مجله محاسبات نرم، جلد ۱۱، شماره ۲ (پاییز و زمستان ۱۴۰۲)، صفحه ۱۳۰ـ۱۴۵
https://doi.org/10.22052/scj.2023.252793.1144

مجله محاسبات نرم
SOFT COMPUTING JOURNAL
تارنمای مجله: scj.kashanu.ac.ir

دانشگاه کاشان
University of Kashan

استراتژی خرید و فروش در بازار بورس ایران با استفاده از مدل‌های یادگیری ماشین همراه با انتخاب ویژگی به کمک الگوریتم جستجوی فاخته✧

عرفان صابری[۱]، دانشجوی کارشناسی ارشد، الناز رادمند[۱]، دانشجوی کارشناسی ارشد، جمشید پیرگزی[۱*]، استادیار، علی کرمانی[۱]، استادیار
[۱] دانشکده مهندسی برق و کامپیوتر، دانشگاه علم و فناوری مازندران، بهشهر، ایران.

اطلاعات مقاله

تاریخچه مقاله:
دریافت ۲ اردیبهشت ماه ۱۴۰۲
پذیرش ۴ مرداد ماه ۱۴۰۲

کلمات کلیدی:
بورس ایران
یادگیری ماشین
طبقه‌بندی
انتخاب ویژگی
الگوریتم تکاملی

چکیده

در کشور ایران، بازار بورس با شرایط متفاوتی نسبت به بازارهای بورس جهان روبه‌رو است. یکی از مهمترین چالش‌های این بازار، عدم شفافیت در اطلاعات بازار و اطلاعات شرکت‌های معامله‌کننده است. همچنین، عدم وجود داده‌های تاریخی مناسب و کامل برای استفاده در الگوریتم‌های پیش‌بینی، از دیگر چالش‌های مهم است. در پیش‌بینی قیمت سهام، با توجه به تعاملات پویای بورس و تغییر قیمت‌ها در بازه‌های زمانی کوتاه، استفاده از هوش مصنوعی به عنوان یک ابزار قدرتمند در پیش‌بینی قیمت‌ها و تصمیم‌گیری‌های مربوط به خرید و فروش سهام مورد استفاده قرار می‌گیرد. در این مقاله، روشی مبتنی بر یادگیری ماشین شامل ۵ مرحله برچسب گذاری داده‌ها، استخراج ویژگی، انتخاب ویژگی، طبقه‌بندی و ارائه سیگنال ارائه شده است. برای این منظور ابتدا استخراج ویژگی‌های مختلف تکنیکالی از داده‌های قیمتی صورت گرفته است و با استفاده از روش برچسب‌گذاری آستانه‌ای، داده‌ها برچسب‌گذاری شده‌اند. سپس مدل‌های یادگیری ماشین مختلف بر روی این داده‌ها آموزش دیده و در خروجی سیگنال خرید و فروش را ارائه می‌دهند. برای بهبود عملکرد مدل یادگیری ماشین، انتخاب ویژگی با استفاده از الگوریتم جستجوی فاخته انجام شده است. به منظور ارزیابی روش پیشنهادی از داده های چند ساله بورس ایران و از شاخص‌های مختلف استفاده شده است. نتایج حاصل از ارزیابی نشان‌دهنده کارایی روش پیشنهادی می‌باشد.

© ۱۴۰۲ نویسندگان. مقاله با دسترسی آزاد تحت مجوز CC-BY

۱. مقدمه

بورس ایران به عنوان یکی از مهم‌ترین بازارهای سهام در خاورمیانه، همواره برای سرمایه‌گذاران جذاب بوده است. این بازار دارای روندهای پیچیده و چندوجهی بوده و پیش‌بینی قیمت سهام در آن به دلیل تاثیرگذاری عوامل مختلف از جمله سیاسی، اقتصادی و حتی آب و هوا بسیار دشوار است.

در سال‌های اخیر، با ورود هوش مصنوعی به دنیای سرمایه‌گذاری، رویکردهای جدیدی برای پیش‌بینی قیمت سهام و خرید و فروش به کار گرفته شده‌اند. از جمله روش‌هایی که با استفاده از هوش مصنوعی در بازار سهام به کار گرفته می‌شوند، شبکه‌های عصبی [۱]- [۵]، الگوریتم‌های یادگیری ماشین [۶]-

✧ نوع مقاله: پژوهشی

* نویسنده مسئول
پست(های) الکترونیک: erfan.saberi97@yahoo.com (صابری)
elnazradmand1377@yahoo.com (رادمند)
j.pirgazi@mazust.ac.ir (پیرگزی)
a_kermani@mazust.ac.ir (کرمانی)

نحوه ارجاع به مقاله: عرفان صابری، الناز رادمند، جمشید پیرگزی، علی کرمانی، «استراتژی خرید و فروش در بازار بورس ایران با استفاده از مدل‌های یادگیری ماشین همراه با انتخاب ویژگی به کمک الگوریتم جستجوی فاخته»، مجله محاسبات نرم، جلد ۱۱، شماره ۲، صفحات ۱۳۰ـ۱۴۵، پاییز و زمستان ۱۴۰۲.

[۱۰] و مدل‌های مبتنی بر درخت تصمیم [۱۱]، [۱۲] هستند.

استراتژی خرید و فروش در بازار بورس ایران نیز بسته به تحلیل تکنیکال و بنیادی سهام متفاوت است. در تحلیل تکنیکال، از داده‌های گذشته قیمت سهام برای پیش‌بینی آینده استفاده می‌شود. برای این کار، از نرم‌افزارهای تحلیل فنی مانند متاتریدر استفاده می‌گردد. در تحلیل بنیادی نیز، از عوامل اقتصادی، سیاسی و اخبار شرکت‌ها برای پیش‌بینی قیمت سهام استفاده می‌شود.

به طور کلی، استفاده از هوش مصنوعی در بازار بورس ایران می‌تواند به عنوان یک راه‌حل نوین در پیش‌بینی قیمت سهام و ارائه استراتژی‌های بهینه خرید و فروش در بازار سرمایه باشد. با توجه به پیشرفت روزافزون فناوری‌ها و از آنجایی که بازار بورس ایران به عنوان یکی از بزرگترین بازارهای مالی در خاورمیانه شناخته می‌شود، پیش‌بینی قیمت سهام از اهمیت بسیاری برخوردار است. در بین روش‌های مبتنی بر هوش مصنوعی اخیر، استفاده از الگوریتم‌های یادگیری عمیق به عنوان یکی از روش‌های مفید در پیش‌بینی قیمت سهام در بازار بورس ایران مطرح شده است. این الگوریتم‌ها، با استفاده از داده‌های تاریخی سهام و با تحلیل الگوهای قیمتی و نمودارها، می‌توانند بهترین استراتژی خرید و فروش سهام را پیدا کنند. یکی دیگر از روش‌های استفاده از هوش مصنوعی در بازار بورس، استفاده از ربات‌های معامله‌گر خودکار است. ربات‌های معامله‌گر خودکار، به دلیل دقت بالا، سرعت بیشتر و قابلیت هماهنگی با داده‌های بازار، از مزایای بسیاری برخوردار هستند. با استفاده از این ربات‌ها، می‌توان به صورت خودکار تصمیمات خرید و فروش سهام را بر اساس داده‌های تاریخی و الگوریتم‌های هوش مصنوعی انجام داد. در این راستا، روش‌های مختلفی برای پیش‌بینی قیمت سهام توسعه داده شده است. این روش‌ها از روش‌های ساده مبتنی بر آمار و الگوریتم‌های پیش‌بینی ساده گرفته تا روش‌های پیچیده مانند شبکه‌های عصبی و الگوریتم‌های یادگیری ماشین را شامل می‌شوند. برای مثال، یکی از روش‌های ساده پیش‌بینی، استفاده از مدل‌های خطی [۱۳]- [۱۷] است که بر اساس تحلیل تاریخی و شاخص‌های اقتصادی، اقدام به پیش‌بینی قیمت سهام در آینده می‌کنند. در حالی که پیش‌بینی دقیق قیمت سهام امری بسیار پیچیده و چالش برانگیز است، استفاده از الگوریتم‌های هوش مصنوعی می‌تواند در این زمینه موثر باشد.

روشی که در این مقاله مورد بررسی قرار گرفته است، یک روش بهینه برای پیش‌بینی قیمت بازار بورس مبتنی بر روش‌های یادگیری ماشین و انتخاب ویژگی است. کاربرد انتخاب ویژگی با الگوریتم جستجوی فاخته در بازار سهام ایران نشان‌دهنده رویکردی نوآورانه برای مقابله با چالش‌های منحصر به فرد این بازار یعنی انتخاب ویژگی‌های مفید و تاثیرگذار است. عدم شفافیت بازار سهام ایران و داده‌های تاریخی ناقص، توسعه الگوریتم‌های پیش‌بینی دقیق را دشوار می‌کند. انتخاب ویژگی با الگوریتم جستجوی فاخته، با شناسایی مرتبط‌ترین ویژگی‌ها برای پیش‌بینی قیمت سهام، با این چالش‌ها مقابله می‌کند. این نوآوری برای بازار ایران مهم است زیرا دقت مدل‌های پیش‌بینی را افزایش می‌دهد و به تصمیم‌گیری مربوط به خرید و فروش سهام کمک می‌کند.

ساختار مقاله در ادامه به این صورت است: در بخش دوم کارهای مرتبط در حوزه پیش‌بینی قیمت سهام مبتنی بر یادگیری ماشین در بازارهای مالی آورده شده است. در بخش سوم روش پیشنهادی مقاله بیان شده است که شامل ۵ بخش برچسب‌گذاری داده‌ها، استخراج ویژگی، انتخاب ویژگی و پیش‌پردازش داده‌ها و آموزش مدل می‌باشد. در بخش چهارم معیارهای ارزیابی و در بخش پنجم تحلیل نتایج آورده شده است و در نهایت در بخش ششم نتیجه‌گیری بیان شده است.

۲. کارهای مرتبط

تاکنون روش‌های زیادی برای پیش‌بینی قیمت سهام مبتنی بر هوش مصنوعی ارائه شده است. بعضی از این روش‌ها در زمینه استخراج ویژگی و انتخاب ویژگی مطالعه کرده‌اند، بعضی دیگر نیز به توسعه مدل‌های یادگیری ماشین کلاسیک یا مدل‌های یادگیری عمیق توجه کرده‌اند.

Wasiat Khan و همکاران در [۱۸] سیستمی را برای پیش‌بینی

بازار سهام پیشنهاد می‌کنند که از طبقه‌بندی‌کننده‌های یادگیری ماشین، رسانه‌های اجتماعی و داده‌های خبری استفاده می‌کند. این سیستم از چندین مرحله شامل جمع‌آوری داده‌ها، پیش‌پردازش، استخراج ویژگی و طبقه‌بندی تشکیل شده است. داده‌های جمع‌آوری شده از رسانه‌های اجتماعی و منابع خبری برای حذف اطلاعات نامربوط پیش‌پردازش شده و از رویکردهای استخراج ویژگی برای استخراج مرتبط‌ترین ویژگی‌ها استفاده می‌شود. نویسندگان از چندین طبقه‌بندی یادگیری ماشین، ازجمله درخت تصمیم[1]، جنگل تصادفی[2] و ماشین بردار پشتیبان[3] برای پیش‌بینی قیمت‌های بازار سهام بر اساس ویژگی‌های استخراج شده استفاده کردند. هدف اصلی این پژوهش بررسی تاثیر داده‌های موجود در رسانه‌های اجتماعی و اخبار مالی بر دقت پیش‌بینی بازار سهام[4] برای ده روز آینده است. برای بهبود عملکرد و دقت پیش‌بینی، انتخاب ویژگی و کاهش توییت‌های هرزنامه در مجموعه داده‌ها انجام می‌شود. همچنین، آزمایش‌هایی برای تحلیل بازارهای سهام که پیش‌بینی آنها دشوار است و بازارهایی که بیشتر تحت تاثیر رسانه‌های اجتماعی و اخبار مالی قرار دارند، انجام شده است. نتایج تجربی نشان می‌دهد که با استفاده از رسانه‌های اجتماعی، دقت پیش‌بینی بالای ۵۳/۸۰٪ حاصل می‌شود و با استفاده از اخبار مالی نیز دقت پیش‌بینی به میزان ۷۵/۱۶٪ می‌رسد. در این مقاله، برای دستیابی به حداکثر دقت پیش‌بینی، از یادگیری عمیق و ترکیب طبقه‌بندها استفاده شده است. طبقه‌بندی‌کننده جنگل تصادفی نیز با دقت بالای ۸۳/۲۲٪ بهترین عملکرد را دارد.

در مرجع [۱۹]، روشی برای پیش‌بینی قیمت سهام با استفاده از یادگیری ماشین و مدل‌های مبتنی بر یادگیری عمیق ارائه شده است. محققان بر روی شاخص NIFTY 50 بورس ملی هند در بازه زمانی ۲۹ دسامبر ۲۰۱۴ تا ۳۱ ژوئیه ۲۰۲۰ کار کرده‌اند. با استفاده از هشت مدل رگرسیون، قیمت بازشدن NIFTY 50 را برای دوره ۳۱ دسامبر ۲۰۱۸ تا ۳۱ جولای ۲۰۲۰ پیش‌بینی کرده‌اند. سپس با استفاده از چهار مدل رگرسیون مبتنی بر یادگیری عمیق شبکه‌های حافظه بلندمدت و کوتاه‌مدت[5]، قدرت پیش‌بینی را افزایش داده‌اند. نتایج نشان می‌دهد که مدل تک متغیره مبتنی بر LSTM و با استفاده از داده‌های یک هفته قبل برای پیش‌بینی قیمت هفته آینده داده‌های سری زمانی NIFTY 50 دقیق‌ترین مدل است. عملکرد هر مدل بر اساس میانگین مربعات خطا، جذر میانگین مربعات خطا، میانگین خطای مطلق و R-squared ارزیابی شده است. نتایج نشان می‌دهد که مدل‌های یادگیری عمیق از نظر دقت، بهتر از الگوریتم‌های یادگیری ماشین کلاسیک عمل می‌کنند و LSTM دوطرفه بهترین مدل در بین همه مدل‌ها است.

در مرجع [۲۰]، استفاده از رویکردهای هوش مصنوعی، به‌ویژه شبکه عصبی مصنوعی[6] و جنگل تصادفی، برای پیش‌بینی بازده بازار سهام مورد بحث قرار گرفته است. این مطالعه بر روی پنج شرکت از بخش‌های مختلف تمرکز دارد و از داده‌های مالی مانند قیمت‌های بازشدن، بالاترین و پایین‌ترین قیمت و همچنین قیمت بسته‌شدن برای ایجاد متغیرهای جدید به عنوان ورودی مدل‌ها استفاده می‌کند. عملکرد مدل‌ها با استفاده از شاخص‌های استاندارد، یعنی خطای جذر میانگین مربعات[7] و میانگین درصد مطلق خطا[8]، با مقادیر کم که نشان‌دهنده پیش‌بینی کارآمد قیمت بسته شدن سهام است، ارزیابی می‌شود. مدل ANN با ساختار سه لایه و متغیرهای جدید به عنوان ورودی، از نظر دقت پیش‌بینی بهتر از مدل جنگل تصادفی عمل می‌کند. نتایج حاصل عملکرد بهتر مدل ANN را با مقادیر RMSE (۰/۴۲)، MAPE (۰/۷۷) و میانگین خطای سوگیری[9] (۰/۰۱۳) نشان می‌دهد. همچنین با توجه به نتایج الگوریتم تقویت گرادیان با کمترین خطا و بیشترین دقت نسبت به سایر مدل‌ها عملکرد بهتری دارد.

در مرجع [۲۱]، یک رویکرد جدید ساخت سبد سهام[10] با استفاده از یک مدل ترکیبی مبتنی بر یادگیری ماشینبرای پیش‌بینی قیمت سهام و مدل میانگین واریانس برای انتخاب سبد

[1] Decision Tree (DT)
[2] Random Forest (RF)
[3] Support vector machine (SVM)
[4] Stock market

[5] Long short-term memory (LSTM)
[6] Artificial neural network (ANN)
[7] Root Mean Square Error (RMSE)
[8] Mean absolute percentage error (MAPE)
[9] Mean bias error (MBE)
[10] Portfolio

سهام توسعه داده شده است. روش پیشنهادی شامل دو مرحله پیش‌بینی قیمت سهام و انتخاب سبد سهام است. در مرحله اول، یک مدل ترکیبی که ترکیبی از XGBoost[1] با الگوریتم کرم شب‌تاب بهبود یافته[2] است، برای پیش‌بینی قیمت سهام برای دوره بعدی استفاده می‌شود. الگوریتم کرم شب‌تاب بهبود یافته فراپارامترهای XGBoost را بهینه می‌کند. در مرحله دوم، سهام با پتانسیل بازده بالاتر انتخاب شده و از مدل میانگین واریانس برای انتخاب سبد سهام استفاده می‌شود. این مطالعه از بورس شانگهای استفاده می‌کند و نشان می‌دهد که روش پیشنهادی از نظر بازده و خطر از رویکردها و معیارهای سنتی بهتر عمل می‌کند.

Patel و همکاران در [۲۲] روشی را برای پیش‌بینی حرکات شاخص قیمت سهام با استفاده از روش‌های آماده‌سازی داده‌های قطعی روند[3] و رویکردهای یادگیری ماشین پیشنهاد می‌کنند. رویکرد پیشنهادی شامل سه مرحله اصلی آماده‌سازی داده، استخراج ویژگی و طبقه‌بندی است. در مرحله آماده‌سازی داده‌ها، نویسندگان از روش آماده‌سازی داده‌های قطعی روند برای حذف نویز و شناسایی روند اساسی در داده‌ها، استفاده می‌کنند. در مرحله استخراج ویژگی، ویژگی‌های آماری مختلفی را از داده‌های پیش‌پردازش شده استخراج می‌کنند. در نهایت، در مرحله طبقه‌بندی، نویسندگان از الگوریتم‌های یادگیری ماشین مانند ماشین بردار پشتیبان، شبکه‌های عصبی مصنوعی و جنگل تصادفی برای پیش‌بینی حرکت قیمت سهام استفاده می‌کنند. نویسندگان آزمایش‌هایی را بر روی داده‌های سهام در دنیای واقعی انجام دادند و نتایج امیدوارکننده‌ای را گزارش کردند و اثربخشی روش پیشنهادی خود را در پیش‌بینی حرکت قیمت سهام نشان داده‌اند.

نویسندگان مرجع [۲۳]، به چالش انتخاب ویژگی برای پیش‌بینی سهام در طراحی محصول سرمایه‌گذاری کمی (QI) می‌پردازند. برخلاف مدل‌های موجود که بر تحلیل همبستگی تکیه می‌کنند، این مطالعه تحلیل علی را بر اساس داده‌های مشاهده‌ای برای شناسایی تاثیرات مستقیم بین ویژگی‌های سهام معرفی می‌کند. آزمایش‌های مقایسه‌ای با استفاده از داده‌های ۱۳ ساله از بورس‌های شانگهای، مقایسه الگوریتم انتخاب ویژگی علی[4] با تجزیه و تحلیل مولفه‌های اصلی[5]، درخت‌های تصمیم (DT؛ CART)، و کمترین انقباض مطلق و اپراتور انتخاب[6] انجام شده است. نتایج نشان می‌دهد که CFS، هنگامی که با مدل‌های پایه مختلف ترکیب می‌شود، به دقت بالاتری دست می‌یابد و ۱۸ ویژگی سازگار مهم را شناسایی می‌کند. یافته‌ها نشان می‌دهد که CFS دارای پتانسیل قابل‌توجهی برای افزایش توسعه محصولات QI با بهبود انتخاب ویژگی در پیش‌بینی سهام است.

در مرجع [۲۴]، به چالش پیش‌بینی دقیق قیمت سهام در محیط پویا و نویزدار بازار سهام می‌پردازد. مدل‌های محاسباتی نرم در گرفتن الگوهای بازار سهام موثر بوده‌اند، اما تعیین معماری مدل بهینه همچنان یک مشکل باز است. برای مقابله با این موضوع، این مطالعه از الگوریتم جستجوی هارمونی برای بهینه‌سازی معماری مدل‌های پیش‌بینی مختلف، از جمله شبکه عصبی، شبکه عصبی بازگشتی جردن[7]، ماشین یادگیری افراطی[8]، ماشین یادگیری افراطی بازگشتی، مدل خطی تعمیم‌یافته[9]، درخت رگرسیون[10] و رگرسیون فرآیند گاوسی[11] استفاده می‌کند. نتایج تجربی بینش‌های ارزشمندی را در مورد رفتار بازار سهام در شرایط مختلف ارائه می‌دهد. این مطالعه نه تنها بهترین مدل پیش‌بینی را شناسایی می‌کند، بلکه چگونگی تاثیر شرایط مختلف بر دقت این مدل‌ها را نیز روشن می‌کند. مدل‌های ترکیبی پیشنهادی، کاربرد عملی را برای سرمایه‌گذاران ارائه می‌دهد و آنها را قادر می‌سازد تا تصمیمات سرمایه‌گذاری آگاهانه و همچنین کاهش خطرهای بالقوه را در بازارهای سهام داشته باشند.

علاوه بر این، در مرجع [۲۵]، پیشنهاد شده که استخراج ویژگی

[1] Extreme Gradient Boosting
[2] Improved firefly algorithm (IFA)
[3] Trend deterministic data

[4] Causal feature selection (CFS)
[5] Principal component analysis (PCA)
[6] LASSO
[7] Jordan Recurrent Neural Network
[8] Extreme Learning Machine
[9] Generalized Linear Model
[10] Regression Tree
[11] Gaussian Process Regression

موثر می‌تواند دقت و پیش‌بینی مدل‌های رگرسیونی را افزایش دهد. این مقاله به چالش پیش‌بینی روندهای بازار سهام بر اساس داده‌های معاملات اخیر می‌پردازد. در حالی که مطالعات قبلی نشان داده‌اند که روش‌های یادگیری ماشین می‌توانند وابستگی‌ها را در توالی‌های قیمت بازار سهام شناسایی کنند. ماهیت پر نوسان و غیرثابت بازار، پیش‌بینی دقیق روند را دشوار می‌کند. برای پرداختن به این موضوع، این مقاله روش جدیدی را معرفی می‌کند که سری‌های زمانی مالی نویزدار را با بازسازی آنها با استفاده از الگوهای مکرر ساده می‌کند. علاوه بر این، یک شبکه عصبی کانولوشن برای به تصویر کشیدن ساختار فضایی سری‌های زمانی استفاده می‌شود. نتایج تجربی اثربخشی روش پیشنهادی را نشان می‌دهد و بهبود قابل‌توجه ۴ تا ۷ درصدی در دقت را در مقایسه با روش‌های پردازش سیگنال سنتی و الگوهای معاملاتی فرکانس رویکرد مدل‌سازی[1] استفاده از یادگیری عمیق برای پیش‌بینی روند سهام نشان می‌دهد. این رویکرد یادگیری ویژگی‌ها را افزایش می‌دهد و از روش‌های موجود بهتر عمل می‌کند و سهم ارزشمندی در زمینه پیش‌بینی بازار سهام ارائه می‌کند.

در مرجع [۲۶]، مطالعه‌ای توسط چن و همکاران صورت گرفت که یک روش انتخاب ویژگی مبتنی بر نظریه مجموعه‌های فازی برای استخراج ویژگی‌های مهم برای پیش‌بینی قیمت سهام استفاده پیشنهاد داده‌اند. این مقاله به موضوع چالش برانگیز پیش‌بینی بازار سهام می‌پردازد و اهمیت پیش‌بینی روندهای آتی بر اساس داده های تاریخی را برجسته می‌کند. با توجه به نوسانات مداوم در قیمت سهام، پیش‌بینی دقیق آنها کار دشواری است. سرمایه‌گذاران و محققان با استفاده از رویکردهای مختلف یادگیری ماشین و داده‌کاوی برای پیش‌بینی قیمت دارایی‌های مالی مانند طلا، صندوق‌های سرمایه‌گذاری مشترک، نفت خام، مبادلات ارزی و مواد معدنی علاقه نشان داده‌اند. در زمینه بازارهای هند، شاخص‌های Sensex و Nifty به عنوان شاخص‌های معیار برای پیش‌بینی قیمت بازار عمل می‌کنند. رویکرد پیشنهادی در این مقاله ترکیبی جدید از نظریه مجموعه‌های خشن[2] و ماشین بردار پشتیبان به نام R-SVM است. روش R-SVM از نظر دقت و پیچیدگی پیش‌بینی، از سایر الگوریتم‌ها مانند درخت‌های تصمیم، Naïve Bayes و شبکه‌های عصبی مصنوعی بهتر عمل می‌کند. این رویکرد برای افرادی که به دنبال تصمیم‌گیری سودآور در معاملات پولی با ارائه پیش‌بینی‌های دقیق‌تر بازار سهام هستند، نویدبخش است.

مطالعه دیگری توسط کیم و همکاران در [۲۷] انجام شده است که از یک الگوریتم ژنتیک برای انتخاب مرتبط‌ترین شاخص‌های فنی برای پیش‌بینی قیمت سهام استفاده کرده‌اند. این مقاله به طور خاص بر روی انتخاب ویژگی برای مدل‌های شبکه عصبی عمیق که برای پیش‌بینی جهت قیمت سهام استفاده می‌شود، تمرکز می‌کند. این مطالعه ۱۲۴ شاخص تحلیل تکنیکال را که به طور معمول در ادبیات و وبگاه‌های تجاری به عنوان متغیرهای توضیحی مورد استفاده قرار می‌گیرند، بررسی می‌کند. سه روش انتخاب ویژگی برای کاهش افزونگی در میان شاخص‌های مشابه اعمال می‌شود. با استفاده از داده‌های روزانه از سهام هفت شاخص بازار جهانی در بازه زمانی ۲۰۰۸ تا ۲۰۱۹، شبکه‌های عصبی با پیکربندی‌های لایه پنهان مختلف و نرخ Dropout آزمایش می‌شوند. این تحلیل معیارهای طبقه‌بندی مختلف، سودآوری و هزینه‌های مبادله را برای ارزیابی سود اقتصادی در نظر می‌گیرد. نتایج نشان می‌دهد که الگوریتم‌های انتخاب ویژگی متغیرها را به طور یکنواخت انتخاب نمی‌کنند و دقت پیش‌بینی خارج از نمونه به دو مقدار متمایز همگرا می‌شود.

علاوه بر این، لیو و همکاران در [۲۸] به محدودیت‌های مدل کلاسیک انتخاب سهام چند عاملی خطی[3] با پیشنهاد یک رویکرد یکپارچه‌تر برای پیش‌بینی روند قیمت سهام می‌پردازند. با شناخت ماهیت آشفته، پیچیده و پویای بازار سهام، این مطالعه، الگوریتم‌های مختلف انتخاب ویژگی و روش‌های پیش‌بینی غیرخطی را بررسی می‌کند. این تحقیق از داده‌های ۸ ساله بازار سهام چین برای انتخاب ویژگی‌ها و تنظیم متغیرها برای مدل‌های پیش‌بینی روند قیمت سهام مبتنی بر یادگیری ماشین با استفاده از

[1] Frequency trading patterns modeling approach

[2] Rough Set Theory

[3] Linear multi-factor

اعتبارسنجی متقابل پنجره زمانی[1] استفاده می‌کند. تجزیه و تحلیل نشان می‌دهد که الگوریتم جنگل تصادفی بهترین عملکرد را در صورت استفاده برای انتخاب ویژگی و پیش‌بینی روند قیمت سهام دارد.

۳. روش پیشنهادی

در این مقاله، به پیچیدگی‌های بازار سهام ایران پرداخته شده است، اکوسیستمی منحصر به فرد که چالش‌های متمایزی را در مقایسه با بازارهای سهام جهانی ارائه می‌دهد. شفافیت اطلاعات بازار و داده‌های شرکت‌های بازرگانی یک مانع همیشگی در بازار ایران است. علاوه بر این، کمبود داده‌های تاریخی کامل، مانع مهمی برای توسعه الگوریتم‌های پیش‌بینی دقیق است. با این حال، با درک ماهیت پویای تعاملات بازار سهام و تغییرات سریع در قیمت‌گذاری، در این مقاله از قدرت هوش مصنوعی برای ابداع راه‌حلی قدرتمند برای پیش‌بینی قیمت سهام و تصمیم‌گیری استفاده شده است. در روش پیشنهادی، نه تنها از ویژگی‌های اندیکاتوری استفاده شده است، بلکه از سیگنال‌های خرید و فروش صادر شده توسط اندیکاتورها و همچنین الگوهای شمعی ژاپنی به عنوان ویژگی‌هایی برای بهبود عملکرد و دقت مدل استفاده شده است. با ترکیب سیگنال‌های خرید و فروش تولید شده توسط خود اندیکاتورها به عنوان ویژگی‌های اضافی، عملکرد و دقت مدل افزایش پیدا کرده است. این یک رویکرد موثر است زیرا ویژگی‌های مرتبط و مناسب به مدل ارائه می‌دهد. به جای این که مدل با تعداد بیش از حد ویژگی‌های خام آموزش داده شود، روی ارائه ورودی‌های با کیفیت بالا به مدل تمرکز شده است. این رویکرد اصلاح ویژگی‌ها در مقایسه با سناریوهایی که از ویژگی‌های خام استفاده می‌شود، پیشرفت‌های قابل‌توجهی را به همراه داشته است. روش پیشنهادی سعی می‌کند که ویژگی‌های مورد استفاده در مدل، بهینه باشند و در نتیجه عملکرد و دقت افزایش یابد.

در این مقاله اطلاعات قیمت ۵ سهم مختلف بورس ایران (خودرو، وخارزم، وتجارت، شپنا و حکشتی) با استفاده از کتابخانه Finpy در پایتون گرفته شده است. پس از استخراج دیتاست از ۴ ویژگی قیمت باز شدن، بالاترین قیمت، پایین‌ترین قیمت و قیمت بسته شدن که به صورت تعدیل شده در یک روز کاری می‌باشند، برای به دست آوردن ویژگی‌های جدید و برچسب‌گذاری داده‌ها استفاده شده است.

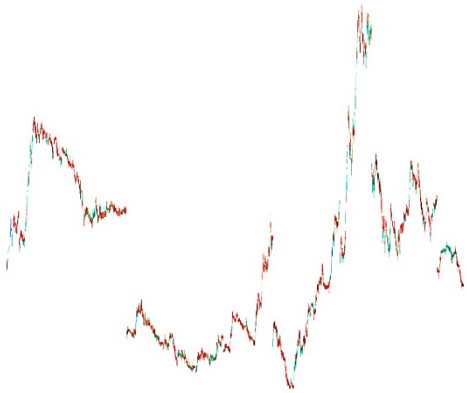
یکی از جنبه‌های مهم در بازار سهام ایران در تمایز بین نمودارهای تعدیل‌شده و تعدیل‌نشده است. نمودار تعدیل‌شده (شکل (۱)) شامل تنظیمات مختلفی مانند تعدیل سود، تعدیل سهام جدید و تعدیل شاخص است. این تعدیل‌ها برای محاسبه تغییرات موجودی سرمایه، تقسیم سهام جدید و سود سهام محاسبه می‌شوند. در مقابل، نمودار تعدیل‌نشده (شکل (۲)) این تعدیل‌ها را در نظر نمی‌گیرد و نشان‌دهنده قیمت اصلی سهام و نقدی است. در حالی که نمودار تعدیل‌نشده تغییرات روزانه قیمت سهام را از طریق نوسانات نمودار منعکس می‌کند، نمودار تعدیل‌شده با ترکیب تعدیلات محاسبه شده نمای جامع‌تری را ارائه می‌دهد. نمودار تعدیل‌شده برای بررسی تغییرات بلند مدت قیمت سهام و مقایسه قیمت‌ها در دوره‌های مختلف بسیار ارزشمند است. از سوی دیگر، نمودار تعدیل‌نشده برای بررسی دقیق تغییرات روزانه قیمت سهام ایده‌آل است. با توجه به نامناسب بودن نمودارهای تعدیل‌نشده برای آموزش مدل به دلیل شکاف قیمتی، از داده‌های قیمت تعدیل‌شده برای آموزش مدل استفاده شده است. برای درک بیشتر از شکل‌های (۱) و (۲) استفاده شده تا تفاوت بین نمودار تعدیل‌شده و تعدیل‌نشده نمایش داده شود.

این رویکرد نوآورانه با ترکیب انتخاب ویژگی با الگوریتم جستجوی فاخته، به چالش‌ها و پیچیدگی‌های منحصر به فرد بازار سهام ایران می‌پردازد. روش ارائه شده در این مقاله نه تنها دقت مدل‌های پیش‌بینی قیمت سهام را افزایش می‌دهد، بلکه به شرکت‌کنندگان بازار بینش قابل اعتمادتری برای تصمیم‌گیری موثر در معاملات سهام می‌دهد. استفاده جامع از داده‌های قیمت تعدیل‌شده و انتخاب دقیق ویژگی‌ها به قابلیت اطمینان و استحکام روش کمک می‌کند و آن را قادر می‌سازد بر محدودیت‌های ناشی از داده‌های تاریخی ناقص و عدم شفافیت غلبه کند. در بازاری

[1] Time-sliding window cross-validation

که با شرایط متمایز آن مشخص می‌شود، این روش به عنوان یک چراغ امید ظاهر می‌شود و شکاف بین تحلیل سنتی بازار سهام و رویکردهای هوش مصنوعی پیشرفته را پر می‌کند.

شکل (۱): نمودار سهم فارس در بازه روزانه در حالت تعدیل عملکردی



شکل (۲): نمودار سهم فارس در بازه روزانه بدون تعدیل قیمتی

۱.۳. روش برچسب‌گذاری آستانه‌ای

در این روش درصد اختلاف قیمت بسته شدن تعدیل‌شده در هر روز با روز آینده محاسبه می‌شود و اگر این درصد اختلاف از یک حد آستانه‌ای بیشتر باشد، برچسب ۲، اگر کمتر از منفی حد آستانه باشد، برچسب ۱ و در غیر این صورت، برچسب ۰ به داده مطابق رابطه‌های زیر تعلق می‌گیرد.

$$\Delta Close = \frac{Close_{t+1} - Close_t}{Close_t} \times 100 \qquad (۱)$$

$$Label_t = \begin{cases} 2 & if\ \Delta Close > T \\ 1 & if\ \Delta Close < -T \\ 0 & otherwise \end{cases} \qquad (۲)$$

در رابطه (۲)، T حد آستانه مورد نظر می‌باشد که در این مقاله ۳/۵ درصد در نظر گرفته شده است. در واقع، در رابطه (۱) اختلاف درصد محاسبه شده است و در رابطه (۲) بر اساس آن برچسب‌گذاری داده‌ها انجام گرفته است که برچسب ۲ به معنای سیگنال خرید، برچسب ۱ به معنای سیگنال فروش و برچسب ۰ به معنای معامله نکردن محسوب می‌شود.

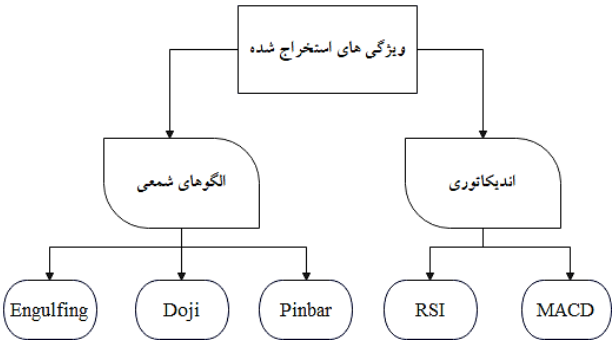
۲.۳. استخراج ویژگی

استخراج ویژگی یک مرحله ضروری در پیش‌بینی قیمت سهام است زیرا کیفیت داده‌های ورودی مدل‌های یادگیری ماشین را تعیین می‌کند. در پیش‌بینی قیمت سهام، شاخص‌های مالی و اقتصادی مانند حجم معاملات، شاخص‌های بازار و اخبار مربوط به شرکت‌ها، متداول‌ترین ویژگی‌های مورد استفاده برای مدل‌سازی هستند.

یکی از روش‌های استخراج ویژگی از داده‌های قیمتی در تحلیل تکنیکال، استفاده از اندیکاتورهای معروفی همچون شاخص قدرت نسبی[۱] و همگرایی و واگرایی میانگین متحرک[۲] می‌باشد. شاخص RSI، اندیکاتوری است که برای اندازه‌گیری قدرت یک روند صعودی یا نزولی در بازار استفاده می‌شود. همچنین، MACD شاخصی است که برای تشخیص تغییر در روند قیمتی بازار به کار می‌رود. در این مقاله، با استفاده از داده‌های قیمتی، اندیکاتورهای RSI و MACD به همراه الگوهای شمعی مختلف استخراج شده و برای پیش‌بینی قیمت سهام به کار گرفته شده‌اند. در نهایت، با ارزیابی عملکرد این اندیکاتورها و الگوهای شمعی، به نتیجه‌گیری در مورد بازدهی آنها در پیش‌بینی قیمت سهام پرداخته شده است.

برای ویژگی‌های از نوع الگوهای شمعی از الگوهای معروفی مانند الگوی پوشای صعودی و نزولی[۳] و انواع الگوهای دوجی[۴] و الگوهای پین بار[۵] استفاده شده است. پس از ایجاد این دو نوع ویژگی، این ویژگی‌ها به دیتاست اصلی که حاوی اطلاعات اولیه قیمت بوده، اضافه شده است و ماتریس ویژگی تشکیل شده است. انواع ویژگی‌های استخراج شده در شکل (۳) نمایش داده شده است.

[1] Relative strength index (RSI)
[2] Moving Average Convergence Divergence (MACD)
[3] Bullish and Bearish Engulfing pattern
[4] Doji
[5] Pin Bar



شکل(۳): انواع ویژگی‌های استخراج شده

برای آموزش دادن مدل یادگیری ماشین می‌توان از این ویژگی‌ها به تنهایی یا ترکیب‌های مختلفی از آنها استفاده شود. برخی از ویژگی‌های تولید شده در ادامه تعریف شده است.

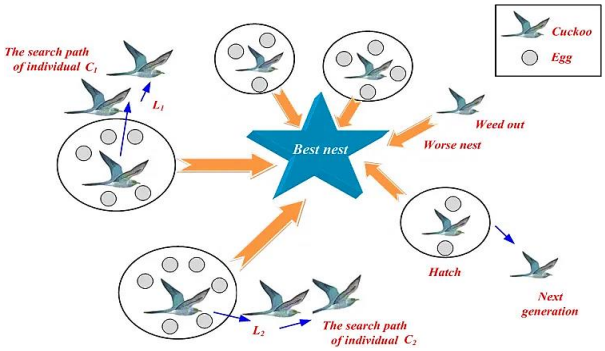
- سیگنال خرید MACD: زمانی که فاز MACD منفی باشد و خط MACD از ناحیه زیر صفر سیگنال را از پایین به بالا قطع کند، سیگنال خرید صادر می‌شود.
- سیگنال فروش MACD: زمانی که خط MACD از ناحیه بالای صفر خط سیگنال را از بالا به پایین قطع کند سیگنال فروش صادر می‌شود و فاز MACD در حالت مثبت باشد.
- سیگنال خرید RSI: هنگامی که RSI در ناحیه اشباع فروش یا ناحیه بیش‌فروش باشد و سطح ۳۰ را از پایین به بالا قطع کند، سیگنال خرید صادر می‌شود.
- سیگنال فروش RSI: هنگامی که RSI در ناحیه بیش خرید یا اشباع خرید باشد و سطح ۷۰ را از بالا به پایین قطع کند، سیگنال فروش صادر می‌شود.

۳.۳. انتخاب ویژگی

انتخاب ویژگی [۲۹]، [۳۰] یک گام مهم در روش‌های یادگیری ماشین [۳۱]، از جمله پیش‌بینی بازار سهام، برای بهبود دقت مدل و کاهش زمان محاسبات است. الگوریتم CS[1] یک الگوریتم بهینه‌سازی مبتنی بر جمعیت است که از رفتار پرندگان فاخته الهام گرفته شده است. این الگوریتم به دلیل کارایی و تاثیر آن در مدیریت مجموعه داده‌های بزرگ به طور گسترده در انتخاب ویژگی استفاده می‌شود [۳۲].

فرآیند انتخاب ویژگی با الگوریتم CS را می‌توان به چند مرحله تقسیم کرد [۳۳]. اولین گام رمزگذاری مساله انتخاب ویژگی به عنوان یک مساله بهینه‌سازی است، که در آن تابع هدف، حداکثر کردن دقت پیش‌بینی مدل و در عین حال به حداقل رساندن تعداد ویژگی‌های انتخاب شده است. سپس، جمعیتی از راه‌حل‌های کاندید، که به صورت رشته‌های باینری هستند و نشان‌دهنده وجود یا عدم وجود هر ویژگی می‌باشند، به طور تصادفی مقداردهی اولیه می‌شوند.

در هر تکرار، با استفاده از الگوریتم CS، یک راه‌حل کاندید جدید تولید می‌شود. این راه‌حل شامل ایجاد یک لانه فاخته جدید است، جایی که یک پرنده فاخته تخم می‌گذارد، که یا نسخه اصلاح شده یکی از لانه‌های موجود است یا یک راه‌حل کاملا جدید می‌باشد. در شکل (۴) شمای کلی این الگوریتم نمایش داده شده است. انتخاب لانه موجود و فرآیند اصلاح بر اساس احتمال کشف و احتمال پذیرش است که توسط الگوریتم CS تعریف شده است.



شکل (۴): انتخاب ویژگی با استفاده از الگوریتم فاخته [۷]

پس از تولید راه‌حل کاندید جدید، تابع هدف برای اندازه‌گیری برازش آن ارزیابی می‌شود. راه‌حل کاندید با بالاترین مقدار تابع برازش به عنوان بهترین راه‌حل و ویژگی‌های آن به عنوان ویژگی‌های انتخابی مدل در نظر گرفته می‌شود. این فرآیند تا زمانی ادامه می‌یابد که یک معیار توقف، مانند رسیدن به حداکثر تعداد تکرارها یا دستیابی به آستانه تابع برازش از پیش تعریف شده، برآورده شود.

در نتیجه، فرآیند انتخاب ویژگی با الگوریتم CS یک ابزار قدرتمند در بهبود عملکرد مدل‌های یادگیری ماشین در پیش‌بینی

[1] Cuckoo search

بازار سهام است. این شامل رمزگذاری مساله انتخاب ویژگی به عنوان یک مساله بهینه‌سازی، تولید جمعیتی از راه‌حل‌های کاندید، استفاده از الگوریتم CS برای ایجاد راه‌حل‌های جدید، ارزیابی برازش هر راه‌حل و انتخاب بهترین راه‌حل با بالاترین مقدار تابع برازش است. متغیرهای الگوریتم فاخته با توجه به جدول (۱) مقداردهی شده است.

جدول ۱. مقادیر متغیرها

متغیر	مقدار
تعداد ویژگی‌های انتخابی	۳
تعداد تکرار بیشینه	۵۰
اندازه جمعیت اولیه	۱۰

دقت داشته باشید که تعداد کل ویژگی‌های استخراج شده ۲۹ می‌باشد که از میان این تعداد ویژگی‌ها، فقط ۳ ویژگی انتخاب شده است که شامل بهترین ویژگی‌ها می‌باشد.

۴.۳. پیش‌پردازش داده‌ها و آموزش مدل

بعد از این که برچسب‌گذاری و استخراج ویژگی انجام شده است و داده‌ها به دو دسته داده‌های آزمون و آموزش تقسیم شده‌اند، به نحوی که ۳۰ درصد از آنها به طور تصادفی به عنوان داده آزمون و بقیه به عنوان داده آموزشی انتخاب شده‌اند. برای نرمال‌سازی ویژگی‌ها از روش مقیاس‌کننده استاندارد[1] استفاده شده است. در یادگیری ماشین از مقیاس‌کننده استاندارد برای تغییر اندازه توزیع مقادیر استفاده می‌شود به طوری که میانگین مقادیر مشاهده شده ۰ و انحراف استاندارد ۱ باشد. مقیاس‌کننده استاندارد یک روش مهم است که به طور عمده به عنوان یک مرحله پیش‌پردازش قبل از بسیاری از مدل‌های یادگیری ماشین انجام می‌شود تا محدوده عملکرد مجموعه داده ورودی را استاندارد کند.

در این مقاله از روش‌های یادگیری ماشین K-نزدیک‌ترین همسایگی[2]، XGBoost، ماشین بردار پشتیبان، جنگل تصادفی، بیز ساده[3] و پرسپترون چند لایه[4] استفاده شده است و مجموعه داده سهم‌های مختلف بر روی این مدل‌ها آموزش دیده‌اند. شمای کلی مراحل مقاله در شکل (۵) نشان داده شده است.

۴. معیارهای ارزیابی

برای معیار ارزیابی مدل‌های یادگیری ماشین از ماتریس تداخل[5]، صحت[6]، حساسیت[7]، معیار $F1$[8]، دقت[9] و منحنی مشخصه عملکرد[10] سیستم استفاده شده است.

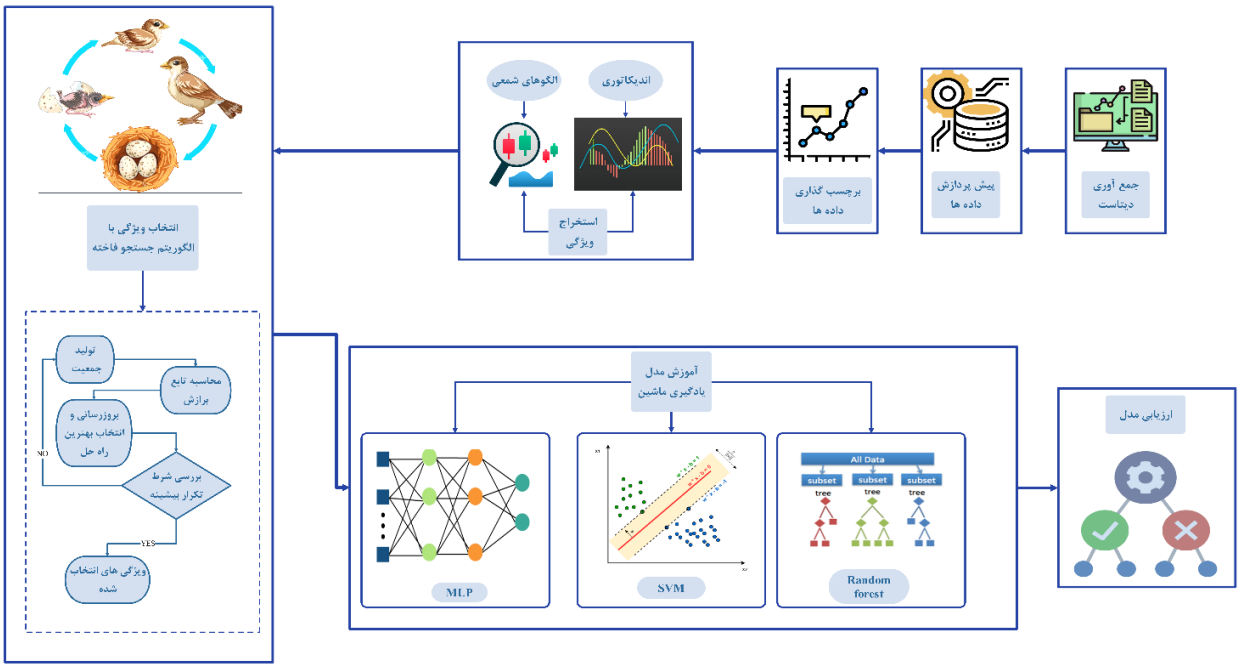
ماتریس تداخل به ما کمک می‌کند تا عملکرد یک مدل یا نحوه پیش‌بینی یک مدل را در یادگیری ماشین نشان دهیم. این ماتریس از ۴ عنصر اصلی تشکیل شده است که معیارهای مختلفی را برای شمارش تعداد پیش‌بینی‌های صحیح و نادرست نشان می‌دهد. اگر برچسب‌های پیش‌بینی شده و واقعی مطابقت داشته باشند، گفته می‌شود که پیش‌بینی صحیح است، اما زمانی که برچسب‌های پیش‌بینی شده و واقعی با هم تطبیق نداشته باشند، آنگاه گفته می‌شود که پیش‌بینی نادرست است. علاوه بر این، مثبت و منفی نشان‌دهنده برچسب‌های پیش‌بینی شده در ماتریس است. شکل (۶) نمای کلی یک ماتریس تداخل را نشان می‌دهد. صحت به عنوان نسبت نمونه‌های مثبت طبقه‌بندی شده به درستی (مثبت واقعی[11]) به تعداد کل نمونه‌های مثبت طبقه‌بندی شده (به درستی یا نادرست) تعریف می‌شود و از رابطه زیر محاسبه می‌شود.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (۳)$$

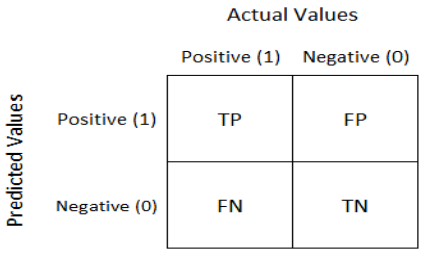
حساسیت به عنوان نسبت بین تعداد نمونه‌های مثبت که به درستی به عنوان مثبت طبقه‌بندی شده‌اند به تعداد کل نمونه‌های مثبت محاسبه می‌شود و طبق رابطه (۴) به دست می‌آید. این معیار توانایی مدل را در تشخیص نمونه‌های مثبت اندازه‌گیری می‌کند.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (۴)$$

[1] StandardScaler
[2] K-Nearest Neighbors
[3] Naive Bayes
[4] Multilayer perceptron
[5] Confusion Matrix
[6] Precision
[7] Recal
[8] F1-score
[9] Accuracy
[10] Receiver Operating Characteristic
[11] True Positive



شکل (۵): شمای کلی مراحل مختلف مدل پیشنهادی



شکل (۶): تصویری از یک ماتریس تداخل

صحت و حساسیت دو عنصر سازنده معیار F1 هستند. هدف معیار F1 ترکیب معیارهای صحت و حساسیت در یک شاخص واحد است. در عین حال، معیار F1 طوری طراحی شده است که روی داده‌های نامتعادل به خوبی کار کند. معیار F1 به عنوان میانگین هارمونیک دقت و حساسیت مطابق رابطه (۵) تعریف می‌شود.

$$F1 = \frac{Precision \times Recall}{Precision + Recall} \times 2 \qquad (۵)$$

منحنی مشخصه عملیاتی سیستم (ROC) یک نمایش گرافیکی از عملکرد یک طبقه‌بندی کننده باینری است زیرا آستانه تشخیص آن متفاوت است. منحنی ROC نرخ مثبت واقعی (TPR)، که به عنوان حساسیت نیز شناخته می‌شود، در برابر نرخ مثبت کاذب (FPR) را ترسیم می‌کند که ۱ منهای نرخ منفی واقعی (TNR) است و به عنوان ویژگی نیز شناخته می‌شود. سطح زیر منحنی ROC (AUC) یک معیار پرکاربرد برای ارزیابی عملکرد طبقه‌بندی‌کننده‌های باینری است. منحنی ROC ابزار مفیدی برای انتخاب آستانه مناسب بر اساس مبادله بین TPR و FPR فراهم می‌کند.

۵. تحلیل نتایج

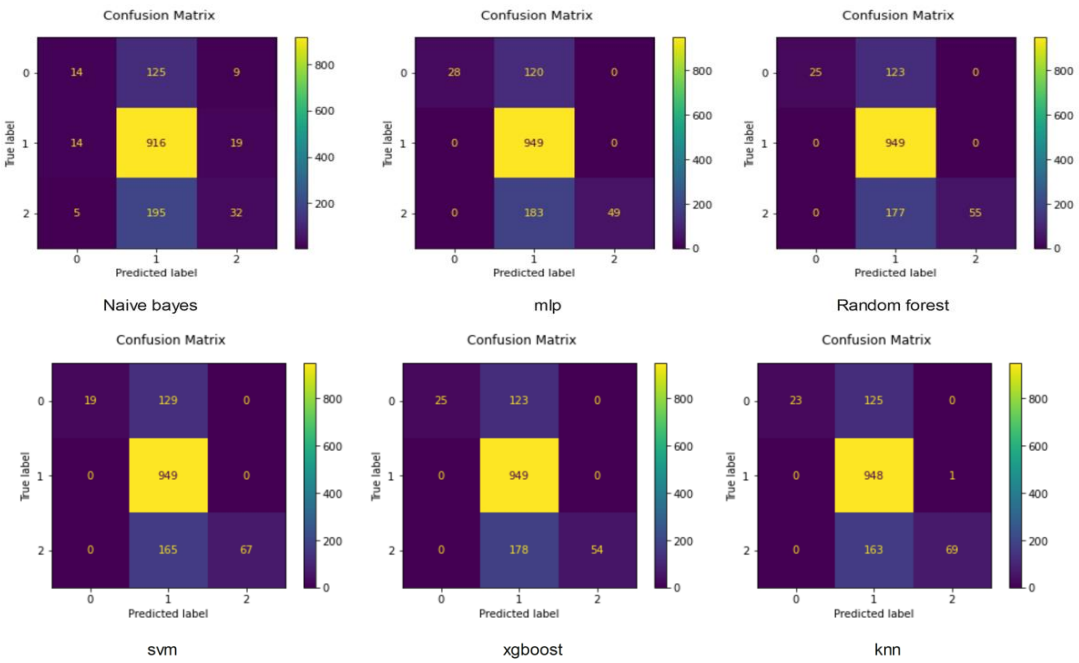
در بخش تحلیل نتایج، نتایج معیارهای ارزیابی مدل‌های یادگیری ماشین بر روی چندین سهم مورد بررسی قرار گرفته است و از ماتریس تداخل و نمودار مشخصه عملکرد سیستم برای نمایش خروجی استفاده شده است. در جداول (۲) و (۳) به بررسی دقت مدل‌های یادگیری ماشین در سهم‌های گوناگون پرداخته شده است. در جدول (۴) دقت تمامی مدل‌ها در سهم‌های گوناگون با یکدیگر مقایسه شده است و در نهایت برای درک بهتر، نمودار مستطیلی مربوط به این جداول ترسیم شده است.

با توجه به جدول (۲) و شکل‌های (۷) و (۸) که به ترتیب ماتریس‌های تداخل و منحنی‌های مشخصه عملکرد سیستم سهم خودرو را نمایش می‌دهد، مدل‌های MLP و KNN تحت معیارهای ارزیابی مختلف، عملکرد بهتری نسبت به بقیه مدل‌ها

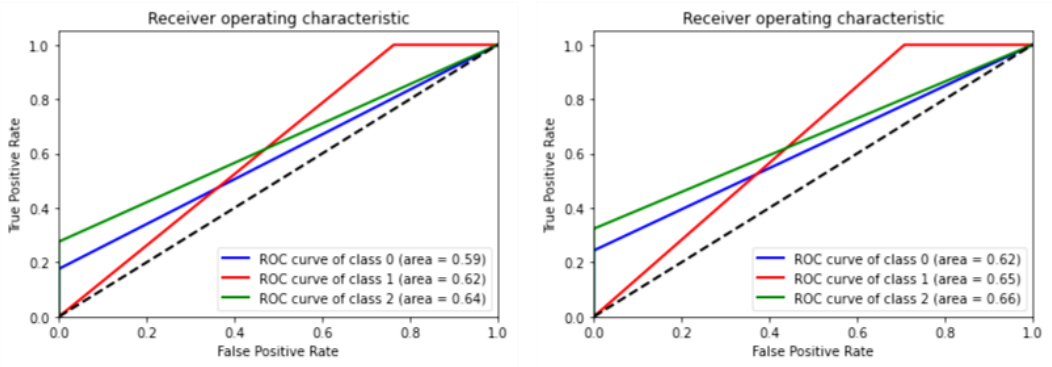
داشته‌اند. اغلب مدل‌های یادگیر در سهم خودرو دقت بالایی داشتند؛ زیرا که در سهم خودرو نسبت به دیگر سهم‌ها داده‌های بیشتری برای آموزش مدل وجود داشت و مدل‌ها توانستند به خوبی آموزش ببینند. تمامی مدل‌های یادگیری ماشین در تمامی سهم‌ها وقتی که هر دو برچسب پیش‌بینی شده و واقعی یک باشند، نتیجه خوبی را نشان می‌دهند.

جدول (۲): نتایج ارزیابی روش‌های یادگیری ماشین در سهم خودرو

مدل یادگیری ماشین	Label	Precision	Recall	F1-score	Accuracy
SVM	۰	۱/۰۰	۰/۱۴	۰/۲۵	۰/۷۸
	۱	۰/۷۷	۱/۰۰	۰/۸۷	
	۲	۰/۹۹	۰/۳۱	۰/۴۷	
XGBoost	۰	۱/۰۰	۰/۱۷	۰/۲۹	۰/۷۷
	۱	۰/۷۶	۱/۰۰	۰/۸۶	
	۲	۱/۰۰	۰/۲۳	۰/۳۸	
KNN	۰	۰/۸۶	۰/۱۷	۰/۲۸	۰/۷۹
	۱	۰/۷۷	۱/۰۰	۰/۸۷	
	۲	۱/۰۰	۰/۳۲	۰/۴۹	
RF	۰	۱/۰۰	۰/۱۷	۰/۲۹	۰/۷۷
	۱	۰/۷۶	۱/۰۰	۰/۸۶	
	۲	۱/۰۰	۰/۲۴	۰/۳۸	
NB	۰	۰/۴۶	۰/۲۳	۰/۳۱	۰/۷۵
	۱	۰/۷۶	۰/۹۵	۰/۸۴	
	۲	۰/۸۶	۰/۲۳	۰/۳۷	
MLP	۰	۱/۰۰	۰/۲۷	۰/۴۳	۰/۸۰
	۱	۰/۷۸	۱/۰۰	۰/۸۸	
	۲	۱/۰۰	۰/۳۱	۰/۴۷	



شکل (۷): ماتریس‌های تداخل مربوط به سهم خودرو برای مدل‌های یادگیری ماشین

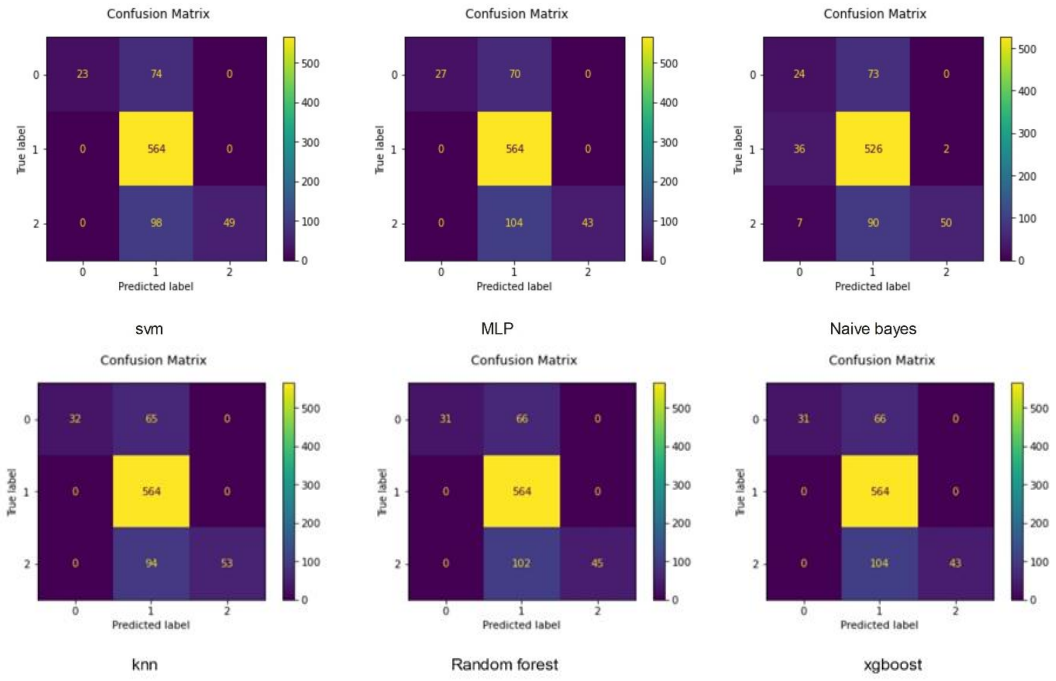


شکل (۸): منحنی مشخصه عملکرد سیستم (ROC) سهم خودرو برای مدل‌های یادگیری ماشین (الف) KNN و (ب) MLP

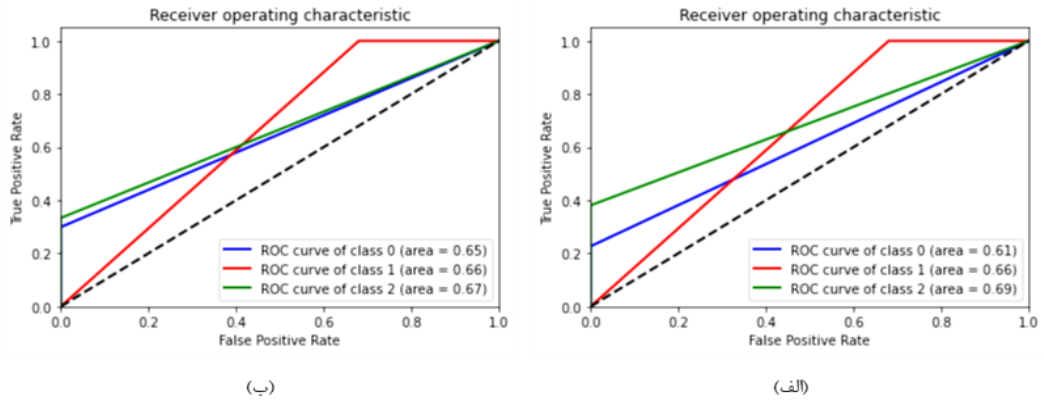
با توجه به جدول (۳) و شکل‌های (۹) و (۱۰) که به ترتیب ماتریس‌های تداخل و منحنی‌های مشخصه عملکرد سیستم سهم وتجارت را نمایش می‌دهد، مدل KNN عملکرد خوبی دارد و دقت مدل برابر ۸۰ درصد بوده و دقت به نسبت بالایی را نسبت به بقیه مدل‌ها دارا می‌باشد. سپس مدل‌های SVM و XGBoost و Random Forest دقت یکسان ۷۹ درصدی داشته‌اند. ماتریس تداخل مربوط به سهم وتجارت در سلول (۱،۱) تعداد موارد مثبت واقعی را نشان می‌دهد که نسبت به بقیه قسمت‌ها مقدار بالایی را داشته است که به این معنی است که مدل نقطه مورد نظر را درست پیش‌بینی کرده است و برچسب مربوط به یک درست پیش‌بینی شده است.

جدول (۳): نتایج ارزیابی روش‌های یادگیری ماشین در سهم وتجارت

مدل یادگیری ماشین	Label	Precision	Recall	F1-score	Accuracy
SVM	۰	۱/۰۰	۰/۲۴	۰/۳۸	۰/۷۹
	۱	۰/۷۷	۱/۰۰	۰/۸۷	
	۲	۱/۰۰	۰/۳۳	۰/۵۰	
XGBoost	۰	۱/۰۰	۰/۳۲	۰/۴۸	۰/۷۹
	۱	۰/۷۷	۱/۰۰	۰/۸۷	
	۲	۱/۰۰	۰/۳۱	۰/۴۷	
KNN	۰	۱/۰۰	۰/۳۳	۰/۵۰	۰/۸۰
	۱	۰/۷۸	۱/۰۰	۰/۸۸	
	۲	۱/۰۰	۰/۳۶	۰/۵۳	
RF	۰	۱/۰۰	۰/۳۲	۰/۴۸	۰/۷۹
	۱	۰/۷۷	۱/۰۰	۰/۸۷	
	۲	۱/۰۰	۰/۲۹	۰/۴۵	
NB	۰	۰/۳۶	۰/۲۵	۰/۲۹	۰/۷۴
	۱	۰/۷۶	۰/۹۳	۰/۸۴	
	۲	۰/۹۶	۰/۳۴	۰/۵۰	
MLP	۰	۱/۰۰	۰/۲۸	۰/۴۴	۰/۷۸
	۱	۰/۷۶	۱/۰۰	۰/۸۷	
	۲	۱/۰۰	۰/۲۹	۰/۴۵	



شکل (۹): ماتریس‌های تداخل مربوط به سهم وتجارت برای مدل‌های یادگیری ماشین

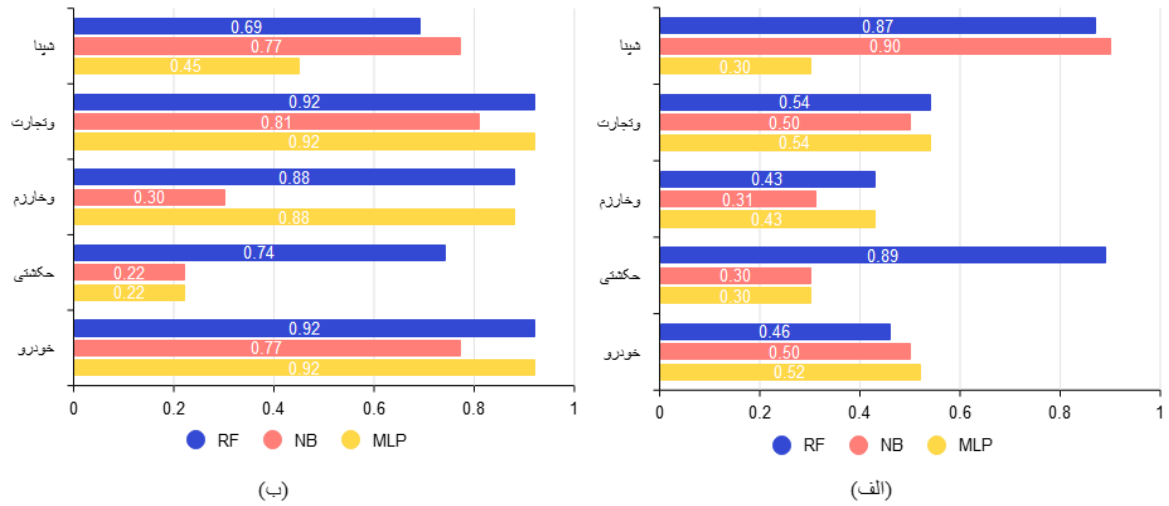


شکل (۱۰): منحنی مشخصه عملکرد سیستم (ROC) سهم وتجارت برای مدل‌های یادگیری ماشین (الف) KNN و (ب) MLP

در جدول (۴) درصد تمامی مدل‌ها در ۵ سهم مختلف مورد مقایسه قرار گرفته است. همان‌طور که مشاهده می‌شود، مدل‌های یادگیری ماشین درصد دقت خوبی در بعضی از سهام داشتند که به ترتیب مدل SVM و XGBoost در سهم وتجارت ۷۹ درصد و مدل KNN در سهم وتجارت درصد ۸۰ و مدل RF در سهم وتجارت درصد ۷۹ و مدل NB در سهم حکشتی ۸۵ درصد و همچنین مدل MLP در سهم خودرو ۸۰ درصد را نشان می‌دهند. در نمودارهای شکل (۱۱) میانگین مقادیر Precision (صحت) و Recall (حساسیت) مدل‌های یادگیری ماشین RF، NB و MLP برای ۵ سهم بورس ایران نمایش داده شده است که با توجه به نمودار (الف) و (ب)، مدل یادگیری ماشین RF در اکثر سهم‌ها توانسته عملکرد بهتری از نظر معیار حساسیت و صحت کسب کند.

جدول (۴): مقایسه دقت مدل‌های یادگیری ماشین در سهم‌های مختلف

مدل	خودرو	حکشتی	وخارزم	وتجارت	شپنا
SVM	۰/۷۸	۰/۶۶	۰/۶۹	۰/۷۹	۰/۶۹
XGBoost	۰/۷۷	۰/۶۳	۰/۶۸	۰/۷۹	۰/۶۴
KNN	۰/۷۹	۰/۶۴	۰/۶۹	۰/۸۰	۰/۶۷
RF	۰/۷۷	۰/۶۶	۰/۶۸	۰/۷۹	۰/۷۳
NB	۰/۷۵	۰/۸۵	۰/۷۸	۰/۷۴	۰/۶۹
MLP	۰/۸۰	۰/۶۶	۰/۷۲	۰/۷۸	۰/۶۹



شکل (۱۱): مقایسه و بررسی میانگین (الف) RECALL و (ب) PRECISION مدل‌های یادگیری ماشین گوناگون در سهم‌های مختلف

۶. نتیجه‌گیری

در پیش‌بینی بازارهای مالی، انتخاب ویژگی به دلیل ابعاد بالای داده‌ها بسیار مهم است و می‌تواند به شناسایی مرتبط‌ترین شاخص‌ها برای پیش‌بینی قیمت سهام کمک کند. در این مقاله، مجموعه‌ای از ویژگی‌های تکنیکال، از جمله اندیکاتورهای RSI، MACD و الگوهای شمعی از داده‌های قیمت سهام استخراج شده است. سپس از الگوریتم جستجوی فاخته برای انتخاب مهم‌ترین ویژگی‌ها برای پیش‌بینی قیمت سهام استفاده شده است. عملکرد ویژگی‌های انتخاب شده با استفاده از مدل‌های مختلف یادگیری ماشین ارزیابی شده است. نتایج تجربی نشان می‌دهد که الگوریتم جستجوی فاخته می‌تواند به طور موثری بهترین ویژگی‌ها را برای پیش‌بینی قیمت سهام انتخاب کند، که منجر به بهبود دقت پیش‌بینی و کاهش زمان محاسبه می‌شود.

تعارض منافع: نویسندگان اعلام می‌کنند که هیچ تعارض منافعی ندارند.

مراجع

[1] D. Selvamuthu, V. Kumar, and A. Mishra, “Indian stock market prediction using artificial neural networks on tick data,” Financ. Innov., vol. 5, p. 16, 2019, doi: 10.1186/s40854-019-0131-7.

[2] J. Qiu, B. Wang, and C. Zhou, “Forecasting Stock Prices with Long-Short Term Memory Neural Network Based on Attention Mechanism,” PLOS ONE, vol. 15, no. 1, 2020, p. e0227222, doi: 10.1371/journal.pone.0227222.

[3] P. Yu and X. Yan, “Stock Price Prediction Based on Deep Neural Networks,” Neural Comput. Appl., vol. 32, no. 6, pp. 1609-1628, 2020, doi: 10.1007/s00521-019-04212-x.

[4] G. Ding and L. Qin. “Study on the Prediction of Stock Price Based on the Associated Network Model of LSTM,” Int. J. Mach. Learn. Cybern., vol. 11, no. 6, pp. 1307-1317, 2020, doi: 10.1007/s13042-019-01041-1.

[5] T. Kim and H.Y. Kim, “Forecasting Stock Prices with a Feature Fusion LSTM-CNN Model Using Different Representations of the Same Data,” PLOS ONE, vol. 14, no. 2, 2019, p. e0212320, doi: 10.1371/journal.pone.0212320.

[6] P. Meesad and R. I. Rasel, “Predicting stock market price using support vector regression,” in Int. Conf. Inf. Electron. Vision (ICIEV), Dhaka, Bangladesh, 2013, pp. 1-6, doi: 10.1109/ICIEV.2013.6572570.

[7] M. Obthong, N. Tantisantiwong, W. Jeamwatthanachai, and G.B. Wills, “A Survey on Machine Learning for Stock Price Prediction: Algorithms and Techniques,” in Proc. 2nd Int. Conf. Fin. Econ. Manag. IT Bus. (FEMIB), Prague, Czech Republic, 2020, pp. 63-71, doi: 10.5220/0009340700630071.

[8] J. Pan, Y. Zhuang, and S. Fong, “The Impact of Data Normalization on Stock Market Prediction: Using SVM and Technical Indicators,” in Proc. 2nd Int. Conf. Soft Comput. Data Sci. (SCDS), Kuala Lumpur, Malaysia,

2016, pp. 72-88, doi: 10.1007/978-981-10-2777-2_7.

[9] P. Ghosh, A. Neufeld, and J.K. Sahoo, "Forecasting Directional Movements of Stock Prices for Intraday Trading Using LSTM and Random Forests," Fin. Res. Letters, vol. 46, p. 102280, 2022, doi: 10.1016/j.frl.2021.102280.

[10] J. Zhang, L. Li, and W. Chen, "Predicting Stock Price Using Two-Stage Machine Learning Techniques," Comput. Econ., vol. 57, pp. 1237–1261, 2021, doi: 10.1007/s10614-020-10013-5.

[11] B.B. Nair, N.M. Dharini, and V.P. Mohandas, "A Stock Market Trend Prediction System Using a Hybrid Decision Tree-Neuro-Fuzzy System," in Int. Conf. Adv. Recent Technol. Commun. Comput., Kottayam, India, 2010, pp. 381-385, doi: 10.1109/ARTCom.2010.75.

[12] S.S. Panigrahi and J.K. Mantri, "Epsilon-SVR and decision tree for stock market forecasting," in Int. Conf. Green Comput. Internet Things (ICGCIoT), Greater Noida, India, 2015, pp. 761-766, doi: 10.1109/ICGCIoT.2015.7380565.

[13] D. Enke, M. Grauer, and N. Mehdiyev, "Stock Market Prediction with Multiple Regression, Fuzzy Type-2 Clustering and Neural Networks," in Proc. Complex Adapt. Syst. Conf., Chicago, Illinois, USA, 2011, pp. 201-206, doi: 10.1016/j.procs.2011.08.038.

[14] Y.E. Cakra and B.D. Trisedya, "Stock price prediction using linear regression based on sentiment analysis," in Int. Conf. Adv. Comput. Sci. Inf. Syst. (ICACSIS), Depok, Indonesia, 2015, pp. 147-154, doi: 10.1109/ICACSIS.2015.7415179.

[15] B. Panwar, G. Dhuriya, P. Johri, S. Singh Yadav, and N. Gaur, "Stock Market Prediction Using Linear Regression and SVM," in Int. Conf. Adv. Comput. Innov. Technol. Eng. (ICACITE), Greater Noida, India, 2021, pp. 629-631, doi: 10.1109/ICACITE51222.2021.9404733.

[16] A. Izzah, Y.A. Sari, R. Widyastuti, and T. A. Cinderatama, "Mobile app for stock prediction using Improved Multiple Linear Regression," in Int. Conf. Sustain. Inf. Eng. Technol. (SIET), Malang, Indonesia, 2017, pp. 150-154, doi: 10.1109/SIET.2017.8304126.

[17] D. Enke and N. Mehdiyev, "Stock Market Prediction Using a Combination of Stepwise Regression Analysis, Differential Evolution-Based Fuzzy Clustering, and a Fuzzy Inference Neural Network," Intell. Autom. Soft Comput., vol. 19, no. 4, pp. 636-48, 2013, doi: 10.1080/10798587.2013.839287.

[18] W. Khan, M.A. Ghazanfar, M.A. Azam, A. Karami, K.H. Alyoubi, and A.S. Alfakeeh, "Stock Market Prediction Using Machine Learning Classifiers and Social Media, News," J. Ambient Intell. Humaniz. Comput., vol. 13, no. 7, pp. 3433-3456, 2022, doi: 10.1007/s12652-020-01839-w.

[19] S. Mehtab, J. Sen, and A. Dutta, "Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models," in Machine Learning and Metaheuristics Algorithms, and Applications. (SoMMA), Commun. Comput. Inf. Sci., vol 1366, Springer, Singapore, 2021, doi: 10.1007/978-981-16-0419-5_8.

[20] M. Vijh, D. Chandola, V.A. Tikkiwal, and A. Kumar, "Stock Closing Price Prediction Using Machine Learning Techniques," Procedia Comput. Sci., vol. 167, pp. 599-606, 2020, doi: 10.1016/j.procs.2020.03.326.

[21] W. Chen, H. Zhang, M.K. Mehlawat, and L. Jia, "Mean–Variance Portfolio Optimization Using Machine Learning-Based Stock Price Prediction," Appl. Soft Comput., vol. 100, p. 106943, 2021, doi: 10.1016/j.asoc.2020.106943.

[22] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting Stock and Stock Price Index Movement Using Trend Deterministic Data Preparation and Machine Learning Techniques," Expert Syst. Appl., vol. 42, no. 1, pp. 259-268, 2015, doi: 10.1016/j.eswa.2014.07.040.

[23] X. Zhang, Y. Hu, K. Xie, S. Wang, E.W.T. Ngai, and M. Liu, "A Causal Feature Selection Algorithm for Stock Prediction Modeling," Neurocomputing, vol. 142, pp. 48-59, 2014, doi: 10.1016/j.neucom.2014.01.057.

[24] M. Gocken, M. Ozcalici, A. Boru, and A.T. Dosdogru, "Stock Price Prediction Using Hybrid Soft Computing Models Incorporating Parameter Tuning and Input Variable Selection," Neural Comput. Appl., vol. 31, no. 2, pp. 577-592, 2019, doi: 10.1007/s00521-017-3089-2.

[25] M. Wen, P. Li, L. Zhang, and Y. Chen, "Stock Market Trend Prediction Using High-Order Information of Time Series," IEEE Access, vol. 7, pp. 28299-28308, 2019, doi: 10.1109/access.2019.2901842.

[26] R.K. Nayak, et al., "Indian Stock Market Prediction Based on Rough Set and Support Vector Machine Approach," in Intell. Cloud Comput. Smart Innov. Syst. Technol., vol 153, Springer, Singapore, 2021, doi: 10.1007/978-981-15-6202-0_35.

[27] Y. Peng, P.H.M. Albuquerque, H. Kimura, and C.A.P.B. Saavedra, "Feature Selection and Deep Neural Networks for Stock Price Direction Forecasting Using Technical Analysis Indicators," Mach. Learn. Appl., vol. 5, p. 100060, 2021, doi: 10.1016/j.mlwa.2021.100060.

[28] X. Yuan, J. Yuan, T. Jiang, and Q.U. Ain, "Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market," IEEE Access, vol. 8, pp. 22672–22685, 2020, doi: 10.1109/access.2020.2969293.

[29] H. Veisi, H.R. Ghaedsharaf, and M. Ebrahimi, "Improving the Performance of Machine Learning Algorithms for Heart Disease Diagnosis by Optimizing Data and Features," Soft Comput. J., vol. 8, no. 1, pp. 70-85, 2019, doi: 10.22052/8.1.70 [In Persian].

[30] A.A.A. Mahdavi and E. Mahdipour, "Prediction of couple relationship during the Covid-19 period using correlation-based feature selection and machine learning," Soft Comput. J., vol. 10, no. 2, pp. 56-71, 2022, doi: 10.22052/scj.2022.243472.1040 [In Persian].

[31] M. Bigdeli, "Classification of transformer faults using frequency response analysis based on cross-correlation technique and support vector machine," Soft Comput. J., vol. 9, no. 1, pp. 2-13, 2020, doi: 10.22052/scj.2021.111448 [In Persian].

[32] X.-S. Yang and S. Deb, "Cuckoo Search: Recent Advances and Applications," Neural Comput. Appl., vol. 24, no. 1, pp. 169-174, 2014, doi: 10.1007/s00521-013-1367-1.

[33] L.A.M. Pereira, "A Binary Cuckoo Search and Its Application for Feature Selection," in Cuckoo Search and Firefly Algorithm. Studies in Computational Intelligence, vol 516, Springer, Cham, 2014, doi: 10.1007/978-3-319-02141-6_7.