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

عرفان صابری[1]، دانشجوی کارشناسی ارشد، الناز رادمند[2]، دانشجوی کارشناسی ارشد، جمشید پیرگزی[3]*، استادیار، دکتر علی کرمانی[4]، استادیار

[1و2و3و4] دانشکده مهندسی برق و کامپیوتر ، دانشگاه علم و فناوری مازندران، بهشهر، ایران

erfan.saberi97@yahoo.com

elnazradmand1377@yahoo.com

j.pirgazi@mazust.ac.ir

a_kermani@mazust.ac.ir

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 شامل برچسب گذاری داده‌ها، استخراج ویژگی، انتخاب ویژگی، طبقه بندی 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 شاخص های مختلف استفاده شده است. نتایج حاصل از ارزیابی نشان دهنده کارایی روش پیشنهادی می باشد.

واژه‌های کلیدی: بورس ایران ، یادگیری ماشین، طبقه بندی، انتخاب ویژگی، الگوریتم تکاملی

* جمشید پیرگزی، **j.pirgazi@mazust.ac.ir**

پذیرفته شده در مجله محاسبات نرم

© پیش نویس مجله علمی محاسبات نرم، برای داوری در مجله

Buying and selling strategy in the Iranian stock market using machine learning models along with feature selection using the Cuckoo Search algorithm

Erfan Saberi [1*], Masters student, Elnaz Radmand [2], Masters student, Jamshid Pirgazi [3], Assistant Professor, Ali Kermani [4], Assistant Professor

[1, 2,3,4] Department of Computer Engineering, University of Science and Technology of Mazandaran, Behshahr, Iran,

erfan.saberi97@yahoo.com

elnazradmand1377@yahoo.com

j.pirgazi@mazust.ac.ir

a_kermani@mazust.ac.ir

Abstract: In Iran, the stock market is facing different conditions compared to the rest of the world. One of the most important challenges in this market is the lack of transparency in market information and information on trading companies. Also, the lack of appropriate and complete historical data for use in forecasting algorithms is another important challenge. In stock price forecasting, due to the dynamic interactions of the stock market and price changes in short periods of time, the use of artificial intelligence is used as a powerful tool in price forecasting and decisions related to buying and selling stocks. In this paper, a method based on machine learning including five steps, including data labeling, feature extraction, feature selection, classification, and signal presentation, is presented. For this purpose, various technical characteristics have been extracted from the price data, and the data has been labeled using the threshold labeling method. Then, various machine learning models are trained on this data and provide buy and sell signals at the output. To improve the performance of the machine learning model, feature selection has been done using the Cuckoo Search algorithm. In order to evaluate the proposed method, several years of Iranian stock market data and various indices have been used. The results of the evaluation show the effectiveness of the proposed method.

Keywords: *Iranian stock market*; *machine learning*; *classification*; *feature selection*; *evolutionary algorithm*

*** Jamshid Pirgazi, j.pirgazi@mazust.ac.ir**

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 چند وجهی بوده و پیش بینی 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 پیش بینی 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 می‌شود، شبکه‌های عصبی [۱-۵]، الگوریتم‌های یادگیری ماشین [۶-۱۰] و مدل‌های مبتنی بر درخت تصمیم‌گیری [۱۱ و ۱۲] است.

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 پیش بینی آینده استفاده می‌شود. برای این کار، از نرم‌افزارهای تحلیل فنی مانند متاتریدر استفاده می‌شود. در تحلیل بنیادی نیز، از عوامل اقتصادی، سیاسی و اخبار شرکت‌ها برای پیش بینی قیمت سهام استفاده می‌شود.

به طورکلی، استفاده از هوش مصنوعی در بازار بورس ایران می‌تواند به عنوان یک راه حل نوین در پیش بینی قیمت سهام و ارائه استراتژی‌های بهینه خرید و فروش در بازار سرمایه باشد. با توجه به پیشرفت روز افزون تکنولوژی ها از آنجایی که بازار بورس ایران به عنوان یکی از بزرگترین بازارهای مالی در منطقه خاورمیانه شناخته می‌شود، پیش بینی قیمت سهام از اهمیت بسیاری برخوردار است.

در بین روش های مبتنی بر هوش مصنوعی اخیر، استفاده از الگوریتم‌های یادگیری عمیق به عنوان یکی از روش‌های مفید در پیش بینی 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 هستند. 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 پیش بینی 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 پیش بینی قیمت بازار بورس مبتنی بر روش های یادگیری ماشین و انتخاب ویژگی می‌باشد. کاربرد انتخاب ویژگی با الگوریتم جستجوی فاخته در بازار سهام ایران نشان دهنده رویکردی نوآورانه برای مقابله با چالش های منحصر به فرد این بازار یعنی انتخاب ویژگی های مفید و تاثیر گذار است. عدم شفافیت بازار سهام ایران و داده های تاریخی ناقص، توسعه الگوریتم های پیش بینی دقیق را دشوار می‌کند. انتخاب ویژگی

با الگوریتم جستجوی فاخته، با شناسایی مرتبط ترین 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 کار های مرتبط در حوزه پیش بینی قیمت سهام مبتنی بر سادگیری ماشین در بازارهای مالی آورده شده است. در بخش سوم روش پیشنهادی مقاله بیان شده است که شامل ۵ بخش برچسب گذاری داده ها، استخراج ویژگی، انتخاب ویژگی و پیش پردازش داده ها و آموزش مدل می‌باشد. در بخش چهارم معیار های ارزیابی و در بخش پنجم تحلیل نتایج آورده شده است و درنهایت در بخش ششم نتیجه گیری بیان شده است.

۲. کارهای مرتبط

تاکنون روش های زیادی برای پیش بینی قیمت سهام مبتنی بر هوش مصنوعی ارائه شده است. بعضی از روش ها در زمینه استخراج ویژگی و انتخاب ویژگی مطالعه کرده اند. بعضی افراد نیز به توسعه مدل های یادگیری ماشین کلاسیک یا مدل های یادگیری عمیق توجه کرده اند.

Wasiat Khan و همکاران در [۱۸] سیستمی را برای پیش‌بینی بازار سهام پیشنهاد می‌کنند که از طبقه‌بندی‌کننده‌های یادگیری ماشین و رسانه‌های اجتماعی و داده‌های خبری استفاده می‌کند. این سیستم از چندین مرحله شامل جمع آوری داده ها، پیش پردازش، استخراج ویژگی و طبقه بندی تشکیل شده است. داده‌های جمع‌آوری‌شده از رسانه‌های اجتماعی و منابع خبری برای حذف اطلاعات نامربوط پیش پردازش شده و از تکنیک‌های استخراج ویژگی برای استخراج مرتبط ترین ویژگی‌ها استفاده می‌شود. نویسندگان از چندین طبقه‌بندی یادگیری ماشین، از جمله درخت تصمیم[1]، جنگل تصادفی[2] و ماشین ماشین بردار پشتیبان[3] برای پیش‌بینی قیمت‌های بازار سهام بر اساس ویژگی‌های استخراج‌شده استفاده کردند. هدف اصلی این پژوهش بررسی تأثیر 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 تأثیر 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 [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 مبتنی بر شبکه‌های حافظه بلندمدت و کوتاه‌مدت[5]، قدرت پیش‌بینی را افزایش داده‌اند. نتایج نشان

[1] Decision Tree (DT)

[2] Random Forest (RF)

[3] Support vector machine (SVM)

[4] Stock market

[5] Long short-term memory (LSTM)

می‌دهد که مدل تک متغیره مبتنی برLSTM و با استفاده از داده‌های یک هفته قبل برای پیش‌بینی قیمت هفته آینده داده های 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 می دهد 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 [۲۰] استفاده از تکنیک‌های هوش مصنوعی، به‌ویژه شبکه عصبی مصنوعی[1]و جنگل تصادفی (RF)، 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 مدل ها با استفاده از شاخص های استاندارد، یعنی خطای جذر میانگین مربعات[2]و میانگین درصد مطلق خطا[3]، با مقادیر کم که نشان دهنده پیش بینی 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 RF عمل می‌کند. نتایج حاصل عملکرد بهتر مدل ANN را با مقادیر RMSE (۰,۴۲)، MAPE (۰,۷۷) و میانگین خطای سوگیری[4](۰,۰۱۳) نشان می‌دهد. همچنین با توجه به نتایج الگوریتم تقویت گرادیان با کمترین خطا و بیشترین دقت نسبت به سایر مدل ها عملکرد بهتری دارد.

در [۲۱] یک رویکرد جدید ساخت سبد سهام[5] با استفاده از یک مدل ترکیبی مبتنی بر یادگیری ماشین برای پیش‌بینی قیمت سهام و مدل میانگین واریانس برای انتخاب سبد سهام توسعه داده شده است. روش پیشنهادی شامل دو مرحله است: پیش‌بینی قیمت سهام و انتخاب سبد سهام. در مرحله اول، یک مدل ترکیبی که ترکیبی از XGBoost [6]با الگوریتم کرم شب تاب بهبود یافته[7] است، برای پیش بینی قیمت سهام برای دوره بعدی استفاده می شود. IFA 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 ریسک 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[8]و تکنیک‌های یادگیری ماشین پیشنهاد می‌کنند. رویکرد پیشنهادی شامل سه مرحله اصلی است: آماده‌سازی داده، استخراج ویژگی و طبقه‌بندی. 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، ویژگی های آماری مختلفی را از داده های پیش پردازش شده استخراج می‌کنند. در نهایت، در مرحله طبقه‌بندی، نویسندگان از الگوریتم‌های یادگیری ماشین مانند

[1] Artificial neural network (ANN)

[2] Root Mean Square Error (RMSE)

[3] Mean absolute percentage error (MAPE)

[4] Mean bias error (MBE)

[5] Portfolio

[6] Extreme Gradient Boosting

[7] Improved firefly algorithm (IFA)

[8] Trend deterministic data

پذیرفته شده در مجله محاسبات نرم

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 داده اند.

نویسندگانی در [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 تأثیرات مستقیم بین ویژگی‌های سهام معرفی می‌کند.آزمایش‌های مقایسه‌ای با استفاده از داده‌های ۱۳ ساله از بورس‌های شانگهای، مقایسه الگوریتم انتخاب ویژگی علی[1] با تجزیه و تحلیل مؤلفه‌های اصلی [2]، درخت‌های تصمیم (DT؛ CART)، و کمترین انقباض مطلق و اپراتور انتخاب[3] 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. یافته ها نشان می‌دهد که CFS دارای پتانسیل قابل توجهی برای افزایش توسعه محصولات QI با بهبود انتخاب ویژگی در پیش بینی سهام است.

در [۲۴] به چالش پیش‌بینی دقیق قیمت سهام در محیط پویا و نویزدار بازار سهام می‌پردازد. مدل‌های محاسباتی نرم در گرفتن الگوهای بازار سهام مؤثر 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[4]، ماشین یادگیری افراطی[5]، ماشین یادگیری افراطی بازگشتی، مدل خطی تعمیم‌یافته[6]، درخت رگرسیون[7] و رگرسیون فرآیند گاوسی[8] استفاده می‌کند. نتایج تجربی بینش های ارزشمندی را در مورد رفتار بازار سهام در شرایط مختلف ارائه می‌دهد. این مطالعه نه تنها بهترین مدل پیش‌بینی را شناسایی می‌کند، بلکه چگونگی تأثیر 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 ریسک‌های بالقوه را در بازارهای سهام داشته باشند.

علاوه بر این، در [۲۵] جیمی تی افیرد پیشنهاد کرد که استخراج ویژگی موثر می تواند دقت و پیش بینی مدل های 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 تکنیک‌های یادگیری ماشین می‌توانند وابستگی‌ها را در توالی‌های قیمت بازار سهام شناسایی کنند. ماهیت پر نوسان و غیر ثابت 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

[1] Causal feature selection (CFS)

[2] Principal component analysis (PCA)

[3] LASSO

[4] Jordan Recurrent Neural Network

[5] Extreme Learning Machine

[6] Generalized Linear Model

[7] Regression Tree

[8] Gaussian Process Regression

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 [۲۶] مطالعه ای 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 پیش بینی بازار سهام می پردازد و اهمیت پیش بینی روندهای آتی بر اساس داده های تاریخی را برجسته می کند. با توجه به نوسانات مداوم در قیمت سهام، پیش بینی دقیق آنها کار دشواری است. سرمایه‌گذاران و محققان با استفاده از تکنیک‌های 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، شاخص های Sensex و Nifty به عنوان شاخص های معیار برای پیش بینی 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 تصمیم گیری سودآور در معاملات پولی با ارائه پیش بینی های دقیق تر بازار سهام هستند، نویدبخش است.

مطالعه دیگری توسط کیم و همکاران در [۲۷] انجام شد که از یک الگوریتم ژنتیک برای انتخاب مرتبط ترین شاخص های فنی برای پیش بینی قیمت سهام استفاده کردند. این مقاله به طور خاص بر روی انتخاب ویژگی برای مدل های شبکه عصبی عمیق که برای پیش بینی جهت قیمت سهام استفاده می‌شود، تمرکز می کند. این مطالعه ۱۲۴ شاخص تحلیل تکنیکال را که معمولاً در ادبیات و وب سایت های 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 طبقه بندی مختلف، سودآوری و هزینه های 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 پارامترها برای مدل‌های پیش‌بینی روند قیمت سهام مبتنی بر یادگیری ماشین با استفاده از اعتبارسنجی متقابل پنجره زمانی [4] 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.

[1] Frequency trading patterns modeling approach

[2] Rough Set Theory

[3] Linear multi-factor

[4] Time-sliding window cross-validation

۳. روش پیشنهادی

در این مقاله، به پیچیدگی های بازار سهام ایران پرداخته شده است، اکوسیستمی منحصر به فرد که چالش های 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 مؤثر 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.

دراین مقاله اطلاعات قیمت ۵ سهم مختلف بورس ایران (خودرو، وخارزم،وتجارت،شپنا، حکشتی) با استفاده از کتابخانه Finpy در پایتون گرفته شده است. پس از استخراج دیتاست از ۴ ویژگی قیمت باز شدن، بالاترین قیمت، پایین ترین قیمت و قیمت بسته شدن که به صورت تعدیل شده در یک روز کاری می‌باشند، برای به دست آوردن ویژگی های جدید و برچسب گذاری داده ها استفاده شده است.

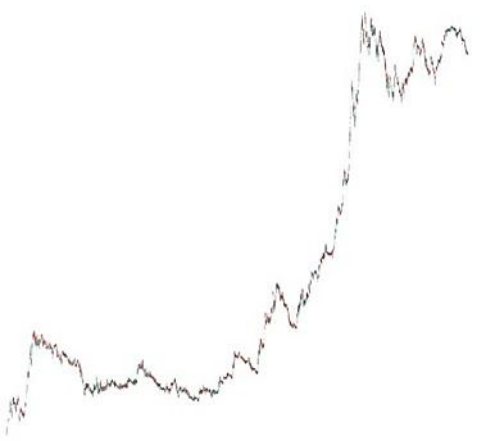
یکی از جنبه های مهم در بازار سهام ایران در تمایز بین نمودارهای تعدیل شده و تعدیل نشده است. نمودار تعدیل شده (شکل ۲) شامل تنظیمات مختلفی مانند تعدیل سود، تعدیل سهام جدید و تعدیل شاخص است. این تعدیل ها برای محاسبه تغییرات موجودی سرمایه، تقسیم سهام جدید و سود سهام محاسبه می‌شوند. در مقابل، نمودار تعدیل نشده (شکل ۱) این تعدیل ها را در نظر نمی‌گیرد و نشان دهنده قیمت اصلی سهام و نقدی است. در حالی که نمودار تعدیل نشده تغییرات روزانه قیمت سهام را از طریق نوسانات نمودار منعکس می کند، نمودار تعدیل شده با ترکیب تعدیلات محاسبه شده نمای جامع تری را ارائه می‌دهد. نمودار تعدیل شده برای بررسی تغییرات بلندمدت قیمت سهام و مقایسه قیمت ها در دوره های مختلف بسیار ارزشمند است. از سوی دیگر، نمودار تعدیل نشده برای بررسی دقیق تغییرات روزانه قیمت سهام ایده آل است. با توجه به نامناسب بودن نمودارهای تعدیل نشده برای آموزش مدل به دلیل شکاف قیمتی، از داده های قیمت تعدیل شده برای آموزش مدل استفاده شده است.برای درک بیشتر از شکل ۱ و شکل ۲ استفاده شده تا تفاوت بین نمودار تعدیل شده و تعدیل نشده نمایش داده شود.

این رویکرد نوآورانه با ترکیب ماهرانه 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 مؤثر در معاملات سهام می‌دهد. استفاده جامع از داده های قیمت تعدیل شده و انتخاب دقیق ویژگی ها به قابلیت اطمینان و استحکام روش کمک می کند و آن را قادر می سازد بر محدودیت های ناشی از داده های تاریخی ناقص و عدم شفافیت غلبه کند. در بازاری که با شرایط متمایز آن مشخص می‌شود، این روش به عنوان یک چراغ امید ظاهر می‌شود و هوش مصنوعی پیشرفته را پر می‌کند.

پیش نویس مجله محاسبات نرم

شکل ۱. نمودار سهم فارس در تایم فریم روزانه در حالت بدون تعدیل قیمتی



شکل ۲.نمودار سهم فارس در تایم فریم روزانه در حالت تعدیل عملکردی

۱،۳. روش برچسب گذاری آستانه ای

در این روش درصد اختلاف قیمت بسته شدن تعدیل شده در هر روز با روز آینده محاسبه می‌شود و اگر این درصد اختلاف از یک حد آستانه ای بیشتر باشد، برچسب ۲، اگر کمتر از منفی حد آستانه باشد، برچسب ۱ و در غیر این صورت، برچسب ۰ به داده مطابق رابطه زیر تعلق می گیرد:

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

که در رابطه ۲، T حد آستانه مورد نظر می باشد که در این مقاله ۳،۵ درصد در نظر گرفته شده است. در رابطه ۱ اختلاف درصد محاسبه شده است و در رابطه ۲ بر اساس آن برچسب گذاری داده ها 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.

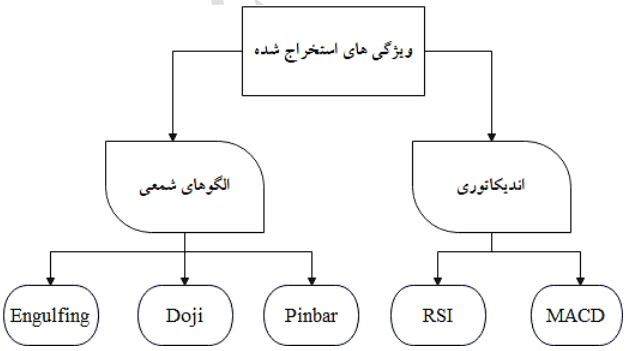
۲،۳. استخراج ویژگی

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[1] و همگرایی و واگرایی میانگین متحرک[1] می‌باشد.

[1] Relative strength index (RSI)

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.



شکل۳.انواع ویژگی های استخراج شده

برای ویژگی‌های از نوع الگوهای شمعی از الگوهای معروفی مانند الگوی پوشای صعودی و نزولی [۲] و انواع الگوهای دوجی [۳] و الگوهای پین بار [۴] استفاده شده است. پس از ایجاد این دو نوع ویژگی، این ویژگی ها 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.

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 ک 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 بیش فروش 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 و ۳۰] یک گام مهم در تکنیک های یادگیری ماشین [۳۱] ، از جمله پیش‌بینی بازار سهام، برای بهبود دقت مدل و کاهش زمان محاسبات است. الگوریتم CS [۵] یک الگوریتم بهینه سازی مبتنی بر جمعیت است که از رفتار پرندگان فاخته الهام گرفته شده است. این الگوریتم به دلیل کارایی و تاثیر آن در مدیریت مجموعه داده های بزرگ به طور گسترده در انتخاب ویژگی استفاده می‌شود [۳۲].

فرآیند انتخاب ویژگی با الگوریتم CS [۳۳] را می توان به چند مرحله تقسیم کرد. اولین گام رمزگذاری مسئله انتخاب ویژگی به عنوان یک مسئله بهینه‌سازی است، که در آن تابع هدف، حداکثر

[1] Moving Average Convergence Divergence (MACD)

[2] Bullish and Bearish Engulfing pattern

[3] Doji

[4] Pin Bar

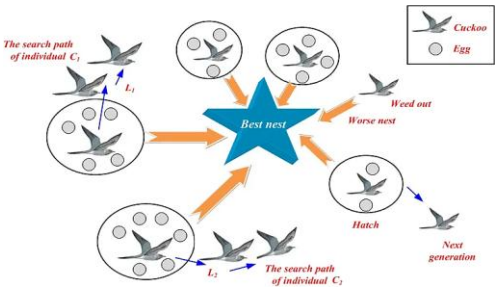
[5] Cuckoo search

کردن دقت پیش‌بینی مدل و در عین حال به حداقل رساندن تعداد ویژگی‌های انتخاب‌شده است. سپس، جمعیتی از راه‌حل‌های کاندید، که به صورت رشته‌های باینری هستند و نشان‌دهنده وجود یا عدم وجود هر ویژگی می‌باشند، به‌طور تصادفی مقداردهی اولیه می‌شوند.

در هر تکرار، با استفاده از الگوریتم CS، یک راه حل کاندید جدید تولید می‌شود. این راه حل شامل ایجاد یک لانه فاخته جدید است، جایی که یک پرنده فاخته تخم می‌گذارد، که یا نسخه اصلاح شده یکی از لانه های موجود است یا یک راه حل کاملاً 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.

پس از تولید راه حل کاندید جدید، تابع هدف برای اندازه گیری برازش آن ارزیابی می‌شود. راه حل کاندید با بالاترین مقدار تابع برازش به عنوان بهترین راه حل و ویژگی‌های آن به عنوان ویژگی‌های انتخابی مدل درنظر 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 بازار سهام است. این شامل رمزگذاری مسئله انتخاب ویژگی به عنوان یک مسئله 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.



شکل۴. انتخاب ویژگی با استفاده از الگوریتم فاخته [۷]

مقادیر پارامترهای الگوریتم فاخته با توجه به جدول ۱ مقدار دهی شده است.

جدول۱. مقادیر متغیر ها

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

تعداد کل ویژگی های استخراج شده ۲۹ می‌باشد که از میان این تعداد ویژگی‌ها، فقط ۳ ویژگی انتخاب شده است که شامل بهترین ویژگی ها می‌باشد.

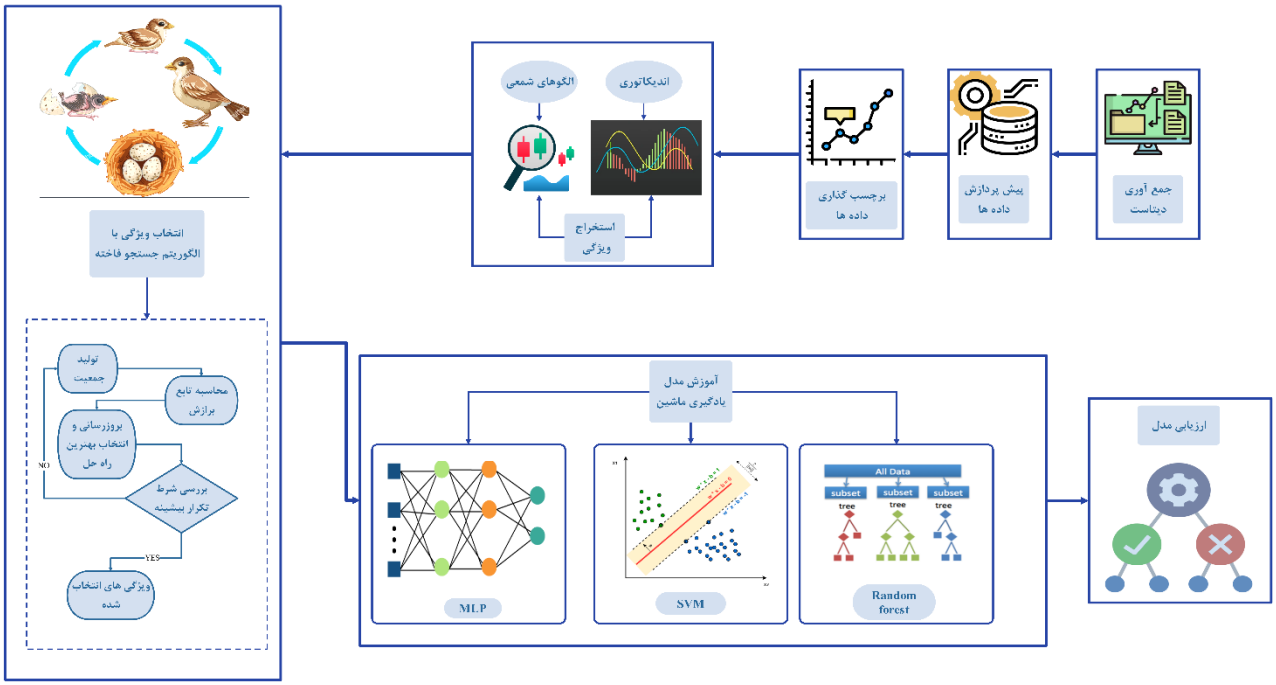
۴،۳. پیش پردازش داده‌ها و آموزش مدل:

بعد از این که برچسب گذاری و استخراج ویژگی انجام شد، داده ها به دو دسته داده‌های تست و آموزش تقسیم شده است، به نحوی که سی درصد از داده ها به طور تصادفی به عنوان داده تست و بقیه داده‌ها به عنوان داده آموزشی انتخاب می‌شوند. برای نرمال‌سازی ویژگی‌ها از روش مقیاس کننده استاندارد[1] استفاده شده است. در یادگیری ماشین از مقیاس کننده استاندارد برای تغییر اندازه توزیع مقادیر استفاده می شود به طوری که میانگین مقادیر مشاهده شده ۰ و انحراف استاندارد ۱ باشد. مقیاس کننده استاندارد یک تکنیک مهم است که عمدتاً به عنوان یک مرحله پیش پردازش قبل از بسیاری از مدل‌های یادگیری

[1] StandardScaler

ماشین انجام می شود تا محدوده عملکرد مجموعه داده ورودی را استاندارد کند.

در این مقاله از روش های یادگیری ماشین K–نزدیک‌ترین همسایگی[1]، XGBoost، ماشین بردار پشتیبان، جنگل تصادفی، بیز ساده[2] و پرسپترون چند لایه[3] استفاده شده است و دیتاست



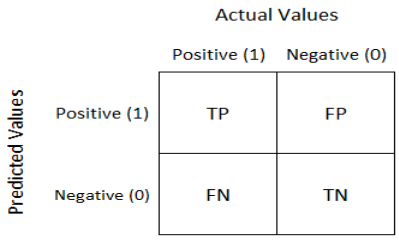
سهم های مختلف بر روی این مدل ها آموزش دیده‌اند. شمای کلی مراحل مختلف مقاله در شکل ۵ نمایش داده شده است.

[1] K-Nearest Neighbors

[2] Naive Bayes

[3] Multilayer perceptron

شکل۵. شمای کلی مراحل مختلف مدل پیشنهادی



شکل۶. تصویری از یک ماتریس تداخل

۴. معیار های ارزیابی

برای معیار ارزیابی مدل های یادگیری ماشین از ماتریس تداخل [۱]، صحت [۲]، حساسیت [۳]، معیار F1 [۴]، دقت [۵] و منحنی مشخصه عملکرد [۶] سیستم استفاده شده است.

۴,۱. Confusion Matrix (ماتریس تداخل)

ماتریس تداخل به ما کمک می‌کند تا عملکرد یک مدل یا نحوه پیش بینی یک مدل را در یادگیری ماشین نشان دهیم. این ماتریس از ۴ عنصر اصلی تشکیل شده است که معیارهای مختلفی را برای شمارش تعداد پیش بینی‌های صحیح و نادرست نشان می‌دهد. اگر برچسب های پیش بینی شده و واقعی مطابقت داشته باشند، گفته می شود که پیش بینی صحیح است، اما زمانی که برچسب های پیش بینی شده و واقعی با هم تطبیق نداشته باشند، آنگاه گفته می شود که پیش بینی نادرست است. علاوه بر این، مثبت و منفی نشان دهنده برچسب های پیش بینی شده در ماتریس است. شکل ۶ نمای کلی یک ماتریس تداخل را نشان می دهد.

۴,۲. Precision (صحت)

صحت به عنوان نسبت نمونه های مثبت طبقه بندی شده به درستی (مثبت واقعی[۷]) به تعداد کل نمونه های مثبت طبقه بندی شده (به درستی یا نادرست) تعریف می‌شود و از رابطه زیر محاسبه می‌شود.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (3)$$

۴,۳. Recall (حساسیت)

حساسیت به عنوان نسبت بین تعداد نمونه های مثبت که به درستی به عنوان مثبت طبقه بندی شده اند به تعداد کل نمونه های مثبت محاسبه می‌شود و طبق رابطه ۴ به دست می‌آید. این معیار توانایی مدل را در تشخیص نمونه های مثبت اندازه گیری می‌کند.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (4)$$

۴,۴. معیار F1

صحت و حساسیت دو عنصر سازنده معیار F1 هستند. هدف معیار F1 ترکیب معیارهای صحت و حساسیت در یک متریک واحد است. در عین حال، معیار F1 طوری طراحی شده است

[1] Confusion Matrix

[2] Precision

[3] Recal

[4] F1-score

[5] Accuracy

[6] Receiver Operating Characteristic

[7] True Positive

که روی داده های نامتعادل به خوبی کار کند. معیار F1 به عنوان میانگین هارمونیک دقت و یادآوری مطابق رابطه ۵ تعریف می‌شود.

$$F1 = \frac{Precision \times Recall}{Precision + Recall} \times 2 \quad (5)$$

۴،۵. منحنی ROC (مشخصه عملکرد سیستم)

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) ترسیم می‌کند که ۱ منهای نرخ منفی واقعی (TNR) است که 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 مقایسه عملکرد طبقه‌بندی‌کننده‌های مختلف و انتخاب آستانه مناسب بر اساس مبادله بین TPR و FPR فراهم می‌کند.

۵. تحلیل نتایج

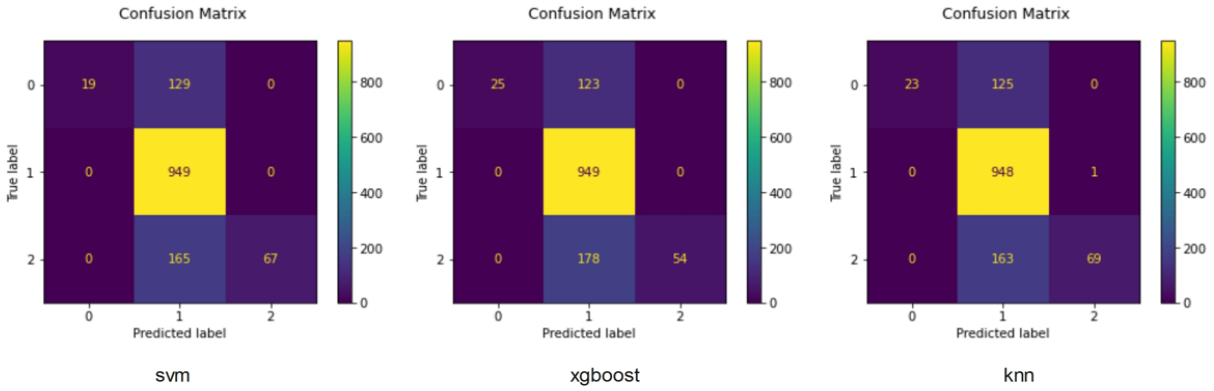
در بخش تحلیل نتایج، نتایج معیار های ارزیابی مدل های یادگیری ماشین بر روی چندین سهم مورد بررسی قرار گرفته است و از ماتریس تداخل و نمودار مشخصه عملکرد سیستم برای نمایش خروجی استفاده شده است. در جداول زیر به بررسی دقت مدل های یادگیری ماشین در سهم های گوناگون پرداخته شده است. در جدول ۴ دقت تمامی مدل ها در سهم های گوناگون با یکدیگر مقایسه شده است و در نهایت برای درک بهتر، نمودار مستطیلی مربوط به این جداول ترسیم شده است.

جدول۲. نتایج ارزیابی روش های یادگیری ماشین در سهم خودرو

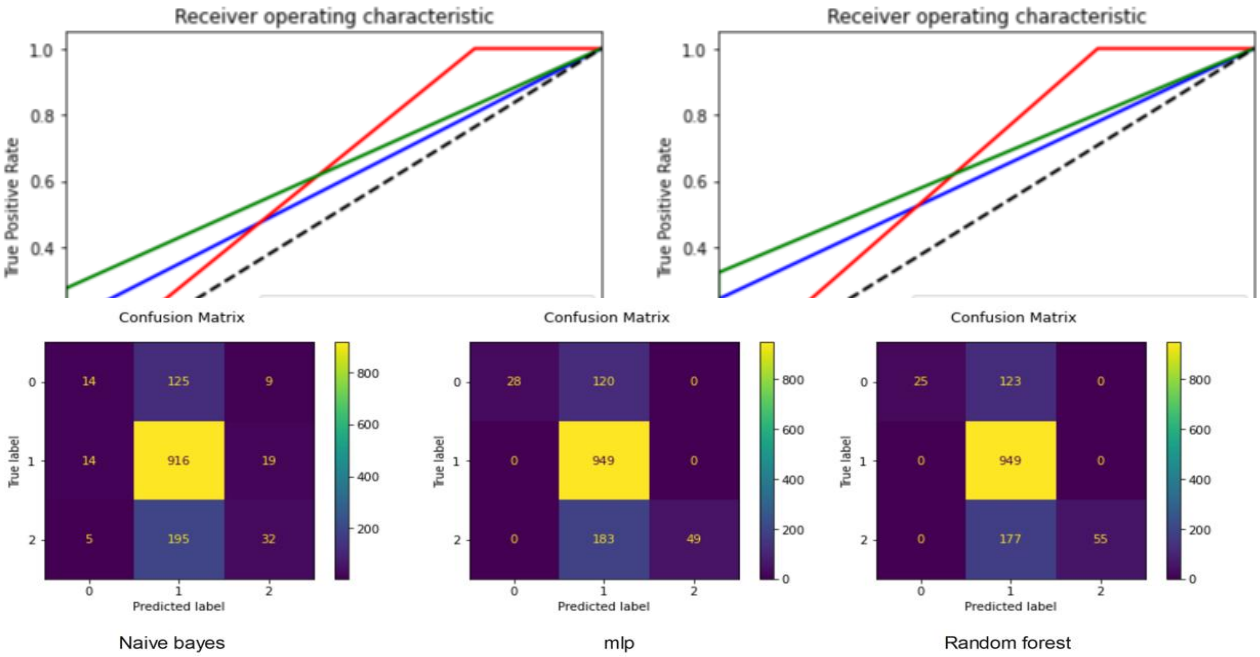
مدل یادگیری ماشین	Label	Precision	Recall	F1-score	Accuracy
SVM	۰	۱،۰۰	۰،۱۴	۰،۲۵	۰،۷۸
	۱	۰،۷۷	۱،۰۰	۰،۸۷	
	۲	۰،۹۹	۰،۳۱	۰،۴۷	
XGBoost	۰	۱،۰۰	۰،۱۷	۰،۲۹	۰،۷۷
	۱	۰،۷۶	۱،۰۰	۰،۸۶	
	۲	۱،۰۰	۰،۲۳	۰،۳۸	
KNN	۰	۰،۸۶	۰،۱۷	۰،۲۸	۰،۷۹
	۱	۰،۷۷	۱،۰۰	۰،۸۷	
	۲	۱،۰۰	۰،۳۲	۰،۴۹	
RF	۰	۱،۰۰	۰،۱۷	۰،۲۹	۰،۷۷
	۱	۰،۷۶	۱،۰۰	۰،۸۶	

پیش نویس مجله محاسبات نرم

	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. ماتریس های تداخل مربوط به سهم خودرو برای مدل های یادگیری ماشین



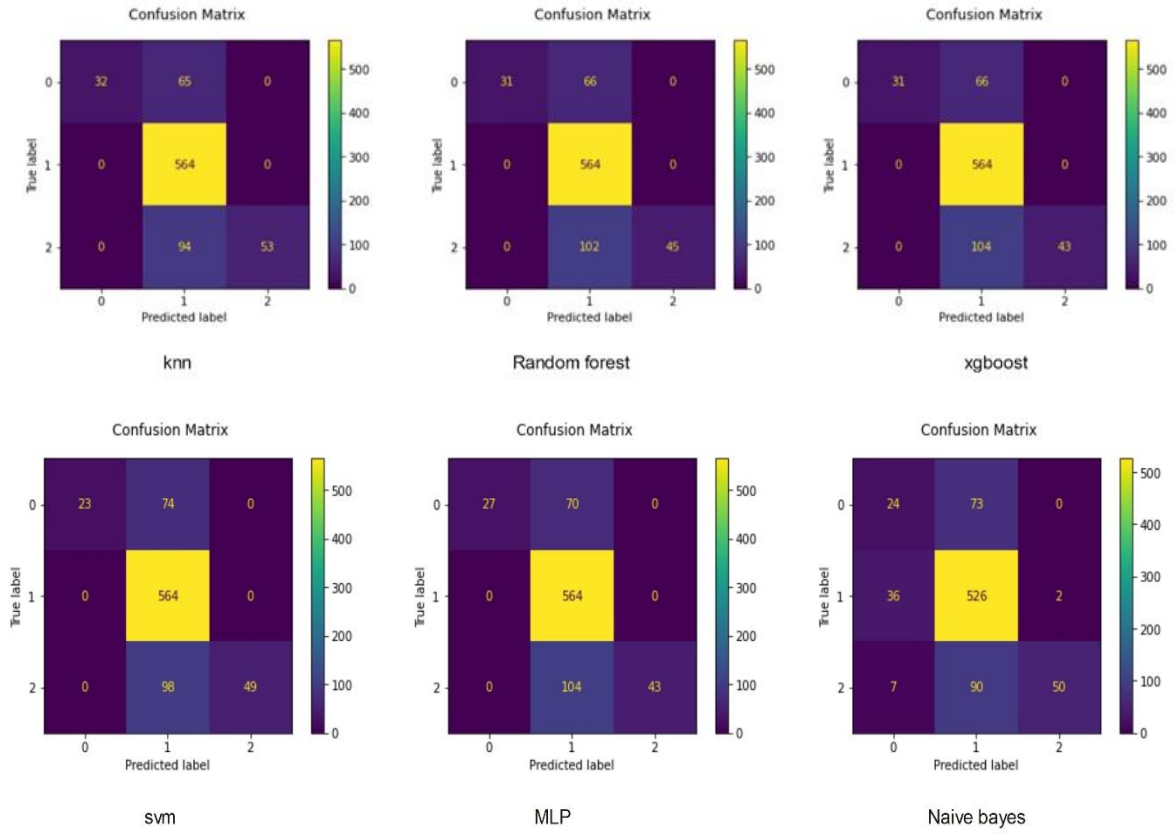
شکل ۸. منحنی مشخصه عملکرد سیستم (ROC) سهم خودرو برای مدل های یادگیری ماشین KNN (شکل الف) و MLP (شکل ب)

با توجه به جدول ۲ و شکل های ۷ و ۸ که به ترتیب ماتریس‌های تداخل و منحنی‌های مشخصه عملکرد سیستم سهم خودرو را نمایش می‌دهد، مدل های MLP و KNN تحت معیار های ارزیابی مختلف، عملکرد بهتری نسبت به بقیه مدل ها داشته اند. اغلب مدل‌های یادگیر در سهم خودرو دقت بالایی داشتند؛ زیرا که در سهم خودرو نسبت به دیگر سهم ها داده های بیشتری برای آموزش مدل وجود داشت و مدل ها توانستند به خوبی آموزش ببینند. تمامی مدل های یادگیری ماشین در تمامی سهم‌ها وقتی که هر دو برچسب پیش بینی شده و واقعی یک باشند نتیجه خوبی را نشان می‌دهند.

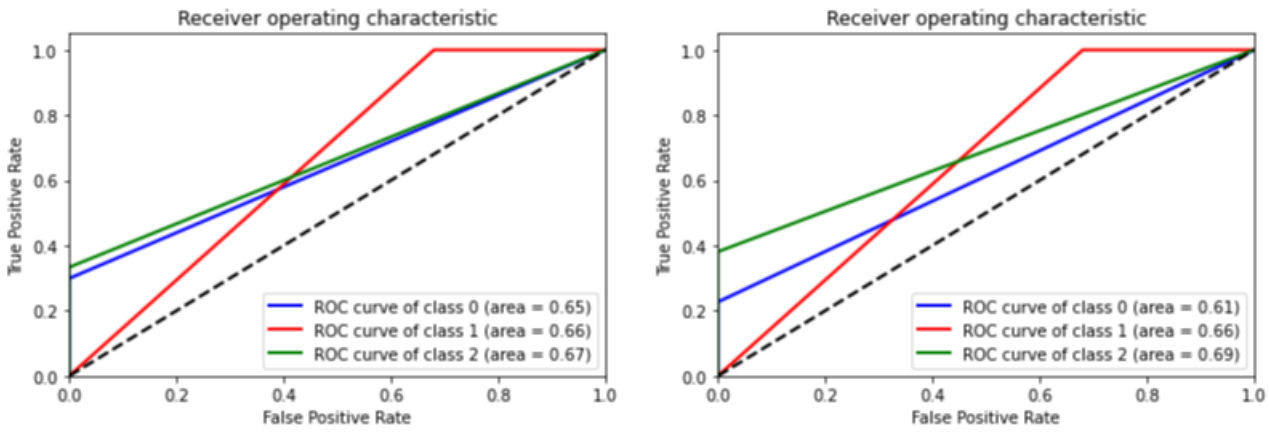
جدول۳. نتایج ارزیابی روش های یادگیری ماشین در سهم وتجارت

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	

پذیرفته شده در مجله محاسبات نرم



شکل ۹. ماتریس های تداخل مربوط به سهم وتجارت برای مدل های یادگیری ماشین



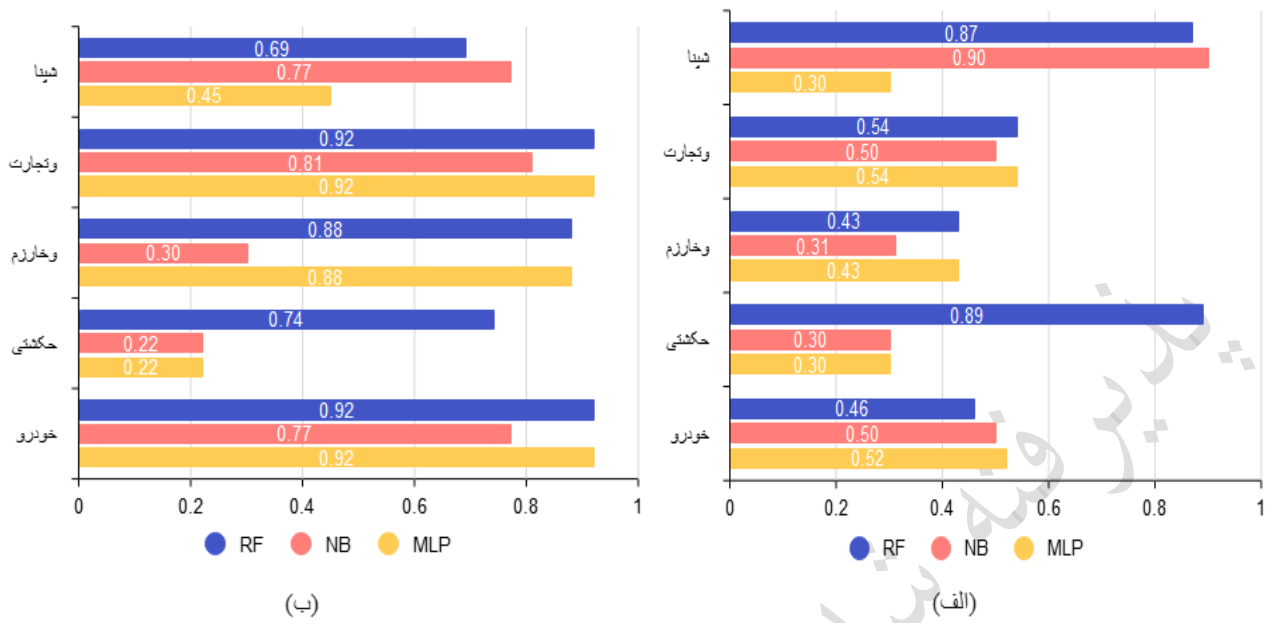
شکل ۱۰. منحنی مشخصه عملکرد سیستم (ROC) سهم وتجارت برای مدل های یادگیری ماشین KNN (شکل الف) و MLP (شکل ب)

با توجه به جدول ۳ و شکل های 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 مدل ها دارا می‌باشد. سپس مدل‌های SVM و XGBoost و Random forest دقت یکسان ۷۹ درصدی داشته اند. ماتریس تداخل مربوط به سهم وتجارت در سلول (۱،۱) تعداد موارد مثبت واقعی را نشان می‌دهد که نسبت به بقیه قسمت ها 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.

در جدول ۴ درصد تمامی مدل ها در ۵ سهم مختلف مورد مقایسه قرار گرفته است. همان طور که مشاهده می‌شود، مدل های 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 درصد ۸۵ و همچنین مدل MLP در سهم خودرو درصد ۸۰ را نشان می‌دهند. در نمودار های شکل ۱۱ میانگین مقادیر Precision (صحت) و Recall (حساسیت) مدل های 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 سهم ها توانسته عملکرد بهتری از نظر معیار حساسیت و صحت کسب کند.

جدول۴. مقایسه دقت مدل های یادگیری ماشین در سهم های مختلف

مدل یادگیری ماشین	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 Recall (شکل الف) و precision (شکل ب) مدل های یادگیری ماشین گوناگون در سهم های مختلف

۶. نتیجه گیری

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، مجموعه ای از ویژگی های تکنیکال، از جمله اندیکاتورهای RSI ، MACD و الگوهای شمعی از داده های قیمت سهام استخراج شده است. سپس از الگوریتم جستجوی فاخته برای انتخاب مهم‌ترین ویژگی‌ها برای پیش‌بینی قیمت سهام استفاده شده است. عملکرد ویژگی های انتخاب شده با استفاده از مدل های 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 منـافعی ندارند.

مراجع

[۱] Selvamuthu, Dharmaraja, et al. “Indian Stock Market Prediction Using Artificial Neural Networks on Tick Data.” *Financial Innovation*, vol. 5, no. 1, Mar. 2019, https://doi.org/10.1186/s40854-019-0131-7.

[۲] Qiu, Jiayu, et al. “Forecasting Stock Prices with Long-Short Term Memory Neural Network Based on Attention Mechanism.” *PLOS ONE*, edited by Tao Song, vol. 15, no. 1, Jan. 2020, p. e0227222, https://doi.org/10.1371/journal.pone.0227222.

[۳] Yu, Pengfei, and Xuesong Yan. “Stock Price Prediction Based on Deep Neural Networks.” *Neural Computing and Applications*, Apr. 2019, https://doi.org/10.1007/s00521-019-04212-x.

[٤] Ding, Guangyu, and Liangxi Qin. "Study on the Prediction of Stock Price Based on the Associated Network Model of LSTM." *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 6, Nov. 2019, pp. 1307–17, https://doi.org/10.1007/s13042-019-01041-1.

[٥] Kim, Taewook, and Ha Young Kim. "Forecasting Stock Prices with a Feature Fusion LSTM-CNN Model Using Different Representations of the Same Data." *PLOS ONE*, edited by Alejandro Raul Hernandez Montoya, vol. 14, no. 2, Feb. 2019, p. e0212320, https://doi.org/10.1371/journal.pone.0212320.

[٦] Meesad, Phayung, and Risul Islam Rasel. "Predicting Stock Market Price Using Support Vector Regression." *2013 International Conference on Informatics, Electronics and Vision (ICIEV)*, May 2013, https://doi.org/10.1109/iciev.2013.6572570. Accessed 5 Oct. 2020.

[٧] Obthong, Mehtabhorn, et al. "A Survey on Machine Learning for Stock Price Prediction: Algorithms and Techniques." *Eprints.soton.ac.uk*, 6 May 2020, pp. 63–71, https://doi.org/10.5220/0009340700630071.

[٨] Pan, Jiaqi, et al. "The Impact of Data Normalization on Stock Market Prediction: Using SVM and Technical Indicators." *Communications in Computer and Information Science*, 2016, pp. 72–88, https://doi.org/10.1007/978-981-10-2777-2_7.

[٩] Ghosh, Pushpendu, et al. "Forecasting Directional Movements of Stock Prices for Intraday Trading Using LSTM and Random Forests." *Finance Research Letters*, July 2021, p. 102280, https://doi.org/10.1016/j.frl.2021.102280.

[١٠] Zhang, Jun, et al. "Predicting Stock Price Using Two-Stage Machine Learning Techniques." *Computational Economics*, July 2020, https://doi.org/10.1007/s10614-020-10013-5. Accessed 10 Aug. 2020.

[١١] Nair, Binoy B., et al. "A Stock Market Trend Prediction System Using a Hybrid Decision Tree-Neuro-Fuzzy System." *2010 International Conference on Advances in Recent Technologies in Communication and Computing*, Oct. 2010, https://doi.org/10.1109/artcom.2010.75. Accessed 30 Nov. 2020.

[١٢] Panigrahi, S. S., and J. K. Mantri. "Epsilon-SVR and Decision Tree for Stock Market Forecasting." *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, Oct. 2015, https://doi.org/10.1109/icgciot.2015.7380565. Accessed 24 Apr. 2022.

[١٣] Enke, David, et al. "Stock Market Prediction with Multiple Regression, Fuzzy Type-2 Clustering and Neural Networks." *Procedia Computer Science*, vol. 6, 2011, pp. 201–6, https://doi.org/10.1016/j.procs.2011.08.038. Accessed 15 Apr. 2019.

[١٤] Cakra, Yahya Eru, and Bayu Distiawan Trisedya. "Stock Price Prediction Using Linear Regression Based on Sentiment Analysis." *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Oct. 2015, https://doi.org/10.1109/icacsis.2015.7415179.

[١٥] Panwar, Bhawna, et al. "Stock Market Prediction Using Linear Regression and SVM." *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Mar. 2021, https://doi.org/10.1109/icacite51222.2021.9404733. Accessed 19 Sept. 2022.

[١٦] Izzah, Abidatul, et al. "Mobile App for Stock Prediction Using Improved Multiple Linear Regression." *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*, Nov. 2017, https://doi.org/10.1109/siet.2017.8304126. Accessed 29 Apr. 2022.

[١٧] Enke, David, and Nijat Mehdiyev. "Stock Market Prediction Using a Combination of Stepwise Regression Analysis, Differential Evolution-Based Fuzzy Clustering, and a Fuzzy Inference Neural Network." *Intelligent Automation & Soft Computing*, vol. 19, no. 4, Dec. 2013, pp. 636–48, https://doi.org/10.1080/10798587.2013.839287. Accessed 28 Feb. 2020.

[١٨] Khan, Wasiat, et al. "Stock Market Prediction Using Machine Learning Classifiers and Social Media, News." *Journal of Ambient Intelligence and Humanized Computing*, Mar. 2020, https://doi.org/10.1007/s12652-020-01839-w.

[١٩] Mehtab, Sidra, et al. "Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models." *Communications in Computer and Information Science*, 2021, pp. 88–106, https://doi.org/10.1007/978-981-16-0419-5_8. Accessed 24 May 2022.

[٢٠] Vijh, Mehar, et al. "Stock Closing Price Prediction Using Machine Learning Techniques." *Procedia Computer Science*, vol. 167, no. 167, 2020, pp. 599–606, https://doi.org/10.1016/j.procs.2020.03.326.

[۲۱] Chen, Wei, et al. “Mean–Variance Portfolio Optimization Using Machine Learning-Based Stock Price Prediction.” *Applied Soft Computing*, vol. 100, Mar. 2021, p. 106943, https://doi.org/10.1016/j.asoc.2020.106943.

[۲۲] Patel, Jigar, et al. “Predicting Stock and Stock Price Index Movement Using Trend Deterministic Data Preparation and Machine Learning Techniques.” *Expert Systems with Applications*, vol. 42, no. 1, Jan. 2015, pp. 259–68, https://doi.org/10.1016/j.eswa.2014.07.040.

[۲۳] Zhang, Xiangzhou, et al. “A Causal Feature Selection Algorithm for Stock Prediction Modeling.” *Neurocomputing*, vol. 142, Oct. 2014, pp. 48–59, https://doi.org/10.1016/j.neucom.2014.01.057. Accessed 2 Apr. 2020.

[۲۴] Göçken, Mustafa, et al. “Stock Price Prediction Using Hybrid Soft Computing Models Incorporating Parameter Tuning and Input Variable Selection.” *Neural Computing and Applications*, vol. 31, no. 2, July 2017, pp. 577–92, https://doi.org/10.1007/s00521-017-3089-2.

[۲۵] Wen, Min, et al. “Stock Market Trend Prediction Using High-Order Information of Time Series.” *IEEE Access*, vol. 7, 2019, pp. 28299–308, https://doi.org/10.1109/access.2019.2901842.

[۲۶] Rudra Kalyan Nayak, et al. *Indian Stock Market Prediction Based on Rough Set and Support Vector Machine Approach.* Aug. 2020, pp. 345–55, https://doi.org/10.1007/978-981-15-6202-0_35. Accessed 4 July 2023.

[۲۷] Peng, Yaohao, et al. “Feature Selection and Deep Neural Networks for Stock Price Direction Forecasting Using Technical Analysis Indicators.” *Machine Learning with Applications*, vol. 5, Sept. 2021, p. 100060, https://doi.org/10.1016/j.mlwa.2021.100060. Accessed 25 July 2021.

[۲۸] Yuan, Xianghui, et al. “Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market.” *IEEE Access*, vol. 8, 2020, pp. 22672–85, https://doi.org/10.1109/access.2020.2969293. Accessed 25 July 2021.

[۲۹] Hadi Veisi; HamidReza Ghaedsharaf; Morteza Ebrahimi. "Improving the Performance of Machine Learning Algorithms for Heart Disease Diagnosis by Optimizing Data and Features", Soft Computing Journal, 8, 1, 2019, 70-85. magiran.com/p2177737

[۳۰] AliAsghar Akhavan Mahdavi; Elham Mahdipour . "Prediction of Couple Relationship in the Covid-19 Era Using Correlation Based Feature Selection and Machine Learning", Soft Computing Journal, 10, 2, 2023, 56-71. magiran.com/p2556161

[۳۱] Mehdi Bigdeli. "Classification of transformer faults using frequency response analysis based on cross-correlation technique and support vector machine", Soft Computing Journal, 9, 1, 2022, 2-13. magiran.com/p2272573

[۳۲] Yang, Xin-She, and Suash Deb. “Cuckoo Search: Recent Advances and Applications.” *Neural Computing and Applications*, vol. 24, no. 1, Mar. 2013, pp. 169–74, https://doi.org/10.1007/s00521-013-1367-1. Accessed 26 Aug. 2021.

[۳۳] Pereira, Luis S., et al. *A Binary Cuckoo Search and Its Application for Feature Selection*. Nov. 2013, pp. 141–54, https://doi.org/10.1007/978-3-319-02141-6_7. Accessed 4 July 2023.

پیش نویس مقاله برای مجله محاسبات نرم