Journal of Intelligent Strategic Management
Vol(4), 447-478
Bumara.ir

A Two-Stage Framework for Portfolio Optimization: Intelligent Stock Pre-Selection Using Support Vector Machine

Ebrahim Nasir-Eslami*

Assistant Professor, Department of Statistics, Faculty of Basic Sciences, Bu-Ali Sina University, Hamadan, Iran.

Abstract

This study investigates the effectiveness of the support vector machine model in portfolio optimization in the Tehran Stock Exchange. The support vector machine model selects a portfolio based on high-frequency data and its risk-adjusted performance with the return of a statistical portfolio predicted by the comparison capital. This model is used in solving classification and regression problems with a supervised learning approach. This research is an applied research that analyzes and models and draws conclusions from empirical evidence based on historical data. The statistical population is the financial and trading information of companies on the Tehran Stock Exchange. The statistical sample with the help of the support vector machine includes a selected portfolio with a volume of 100 companies from 387 companies active in the stock market between the period 1400 and 1401. The data in terms of preparation includes daily, monthly and annual periods. The SVM model, as an efficient tool in predicting returns and managing portfolio risk, can be a suitable alternative to traditional models such as CAPM. This model has the ability to better interpret results and higher accuracy in predicting stock returns. The results showed that the portfolio formed with SVM with an average return of 3.3 percent and a standard deviation of 1.2 percent, has a much better performance than the portfolio based on the traditional CAPM model with an average return of -1 percent and a standard deviation of 5.5 percent. The Sharpe ratio in the SVM model was significantly higher, indicating the higher efficiency of this model in managing stock portfolios.

Keywords: Stock portfolio optimization, support vector machine, CAPM model, Tehran Stock Exchange, return prediction

How to Cite: Nasirel-Islami,E. (2025). A Two-Stage Framework for Portfolio Optimization: Intelligent Stock Pre-Selection Using Support Vector Machine. Journal of Intelligent Strategic Management, 4(1), 447-478.

doi: bumara .3.2.15564.35879847070

Intelligent Strategic Management (JISM) in Development and Evolution is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.

© Authors

* Corresponding Author: e.nasiroleslami@basu.ac.ir

Original Research
Received : 2025/01/14
Review: 2025/03/07
Accepted: 2025/04/19
eISSN: 2891-4128

نشریه مدیریت استراتژیک هوشمند
دوره ۴ شماره۱، بهار، ۱۴۰۴، ۴۷۸-۴۴۷
Bumara.ir

یک چارچوب دو مرحله‌ای برای بهینه‌سازی سبد سهام: پیش‌انتخاب هوشمند سهام با استفاده از ماشین بردار پشتیبان

ابراهیم نصیرالاسلامی * استادیار گروه آمار دانشکده علوم پایه، دانشگاه بوعلی سینا، همدان، ایران.

مقاله پژوهشی
تاریخ دریافت:۱۴۰۳/۱۰/۲۵
بازنگری:۱۴۰۳/۱۲/۱۷
تاریخ پذیرش: ۱۴۰۴/۰۱/۳۰
eISSN:2891-4128

چکیده

این پژوهش به بررسی کارایی مدل ماشین بردار پشتیبان در بهینه‌سازی سبد سهام در بورس اوراق بهادار تهران می‌پردازد. مدل ماشین بردار پشتیبان برای تشکیل یک سبد بر اساس داده‌های فرکانس بالا و عملکرد تعدیل شده با ریسک آن با بازده یک سبد آماری پیش‌بینی‌شده توسط سرمایه مقایسه اقدام به انتخاب سبد سهام می کند. این الگو با رویکرد یادگیری با نظارت در حل مسائل رده بندی و رگرسیون مورد استفاده قرار می گیرد. این پژوهش از جنس پژوهش های کاربردی است که بر اساس داده های تاریخی اقدام به تحلیل و مدلسازی و نتیجه گیری از شواهد تجربی می کند. جامعه آماری اطلاعات مالی و معاملاتی شرکت های بورس اوراق بهادار تهران می باشد. نمونه آماری با کمک ماشین بردار پشتیبان شامل سبد انتخابی با حجم ۱۰۰ شرکت از بین ۳۸۷ شرکت فعال در تالار بورس بین دوره زمانی ۱۴۰۰ تا ۱۴۰۱ می باشد. که داده ها بر حسب آماده سازی شامل دوره های روزانه، ماهانه و سالانه است. مدل SVM به عنوان یک ابزار کارآمد در پیش‌بینی بازدهی و مدیریت ریسک سبد سهام، می‌تواند جایگزین مناسبی برای مدل‌های سنتی مانند CAPM باشد. این مدل قابلیت تفسیر بهتر نتایج و دقت بالاتری در پیش‌بینی بازدهی سهام دارد. نتایج نشان داد سبد تشکیل‌شده با SVM با میانگین بازدهی ۳/۳ درصد و انحراف معیار ۱/۲ درصد، عملکرد به‌مراتب بهتری نسبت به سبد مبتنی بر مدل سنتی CAPM با میانگین بازدهی ۱- درصد و انحراف معیار ۵/۵ درصد دارد. نسبت شارپ در مدل SVM به‌طور معناداری بالاتر بود که نشان‌دهنده کارایی بالاتر این مدل در مدیریت سبد سهام است.

کلیدواژه‌ها: بهینه‌سازی سبد سهام، ماشین بردار پشتیبان، مدل CAPM، بورس تهران، پیش‌بینی بازدهی

استناد به این مقاله: نصیرالاسلامی، ابراهیم . (۱۴۰۴). یک چارچوب دو مرحله‌ای برای بهینه‌سازی سبد سهام: پیش‌انتخاب هوشمند سهام با استفاده از ماشین بردار پشتیبان. مدیریت استراتژیک هوشمند, ۴(۱), ۴۴۷-۴۷۸.

مدیریت استراتژیک هوشمند (JISM) در توسعه و تکامل تحت مجوز بین‌المللی کریتیو کامنز با شرایط انتساب-غیرتجاری ۴,۰ منتشر می‌شود.

©نویسندگان

* نویسنده مسئول: e.nasiroleslami@basu.ac.ir

مقدمه

مکانیسم‌های پیش‌بینی قیمت سهام برای شکل‌گیری استراتژی‌های سرمایه‌گذاری و توسعه مدل‌های مدیریت ریسک، اساسی هستند. با این حال، فرضیه بازار کارآمد بیان می‌کند که نمی‌توان به طور مداوم بازده تعدیل‌شده بر اساس ریسک را بالاتر از سودآوری کل بازار به دست آورد. پیشرفت‌های محاسباتی منجر به استفاده از چندین الگوریتم یادگیری ماشینی برای پیش‌بینی مداوم حرکات بازار شده است. بنابراین ارزش دارایی های آتی مانند قیمت سهام شرکت را به راحتی می توان برآورد کرد. مدل‌های مبتنی بر ماشین بردار پشتیبان (SVM) جزو پرکاربردترین تکنیک‌ها هستند.

اطلاعات یک منبع ارزشمند در هنگام ساخت مدل‌های پیش‌بینی در تعقیب سیستم‌های معاملاتی بازار مالی سودآور است. با توجه به ویژگی‌های سری زمانی مالی، هنگام توسعه سیستم‌های پیش‌بینی قیمت باید با چالش‌های مختلفی مواجه شد. از نقطه نظر ادبیات موضوع، بر اساس فرضیه بازار کارآمد ، اطلاعات مربوط به طور گسترده در دسترس همه فعالان بازار خواهد بود و بلافاصله در قیمت منعکس می شود، طبق ادعای مالکیل و فاما (۱۹۷۰) فرضیه بازار کارآمد که غیرممکن است، به طور مداوم و بیش از حد در بلندمدت، برای دستیابی به بازده بالاتر از بازار که با سطح ریسک مفروض تنظیم شده است. همانطور که توسط ماکیل(2003) خلاصه شده است ؛ فرضیه بازار کارآمد از زمان معرفی آن، به ویژه با توسعه وی مورد سوال قرار گرفته است. همانطور که در مطالعات مبتنی بر ماشین بردار پشتیبان و سایر الگوریتم‌ها نشان داده شده است، فرضیه بازار کارآمد از زمان معرفی، به ویژه با توسعه سیستم‌های پیش‌بینی، مورد سؤال قرار گرفته است(کیو و زانگ،۲۰۱۶)[1] ؛ که می تواند در بلندمدت سود ایجاد کند. با این حال، مالکیل و فاما (۱۹۷۰)، استدلال می‌کنند که بازار یک راه رفتن تصادفی را دنبال می‌کند و تلاش برای پیش‌بینی حرکات آن به شیوه‌ای ثابت بیهوده خواهد بود. پیشرفت های محاسباتی منجر به معرفی تکنیک های یادگیری ماشین برای سیستم های پیش بینی در بازارهای مالی شده است. در بررسی مقالاتی در مورد سیستم های پیش بینی، هسو[2] و همکاران (۲۰۱۶) مشاهده کردند که استفاده از سری های مالی برای اندازه گیری کارایی الگوریتم های پیش بینی و طبقه بندی کننده ها در یادگیری ماشین رایج است. طبقه‌بندی‌کننده‌ها سیستم‌هایی هستند که می‌توانند از طریق آموزش، تشخیص الگوها و در نتیجه اختصاص یک کلاس به داده‌های جدید را

1 Qu and Zhang
2 Hsu

بیاموزند. به عنوان مثال، الگوریتم های یادگیری ماشین را می توان برای پیش بینی ورشکستگی، همانطور که توسط ژو و همکاران مشاهده شد، استفاده کرد. لی و همکاران (۲۰۱۲) در چنین مواردی، هدف طبقه بندی شرکت هایی با بیشترین احتمال ورشکستگی، بر اساس الگوریتم طبقه بندی خودکار است. نمونه های دیگر اندازه گیری ریسک اعتباری است، همانطور که در پژوهش لی و همکاران (۲۰۰۶) و الگوی پیش بینی قیمت دارایی توسط کائو [1] و همکاران پیشنهاد شده است؛ علاوه بر توسعه استراتژی‌های معاملاتی، پیشرفت در سیستم‌های اطلاعات محاسباتی امکان انجام معاملات الکترونیکی سریع در بازارهای مالی را فراهم کرده است. بر اساس الگوریتم های معاملاتی فرکانس بالا، ارائه و اجرای سفارش های خرید، فروش یا لغو را می توان در ثانیه و میکروثانیه انجام داد، همانطور که گلدشتاین و همکاران (۲۰۱۴) بیان کرد استفاده شدید از سیستم‌های محاسباتی سریع توسط برخی از فعالان بازار ممکن است سودآوری را افزایش دهد، اما اثرات آن بر عملکرد عادی بازار مشکوک است، زیرا همه شرکت‌کنندگان به این نوع فناوری دسترسی ندارند. به هر جهت الگوهای ماشین بردار می تواند نسبت به روش های سنتی ارزیابی و انتخاب سبد سهام از نظر دقت و تکنیک تفاوت معنی داری داشته باشد. از این رو این تحقیق به بررسی همین مسئله با استفاده از داده های بازار سرمایه ایران خواهد پرداخت.

در سال های اخیر، الگوریتم های مختلف یادگیری ماشین برای پیش بینی قیمت سهام با نتایج امیدوارکننده ای به کار گرفته شده است (ویج، ۲۰۲۰). با این حال، نویسندگان در این مطالعات اغلب از مجموعه داده‌های متشکل از مقادیر باز، زیاد، کم، بسته، حجم استفاده می کنند. این مجموعه داده ها به طور کلی و به راحتی در دسترس هستند، که ممکن است دلیل محبوبیت آنها باشد. متأسفانه، از آنجایی که این مقادیر در مجموعه داده‌های عموماً در دسترس از قبل تجمیع شده‌اند، نمی‌توانند اطلاعاتی در مورد معاملات فردی یا لحظات خاصی که در آن معاملات رخ داده‌اند ارائه دهند. این امر دقت پیش‌بینی‌های انجام‌شده با استفاده از این مجموعه‌های داده را محدود می کند. در واقع، به طور شهودی توجیه می‌شود که داده‌های مستقیم و با وضوح بالا می‌توانند بینش بهتری نسبت به عملکرد درونی توسعه قیمت ارائه دهند. بازار سهام یک سیستم پیچیده است که تحت تأثیر طیف گسترده ای از عوامل، از جمله شاخص های اقتصادی، عملکرد شرکت، و رویدادهای جهانی و همچنین اثرات لحظه ای بی شماری مانند: برخاستن، اولین فروش، تله خرس،

1 Kao

توجه رسانه ها، اشتیاق، طمع، توهم، پارادایم جدید، انکار، بازگشت معنایی، ترس، تسلیم شدن، یا ناامیدی (جاکوبز، ۱۹۸۹). روش‌های سنتی قیمت‌گذاری دارایی‌ها مانند مدل قیمت‌گذاری دارایی سرمایه (CAPM) به طور گسترده برای تخمین بازده مورد انتظار یک سرمایه‌گذاری بر اساس بازده مورد انتظار بازار و ریسک دارایی به طور گسترده برای تعیین بازده مورد انتظار استفاده می‌شوند. یک مدل سرمایه گذاری با توجه به طراحی آن از یک متغیر منفرد، بتا بازار، برای نشان دادن ریسک یک دارایی استفاده می کند. از سوی دیگر، در یادگیری ماشینی، ما می‌توانیم داده‌های با ابعاد بالا و غیرخطی را مدیریت کنیم، که آن را برای پیش‌بینی داده‌های با کارایی بالا، مانند بازار سهام، مناسب می‌سازد. هدف اصلی این مقاله، به چالش کشیدن شکل ضعیف بازار است. کارایی با استفاده از تکنیک یادگیری ماشین SVM برای پیش‌بینی قیمت سهام (و حرکت آن) شگفت انگیز است. شکاف اصلی تحقیقاتی که این مطالعه برای آن پر می کند، مقایسه استراتژی تجارت مبتنی بر الگوریتم یادگیری ماشین غیر تکراری با پورتفولیوی ایستا متوازن است. فرضیه این مطالعه این است که یک الگوریتم ماشینی می‌تواند منجر به سطح ریسک پایین‌تری در مقایسه با پرتفوی استاتیک متوازن شود، همانطور که توسط مدل قیمت‌گذاری دارایی سرمایه (CAPM) پیش‌بینی شده است. برای دستیابی به این هدف، نویسندگان یک بررسی جامع از کاربرد ماشین‌های بردار پشتیبان در پیش‌بینی قیمت سهام انجام دادند و عملکرد آن را با مدل سنتی CAPM مقایسه کردند. علاوه بر این، این مطالعه عوامل کلیدی مؤثر در عملکرد برتر الگوریتم را مورد بحث قرار می‌دهد و بر اهمیت انتخاب ویژگی، پیش پردازش داده‌ها و انتخاب مدل در دستیابی به عملکرد معاملاتی بهینه تأکید می کند.

مبانی نظری و پیشینه تحقیق

بر اساس تئوری بازار کارا (جنسن، ۱۹۷۸)، بازارهای مالی قابل پیش بینی نیستند. با این حال، برخی از مطالعات با بهره گیری از برخی ناکارآمدی ها در تمامی بازارهای مالی، به دقت رضایت بخشی در پیش بینی سری های زمانی مالی دست یافته اند. مدل های پیش بینی سری زمانی مالی دو دسته اصلی هستند: ۱- مدل های آماری که به دو گروه تقسیم می شوند. گروه اول مدل های میانگین گرا مانند میانگین متحرک یکپارچه رگرسیون خودکار (ARIMA) (باکس و جنکینز، ۱۹۷۶) هستند، در حالی که گروه دوم بر اساس نوسانات یا واریانس هستند، مانند ناهمسانی شرطی اتورگرسیو (ARCH)، و ناهمسانی

شرطی خودبازگشتی تعمیم یافته(GARCH) شکل گرفته اند. برخی از مدل‌های ماشین یادگیری مانند رگرسیون بردار پشتیبان (SVR) توسط راسل و همکاران(۲۰۱۵) شکل گرفتند. همچنین جنگل تصادفی توسط بالینگ و همکاران(2015) توسعه یافت.[2]از طرفی الگوهای حافظه بلند مدت ماشین بردار توسطروندیوالا و همکاران(۲۰۱۷) معرفی شد، شبکه های عصبی کانولوشن (CNN) و روش های یادگیری مجموعه ای مانند درخت تصمیم تقویت شده با گرادیان (GBDT) بعدا معرفی شد(ژو و همکاران، ۲۰۱۹).. در میان مدل های یادگیری ماشین به دلیل در نظر گرفتن رابطه بین ویژگی های ورودی از طریق هسته های کانولوشن، پیچیدگی بازارهای مالی را بهتر از سایر مدل ها مثل الگوهای مارکویز و یا CAPM در نظر می گیرند و عملکرد بالقوه آنها در پیش بینی سری های زمانی مالی در بسیاری از مطالعات تایید شده است (سزار و اوزبایلو[3]۲۰۱۸). اگرچه انتخاب نوع مدل پیش‌بینی بسیار مهم است، روش‌های مختلفی مانند بهینه‌سازی متا پارامترهای مدل‌های مختلف، نحوه آموزش مدل‌ها و استفاده از روش‌های انتخاب ویژگی، نقش مهمی در بهبود عملکرد مدل‌ها دارند. مدل‌های ماشین بردار به طور فزاینده‌ای پیچیده‌تر می‌شوند، با پارامترهای زیادی که تأثیر زیادی بر عملکرد دارند، که به آنها هایپر پارامتر[4]گفته می‌شود. برای دستیابی به عملکرد بهینه این مدل‌ها، محققان چندین الگوریتم فراابتکاری را پیشنهاد کرده‌اند که برای جستجوی بهترین مجموعه از فراپارامترها طراحی شده‌اند. این الگوریتم ها شامل الگوریتم کرم شب تاب (نایاک[5]۲۰۲۱) ، الگوریتم ازدحام سالپ توسط لی و همکاران[6](2021) ، الگوریتم جستجوی گرانشی توسط (ایزات و همکاران،۲۰۲۱)[7]هستند. این محققان برای بهینه‌سازی فراپارامترها استفاده کردند و تطبیق‌پذیری الگوریتم‌های فراابتکاری را به عنوان ابزاری برای تنظیم فراپارامتر برجسته کردند. این رویکردها بسیار انعطاف‌پذیر هستند و می‌توانند پارامترهای مختلفی از جمله ضرایب منظم‌سازی، نرخ یادگیری، تعداد لایه‌های پنهان و اندازه لایه‌ها را بهینه کنند.

راه دیگر برای بهبود عملکرد مدل های یادگیری، استفاده از مدل های انتخاب ویژگی برای انتخاب تعداد مناسب ویژگی های ورودی است. انتخاب ویژگی به طور موثر داده های نامربوط و اضافی را حذف می کند، زمان محاسبات را کاهش می دهد، دقت یادگیری را

1 Rasel et al
2 Ballings et al
3 Sezer and Ozbayoglu
4 hyperparameters
5 Nayak et al
6 Li et al.
7 Ezzat et al

بهبود می بخشد و درک بهتر مدل یا داده های یادگیری را تسهیل می کند (سای و همکاران،۲۰۱۸[1]). علاوه بر این، روش آموزش مدل به طور قابل توجهی پیچیدگی محاسباتی مدل را کاهش می دهد. در بیشتر مطالعات، از داده های سهام فردی برای آموزش مدل ها استفاده شده است. این روش آموزش پیچیدگی مدل را در مسائل بهینه سازی پورتفولیو که در آن تعداد زیادی سهام ارزیابی می شود، به طور چشمگیری افزایش می دهد. بنابراین برای غلبه بر این مشکل، قبل از آموزش مدل‌ها بر روی تک سهام، می‌توان آنها را با روش‌های خوشه‌بندی سهام خوشه‌بندی کرد و آموزش مدل‌ها بر روی نماینده هر خوشه انجام می‌شود. در این مقاله مبانی نظری کاربرد تکنیک های فوق در راستای انتخاب سبد بهینه سهام دسته بندی شده است.

مبانی نظری مدل های پیش بینی کننده قیمت سهام

نقش بورس در اقتصاد امروز بسیار ارزشمند است. سرمایه گذاران دائما به دنبال راه هایی برای پیش بینی قیمت سهام ترجیحی خود هستند. با پیش بینی های دقیق، سرمایه گذاران سود قابل توجهی در بورس به دست می آورند. با این حال، این کار یک کار ساده نیست. بازار سهام بسیار نوسان و مستعد عوامل مختلفی از جمله رویدادهای سیاسی، شرایط اقتصادی، نگرش معامله‌گران و رفتار است که می‌تواند به طور ناگهانی تغییر کند. علاوه بر این، به دلیل ماهیت غیر خطی و پیچیده آن، پیش بینی قیمت سهام یک چالش مهم باقی می ماند.

فرضیه گام تصادفی و فرضیه بازار کارآمد این مفهوم را ایجاد کرده‌اند که پیش‌بینی بازار سهام یک کار غیرعملی است. فرضیه Wisdom of Crowd پیشنهاد می کند که نظرات جمعی ممکن است پیش بینی های دقیقی ارائه دهند، اما کارایی محدودی در پیش بینی بازده بازار سهام نشان داده است. علاوه بر این، ماهیت پیچیده پیش‌بینی بازار سهام توسط متغیرهای متعدد و عدم قطعیت‌هایی که بر قیمت سهام تأثیر می‌گذارند، پیچیده‌تر می‌شود. با این حال، علیرغم این چالش ها، افراد خاص و سرمایه گذاران نهادی موفق شده اند از بازار پیشی بگیرند و به نتایج سودآوری دست یابند (آوری و همکاران،۲۰۱۶[2]).

وجود الگوهای قابل پیش بینی در قیمت سهام توسط اقتصاددانان مالی و آماردانان تایید شده است که منجر به ادعاهای بحث برانگیز مبنی بر اینکه سرمایه گذاران می توانند بازده

1 Cai et al
2 Avery

تعدیل شده با ریسک اضافی کسب کنند، شده است (کریستوفرن،۲۰۰۶[1]). با این حال، وابستگی‌های آماری که باعث ایجاد شتاب می‌شوند، بسیار کوچک هستند و درک بازده مازاد را برای سرمایه‌گذاران دشوار می‌کنند. در حالی که ناهنجاری های خاصی در بازده سهام وجود دارد، آنها معمولاً در زمینه های خاص مدل می شوند و نمی توان آنها را تعمیم داد. تلاش برای حدس زدن اینکه بازارهای سهام به کدام سمت می‌روند، هم برای سرمایه‌گذاران و هم برای محققین مشکلی بسیار سخت و سخت برای حل آن بوده است. کارشناسان در این زمینه با کمک ریاضیات پیچیده، علوم کامپیوتر، اقتصاد و سایر زمینه های دانش به روندهای بازار سهام نگاه می کنند. به طور مشخص، تحلیل فنی، پیش‌بینی سری زمانی، یادگیری ماشین و داده کاوی و مدل‌سازی و پیش‌بینی نوسانات سهام با استفاده از معادلات دیفرانسیل روش‌هایی هستند که برای پیش‌بینی رفتار قیمت سهام استفاده می‌شوند(ساها، ۲۰۱۴[2]).

در سال‌های اخیر تحقیق و استفاده از روش‌های یادگیری ماشین برای پیش‌بینی قیمت سهام افزایش یافته است که جرقه‌ای از کار تأثیرگذار هلستروم[3](۱۹۹۸) بوده است. اینها شامل مدل‌های یک شبکه عصبی مصنوعی، ماشین‌های بردار پشتیبان (SVMs)، میانگین متحرک یکپارچه اتورگرسیو (ARIMA)، هموارسازی نمایی تطبیقی و رگرسیون چندگانه است. اخیراً بساک[4]و همکاران(۲۰۱۹) نشان داد که جنگل‌های تصادفی و درخت‌های تصمیم تقویت‌شده با گرادیان نسبت به روش‌های پیش‌بینی قبلی بهبود یافته‌اند. پاگلیرو[5](۲۰۲۳) هدف اصلی مرور جامع خود را استفاده از مدل‌های پیشرفته یادگیری ماشینی بیان نمود که برخی از آن‌ها را قبلاً در سایر زمینه‌های تحقیقاتی استفاده شده بود. برای توسعه یک مدل پیش‌بینی دقیق برای بازده کوتاه‌مدت بازار سهام با استفاده از الگوریتم طبقه‌بندی درخت اضافی، به طور خاص، پیش‌بینی نمود که آیا بازدهی پس از ۱۰ روز معاملاتی (۲ هفته تقویمی) با تجزیه و تحلیل دوره‌های قبلی قابل توجه باشد. اهداف تحقیق وی در راستای ایجاد یک مدل طبقه‌بندی کننده درخت اضافی که برای پیش‌بینی بازده سهام بهینه شده بود. همچنین برای ارزیابی عملکرد آن در برابر مدل‌های معیار مانند جنگل‌های تصادفی و ارزیابی قابلیت آن به عنوان یک تکنیک بهبود یافته برای پیش بینی بازار سهام در مقایسه با روش های سنتی ، مورد بررسی قرار گرفت.

1 Christoffersen
2 Saha
3 Hellstrom
4 Basak
5 Pagliaro

فرضیه کلیدی این است که یک مدل طبقه‌بندی کننده درخت اضافی می‌تواند بازده سهام کوتاه‌مدت را در مقایسه با روش‌های معیاری مانند جنگل‌های تصادفی و مدل‌های رگرسیون سنتی پیش‌بینی کند. مطالعات قبلی که از یادگیری ماشین برای پیش‌بینی سهام استفاده می‌کردند، عمدتاً بر روی مدل‌های مبتنی بر رگرسیون تمرکز داشتند. مدل‌های طبقه‌بندی علیرغم نقاط قوتشان در مدیریت داده‌های با ابعاد بالا و کاهش بیش‌برازش، نسبتاً کمتر مورد بررسی قرار گرفته‌اند. این یک شکاف در ارزیابی الگوریتم‌های طبقه‌بندی مانند درختان اضافی برای پیش‌بینی بازار سهام ایجاد می‌کند. فرضیه درختان اضافی به طور خاص با آموزش سریع‌تر و استحکام در برابر نویز بر معایب جنگل‌های تصادفی غلبه می‌کند و آن را برای داده‌های مالی پیچیده مناسب می‌سازد.

مبانی نظری روش های یادگیری ماشینی

در این بخش سعی شده است مدل‌های یادگیری ماشینی که در مطالعات قبلی برای پیش‌بینی و پیش‌بینی سهام استفاده شده‌اند، خلاصه شود. پس از اینکه داده ها از قبل پردازش شدند و به یک نمایش استاندارد تبدیل شدند، برای پردازش بیشتر به مدل های یادگیری ماشینی داده می شوند. بخش زیر به طور خلاصه رویکردهای مختلف یادگیری ماشین ارائه شده را خلاصه می کند:

شبکه های عصبی مصنوعی : شبکه های عصبی مصنوعی یک تکنیک بیولوژیکی الهام گرفته از مغز است که در آن تعداد زیادی نورون مصنوعی به منظور حل مشکلات پیچیده به شدت به هم متصل می شوند (شارما[1]،۲۰۲۱). این مدل‌ها زمینه یک مسئله را با ایجاد تبدیل‌های متعدد در فضای ویژگی و به دنبال آن غیرخطی بودن، برای ایجاد نمایش‌های ساده‌شده آن درک می‌کنند (گنسر[2]،۲۰۲۱). مطالعات متعددی از مدل های شبکه های عصبی مصنوعی برای پیشبینی بازار سهام استفاده کرده اند. به عنوان مثال، محققان از شبکه های عصبی مصنوعی برای پیش‌بینی روند روزانه شاخص S&P 500 استفاده کردند. تکنیک‌های کاهش سه‌بعدی – به‌عنوان مثال، PCA، تجزیه و تحلیل مؤلفه‌های قوی فازی (FRPCA) و تجزیه و تحلیل مؤلفه اصلی مبتنی بر هسته – برای ساده‌سازی مجموعه داده‌ها در ادبیات قبلی اعمال شد. نتایج پژوهش رئوف (۲۰۲۱) نشان داد که ترکیب شبکه های

1 Sharma
2 Ganser
3 Rouf

عصبی مصنوعی با الگوریتم PCA کارآمدتر است. علاوه بر این، انتخاب یک تابع هسته مناسب به طور مستقیم بر عملکرد KPCA تأثیر می گذارد.

پرسپترون چندلایه یک تکنیک متداول برای پیشبینی بازار سهام است . پرسپترون چندلایه یک شبکه های عصبی مصنوعی با یک لایه ورودی و خروجی و یک یا چند لایه میانی است. به طور کلی، پرسپترون چندلایه از روش انتشار پس‌انداز برای آموزش استفاده می‌کند، که در آن خطاهای پیش‌بینی‌شده از لایه خروجی به لایه ورودی مجدد منتشر می‌شوند تا خطاها را به حداقل برسانند (خاشعی،۲۰۱۷[2]).

یک مطالعه سه مدل شبکه های عصبی مصنوعی - پرسپترون چندلایه ، شبکه عصبی مصنوعی پویا (DAN2) و ناهمسانی شرطی اتورگرسیو (GARCH) – را برای پیش‌بینی قیمت NASDAQ مقایسه کرد (گورسن،۲۰۱۸[3]). هر سه مدل با استفاده از میانگین انحراف مطلق و میانگین مربع خطا را مورد ارزیابی قرار گرفتند. نتایج نشان داد که پرسپترون چندلایه بهتر از شبکه عصبی مصنوعی پویا و ناهمسانی شرطی اتورگرسیو بوده و پرسپترون چندلایه بهتر عمل می کند. علاوه بر این، با پیشنهاد اینکه آنها بر روی یافتن اینکه آیا GARCH تأثیر اصلاحی بر پیش‌بینی‌ها یا سایر متغیرهای همبسته‌ای که تأثیر اصلاح‌کننده بر پیش‌بینی‌ها دارند، تمرکز می‌کنند، جهت آینده را فراهم می‌کند(رئوف،۲۰۲۳).

بینگ و همکاران (۲۰۱۲) در مطالعه ای از مدل‌های پیش بینی کننده تعمیم‌یافته (GFF) و پرسپترون چندلایه برای پیش‌بینی شاخص بازار بورس استانبول استفاده کردند، جایی که داده‌ها از بانک مرکزی ترکیه گرفته شد. در مجموع هشت مجموعه پیش‌بینی (شش شبکه های عصبی مصنوعی و دو میانگین متحرک) با تغییر تعداد لایه‌های پنهان انجام شد. دو مجموعه از پیش بینی ها بر اساس میانگین متحرک بود. دقت پیش‌بینی با استفاده از ضریب تعیین محاسبه شد و بالاترین دقت برای هر دو پرسپترون چندلایه و مدل‌های پیش بینی کننده تعمیم‌یافته با استفاده از یک لایه پنهان به دست آمد.

علاوه بر این، تکنیک شبکه های عصبی مصنوعی اغلب مورد استفاده برای پیشبینی بازار سهام، شبکه تابع پایه شعاعی است. این یک شبکه لایه ای است که در آن لایه های پنهان از یک تابع فعال سازی شعاعی استفاده می کنند (هوتا،۲۰۱۷). به عنوان مثال، گنسر(۲۰۲۱)

---

1 MLP
2 Khashei
3 Guresen

از شبکه عصبی RBF برای پیش‌بینی شاخص شانگهای و NASDAQ با استفاده از پسوند "Locality Preserving Projection" معروف به LPP دو بعدی برای انتخاب بیشتر ویژگی‌های مرتبط برای پیش‌بینی استفاده کردند. روش پیشنهادی در هر دو شاخص بازار عملکرد خوبی داشت.

مدل بیز ساده : مدل بیز ساده یک روش طبقه بندی است که نقاط داده را بر اساس قضیه احتمال بیزی طبقه بندی می کند. این روش طبقه بندی بسیار سریع است و می تواند در مجموعه داده های بزرگ مقیاس شود. این رویکرد طبقه بندی به طور گسترده برای پیشبینی بازار سهام استفاده شده است. برای مثال، یو(۲۰۱۳) از الگوریتم ساده بیز برای تحلیل احساسات داده های متنی از منابع متعدد استفاده کردند. نویسندگان تأثیر منابع داده های معمولی و رسانه های اجتماعی را بر شرکت های مختلف و ارتباط متقابل آنها مقایسه کردند.

الگوریتم ژنتیک : الگوریتم ژنتیک یک رویکرد اکتشافی برای حل مسئله است که فرآیند تکامل طبیعی را تقلید می کند. الگوریتم ها مفهوم انتخاب طبیعی را برای انتخاب راه حل ممکن بهینه اعمال می کنند. در پیشبینی بازار سهام، الگوریتم ژنتیک برای تنظیم دقیق پارامترها برای تولید بهترین قانون معاملاتی استفاده می شود. مطالعات متعددی از الگوریتم ژنتیک برای افزایش دقت پیشبینی بازار سهام استفاده کرده اند (استرادر، ۲۰۲۰)[2] به عنوان مثال، در مطالعه کیم(۲۰۱۲) یک سیستم پشتیبانی تصمیم گیری هوشمند را برای معاملات سهام توسعه دادند. این مطالعه از مجموعه‌های خشن و الگوریتم ژنتیک برای داده‌های غیرخطی و پیچیده سهام برای یافتن ویژگی‌هایی استفاده کرد که می‌توان از آنها برای ایجاد قوانین معاملاتی بهینه استفاده کرد. این قوانین برای ایجاد استراتژی های خرید یا فروش بهینه اعمال می شوند.

الگوریتم های فازی : منطق فازی یک روش مبتنی بر استدلال انسانی است که در آن از تمام احتمالات میانی بین ۰ و ۱ برای تصمیم گیری استفاده می شود. این یک رویکرد قدرتمند است که در آن میزان تعلق به یک دسته خاص در نظر گرفته می شود. سیستم استنتاج عصبی فازی تطبیقی (ANFIS) محبوب ترین الگوریتم فازی است که برای پیشبینی بازار سهام استفاده می شود. برخی از مطالعات نمونه که در آن سیستم استنتاج عصبی فازی تطبیقی برای پیشبینی بازار سهام به کار گرفته شده است؛ شامل

1 Yu
2 Strader

است(چاندر،۲۰۱۷)[1]. در مطالعه هاولز(۲۰۱۷)[2]یک رویکرد منطق فازی برای تجزیه و تحلیل احساسات در رسانه های اجتماعی برای پیشبینی بازار سهام توسعه دادند. علاوه بر این، چندین مطالعه از رویکردهای فازی ترکیبی برای پیشبینی بازار سهام استفاده کرده اند(رجب،۲۰۱۹)[3]. به عنوان نمونه ای از رویکرد فازی ترکیبی، رجب و همکاران (۲۰۱۹) یک مدل جدید برای پیش‌بینی با استفاده از کلونی زنبورهای مصنوعی پیشنهاد کرد. در این مطالعه، داده های ۵۰ شرکت از بورس اوراق بهادار ایالات متحده از سال ۲۰۰۸ تا ۲۰۱۸ گرفته شده است. مدل از ۲۰ شاخص فنی به عنوان ورودی استفاده کرد. معیار سنجش عملکرد، دقت و کیفیت بود. علاوه بر این، مدل از دقت پیش‌بینی دقیق‌تری نسبت به مدل‌های دیگر برخوردار بود.

شبکه های عصبی یادگیری عمیق : شبکه های عصبی یادگیری عمیق پیشرفتی نسبت به شبکه‌های عصبی معمولی است که در آن لایه‌ها و نورون‌های پنهان بیشتری برای استخراج و تبدیل خودکار ویژگی‌ها استفاده می‌شوند. افزایش تعداد لایه‌های پنهان با واحدهای پردازش غیرخطی، کارایی یادگیری از داده‌های خام را بهبود می‌بخشد (لکن،۲۰۱۵[4]). شبکه های عصبی یادگیری عمیق اغلب برای پیش بینی های مالی با استفاده از داده های متنی و عددی استفاده شده است (واو،۲۰۲۰[5]). مطالعات مختلف از الگوریتم‌های شبکه های عصبی یادگیری عمیق مانند شبکه‌های عصبی کانولوشن ، حافظه بلند مدت و شبکه‌های باور عمیق[6] استفاده کرده‌اند(مثل ژانگ،۲۰۱۸[7]). به عنوان مثال، یک مطالعه اخیر توسط هوشانی(۲۰۱۸)[8] مقایسه ای از چهار مدل پیش بینی برای پیش بینی قیمت بازار سهام انجام داد که شامل میانگین متحرک یکپارچه رگرسیون خودکار، رگرسیون خودکار برداری، LSTM و مدل غیرخطی خودکار رگرسیو با ورودی های برون زا[9] را انجام داد؛ که عملکرد مدل با استفاده از متریک دقت ارزیابی شد. داده های مورد استفاده برای تجزیه و تحلیل، قیمت پایانی NASDAQ بود. نتایج نشان داد که مدل غیرخطی خودکار رگرسیو با ورودی های برون زا پیش‌بینی‌های دقیقی برای کوتاه‌مدت انجام داده است؛ اما در پیش‌بینی‌های بلندمدت شکست خورده است. همچنین نتیجه گیری شد که مدل‌هایی که

1 Chandar
2 Howells
3 Rajab
4 LeCun
5 Wu
6 DBNs
7 Zhang
8 Hushani
9 NARX

یادگیری ماشین و شاخص‌های فنی را ادغام می‌کنند، می‌توانند با دقت بیشتری پیش‌بینی کنند. شبکه‌های LSTM قادر به یادگیری وابستگی‌های بلندمدت هستند، به گونه‌ای که تأثیری هوشیارانه بر پیش‌بینی سری‌های زمانی دارند. علاوه بر این، دی پرسیو(۲۰۱۷)[1]سه مدل شبکه عصبی بازگشتی را بر روی داده های قیمت سهام گوگل، یعنی مدل شبکه عصبی بازگشتی پایه، واحد بازگشتی دردارو LSTM[2] مقایسه کردند. نتایج نشان داد که مدل LSTM از سایر تکنیک ها بهتر عمل کرده و در یک افق ۵ روزه به دقت ۷۲ درصد دست یافته است. علاوه بر این، نگویین(۲۰۱۹)[3]شبکه پویا LSTM را برای پیش‌بینی قیمت‌های Nifty با استفاده از ویژگی‌های قیمت باز، بالاترین قیمت، پایین ترین قیمت و قیمت بسته شدن به‌عنوان ویژگی‌ها استفاده کردند و به یک ریشه میانگین مربع خطا 0.00859 (RMSE) بر حسب درصد تغییرات روزانه دست یافتند(رئوف،۲۰۲۳).

ماشین بردار پشتیبانی (SVM): ماشین بردار پشتیبانی یک تکنیک یادگیری ماشینی نظارت شده است که خطا را محدود می کند و حاشیه های هندسی را افزایش می دهد. این یک الگوریتم طبقه بندی الگو و رگرسیون است که توسط (کورتز،۱۹۹۵)[4] ارائه شده است. از نظر دقت، ماشین بردار پشتیبان یک الگوریتم جداسازی خطی مهم در مقایسه با سایر طبقه‌بندی‌کننده‌ها است (سرسیلوا،۲۰۱۰)[5]. نایاک(۲۰۱۶)[6]یک مدل پیشبینی بازار سهام روزانه و ماهانه را با استفاده از داده های تاریخی و احساساتی برای بخش های بانک، معدن و نفت توسعه دادند. قیمت‌های تاریخی از یاهو فاینانس به‌دست آمد و یک مجموعه داده احساسات با استفاده از اخبار و توییت‌ها به مدت یک سال ایجاد شد. PCA با عوامل متعدد برای مجموعه داده پراکنده در نظر گرفته شده برای تجزیه و تحلیل احساسات اعمال شد. در این مطالعه، سه الگوریتم - یعنی درخت تقویت شده با تصمیم، SVM و رگرسیون لجستیک - مقایسه شدند و دقت به عنوان یک معیار عملکرد استفاده شد. درخت تقویت تصمیم از رگرسیون لجستیک و SVM دور شد. درخت تقویت‌شده تصمیم‌گیری به ترتیب به دقت ۵۴٫۸، ۷۶، و ۷۶٫۹ درصد برای بخش‌های بانک، معدن و نفت دست یافت. رگرسیون لجستیک به ترتیب به دقت ۶۵٫۴٪، ۶۱٪ و ۴۴٫۲٪ و SVM به دقت ۵۱٪، ۵۹٪ و ۴۴٫۲٪ برای بخش‌های مربوطه دست یافت.

1 Di Persio
2 GRU
3 Nguyen
4 Cortes
5 Srivastava
6 Nayak

الگوریتم یادگیری ماشین پایه: رگرسیون خطی برای پیش‌بینی سهام یا بازار مالی برای پیش‌بینی قیمت آتی رگرسیون سهام استفاده می‌شود و از مدلی مبتنی بر یک یا چند ویژگی مانند قیمت بسته، قیمت باز، حجم و غیره برای پیش‌بینی قیمت سهام استفاده می‌کند. هدف مدل‌سازی رگرسیون شبیه‌سازی رابطه خطی بین متغیرهای وابسته و مستقل است. مدل رگرسیون خطی بهترین خط را تولید می کند که ارتباط بین عوامل مستقل و متغیر وابسته را توصیف می کند.

در این تکنیک، یک خط مستقیم که با معادله (۱) نشان داده می‌شود، ترسیم می‌شود و اطمینان حاصل می‌کند که خط از بالاترین تعداد ممکن از نقاط داده مجموعه داده عبور می‌کند. هنگام ترسیم مقادیر مجموعه داده بر روی یک نمودار، یک خط مستقیم از نظر ریاضی بین نقاط برازش داده می شود تا مجذور فاصله یا اختلاف بین هر نقطه و خط تا حد امکان کوچک باشد. برای هر x داده شده، از خط فرضیه برای پیش بینی مقدار y استفاده می شود. این روش پیش بینی به عنوان رگرسیون خطی شناخته می شود. برای ارزیابی نتایج و بررسی میزان تناسب مدل با خط، از پارامترهایی مانند RMSE، MAE، MSE و R-squared استفاده می‌شود (رئوف،۲۰۲۳):

$$O=Sx+K$$

جایی که O خروجی است، Sx نشان دهنده شیب است و K ثابت است.

K نزدیکترین همسایه (KNN)

KNN یک تکنیک طبقه بندی و رگرسیون است و به عنوان یک یادگیرنده تنبل نامیده می شود، زیرا به دوره زمانی زیادی برای یادگیری نیاز ندارد. یکی از مزایای KNN این است که یکی از ساده ترین الگوریتم های ML است. تنها اقدامی که برای KNN باید انجام شود محاسبه مقدار K و فاصله اقلیدسی است. جنبه یادگیری آهسته این الگوریتم آن را سریعتر از سایر الگوریتم ها می کند. ممکن است برای کلان داده ها به خوبی تعمیم نیابد زیرا از مرحله یادگیری رد می شود. محاسبه فاصله اقلیدسی در رابطه (۲) آورده شده است.

$$(hi,pr)= \sqrt{\sum} (Pr-hi)2$$

(۲) که در آن Pr مقدار پیش بینی شده را نشان می دهد و hi نشان دهنده مقدار داده است.

که در آن Pr نشان دهنده مقدار پیش بینی شده و hi نشان دهنده مقدار داده است.

پیش‌بینی بازار سهام با استفاده از SVM می‌تواند مفیدترین تکنیک برای پیش‌بینی قیمت سهام باشد، زیرا می‌توان از آن به عنوان یک الگوریتم طبقه‌بندی و رگرسیون استفاده کرد.

مقایسه SVM و انواع آن، مانند "Peeling + SVM" و "CC + SVM" نشان می‌دهد که می‌توان پیش‌بینی آن را با روش‌های پیشرفته SVM بهبود بخشید. ماشین بردار پشتیبان شامل یادگیری نظارت شده است که برای دسته بندی جنبه ها با استفاده از جداکننده استفاده می شود. سپس زمانی که داده ها در ابتدا به یک فضای ویژگی با ابعاد بالا نگاشت می شوند، جداکننده کشف می شود. دسته بندی نقاط داده ای را که در فضای n بعدی رخ می دهند پیدا می کند و ابر صفحه بهینه را پیدا می کند. نقاط داده با توجه به موقعیت آنها در رابطه با ابرصفحه ها گروه بندی می شوند. عملکرد الگوریتم SVM را می توان با تنظیم پارامترهایی مانند تنظیم، گاما و پارامترهای هسته SVM افزایش داد. همچنین می تواند برای تجزیه و تحلیل احساسات برای ارزیابی احساسات سرمایه گذاران استفاده شود، که به طور غیر مستقیم بر شرایط بازار تأثیر می گذارد. هم برای مجموعه داده های با ابعاد بالا و هم برای مجموعه داده های مقیاس کوچک مناسب است.

فرضیه بازار کارا و الگوی ماشین بردار پشتیبان

فاما و ماکیل (۱۹۷۰) با توصیف فرضیه بازار کارآمد اظهار می دارند که در حالت تعادل، قیمت ها منعکس کننده تمام اطلاعات مرتبط موجود هنگام قیمت گذاری یک دارایی هستند. این فرضیه از مشاهدات تجربی تغییرات در سری‌های زمانی قیمت‌ها ناشی می‌شود که بسیار شبیه به یک فرآیند گام تصادفی است. به عقیده این نویسندگان، حتی سیستمی که در آن تعدادی سفارش خرید و فروش در کوتاه مدت تولید می شود، به دلیل هزینه های مبادله و پورسانت های مالکیل و فاما (۱۹۷۰) سودآور نیست(برانو[1] و همکاران،۲۰۱۸). علی‌رغم شواهد به‌دست‌آمده برای کارایی بازار، مالکیل و فاما (۱۹۷۰) جستجو برای داده‌های بیشتری را که فرضیه آنها را تأیید یا رد می‌کنند، تشویق می‌کنند. از آن زمان، مقالات دانشگاهی به دنبال این بودند که نشان دهند که قیمت‌های بازار سهام تا حدی قابل پیش‌بینی هستند. ماکیل (۲۰۰۳) نتیجه می گیرد که همه شرکت کنندگان در بازار منطقی نیستند و شکل گیری قیمت نامنظمی وجود دارد که منجر به الگوهای بازده قابل بهره برداری در دوره های زمانی کوتاه می شود. این احتمال را در نظر بگیرید که یک سیستم پیش بینی سودآور ممکن است وجود داشته باشد اما فقط تا نقطه کشف آن. در این حالت، عملکرد آن سیستم زمانی بدتر می شود که شرکت کنندگان بیشتری در بازار شروع به استفاده از آن کنند(برانو و همکاران،۲۰۱۸).

1 Bruno

همانطور که توسط هسو و همکاران(۲۰۱۶)[1]پیشنهاد شده است، توسعه سیستم‌های سودآور مداوم ممکن است شواهدی علیه فرضیه بازار کارآمد باشد. چنین سیستم‌هایی ممکن است از تکنیک‌های محاسباتی فشرده، مانند آن‌هایی که از الگوریتم‌های یادگیری ماشین بهره‌برداری می‌کنند، بهره‌مند شوند. هسو و همکاران (۲۰۱۶) نشان می دهد که الگوریتم های یادگیری ماشین معمولاً از سری های زمانی مالی برای ارزیابی قابلیت های پیش بینی خود استفاده می کنند. در این زمینه بالینگ و همکاران(۲۰۱۵) و گرلین[2]و همکاران (۲۰۱۶) نتایج خوبی در پیش بینی ها هنگام استفاده از طبقه بندی کننده هایی مانند لگوهای SVM، k-Nearest Neighbors (KNN) و شبکه های عصبی و درخت های تصمیم به دست آورده اند.

زبیکوفسکی[3] (۲۰۱۵) برای مثال، ازالگوی ماشین بردار پشتیبان و یک روش انتخاب متغیر پیش‌بینی‌کننده در شاخص‌های تحلیل فنی استفاده شده است. در تلاش برای توسعه یک استراتژی معاملات بازار بهینه، ۵ از n=k، میانگین برای پیش بینی نوسانات بازار و الگوی ساختاری برداری پشتیبان بُرای پیش بینی قیمت ها در بازار سهام هند استفاده کرد. نویسندگان داده های روزانه را برای تخمین قیمت ها به مدت دو روز تجزیه و تحلیل کردند.۲ جهت بازار سهام را مورد مطالعه قرار دادند و در این کار، طبقه بندی کننده ها را ارزیابی کردند. در مطالعه بالینگز و همکاران(۲۰۱۵) ، داده های سالانه بیش از ۵۰۰۰ شرکت اروپایی در طبقه بندی کننده هایی مانند رگرسیون لجستیک، شبکه های عصبی، KNN و SVM استفاده شد. برای تعیین جهت سهام شرکت مربوطه در سال بعد از طبقه بندی کننده ها استفاده می شود. نویسندگان نتایج این طبقه‌بندی‌کننده‌ها را با رویکردهای مجموعه‌ای، مانند جنگل‌های تصادفی و کارخانه هسته مقایسه کردند. به اصطلاح تکنیک‌های گروهی شامل طبقه‌بندی‌کننده‌های متعدد، معمولاً از یک نوع یا الگوریتم، می‌شود که منجر به طبقه‌بندی‌های مستقل می‌شود اما با برخی روش‌های تصمیم‌گیری برای تعیین یک طبقه‌بندی نهایی منفرد. با به کارگیری این تکنیک ها، متغیرهای پیش بینی آن ها با در نظر گرفتن ترازنامه و صورت های مالی شرکت ها محاسبه شد. بر اساس پیش‌بینی جهتی که یک سهام خاص در طول سال در پیش می‌گیرد، نویسندگان نشان دادند که چگونه می‌توان یک استراتژی سودآور ساخت(برانو،۲۰۱۸). همچنین کومار[5] و همکاران

1 Hsu
2 Gerlein
3 Zbikowski
4 SVR
5 Kumar

(۲۰۱۶) به دنبال پیش بینی جهت قیمت سهام هستند. باراک و مدرس (۱۳۹۴) با انتقاد از رویکردهای طبقه‌بندی تنها بر اساس جهت‌گیری بازار سهام، سعی در پیش‌بینی ریسک‌ها و بازده‌های سهام تهران داشت. نویسندگان یک روش انتخاب متغیر پیش‌بینی‌کننده و درخت تصمیم و مکانیسم‌های شبکه عصبی را پیشنهاد کردند. متغیرهای پیش بینی استفاده شده توسط باراک و مدرس (۱۳۹۴) از صورت های مالی منتشر شده توسط شرکت ها ساخته شده اند. زبیکوفسکی(۲۰۱۵) از مقادیر شاخص TA به عنوان متغیرهای پیش بینی در مدل پیش بینی روند کوتاه مدت خود استفاده کرد. زبیکوفسکی (۲۰۱۵) یک الگوی SVM اصلاح شده با انتخاب متغیر تعیین شده توسط نمرات فیشر پیشنهاد کرد، روشی که برای رتبه‌بندی تعاریف به کلاس‌ها بر اساس متغیرهای پیش‌بینی‌کننده خاص استفاده می‌شود. متغیرهای فنی مورد استفاده توسط زبیکوفسکی (۲۰۱۵) شامل حجم تعادل (OBV)، شاخص قدرت نسبی (RSI) و نوسانگر ویلیامز است. نتایج به‌دست‌آمده از طریق رویکرد SVM از نتایج مبتنی بر استراتژی خرید و نگه‌داری در مطالعه زبیکوفسکی فراتر رفت. یه و همکاران (۲۰۱۱) و لو[1] و همکاران (۲۰۰۹) سیستم های مبتنی بر SVR را برای پیش بینی شاخص های TAIEX و Nikkei 225، هر دو با استفاده از داده های روزانه، به کار برد.

برخی از نویسندگان از الگوریتم های یادگیری ماشین ترکیبی برای افزایش عملکرد پیش بینی استفاده کرده اند. شیائو و همکاران (۲۰۱۳) برای مثال، ادغام شبکه‌های عصبی مختلف با شبکه بردار پشتیبان برای پیش‌بینی مقادیر روزانه شاخص‌های سهام را پیشنهاد کرد. نایاک و همکاران(۲۰۱۶) با استفاده از شاخص های تحلیل تکنیکال به عنوان متغیرهای ورودی؛ یک سیستم ترکیبی SVM و KNN برای پیش‌بینی مقادیر شاخص پیشنهاد کرد. نتایج نایاک(۲۰۱۶) بهتر از شبکه های عصبی سنتی بود.

پاتل[2] و همکاران(۲۰۱۵) کارایی پیش بینی های قیمت بسته شدن روزانه انجام شده توسط الگوی های SVM، ، جنگل تصادفی و هیبریدهای شبکه عصبی را ارزیابی کرد و نتایج را با استفاده ایزوله از همین الگوریتم ها مقایسه کرد. وی به این نتیجه رسیدند که استفاده ترکیبی از این الگوریتم ها نتایج بهتری ارائه می دهد. در نهایت، دش[3] (۲۰۱۶) رویکردی به شبکه‌های عصبی با شاخص های تحلیل تکنیکال را در مقایسه با الگوریتم‌های یادگیری

1 Lu
2 Patel
3 Dash

ماشین سنتی در تصمیم‌گیری‌های معاملات سهام بررسی کردند و به این نتیجه رسیدند که فرضیه بازار کارا با پیشبینی سبد سهام در الگوی SVM نسبت به سایر الگوهای یادگیری ماشین بهتر عمل می کند.

مقایسه الگوی ماشین بردار پشتیبان با سایر الگوهای یادگیری ماشین

توسعه محاسبات نه تنها توسعه الگوریتم‌های پیش‌بینی پیچیده را امکان‌پذیر کرده است، بلکه تجزیه و تحلیل کمی داده‌های با فرکانس بالا را نیز ممکن کرده است، همانطور که در مطالعه براونلی[1] و گالو (۲۰۰۶) انجام شد. این نوع داده نیاز به درمان متفاوتی نسبت به داده‌های فرکانس پایین‌تر سنتی دارد. براونلی و گالو (۲۰۰۶) راه‌هایی را برای درمان داده‌های فرکانس بالا با توجه به نقاط پرت و بی‌نظمی‌های زمانی پیشنهاد می کنند. پس از درمان، داده‌های فرکانس بالا می‌توانند برای توسعه استراتژی‌های معاملاتی بازار بسیار سریع، به نام تجارت با فرکانس بالا مورد استفاده قرار گیرند. این نوع معامله در بازارهای مالی امروزی رایج شده است و اثرات آن توسط نویسندگانی مانند لی و همکاران(۲۰۱۳)[2] مورد مطالعه قرار گرفته است.

با توجه به پیشرفت های استفاده از ماشین بردار پشتیبان که توسط آراجو(۲۰۱۵)[3] و همکاران مورد بحث قرار گرفت. (۲۰۱۵، ص ۴۰۸۲)، فضای زیادی برای توسعه مدل ها در این زمینه وجود دارد. به عنوان مثال، Araújo و همکاران. (۲۰۱۵) یک مدل ریاضی جدید برای پیش بینی تغییرات قیمت شرکت های فهرست شده که در ثانیه اندازه گیری می شود، توسعه داد. در همین حال، ماناهوف و همکاران(۲۰۱۴) سودآوری یک الگوریتم یادگیری ژنتیکی را برای بازار فارکس بررسی کردند که برای تغییرات قیمت اندازه‌گیری شده در دقیقه اعمال می‌شود. همچنین مدل SVR برای قیمت های لحظه ای در بازار چین در مطالعه کیو و ژانگ[4](2016) اعمال می شود. مطالعات دیگر تکنیک‌های پیش‌بینی را برای داده‌های فرکانس بالاتر نیز اعمال کرده‌اند. براونلی و گالو (۲۰۰۶) به عنوان مثال، استفاده از مدل یادگیری بردار پشتیبان را در تغییرات قیمت فوری پیشنهاد داد، یعنی برای تیک هایی که بر اساس هر معامله خرید یا فروش انجام شده متفاوت است.

مقایسه الگوی بردار پشتیبان و الگوی میانگین متحرک

1 Brownlees
2 Lee
3 Araujo
4 Qu and Zhang

تحقیقات متعددی حول بهبود اثربخشی مدل میانگین متحرک برای بهینه‌سازی پورتفولیوها متمرکز شده‌اند. همانطور که مدل‌های ماشین یادگیری پیشرفت کرده‌اند، چند مطالعه ادغام آنها با مدل‌های میانگین متحرک را بررسی کرده‌اند. پایوا[1] و همکاران (۲۰۱۹) مدلی را با استفاده از طبقه‌بندی‌کننده ماشین بردار پشتیبان (SVM) و مدل میانگین متحرک برای انتخاب پرتفوی برای سرمایه‌گذاری‌های معاملاتی روزانه پیشنهاد کرد. این مدل با استفاده از شاخص بورس سائوپائولو ارزیابی شد، که از نظر بازده و ریسک از سایر مدل‌های معیار بهتر عمل کرد. این مطالعه همچنین تأثیر هزینه‌های کارگزاری را بر خرید و فروش سهام در بازار برزیل و عملکرد طبقه‌بندی کننده در بهینه‌سازی پرتفوی ارزیابی کرد.

وانگ[2] و همکاران (۲۰۲۰) از مدل‌های حافظه بلندمدت ماشین یادگیری بردار پشتیبان و میانگین متحرک برای ساخت مجموعه‌ای استفاده کرد که وابستگی‌های بلندمدت داده‌های سری زمانی مالی را جمع‌آوری می‌کند و دارایی‌هایی با بازده بالقوه بالاتر را انتخاب می‌کند. نتایج مدل آنها در شاخص بورس ۱۰۰ بریتانیا (FTSE 100) بین مارس ۱۹۹۴ و مارس ۲۰۱۹ نشان می دهد که مدل پیشنهادی ترکیب حافظه بلندمدت ماشین یادگیری بردار پشتیبان و میانگین متحرک در مقایسه با سایر مدل های معیار عملکرد برتر در بازده انباشته، نسبت شارپ و متوسط بازده به ریسک دارد.

چن و همکاران (۲۰۲۱) ادغام نتایج پیش‌بینی مدل تقویت شیب شدید هیبریدی ژا[3] بررسی کرد که پارامترهای آن با الگوریتم کرم شب تاب بهبود یافته با مدل میانگین متحرک در بازار سهام شانگهای بهینه‌سازی شده‌اند. نتایج نشان دهنده برتری مدل آنها از نظر بازده و ریسک در مقایسه با مدل های معیار است. چایسری[4] (۲۰۲۲) یک مدل جدید ساخت سبد دارای ترکیب CNN و LSTM دو طرفه برای پیش‌بینی قیمت سهام و بهینه‌سازی پرتفوی پیشنهاد کردند. نتایج شاخص بورس تایلند ۵۰ (SET50) نشان می دهد که مدل پیشنهادی نسبت به مدل های مقایسه در مورد ریسک، بازده و نسبت شارپ بهتر عمل می کند. تا و همکاران (۲۰۲۲) استفاده از LSTM را برای پیش‌بینی حرکات سهام و ایجاد یک پرتفوی کارآمد پیشنهاد کرد. مدل‌های بهینه‌سازی پورتفولیو برای بررسی عملکرد، از جمله مدل‌سازی هم وزن و مدل‌سازی بهینه‌سازی بهینه‌سازی میانگین متحرک استفاده شد. نتایج نشان داد که مدل پیش‌بینی LSTM از دقت بالایی برخوردار بوده و از سایر مدل‌های

1 Paiva
2 Wang
3 XGBoost
4 Chaysiri

پیش‌بینی بهتر عمل می‌کند. آنها تایید کردند که ترکیب LSTM با مدل میانگین متحرک برای ساخت نمونه کارها مناسب است.

جدای از استفاده از مدل‌های یادگیری ماشین برای بهینه‌سازی پورتفولیو، تحقیقات دیگری به بهبود مدل میانگین متحرک اختصاص یافته است. فریتاس و همکاران (۲۰۰۹) یک مدل جدید بهینه‌سازی پورتفولیو را پیشنهاد کرد که از پیش‌بینی‌کننده‌های شبکه عصبی برای گرفتن فرصت‌های سرمایه‌گذاری کوتاه‌مدت استفاده می‌کند. این مدل یک معیار ریسک را بر اساس خطاهای پیش‌بینی استخراج می‌کند و پیش‌بینی‌کننده‌هایی را با پروفایل‌های خطای زوجی کم و مکمل انتخاب می‌کند تا تنوع کارآمد را ممکن کند. ارزیابی مدل با استفاده از داده‌های واقعی از بازار سهام برزیل نشان داد که با بهره‌گیری از فرصت‌های کوتاه‌مدت و ایجاد خطاهای پیش‌بینی عادی، علی‌رغم غیرعادی بودن سری‌های زمانی بازده سهام، از مدل میانگین متحرک و شاخص بازار بهتر عمل می‌کند.

تا و همکاران (۲۰۲۱) از پنج مدل مختلف پیش‌بینی شامل: RF، SVR، LSTM، پرسپترون چند لایه عمیق (Dپرسپترون چندلایه) و CNN استفاده کرد. این مدل‌ها برای پیش‌گزینی سهام برای بهینه‌سازی پرتفوی مورد استفاده قرار گرفتند و نتایج پیش‌بینی‌کننده در یک مدل میانگین متحرک با پیش‌بینی (MVF) گنجانده شد. این تحقیق داده‌های تاریخی سهام مؤلفه شاخص ۱۰۰ اوراق بهادار چین (CSI 100) را از سال ۲۰۰۷ تا ۲۰۱۵ تجزیه و تحلیل کرد. مطالعه تا (۲۰۲۱) به این نتیجه رسید که مدل ترکیب SVM و میانگین متحرک است. نتایج آنها برای این ترکیب بیان نمود که این الگو مناسب‌ترین مدل برای معاملات سرمایه‌گذاری روزانه است. برانو و همکاران(۲۰۱۸) از تحلیل الگوی SVM با رویکرد استفاده از خودرگرسیون ساختاری برای دوره یادگیری ؛ بیان نمود که توسعه مدل‌های قیمت پیش‌بینی‌کننده برای بازار سهام چالش برانگیز است، اما در هنگام ایجاد استراتژی‌های معاملاتی بازار مالی سودآور، وظیفه مهمی است. روش‌های محاسباتی فشرده، با استفاده از قیمت‌های گذشته، برای تسهیل مدیریت بهتر ریسک بازار برای سرمایه‌گذاران و سفته‌بازان توسعه یافته‌اند. از تکنیک‌های یادگیری ماشینی موجود، این مطالعه از SVR استفاده می‌کند و عملکرد آن را در سهام‌های مختلف برزیل، آمریکا و چین با ویژگی‌های مختلف، به‌عنوان مثال، تراشه‌های کوچک یا بلو چیپ اندازه‌گیری می‌کند. متغیرهای پیش بینی با استفاده از شاخص های TA بر روی قیمت دارایی ها محاسبه می شوند. نتایج نشان‌دهنده بزرگی میانگین مربعات خطا برای سه هسته رایج در ادبیات، با

استفاده از استراتژی‌های آموزش الگوریتم خاص با فرکانس‌های قیمتی متفاوت در روز و دقیقه است. نتایج با نتایج یک مدل مبتنی بر گام تصادفی در تضاد است(برانو و همکاران،۲۰۱۸).

مطالعه برانو(۲۰۱۸) نشان می‌دهد که با استفاده از یک مجموعه آموزشی ثابت بر روی قیمت‌های روزانه، می‌توان خطاهای پیش‌بینی کمتری را در مجموعه تست نسبت به مجموعه آموزشی هنگام استفاده از یک هسته خطی به دست آورد. علاوه بر این، این هسته برای پیش‌بینی قیمت نسبت به هسته‌های شعاعی و چند جمله‌ای در مورد قیمت‌های روزانه و مدل‌های آموزشی ثابت مناسب‌تر بود و از مدل تصادفی برای برخی از سهام طبقه‌بندی‌شده به‌عنوان تراشه‌های آبی و کوچک در سه کشور مورد مطالعه عملکرد بهتری داشت. با این حال، افزایش فرکانس قیمت به دقیقه قدرت پیش بینی مدل را با استفاده از یک دوره آموزشی ثابت کاهش داد. به طور خاص، SVR نتایج پیش‌بینی پایین تری نسبت به مدل گام تصادفی برای تقریباً تمام سهام مورد مطالعه در قیمت‌های لحظه‌ای، با استفاده از آموزش ثابت، بدون توجه به تابع هسته اتخاذ شده، به‌دست آورد. وی بیان نمود مدل های به روز شده دوره ای شواهد مهمی ارائه کردند. در این موارد، استفاده از هسته‌های خطی و شعاعی منجر به خطاهای کوچک‌تری نسبت به مدل گام تصادفی برای تقریباً تمام قیمت‌های روزانه سهام شد. تنها استثنا سهامی بود که نرخ داده گمشده بالایی داشت. به روز رسانی مداوم مدل همچنین در فرکانس قیمت لحظه ای مفید بود و مدل های SVR با هسته های خطی و شعاعی نتایج بهتری نسبت به مدل پیاده روی تصادفی در هنگام استفاده از این استراتژی به دست آوردند. برای تأکید بر ثبات پیش‌بینی‌ها در بلندمدت، ما یک دوره قیمتی ۲ ساله را برای سهام منتخب برزیل پردازش کردیم که نتایج بهتری را با مدلی که دائماً به‌روز می‌شود تأیید کرد. تحلیل های ارائه شده در این مطالعه نشان می دهد که به روز رسانی دوره ای مدل SVR در مقایسه با استفاده از یک مدل صلب بدون به روز رسانی دوره ای، میانگین مربعات خطا را کاهش می دهد. این نتیجه با نتیجه هسو و همکاران (۲۰۱۶)در تضاد است؛ که هنگام استفاده از پنجره کشویی روی داده های آموزشی به عملکرد بهتری نرسیدند.

بر پایه دسته بندی الگوهای یادگیری ماشین ، سهم مهم این مقاله مقایسه نتایج پیش‌بینی قیمت مدل‌های SVR ارائه‌شده با مدل های دیگر و مدل های بر پایه گام تصادفی است که بر اساس آن بازارها در بلندمدت غیرقابل پیش‌بینی هستند. در این رابطه، نتایج ارائه شده

در اینجا نشان می‌دهد که برخی از مدل‌های SVR، با به‌روزرسانی‌های دوره‌ای یا ثابت، ممکن است عملکرد بهتری نسبت به پیش‌بینی تصادفی، به ویژه با استفاده از هسته خطی داشته باشند.

از آنجایی که تمرکز مطالعه حاضر بر تجزیه و تحلیل خطاهای پیش‌بینی قیمت دارایی است، می‌توان مدل‌های مدیریت ریسک را با استفاده از برآوردهای مبتنی بر SVR ساخت. محدودیت های مواجهه ممکن است با ارزیابی خطاهای مدل به دست آید. بنابراین این مطالعه مبنایی را برای ساخت سیستم‌هایی فراهم می‌کند که در عین اینکه مستقیماً فرضیه بازار کارآمد را ارزیابی نمی‌کنند، مطالعه کارایی بازار و تحلیل ریسک را ممکن می‌سازند. این مطالعه به مانند برانو و همکاران(۲۰۱۸) نتایجی را با استفاده از SVR به دست آورد که بهتر از مدل تصادفی میانگین بازده صفر بود.

علیرغم مقایسه نرخ های روزانه در مطالعه برانو(۲۰۱۸) با استفاده از معاملات لحظه به لحظه با فرکانس بالا، این مقاله تنها یک الگوریتم پیش بینی مبتنی بر یادگیری ماشین را در نظر می گیرد. علاوه بر این، مدل SVR امکان آزمایش بسیاری از توابع هسته را فراهم می کند، در حالی که این مطالعه تنها به سه مورد از رایج ترین در ادبیات محدود می شود. لازم به ذکر است که برای شبیه‌سازی قوی‌تر استراتژی‌های بورس با فرکانس بالا، باید هزینه‌های مبادله، تاخیرهای شبکه ارتباطی، تمایز بین قیمت بازار و ارزش واقعی معامله خرید یا فروش (لغزش) و معامله را شامل شود. دراپ و همکاران(۲۰۲۳) بیان نمود که مطالعه حاضر از نه روش یادگیری ماشینی (ML) برای پیش‌بینی جریان موج در یک روش ابتکاری و جامع استفاده می‌کند. بر خلاف تحقیقات قبلی، که اغلب عوامل در نظر گرفته شده هنگام استفاده از روش‌های ML برای پیش‌بینی جریان موج را محدود می‌کردند، این رویکرد یک دیدگاه کل‌نگر دارد. این تجزیه و تحلیل طیف گسترده ای از پارامترهای ساحلی مهم را در نظر می گیرد، از جمله ارزش مازاد بر ۲٪ برای راه اندازی، راه اندازی، گشت و گذار کل سواش، سوش تصادفی، موج زیر گرانی، ارتفاع موج قابل توجه، دوره موج اوج، شیب ساحلی، و اندازه متوسط رسوب. . عملکرد مدل، تفسیرپذیری، و عملی بودن ارزیابی شد. یافته‌های این مطالعه نشان می‌دهد که مدل‌های خطی، در حالی که در بسیاری از کاربردها ارزشمند هستند، در درک پیچیدگی این مجموعه داده کافی نیستند. از سوی دیگر، ما متوجه شدیم که مدل‌های غیرخطی برای دستیابی به پیش‌بینی‌های دقیق موج ضروری هستند، که بر اهمیت آنها در زمینه تحقیق تأکید می‌کند. در چارچوب این

بررسی، مشخص شد که روند موج تحت‌تاثیر اندازه رسوب متوسط، ارتفاع موج قابل توجه و شیب ساحلی قرار می‌گیرد. مهندسان و مدیران ساحلی می توانند از این یافته ها برای طراحی سازه های ساحلی انعطاف پذیرتر و ارزیابی خطرات ناشی از خطرات ساحلی استفاده کنند. برای بهبود دقت پیش‌بینی، این تحقیق بر انتخاب ویژگی و مدیریت پیچیدگی مدل تأکید کرد. این تحقیق ثابت می‌کند که الگوریتم‌های یادگیری ماشینی می‌توانند جریان موج را پیش‌بینی کنند و به مهندسی و مدیریت ساحلی کمک کنند. این مدل ها به ساخت زیرساخت های ساحلی و پیش بینی خطرات ساحلی کمک می کنند.

اشرف زاده و همکاران(۲۰۲۳) بیان نمود که گنجاندن پیش‌بینی بازده در بهینه‌سازی پرتفوی می‌تواند با انتخاب سهامی که انتظار می‌رود در آینده عملکرد خوبی داشته باشند، بهینه‌سازی سبد را کارآمدتر کند. این مقاله یک روش ترکیبی را پیشنهاد می‌کند که یک شبکه عصبی کانولوشنال (CNN)را با فراپارامترهای بهینه‌شده توسط بهینه‌سازی ازدحام ذرات (PSO)برای پیش‌انتخاب سهام و یک مدل میانگین واریانس با پیش‌بینی (MVF)برای بهینه‌سازی پورتفولیو ادغام می‌کند. در مرحله پیش انتخاب سهام، برای کاهش پیچیدگی محاسباتی مدل، شبکه CNN بر روی سهام خوشه‌بندی شده از طریق روش K-means به جای آموزش روی هر سهام آموزش داده می‌شود. مدل پیشنهادی همچنین شامل یک روش انتخاب ویژگی جدید است که ویژگی‌ها را بر اساس تأثیر آنها بر پیش‌بینی بازده سهام برای پیش‌بینی‌های دقیق‌تر وزن می‌کند.

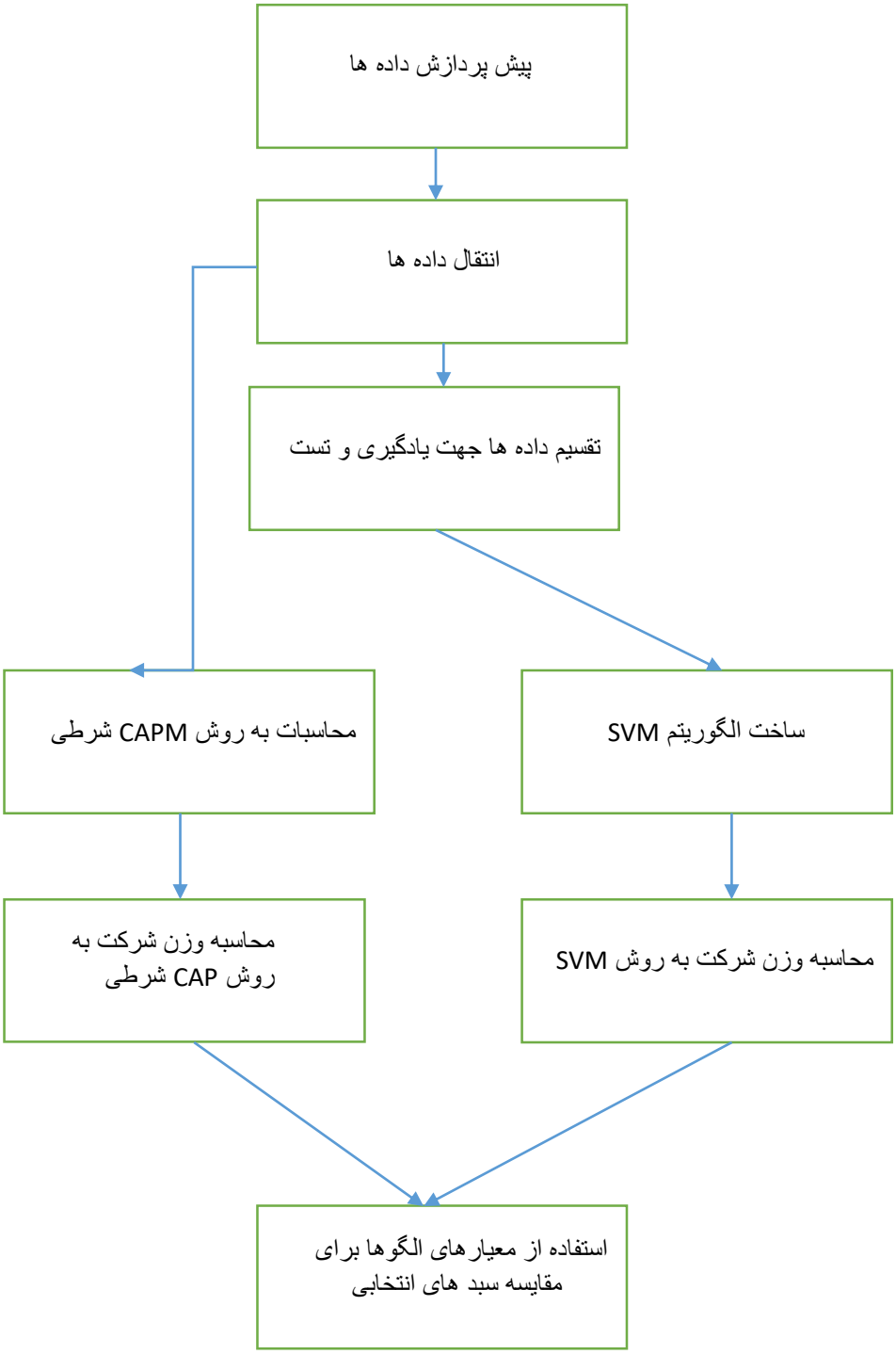
روش شناسی و معرفی الگو

انتخاب سبد بهبنه سهام ، پیشبینی بازار سهام و یا تحلیل قیمت سهام شرکت از جمله مواردی است که طیفی از الگوهای رگرسیون و الگوهای ماشین یادگیری مثل الگوی ماشین بردار پشتیبان در خلال ادبیات موضوع به تحلیل ها بر اساس این الگوها پرداخته اند. اما آنچه در این مطالعه مورد توجه است؛ معرفی روش شناسی تحقیق ، تحلیل یافته ها و نتیجه گیری از داده کاوی بازار سرمایه ایران طی دوره ۱۴۰۰ تا ۱۴۰۱ است.

این پژوهش از جمله مطالعات کاربردی است که با استفاده از شواهد تجربی اقدام به آزمون فرضیه نموده است. فرضیه پژوهش مطرح نموده که سبد انتخابی به روش ماشین بردار پشتیبان نسبت به سبد انتخابی از روش رگرسیون الگوی قیمتگذاری دارایی های سرمایه ای؛ بهینه تر و با بازدهی سهام بیشتر نسبت به الگوهایرگرسیونCAPM انتخاب خواهد نمود.

روش ماشین بردار پشتیبان یکی از روش های یادگیری با نظارت است که از آن برای طبقه بندی و رگرسیون استفاده می کنند.. این روش از جمله روش های نسبتاً جدیدی است که در سال های اخیر کارایی خوبی نسبت به روش های قدیمی تر برا ی طبقه بندی نشان داده است. مبنا ی کاری دسته بند ی کننده SVM دسته بندی خطی داده ها است و در تقسیم خطی داده ها سعی می کنیم خطی را انتخاب کنیم که حاشیه اطمینان بیشتری داشته باشد. حل معادله پیدا کردن خط بهینه برای داده ها به وسیله روش های یادگیری ماشین برای طبقه بندی متغیر هدف که روش های شناخته شده ای در حل مسائل محدودیت دار هستند صورت می گیرد. قبل از تقسیمِ خطی برای اینکه ماشین بتواند داده های با پیچیدگی بالا را دسته بندی کند داده ها را به وسیله تابعِ phi به فضا ی با ابعاد خیلی بالاتر می بریم. برا ی اینکه بتوانیم مسئله ابعاد خیلی بالا را با استفاده از این روش ها حل کنیم از قضیه دوگانی لاگرانژ برا ی تبدیلِ مسئله مینیمم سازی مورد نظر به فرم دوگانی آن که در آن به جا ی تابع پیچیده phi که ما را به فضایی با ابعاد بالا می برد، تابعِ ساده تری به نامِ تابع هسته که ضرب بردار ی تابع phi است ظاهر می شود استفاده می کنیم. از توابع هسته مختلفی از جمله هسته های نمایی، چندجمله ای و سیگموید می توان استفاده نمود.

جهت انتخاب بهینه سبد سهام از دو رویکرد CAPM و SVM برای انتخاب وزن اختصاص یافته و از روش های تحلیلی برداری با درنظر گرفتن سه لایه بردار یادگیری برای الگوی ماشین بردار پژوهش استفاده شده است . در شکل ۱مراحل تحلیلی انتخاب سبد بهینه سهام بیان شده است.



شکل ۱: مراحل تحلیل الگوی پژوهش

نوآوری پژوهش

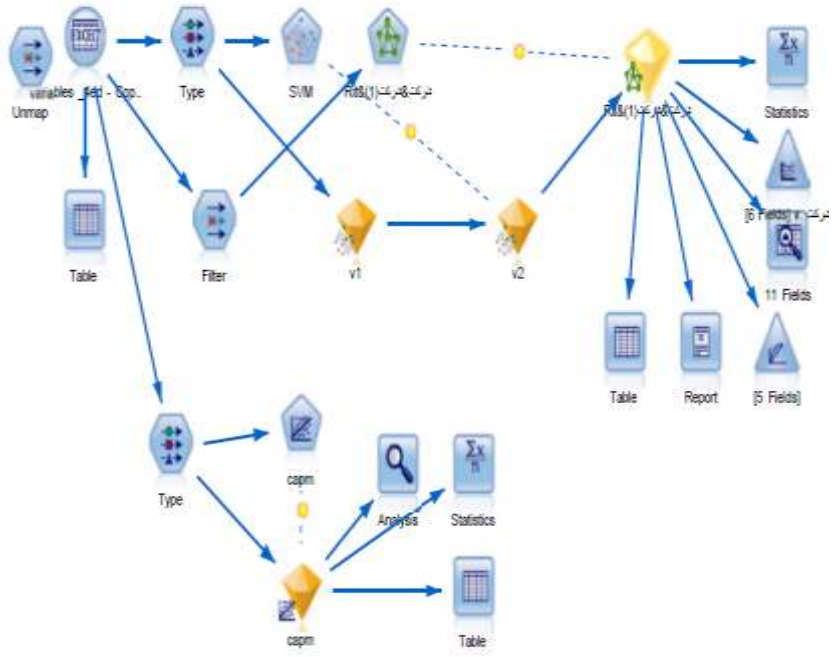
الگوهای قیمت گذاری دارایی های سرمایه ای جهت بهینه سازی سبد سهام در ادبیات قبلی مورد سنجش و ارزیابی قرار گرفته است. همچنین در زمینه بهینه سازی سبد سهام از روش های مختلفی مثل انواع الگوهای فاما و شبکه عصبی قبلا صورت پذیرفته است. الگوی ماشین بردار جهت بهینه سازی سبد سهام از جمله الگوهایی است که طبق پژوهش وانگ و همکاران(۲۰۲۳) قابلیت تفسیر بهتر نتایج را دارا می باشد. اما وانگ و همکاران(۲۰۲۳) به مقایسه این الگو با الگوی قیمت گذاری دارایی های سرمایه ای نپرداختند. بنابراین این پژوهش از جنبه تحلیل بهینه سازی سبد سهام با استفاده از ماشین بردار پشتیبان برای داده های بازار سرمایه ایران و نیز مقایسه قدرت پیشبینی بازدهی سبد سهام با استفاده از روش قیمت گذاری دارایی های سرمایه ای دارای نوآوری بوده و به ادبیات موضوع خواهد افزود.

نتایج تحلیل الگوی پژوهش

در ابتدا مجموعه داده جمع آوری شده و به آماده ساز ی و پیش پردازش داده ها پرداخته می شود. در آماده ساز ی و پیش پردازش داده ها از روش های مختلفی استفاده می شود. اول این که برخی ویژگی ها دارا ی مقادیر منحصر به فرد هستند. این ویژگی ها نمی توانند دانش مفیدی را در مجموعه داده ایجاد کنند. لذا این مجموعه ویژگی ها باید از داده ها حذف شوند. به طور نمونه می توان به ویژگی نام شرکت اشاره نمود. همچنین ممکن است برخی تراکنش ها دارا ی مقادیر مفقود فراوان باشند. لذا این تراکنش ها نیز باید از مجموعه داده ها حذف شوند. از طرفی ممکن است، مقادیر برخی ویژگی ها دارای مقادیر نویز و مفقود باشند لذا این مقادیر نیز باید در مجموعه داده اصالح شوند. مرحله بعدی نوبت به استفاده از ابزار کشف آنومالی پرداخته می شود. داده هایی که در نقاط خارج از قانون مجموعه داده قرار دارند شناسایی شده و ح ذف می شوند. برا ی اینکه بتوان رو ی داده ها به عنوان ورود ی کار کرد می بایست ویژگی هایی را از آن ها استخراج نمود. به طور معمول پیش از انتخاب و استخراج ویژگی ها، برخی عملیات پیش پردازش بر رو ی داده ها انجام می شود.

در مرحله انتقال داده ها داده ها در دامنه های درست قرار می گیرند . بدین معنا که داده ها باید به رنج هایی که در سیستم مشخص شده است منتقل شوند و داده های خارج از رنج، داده ها ی مشکل دار بوده و می بایست حذف شوند. داده های می بایست در محدوده

درست قرار بگیرند همچنین در این مرحله متغیرهای ورودی و متغیر مورد هدف برای یادگیری و انتخاب شرکت با استفاده از وزن دهی انتخاب می شود. که در شکل ۲ که مدل بردار پشتیبان پژوهش را نشان می دهد بیان شده است.



شکل۲: الگوی طراحی شده پژوهش

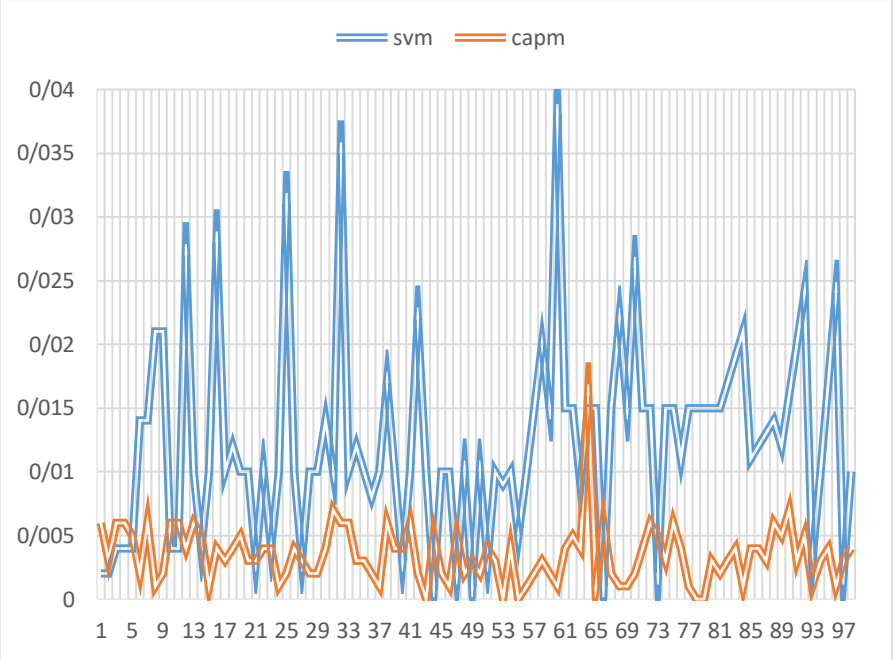
الگوی ماشین بردار پشتیبان پژوهش از سه بردار پشتیبان و یادگیری و تحلیل v1 ، v2 و بردار شرکت ها تشکیل شده است. نتایج تحلیل الگوی SVM در جدول ۱ گزارش شده است:

جدول۱: نتایج محاسبات سبد های انتخابی به روش بردار پشتیبان(SVM) و الگوی CAPM

مشخصات سبد انتخابی	الگوی SVM	الگوی CAPM
تعداد انتخابی	101	101
میانگین بازدهی	0.033	-0.01
حداقل	-0.059	-0.327
حداکثر	0.65	0.155
دامنه	0.109	0.482
واریانس	0.0001	0.003
شاخص SD	0.021	0.055
شاخص SEM	0.001	0.005

همانگونه که در جدول ۱ مشخص شده است میانگین بازدهی سبد انتخابی روش SVM بیش از میانگین انتخابی سبد CAPM می باشد. از طرفی سبد انتخابی در روش ماشین بردار پشتیبان از نظر دو شاخص SD و SEM با توجه به اعداد کوچکتر بدست آمده، بهتر و کاراتر انتخاب شده است.

در نمودار ۳ نتایج بازدهی بازار ، بازدهی سبد سهام در روش ماشین بردار و بازدهی سهام در روش CAPM به تصویر کشیده شده است:



نمودار ۱: مقایسه سبد های سهام انتخابی در دو روش SVM و CAPM

همانگونه که در نمودار ۳ مشخص است، میانگین بازدهی سالانه برای سبد سهام انتخابی بر مبنای ماشین بردار پشتیبان بیشتر از میانگین بازدهی سالانه برای سبد سهام انتخابی بر مبنای الگوی قیمت گذاری دارایی های سرمایه ای است.

نتایج آزمون فرضیه

سوال پژوهش مطرح نمود که بر اساس پیش‌بینی‌های SVM، می‌توان یک الگوریتم معاملاتی غیرمعمول منجر به بازدهی بالاتری نسبت به یک سبد استاتیک متوازن می شود.

در پاسخ به این سوال طبق معیارهای محاسبه شده در جدول ۳-۱ مشخص شد که روش ماشین بردار پشتیبان دارای شاخص SD کمتر و عدد شاخص SEM کمتر نسبت به سبد بهینه شده الگوی CAPM استو براین اساس نتیجه گرفته می شود که استراتژی معاملاتی

بر مبنای مدلSVM منجر به انتخاب سبد سهام با بازدهی بالاتر می شود. این نتیجه از مقایسه روند بازدهی دو سبد در ۱۰۱ انتخاب توسط ۲ الگو در نمودار ۳ نشان داده شده است.

نتیجه گیری و پیشنهادها

با توجه به گسترش بازارهای مالی و توجه زیاد محققین در این زمینه به منظور استفاده از سیستم های قاعده محور به منظور تصمیم گیری ها ی سریع و با ریسک حداقل و به دور از اشتباهات انسانی، طراحی و توسعه و یا بهبود این سیستم ها می تواند مزیت رقابتی بسیار خوبی برا ی سرمایه گذاران باشد. از این رو، در این تحقیق، یک روش مبتنی بر مجموعه راف و تحلیل سلسله مراتبی برا ی تشخیص پارامترهای تاثیرگذار و پیش بینی انتخاب مناسب سهام برای سبد بهینه سهام ارائه شده است. در پژوهش ارائه شده در این تحقیق تعداد ویژگی ها کاهش یافته است تا سربار محاسباتی کاهش یابد و همچنین در مواردی استفاده از ویژگی هایی که تاث یری در خروج ی ندارند و ممکن است باعث ایجاد نویز شود، از الگوریتم SVM با بردار شارپ در جهت تشخیص ویژگی های موثر استفاده شده است و تنها ویژگی هایی انتخاب شده اند که دارا ی بیشترین تاثیرگذاری در خروج ی باشند و بتوان نتایج دقیق تر را با سربار محاسباتی پایین تر ی دست پیدا کرد. از طرف دیگری در این تحقیق از روش بهینه سازی سبد سهام CAPM استفاده شده است که نتایج آن با الگوی SVM در نمودار ۳-۳ مقایسه شد. در این مطالعه، ما یک رویکرد بهینه‌سازی سبد بر اساس پیش‌انتخاب سهام با استفاده از مدل SVM ارائه کرده‌ایم. همانطور که در بخش نتایج نشان داده شد، روش طبقه بندی ما به نرخ دقت ۰٫۰۱٪ دست یافته است. پس از ارزیابی مدل، دارایی‌ها انتخاب شدند و بهینه‌سازی پورتفولیو انجام شد. نتایج بهینه‌سازی با یک معیار بازار و یک نمونه کار ساده با در نظر گرفتن شاخص‌های عملکرد متفاوت مقایسه شد. پورتفولیوی بهینه‌سازی شده SVM با حداکثرسازی شارپ، جایگزین کارایی بالاتری را هم در فراتر از بازده مورد انتظار و هم در بازیابی سریع‌تر پس از دوره‌های کاهش سرمایه نشان می‌دهد. ما در نظر داریم که این مطالعه با استفاده از مدل SVM برای پیش‌انتخاب یک سبد سرمایه‌گذاری بر اساس شاخص‌های مالی بنیادی قبل از بهینه‌سازی نسبت شارپ، مشارکت‌های تجربی را در حوزه مالی ارائه می‌کند. بنابر این پیشنهادها به صورت زیر ارائه می گردد:

-انتخاب سبد های بهینه سهام با استفاده از الگوی ماشین بردار پشتیبان می تواند بازده بیشتر سبد سهام نسبت به الگوهای سنتی را تضمین کند براین اساس پیشنهاد می شود برنامه نویسان و کارگزاری های بورس اوراق بهادار از این تکنیک در انتخاب سبد های سرمایه گذاری استفاده کنند.

-پیشنهاد می شود در ارزیابی ریسک برای سبد سهام از بردارهای پشتیبان ماشین یادگیری با هسته های یادگیری حداقل سازی ریسک انتخاب سبد سهام توسط سرمایه گذاران عمده صورت گیرد.

-از آنجایی که الگوهای ماشین بردار پشتیبان می تواند تحلیل های داده کاوی حجیم را جهت انتخاب سبد سهام بهینه پوشش دهد، پیشنهاد می شود برنامه نویسان سازمانی این الگوریتم را جایگزین الگوهای قدیمی تر بهینه یابی سبد سهام مثل انواع الگوهای CAPM کنند.

منابع:

خنده خوش، مژگان؛ حقیقی نیت،رضا(۱۳۹۵)، در نظر گرفتن عوامل مؤثر در پیش‌بینی شاخص قیمت بورس تهران با بهبود الگوریتم بهینه‌سازی ملخ در انتخاب بهترین نمونه‌ها در مدل آموزش چندتایی شبکه عصبی، سومین کنفرانس بین المللی مدیریت، اقتصاد و حسابداری.

فرشیدورد، آیدا و هوشمندخلیق، فرناز و میرحسنی، سیدعلی،۱۴۰۰،رویکرد کارا مبتنی بر دسته بندی SVM و مدل تعادل ریسک برای تشکیل سبد سهام، چهاردهمین کنفرانس بین المللی انجمن ایرانی تحقیق در عملیات،مشهد،https://civilica.com/doc/1365965

میرعلوی، سیدحسین؛ پرزمانی، زهرا. مدلی جهت پیش بینی قیمت سهام با استفاده از روش های فرا ابتکاری و شبکه های عصبی، فصلنامه مهندسی مالی و مدیریت اوراق بهادار، ۱۳۹۸، شماره ۴۰.

یزدانی خداشهری, محمدباقر, نسل موسوی, سید حسین & ,حسینی شیروانی, میر سعید. (۱۴۰۲). انتخاب بهینه سبدسهام با استفاده از الگوریتم های یادگیری ماشین .دانش سرمایه‌گذاری.538-511 ,(48)12 ,

Avery, C.N.; Chevalier, J.A.; Zeckhauser, R.J. The CAPS prediction system and stock market returns. Rev. Financ. 2016, 20, 1363–1381.

Basak, S.; Kar, S.; Saha, S.; Khaidem, L.; Dey, S.R. Predicting the direction of stock market prices using tree-based classifiers. N. Am. J. Econ. Financ. 2019, 47, 552–567.

Brownlees C, Gallo G. Financial econometric analysis at ultra-high frequency: data handling concerns. Comput Stat Data Anal. 2006;51(4):2232e2245.

Chandar, S.K. Stock market prediction using subtractive clustering for a neuro fuzzy hybrid approach. Clust. Comput. 2017, 22, 13159–13166

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

Di Persio, L.; Honchar, O. Recurrent neural networks approach to the financial forecast of Google assets. Int. J. Math. Comput. Simul. 2017, 11, 7–13

Durap, A. A comparative analysis of machine learning algorithms for predicting wave runup. Anthropocene Coasts 6, 17 (2023). https://doi.org/10.1007/s44218-023-00033-7

Ganser, A.; Hollaus, B.; Stabinger, S. Classification of Tennis Shots with a Neural Network Approach. Sensors 2021, 21, 5703.

Hota, H.S.; Handa, R.; Shrivas, A.K. Time Series Data Prediction Using Sliding Window Based RBF Neural Network. Int. J. Comput. Intell. Res. 2017, 13, 1145–1156.

Hushani, P. Using Autoregressive Modelling and Machine Learning for Stock Market Prediction and Trading. In Third International Congress on Information and Communication Technology; Springer: Singapore, 2018; pp. 767–774.

Jalota, Hemant & Mandal, Pawan & Thakur, Manoj & Mittal, Garima. (2022). A novel approach to incorporate investor's preference in fuzzy multi-objective portfolio selection problem using credibility measure. Expert Systems with Applications. 212. 118583. 10.1016/j.eswa.2022.118583.

Khashei, M.; Hajirahimi, Z. Performance evaluation of series and parallel strategies for financial time series forecasting. Financ. Innov. 2017, 3, 24.

Melo, Maisa & Cardoso, Rodrigo & Jesus, Tales. (2022). Multiobjective Model Predictive Control for portfolio optimization with cardinality constraint. Expert Systems with Applications. 205. 117639. 10.1016/j.eswa.2022.117639.

Nguyen, D.H.D.; Tran, L.P.; Nguyen, V. Predicting Stock Prices Using Dynamic LSTM Models. Int. Conf. Appl. Inform. 2019, 6, 199–212.

Obthong, Mehtabhorn, Nongnuch Tantisantiwong, Watthanasak Jeamwatthanachai, and Gary Wills. 2020. A Survey on Machine Learning for Stock Price Prediction: Algorithms and Techniques. Paper presented at 2nd International Conference on Finance, Economics, Management and IT Business, Prague, Czech Republic, May 5–6.

Pagliaro, Antonio. 2023. "Forecasting Significant Stock Market Price Changes Using Machine Learning: Extra Trees Classifier

Leads" Electronics 12, no. 21: 4551. https://doi.org/10.3390/electronics12214551

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 Syst. Appl. 115, 635–655.

Rajab, S.; Sharma, V. An interpretable neuro-fuzzy approach to stock price forecasting. Soft Comput. 2019, 23, 921–936.

Rouf, Nusrat, Majid Bashir Malik, Tasleem Arif, Sparsh Sharma, Saurabh Singh, Satyabrata Aich, and Hee-Cheol Kim. 2021. "Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions" Electronics 10, no. 21: 2717.

Sharma, S.; Ahmed, S.; Naseem, M.; Alnumay, W.S.; Singh, S.; Cho, G.H. A Survey on Applications of Artificial Intelligence for Pre-Parametric Project Cost and Soil Shear-Strength Estimation in Construction and Geotechnical Engineering. Sensors 2021, 21, 463.

Strader, T.J.; Rozycki, J.J.; Root, T.H.; Huang, Y.H.J. Machine Learning Stock Market Prediction Studies: Review and Research Directions. J. Int. Technol. Inf. Manag. 2020, 28, 63–83.

Wang, Xianhe & Ouyang, Yuliang & Li, You & Liu, Shu & Teng, Long & Wang, Bo. (2023). Multi-objective portfolio selection considering expected and total utility. Finance Research Letters. 58. 104552. 10.1016/j.frl.2023.104552.

Wang, Y., Zhang, H., Zhang, G., 2019. cPSO-CNN: An efficient PSO-based algorithm for fine-tuning hyper-parameters of convolutional neural networks. Swarm Evol. Comput. 49, 114–123

Wu, H.; Liu, Y.; Wang, J. Review of Text Classification Methods on Deep Learning. Comput. Mater. Contin. 2020, 63, 1309–1321.

Zhang, J.; Teng, Y.-F.; Chen, W. Support vector regression with modified firefly algorithm for stock price forecasting. Appl. Intell. 2018, 49, 1658–1674.