشاهده‌ای به نام مولفه‌های اصلی تبدیل می‌شوند. این مولفه‌ها که ترکیبات خطی از متغیرهای اولیه هستند، با یکدیگر همبستگی ندارند و به گونه‌ای محاسبه می‌شوند که هر یک، بخش بیشتری از واریانس باقی‌مانده را پوشش دهند. نتیجه این فرآیند، ایجاد چند مولفه اصلی است که بخش اعظم اطلاعات موجود در مجموعه داده را در خود جای می‌دهند. معمولاً، یک تا سه مولفه اصلی اول، مقدار زیادی از واریانس را توضیح می‌دهند. به این ترتیب، تحلیل مولفه‌های اصلی، مولفه‌هایی را ایجاد می‌کند که قدرت توضیحی مجموعه متغیرها را به حداکثر می‌رسانند (اصغری و همکاران، ۱۴۰۱).

نقشه‌های خودسازمان‌دهنده تیوکوهنن[1] به عنوان یک ابزار قدرتمند در تحلیل داده‌های چندمتغیره شناخته شده است. این تکنیک با کاهش ابعاد داده‌ها و حفظ ساختار همسایگی، امکان استخراج اطلاعات پنهان در داده‌ها را فراهم می‌کند. نقشه‌های خودسازمان‌دهنده از طریق فرایند یادگیری خودکار، داده‌ها را به صورت خوشه‌هایی سازمان‌دهی می‌کند که هر خوشه نماینده یک گروه از داده‌های مشابه است. این الگوریتم با استفاده از مکانیزم کوانتیزاسیون برداری و حفظ همسایگی، توانایی بالایی در تشخیص الگوها و ساختارهای نهفته در داده‌ها دارد. کاربردهای نقشه‌های خودسازمان‌دهنده در حوزه‌های مختلفی از جمله اقتصاد، مالی، مهندسی و علوم زیست‌پزشکی بسیار گسترده است (شانموگاناتان[2]، ۲۰۱۸).

### ۲-۱-۶. یادگیری گروهی

در حوزه یادگیری ماشین، طبقه‌بندی‌کننده‌های گروهی یا چندگانه، با بهره‌گیری از تکنیک‌های یادگیری گروهی، به عنوان رویکردی مؤثر مطرح شده‌اند. در این روش، با ساخت چندین طبقه‌بندی‌کننده و ترکیب نتایج آنها، خروجی نهایی بهبود می‌یابد. این تکنیک، با استفاده از روش‌هایی نظیر بگینگ و بوستینگ ، قادر به ارائه نتایج دقیق‌تر و قابل اعتمادتر در مقایسه با طبقه‌بندی‌کننده‌های منفرد است. در روش بگینگ، با نمونه‌برداری

[1] . Teuvo Kohonen
[2] Shanmuganathan

از مجموعه داده آموزشی اصلی و ایجاد زیرمجموعه‌های متعدد، مدل‌های گوناگونی ساخته می‌شوند. سپس، با ترکیب خروجی این مدل‌ها از طریق میانگین‌گیری یا رای‌گیری، نتیجه نهایی حاصل می‌شود. این روش، با کاهش واریانس مدل‌ها، به بهبود عملکرد طبقه‌بندی‌کننده‌ها کمک می‌کند. در مقابل، روش بوستینگ با آموزش تکراری طبقه‌بندی‌کننده‌های ضعیف و تبدیل آنها به دسته‌بندی‌کننده‌های قوی، عمل می‌کند. در این روش، به داده‌های دسته‌بندی‌شده نادرست وزن بیشتری اختصاص داده می‌شود تا دسته‌بندی‌کننده‌های بعدی بر روی آنها تمرکز کنند. این رویکرد، با کاهش بایاس مدل‌ها، دقت دسته‌بندی را افزایش می‌دهد (تاسی و همکاران، ۲۰۲۱).

## ۲-۲. پیشینه پژوهش

### ۲-۲-۱. مطالعات خارجی

لیانگ و همکاران (۲۰۲۴) طی پژوهشی به پیش‌بینی عملکرد مالی شرکت‌ها با استفاده از یادگیری عمیق و یک روش قابل تفسیر پرداختند. آن‌ها شبکه‌های عصبی کانولوشنال[1] یک‌بعدی و مدل‌های یادگیری عمیق حافظه طولانی کوتاه مدت[2] را برای بررسی امکان پیش‌بینی عملکرد مالی شرکت‌ها با مدل‌های یادگیری عمیق، با استفاده از ویژگی‌های مالی شرکت و شاخص رتبه‌بندی محیط، اجتماعی و حاکمیت و داده‌های شرکت‌های چینی پذیرفته شده از سال ۲۰۱۵ تا ۲۰۲۱ ساختند. نتایج آن‌ها نشان داد که مدل‌های یادگیری عمیق می‌توانند به طور مؤثری اطلاعات سری زمانی را از داده‌های شرکت استخراج کنند و در نتیجه پیش‌بینی عملکرد مالی شرکت را با دقت بالا انجام دهند. علاوه بر این نشان دادند که معرفی اطلاعات محیط، اجتماعی و حاکمیت به طور قابل توجهی به دقت پیش‌بینی عملکرد مالی شرکت‌ها کمک می‌کند.

سون و دانگ[3] (۲۰۲۴) طی پژوهشی یک مدل یادگیری ماشینی برای پیش‌بینی شاخص بازده دارایی‌ها معرفی کردند. آن‌ها داده‌ها را از ۷۸ شرکت فهرست شده در بورس‌های اوراق بهادار ویتنام در بازه زمانی ۲۰۱۲ تا ۲۰۲۲ جمع‌آوری کردند. مدل جنگل تصادفی با استفاده از مجموعه داده کسب‌وکارهای منتخب ویتنام در سال ۲۰۲۳ آزمایش شد. نتایج مقدار ضریب تعیین برابر با ۰/۹۷۶۲ و ریشه میانگین مربعات خطا برابر با ۰/۵۸۲۶ را نشان داد که نشان دهنده کاربردهای بالقوه در دنیای واقعی در پیش‌بینی عملکرد کسب‌وکار هستند.

پاپیکووا و همکاران[4] (۲۰۲۲) طی پژوهشی با هدف پیش‌بینی ورشکستگی شرکت‌های کوچک و متوسط، داده‌های مالی تعداد زیادی از این شرکت‌ها را مورد تحلیل قرار دادند. محققان با استفاده از مدل‌های پیش‌بینی مختلف، به دنبال یافتن بهترین روش برای شناسایی شرکت‌هایی بودند که احتمال ورشکستگی آن‌ها بیشتر است. نتایج نشان داد که الگوریتم CatBoost در مقایسه با سایر روش‌ها، دقت بالاتری در پیش‌بینی ورشکستگی داشته است. همچنین، این مطالعه نشان داد که حتی با وجود عدم تعادل در داده‌ها (تعداد

[1] convolutional
[2] long short term memory
[3] Son & Duong
[4] Papíková et al.

شرکت‌های ورشکسته بسیار کمتر از شرکت‌های سالم است)، الگوریتم‌های پیش‌بینی می‌توانند نتایج قابل‌قبولی ارائه دهند.

مانوگنا و همکاران(۲۰۲۰) طی پژوهشی با عنوان اندازه‌گیری عملکرد مالی شرکت‌های تولیدی هند: کاربرد الگوریتم‌های درخت تصمیم، به شناسایی مهم‌ترین معیارهای مالی مؤثر بر عملکرد شرکت‌ها، با بررسی داده‌های مالی شرکت‌های تولیدی هند در سال‌های ۲۰۱۱ تا ۲۰۱۸ پرداختند. محققان با استفاده از تحلیل عاملی اکتشافی و سپس مدل‌های درخت تصمیم، به دنبال کشف ارتباط بین این معیارها و عملکرد شرکت‌ها بوده‌اند. نتایج نشان داد که از بین مدل‌های درخت تصمیم، الگوریتم‌های C5.0 و CHAID بهترین عملکرد را داشته‌اند و مهم‌ترین عوامل مؤثر بر عملکرد شرکت‌ها، حاشیه سود خالص و نرخ گردش دارایی‌های کل شناسایی شدند.

### ۲-۲-۲. مطالعات داخلی

میرزایی و همکاران (۱۴۰۲) طی پژوهشی به مقایسه عملکرد مدل‌های یادگیری ماشین و مدل‌های آماری در پیش‌بینی سود و جریان نقد عملیاتی با استفاده از مجموعه متغیرهای تعهدی و نقدی پرداختند. آن‌ها داده‌های ۱۸۴ شرکت بورس اوراق بهادار تهران طی بازه زمانی ۱۳۹۱ تا ۱۴۰۰ را مورد بررسی قراردادند. محققان نشان دادند که اطلاعات تعهدی نسبت به اطلاعات نقدی توانایی بهتری در پیش‌بینی دارند. همچنین، مدل‌های یادگیری ماشین به طور کلی عملکرد بهتری نسبت به مدل‌های آماری سنتی نشان داده‌اند. با این حال، نتایج نشان داد که در برخی موارد، مدل‌های آماری نیز می‌توانند نتایج قابل قبولی ارائه دهند. به طور کلی، این پژوهش نشان می‌دهد که استفاده از روش‌های یادگیری ماشین و اطلاعات تعهدی می‌تواند به بهبود دقت پیش‌بینی‌های مالی کمک کند.

جعفری و همکاران (۱۴۰۲) طی پژوهشی به پیش‌بینی دقیق‌تر بازده سهام در بورس اوراق بهادار تهران پرداختند. محققان با بررسی داده‌های هفت ساله شرکت‌های بورسی و استفاده از روش‌های یادگیری ماشین، به دنبال یافتن بهترین مدل برای پیش‌بینی بازده بوده‌اند. نتایج نشان داد که برخی متغیرهای مالی مانند نسبت قیمت به سود و اهرم مالی، تاثیر قابل توجهی بر بازده سهام دارند. همچنین، الگوریتم‌های یادگیری ماشین مانند PINSVR و CART توانایی بسیار خوبی در پیش‌بینی بازده نشان دادند. این پژوهش نشان داد که استفاده از هوش مصنوعی و روش‌های پیشرفته آماری می‌تواند به سرمایه‌گذاران و تصمیم‌گیران کمک کند تا تصمیمات سرمایه‌گذاری بهتری اتخاذ کنند.

هارونکلایی و برزگر (۱۴۰۲) طی پژوهشی به شناسایی عواملی که می‌توانند احیای مالی شرکت‌ها را پیش‌بینی کنند پرداختند. محققان با بررسی اطلاعات مالی ۱۷۳ شرکت که در گذشته با مشکلات مالی مواجه شده بودند، به دنبال یافتن مهم‌ترین متغیرهای مالی تأثیرگذار بر احیای این شرکت‌ها بوده‌اند. پس از شناسایی این متغیرها از الگوریتم‌های یادگیری ماشین برای پیش‌بینی احیای شرکت‌ها استفاده شده است. نتایج نشان می‌دهد که برخی متغیرهای مالی خاص به طور قابل‌توجهی در پیش‌بینی احیا مؤثر هستند و الگوریتم ماشین بردار پشتیبان نسبت به الگوریتم درخت تصمیم دقت بالاتری در این پیش‌بینی دارد.

عباسیان و همکاران (۱۴۰۲) طی پژوهشی برای بهبود پیش‌بینی درماندگی مالی شرکت‌ها، روشی نوین پیشنهاد کردند. در این روش، علاوه بر نسبت‌های مالی سنتی، از متغیرهای شبکه‌ای که نشان‌دهنده ارتباطات بین شرکت‌ها هستند، نیز استفاده شده است. با ترکیب این متغیرها و بکارگیری الگوریتم درخت تصمیم تقویت گرادیان که با روش بهینه‌سازی گرگ خاکستری تنظیم شده است، مدل پیش‌بینی قدرتمندتری نسبت به روش‌های سنتی مانند k نزدیک‌ترین همسایه و رگرسیون لجستیک ارائه شده است. نتایج نشان می‌دهد که شرکت‌هایی که در شبکه‌های مالی موقعیت مرکزی‌تری دارند، کمتر در معرض خطر درماندگی مالی قرار دارند.

## ۳. روش‌شناسی پژوهش

در این پژوهش، تمامی شرکت‌های پذیرفته‌شده در بورس تهران طی سال‌های ۱۳۸۹ تا ۱۴۰۲ به عنوان جامعه آماری در نظر گرفته شدند. برای انتخاب نمونه، از روش حذف سیستماتیک استفاده شد. به عبارت دیگر، شرکت‌هایی که تمامی شرایط مشخص شده را دارا بودند، به عنوان نمونه پژوهش انتخاب شدند. این شرایط شامل عدم عضویت در گروه‌های سرمایه‌گذاری یا واسطه‌گری مالی، دسترسی به اطلاعات مالی کامل در بازه زمانی مورد مطالعه، عدم تغییر سال مالی در طول دوره پژوهش، حضور مستمر در بورس و پایان سال مالی در تاریخ ۲۹ اسفند هر سال بود. در نهایت، ۱۷۱ شرکت به عنوان نمونه نهایی انتخاب شدند. اطلاعات مورد نیاز از پایگاه‌های داده‌ای بورس ویو، ره‌آورد نوین و کدال گردآوری شد.

در گام اول یک مجموعه داده با استفاده داده شرکت‌های بورس تهران ایجاد گردید. سپس در گام دوم به دلیل وجود داده‌های پرت، داده‌های موجود وینسورایز گردید. در گام سوم، چهار روش انتخاب ویژگی و دو روش استخراج ویژگی برای انتخاب روش بهینه اتخاذ شد. در گام چهارم، ده مدل یادگیری ماشین اعمال شد. پنجم مدل‌های برتر در مرحله قبلی با استفاده از روش‌های بوستینگ و بگینگ تقویت شدند. ششم، مدل‌ها از طریق معیار دقت و AUC مورد مقایسه قرار گرفتند. در نهایت بهترین مدل‌ها از نظر عملکرد بر روی پرتفوی سرمایه‌گذاری هلدینگ غدیر آزمایش شد. داده‌های گردآوری‌شده با استفاده از نرم‌افزار یادگیری ماشین رپیدماینر[1] و نرم‌افزار اقتصادسنجی ایویوز[2] مورد تجزیه‌وتحلیل قرار گرفته و همچنین برای مرتب‌سازی و دسته‌بندی داده‌ها از اکسل[3] استفاده شده است. هر مجموعه داده بر اساس روش اعتبارسنجی متقاطع ۱۰ برابری به ۹۰ درصد زیر مجموعه آموزش و ۱۰ درصد آزمایش تقسیم شد. میانگین ۱۰ نتیجه برای مقایسه عملکرد بین رویکردهای مختلف گزارش شده است.

### ۳-۱. متغیر و مدل

در پژوهش حاضر از دادهای مربوط به ۳۲ نسبت مالی که در پژوهش‌های گذشته برای پیش‌بینی مورد استفاده قرارگرفته، استفاده شده است که تعاریف عملیاتی متغیرها در جدول شماره (۱) بیان شده است. مدل‌های

[1] RapidMiner
[2] EViews
[3] Excel

انتخاب شده اگرچه فهرست جامعی نیستند، اما ۱۰ مدل انتخاب‌شده برای این پژوهش به‌طور گسترده‌ای نمایانگر پرکاربردترین و مورد استنادترین مدل‌های طبقه‌بندی در ادبیات هستند (جونز و همکاران، ۲۰۱۵؛ هستی و همکاران[1]، ۲۰۰۹) و تاکنون پژوهشی با این وسعت در حوزه پیش‌بینی عملکرد شرکت انجام نشده است.

جدول (۱) تعاریف عملیاتی متغیرهای پژوهش

| نام متغیر | | تعریف عملیاتی | منبع |
|---|---|---|---|
| نسبت سریع X1 | نسبت‌های نقدینگی | (دارایی‌های جاری – موجودی کالا) ÷ بدهی‌های جاری | فاروق و قمر[2] (۲۰۱۹) |
| نسبت نقدینگی X2 | | دارایی‌های جاری ÷ بدهی‌های جاری | تومپاچ و همکاران[3] (۲۰۲۰) |
| نسبت نقد X3 | | وجه نقد و معادل‌های نقد ÷ بدهی‌های جاری | لیانگ و همکاران[4] (۲۰۱۶) |
| نسبت جریان نقد عملیاتی X4 | | جریان نقد عملیاتی ÷ بدهی جاری | لیانگ و همکاران (۲۰۲۴) |
| نسبت گردش حساب‌های دریافتنی X5 | نسبت‌های بهره‌وری یا گردش دارایی | فروش ÷ حساب‌های دریافتنی | فاروق و قمر (۲۰۱۹) |
| نسبت گردش موجودی کالا X6 | | بهای تمام شده کالای فروش رفته ÷ موجودی کالا | فاروق و قمر (۲۰۱۹) |
| نسبت گردش سرمایه در گردش خالص X7 | | فروش ÷ (دارایی‌های جاری – بدهی‌های جاری) | فاروق و قمر (۲۰۱۹) |
| نسبت گردش دارایی X8 | | فروش ÷ مجموع دارایی‌ها | تانگ و همکاران[5] (۲۰۲۰) |
| نسبت گردش حقوق صاحبان سهام X9 | | فروش ÷ حقوق صاحبان سهام | مانوگنا و میشرا (۲۰۲۲) |
| نسبت گردش دارایی‌های ثابت X10 | | فروش ÷ دارایی‌های ثابت | تانگ و همکاران (۲۰۲۰) |
| نسبت گردش دارایی‌های بلندمدت X11 | | فروش ÷ دارایی‌های بلندمدت | مانوگنا و میشرا (۲۰۲۲) |
| نسبت گردش دارایی‌های جاری X12 | | فروش ÷ دارایی‌های جاری | مانوگنا و میشرا (۲۰۲۲) |
| نسبت حاشیه سود ناخالص X13 | نسبت‌های سودآوری | سود ناخالص ÷ فروش | مانوگنا و میشرا (۲۰۲۲) |
| نسبت حاشیه سود قبل از بهره، مالیات، استهلاک و کاهش ارزش X14 | | سود قبل از بهره، مالیات، استهلاک و کاهش ارزش ÷ فروش | مانوگنا و میشرا (۲۰۲۲) |
| نسبت حاشیه سود خالص X15 | | سود خالص ÷ فروش | تانگ و همکاران (۲۰۲۰) |

[1] Hastie et al
[2] Farooq & Qamar
[3] Tumpach et al
[4] Liang et al
[5] Tang et al

| نام متغیر | | تعریف عملیاتی | منبع |
|---|---|---|---|
| | نسبت سود قبل از مالیات به حقوق صاحبان سهام X16 | سود قبل از مالیات ÷ حقوق صاحبان سهام | مانوگنا و میشرا (۲۰۲۲) |
| | نسبت هزینه‌های عملیاتی به فروش خالص X17 | هزینه‌های عملیاتی ÷ فروش خالص | مانوگنا و میشرا (۲۰۲۲) |
| | سود هر سهم X18 | سودخالص ÷ تعداد سهام عادی | اولاییِنکا[1] (۲۰۲۲) |
| نسبت‌های رشد | نرخ رشد دارایی‌ها X19 | (کل دارایی‌ها$_t$ − کل دارایی‌ها$_{t-1}$) ÷ کل دارایی‌ها$_{t-1}$ | مانوگنا و میشرا (۲۰۲۲) |
| | نرخ رشد سود خالص X20 | (سود خالص$_t$ − سود خالص$_{t-1}$) ÷ سودخالص$_{t-1}$ | مانوگنا و میشرا (۲۰۲۲) |
| | نرخ رشد فروش X21 | (فروش$_t$ − فروش$_{t-1}$) ÷ فروش$_{t-1}$ | تانگ و همکاران (۲۰۲۰) |
| | نرخ رشد خود پایدار X22 | بازده حقوق صاحبان سهام * نرخ انباشت | لیانگ و همکاران (۲۰۲۴) |
| نسبت‌های ساختار دارایی | نسبت دارایی‌های جاری به مجموع دارایی‌ها X23 | دارایی‌های جاری ÷ مجموع دارایی‌ها | دیلین و همکاران[2] (۲۰۱۳) |
| | نسبت موجودی کالا به دارایی‌های جاری X24 | موجودی کالا ÷ دارایی‌های جاری | دیلین و همکاران (۲۰۱۳) |
| | نسبت وجه نقد و معادل‌های نقدی به دارایی‌های جاری X25 | وجه نقد و معادل‌های نقدی ÷ دارایی‌های جاری | مانوگنا و میشرا (۲۰۲۲) |
| | نسبت دارایی‌های بلندمدت به مجموع دارایی‌ها X26 | دارایی‌های بلندمدت ÷ مجموع دارایی‌ها | مانوگنا و میشرا (۲۰۲۲) |
| نسبت‌های توان پرداخت بدهی | نسبت بدهی مالی کوتاه‌مدت به مجموع بدهی‌ها X27 | بدهی مالی کوتاه‌مدت ÷ مجموع بدهی‌ها | مانوگنا و میشرا (۲۰۲۲) |
| | نسبت بدهی کوتاه‌مدت به بدهی کل X28 | بدهی‌های جاری ÷ مجموع بدهی‌ها | مانوگنا و میشرا (۲۰۲۲) |
| | نسبت پوشش بهره X29 | سود قبل از بهره و مالیات ÷ هزینه بهره | تومپاچ و همکاران[3] (۲۰۲۰) |
| | نسبت بدهی X30 | مجموع بدهی‌ها ÷ حقوق صاحبان سهام | تانگ و همکاران (۲۰۲۰) |
| | نسبت اهرم مالی X31 | مجموع بدهی‌ها ÷ مجموع دارایی‌ها | فاروق و قمر (۲۰۱۹) |
| | نسبت کل بدهی مالی به بدهی کل X32 | کل بدهی مالی ÷ مجموع بدهی‌ها | مانوگنا و میشرا (۲۰۲۲) |

منبع: مگارندگان پژوهش

## ۳-۲. معیارهای ارزیابی

در این مطالعه، از معیارهای دقت و AUC برای اندازه‌گیری عملکرد مدل استفاده می‌شود. محاسبه دقت بر اساس یک ماتریس درهم‌ریختگی حاوی طبقه‌بندی مقادیر واقعی و پیش‌بینی شده در دسته‌های، مثبت

[1] Olayinka
[2] Delen et al
[3] Tumpach et al

واقعی، منفی واقعی، مثبت کاذب و منفی کاذب است. همچنین باتوجه‌به ویژگی‌های وظایف طبقه‌بندی، AUC به‌عنوان یک معیار دیگر برای ارزیابی عملکرد مدل‌ها انتخاب شد. ROC منحنی است که به ترتیب با نرخ مثبت کاذب به‌عنوان محور x و نرخ مثبت واقعی به‌عنوان محور y ترسیم شده است، درحالی‌که AUC با جمع‌کردن مساحت زیر منحنی ROC به دست می‌آید (لیانگ و همکاران، ۲۰۲۴).

## ۴. تجزیه و تحلیل داده‌ها و یافته‌های پژوهش

### ۴-۱. آمار توصیفی

یکی از کاربردهای مهم جدول آمار توصیفی بیان نوع توزیع داده‌ها است. از آن جایی که متغیر هدف پژوهش یعنی بازده دارایی‌ها یک متغیر باینری است و متغیرهای مستقل کمی پیوسته هستند، آمار توصیفی در دو جدول مجزا شماره (۲) و (۳) ارائه شده است. در این پژوهش، برای کاهش اثر نقاط پرت بر نتایج تحلیل، از روش وینسورایز[1] (در سطح ۱ و ۹۹ درصد) استفاده شد.

جدول شماره (۲) آمار توصیفی متغیرهای پژوهش

| نام متغیر | کشیدگی | چولگی | انحراف معیار | کمینه | بیشینه | میانه | میانگین | آماره جارک برا |
|---|---|---|---|---|---|---|---|---|
| X1 | ۳/۸۲۲ | ۱/۱۳۵ | ۰/۵۳۵ | ۰/۲۸۹ | ۲/۳۷۸ | ۰/۸۵۵ | ۰/۹۶۱ | ۵۳۹/۵۶۶ |
| X2 | ۳/۸۳۶ | ۱/۱۷۳ | ۰/۷۲۲ | ۰/۵۹۹ | ۳/۴۳۷ | ۱/۳۵۵ | ۱/۵۳۵ | ۵۷۴/۲۹۵ |
| X3 | ۵/۳۶۷ | ۱/۸۱۵ | ۰/۲۲۹ | ۰/۰۰۹ | ۰/۸۶۳ | ۰/۰۸۹ | ۰/۱۸۶ | ۱۷۳۹/۰۵۱ |
| X4 | ۴/۱۱۸ | ۱/۳۴۲ | ۰/۴۰۸ | -۰/۱۵۴ | ۱/۴۳۷ | ۰/۲۰۸ | ۰/۳۳۷ | ۷۸۲/۵۵۵ |
| X5 | ۲/۱۸۷ | ۰/۷۸۷ | ۳/۱۸۸ | ۱/۳۷۶ | ۱۰/۳۸۶ | ۳/۴۵۰ | ۴/۵۹۳ | ۲۹۰/۶۷۷ |
| X6 | ۱/۹۳۸ | ۰/۶۶۷ | ۱/۸۹۱ | ۱/۴۲۲ | ۶/۶۲۱ | ۲/۶۱۸ | ۳/۳۷۵ | ۲۶۹/۴۱۶ |
| X7 | ۲/۰۸۲ | ۰/۱۸۰ | ۴/۱۱۲ | -۲/۶۸۹ | ۱۰/۰۳۶ | ۲/۸۵۷ | ۳/۴۰۳ | ۸۹/۹۸۴ |
| X8 | ۳/۳۹۱ | ۱/۰۲۰ | ۰/۵۲۲ | ۰/۲۸۰ | ۲/۲۶۷ | ۰/۸۲۰ | ۰/۹۵۰ | ۳۹۹/۶۵۵ |
| X9 | ۲/۱۵۴ | ۰/۷۴۲ | ۱/۴۴۲ | ۰/۸۶۲ | ۴/۹۷۱ | ۱/۸۷۷ | ۲/۳۷۴ | ۲۷۰/۰۴۹ |
| X10 | ۲/۰۳۳ | ۰/۶۷۲ | ۴/۱۴۸ | ۱/۲۵۱ | ۱۲/۹۳۶ | ۴/۲۶۱ | ۵/۶۳۶ | ۲۵۳/۹۹۴ |
| X11 | ۱/۹۱۷ | ۰/۶۲۲ | ۲/۷۵۴ | ۰/۹۸۱ | ۸/۶۱۴ | ۳/۰۳۳ | ۳/۹۱۴ | ۲۵۱/۹۴۳ |
| X12 | ۲/۹۶۰ | ۰/۸۳۱ | ۰/۷۶۶ | ۰/۴۸۴ | ۳/۳۱۹ | ۱/۳۳۹ | ۱/۴۹۸ | ۲۵۶/۲۱۸ |
| X13 | ۲/۰۹۲ | ۰/۳۲۷ | ۰/۱۵۶ | ۰/۰۲۶ | ۰/۵۷۰ | ۰/۲۴۷ | ۰/۲۶۶ | ۱۱۵/۸۸۳ |
| X14 | ۲/۲۸۰ | ۰/۳۶۲ | ۰/۱۶۸ | -۰/۰۹۲ | ۰/۵۲۹ | ۰/۱۶۳ | ۰/۱۹۳ | ۹۶/۷۵۰ |

[1] Winsorizing

| نام متغیر | کشیدگی | چولگی | انحراف معیار | کمینه | بیشینه | میانه | میانگین | آماره جارک برا |
|---|---|---|---|---|---|---|---|---|
| X15 | ۲/۶۳۹ | ۰/۵۶۹ | ۰/۱۷۸ | -۰/۱۲۳ | ۰/۵۶۵ | ۰/۱۳۶ | ۰/۱۷۲ | ۱۳۲/۱۷۹ |
| X16 | ۲/۱۲۵ | ۰/۰۴۶ | ۰/۲۹۰ | -۰/۱۷۰ | ۰/۸۸۳ | ۰/۳۵۶ | ۰/۳۶۳ | ۷۱/۷۴۵ |
| X17 | ۲/۱۷۳ | -۰/۳۵۲ | ۰/۱۵۷ | ۰/۵۰۱ | ۱/۰۶۴ | ۰/۸۳۱ | ۰/۸۰۸ | ۱۰۹/۳۰۸ |
| X18 | ۴/۷۶۶ | ۱/۵۸۱ | ۱۳۲۴/۰۵۳ | -۲۳۹/۲۰۸ | ۴۸۵۴/۰۴۹ | ۵۶۲/۲۰۹ | ۱۰۵۶/۹۸۹ | ۱۲۱۴/۷۸۲ |
| X19 | ۳/۴۸۰ | ۱/۰۸۳ | ۰/۳۲۳ | -۰/۰۷۸ | ۱/۱۳۶ | ۰/۲۲۳ | ۰/۳۱۱ | ۴۵۶/۲۸۰ |
| X20 | ۱/۹۴۶ | ۰/۴۵۰ | ۰/۸۲۸ | -۰/۵۵۳ | ۱/۸۸۳ | ۰/۳۱۴ | ۰/۴۸۸ | ۱۷۷/۷۹۹ |
| X21 | ۲/۶۴۹ | ۰/۵۳۲ | ۰/۳۸۸ | -۰/۲۵۲ | ۱/۲۱۰ | ۰/۳۰۴ | ۰/۳۴۹ | ۱۱۶/۳۲۵ |
| X22 | ۳/۲۲۹ | ۰/۳۳۴ | ۰/۱۴۵ | -۰/۲۰۴ | ۰/۴۲۴ | ۰/۰۷۲ | ۰/۰۹۹ | ۴۶/۰۷۴ |
| X23 | ۲/۰۶۷ | -۰/۳۸۹ | ۰/۱۸۹ | ۰/۲۸۹ | ۰/۹۳۲ | ۰/۶۸۹ | ۰/۶۶۱ | ۱۳۶/۸۴۴ |
| X24 | ۲/۰۷۸ | ۰/۲۴۲ | ۰/۱۶۹ | ۰/۱۰۳ | ۰/۶۹۱ | ۰/۳۵۵ | ۰/۳۷۱ | ۱۰۰/۴۴۱ |
| X25 | ۳/۸۹۷ | ۱/۳۵۷ | ۰/۱۰۵ | ۰/۰۰۹ | ۰/۳۸۴ | ۰/۰۶۹ | ۰/۱۰۹ | ۷۵۶/۶۵۶ |
| X26 | ۲/۰۶۷ | ۰/۳۸۹ | ۰/۱۸۹ | ۰/۰۶۸ | ۰/۷۱۱ | ۰/۳۱۱ | ۰/۳۳۹ | ۱۳۶/۸۴۴ |
| X27 | ۱/۹۸۶ | ۰/۳۹۵ | ۰/۲۰۰ | ۰/۰۰۰ | ۰/۶۴۱ | ۰/۲۳۲ | ۰/۲۵۹ | ۱۵۲/۸۳۱ |
| X28 | ۳/۵۵۲ | -۱/۲۰۱ | ۰/۱۰۴ | ۰/۶۱۴ | ۰/۹۹۱ | ۰/۹۲۱ | ۰/۸۸۴ | ۵۶۲/۵۵۶ |
| X29 | ۳/۰۴۶ | ۱/۳۰۹ | ۲۸/۰۲۹ | ۰/۸۷۱ | ۷۸/۳۶۱ | ۵/۷۳۱ | ۲۰/۷۵۴ | ۶۳۵/۳۲۶ |
| X30 | ۱/۹۲۴ | ۰/۵۹۲ | ۰/۹۷۹ | ۰/۴۴۹ | ۳/۱۹۸ | ۱/۲۱۴ | ۱/۵۲۷ | ۲۳۷/۰۷۵ |
| X31 | ۲/۰۸۶ | -۰/۰۷۳ | ۰/۱۹۵ | ۰/۲۱۱ | ۰/۸۹۶ | ۰/۵۵۹ | ۰/۵۶۰ | ۷۹/۲۸۴ |
| X32 | ۱/۸۳۴ | ۰/۲۲۶ | ۰/۲۲۰ | ۰/۰۰۰ | ۰/۶۹۴ | ۰/۲۸۵ | ۰/۳۰۴ | ۱۴۴/۸۷۲ |

منبع: یافته‌های پژوهش

جدول شماره (۳) آمار توصیفی متغیر وابسته پژوهش

| نام متغیر | مقدار | تعداد | درصد |
|---|---|---|---|
| بازده دارایی | ۰ | ۱۰۲۶ | ۵۰ |
| بازده دارایی | ۱ | ۱۰۲۶ | ۵۰ |

منبع: یافته‌های پژوهش

بر اساس تحلیل جدول شماره (۲) و نتایج آزمون جارک-برا، می‌توان نتیجه گرفت که توزیع بسیاری از نسبت‌های مالی مورد مطالعه، از توزیع نرمال فاصله دارد. وجود چولگی مثبت و کشیدگی بیش از حد در این متغیرها، این عدم نرمال بودن را تایید می‌کند. این یافته در مطالعات مالی، امری متداول است. میانگین حاشیه

سود ناخالص ۰/۲۶۶ است، در حالی که حاشیه سود خالص ۰/۱۷۲ است. این نشان می‌دهد که شرکت‌ها به طور متوسط ۱۷/۲ درصد از فروش خود را پس از تمام هزینه‌ها به عنوان سود خالص حفظ می‌کنند. حاشیه سود قبل از بهره، مالیات و استهلاک میانگین ۰/۱۹۳ را نشان می‌دهد که بینشی از سودآوری عملیاتی قبل از تأثیر تصمیمات مالی و حسابداری ارائه می‌دهد. میانگین نسبت سریع ۰/۹۶۱ است، که نشان می‌دهد، به طور متوسط، شرکت‌ها می‌توانند ۹۶/۱ درصد از بدهی‌های جاری خود را با نقدشوندگی‌ترین دارایی‌های خود پوشش دهند. نسبت نقدینگی میانگین ۱/۵۳۵ را نشان می‌دهد که نشان می‌دهد شرکت‌ها عموما دارایی‌های جاری کافی برای پوشش بدهی‌های جاری خود دارند. با این حال، نسبت نقدی دارای میانگین پایین‌تر ۰/۱۸۶ است، که نشان می‌دهد شرکت‌ها در صورت نیاز به پرداخت فوری تمام بدهی‌های جاری تنها با استفاده از پول نقد و معادل‌های نقدی، ممکن است با چالش‌هایی مواجه شوند.

جدول شماره (۳) نشان می‌دهد که بازده دارایی‌ها به طور مساوی بین دو دسته توزیع می‌شود، به طوری که ۵۰ درصد مشاهدات بالاتر از میانه و ۵۰ درصد کمتر هستند. این توزیع متعادل تضمین می‌کند که مدل‌های پیش‌بینی ما نسبت به یک کلاس تعصب نخواهند داشت.

## ۴-۲. برآورد مدل‌های پژوهش

جدول شماره (۴) نتایج مدل‌های یادگیری ماشین را بدون استفاده از هیچ روش انتخاب ویژگی ارائه می‌دهد. این مقایسه پایه، بینش‌های ارزشمندی را در مورد قدرت پیش‌بینی ذاتی الگوریتم‌های مختلف در مواجهه با مجموعه کامل ویژگی‌ها فراهم می‌کند. نتایج جدول شماره (۴) نشان می‌دهد که ماشین‌های بردار پشتیبان بالاترین دقت (۸۰/۰۵ درصد) را نشان دادند، در حالی که جنگل تصادفی بهترین AUC (۰/۸۸۳) را به دست آورد. عملکرد قوی این مدل‌ها نشان می‌دهد که آنها به‌ویژه برای درک روابط پیچیده و بالقوه غیرخطی در داده‌های بازار سهام مناسب هستند. جالب توجه است که مدل‌های ساده‌تر مانند رگرسیون لجستیک به صورت رقابتی (دقت ۷۸/۸۶ درصد، AUC ۰/۷۸۴) عمل کردند، که نشان می‌دهد رابطه بین ویژگی‌های پیش‌بینی‌کننده و متغیر هدف ممکن است دارای مولفه‌های خطی باشد. این یافته بر اهمیت نادیده گرفتن مدل‌های ساده‌تر به نفع مدل‌های پیچیده‌تر تأکید می‌کند، زیرا گاهی اوقات می‌توانند جنبه‌های مهم داده‌ها را به تصویر بکشند. شبکه‌های عصبی، که اغلب به دلیل توانایی‌شان در مدل‌سازی روابط پیچیده تبلیغ می‌شوند، عملکرد خوب اما نه برتر را نشان دادند (دقت ۷۶/۵۲ درصد، AUC ۰/۸۵۱). این ممکن است نشان دهد که معماری پیش فرض استفاده شده برای این مشکل خاص بهینه نبوده است، یا اینکه مجموعه ویژگی خام حاوی نویز است که بر یادگیری شبکه تأثیر می‌گذارد. عملکرد مدل‌های مبتنی بر درخت (درخت تصمیم، جنگل تصادفی، درختان تقویت کننده گرادیان، تقویت گرادیان شدید) به طور مداوم با دقت‌های ۷۸/۱۰ درصد تا ۷۸/۹۱ درصد و AUC ها از ۰/۷۹۳ تا ۰/۸۸۳ قوی بود. این نشان می‌دهد که داده‌ها ممکن است ساختارهای سلسله مراتبی یا مبتنی بر تصمیم ذاتی داشته باشند که این مدل‌ها در درک آن‌ها خوب هستند.

جدول شماره (۴) عملکرد طبقه‌بندی کننده‌ها بدون استفاده از تکنیک‌های انتخاب ویژگی

| مدل | دقت | رتبه دقت | AUC | رتبه AUC |
|---|---|---|---|---|
| شبکه عصبی | ۷۶/۵۲% | ۹ | ۰/۸۵۱ | ۷ |
| درخت تصمیم | ۷۸/۱۰% | ۷ | ۰/۷۹۳ | ۸ |
| رگرسیون لجستیک | ۷۸/۸۶% | ۳ | ۰/۷۸۴ | ۱۰ |
| ماشین بردار پشتیبان | ۸۰/۰۵% | ۱ | ۰/۸۷۹ | ۲ |
| یادگیری عمیق | ۷۷/۷۷% | ۸ | ۰/۸۶۷ | ۵ |
| جنگل تصادفی | ۷۸/۵۹% | ۵ | ۰/۸۸۳ | ۱ |
| تقویت گرادیان درختی | ۷۸/۹۱% | ۲ | ۰/۸۷۴ | ۳ |
| بیز ساده | ۷۸/۱۵% | ۶ | ۰/۸۵۹ | ۶ |
| کا نزدیک‌ترین همسایه | ۷۲/۵۵% | ۱۰ | ۰/۷۹۲ | ۹ |
| تقویت گرادیان شدید | ۷۸/۸۰% | ۴ | ۰/۸۷۲ | ۴ |

منبع: یافته‌های پژوهش

نتایج جدول شماره (۵) نشان می‌دهد که انتخاب روبه‌جلو به طور قابل توجهی عملکرد همه مدل‌ها را بهبود بخشید و شبکه‌های عصبی بیشترین پیشرفت را نشان دادند (دقت ۸۲/۶۶ درصد، AUC ۰/۹۰۱). این پیشرفت چشمگیر نشان می‌دهد که برخی از ویژگی‌های مجموعه داده اصلی ممکن است اضافی یا دارای نویز بوده باشند، و شبکه‌های عصبی به ویژه به کیفیت ویژگی حساس هستند. بهبود مداوم در تمام مدل‌ها (از ۲/۳۳ تا ۷/۶۱ درصد در دقت) اهمیت انتخاب ویژگی در وظایف پیش‌بینی مالی را نشان می‌دهد.

جدول شماره (۵) نتایج طبقه‌بندی کننده‌ها با استفاده از تکنیک انتخاب ویژگی انتخاب روبه‌جلو

| مدل | دقت | رتبه دقت | AUC | رتبه AUC |
|---|---|---|---|---|
| شبکه عصبی | ۸۲/۶۶% | ۱ | ۰/۹۰۱ | ۱ |
| درخت تصمیم | ۸۰/۴۳% | ۹ | ۰/۸۱۶ | ۱۰ |
| رگرسیون لجستیک | ۸۱/۴۷% | ۵ | ۰/۸۹ | ۴ |
| ماشین بردار پشتیبان | ۸۱/۶۸% | ۴ | ۰/۸۷۴ | ۷ |
| یادگیری عمیق | ۸۱/۸۵% | ۳ | ۰/۸۸۶ | ۵ |
| جنگل تصادفی | ۸۲/۲۳% | ۲ | ۰/۸۹۲ | ۲ |
| تقویت گرادیان درختی | ۸۰/۸۷% | ۷ | ۰/۸۹۱ | ۳ |
| بیز ساده | ۸۱/۴۱% | ۶ | ۰/۸۷۳ | ۸ |
| کا نزدیک‌ترین همسایه | ۸۰/۱۶% | ۱۰ | ۰/۸۴۵ | ۹ |
| تقویت گرادیان شدید | ۸۰/۵۴% | ۸ | ۰/۸۷۹ | ۶ |

منبع: یافته‌های پژوهش

نتایج جدول شماره (۶) نشان می‌دهد که حذف روبه‌عقب همچنین عملکرد مدل را افزایش داد که ماشین بردار پشتیبان و جنگل تصادفی به ترتیب بالاترین دقت (۸۲/۶۶ درصد) و AUC (۰/۹۰۵) را به دست آوردند. بهبودها عموماً کوچکتر از مواردی بود که با انتخاب روبه‌جلو مشاهده شد که ممکن است نشان دهد که حذف

به عقب در حذف ویژگی محافظه کارانه تر است. به نظر می‌رسد که این روش به طور بالقوه با حذف ویژگی‌هایی که باعث بیش برازش می‌شدند، موجب عملکرد بهتر ماشین بردار پشتیبان شده باشد.

جدول شماره (۶) نتایج طبقه‌بندی کننده‌ها با استفاده از تکنیک انتخاب ویژگی حذف روبه‌عقب

| مدل | دقت | رتبه دقت | AUC | رتبه AUC |
|---|---|---|---|---|
| شبکه عصبی | ۸۰/۴۳% | ۸ | ۰/۸۸۱ | ۷ |
| درخت تصمیم | ۷۹/۶۷% | ۹ | ۰/۸۱۵ | ۱۰ |
| رگرسیون لجستیک | ۸۲/۱۱% | ۴ | ۰/۸۹۴ | ۵ |
| ماشین بردار پشتیبان | ۸۲/۶۶% | ۱ | ۰/۸۹۹ | ۳ |
| یادگیری عمیق | ۸۲/۰۷% | ۵ | ۰/۹ | ۲ |
| جنگل تصادفی | ۸۲/۳۹% | ۲ | ۰/۹۰۵ | ۱ |
| تقویت گرادیان درختی | ۸۲/۱۲% | ۳ | ۰/۸۹۷ | ۴ |
| بیز ساده | ۸۰/۹۲% | ۷ | ۰/۸۷۸ | ۸ |
| کا نزدیک‌ترین همسایه | ۷۹/۰۸% | ۱۰ | ۰/۸۳۱ | ۹ |
| تقویت گرادیان شدید | ۸۱/۷۹% | ۶ | ۰/۸۹۳ | ۶ |

منبع: یافته‌های پژوهش

نتایج جدول شماره (۷) نشان می‌دهد که روش هدایت وزنی پیشرفت‌های کمتری را نشان داد، اما همچنان عملکرد را برای اکثر مدل‌ها در مقایسه با عدم انتخاب ویژگی افزایش داد. جنگل تصادفی عملکرد قوی خود را حفظ کرد (دقت ۸۲/۲۳ درصد، ۰/۸۹۳ AUC). دستاوردهای کمتر در اینجا احتمالاً به دلیل ایجاد تعادلی بین کاهش ویژگی و حفظ اطلاعات توسط این روش است.

جدول شماره (۷) نتایج طبقه‌بندی کننده‌ها با استفاده از تکنیک انتخاب ویژگی هدایت وزن‌دار

| مدل | دقت | رتبه دقت | AUC | رتبه AUC |
|---|---|---|---|---|
| شبکه عصبی | ۷۹/۶۷% | ۶ | ۰/۸۶۵ | ۶ |
| درخت تصمیم | ۷۹/۳۵% | ۷ | ۰/۸۳۵ | ۹ |
| رگرسیون لجستیک | ۸۰/۲۶% | ۵ | ۰/۸۷۴ | ۴ |
| ماشین بردار پشتیبان | ۸۰/۷۱% | ۲ | ۰/۸۸۱ | ۲ |
| یادگیری عمیق | ۷۸/۷۰% | ۸ | ۰/۸۵۹ | ۷ |
| جنگل تصادفی | ۸۲/۲۳% | ۱ | ۰/۸۹۳ | ۱ |
| تقویت گرادیان درختی | ۸۰/۲۷% | ۴ | ۰/۸۷۳ | ۵ |
| بیز ساده | ۷۷/۶۱% | ۱۰ | ۰/۸۵۸ | ۸ |
| کا نزدیک‌ترین همسایه | ۷۷/۷۲% | ۹ | ۰/۸۲۷ | ۱۰ |
| تقویت گرادیان شدید | ۸۰/۳۳% | ۳ | ۰/۸۷۵ | ۳ |

منبع: یافته‌های پژوهش

نتایج جدول شماره (۸) نشان می‌دهد که الگوریتم ژنتیک به طور ویژه موثر بود و بالاترین دقت کلی را با جنگل تصادفی (۸۳/۲۱ درصد) و بالاترین AUC با ماشین بردار پشتیبان (۰/۹۰۷) را به همراه داشت. این

نشان می‌دهد که الگوریتم ژنتیک قادر به شناسایی زیرمجموعه‌های بسیار آموزنده از ویژگی‌ها بوده و به طور بالقوه تعاملات پیچیده‌ای را که سایر روش‌ها از دست داده‌اند، ثبت می‌کند. عملکرد قوی در مدل‌های مختلف نشان می‌دهد که این زیرمجموعه از ویژگی نه فقط برای یک الگوریتم خاص بلکه برای همه الگوریتم‌ها به خوبی پیش‌بینی‌کننده بود.

جدول شماره (۸) نتایج طبقه‌بندی کننده‌ها با استفاده از تکنیک انتخاب ویژگی الگوریتم ژنتیک

| مدل | دقت | رتبه دقت | AUC | رتبه AUC |
|---|---|---|---|---|
| شبکه عصبی | ۸۲/۱۲% | ۵ | ۰/۸۹۴ | ۷ |
| درخت تصمیم | ۸۰/۳۸% | ۹ | ۰/۸۳۲ | ۱۰ |
| رگرسیون لجستیک | ۸۲/۹۹% | ۲ | ۰/۸۹۹ | ۳ |
| ماشین بردار پشتیبان | ۸۱/۹۶% | ۸ | ۰/۹۰۷ | ۱ |
| یادگیری عمیق | ۸۲/۵۵% | ۳ | ۰/۸۹۸ | ۴ |
| جنگل تصادفی | ۸۳/۲۱% | ۱ | ۰/۹ | ۲ |
| تقویت گرادیان درختی | ۸۲/۱۷% | ۴ | ۰/۸۹۵ | ۶ |
| بیز ساده | ۸۲/۰۷% | ۶ | ۰/۸۷۶ | ۸ |
| کا نزدیک‌ترین همسایه | ۷۹/۱۳% | ۱۰ | ۰/۸۳۹ | ۹ |
| تقویت گرادیان شدید | ۸۲/۰۱% | ۷ | ۰/۸۹۶ | ۵ |

منبع: یافته‌های پژوهش

قابل‌توجه است که تجزیه‌وتحلیل مؤلفه اصلی جدول شماره (۹) و نقشه‌های خودسازمان دهنده جدول شماره (۱۰) منجر به کاهش عملکرد در همه مدل‌ها شد. این یک یافته بسیار مهم است، زیرا برخلاف این فرض رایج است که تکنیک‌های کاهش ابعاد همیشه عملکرد مدل را بهبود می‌بخشد. برای تحلیل مؤلفه اصلی، افت دقت بین ۴/۷۸ تا ۸/۹۱ درصد بود، در حالی که برای نقشه‌های خودسازمان دهنده، افت بین ۳/۰۵ تا ۹/۶۲ درصد بود. کاهش عملکرد ممکن است نشان دهنده این باشد که این تکنیک‌های کاهش ابعاد، اطلاعات مهمی را در فرآیند ایجاد ویژگی‌های جدید از دست داده‌اند، یا اینکه ویژگی‌های اصلی از قبل برای وظیفه پیش‌بینی بهینه بوده‌اند. ممکن است روابط غیرخطی در داده‌ها توسط این روش‌ها به خوبی درک نشده باشند، یا اینکه ویژگی‌های جدید ایجاد شده برای مدل‌ها کمتر قابل تفسیر بوده‌اند. این بر اهمیت دقت در در نظر گرفتن ماهیت داده‌ها و وظیفه پیش‌بینی هنگام انتخاب روش‌های مهندسی ویژگی تأکید می‌کند.

جدول شماره (۹) نتایج طبقه‌بندی کننده‌ها با استفاده از تکنیک استخراج ویژگی تحلیل مولفه‌های اصلی

| مدل | دقت | رتبه دقت | AUC | رتبه AUC |
|---|---|---|---|---|
| شبکه عصبی | ۷۵/۲۷% | ۱ | ۰/۸۳ | ۱ |
| درخت تصمیم | ۷۳/۶۴% | ۲ | ۰/۷۶۷ | ۹ |
| رگرسیون لجستیک | ۷۲/۶۱% | ۷ | ۰/۸۲۱ | ۴ |
| ماشین بردار پشتیبان | ۷۲/۴۵% | ۸ | ۰/۸۲۲ | ۳ |
| یادگیری عمیق | ۷۳/۴۲% | ۳ | ۰/۸۲ | ۵ |
| جنگل تصادفی | ۷۲/۹۹% | ۵ | ۰/۸۱۱ | ۶ |

| مدل | دقت | رتبه دقت | AUC | رتبه AUC |
|---|---|---|---|---|
| تقویت گرادیان درختی | ۷۲/۸۸% | ۶ | ۰/۸۱ | ۷ |
| بیز ساده | ۶۹/۹۵% | ۱۰ | ۰/۸۲۹ | ۲ |
| کا نزدیک‌ترین همسایه | ۷۰/۸۷% | ۹ | ۰/۷۶۳ | ۱۰ |
| تقویت گرادیان شدید | ۷۳/۳۲% | ۴ | ۰/۸۰۲ | ۸ |

منبع: یافته‌های پژوهش

جدول شماره (۱۰) نتایج طبقه‌بندی کننده‌ها با استفاده از تکنیک استخراج ویژگی نقشه خود سازمان دهنده

| مدل | دقت | رتبه دقت | AUC | رتبه AUC |
|---|---|---|---|---|
| شبکه عصبی | ۰/۷۳۹۷ | ۴ | ۰/۷۹ | ۶ |
| درخت تصمیم | ۰/۷۵۰۵ | ۲ | ۰/۸۱ | ۳ |
| رگرسیون لجستیک | ۰/۶۹۲۴ | ۱۰ | ۰/۷۳۴ | ۱۰ |
| ماشین بردار پشتیبان | ۰/۶۹۷۸ | ۹ | ۰/۷۴ | ۹ |
| یادگیری عمیق | ۰/۷۲۰۱ | ۶ | ۰/۷۹۹ | ۵ |
| جنگل تصادفی | ۰/۷۴۲۹ | ۳ | ۰/۸۰۹ | ۴ |
| تقویت گرادیان درختی | ۰/۷۳۷ | ۵ | ۰/۸۱۲ | ۲ |
| بیز ساده | ۰/۷۰۸۲ | ۸ | ۰/۷۶۵ | ۸ |
| کا نزدیک‌ترین همسایه | ۰/۷۱۶۸ | ۷ | ۰/۷۶۹ | ۷ |
| تقویت گرادیان شدید | ۰/۷۵۶۵ | ۱ | ۰/۸۱۹ | ۱ |

منبع: یافته‌های پژوهش

جدول شماره (۱۱) نشان می‌دهد که ترکیب شبکه‌های عصبی با انتخاب پیشرو و بگینگ بیشترین دقت (۸۳/۴۲ درصد) را به دست آورد، در حالی که جنگل تصادفی با الگوریتم ژنتیک و بگینگ بالاترین AUC (۰/۹۰۷) را به دست آورد. این نتایج نشان می‌دهد که روش‌های گروهی می‌توانند قدرت پیش‌بینی مدل‌های از قبل قوی را بیشتر افزایش دهند، به ویژه هنگامی که با انتخاب ویژگی مؤثر ترکیب شوند.

به نظر می‌رسد که بگینگ در این زمینه عموماً مؤثرتر از بوستینگ باشد. به‌عنوان‌مثال، ماشین بردار پشتیبان با حذف روبه‌عقب AUC برابر ۰/۹ با بگینگ در مقایسه با ۰/۸۶۲ با بوستینگ را نشان داد. این ممکن است نشان دهد که کاهش واریانس مدل سودمندتر از تمرکز بر کاهش سوگیری برای این کار پیش‌بینی خاص بود. عملکرد قوی مدل‌های بگینگ، به‌ویژه هنگامی که با انتخاب ویژگی مؤثر ترکیب می‌شود، نشان می‌دهد که قدرت پیش‌بینی قابل‌توجهی در زیرنمونه‌گیری نمونه‌های داده و فضای ویژگی وجود دارد. این می‌تواند به‌ویژه در زمینه بازار سهام مرتبط باشد، جایی که زیرمجموعه‌های مختلف ویژگی‌ها ممکن است در زمان‌های مختلف یا برای انواع مختلف شرکت‌ها بیشتر یا کمتر مرتبط باشند.

جدول شماره (۱۱) نتایج ترکیب مدل هایب با عملکرد برتر با روش‌های بوستینگ و بگینگ

| مدل | دقت | رتبه دقت | AUC | رتبه AUC |
|---|---|---|---|---|
| ماشین بردار پشتیبان + بگینگ | ۷۹/۵۷ | ۱۳ | ۰/۸۷۶ | ۸ |

| مدل | دقت | رتبه دقت | AUC | رتبه AUC |
|---|---|---|---|---|
| جنگل تصادفی + بگینگ | ۷۹/۰۸ | ۱۴ | ۰/۸۸۶ | ۷ |
| ماشین بردار پشتیبان + بوستینگ | ۷۹/۷۸ | ۱۲ | ۰/۸۳۷ | ۱۴ |
| جنگل تصادفی + بوستینگ | ۷۸/۹۷ | ۱۵ | ۰/۷۹۷ | ۱۷ |
| شبکه عصبی +انتخاب روبه‌جلو + بگینگ | ۸۳/۴۲ | ۱ | ۰/۸۹۲ | ۵ |
| شبکه عصبی +انتخاب روبه‌جلو + بوستینگ | ۸۲/۲۸ | ۹ | ۰/۸۶۵ | ۱۰ |
| ماشین بردار پشتیبان + حذف روبه‌عقب + بگینگ | ۸۲/۶۱ | ۷ | ۰/۹ | ۴ |
| جنگل تصادفی + حذف روبه‌عقب + بگینگ | ۸۲/۵ | ۸ | ۰/۹۰۳ | ۲ |
| ماشین بردار پشتیبان + حذف روبه‌عقب + بوستینگ | ۸۲/۲۸ | ۹ | ۰/۸۶۲ | ۱۱ |
| جنگل تصادفی + حذف روبه‌عقب + بوستینگ | ۸۲/۶۶ | ۶ | ۰/۸۳۹ | ۱۳ |
| جنگل تصادفی + هدایت وزن‌دار + بگینگ | ۷۸/۰۴ | ۱۶ | ۰/۸۷۵ | ۹ |
| جنگل تصادفی + هدایت وزن‌دار + بوستینگ | ۸۱/۳ | ۱۱ | ۰/۸۸۹ | ۶ |
| جنگل تصادفی + الگوریتم ژنتیک + بگینگ | ۸۳/۱ | ۴ | ۰/۹۰۷ | ۱ |
| ماشین بردار پشتیبان + الگوریتم ژنتیک + بگینگ | ۸۳/۳۲ | ۳ | ۰/۹۰۱ | ۳ |
| جنگل تصادفی + الگوریتم ژنتیک + بوستینگ | ۸۲/۸۸ | ۵ | ۰/۸۳۴ | ۱۵ |
| ماشین بردار پشتیبان + الگوریتم ژنتیک + بوستینگ | ۸۳/۳۷ | ۲ | ۰/۸۵۳ | ۱۲ |
| شبکه عصبی + تحلیل مؤلفه اصلی + بگینگ | ۷۳/۳۷ | ۱۸ | ۰/۸۲۲ | ۱۶ |
| شبکه عصبی + تحلیل مولفه اصلی + بوستینگ | ۷۴/۶۲ | ۱۷ | ۰/۷۸۶ | ۱۸ |

منبع: یافته‌های پژوهش

جدول شماره (۱۲) نشان می‌دهد که ارزیابی نهایی پرتفوی هلدینگ سرمایه گذاری غدیر تعمیم عالی مدل‌های با عملکرد برتر را نشان داد. ماشین‌های بردار پشتیبان با انتخاب ویژگی الگوریتم ژنتیک بالاترین دقت (۹۵/۲۶ درصد) را به دست آوردند، در حالی که جنگل تصادفی با الگوریتم ژنتیک و بگینگ بالاترین AUC (۰/۹۴۵) را به دست آورد. این نتایج با توجه به اینکه عملکرد خارج از نمونه را در یک سبد سرمایه گذاری خاص نشان می‌دهند، بسیار چشمگیر هستند. دقت بالا و مقادیر AUC نشان می‌دهد که مدل‌ها، نه تنها ویژگی‌های خاص داده‌های آموزشی بلکه الگوهای تعمیم‌پذیری را در پیش‌بینی عملکرد ثبت کرده‌اند. عملکرد قوی ماشین بردار پشتیبان و جنگل تصادفی، هر دو با الگوریتم ژنتیک برای انتخاب ویژگی، اثربخشی این روش انتخاب ویژگی را تقویت می‌کند. این نشان می‌دهد که الگوریتم ژنتیک توانسته است مجموعه‌ای از ویژگی‌ها را شناسایی کند که به طور قوی برای زیرمجموعه‌های مختلف شرکت‌ها پیش‌بینی‌کننده هستند. این واقعیت که این مدل‌ها به خوبی به بافت خاص هلدینگ سرمایه گذاری غدیر منتقل می‌شوند، پیامدهای عملی قابل توجهی دارد. این نشان می‌دهد که شرکت‌های سرمایه‌گذاری به طور بالقوه می‌توانند از رویکردهای مدل‌سازی مشابه برای به دست آوردن بینشی در مورد سبد دارایی خاص خود استفاده کنند.

جدول شماره (۱۲) نتایج آزمایش طبقه‌بندی کننده‌های برتر بر روی پرتفوی هلدینگ سرمایه گذاری غدیر

| مدل | دقت | رتبه دقت | AUC | رتبه AUC |
|---|---|---|---|---|
| شبکه عصبی + انتخاب روبه‌جلو + بگینگ | ۹۳/۵۶ | ۳ | ۰/۸۹۱ | ۳ |

| مدل | دقت | رتبه دقت | AUC | رتبه AUC |
|---|---|---|---|---|
| جنگل تصادفی + الگوریتم ژنتیک + بگینگ | ۹۳/۹۷ | ۲ | ۰/۹۴۵ | ۱ |
| ماشین بردار پشتیبان + الگوریتم ژنتیک | ۹۵/۲۶ | ۱ | ۰/۹۴۱ | ۲ |

منبع: یافته‌های پژوهش

## ۵. نتیجه‌گیری و پیشنهادها

این پژوهش با بررسی کاملی در مورد کاربرد تکنیک‌های یادگیری ماشین برای پیش‌بینی عملکرد شرکت در بورس اوراق بهادار تهران انجام شد. از طریق ارزیابی سیستماتیک مدل‌های مختلف، روش‌های انتخاب ویژگی و تکنیک‌های گروهی، پژوهش حاضر چندین بینش کلیدی را ارائه کرد که به درک نظری و کاربرد عملی یادگیری ماشین در بخش مالی کمک می‌کند.

یافته‌های این پژوهش بر نقش حیاتی انتخاب ویژگی در افزایش عملکرد پیش‌بینی تاکید می‌کند. الگوریتم ژنتیک و روش‌های انتخاب رو به جلو به طور مداوم دقت مدل و AUC را در الگوریتم‌های مختلف بهبود می‌بخشد و اهمیت شناسایی مرتبط‌ترین ویژگی‌ها را برای وظایف پیش‌بینی مالی برجسته می‌کند. این امر بر نیاز به مهندسی ویژگی‌های دقیق در مطالعات آینده و کاربردهای عملی تأکید می‌کند.

باتوجه‌به نتایج این پژوهش پیشنهاد می‌شود که شرکت سرمایه گذاری غدیر، از مدل‌های یادگیری ماشینی مانند ماشین‌های بردار پشتیبان و جنگل تصادفی برای پیش‌بینی عملکرد شرکت استفاده کنند. استفاده از روش‌های انتخاب ویژگی، به‌ویژه الگوریتم ژنتیک، در ترکیب با تکنیک‌های گروهی می‌تواند با بهبود دقت پیش‌بینی‌ها، تصمیم‌گیری سرمایه‌گذاری را به میزان قابل‌توجهی بهبود دهد. همچنین مدل‌های توسعه‌یافته در این پژوهش را می‌توان در سیستم‌های مدیریت ریسک در بخش مالی گنجاند. با پیش‌بینی نتایج مالی بالقوه، شرکت‌ها می‌توانند ریسک‌ها را بهتر ارزیابی کنند و تصمیمات استراتژیک آگاهانه‌تری در مورد سرمایه‌گذاری، تخصیص دارایی‌ها و تملک شرکت‌ها اتخاذ کنند. پیشنهاد می‌شود محققان سایر روش‌های گروهی مانند انباشتن و ترکیب را بررسی کنند که چندین طبقه‌بندی‌کننده یا مدل‌های رگرسیون را به روش‌های مختلف ترکیب می‌کنند. این رویکردها ممکن است بهبودهای بیشتری را در دقت پیش‌بینی و تعمیم، به‌ویژه در حوزه‌های مالی بسیار ناپایدار یا پیچیده ارائه دهند. همچنین محققان می‌توانند تحلیل احساسات، داده‌های خبری یا شاخص‌های کلان اقتصادی را در مدل‌های یادگیری ماشین بگنجانند. این امر به درک جامع‌تری از اینکه چگونه عوامل خارجی بر عملکرد شرکت تأثیر می‌گذارد و منجر به پیش‌بینی‌های دقیق‌تر می‌شود، می‌انجامد.

## منابع و مآخذ

### منابع فارسی


- اصغری، ایرج؛ شکرخواه، جواد؛مرفوع، محمد و سلیمی، محمد جواد. (۱۴۰۱). داده‌کاهی متغیرهای نمایندهٔ نقدشوندگی با استفاده از روش تحلیل مؤلفه‌های اصلی در بازار اوراق بهادار تهران، *مدیریت دارایی و تامین مالی مدیریت دارایی و تامین مالی*، ۱۰(۴)، ۴۷-۶۶.
- جعفری، علی؛ منصوری‌خواه، مصطفی و پورآقاجان، عباسعلی. (۱۴۰۲). پیش‌بینی بازده سهام با تأکید بر نقش معیارهای مالی و نظارتی با استفاده از روش‌های یادگیری ماشین، *مطالعات حسابداری و حسابرسی*، ۱۲(۴۵)، ۱۲۵-۱۴۶.
- عباسیان، عزت‌اله؛ شهرکی، کاوه؛ فلاح پور، سعید و نمکی، علی. (۱۴۰۲). رویکردی نوین در پیش‌بینی درماندگی مالی با به‌کارگیری اطلاعات مبتنی‌بر شبکهٔ مالی و روش ترکیبی درخت تصمیم تقویت گرادیان، *مدیریت دارایی و تامین مالی*، ۱۱(۳)، ۱۱۳- ۱۴۰
- میرزایی، سجاد؛ آشتاب، علی و زواری رضائی، اکبر. (۱۴۰۲). مقایسهٔ کارایی مدل‌های آماری و یادگیری ماشین و انتخاب مدل بهینه در پیش‌بینی سود خالص و جریان‌های نقدی عملیاتی، *مدیریت دارایی و تأمین مالی*، ۱۱(۲)، ۵۳-۷۴.
- میرزایی، سجاد؛ محمدی، مهدی و منصور فر، غلامرضا. (۱۴۰۲). مقایسه دقت مدل‌های آماری و یادگیری ماشین برای پیش‌بینی نگهداشت وجه نقد و ارائه مدل بهینه، *راهبرد مدیریت مالی*، ۱۱(۳)، ۱-۲۸.
- هارونکلایی، کاظم و برزگر، قدرت الله. (۱۴۰۲). تبیین متغیرهای مالی مؤثر در پیش‌بینی احیای مالی با استفاده از رویکرد هوش مصنوعی .*مدلسازی اقتصادی*، ۱۷(۶۱)، ۸۹-۱۰۴.


### منابع لاتین


- Albuquerque, R., Koskinen, Y., & Zhang, C. (2019). Corporate social responsibility and firm risk: Theory and empirical evidence. *Management Science, 65*(10), 4451–4469.
- Anand, V., Brunner, R., Ikegwu, K., & Sougiannis, T. (2019). Predicting profitability using machine learning. *SSRN*.
- Attewell, P., & Monaghan, D. (2015). *Data mining for the social sciences: An introduction*. University of California Press.
- Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: Artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications, 19*(8), 1165–1195.

- Ben Jabeur, S., Stef, N., & Carmona, P. (2023). Bankruptcy prediction using the XGBoost algorithm and variable importance feature engineering. *Computational Economics, 61*(2), 715–741.
- Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and Information Systems, 34*, 483–519.
- Chen, Y., & Hao, Y. (2017). A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. *Expert Systems with Applications, 80*, 340–355.
- Cortes, C., & Vapnik, V. (1995). Support vector networks. *Machine Learning, 20*, 273–297.
- Delen, D., Kuzey, C., & Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert Systems with Applications, 40*(10), 3970–3983.
- Farooq, U., & Qamar, M. A. J. (2019). Predicting multistage financial distress: Reflections on sampling, feature and model selection criteria. *Journal of Forecasting, 38*(7), 632–648.
- Freeman, R. N., Ohlson, J. A., & Penman, S. H. (1982). Book rate-of-return and prediction of earnings changes: An empirical investigation. *Journal of Accounting Research, 20*(2), 639–653.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies, 33*(5), 2223–2273.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Issah, M., & Antwi, S. (2017). Role of macroeconomic variables on firms' performance: Evidence from the UK. *Cogent Economics & Finance, 5*(1), 1405581.
- Jan, C. L. (2018). An effective financial statements fraud detection model for the sustainable development of financial markets: Evidence from Taiwan. *Sustainability, 10*(2), 513.
- Jones, S., Johnstone, D., & Wilson, R. (2015). An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes. *Journal of Banking & Finance, 56*, 72–85.
- Katuwal, R., Suganthan, P. N., & Zhang, L. (2020). Heterogeneous oblique random forest. *Pattern Recognition, 99*, 107078.
- Li, Y., Li, T., & Liu, H. (2017). Recent advances in feature selection and its applications. *Knowledge and Information Systems, 53*, 551–577.

- Liang, D., Lu, C. C., Tsai, C. F., & Shih, G. A. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research, 252*(2), 561–572.
- Liang, D., Tsai, C. F., & Wu, H. T. (2015). The effect of feature selection on financial distress prediction. *Knowledge-Based Systems, 73*, 289–297.
- Liang, L., Liu, B., Su, Z., & Cai, X. (2024). Forecasting corporate financial performance with deep learning and interpretable ALE method: Evidence from China. *Journal of Forecasting*.
- Lu, X., Chen, X. Y., Zhang, L. S., & Liu, H. X. (2001). Prediction ability of basic financial information of Chinese listed companies on future earnings. *Economic Science, 6*, 53–62.
- Oeyono, J., Samy, M., & Bampton, R. (2011). An examination of corporate social responsibility and financial performance: A study of the top 50 Indonesian listed corporations. *Journal of Global Responsibility, 2*(1), 100–112.
- Olayinka, A. A. (2022). Financial statement analysis as a tool for investment decisions and assessment of companies' performance. *International Journal of Financial, Accounting, and Management, 4*(1), 49–66.
- Ou, J. A., & Penman, S. H. (1989). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics, 11*(4), 295–329.
- Oz, I. O., Yelkenci, T., & Meral, G. (2021). The role of earnings components and machine learning on the revelation of deteriorating firm performance. *International Review of Financial Analysis, 77*, 101797.
- Papíková, L., & Papík, M. (2022). Effects of classification, feature selection, and resampling methods on bankruptcy prediction of small and medium-sized enterprises. *Intelligent Systems in Accounting, Finance and Management, 29*(4), 254–281.
- Sermpinis, G., Stasinakis, C., & Hassanniakalager, A. (2017). Reverse adaptive krill herd locally weighted support vector regression for forecasting and trading exchange traded funds. *European Journal of Operational Research, 263*(2), 540–558.
- Shanmuganathan, M. (2018). Visualized financial performance analysis: Self-organizing maps (MS). *Research Journal of Finance and Accounting, 9*(12), 27–34.
- Shen, W., Guo, X., Wu, C., & Wu, D. (2011). Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm. *Knowledge-Based Systems, 24*(3), 378–385.

- Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting, 36*(1), 75–85.
- Son, P. V. H., & Duong, L. T. (2024). Research on applying machine learning models to predict and assess return on assets (ROA). *Asian Journal of Civil Engineering, 25*, 1–11.
- Soumm, M. (2024). Causal inference tools for a better evaluation of machine learning. *arXiv*.
- Sun, J., Li, H., Fujita, H., Fu, B., & Ai, W. (2020). Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting. *Information Fusion, 54*, 128–144.
- Tang, X., Li, S., Tan, M., & Shi, W. (2020). Incorporating textual and management factors into financial distress prediction: A comparative study of machine learning methods. *Journal of Forecasting, 39*(5), 769–787.
- Tsai, C. F., Sue, K. L., Hu, Y. H., & Chiu, A. (2021). Combining feature selection, instance selection, and ensemble classification techniques for improved financial distress prediction. *Journal of Business Research, 130*, 200–209.
- Tumpach, M., Surovičová, A., Juhászová, Z., Marci, A., & Kubaščíková, Z. (2020). Prediction of the bankruptcy of Slovak companies using neural networks with SMOTE. *Ekonomický časopis, 68*(10), 1021–1039.
- Tutcu, B., Kayakuş, M., Terzioğlu, M., Ünal Uyar, G. F., Talaş, H., & Yetiz, F. (2024). Predicting financial performance in the IT industry with machine learning: ROA and ROE analysis. *Applied Sciences, 14*(17), 7459.
- Wang, Z., & Sarkis, J. (2017). Corporate social responsibility governance, outcomes, and financial performance. *Journal of Cleaner Production, 162*, 1607–1616.
- Wong, Z., Chen, A., Taghizadeh-Hesary, F., Li, R., & Kong, Q. (2022). Financing constraints and firm's productivity under the COVID-19 epidemic shock: Evidence of A-shared Chinese companies. *The European Journal of Development Research, 35*(1), 167–193.
- Zhou, F., Zhang, Q., Sornette, D., & Jiang, L. (2019). Cascading logistic regression onto gradient boosted decision trees for forecasting and trading stock indices. *Applied Soft Computing, 84*, 105747.