Journal of Advances in Finance and Investment
Volume 4, Issue 4, 2023
pp. 177-206.
Paper type: Research paper

The use of support vector machine and Naive Bayes algorithms and its combination with risk measure and fuzzy theory in the selection of stock portfolio

Danial Mohammadi[1], Emran Mohammadi[1], Naeim Shokri[2], Nima Heidari[3]

Received: 21/10/2023 Accepted: 22/12/2023

Extended Abstract

Introduction

Choosing the right investment portfolio makes people earn more profit by investing in the right fields. Therefore, studying ways to determine the optimal stock portfolio is of great importance and necessity. In recent decades, special attention has been paid to the issue of stock portfolio in financial engineering. Many researchers researched this issue and proposed models to optimize the investment portfolio, in which they tried to improve the previous models. On the other hand, considering that the uncertainty in the future economic conditions plays a key role in financial decisions, especially the issues of stock portfolio selection, stock portfolio optimization techniques should be studied along with risk measurement and contingency planning techniques. Both in the part of classification by machine learning algorithms to separate data and in the part of selecting the optimal basket and portfolio, the research gap can be checked. In the data classification part, methods such as random or random classification have been used so far, in this study artificial intelligence method has been used, and more importantly in the part of selecting and optimizing the capital portfolio. In this study, it is tried to use the ability of neural network (machine learning) to create a relationship between different variables, portfolio using machine learning methods (support vector machine and simple Bayes) as well as value at risk and value at conditional risk and it should be combined with the fuzzy theory, which

۱ Department of Financial Engineering, Faculty of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran.
۲ Department of Economic Development and Planning, Faculty of Management and Economics, Tarbiat Modares University, Tehran, Iran. (Corresponding Author). n.shokri@modares.ac.ir
۳ Department of Industrial Engineering, Faculty of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran.

How to cite this paper: Mohammadi, D., Mohammadi, E., Shokri, N., & Heidari, N. (2023). The use of support vector machine and Naive Bayes algorithms and its combination with risk measure and fuzzy theory in the selection of stock portfolio. *Advances in Finance and Investment, 4*(4), 177-206. [In Persian]

https://doi.org/10.30495/afi.2023.1995691.1257

has a better return than the average return of the market, and this method reveals the innovation of the current research compared to other researches in this field.

Literature Review

Selection of a stock portfolio is a crucial aspect of investment decision-making. Various researchers such as Bleuler *et al.* (2001), Huang (2012), Lin and Gen (2007), Qiu *et al.* (2015), Hastie *et al.* (2009), Lai *et al.* (2006), Linoff and Berry (2011), Goykhman and Teimouri (2018), Chowdhury *et al.* (2020), Wang *et al.* (2002), Yamai and Yoshiba (2002), Bi and Zhang (2004), Ravi and Ravi (2017), Zhong and Enke (2017), Schlottmann and Seese (2001), Zolfaghari and Gholami (2021), and Wang *et al.* (2020) have conducted studies in this field. Their research focuses on developing effective methods and techniques for selecting stock portfolios. They analyze various factors such as risk, return, market conditions, and investor preferences to optimize the portfolio's performance. The studies explored different approaches, including utilizing mathematical models, heuristic algorithms, and portfolio selection theories such as the Markowitz model. Researchers have identified that real-world constraints, such as a large number of assets and weight value limitations, pose challenges in applying precise mathematical algorithms. Therefore, the use of heuristic algorithms has gained significance in constructing optimal portfolios. The studies also emphasize the importance of considering market dynamics, investor behavior, and incorporating diverse assets to enhance portfolio diversification. Overall, the researches conducted by these scholars contribute to the understanding and development of effective stock portfolio selection strategies, aiming to improve investment performance and achieve desired risk-return trade-offs.

Research Methodology

In this study, an attempt is made to design a model using the ability of neural networks (machine learning) to establish a relationship between different variables in order to create an optimal stock portfolio. Before delving into the optimization of the stock portfolio, the data is first segregated and classified using machine learning techniques. Then, the filtered data will be optimized based on the type of available information (Raei and Beigi, 2010). The machine learning-based models are artificial intelligence models that enable using various types of data. The main objective of this study is to construct a decision-making model that is integrated using machine learning techniques (support vector machine and naive Bayes). Ultimately, it will result in an optimal portfolio (stock portfolio) from the desired class, exposed to value at risk (VaR) and conditional value at risk (C_VaR) measures combined with fuzzy theory. In this regard, after obtaining the information (prices) from the official website of the stock exchange organization, they are classified using a series of features and artificial intelligence techniques (a filtering process is applied). These stocks are divided into two classes: positive and negative, with the aim of creating an optimal portfolio from the positive class

based on the risk measures, C_VaR and VaR. Finally, using the Trainer criterion, the generated portfolios will be compared to the market returns.

Results

In the following, we will examine the results of using the examined models and the limitations caused by them. In this research, annual risk-free return is considered to be 23%, which is estimated at 1.9% for one month and is used in comparisons. On the other hand, the monthly return of the market (total index) in 2023 between April and May is 17%. In Table (18), the intended outputs of the algorithms as well as the risk metrics are ranked along with different levels of confidence to convey a clear and specific view to the readers. It is clear that the CVAR risk measure has a better capability and result than the VAR risk measure. Also, the support vector machine algorithm has achieved better results.In the analysis of the results, the actual return of the investigated stocks with the output weight of the algorithm has been split and ranked. Finally, the efficiency of each of the methods is compared to the market efficiency with Trainer's criterion and is listed in Table (18).

Discussion and Conclusion

Nowadays, investors use various metrics to measure returns and risks. These metrics are selected based on the investors' behavior in the capital market and their level of knowledge and expertise in financial matters. The topic of utilizing risk in portfolio analysis has been addressed, and investors, regardless of risk aversion, have always tried to optimize the relationship between return and the risk associated with their activities. The foundation of this research is based on two variables: return rate and risk, along with machine learning algorithms and the use of different risk measures. Each of these two macro variables, as well as price prediction, are considered decision-making criteria in the investment process. According to the obtained results, the CVaR risk measure has demonstrated better capability and results compared to the VAR risk measure. Additionally, the support vector machine algorithm has achieved superior results. Based on the results, it is recommended that researchers utilize other machine learning algorithms, both in classification and prediction, and also benefit from other risk measures, in addition to value-at-risk and conditional value-at-risk.

Conflict of Interest

The author (s) of this article declared no conflict of interest regarding the authorship or publication of this article.

Keywords: Conditional Value at Risk; Machine Learning; Stock portfolio; Tehran security exchange.

JEL Classification: G01, G32, E44.

پیشرفت‌های مالی و سرمایه‌گذاری
سال چهارم، زمستان ۱۴۰۲- شماره ۴
صفحات ۲۰۶-۱۷۷
نوع مقاله: پژوهشی

استفاده از الگوریتم‌های ماشین بردار پشتیبان و بیز ساده و ترکیب آن با سنجه ریسک و نظریه فازی در انتخاب سبد سهام

دانیال محمدی[۱]، عمران محمدی[۱]، نعیم شکری[۲]، نیما حیدری[۳]

تاریخ دریافت: ۱۴۰۲/۰۷/۲۹ تاریخ پذیرش: ۱۴۰۲/۱۰/۰۱

چکیده

هدف: هدف پژوهش حاضر تشکیل پرتفوی بهینه با استفاده از الگوریتم‌های یادگیری ماشین و نظریه فازی است که بازدهی بهتر از بازده میانگین بازار (شاخص کل بورس اوراق بهادار) داشته باشد.

روش‌شناسی پژوهش: در پژوهش حاضر، سهام شرکت‌های منتخب در مرحله اول با استفاده از دو الگوریتم معرفی شده، کلاس‌بندی می‌شوند. در مرحله بعد سهامی که وارد کلاس مثبت شده‌اند، با کمک الگوریتم جنگل تصادفی، برای روز معاملاتی بعد پیش‌بینی می‌شوند. برای هریک از شرکت‌ها سه پیش‌بینی صورت می‌گیرد که ورودی‌های بهینه‌سازی با روش فازی هستند. بهینه‌سازی با هدف کمینه‌کردن ریسک با سنجه‌های ریسک ارزش در معرض خطر و ارزش در معرض خطر شرطی صورت می‌گیرد. اطلاعات سهم‌ها پنج‌ساله، به‌صورت روزانه و بازه زمانی آن از ابتدای سال ۱۳۹۷ تا پایان سال ۱۴۰۱ می‌باشد.

یافته‌ها: در پایان، هرکدام از الگوریتم‌ها و سنجه ریسک مورداستفاده آن، با بازده واقعی بازار سنجیده و مقایسه گردید. بر اساس نتایج به‌دست‌آمده سنجه ریسک CVAR قابلیت و نتیجه بهتری را نسبت به سنجه ریسک VAR داشته است و الگوریتم ماشین بردار پشتیبان نیز، عملکرد بهتری را در انتخاب سبد سرمایه‌گذاری رقم زده است.

اصالت / ارزش‌افزوده علمی: این پژوهش با یکپارچه‌سازی روش‌های یادگیری ماشین و سنجه‌های ریسک، به شکل یک نمونه سرمایه بهینه می‌شود. اضافه‌کردن سنجه‌های ریسک VaR و CVaR فرآیند تصمیم‌گیری در خصوص کاهش ریسک را تقویت می‌کند. پیش‌بینی به کمک جنگل تصادفی و استفاده از رویکردی بر پایه نظریه فازی برای تحلیل ریسک و ارزش، پژوهش را به یک چشم‌انداز نوآورانه در تشکیل پرتفوی می‌بخشد. یافته‌ها، به سرمایه‌گذاران و پژوهشگران در جستجوی استراتژی‌های سرمایه‌گذاری بهتر، یافته‌های ارزشمندی ارائه می‌دهد.

کلیدواژه‌ها: ارزش در معرض ریسک شرطی، بورس اوراق بهادار تهران، سبد سهام، یادگیری ماشین.
طبقه‌بندی موضوعی: G01, G32, E44.

۱. گروه مهندسی مالی، دانشکده مهندسی صنایع، دانشگاه علم و صنعت ایران، تهران، ایران.
۲. گروه توسعه و برنامه ریزی اقتصادی، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران، ایران. (نویسنده مسئول). n.shokri@modares.ac.ir
۳. گروه مهندسی صنایع، دانشکده مهندسی صنایع، دانشگاه علم و صنعت ایران، تهران، ایران.

استناد: محمدی، دانیال؛ محمدی، عمران؛ شکری، نعیم؛ حیدری، نیما. (۱۴۰۲). استفاده از الگوریتم‌های ماشین بردار پشتیبان و بیز ساده و ترکیب آن با سنجه ریسک و نظریه فازی در انتخاب سبد سهام. *پیشرفت‌های مالی و سرمایه‌گذاری*، ۴(۴). ۲۰۶-۱۷۷.
https://doi.org/10.30495/afi.2023.1995691.1257

۱- مقدمه

انتخاب سبد مطلوب سهام[1]باعث می‌شود افراد با سرمایه‌گذاری در بازار سرمایه، سود بیشتری کسب کنند. بنابراین مطالعه راه‌های تعیین سبد بهینه از اهمیت بالایی برخوردار است. پژوهش‌گران بسیاری بر روی این مسئله پژوهش کردند و مدل‌هایی برای بهینه‌سازی سبد سهام پیشنهاد نمودند، که در آن‌ها سعی شده مدل‌های قبلی بهبود داده شود. هدف از حل مسئله بهینه‌سازی سبد سهام، ارائه ابزاری مناسب برای سرمایه‌گذاران جهت انتخاب سبد سهام مناسب است (Zare Mehrjerdi *et al.*, 2013). یادگیری ماشین[2]در تحقیقات مالی مدرن، به‌عنوان یک رویکرد مقیاس‌پذیر برای انتخاب پورتفولیو توجه بالایی را به خود جلب کرده است. یکی‌از چالش‌های تکراری در این زمینه، بهینه‌سازی پرتفوی از طریق انتخاب و تخصیص مناسب سهام است (Paula & Vitor Venceslau Curtis, 2023). باتوجه به این موضوع که عدم‌قطعیت در شرایط اقتصادی آینده نقشی مهمی را در تصمیم-گیری‌های مالی به‌ویژه مسائل انتخاب سبد سهام بازی می‌کنند، باید تکنیک‌های بهینه‌سازی سبد سرمایه‌گذاری را در کنار تکنیک‌های سنجه ریسک و برنامه‌ریزی احتمالی مطالعه کرد. در مرحله طبقه-بندی[3]نمونه‌ها (سهام) تاکنون از روش‌هایی چون کلاس‌بندی تصادفی استفاده می‌شد که در این پژوهش از الگوریتم‌های ماشین بردار پشتیبان و بیز ساده استفاده شده است و از آن مهم‌تر در قسمت انتخاب و بهینه‌کردن سبد سهام. تحقیقات پیشین همگی برفرض نرمال بودن دادخ‌ها و استفاده از مدل پایه مارکوییتز بنا شده بود (Sultan Panah *et al.*, 2013). انتخاب سبد سهام و مدیریت آن از مهم-ترین حوزه‌های تصمیم‌گیری مالی می‌باشد. در این راستا، بررسی و مطالعه سرمایه‌گذاران جهت تشکیل بهترین سبد باتوجه به میزان ریسک و بازده آن انجام می‌شود. انتشار نظریه انتخاب سبد سهام مارکوییتز[4] اصلی‌ترین موفقیت در این زمینه است (Farahabadi *et al.*, 2022). مدل پایه مارکوییتز با استفاده از مدل‌های برنامه‌ریزی ریاضی قابل‌حل است، ولی وقتی محدودیت‌های واقعی هم‌چون تعداد زیاد سرمایه، به آن افزوده می‌شود، استفاده از الگوریتم‌های دقیق ریاضی غیرممکن شده و از این روست که الگوریتم-های فراابتکاری جایگاه ویژه‌ای یافته‌اند. امروزه استفاده از شبکه‌های عصبی ازجمله روش‌هایی است که برای حل مسائل بهینه‌سازی بسیار موردتوجه می‌باشد (Fazlzadeh *et al.*, 2012). سرعت محاسباتی بسیار بالا ازجمله مزیت‌های این روش محسوب می‌شود، که باعث اقبال پژوهش‌گران به آن شده است. شایان ذکر است که مدل‌های میانگین–واریانس و میانگین–انحراف‌معیار مطلق، در بسیاری

1. Stock
2. Machine Learning
3. Classification
4. Markowitz

از بازارهای سهام، زمانی که بازده پرتفوی متقارن است، مؤثر است. با این‌حال، اگر بازده پرتفوی نامتقارن باشد، ناکارامد است، زیرا واریانس و انحراف مطلق، بازده مناسب را هم درنظر می‌گیرند که سرمایه‌گذاران آن‌را مطلوب می‌دانند، دقیقا به همان اندازه که بازده نامناسب است، که سرمایه‌گذاران نمی‌خواهند (Behera *et al.*, 2023). از این‌رو از دو سنجه ریسک VaR و CVaR استفاده شده تا صرفا انحرافات منفی از بازده را ارزیابی کند. در این مطالعه سعی می‌شود تا با استفاده از توانایی یادگیری ماشین در ایجاد رابطه بین متغیرهای مختلف، پرتفویی با استفاده از روش‌های یادگیری ماشین (ماشین بردار پشتیبان[1] و بیز ساده[2]) و همچنین ارزش در معرض ریسک[3] و ارزش در معرض ریسک شرطی[4] و ترکیب آن با نظریه فازی تشکیل شود، که بازده‌ای بهتر از بازده میانگین بازار داشته باشد. رویکرد بسیاری از مقالات، ساخت سبد بهینه مبتنی‌بر یادگیری ماشین، استفاده از یک مدل پیش‌بینی بازده سهم برای شناسایی سهم‌های پربازده و قرار دادن آن‌ها در سبد سهام و تقسیم سرمایه به طور برابر میان سهام منتخب است (Haratizadeh and Rezaee, 2023). نوآوری اصلی این پژوهش را می‌توان از دو منظر بررسی کرد. اول؛ مقایسه عملکرد دو الگوریتم؛ ماشین بردار پشتیبان و بیز ساده در قدرت تشخیص روند و به‌تبع آن کلاس‌بندی سهام است. در پژوهش پایوا و همکاران (Paiva *et al.*, 2019) که با الگوریتم ماشین بردار پشتیبان، کلاس‌بندی را انجام داده و با مدل اولیه مارکوییتز، سبد سهام را بهینه کرده‌اند، از سنجه ریسک واریانس استفاده کرده که در پژوهش حاضر از دو سنجه ریسک ارزش در معرض خطر و ارزش در معرض خطر شرطی که تطابق بهتری با واقعیت بازار دارد، استفاده شد. همچنین این دو سنجه ریسک، باتوجه به عدم‌قطعیتی که راجب بازده‌ها وجود داشت، با نظریه فازی ادغام شده، درنتیجه بهینه‌سازی قابل‌اعتمادتری را به پژوهش‌گران منتقل خواهد کرد. دوم؛ استفاده هم‌زمان ظرفیت الگوریتم‌ها، یعنی کلاس‌بندی و پیش‌بینی. در این تحقیق هم‌زمان از دو قابلیت مذکور استفاده شده است که اولی برای طبقه‌بندی سهام پربازده و دومی برای پیش‌بینی روز معاملاتی بعد و ورودی برای مرحله بهینه‌سازی فازی مسئله می‌باشد. به‌طورکلی استفاده هم‌زمان از الگوریتم‌های یادگیری ماشین، عدم‌قطعیت و سنجه‌های ریسک مناسب و قیاس عملکردی آن‌ها، تحقیق حاضر را برروی سهام شرکت-های بورس اوراق بهادار، منحصر به فرد کرده است.

1. Support Vector Machine
2. Naïve Bayes
3. Value at Risk
4. Conditional Value at Risk

۲- مبانی نظری و پیشینه پژوهش

تغییر پورتفوی[1] یک استراتژی سرمایه‌گذاری است که با تنظیم یک سبد برای تولید ارزش بیشتر نسبت به حرکت بازار پاسخ می‌دهد (Reddy and Nagarjuna, 2023). ازجمله دغدغه‌های سرمایه-گذاران در بازار سرمایه، سرمایه‌گذاری در سبد اوراقی و قراردادهایی است که پر بازده‌تر باشد. یکی از وظایفی که متخصصین مالی با آن مواجه هستند، کنترل و مشاوره برای ایجاد یک زمینه‌ی مناسب برای سرمایه‌گذاری با هدف دستیابی به سبدی بهینه از اوراق بهادار می‌باشد که مهندسی مالی توانسته است با ارائه‌ی راهکارهای مناسب این نیاز سرمایه‌گذاران را برطرف کند (Rai and Pouyanfar, 2017). بیشتر مسائل بهینه‌سازی که در جهان واقعی با آن ها مواجه هستیم، بیش‌از یک هدف را در بر می‌گیرند. نخستین قدم در سنجش عملکرد سبد سرمایه بهینه، محاسبه بازدهی به‌دست‌آمده باتوجه به هدف‌های از قبل تعیین شده، برای آن می‌باشد و ازآنجا که افزایش ثروت سهامدار به‌عنوان هدف اصلی تشکیل سبد است، باید به‌گونه‌ای عمل کند که با کسب بازده مناسب، ثروت خودرا را بهبود بخشد (Taghizadeh *et al.*, 2022). قیمت سهام، یک منبع اطلاعاتی مهم و مؤثر تصمیم‌گیری برای افراد حقوقی و حقیقی در بازار سرمایه محسوب می‌شود. به‌تبع آن پیش‌بینی تغییرات قیمت نیز از اهمیت بالایی برخوردار خواهد بود. با وجود ابزارهای تحلیلی جهت پیش‌بینی قیمت سهام، عامل اول تعیین شاخص‌های مؤثر بر آن است. در این راستا مطالعات متعددی صورت گرفته است و اجماعی روی آن وجود ندارد (Ebadati *et al.*, 2022). در این‌گونه مسائل، که با عنوان بهینه‌سازی چندهدفه شناخته شده است، ناگزیریم چندین تابع هدف یا شاخص عملکرد را تعریف نماییم و به طور هم‌زمان و موازی مقدار همه‌ی آن‌ها را بهینه نماییم. از آنجایی‌که روش‌های بهینه‌سازی تک‌هدفه در هر مرحله از اجرا تنها یک پاسخ را می‌دهند، نمتوان برای یافتن مجموعه‌ای از پاسخ‌های بهینه مطلوب باشد، بنابراین باتوجه به بالابودن درجه‌ی پیچیدگی، امروزه الگوریتم‌های تکاملی چندهدفه و استفاده از هوش‌مصنوعی ابزاری مطلوب برای حل مسائل بهینه‌سازی چندهدفه درنظر گرفته می‌باشد (Sinaei and Zamani, 2014).

در جدول (۱) اسامی افرادی‌که در حوزه کاربرد یادگیری ماشین و شبکه عصبی در مبحث مالی و سبد سهام فعالیتی را صورت داده‌اند، ارائه شده. بعداز مطالعه و بررسی آن، به این نتیجه رسیده می‌شود که بالاترین الگوریتم مورداستفاده توسط پژوهش‌گران SVM و بعد از آن شبکه‌های عصبی است.

1. Portfolio Change

جدول (۱) پیشینه پژوهش

Table (1) research background

ردیف	نویسنده	سال	SVM[8]	DL[7]	ANN[6]	GA[5]	E[4]	B[3]	DT[2]	O[1]
۱	بلویلر و همکاران (Bleuler *et al.*, 2001)	۲۰۰۱	■		■					
۲	هوانگ (Huang, 2012)	۲۰۱۲	■				■		■	
۳	لین و ژن (Lin and Gen, 2007)	۲۰۰۷			■					
۴	کیو و همکاران (Qiu *et al.*, 2015)	۲۰۱۵		■						
۵	هستی و همکاران (Hastie *et al.*, 2009)	۲۰۰۹					■			
۶	لای و همکاران (Lai *et al.*, 2006)	۲۰۰۶					■			
۷	لینوف و بری (Linoff and Berry, 2011)	۲۰۱۱		■						■
۸	گویخمان و ایموری (Goykhman and Teimouri, 2018)	۲۰۱۸						■		
۹	چاودری و همکاران (Chowdhury *et al.*, 2020)	۲۰۲۰	■							
۱۰	وانگ و همکاران (Wang *et al.*, 2002)	۲۰۰۲	■	■	■		■			
۱۱	یامی و یوشیبا (Yamai and Yoshiba, 2002)	۲۰۰۲					■			
۱۲	بی و ژانگ (Bi and Zhang, 2004)	۲۰۰۴							■	
۱۳	راوی و راوی (Ravi and Ravi, 2017)	۲۰۱۷			■					
۱۴	ژونگ و انکه (Zhong and Enke, 2017)	۲۰۱۷			■					
۱۵	شلاتمان و زیس (Schlottmann and Seese, 2001)	۲۰۰۱				■				
۱۶	ذوالفقاری و غلامی (Zolfaghari and Gholami, 2021)	۲۰۲۱	■							
۱۷	وانگ و همکاران (Wang *et al.*, 2020)	۲۰۲۰	■		■					

1. Other Category
2. Decision Tree
3. Bayasian Models
4. Ensemble Models
5. Genetic Algorithms
6. Artificial Neutral Networks
7. Deep Learning
8. Support Vector Machine

۳- روش‌شناسی پژوهش

در این پژوهش سعی شد تا با کمک دو الگوریتم یادگیری ماشین در ساخت رابطه‌ای بین متغیرهای متفائت، مدلی طراحی شود تا بتواند یک سبد سرمایه‌گذاری بهینه را طراحی کند. قبل از ورود به بحث بهینه‌سازی سبد، در ابتدا به‌وسیله الگوریتم‌ها، طبقه‌بندی[1]سهام صورت گرفت و سپس داده‌های انتخاب شده برحسب نوع اطلاعاتی که وجود دارد، بهینه شدند (Raei and Beigi, 2010). مدل‌های مبتنی‌بر الگوریتم‌های یادگیری ماشین، مدل‌های مبتنی‌بر هوش مصنوعی هستند که استفاده از انواع مختلف داده‌ها را ممکن می‌سازند. هدف اصلی از پژوهش حاضر ساختن مدلی تصمیم‌گیری است که یکپارچه-سازی آن به‌وسیله الگوریتم‌های یادگیری ماشین انجام شده و درنهایت از کلاس مثبت، یک سبد بهینه با ارزش در معرض ریسک و ارزش در معرض ریسک شرطی که با نظریه فازی ترکیب شده ، بدست آمد. در این راستا بعداز دریافت اطلاعات (قیمت‌ها) از سایت رسمی سازمان بورس اوراق بهادار، آن‌ها به‌وسیله یک سری ویژگی‌ها و به‌کمک یادگیری ماشین طبقه‌بندی می‌شوند (نوعی فیلتر صورت گرفت). این سهم‌ها به دو کلاس مثبت و منفی تقسیم و هدف آن است که از بین سهم‌های کلاس مثبت یک سبد بهینه به کمک سنجه ریسک C_VAR و VaR ایجاد شود. درنهایت با استفاده از معیار بررسی بهینگی ترینر، پرتفوهای ایجادشده با بازدهی صورت‌گرفته از طریق بازار مقایسه شدند. در سال ۱۹۶۵ گامی بسیار ارزشمند در طراحی دسته‌بندی برداشت شد و نظریه آماری یادگیری را به‌صورت مستحکم-تری بنا نهاد و ماشین‌های بردار پشتیبان را بر این اساس ارائه داد. رویکرد ماشین بردار پشتیبان به‌این صورت است که در مرحله آموزش سعی خواهد کرد که مرز تصمیم‌گیری[2]به‌گونه‌ای انتخاب گردد که حداقل فاصله آن با هریک از دسته‌های مورد نظر ماکزیمم گردد. این نوع انتخاب باعث می‌شود که تصمیم‌گیری در عمل پاسخ‌دهی خوبی داشته باشد، این نحوه انتخاب مرز براساس نقاطی به‌نام بردارهای پشتیبان انجام می‌شود.

از جمله روش‌های نسبتاً جدید در دسته‌بندی و شناسایی الگوها، بیز ساده و جنگل تصادفی می‌باشد. بیز ساده یک الگوریتم طبقه‌بندی ساده اما مؤثر و متداول یادگیری ماشین است که در دستهٔ یادگیری با ناظر[3]جای می‌گیرد. بیز ساده الگوریتمی احتمالی است که براساس نظریه‌ی بیز برای طبقه‌بندی استفاده می‌شود.

1. Classification
2. Decision Boundary
3. Supervised Learning

ازطرفی CVaR معیارمنسجمی برای اندازه‌گیری ریسک است و خصوصیات جالبی را از خود بروز می‌دهد، اول این که، CVaR نسبت‌به آنالیز میانگین – واریانس در مواجه با توزیع نامتقارن بازدهی سرمایه، بهتر عمل می‌کند و دوم آن‌که، حداقل‌سازی CVaR معمولا در نتیجه حل یک مسئله برنامه-ریزی همگرا (محدب) مانند مسائل برنامه‌ریزی خطی است که به تصمیم‌گیرنده اجازه می‌دهد به‌صورت کاراتر و بهتری از عهده یک مسئله پرتفوی با مقیاس بزرگ برآید (Feizi and Foroush Bastani, 2012).

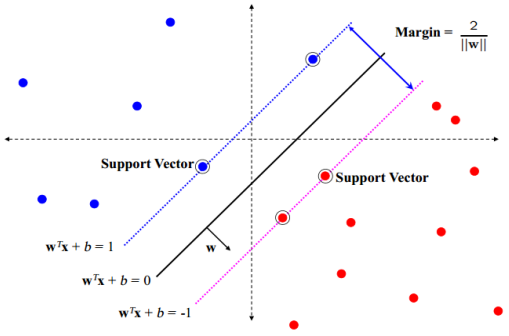
فرآیند استخراج اطلاعات از داده‌ها، یک عمل پیچیده است که می‌توان از کارهایی که انجام می‌شود، به بحث دسته‌بندی و پیش‌بینی، کاهش حجم داده‌ها، کاهش پیچیدگی داده و استخراج اصلی-ترین ویژگی‌ها اشاره کرد، که می‌توان به‌وسیله آن‌ها حجم بزرگی از داده‌های پخش شده در اطرافمان را براساس الگوهایی که دارند، به‌درستی طبقه‌بندی کرد و اطلاعات و دانش مورد نیاز خود را از آن‌ها استخراج کنیم. این کار یک مزیت دیگر نیز دارد که آن، کاهش نوسان داده‌ها است، که باعث خواهد شد درهنگام بهره‌برداری از داده‌ها، دچار درماندگی نشویم. درحقیقت، ما به بررسی یکی‌از ابزارهای جدید و در عین‌حال کارامد در حوزه داده‌کاوی پرداختیم، که قابلیت زیادی دارد و می‌تواند بهترین نتیجه را به نمایش بگذارید. در الگوریتم ماشین بردار پشتیبان، مهم است که بتوانیم مقدار بهینه‌ای برای وزن (شیب خطوط) و پارامتر عرض از مبدا (بایاس) پیدا کنیم. می‌توان گفت، به دنبال خطی هستیم که بهترین جداسازی بین داده‌ها را انجام دهد، که به این خط در n بعد، ابرصفحه گویند. بهترین ابرصفحه را ابرصفحه[1]جداکننده با بیشترین حاشیه نام‌گذاری می‌کنیم (Beale and Jackson, 2007).

ماشین بردار پشتیبان[2]درواقع یک کلاس‌بند باینری است که دو کلاس را به‌کمک یک مرز خطی از هم جدا می‌کند. در این روش با استفاده از یک الگوریتم بهینه‌سازی، نمونه‌هایی که مرز کلاس‌ها را تشکیل می‌دهند، به‌دست می‌اورد، این نمونه‌ها را بردارهای پشتیبان گویند. تعدادی از داده‌ها که کم‌ترین فاصله تا مرز تصمیم‌گیری را دارند، می‌تواند به‌عنوان زیرمجموعه‌ای برای تعریف مرزهای تصمیم‌گیری و همچنین برای بردار پشتیبان درنظر گرفته شوند (Bishop and Tipping, 2013). ویژگی‌های زیر از SVM را درنظر بگیرید (Paiva *et al.*, 2019):

- محدب بودن تابع هدف یکی از مزایای روش SVM است، زیرا آموزش این روش معادل حل مسئله برنامه‌ریزی درجه دوم است که به‌موجب آن راحل مسئله همیشه منحصربه‌فرد و به‌طورکلی بهتر خواهد بود.

1. Hyper Plan
2. Support Vector Machine

- روش SVM نسبت به سایر روش‌های پیش‌بینی فردی دقت بیشتری دارد. عملکرد برتری دارد، زیرا برای به‌حداقل‌رساندن ریسک ساختاری طراحی شده است، درحالی‌که روش‌های دیگر بر به‌حداقل‌رساندن ریسک تجربی تمرکز دارند. به‌عبارت‌دیگر، روش SVM به دنبال به‌حداقل‌رساندن حد بالای خطای تعمیم به ضرر خطای آموزشی است.



شکل (۱) ماشین بردار پشتیبان

Figure (1) support vector machine

فرض کنید داده‌ها از دو کلاس تشکیل شده و کلاس‌ها در مجموع دارای $X_i \ , (i = 1,2, \dots, n)$ نقطه داده شده باشند، که X_i بردار است. این دو کلاس با $Y_i = \pm 1$ برچسب زده می‌شوند. برای محاسبه مرز تصمیم‌گیری دو کلاس کاملا جدا از هم، از روش حاشیه بهینه استفاده می‌کنیم. در این روش مرز خطی بین دو کلاس طوری محاسبه می‌شود که:

۱. تمام نمونه‌های کلاس ۱+ در یک‌طرف مرز و تمام نمونه‌های کلاس ۱− در طرف دیگر مرز واقع گردد.

۲. مرز تصمیم‌گیری، به گونه‌ای باشد که فاصله نزدیک‌ترین نمونه‌های هر دو کلاس از یکدیگر در راستای عمود بر مرز تصمیم‌گیری تا جایی که ممکن است، بیشینه شوند.

فرض کنید تعدادی بردار یا الگوهای آموزشی به‌صورت $\{x_1, x_2, \dots, x_N\}$ داریم که هرکدام یک بردار ویژگی d بعدی بوده و دارای برچسب $y_i \in \{-1, +1\}$ است. هدف حل یک مسئله کلاس‌بندی دوکلاسه به‌صورت بهینه است. فرض کنید این دو کلاس را با تابع تمایز $f(x)$ و با یک ابرصفحه H با معادله زیر بخواهیم از هم جدا کنیم:

$$H: w.x + b = 0$$ رابطه (۱)

$$f(x) = sign(w.x + b)$$ رابطه (۲)

بردار وزن w بردار عمود بر ابر صفحهٔ تفکیک‌کننده است و منظور از w.x حاصل‌ضرب داخلی و b مقدار اریبی می‌باشد. وپنیک[1] ثابت کرد که بعد VC برای طبقه‌بندی کننده‌هایی از نوع ابرصفحه کانونی دارای یک کران بالا است که آن با توان دوم نرم وزن یعنی $||W||^2$ نسبت مستقیم دارد که:

رابطه (۳)
$$||W||^2 = \sum_{i=1}^{n} w_i^2$$

بعد VC یک مجموعه از توابع، ماکزیمم تعداد نقاطی است که می‌تواند در روش‌های ممکن مجزا شوند. درواقع اگر $||W||^2$ را مینیمم کنیم بعد VC طبقه‌بندی کننده را مینیمم کرده‌ایم و تخمین از مقدار ریسک واقعی به‌صورت احتمالی دقیق‌تر بوده و خاصیت تعمیم کلاس‌بندی بیشتر خواهد شد. فرض کنید داده‌های دو کلاس تفکیک‌پذیری باشند و بردارهای ویژگی مرزی کلاس اول روی ابرصفحه H^+ و بردارهای ویژگی مرزی کلاس دوم روی ابرصفحه H^- قرار گیرند.

رابطه (۴)
$$\begin{cases} H^+ : w.x + b = +1 \\ H^- : w.x + b = -1 \end{cases}$$

الگوهایی که بر روی ابر صفحات H^+ و H^- قرار می‌گیرند، بردار پشتیبان نامیده می‌شوند. ناحیه بین دو ابرصفحه را حاشیه مرزی می‌گویند، همچنین فاصله بین دو ابرصفحه برابر $\frac{2}{||W||}$ است.

طراحی ابر صفحه با ماکزیمم عرض ناحیه مرزی بهینه بر این استوار است که با شرط درست طبقه‌بندی‌شدن الگوها، عرض ناحیه مرزی حداکثر شود یعنی $\frac{2}{||W||}$ و $||W||$ یا $||W||^2$ کمینه گردد. هدف این است که اولا الگوها درست طبقه‌بندی گردند و ثانیا بر روی و یا خارج از ناحیه مرزی واقع شوند، یعنی:

رابطه (۵)
$$y_i(w.x_i + b) \geq 1 \qquad i = 1.2. \ldots . N$$

پس درواقع طراحی یک دسته‌بندی‌کننده ابر صفحه‌ای با ناحیه مرزی بهینه به‌صورت مسئله زیر خواهد بود:

رابطه (۶)
$$minimize \quad \frac{||W||^2}{2}$$
$$s.t: \quad y_i(w.x_i + b) \geq 1$$
$$i = 1.2. \ldots . N$$

واضح است که داریم:

$$W = [w_1.w_2. \ldots . w_d]^t \quad ||W||^2 = w.w^t \quad X_i = [x_{i1}.x_{i2}. \ldots . x_{id}]^t$$

1. Vapenic

مسئله قبل یک مسئله بهینه‌سازی مقید از نوع محدب و درجه دوم می‌باشد. برای حل این مسئله تابع لاگرانژ[1] زیر را تشکیل داده و ضرایب تابع لاگرانژ α_i را به‌دست خواهیم آورد.

$$L(w.b.\alpha) = \frac{1}{2} w.w^t - \sum_{i=1}^{N} \alpha_i (y_i (w.x_i + b) - 1)$$ رابطه (۷)

برای این‌که $(w.b.\alpha)$ جواب مسئله باشند، این جواب باید در شرایط KKT صدق کند یعنی، به‌ازای جواب به‌دست‌آمده مشتق L نسبت به w، b و α برابر صفر باشد، با مساوی صفر قرار دادن مشتق به این معاملات خواهیم رسید:

$$W = \sum_{i=1}^{N} \alpha_i y_i x_i$$ رابطه (۸)

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$ رابطه (۹)

با قراردادن مقدار w از رابطه فوق در $L(w.b.\alpha)$ به مسئله دوگان، برای بهینه‌سازی مقید خواهیم رسید:

$$\begin{aligned} &maximize \quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^{N} \alpha_i \\ &s.t: \qquad \alpha_i \geq 0 \qquad i = 1.2.\ldots.N \\ &\sum_{i=1}^{N} \alpha_i y_i = 0 \end{aligned}$$ رابطه (۱۰)

پس از حل این مسئله دوگان، ضریب لاگرانژ α_i به‌دست می‌آید. درواقع هرکدام از ضرایب α_i متناظر با یکی از الگوهای $\mathbf{x_i}$ می‌باشند. الگوهای $\mathbf{x_i}$ را که متناظر با ضرایب $\alpha_i > 0$ بردارهای پشتیان (sv_i) می‌نامیم. مقدار بردار w و b از این روابط به‌دست خواهند آمد.

$$W = \sum_{i=1}^{N_{sv}} \alpha_i y_i sv_i$$ رابطه (۱۱)

$$b_j = y_j - \sum_{i=1}^{N_{sv}} \alpha_i y_i sv_i sv_j$$ رابطه (۱۲)

$$b = \frac{1}{N_{sv}} \sum_{i=1}^{N_{sv}} b_j$$ رابطه (۱۳)

تابع تمایز برای دسته‌بندی یک الگوی ورودی x به‌صورت زیر می‌باشد:

$$f(x) = sign\left(\sum_{i=}^{N_{sv}} \alpha_i y_i x.sv_i + b\right)$$ رابطه (۱۴)

بنابراین، اگر $f(\mathbf{x}) < 0$ باشد، x متعلق به دسته اول و اگر $f(\mathbf{x}) > 0$ باشد، x متعلق به دسته دوم خواهد بود.

1. Lagrange Function

بیز ساده[1]

گاهی سعی می‌کنیم که از بین فضای فرضیه‌های H بهترین فرضیهٔ سازگار با نمونه‌های آموزشی D را پیدا کنیم. چندین راه برای تعریف "بهترین" در این جمله وجود دارد، یکی از این تعاریف "محتمل-ترین" است، با دردست‌داشتن داده‌های D بدون نیاز به هیچ اطلاعات اولیهٔ دیگر نمی‌توان محتمل‌ترین فرضیه را انتخاب کرد. قضیهٔ بیز روشی مستقیم برای محاسبهٔ احتمالات فرضیه‌های موجود در H ارائه می‌کند. به‌عبارت‌دیگر، قضیهٔ بیز متدی برای محاسبهٔ احتمال یک فرضیه بر اساس احتمال قبلی‌اش، احتمال مشاهدهٔ داده‌های سازگار با فرض درستی این فرضیه و احتمال خود داده‌های مشاهده شده ارائه می‌کند.

برای نشان‌دادن احتمال اولیهٔ فرضیهٔ h احتمال قبل از مشاهدهٔ داده‌های آموزشی، از P(h) استفاده می‌کنیم. به P(h) احتمال ابتدایی h نیز می‌گویند، این احتمال از اطلاعات قبلی ای که در مورد احتمال درستی فرضیهٔ h داریم تأثیر می‌پذیرد. به طور مشابه از P(D) برای نمایش احتمال اولیهٔ مشاهدهٔ نمونه‌های آموزشی D استفاده می‌کنیم (مثلاً احتمال مشاهدهٔ D بدون داشتن هیچ اطلاعات قبلی در مورد اینکه با چه فرضیه‌هایی سازگار است). برای نشان‌دادن احتمال مشاهدهٔ D در جایی که فرضیهٔ h درست است از P(D|h) استفاده می‌کنیم. در حالت کلی، از p(x|y) برای نمایش احتمال x با فرض وقوع y استفاده می‌کنیم. در این نوع، علاقه‌ی ما به احتمال P(h|D) است که در آن h یک فرضیه و D نمونه‌های آموزشی مشاهده شده هستند. به P(h|D) احتمال ثانویه h نیز می‌گویند، زیرا که اطمینان ما به فرضیهٔ h بعد از مشاهده‌ی نمونه‌های آموزشی D را نشان می‌دهد. توجه داشته باشید که احتمال ثانویه P(h|D) بر خلاف احتمال اولیه P(h) که از نمونه‌های آموزشی مستقل است، از نمونه‌های آموزشی D تأثیر می‌پذیرد.

قضیه بیز، اساس متدهای یادگیری بیز است، زیرا که راهی برای محاسبهٔ احتمال ثانویه P(h|D) از P(h)، P(D) و P(D|h)

$$P(h|D) = \frac{p\ (D|h)P(h)}{p(D)}$$ رابطه (۱۵)

بر اساس قضیهٔ بیز P(h|D) با افزایش P(h) و P(D|h) افزایش می‌یابد. همچنین منطقی است که P(h|D) با افزایش P(h) کاهش می‌یابد، زیرا که هر چه که احتمال مشاهدهٔ D به طور مستقل از h بالاتر رود دیگر D مدرکی برای درستی h وجود ندارد.

1. Naïve Bayes

در بسیاری از مسائل یادگیری، یادگیر مجموعهٔ فرضیه‌هایی مثل H را در نظر می‌گیرد و در بین آن‌ها به دنبال محتمل‌ترین فرضیهٔ hϵH باتوجه‌به نمونه‌های آموزشی D می‌گردد. هر کدام از این محتمل‌ترین، فرضیه با حداکثر احتمال ثانویه یا (MAP) نامیده می‌شود. فرضیه‌های MAP را می‌توان با استفاده از قضیهٔ بیز برای محاسبهٔ احتمال ثانویه هر فرضیه مشخص کرد.

در بعضی موارد، فرض می‌کنیم که هر فرضیه در H احتمال اولیهٔ مساوی ای دارد (برای هر hj و hi در H داریم که P(hi)= P(hj) در این شرایط می‌توان فقط عبارت P(D|h) را برای پیداکردن محتمل‌ترین فرضیه در نظر بگیریم. P(D|h) گاهی محتمل بودن داده‌های D برای h نیز نامیده می‌شود و هر فرضیه‌ای که P(D|h) را ماکزیمم کند (ML)، hML نامیده می‌شود.

برای مشخص‌شدن رابطه با مسائل یادگیری ماشین، ابتدا قضیهٔ بیز را باتوجه‌به نمونه‌های D و فضای فرضیه‌ای H معرفی کردیم. در واقع قضیهٔ بیز کلی‌تر از آنچه در بالا گفته شد است. از قضیهٔ بیز می‌توان برای هر زیرمجموعهٔ H که ناسازگارند استفاده کرد. در این فصل، در اکثر موارد فرض خواهیم کرد که H فضای فرضیه‌ای که تابع هدف را شامل می‌شود است و D نمونه‌های آموزشی هستند. در مواقع دیگر فرض می‌کنیم که H مجموعهٔ دیگر ناسازگاری با یکدیگر از فرضیه‌هاست و D نیز مجموعهٔ دیگری از داده‌هاست.

بررسی بهینگی پرتفوی

روش‌های مختلفی برای بررسی بهینه بودن سبد وجود دارد. در این پژوهش از معیار ترینر[1]جهت بررسی بهینگی سبد سهام استفاده شده است. این روش که به معیار بازده به نوسان پذیری نیز معروف است، رابطه را برای پرتفوی موردنظر و بازار محاسبه کرده و به مقایسه آن‌ها می‌پردازد اگر این مقدار برای پرتفوی موردنظر از مقدار محاسبه شده برای بازار بیشتر باشد، نشان‌دهنده بهینگی نسبی پرتفوی است. این معیار به شکل رابطه زیر است:

$\frac{R_p - R_f}{\beta_p}$ رابطه (۱۶)

$R_i = R_f + \beta_i(R_m - R_f)$ رابطه (۱۷)

$\beta_i = \frac{R_i - R_f}{R_m - R_f}$ رابطه (۱۸)

$R_p = \frac{\sum_{i=1}^{n} R_i}{n}$ رابطه (۱۹)

1. Treyner

ارزش در معرض ریسک و ارزش در معرض ریسک شرطی تحت نظریه اعتبار

در این بخش که قسمتی از نوآوری تحقیق می‌باشد، سنجه ریسک‌های ارزش در معرض ریسک و ارزش در معرض ریسک شرطی با نظریه فازی ادغام شده و برای عمل بهینه‌سازی و ساخت سبد سهام مورداستفاده قرار گرفته است. باتوجه‌به غیرقطعی بودن داده‌ها، استفاده از روش‌های فازی باعث دقت بالاتری در مدل‌سازی می‌شود. همچنین استفاده از سنجهٔ ریسک CVaR به این‌که اندازه ضرر را به سرمایه‌گذار نشان می‌دهد، به تصمیم‌گیری بهتر کمک می‌کند. در این پژوهش با استفاده از سنجه ارزش در معرض ریسک مشروط و تخمین آن به‌وسیله نظریه اعتبار فازی، اقدام به بهینه‌سازی سبد سهام شده است. به این منظور، بازده انتظاری پرتفوی به‌وسیله میانگین اعتبار فازی به‌دست‌آمده و سپس ارزش در معرض ریسک مشروط به‌وسیله همین نظریه تخمین زده شده است. محدودیت‌های در نظر گرفته شده برای مدل بهینه‌سازی، شامل محدودیت کف و سقف نسبت خرید هر سهم که حداقل آن صفر و حداکثر آن ۷۰ درصد است، همچنین محدودیت حداقلی از بازده می‌باشد.

رابطه (۲۰)
$$Var(\alpha) = -\varphi^{-1}(\alpha) = \begin{cases} 2(r_1 - r_2)\alpha - r_2 & \alpha \le 0.5 \\ 2(r_2 - r_3)\alpha + r_3 - 2r_2 & \alpha > 0.5 \end{cases}$$

رابطه (۲۱)
$$Min - \sum_{i=1}^{n}[VaR(\alpha)]x_i$$
$$s.t: \frac{1}{4}\sum_{i=1}^{n}[r_1 + 2r_2 + r_3]x_i \ge \mu_p$$
$$\sum_{j=1}^{n} x_j = 1 \qquad j = 1.2.3. \ldots . n$$

رابطه (۲۲)
$$CVar(\alpha) = \frac{1}{1-\alpha}\int_{\alpha}^{1} Var(r)dr = \begin{cases} \alpha r_1 - (1+\alpha)r_2 & \alpha \le 0.5 \\ (\alpha - 1)r_2 - \alpha r_3 & \alpha > 0.5 \end{cases}$$

رابطه (۲۳)
$$Min - \sum_{i=1}^{n}[CVaR(\alpha)]x_i$$
$$s.t: \ \frac{1}{4}\sum_{i=1}^{n}[r_1 + 2r_2 + r_3]x_i \ge \mu_p$$
$$\sum_{j=1}^{n} x_j = 1 \qquad j = 1.2.3. \ldots . n$$
$$l_i y_i \le x_i \le u_i y_i \qquad y_i = \{0.1\}$$
$$x_i \ge 0$$

٤- تجزیه‌وتحلیل داده‌ها

در ابتدا، اطلاعات ۵ساله هریک از سهام انتخابی (بازدهی روزانه آن‌ها) با کمک دو الگوریتم ماشین بردار پشتیبان و بیز ساده به‌منظور کلاس‌بندی، استفاده می‌شود. این کلاس‌بندی طبق ۱۴ ویژگی[1] بنیادی و تکنیکال می‌باشد. در فاز بعد، کلاس مثبت هر یک از الگوریتم‌ها، به‌وسیله الگوریتم جنگل

1. Feature

تصادفی[1] پیش‌بینی می‌شود. پیش‌بینی برای هر سهم و هر خروجی الگوریتم، ۳ مرتبه صورت گرفت و علت آن نیز استفاده از نظریه فازی و ترکیب آن با سنجه‌های ریسک ارزش در معرض خطر و ارزش در معرض خطر شرطی برای تشکیل پرتفوی می‌باشد. در پایان، خروجی‌های هر الگوریتم یک‌مرتبه به‌وسیله سنجه ریسک VAR و یک‌مرتبه نیز به‌وسیله سنجه ریسک CVAR بهینه خواهد شد و در پایان خروجی‌ها با یکدیگر قیاس می‌شوند. برای بهینه‌سازی سبد سهام به‌وسیله دو سنجه ریسک مطرح شده، از دو سطح اطمینان ۹۰ و ۹۵ درصد استفاده شده و در مرحله آخر سبدهای تشکیل شده با واقعیت بازار و استفاده از معیار ترینر برای بازدهی آن، مقایسه می‌شوند.

شرکت‌های مورد بررسی در پژوهش

فرآیند انتخاب شرکت‌های مورد بررسی در این تحقیق، با رویکرد تنوع صنایع مختلف بورس و همچنین عمق بازار این سهم‌ها (حجم معاملات هریک از شرکت‌ها در صنعت خود) می‌باشد. در این راستا طبق گزارش‌های سایت رسمی سازمان بورس اوراق بهادار مبنی‌بر حجم معاملات صنایع مختلف بورسی، صنایع "خودرو و ساخت قطعات"، "فلزات اساسی"، "بانک‌ها و مؤسسات اعتباری"، "فرآورده-های نفتی، کک و سوخت هسته‌ای"، "بیمه و صندوق بازنشستگی به‌جز تأمین اجتماعی"، "محصولات شیمیایی"، "سرمایه‌گذاری‌ها"، "مواد و محصولات دارویی"، "سیمان، آهک و گچ" و چند صنعت دیگر با هدف بهره‌گیری تقریبی از جمیع صنایع ازیک‌طرف و نتیجه‌گیری قابل‌اتکا از طرف دیگر انتخاب شده است. بعد از انتخاب صنایع پرطرفدار معاملاتی از طرف معامله‌گران، شرکت‌های مربوط به هر گروه، از طریق نظرات خبرگان و روش سلسله‌مراتبی شرکت‌ها انتخاب شده و نتایج در قالب جدول (۲) ارائه شده است.

جدول (۲) شرکت‌های مورد بررسی در پژوهش

Table (2) companies examined in the research

ردیف	نماد	شرکت	ردیف	نماد	شرکت
۱	اخابر	مخابرات ایران	۱۶	زاگرس	پتروشیمی زاگرس
۲	همراه	شرکت ارتباطات سیار ایران	۱۷	شبریز	پالایشگاه نفت تبریز
۳	وسینا	بانک سینا	۱۸	شبندر	پالایشگاه نفت بندرعباس
۴	وبملت	بانک ملت	۱۹	ولساپا	لیزینگ رایان سایپا
۵	پارسیان	بیمه پارسیان	۲۰	وایران	لیزینگ ایرانیان
۶	آسیا	بیمه اسیا	۲۱	وصنعت	سرمایه‌گذاری توسعه صنعت و تجارت
۷	خبهمن	گروه بهمن	۲۲	وخارزم	سرمایه‌گذاری خوارزمی

1. Random Forest

۸	خرینگ	رینگ‌سازی مشهد	۲۳	ومعادن	توسعه معادن و فلزات
۹	خزامیا	زامیاد	۲۴	ثفارس	عمران و توسعه فارس
۱۰	فولاد	فولاد مبارکه اصفهان	۲۵	رکیش	کارت اعتباری ایران کیش
۱۱	فملی	ملی صنایع مس ایران	۲۶	فاراک	ماشین‌سازی اراک
۱۲	سغرب	سیمان غرب	۲۷	دارو	کارخانجات داروپخش
۱۳	سکرد	سیمان کردستان	۲۸	قزوین	کارخانجات قند قزوین
۱۴	شفن	پتروشیمی فناوران	۲۹	غشاذر	پگاه آذربایجان غربی
۱۵	کرماشا	صنایع پتروشیمی کرمانشاه	۳۰	حکشتی	کشتی‌رانی جمهوری اسلامی ایران

ویژگی‌های مورد بررسی در الگوریتم‌های مورداستفاده

در این قسمت از پژوهش، ویژگی‌ها استفاده شده در الگوریتم‌های مورد بررسی ارایه شده است. این ویژگی‌ها ترکیبی از شاخص‌های تکنیکال و بازدهی های مختلف است. در جدول (۳) و (۴) ویژگی-های مورد بررسی در الگوریتم‌ها ذکر شده است.

جدول (۳) جدول راهنمای علائم

Table (3) Guide table of symptoms

ردیف	متغیر–پارامتر	توضیحات	ردیف	متغیر–پارامتر	توضیحات
۱	EMA	میانگین متحرک نمایی	۸	OBV	اندیکاتور حجم و قیمت
۲	ADX	شاخص میانگین حرکت جهت-دار	۹	TRIX	اندیکاتور نواحی اشباع خرید/فروش
۳	RSI	شاخص قدرت نسبی	۱۰	R(-i)	بازدهی قیمتی روز گذشته i نسبت به یک روز قبل آن
۴	Stoch	شاخص آماری	۱۱	Close	قیمت پایانی
۵	MACD	واگرایی/هم‌گرایی میانگین متحرک	۱۲	High	بیشترین قیمت
۶	CCI	اندیکاتور تغییرات قیمت در بازار	۱۳	Low	کمترین قیمت
۷	ATR	اندیکاتور اندازه‌گیری نوسانات قیمت	۱۴	Open	قیمت بازگشایی

جدول (۴) ویژگی‌های مورد بررسی در الگوریتم‌ها

Table (4) characteristics investigated in algorithms

ردیف	ویژگی	پارامترهای مورد نیاز	فرمول	منبع
۱	EMA	close (i,i-1,i-2,…)	$\left(\text{Value}_i * (\frac{\text{smoothing}}{1+\text{days}})\right) + \text{EMA}_{i-1}(1-(\frac{\text{smoothing}}{1+\text{days}}))$	https://www.cse.ust.hk
۲	ADX	DI^+, DI^-, ADX	$\text{TR}_{i+1} = \text{TR}_{i-1} - \frac{\text{TR}_{i-1}}{14} + \text{TR}_i$	https://dl.abcbourse.ir
۳	RSI	Profit , loss	$100 - \frac{100}{1+\text{RS}}$	https://cse.hkust.edu.hk
۴	Stoch	Close , low , high	$100 * \frac{C - L_{14}}{H_{14} - L_{14}}$	https://chartnexus.com
۵	MACD	EMA	$\text{EMA}(10) - \text{EMA}(20)$	https://core.ac.uk/
۶	CCI	High , low , close	$\frac{\text{price} - \text{ma}}{0.015 * \text{d}}$	https://www.courtneycapital.co.za
۷	ATR	High , low	$H - L$	https://www.fidelity.com
۸	OBV	Volume	$\text{OBV(t)} = \text{OBV(t-1)} \pm \text{Vol}$	https://www.mql5.com
۹	TRIX	EMA	EMA(a. b. c)	https://www.ensignsupport.org
۱۰	R(-1)	Close i , close i-1	$\ln(\frac{\text{close i}}{\text{close i}-1})$	بازدهی روزهای مختلف نسبت به‌هم
۱۱	R(-2)	Close i-1 , close i-2	$\ln(\frac{\text{close i}-1}{\text{close i}-2})$	
۱۲	R(-3)	Close i-2 , close i-3	$\ln(\frac{\text{close i}-2}{\text{close i}-3})$	
۱۳	R(-4)	Close i-3 , close i-4	$\ln(\frac{\text{close i}-3}{\text{close i}-4})$	
۱۴	R(-5)	Close i-4 , close i-5	$\ln(\frac{\text{close i}-4}{\text{close i}-5})$	

در این راستا در جدول (۵) ویژگی های مدل‌های مورداستفاده در پژوهش به‌صورت خلاصه قابل ملاحظه می‌باشد.

جدول (۵) خلاصه ویژگی‌های مدل مورداستفاده

Table (5) Summary of the features of the used model

۱	الگوریتم‌های مورداستفاده در مرحله طبقه‌بندی	ماشین بردار پشتیبان و بیز ساده
۲	الگوریتم مورداستفاده در مرحله پیش‌بینی	جنگل تصادفی
۳	سنجه ریسک مورداستفاده در مرحله بهینه‌سازی	VaR و CVaR
۴	سطح اطمینان مورداستفاده برای پیاده‌سازی دو سنجه ریسک	۹۹ ٪ و ۹۵ ٪

ماشین بردار پشتیبان

در این بخش، هدف آن است که شرکت‌های معرفی‌شده به کمک الگوریتم ماشین بردار پشتیبان طبقه‌بندی شوند. خروجی موردنظر بر اساس کلاس مثبت و منفی در قالب جدول ۶ ارائه می‌گردد. ملاحظه می‌شود ۸ سهم طبق این الگوریتم، مجوز حضور در کلاس مثبت را اخذ کرده‌اند.

در جدول (۶) خروجی الگوریتم ماشین بردار پشتیبان ارائه شده است.

جدول (۶) خروجی الگوریتم ماشین بردار پشتیبان

Table (6) output of support vector machine algorithm

	اخابر	همراه	وسینا	وبملت	پارسیان	آسیا	خبهمن	خرینگ	خزامیا	فولاد
کلاس	+	−	−	−	−	−	+	−	−	−
دقت الگوریتم	٪ ۷۶	٪ ۸۲	٪ ۷۰	٪ ۸۲	٪ ۹۱	٪ ۷۳	٪ ۸۸	٪ ۸۶	٪ ۷۶	٪ ۷۳
	فملی	سغرب	سکرد	شفن	کرماشا	زاگرس	شبریز	شبندر	ولساپا	وایران
کلاس	−	−	−	−	+	−	+	+	−	−
دقت الگوریتم	٪ ۸۹	٪ ۸۴	٪ ۸۲	٪ ۷۷	٪ ۷۶	٪ ۹۴	٪ ۸۱	٪ ۶۸	٪ ۷۴	٪ ۸۶
	وصنعت	وخارزم	ومعادن	ثفارس	رکیش	فاراک	دارو	قزوین	غشاذر	حکشتی
کلاس	−	+	−	−	−	−	−	+	−	+
دقت الگوریتم	٪ ۹۲	٪ ۷۹	٪ ۷۸	٪ ۸۷	٪ ۸۲	٪ ۸۳	٪ ۷۴	٪ ۷۷	٪ ۷۶	٪ ۸۲

بیز ساده

در این بخش، هدف آن است که شرکت‌های معرفی‌شده به‌کمک الگوریتم بیز ساده طبقه‌بندی شوند. خروجی موردنظر بر اساس کلاس مثبت و منفی در قالب جدول (۷) ارائه می‌گردد. ملاحظه می‌شود ۱۲ سهم طبق این الگوریتم، مجوز حضور در کلاس مثبت را اخذ کرده‌اند.

در جدول (۷) خروجی الگوریتم بیز ساده ارائه شده است.

جدول (۷) خروجی الگوریتم بیز ساده

Table (7) output of simple Bayes algorithm

	اخابر	همراه	وسینا	وبملت	پارسیان	آسیا	خبهمن	خرینگ	خزامیا	فولاد
کلاس	−	+	−	−	−	−	+	−	+	−
دقت الگوریتم	٪ ۷۸	٪ ۷۶	٪ ۷۴	٪ ۷۴	٪ ۸۲	٪ ۸۱	٪ ۸۲	٪ ۹۴	٪ ۷۸	٪ ۷۳

	فملی	سغرب	سکرد	شفن	کرماشا	زاگرس	شبریز	شبندر	ولساپا	وایران
کلاس	−	+	+	−	−	−	+	+	−	−
دقت الگوریتم	٪ ۷۶	٪ ۸۰	٪ ۷۸	٪ ۸۰	٪ ۹۵	٪ ۸۴	٪ ۹۴	٪ ۹۱	٪ ۸۳	٪ ۸۷
	وصنعت	وخارزم	ومعادن	ثفارس	رکیش	فاراک	دارو	قزوین	غشاذر	حکشتی
کلاس	−	+	−	−	+	−	+	+	+	−
دقت الگوریتم	٪ ۸۷	٪ ۹۲	٪ ۸۵	٪ ۷۴	٪ ۸۴	٪ ۸۲	٪ ۷۱	٪ ۷۰	٪ ۷۴	٪ ۷۳

دقت الگوریتم‌ها مستخرج از فرآیند یادگیری آن‌ها در دوره آموزش[1] و آزمودن عملکرد آن‌ها در مرحله آزمون[2] است. در تحقیق حاضر، از دو تابع کرنل سیگوئید و گاوسی که بیشترین انطباق را در بین سه تابع کرنل چندجمله‌ای،[3] گاوسی[4] و سیگموئید[5] با نمونه‌ها داشته‌اند، استفاده شده است.

پیش‌بینی با استفاده از الگوریتم جنگل تصادفی

در این بخش کلیه خروجی‌های الگوریتم‌ها، که برای طبقه‌بندی استفاده شده‌اند، به‌وسیله الگوریتم جنگل تصادفی در جدول (۸) و (۹) پیش‌بینی می‌شوند. استفاده از نظریه فازی و سه پیش‌بینی خوش-بینانه، محتمل و بدبینانه باعث نوآوری جدیدی در این پژوهش شده است.

جدول (۸) پیش‌بینی بازده براساس خروجی الگوریتم ماشین بردار پشتیبان

Table (8) prediction based on the output of the support vector machine algorithm

ردیف	سهم	r_1	r_2	r_3
۱	اخابر	٪۰/۴	٪۰/۹۶	٪۲/۱۱
۲	خبهمن	−٪۰/۹	٪۲/۲۴	٪۲/۷۸
۳	کرماشا	−٪۱	۰	٪۰/۴۴
۴	شبریز	٪۱/۲۵	٪۱/۸۶	٪۲/۲۸
۵	شبندر	−٪۲/۱۳	−٪۰/۹	٪۰/۶۲
۶	وخارزم	٪۱/۲۴	٪۱/۳	٪۱/۸۴
۷	قزوین	٪۰/۹	٪۱/۵۷	٪۲/۳۴
۸	حکشتی	٪۰/۴۵	٪۱/۲۵	٪۲/۱

1. Train
2. Test
3. Polynominal
4. Gaussian
5. Sigmoid

جدول (۹) پیش‌بینی بازده براساس خروجی الگوریتم بیز ساده

Table (9) prediction based on the output of simple Bayes algorithm

ردیف	سهم	r_1	r_2	r_3
۱	همراه	٪ ۰/۶	٪ ۰/۹	٪ ۱/۷۸
۲	خبهمن	٪ ۰/۱	٪ ۱/۸۴	٪ ۲/۲۳
۳	خزامیا	-٪ ۰/۷	۰	٪ ۱/۹۲
۴	سغرب	٪ ۰/۴	٪ ۰/۸	٪ ۲/۱۳
۵	سکرد	٪ ۰/۰۴	٪ ۰/۹۲	٪۱
۶	شبریز	٪ ۰/۷۵	٪ ۲/۱	٪ ۲/۹۹
۷	شبندر	٪ ۱/۲	٪ ۱/۵۵	٪ ۲/۳۴
۸	وخارزم	-٪ ۱/۸۴	-٪ ۰/۲۲	٪ ۱/۴۶
۹	رکیش	٪ ۰/۹۶	٪ ۲/۲۳	٪ ۲/۷۷
۱۰	دارو	-٪ ۰/۹۲	٪ ۰/۲	٪۲
۱۱	قزوین	-٪ ۰/۶	٪ ۱/۵۸	٪ ۱/۸۸
۱۲	غشاذر	٪۰/۸۲	٪ ۱/۱۲	٪ ۱/۹۲

جداول بهینه‌سازی خروجی الگوریتم‌ها

در جدول (۱۰) و (۱۱) مقادیر بهینه‌سازی خروجی الگوریتم ماشین بردار پشتیبان، با استفاده از ارزش در معرض ریسک، ارایه شده است.

جدول (۱۰) جدول بهینه‌سازی خروجی الگوریتم ماشین بردار پشتیبان

Table (10) optimization table of support vector machine algorithm output

ردیف	متغیر	نماد	$W(\alpha_1 = 0.05)$	$W(\alpha_2 = 0.01)$
۱	X_1	اخابر	۰/۱۱	۰
۲	X_2	خبهمن	۰	۰/۱۳
۳	X_3	کرماشا	۰	۰
۴	X_4	شبریز	۰	۰
۵	X_5	شبندر	۰/۴۲	۰/۱۴
۶	X_6	وخارزم	۰/۳۲	۰/۵۱
۷	X_7	قزوین	۰/۰۱۸	۰
۸	X_8	حکشتی	۰/۱۳۲	۰/۲۲

$$R_{Index} = w_1R_1 + w_2R_2 + \cdots + w_nR_n$$ رابطه (۲۴)

جدول (۱۱) مقایسه سطح‌اطمینان‌های متفاوت با بازده واقعی درصورت وجود وزن‌دهی پیشنهادی

Table (11) comparison of different assurance levels with actual efficiency in case of proposed weighting

$\alpha_2 = 0.01$	$\alpha_1 = 0.05$	
٪۱۷/۹	٪۲۵/۳	R_{Index}

در جدول (۱۲) و (۱۳) مقادیر بهینه‌سازی خروجی الگوریتم بیز ساده با استفاده از ارزش در معرض ریسک، ارایه شده است.

جدول (۱۲) جدول بهینه‌سازی خروجی الگوریتم بیز ساده

Table (12) of the output optimization table of the simple Bayes algorithm

ردیف	متغیر	نماد	$W(\alpha_1 = 0.05)$	$W(\alpha_2 = 0.01)$
۱	X_1	همراه	۰/۱۳	۰
۲	X_2	خبهمن	۰	۰
۳	X_3	خزامیا	۰/۰۲۴	۰
۴	X_4	سغرب	۰	۰/۱۲
۵	X_5	سکرد	۰/۳۳	۰/۲۴
۶	X_6	شبریز	۰	۰
۷	X_7	شبندر	۰/۱۸	۰/۴۲
۸	X_8	وخارزم	۰	۰
۹	X_9	رکیش	۰	۰/۲۲
۱۰	X_{10}	دارو	۰/۲۲	۰
۱۱	X_{11}	قزوین	۰	۰
۱۲	X_{12}	غشاذر	۰/۱۱۶	۰

جدول (۱۳) مقایسه سطح‌اطمینان‌های متفاوت با بازده واقعی درصورت وجود وزن‌دهی پیشنهادی

Table (13) comparison of different assurance levels with actual efficiency in case of proposed weighting

$\alpha_2 = 0.01$	$\alpha_1 = 0.05$	
٪۱۴/۳	٪۲۱/۴	R_{Index}

در جدول (۱۴) و (۱۵) مقادیر بهینه‌سازی خروجی الگوریتم ماشین بردار پشتیبان، با استفاده از ارزش در معرض ریسک شرطی، ارایه شده است.

جدول (۱۴) جدول بهینه‌سازی خروجی الگوریتم ماشین بردار پشتیبان

Table (14) optimization table of support vector machine algorithm output

ردیف	متغیر	نماد	$W(\alpha_1 = 0.05)$	$W(\alpha_2 = 0.01)$

۱	X_1	اخابر	۰/۱۲	۰
۲	X_2	خبهمن	۰	۰/۲۳
۳	X_3	کرماشا	۰/۲۴	۰/۳۸
۴	X_4	شبریز	۰/۰۸	۰
۵	X_5	شبندر	۰/۳۴	۰
۶	X_6	وخارزم	۰/۰۶	۰/۲۸
۷	X_7	قزوین	۰	۰
۸	X_8	حکشتی	۰/۱۶	۰/۱۱

جدول (۱۵) مقایسه سطح‌اطمینان‌های متفاوت با بازده واقعی درصورت وجود وزن‌دهی پیشنهادی

Table (15) comparison of different assurance levels with actual efficiency in case of proposed weighting

	$\alpha_1 = 0.05$	$\alpha_2 = 0.01$
R_{Index}	٪۳۱/۲	٪۲۸/۷

در جدول (۱۶) و (۱۷) مقادیر بهینه‌سازی خروجی الگوریتم بیز ساده با استفاده از ارزش در معرض ریسک شرطی، ارایه شده است.

جدول (۱۶) جدول بهینه‌سازی خروجی الگوریتم بیز ساده

Table (16) of the optimization table of the output of the simple Bayes algorithm

ردیف	متغیر	نماد	$W(\alpha_1 = 0.05)$	$W(\alpha_2 = 0.01)$
۱	X_1	همراه	۰/۱۳	۰
۲	X_2	خبهمن	۰	۰
۳	X_3	خزامیا	۰/۴۴	۰
۴	X_4	سغرب	۰	۰/۲۴
۵	X_5	سکرد	۰/۱	۰/۱
۶	X_6	شبریز	۰	۰
۷	X_7	شبندر	۰/۱۸	۰/۳
۸	X_8	وخارزم	۰	۰
۹	X_9	رکیش	۰	۰/۱۴
۱۰	X_{10}	دارو	۰/۱۵	۰
۱۱	X_{11}	قزوین	۰	۰
۱۲	X_{12}	غشاذر	۰	۰

جدول (۱۷) مقایسه سطح‌اطمینان‌های متفاوت با بازده واقعی درصورت وجود وزن‌دهی پیشنهادی

Table (17) comparison of different assurance levels with actual efficiency in case of proposed weighting

R_{Index}	$\alpha_1 = 0.05$	$\alpha_2 = 0.01$
٪۱۴/۲	٪۱۶/۱	

در پایان به بررسی نتایج حاصل از کاربرد مدل‌های بررسی‌شده و محدودیت‌های ناشی‌از آن‌ها خواهیم پرداخت. در این تحقیق بازدهی بدون ریسک سالیانه ۲۳ درصد در نظر گرفته شده که برای یک ماه ۹/۱ درصد برآورد شد و در مقایسات استفاده می‌شود. ازطرفی بازده ماهانه بازار (شاخص کل بورس) در سال ۱۴۰۲ بین ماه‌های فروردین و اردیبهشت، ۱۷ درصد می‌باشد. در جدول (۱۸) خروجی-های موردنظر از الگوریتم‌ها و همچنین سنجه‌های ریسک، همچنین سطوح‌اطمینان مختلف رتبه‌بندی شده تا دیدی مشخص را به پژوهش‌گران انتقال دهد. پر واضح است که سنجه ریسک ارزش در معرض خطر شرطی قابلیت و نتیجه بهتری را نسبت به سنجه ریسک ارزش در معرض خطر داشته است. همچنین الگوریتم ماشین بردار پشتیبان، نتایج بهتری را رقم زده است. در تحلیل نتایج، بازدهی واقعی سهام موردبررسی با وزن خروجی از الگوریتم ترکیب‌شده و رتبه‌بندی صورت گرفته است. درآخر بازدهی هرکدام از روش‌ها با معیار ترینر نسبت به بازده بازار قیاس شده و به‌ترتیب در جدول (۱۸) آورده شده است.

جدول (۱۸) رتبه‌بندی الگوریتم‌ها و سطح‌اطمینان‌های مختلف به تفکیک بازدهی

Table (18) ranking of different algorithms and assurance levels by efficiency

ردیف	بازدهی	الگوریتم	سنجه ریسک	سطح اطمینان	$R_p - R_f$	β_p	ترینر پرتفوی	ترینر بازار
۱	٪ ۳۱/۲	ماشین بردار پشتیبان	CVAR	α_1	٪ ۲۹/۳	۱/۸۲	٪ ۱۶/۱	٪ ۷/۳۸
۲	٪ ۲۸/۷	ماشین بردار پشتیبان	CVAR	α_2	٪ ۲۶/۸	۱/۸۲	٪ ۱۴/۷	٪ ۷/۳۸
۳	٪ ۲۵/۳	ماشین بردار پشتیبان	VAR	α_1	٪ ۲۳/۴	۱/۸۲	٪ ۱۲/۸	٪ ۷/۳۸
۴	٪ ۲۱/۴	بیز ساده	VAR	α_1	٪ ۱۹/۵	۲/۱۳	٪ ۹/۱۵	٪ ۷/۳۸
۵	٪ ۱۷/۹	ماشین بردار پشتیبان	VAR	α_2	٪ ۱۶	۱/۸۲	٪ ۸/۷۹	٪ ۷/۳۸
۶	٪ ۱۶/۱	بیز ساده	CVAR	α_1	٪ ۱۴/۲	۲/۱۳	٪ ۶/۶۶	٪ ۷/۳۸
۷	٪ ۱۴/۳	بیز ساده	VAR	α_2	٪ ۱۲/۴	۲/۱۳	٪ ۵/۸۲	٪ ۷/۳۸
۸	٪ ۱۴/۲	بیز ساده	CVAR	α_2	٪ ۱۲/۳	۲/۱۳	٪ ۵/۷۷	٪ ۷/۳۸

۵- بحث و نتیجه‌گیری

بر اساس نتایج بدست آمده سنجه ریسک ارزش در معرض خطر شرطی قابلیت و نتیجه بهتری را نسبت به سنجه ریسک ارزش در معرض خطر داشته است. همچنین الگوریتم ماشین بردار پشتیبان، نتایج بهتری را رقم زده‌اند. همان‌طور که می‌بینیم در ۵ سطر ابتدایی جدول ۱۸، روش مطرح شده در پژوهش توانسته بازدهی بالاتری را نسبت به شاخص کل به‌عنوان یک مرجع معتبر رشد بازار، کسب کند. سطر اول این جدول بدان معانست که اگر معامله‌گر برای طبقه‌بندی سهام خود از روش ماشین بردار پشتیبان استفاده می‌کرد و سپس با سطح اطمینان ۹۵ درصد اقدام به بهینه‌سازی سبد سهام از طریق سنجه ریسک ارزش در معرض خطر شرطی داشته داشت، می‌توانست حدود ۱۴ درصد بازدهی بیشتری کسب کند، تا حالتی‌که سهام موجود در شاخص کل را در پرتفو خود نگهداری می‌کرد. در سه سطر آخر نیز مشاهده می‌شود که الگوریتم‌ها نتوانسته‌اند بازدهی بهتری نسبت به شاخص کل کسب کنند. سه سطر آخر وضعیت غیر بهینه را نشان می‌دهند، یعنی وضعیتی‌که پرتفوی تشکیل شده نتیجه بدتری نسبت به بازدهی شاخص کل داشته است. این بدان معناست که به‌عنوان مثال اگر معامله‌گری در مرحله طبقه‌بندی سهام از الگوریتم بیز ساده استفاده کند و در مرحله بهینه‌سازی سبد سهام، مسئله را با سنجه ریسک ارزش در معرض ریسک شرطی با سطح اطمینان ۹۵ درصد یا ۹۹ درصد حل کند، بازدهی مطلوبی کسب نکرده و اگر به‌جای آن سهام‌های تشکیل دهنده شاخص کل را بخرد، بازدهی بهتری کسب خواهد کرد. باتوجه به روند صعودی بازار بورس از انتهای سال ۱۳۹۸ تا مردادماه سال ۱۳۹۹ و پس‌از آن روند کاملا نزولی بازار، به‌نظر می‌رسد الگوریتم‌ها نمره قابل‌قبولی را باتوجه به هیجانی بودن بازار در برخی از مقاطع زمان بگیرند. در این پژوهش سعی شد یاگیری ماشین، نظریه فازی و هم‌زمان با آن دو سنجه ریسک مهم را دخالت دهد تا بتواند نتایج را قابل اعتمادتر درنظر بگیرد، درحالیکه در اکثر مطالعات هریک از موارد ذکر شده به‌تنهایی ملاک پژوهش قرار گرفتند.

پیشنهادهای مناسبی را می‌توان ارائه داد که در ادامه برخی از آن‌ها مطرح می‌شود:

استفاده از انواع کرنل‌ها و پارامترهای آن، به‌منظور شناخت بهتر دادها در ابعاد بالاتر و یادگیری دقیق‌تر الگوریتم و به‌تبع آن طبقه‌بندی مناسب‌تر.

استفاده از دیگر محدودیت‌های کاربردی، مانند هزینه‌های معاملاتی در مرحله بهینه‌سازی سبد سهام.

استفاده از هوش مصنوعی در مرحله طبقه‌بندی، به‌عنوان رقیبی جدی برای الگوریتم‌های یادگیری ماشین و بررسی عملکرد هریک از آن‌ها.

به‌کارگیری دیگر اوراق بهادار و قراردادهای مالی در ساخت پرتفوی بهینه همچون قراردادهای اختیار معامله و قراردادهای آتی.

در فرآیند اجرای پژوهش حاضر محدودیت‌هایی وجود داشته که به ذکر چندی از آن‌ها اشاره خواهد شد:

محدودیت در جمع‌آوری داده‌های آماری. بسیاری از شرکت‌های بازار سرمایه قدمت چندانی نداشته و به‌تبع استفاده از الگوریتم‌های یادگیری ماشین را با چالشی جدی مواجه می‌کرد.

محدودیت در انتخاب الگوریتم‌ها و پارامترهای آن. انتخاب الگوریتم، مرحله بسیار جدی در شروع مقاله حاضر بوده و سعی شده از الگوریتم‌هایی بهره‌گیری شود که در اکثر مواقع نتایج قابل‌قبولی به همراه داشته باشند. طبیعی است تعداد این الگوریتم‌ها بالا بوده و ممکن است الگوریتمی دیگر نتایج مطلوب‌تری را پیشنهاد دهد.

محدودیت بهینه‌سازی. در این فاز، دو محدودیت اساسی وجود داشته و آن نرمال یا شبه نرمال فرض‌شدن توزیع داده‌ها بوده. زیاد دو سنجه ریسک استفاده شده، مسئله را با فرض نرمال بودن داده‌ها حل می‌کنند. ازطرفی محاسبه ریسک از طریق این دو سنجه، با سطح اطمینان مشخصی صورت‌گرفته که خیلی هم دقیق نمی‌باشد. عدم استفاده از مسئله‌ای دوهدفه در مدل، محدودیت دیگر بوده که ما را از جواب‌هایی دقیق‌تر بازداشته است.

٦- تعارض منافع

هیچ‌گونه تعارض منافع در این پژوهش وجود ندارد.

۷- منابع

Beale, R., & Jackson, T. (2007). Introduction to Neural Networks. (M. Alborzi, Trans.). *Tehran: Sharif University of Technology, Scientific Publications Institute*. [In Persian]

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.

Bi, J., & Zhang, T. (2004). Support vector classification with input data uncertainty. *Advances in neural information processing systems, 17*.

Bishop, C. M., & Tipping, M. (2013). Variational relevance vector machines. *arXiv preprint arXiv:1301.3838*.

Bleuler, S., Brack, M., Thiele, L., & Zitzler, E. (2001). Multiobjective genetic programming: Reducing bloat using SPEA2. In *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No. 01TH8546)* (Vol. 1, pp. 536-543). IEEE.

Chowdhury, R., Mahdy, M. R. C., Alam, T. N., Al Quaderi, G. D., & Rahman, M. A. (2020). Predicting the stock price of frontier markets using machine learning and modified Black–Scholes Option pricing model. *Physica A: Statistical Mechanics and its Applications*, *555*, 124444.

Ebadati, O. M., Jafari, M. A., & Davoodifar, N. (2022). Forecasting Stocks in the Financial Market by Using GA-SVM Hybrid Algorithm. *Advances in Finance and Investment*, *2*(5), 1-22. [In Persian]

Farahabadi, M., Eyvazloo, R., Safari, H. (2022). An Artificial Neural Network and Bayesian Network model for liquidity risk assessment in banking. *Journal of Securities Exchange*, *15*(59), 121-156. [In Persian]

Fazlzadeh, A., Ranjpour, R., Tohidi, R. (2012). The Investigation of the Ability of Single-Index Sharp and DEA Models for Choosing Efficient Portfolio in Tehran Stock Exchange. *Journal of Securities Exchange*, *5*(18), 39-59. [In Persian]

Feizi, J., & Foroush Bastani, A. (2012). Investigating Monte Carlo Methods for Approximating Value at Risk (VaR) and Conditional Value at Risk (CVaR). In Proceedings of the Third Conference on Financial Mathematics and Applications. *Zanjan University, Faculty of Mathematics, Department of Financial Mathematics.* [In Persian]

Goykhman, M., & Teimouri, A. (2018). Machine learning in sentiment reconstruction of the simulated stock market. *Physica A: Statistical Mechanics and its Applications*, *492*, 1729-1740.

Haratizadeh, S., & Rezaee, F. (2023). A novel machine learning approach for portfolio optimization. *Journal of Decisions and Operations Research*, *8*(2), 527-539. [In Persian]

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.

Huang, C. F. (2012). A hybrid stock selection model using genetic algorithms and support vector regression. *Applied Soft Computing*, *12*(2), 807-818.

Lai, K. K., Yu, L., Wang, S., & Zhou, C. (2006). A double-stage genetic optimization algorithm for portfolio selection. In *International Conference on Neural Information Processing* (pp. 928-937). Berlin, Heidelberg: Springer Berlin Heidelberg.

Lin, C. M., & Gen, M. (2007). An effective decision-based genetic algorithm approach to multiobjective portfolio optimization problem. *Applied mathematical sciences*, *1*(5), 201-210.

Linoff, G. S., & Berry, M. J. (2011). *Data mining techniques: for marketing, sales, and customer relationship management.* John Wiley & Sons.

Paiva, F. D., Cardoso, R. T. N., Hanaoka, G. P., & Duarte, W. M. (2019). Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. *Expert Systems with Applications*, *115*, 635–655.

Paula, A., & Vitor Venceslau Curtis. (2023). A Model-Based Approach Machine Learning to Scalable Portfolio Selection. *International Journal on Cybernetics & Informatics*, *12*(3), 23–40.

Qiu, H., Han, F., Liu, H., & Caffo, B. (2015). Robust portfolio optimization. *Advances in Neural Information Processing Systems*, 28.

Raei, R., & Beigi, H.A. (2010). Portfolio optimization using particle swarm optimization method. *Financial Research Journal*, *12*(29), 21-40. [In Persian]

Rai, R., & Pouyanfar, A. (2017). Advanced Investment Management. Institute for Research and Development in the Humanities (SAMT) Publications. [In Persian]

Ravi, K., & Ravi, V. (2017). A novel automatic satire and irony detection using ensembled feature selection and data mining. *Knowledge-based systems*, *120*, 15-33.

Reddy, E. U., & Nagarjuna, N. (2023). Applying Machine Learning for Portfolio Switching Decisions. *Lecture Notes in Computer Science*, 399–406.

Schlottmann, F., & Seese, D. (2001). A hybrid genetic-quantitative method for risk-return optimisation of credit portfolios. In *Proceedings of the Conference of Quantitative Methods in Finance, University of Technology, Sydney, Australia* (Vol. 55).

Sinaei, H., Zamani, S. (2014). Decision-Making for Stock Portfolio Selection: A Comparison of Genetic Algorithms and Bee Algorithm. *Executive Management Research*, *6*(11), 83-102. [In Persian]

Sultan Panah, H., Fatemi, A., Nazaripour, M., & Bigoli, B. (2013). Comparing Stock Selection Models for Portfolio Formation in Terms of Expected Return Estimation, Actual Future Returns, and Their Risks in the Tehran Stock Exchange. In Proceedings of the Sixth Conference on the Development of the Financial System in Iran. *Islamic Azad University, Sanandaj Branch; University of Kurdistan.* [In Persian]

Taghizadeh, K., Mullah Alizadeh Zavardehi, S., Salehi, A. K., & Mahmoudi Rad, A. (2022). Evaluation of the optimal portfolio portfolio using market criteria using multi-criteria decision criteria under conditions of uncertanty in the

Iranian capital market. *Advances in Finance and Investment, 3*(6), 101-128. [In Persian]

Wang, S., Xia, Y., Wang, S., & Xia, Y. (2002). Mean-variance-skewness model for portfolio selection with transaction costs. *Portfolio Selection and Asset Pricing*, 129-144.

Wang, W., Li, W., Zhang, N., & Liu, K. (2020). Portfolio formation with preselection using deep learning from long-term financial data. *Expert Systems with Applications*, *143*, 113042.

Yamai, Y., & Yoshiba, T. (2002). Comparative analyses of expected shortfall and value-at-risk (2): expected utility maximization and tail risk. *Monetary and economic studies*, *20*(2), 95-115.

Zare Mehrjerdi, Y., Shahmohammadi, M., & Emami Maibodi, L. (2013). A Hybrid Intelligent Algorithm for Portfolio Selection using Fuzzy Mean-Variance-Skewness. *International Journal of Industrial Engineering and Production Management*, *23*(4), 447-458. [In Persian]

Zhong, X., & Enke, D. (2017). A comprehensive cluster and classification mining procedure for daily stock market return forecasting. *Neurocomputing*, *267*, 152-168.

Zolfaghari, M., & Gholami, S. (2021). A hybrid approach of adaptive wavelet transform, long short-term memory and ARIMA-GARCH family models for the stock index prediction. *Expert Systems with Applications*, *182*, 115149.

COPYRIGHTS

© 2023 by the authors. Published by Islamic Azad University, Esfarayen Branch. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) (http://creativecommons.org/licenses/by/4.0/).