سامانه‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)

Summer 2024, 5 (2), 1-9
DOR:

Received: 13 Apr 2024
Accepted: 1 Jun 2024

مقاله پژوهشی

Leveraging Machine Learning for Optimal Trade Entry Point Identification in the Cryptocurrency Market

Abulfazl Yavari[1*], Hasan Aama[2], Mohammad Reza Hashemi[3]

1. Assistant Professor, Faculty of Computer Engineering and IT, Payam-e Noor University, Tehran, Iran.
**Corresponding Author*, A.Yavari@pnu.ac.ir
2. Assistant Professor, Faculty of Management Economics and Accounting, Payam-e Noor University, Tehran, Iran.
3. Assistant Professor, Faculty of Computer Engineering, Skill National University, Tehran, Iran.

Abstract

Introduction: In the domain of financial forecasting, machine learning (ML) models have garnered significant attention in recent years. One prominent application lies in the cryptocurrency market, where intelligent trading bots facilitate a substantial portion of daily transactions.

Method: This paper investigates the efficacy of ML-based methods for identifying optimal trade entry points. Specifically, we employ the Relative Strength Index (RSI) and Simple Moving Average (SMA) indicators to extract a set of trend and crossover features. Subsequently, these features are utilized to train multilayer neural network, support vector machine, and nearest neighbor learning models. The performance of each model is then evaluated using digital currency market data encompassing the first seven months of 2023 for a variety of cryptocurrencies.

Results: Our findings demonstrate that, firstly, the extracted features exhibit promising efficacy. Secondly, the nearest neighbor model achieved the highest profitability during the evaluation period compared to the other investigated models.

Discussion: In the research conducted in this field, technical indicators are often used directly in market forecasting but in the proposed method of this article, instead of directly using the values of the indicators as the input of the classification methods, trend behaviors and their intersections have been used. In the continuation of this research, it is possible to determine the best exit points from the transaction with the help of machine learning and the use of volume indicators in the learning process.

Keywords: Cryptocurrency market, market forecasting, artificial neural network, support vector machine, nearest neighbor, technical analysis.

سامانه‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)
انجمن علمی تجارت الکترونیکی ایران
واحد زنجان

شناسایی نقطه ورود بهینه معاملات در بازار ارزهای دیجیتال با استفاده از یادگیری ماشین

تاریخ دریافت: ۲۵/ ۰۱/ ۱۴۰۳
تاریخ پذیرش: ۱۲/ ۰۳/ ۱۴۰۳

دورهٔ پنجم، تابستان ۱۴۰۳
شمارهٔ دوم، صص: ۱-۹

ابوالفضل یاوری[1*]، حسن اعمی [2]، محمدرضا هاشمی [3]

۱- استادیار، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه پیام نور، تهران، ایران. (نویسندهٔ مسئول) A.Yavari@pnu.ac.ir
۲- استادیار، گروه مدیریت اقتصاد و حسابداری، دانشگاه پیام نور، تهران، ایران. Hasanaama@pnu.ac.ir
۳- گروه مهندسی کامپیوتر، دانشگاه ملی مهارت، تهران، ایران. Smr.hashemi@shahroodut.ac.ir

چکیده: در سال‌های اخیر استفاده از مدل‌های یادگیری ماشین در حوزه‌های مختلفی مورد استفاده قرارگرفته‌است. از جمله این حوزه‌ها می‌توان به بازار رمزارزها اشاره‌کرد که بخش مهمی از معاملات روزانه در آن توسط ربات‌های معامله‌گر هوشمند انجام‌می‌شود. هدف اصلی‌این مقاله، یافتن بهترین نقطه ورود به یک معامله با کمک روش‌های یادگیری ماشینی است. دراین مقاله از اندیکاتورهای شاخص قدرت نسبی (RSI) و میانگین متحرک ساده (SMA) برای استخراج چندین ویژگی روندی و تلاقی استفاده شده‌است. سپس با استفاده ازاین ویژگی‌ها، مدل‌های یادگیری شبکه عصبی چندلایه، ماشین بردار پشتیبان و نزدیکترین همسایه آموزش‌داده‌می شوند. در نهایت، عملکرد هریک از مدل‌ها بر روی داده‌های بازار ارزهای دیجیتال در بازه هفت ماه نخست سال ۲۰۲۳ برای چندین ارز دیجیتال ارزیابی شده‌است. نتایج نشان‌می‌دهند که اولاً ویژگی‌های استخراج شده از کارایی مناسبی برخوردارند و ثانیاً مدل نزدیکترین همسایه، نسبت به مدل‌های دیگر، بیشترین سود را در این بازه زمانی کسب‌کرده‌است.

واژه‌های کلیدی: بازار رمزارزها، پیش‌بینی بازار، شبکه عصبی مصنوعی، ماشین بردار پشتیبان، نزدیکترین همسایه، تحلیل تکنیکال.

۱. مقدمه

رمزارز[1] شکلی از پول است که به‌صورت دیجیتال یا مجازی ایجادمی‌شود و برای مبادله می‌توان از آن استفاده‌کرد. رمزارزها تمام ویژگی‌های پول‌های رایج را به‌جز ماهیت فیزیکی دارا هستند و از تکنولوژی رمزنگاری برای ایمن‌سازی و تأیید تراکنش‌ها استفاده‌می‌کنند. می‌توان آن‌ها را به‌سادگی مثل تراکنش‌های بانکی و با محدودیت‌های بسیار کمتر انتقال‌داد، با ارزهای دیگر مبادله‌کرد و خرید آنلاین انجام‌داد. اولین ارز دیجیتال بیت‌کوین بود که در سال ۲۰۰۸ توسط یک فرد ناشناس به جهان معرفی شد. در سال‌های اخیر، بازار رمزارزها رشد گسترده‌ای یافته‌است و ضریب نفوذ بالایی را بین افراد مختلف جامعه تجربه‌کرده‌است. علاوه براین صندوق‌های تأمینی و شرکت‌های بزرگ متعددی دارایی‌های مرتبط با ارزهای دیجیتال را در سبد سرمایه‌گذاری و استراتژی‌های معاملاتی خود گنجانده‌اند.

دسترسی راحت به داده‌های معاملاتی همچون قیمت و حجم معاملات در دوره‌های زمانی[2] مختلف، پژوهشگران زیادی را علاقمند به کار دراین زمینه کرده‌است. درواقع، انگیزه اصلی می‌تواند یافتن بهترین موقعیت‌های خرید و فروش و کسب سود مداوم از بازار باشد. به عبارت دیگر، پیش‌بینی دقیق روند حرکتی آینده بازار از مهمترین اهداف پژوهش‌ها دراین زمینه است. اما آیا به‌راستی بازار قابل پیش‌بینی است؟

در مورد قابل پیش‌بینی بودن بازار سهام، نظریه‌های متعددی وجوددارد. فاما در سال ۱۹۷۰، نظریه بازار کارامد[3] را مطرح‌کرد[1]. طبق‌این نظریه، قیمت بازار از یک مدل راهپیمایی تصادفی[4] پیروی‌می‌کند. به عبارت دیگر، تغییرات آتی قیمت بازار را نمی‌توان با استفاده از اطلاعات موجود، پیش‌بینی‌نمود. وجود ناهنجاری‌های متعددی در بازار سرمایه مانند واکنش‌پذیری ضعیف یا گاه بیش از حد، وجود جهش‌های کوتاه‌مدت و معکوس‌شدن بلندمدت بازار و همچنین نوسانات شدید قیمتی، نظریه بازار کارامد را تأییدمی‌کند. در پاسخ به‌این ناهنجاری‌ها، فرضیه بازار تطبیقی[5] مطرح‌شد که روانشناسی انسان و نظریه مالی رفتاری را به‌کارمی‌گیرد. اگرچه ناهنجاری‌ها در بازارهای مالی بوده و خواهندبود، اما نظریه داو را نمی‌توان نادیده‌گرفت. داو معتقدبود که کلیه بازارهای مالی از مجموعه قوانین و الگوهایی تکرارشونده پیروی‌می‌کنند[۲]. به عبارت دیگر، تاریخ تکرار خواهدشد و این باعث‌می‌شود تا اطلاعات تاریخی بازار همواره مورد توجه معامله‌گران قرارگیرد.

به‌طورکلی دو رویکرد تحلیلی در پیش‌بینی قیمت‌ها مورد استفاده قرارمی‌گیرند: تحلیل بنیادی و تحلیل تکنیکال. تحلیل بنیادی بر اطلاعات بنیادی شرکت‌ها مانند درآمدها، هزینه‌ها، طرح‌های توسعه و موقعیت در بازار اشاره‌دارد. اما تحلیل تکنیکال تنها به مطالعه داده‌های تاریخی قیمت و حجم سهام برای پیش‌بینی حرکات قیمتی استوار است.

از طرفی روش‌های پیش‌بینی روند حرکتی آینده بازار را می‌توان به دو دسته روش‌های آماری و روش‌های مبتنی بر یادگیری ماشین تقسیم‌کرد. اکثر مطالعات قدیمی‌تر از روش‌های آماری سری زمانی مبتنی بر داده‌های تاریخی برای پیش‌بینی قیمت استفاده‌کرده‌اند. دراین میان، مدل ناهمگونی شرطی خودرگرسیون[6]، مدل میانگین متحرک رگرسیون خودکار[7]، مدل میانگین متحرک یکپارچه رگرسیون خودکار[8]، مدل میانگین متحرک، فیلتر کالمن و هموارسازی نمایی از محبوبترین تکنیک‌های آماری به‌حساب‌می‌آیند.

بعدها با معرفی هوش مصنوعی، مدل‌های یادگیری ماشین در مطالعات پیش‌بینی بازارها بیشتر مورد توجه قرارگرفتند. زیرا برخلاف روش‌های سری زمانی سنتی، این تکنیک‌ها می‌توانند داده‌های غیرخطی، آشفته، نویزی و پیچیده بازار را مدیریت‌کنند و با دقت و کارایی بیشتری به پیش‌بینی بپردازند.

دراین مقاله، یک روش مبتنی بر تحلیل تکنیکال با استفاده از مدل‌های یادگیری ماشینی به‌منظور تعیین بهترین موقعیت‌های خرید در بازار ارزهای دیجیتال پیشنهادشده‌است. درواقع ابتدا بر اساس اندیکاتورهای RSI و SMA یک‌سری ویژگی‌های جدید استخراج‌می‌شوند و سپس از مدل‌های یادگیری ماشینی نزدیکترین همسایه، ماشین بردار پشتیبان و شبکه عصبی مصنوعی برای یادگیری موقعیت خرید مناسب استفاده‌می‌شود. درنهایت مدل‌های آموزش دیده، بر روی داده‌های واقعی در بازه زمانی هفت ماه نخست ۲۰۲۳ ارزیابی‌شده‌اند. نتایج حاصل نشان‌می‌دهد که ویژگی‌های استخراج‌شده، جهت گرفتن بیشترین سود، کارایی کافی را دارند و مدل نزدیکترین همسایه بهترین عملکرد را دارا می‌باشد.

اولین نوآوری این مقاله روش برچسب‌گذاری داده‌های اولیه است. درواقع متناسب با میزان نوسانی بودن رمزارز می‌توان بر مبنای اندازه سود مورد انتظار در چند دوره زمانی بعدی، داده‌ها را برچسب‌گذاری نمود. دومین نوآوری این است که به جای استفاده مستقیم از اندیکاتورهای شناخته‌شده، رفتارهای روندی و تلاقی آن‌ها با یکدیگر، به‌عنوان ورودی رده‌بندها استفاده‌می‌شود. به عبارت دیگر به جای داده‌ها، استراتژی‌های معاملاتی به الگوریتم‌های رده‌بندی داده‌می‌شود.

در ادامه ساختار مقاله بدین شرح است: در بخش ۲، تاریخچه‌ای مختصر در مورد پژوهش‌ها در زمینه معامله‌گری در بازارهای سهام آورده‌شده‌است. در بخش ۳ روش پیشنهادی به تفصیل شرح‌داده‌می شود. آزمایش‌ها و نتایج ارزیابی در بخش ۴ آمده‌است. در بخش ۵ نتیجه‌گیری و کارهای پیشنهادی بعدی آورده‌شده‌است.

۲. پیشینه تحقیق

هرچند استفاده از یادگیری ماشین برای پیش‌بینی بازار سهام سابقه طولانی دارد و پژوهش‌های زیادی درباره آن موجود است، اما تلاش‌های کمی برای تطبیق سیستم‌های معاملاتی فعلی با بازار ارزهای دیجیتال شده‌است[3]. در برخی پژوهش‌ها، عوامل مؤثر بیرونی بر بازار ارز دیجیتال را بررسی‌کرده‌اند. در [۴] نشان‌دادند که بازار سهام ایالات متحده و شاخص بورس جهانی هیچ تأثیری بر حرکت بیت‌کوین به‌عنوان معتبرترین ارز دیجیتال ندارد. لیو و همکاران در [۵] به این

نتیجه رسیدند که عوامل کلان اقتصادی که به‌طورکلی بر پویایی بازارهای ارز، سهام و کالا مؤثر است، تاثیر قابل‌توجه‌ای بر پویایی بازار ارزهای دیجیتال ندارند. بدین‌ترتیب به‌نظرمی‌رسد که جهت پیش‌بینی در بازار رمزارزها بیشتر باید به عوامل درونی این بازار توجه‌داشت.

در [6-8] نشان‌دادند که می‌توان با داشتن یک استراتژی معاملاتی مبتنی بر تحلیل شبکه های اجتماعی، عملکردی بهتر نسبت به استراتژی خرید و نگهداری کسب‌نمود. نقش حجم معاملات در پیش‌بینی بازار سهام در [9] مورد بررسی قرارگرفت و نشان‌دادند که حجم در پیش‌بینی بازدهی ارز های دیجیتال مفید است، اما در پیش‌بینی نوسانات قیمت‌ها چندان اهمیتی ندارد. در [10] دو مدل رگرسیون یادگیری ماشین پیشرفته با یک شبکه عصبی مصنوعی کاملاً متصل و یک شبکه حافظه کوتاه‌مدت[9] برای پیش‌بینی قیمت ۶ ارز دیجیتال اصلی ساختند. آن‌ها دریافتند که شبکه عصبی مصنوعی به‌طورکلی بهتر از شبکه حافظه کوتاه‌مدت عمل‌می‌کند. علاوه براین آن‌ها نشان‌دادند که پیش‌بینی همزمان داده‌های سری زمانی چند ارزدیجیتال توانایی پیش‌بینی را در مقایسه با آموزش روی هر ارز دیجتیال به‌طور جداگانه بهبود می‌بخشد.

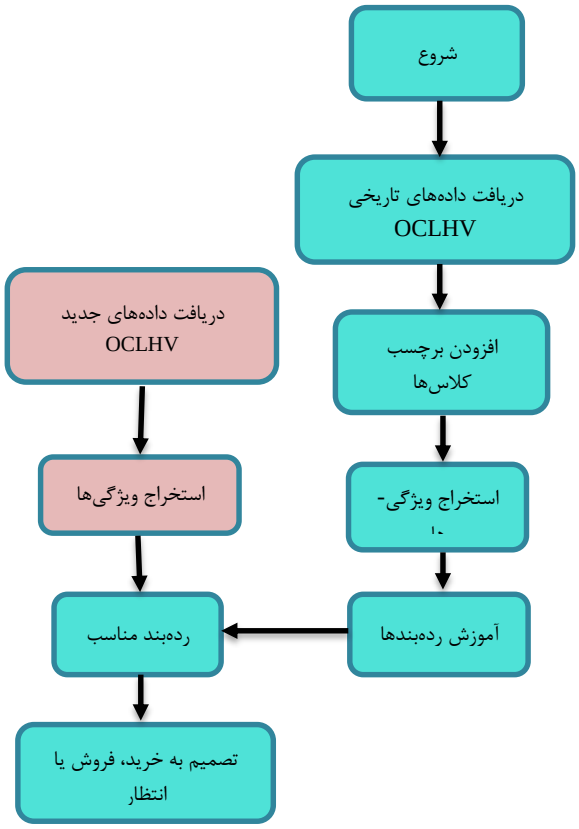
روش‌های یادگیری ماشین متعددی به‌منظور پیش‌بینی رویدادها وجوددارد [۱۱،۱۲]. کوکر و همکاران [۱۳] مدلی را با استفاده از یادگیری تقویتی برای معاملات فعال ۵ ارز دیجیتال اصلی معرفی‌کردند و نشان‌دادند که نسبت به استراتژی خرید و نگهداری بازدهی بهتری دارد. در [۱۴] استفاده از روش‌های یادگیری ماشین را در مقایسه با تکنیک‌های پیش‌بینی سری زمانی در مورد پیش‌بینی قیمت روز بعد چند ارز دیجیتال بررسی‌کردند. هرچند سیگنال‌های خرید و فروش بیشتری در مدل‌های پیش‌بینی سری زمانی‌ایجادشدند، اما بازدهی معاملات در مدل‌های یادگیری ماشین بیشتر بود.

در [۱۵] با استفاده از چندین الگوریتم رده‌بندی، به پیش‌بینی بازار در دوره های زمانی روزانه و دقیقه‌ای پرداختند. آن ها از قیمت های گذشته بازار و اندیکاتورهای تکنیکالی به‌عنوان ورودی رده‌بندهای مختلف استفاده‌نمودند و توانستند دقت پیش‌بینی ۶۵ درصد را به‌دست‌آورند.

غالب این مطالعات نشان‌می‌دهند که داشتن یک استراتژی مبتنی بر یادگیری ماشین منجر به بازدهی بهتری نسبت به استراتژی خرید و نگهداری‌می‌شود. اما تنها از ویژگی‌های شناخته‌شده‌ای همچون اندیکاتورهای تکنیکالی و داده‌های قیمت و حجم استفاده‌نموده‌اند. درحالی‌که در این مقاله به‌جای استفاده مستقیم از مقادیر اندیکاتورها، رفتارهای روندی و تلاقی آن‌ها مورد استفاده قرارگرفته‌است.

۳. روش پیشنهادی

دراین بخش روش پیشنهادی به منظور یافتن بهترین نقطه خرید آورده‌شده‌است. به‌طورکلی هدف استفاده از یادگیری ماشین در کاربردهای مختلف، کشف دانش نهفته در بین داده‌ها ست. یادگیری با نظارت پراستفاده‌ترین تکنیک یادگیری در پیش‌بینی بازارهای مالی است[۱۶]. ازاین‌رو دراین پژوهش نیز از آن استفاده شده‌است. ساختار کلی روش پیشنهادی در شکل ۱ آمده‌است.



شکل ۱: ساختار کلی روش پیشنهادی

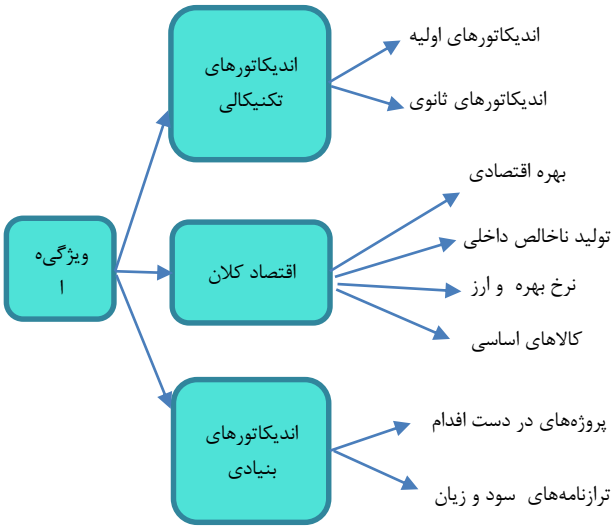
فرایند کلی با کسب داده‌های بازار که معمولاً قیمت باز و بسته شدن، بیشترین و کمترین قیمت و حجم[1] در یک دوره زمانی می‌باشد، آغازمی شود. روش کسب داده و ویژگی‌های اولیه در بخش ۴ آورده شد است. در گام دوم برچسب منا سب کلاس‌ها به داده‌ها اضافه‌می شود. برچسب‌ها شامل "خرید" و "انتظار" می‌باشند. سپس استخراج ویژگی‌های جدید برای تصمیم‌گیری انجام‌می‌شود. گام چهارم به آموزش مدل‌های یادگیری ماشین اختصاص دارد. نهایتاً توانایی مدل‌ها در مواجهه با داده‌های جدید بازار مورد ارزیابی قرارمی‌گیرند. در ادامه نحوه استخراج ویژگی و برچسب‌گذاری داده‌ها و تنظیمات مدل‌های رده‌بندی با جزئیات بیشتر آورده‌شده‌است.

۱،۳. استخراج ویژگی‌های جدید

در پژوهش‌های مختلفی که بین سال‌های ۲۰۰۸ تا ۲۰۲۱ در مورد بازارهای سهام انجام‌شده‌است، تعداد ۲۱۷۳ ویژگی منحصربه‌فرد در فرایندهای یادگیری ماشین مورد استفاده قرارگرفته‌اند [۱۷]. این ویژگی‌ها به‌طورکلی به ۳ گروه و چندین زیر گروه تقسیم‌می‌شوند (شکل ۲).

گروه اول، اندیکاتورهای تکنیکالی می باشند که پرکاربردترین ویژگی‌های مورد استفاده در پژوهش‌های پیش‌بینی بازار سرمایه

هســـتند. این ویژگی ها به دو زیر گروه اندیکاتورهای اولیه شـــامل OCLHV[10] و اندیکاتورهای ثانویه شامل تمامی‌اندیکاتورهای شناخته شـــده در بازار ســـرمایه مانند انواع میانگین‌های تحرک، RSI و صـــدها اندیکاتور شناخته‌شده دیگر تقسیم‌می‌شوند.



شکل ۲: انواع ویژگی‌های مورد استفاده در پژوهش‌ها

گروه دوم ویژگی ها در شـــکل ۲ داده های اقتصـــاد کلان را دربرمی‌گیرند. برخی از زیرگروه‌های اقتصاد کلان شامل بهره اقتصادی، تولید ناخالص داخلی، نرخ بهره، قیمت ارز و کالاهای اساسی می‌باشند. گروه سوم، شامل ویژگیهای بنیادی همچون پروژه‌های در دست اقدام، و ترازنامه‌های سود و زیان می‌باشند.

دراین مقاله از اندیکاتور های RSI و SMA برای اســـتخراج یک سری ویژگی جدید استفاده شده است. اندیکاتور RSI نشان‌دهنده قدرت بازار و قدرت خریداران و فروشـــندگان نســـبت به هم اســـت. این اندیکاتور همواره بین صـــفر تا ۱۰۰ در نوســـان اســـت و در خانواده اســـیلاتورها قراردارد. نحوه محاســـبه آن در رابطه (۱) و رابطه (۲) آورده‌شده‌است.

$$RS(n) = \frac{MG(n)}{ML(n)} \quad (۱)$$

$$RSI = 100 - \left[\frac{100}{1 - RS}\right] \quad (۲)$$

در رابطه (۱)، $MG(n)$ میانگین میزان سود در دوره زمانی به طول n است. دراین مقاله از چهار دوره زمانی متداول شامل ۵، ۹، ۱۴ و ۲۱ جهت استخراج ویژگی استفاده‌شده‌است.

اندیکاتور SMA ســـاده‌ترین اندیکاتور در خانواده میانگین‌های متحرک محســـوب‌می‌شـــود. برای محـــاســـبـــه آن از رابطـه (۳) می‌توان‌استفاده‌کرد.

$$SMA(n) = \frac{\sum_1^n C}{n} \quad (۳)$$

رابطه (۳) یک میانگین‌گیری ساده را بر روی قیمت‌های بسته‌شدن در هر کندل قیمتی در یک دوره زمانی به طول n نشان‌می‌دهد. دراین مقاله از دوره‌های زمانی به طول ۹، ۱۴، ۲۱، ۵۰ و ۱۰۰ اســـتفاده‌شـــده است.

دو اندیکاتور فوق را چنانچه از نظر روندی و تلاقی بررســـی‌کنیم، چهار دســـته ویژگی به‌دســـت‌می‌آیند. ویژگی های روندی SMA نشـــان‌می‌دهند که SMA های مختلف در دوره‌های زمانی ۹، ۱۴، ۲۱، ۵۰، ۱۰۰ در طول یک پنجره زمانی دارای چه روندی بوده‌اند. یعنی مثلاً مشـــخص‌شـــود که در پنجره زمانی ۱۲ کندلی، روند SMA(9) افزایشی، کاهشی یا بدون تغییر بوده‌است. ویژگی‌های تلاقی SMA نیز نشـــان‌می‌دهند که SMA(9)، SMA(14) را در طول یک پنجره زمانی رو به بالا قطع‌کرده‌اســـت یا رو به پایین یا بدون تلاقی بوده‌ا ست. ایـــن تـــلاقـــی بـــرای SMA(14) بـــا SAM(21)، SMA(21) بـــا SMA(50) و SMA(50) با SMA(100) نیز بررســـی‌می‌شـــوند. بنابراین ۹ ویژگی روندی و تلاقی برای SMA تشکیل‌می‌شود که وقوع هر کدام ازاین موارد برای یک معامله‌گر خبره در بازارهای ســـرمایه با معنا ا ست. به‌عنوان مثال چنانچه SMA(50)، SMA(100) را رو به بالا قطع کند، می‌تواند نشانه رشد قیمتی در آینده باشد.

ویژگی‌های روندی RSI نشـــان‌می‌دهند که RSI های مختلف در دوره‌های زمانی 5، 9، 14 و 21 در طول یک پنجره زمانی، دارای چه روندی هســـتند. و ویژگی های تلاقی RSI نیز نشـــان‌می‌دهند که RSI(5)، RSI(9) را در طول پنجره زمانی رو به بالا قطع‌کرده‌اســـت یا رو به پایین یا بدون تلاقی بوده‌اســـت. این محاســـبات برای تلاقی RSI(9) با RSI(14) و RSI(14) با RSI(21) نیز صورت‌می‌پذیرد. بنابراین ۷ ویژگی روندی و تلاقی برای RSI تشـــکیل‌می‌شـــود. نهایتاً با اســـتفـــاده از SMA و RSI، ۱۶ ویژگی رونـــدی و تلاقی جـــدیـــد ایجادشـــده‌است که در فرایند آموزش مدل‌های یادگیری ماشـــین مورد استفاده قرارمی‌گیرند.

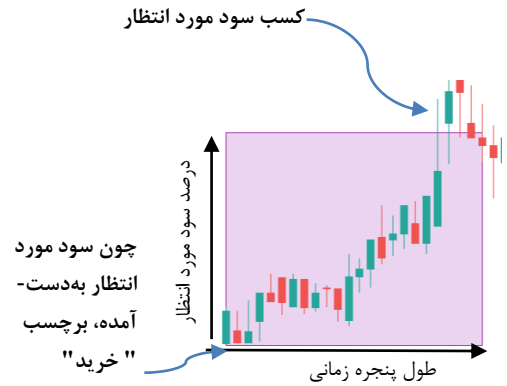
۲،۳. برچسب گذاری داده ها برای آموزش مدل‌های رده‌بند

در ابتدا نیاز اســـت تا داده‌های دریافتی برچســـب‌گذاری شـــوند. روند برچســـب‌گذاری داده‌ها یکی از نوآوری‌های این پژوهش اســـت. برای برچسب‌گذاری ابتدا باید مقدار دو پارامتر مشخص‌شوند. پارامتر نخست میزان درصـــد ســـود مورد انتظار(EP) و پارامتر دوم، طول پنجره زمانی اســـت که معامله‌گر می‌خواهد به ســـود مورد انتظار دســـت‌یابد(WL). به‌عنوان مثال چنانچه WL=24، EP=3% و دوره زمانی نیز ۶۰ دقیقه باشد، بدین معنی است که داده‌ها به‌نحوی برچسب‌گذاری شوند تا بهترین زمان‌های خرید برای کسب سه درصد سود در ۲۴ ساعت را نشان‌دهند.

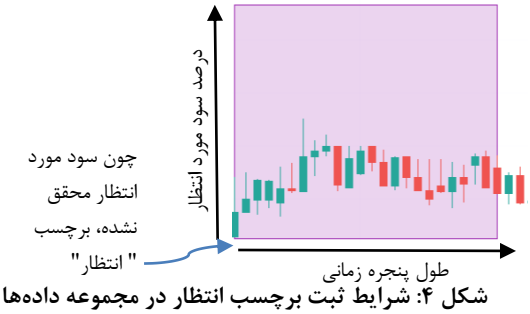
اگر چنین شـــرایطی رخ‌دهد، برچســـب "خرید"، برای زمان ابتدای پنجره لحاظ خواهد شد. شکل 3 این شرایط را نشان‌می‌دهد. اما اگر در طول یک پنجره زمانی، چنین رشد مورد انتظاری حاصل‌نشود، برچسب "انتظار" به معنای عدم خرید ثبت‌می‌شود (شکل ۴).

با برچســـب‌گذاری و اســـتخراج ویژگی‌ها، همه چیز برای آموزش مدل‌های یادگیری مهیا اســـت. دراین مقاله از ســـه مدل شـــبکه عصـــبی

مصـــنوعی، ماشـــین بردار پشـــتیبان و k-نزدیکترین همســـایه استفاده‌شده‌است. تنظیمات هر مدل در جدول ۱ آورده‌شده‌است.



شکل ۳: شرایط ثبت برچسب خرید در مجموعه داده‌ها



شکل ۴: شرایط ثبت برچسب انتظار در مجموعه داده‌ها

۴. آزمایش‌ها

دراین بخش نحوه دریافت داده‌ها و خصـــوصـــیات آن‌ها، آزمایش‌های مختلف جهت ارزیابی مدل‌های رده‌بندی و نهایتاً مقایســـه آن‌ها از نظر سوددهی با استراتژی خرید و نگهداری آورده‌شده‌است.

۴ ، ۱. دریافت داده‌ها

ابتدایی‌ترین داده‌های معاملاتی که در مورد یک نماد در هر بازاری از جمله بازار ارزهای دیجیتال قابل‌دریافت است، OCLHV می‌باشند. این داده‌ها مشخص‌می‌کنند که در یک دوره زمانی خاص مثلاً یک ساعته، قیمت اولین خرید (O)، قیمت آخرین خرید (C)، کمترین قیمت (L)، بیشـــترین قیمت (H) و کل حجم معامله‌شـــده (V) چقدر بوده‌اســـت. داده‌های مربوطه دراین مقاله از طریق API های موجود در ســـایت صرافی نوبیتکس[11] و با زبان برنامه‌نویسی پایتون دانلودشده‌اند. به‌منظور دانلود داده‌ها باید نام رمزارز، تایم فریم مورد نظر، تاریخ شـــروع و تاریخ پایان برای API مشـــخص باشـــد. به‌منظور برچســـب‌گذاری و آموزش مدل‌های رده‌بندی، داده‌های بازار از تاریخ ۲۰۲۱/۱/۱ تا ۲۰۲۲/۱۲/۳۰ در دوره زمانی یک ساعته برای نمادهای بیت‌کوین، اتریوم، ریپل، ترون و تتر دانلودشده‌اند. ۵ ردیف از نمونه داده‌های اولیه برای نماد بیت‌کوین در جدول ۲ آورده‌شده‌است.

پس از دانلود داده‌های اولیه، برچسب‌گذاری داده‌ها طبق روشی که شرح آن در بخش **۲,۳** آمد، انجام‌می‌شود. ۵ ردیف داده برچسب‌گذاری شده برای بیت‌کوین در جدول ۳ آورده‌شده‌است.

مرحله بعدی ا ستخراج ویژگی‌های روندی و تلاقی ا ست که مطابق با روش شـــرح‌داده‌شـــده در بخش ۱,۳ بر اســـاس اندیکاتورهای RSI و SMA و بـــا اســـتفـــاده از کتـــابخـــانـــه pandas_ta در پـــایتون محاسبه‌می‌شوند. (جدول ۴)

اکنون ویژگی‌ها و برچ سب کلاس‌ها، آماده ا ست و با تق سیم‌بندی داده‌ها به مجموعه آموزشـــی و آزمون، مدل‌های رده‌بندی مورد نظر آموزش داده‌می‌شوند. دقت هر مدل در جدول ۵ آورده‌شده‌است.

لازم به ذکر اســـت که، طول پنجره زمانی و ســـود مورد انتظار نمی‌تواند برای همه رمزارزها یکســـان باشـــد. مثلاً اگر برای بیت‌کوین، طول پنجره ۱۲ و درصد ر شد مورد انتظار ۳٪ تنظیم شود، مشکل عدم توازن داده پیش‌خواهدآمد و تعداد برچ سب‌های "خرید" در مقای سه با تعداد برچســـب "انتظار" بســـیار اندک خواهدبود. زیرا اصـــولاً رمزارز بیت‌کوین آنچنان نوســـان زیادی ندارد که بارها و بارها در طول ۱۲ ساعت سه در صد ر شد قیمتی را تجربه‌کند. بنابراین باید با یک روش آزمون و خطا، طول پنجره و درصد رشد مناسب را تعیین‌کرد طوری که بین ۴۰ تا ۶۰ درصد داده‌ها، برچسب خرید داشته‌باشند.

جدول ۱: پارامترهای مدل‌های یادگیری

نام مدل	پارامترها	مقادیر
نزدیکترین همسایه	Weigth	Uniform
	K	5
	Algorithm	Ball_tree
شبکه عصبی چندلایه	Hidden Layer Size	(3,5)
	Activation Function	Tanh
	Solver	lbfgs
	Learning Rate	constant, 2e-5
ماشین بردار پشتیبان	Kernel	Poly
	Degree	4
	C	1

جدول ۲: پنج ردیف از دادهای OCLHV برای بیت‌کوین

Date	Open	high	low	Close	Volume
2021-01-01 00:30:00	768700000	773995986	765501000	773500000	3.21494246
2021-01-01 01:30:00	773921000	774999900	770788301	773314200	4.768125762
2021-01-01 02:30:00	773314200	775000000	770000000	773998999	1.382566634
2021-01-01 03:30:00	773998999	773998999	770000000	770599001	2.713378
2021-01-01 04:30:00	770598002	771499000	767550000	770998000	1.895207556

جدول ۳: پنج ردیف از داده‌های برچسب‌گذاری‌شده برای بیت‌کوین

Date	Open	high	low	Close	Volume	Label
2021-01-01 00:30:00	768700000	773995986	765501000	773500000	3.21494246	W
2021-01-01 01:30:00	773921000	774999900	770788301	773314200	4.768125762	W
2021-01-01 02:30:00	773314200	775000000	770000000	773998999	1.382566634	W
2021-01-01 03:30:00	773998999	773998999	770000000	770599001	2.713378	B
2021-01-01 04:30:00	770598002	771499000	767550000	770998000	1.895207556	B

جدول ۴: پنج ردیف از ویژگی‌های استخراج‌شده برای بیت‌کوین

ویژگی استخراج‌شده	2021-01-01 00:30:00	2021-01-01 01:30:00	2021-01-01 02:30:00	2021-01-01 03:30:00	2021-01-01 04:30:00
SMA_9_Trend	Down	Down	Up	Up	Down
SMA_14_Trend	Down	Down	Down	Up	Down
SMA_21_Trend	Down	Down	Down	Down	Down
SMA_50_Trend	Up	Down	Down	Down	Up
SMA_100_Trend	Down	Down	Up	Down	Down
SMA_9_cross_14	No	No	Up	Down	Down
_cross_21۱۴SMA_	Up	No	Up	No	Up
_cross_50۲۱SMA_	No	No	No	Up	Up
_cross_100۵۰SMA_	No	No	No	No	No
RSI_5_Trend	Down	Down	Down	Down	Down
RSI_9_Trend	Down	Down	Down	Down	Down
RSI_14_Trend	Up	Up	Up	Up	Up
RSI_21_Trend	Up	Up	Up	Up	Up
RSI_5_cross_9	Down	Down	Down	Down	Down
_cross_14۹RSI_	Down	Down	Down	Down	Down
_cross_21۱۴RSI_	Up	Up	Up	Up	Up
Label	w	w	w	B	B

جدول ۵: دقت مدل‌های رده‌بندی

	نزدیکترین همسایه	ماشین بردار پشتیبان	شبکه عصبی
دقت رده بندی	62%	51%	56%

۲،۴. ارزیابی

پس از آموزش مدل‌های یادگیری، بر روی داده‌های جدید شامل هفت ماه نخست سال ۲۰۲۳، بک تست انجام‌می‌شود. عملکرد مدل‌ها با یکدیگر و با استراتژی خرید و نگهداری مقایسه شده‌است. منظور از استراتژی خرید و نگهداری این است که اگر رمزارز مورد نظر در ۲۰۲۳/۱/۱ خریداری‌شود و تا ۲۰۲۳/۰۷/۳۱ نگهداری‌گردد چه میزان سود یا ضرر کسب‌شده‌است. نتایج در جدول ۶ آورده‌شده‌است.

با توجه به جدول ۶، روش نزدیکترین همسایه بهترین عملکرد را بین تمامی روش‌های دیگر داشته‌است. علاوه براین، تمامی مدل‌های یادگیری عملکرد بهتری نسبت به استراتژی خرید و نگهداری داشته‌اند که می‌تواند نشان‌دهنده انتخاب مناسب ویژگی‌های استخراج‌شده باشد.

جدول ۶: میزان درصد سود یا زیان کسب‌شده برای ۵ رمزارز مختلف

	نزدیکترین همسایه	ماشین بردار پشتیبان	شبکه عصبی	خرید و نگهداری
بیتکوین	186%	131%	148%	116%
اتریوم	146%	102%	109%	88%
ترون	128%	95%	108%	75%
ریپل	156%	161%	133%	151%
تتر	76%	31%	42%	23%

لازم به ذکر است که چون در صرافی‌های ایران، امکان‌ایجاد معاملات در جهت فروش فراهم نیست، فروش هر معامله بازشده بر اساس تنظیم حد ضرر ۳ درصد و حد سود ۴ درصد انجام‌شده‌است. با رسیدن به حد سود موردنظر، مجدد مقادیر حد سود و حد ضرر به‌روزرسانی می شوند تا زمانی که حد ضرر فعال شود و فروش رمزارز انجام‌گیرد. در محاسبات درصد سود نهایی جدول ۶، ۰٫۳ درصد کارمزدهای انجام معامله در صرافی نیز لحاظ‌شده‌است.

۵. نتیجه‌گیری و کارهای بعدی

هدف ازاین پژوهش، یافتن بهترین نقاط ورود به یک معامله ارز دیجیتال به کمک مدل‌های یادگیری ماشین است. در پژوهش‌های انجام شده دراین حوزه، غالباً از اندیکاتورهای تکنیکالی به‌طور مستقیم در پیش‌بینی بازار استفاده‌می‌شود. اما در روش پیشنهادی این مقاله، به‌جای استفاده مستقیم از مقادیر اندیکاتورها به‌عنوان ورودی روش‌های رده‌بندی، از رفتارهای روندی و تلاقی آن‌ها استفاده شده‌است. درواقع روندهای افزایشی یا کاهشی و نحوه تلاقی این اندیکاتورها، برای برخی معامله‌گران دراین حوزه به‌عنوان سیگنال‌های ورود و خروج از معامله تعبیرمی‌شوند. بنابراین می‌توان‌گفت که استراتژی‌های معاملاتی به‌عنوان ورودی رده‌بندها درنظرگرفته‌شده‌است.

علاوه براین دراین مقاله روشی برای برچسب‌گذاری داده‌های اولیه پیشنهاد شده‌است که به‌واسطه آن می‌توان بهترین نقاط ورودی از نظر میزان سود مورد انتظار را به‌طور مناسب برچسب‌گذاری نمود. نتایج حاصل بر روی داده‌های هفت ماه نخست سال ۲۰۲۳ نشان‌می‌دهند که ویژگی‌های استخراجی کارایی لازم برای تشخیص بهترین نقاط ورود را داشته‌اند و روش نزدیکترین همسایه بیشترین سوددهی دراین بازه زمانی را کسب‌کرده‌است. در ادامه این پژوهش، تعیین بهترین نقاط خروج از معامله با کمک یادگیری ماشین و استفاده از اندیکاتورهای حجمی‌در فرایند یادگیری، قابل‌بررسی است.

References

[1] E. Fama. “Efficient capital markets: A review of the theory”. *The Journal of Finance, 25*, 383–417, 1970

[2] J. Almeida, T. Cruz,”A systematic literature review of investor behavior in the cryptocurrency markets”, Journal of Behavioral and Experimental Finance, Volume 37, 2023

[3] V. Derbentsev, V. Babenko, K. Khrustalev, H. Obruch and S. Khrustalova, “Comparative performance of machine learning ensemble algorithms for forecasting cryptocurrency prices”. *International Journal of Engineering, 34*, 140–148, 2021

[4] C. Conrad, A. Custovic and E. Ghysels, “Long-and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis”. *Journal of Risk and Financial Management, 11*(2), 23, 2018

[5] Y. Liu and A. Tsyvinski, “Risks and returns of cryptocurrency”. *The Review of Financial Studies, 34*(6), 2689–2727, 2021

[6] R. C. Phillips and D. Gorse, “Predicting cryptocurrency price bubbles using social media data and epidemic modelling”. In 2017 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1-7). IEEE, 2017

[7] F. Valencia, A. Gomez-Espinosa and B. Valdes-Aguirre, “Price movement prediction of cryptocurrencies using sentiment analysis and machine learning”. Entropy, 21(6), 589, 2019

[8] A. Yavari, H. Hassanpour, B. Rahimpour and M. Mahdavi, “Event prediction in social network through Twitter messages analysis. Social Network Analysis and Mining”. 12. 10.1007/s13278-022-00911-x. 2022

[9] E. Bouri, C. K. Lau, B. Lucey and D. Roubaud, “Trading volume and the predictability of return and volatility in the cryptocurrency market”. *Finance Research Letters, 29*, 340–346, 2019

[10] W. Zhengyang, L. Xingzhou, R. Jinjin and K. Jiaqing, “Prediction of cryptocurrency price dynamics with multiple machine learning techniques”. *In Proceedings of the 2019 4th International Conference on Machine Learning Technologies* (pp. 15–19), 2019

[11] Hajighorbani, Hashemi, Broumandnia, faridpour, “a review of some semi-supervised learning methods”, Journal of Knowledge-Based Engineering and Information”, 2(4), 2013

[12] S. Ansaripour, T. J. Gandomani, “Comparison of the classification methods in software development effort estimation”, Intelligent Multimedia Processing and Communication Systems(IMPCS), 3(2), 11-18, 2022

[13] T. E. Koker and D. Koutmos, “Cryptocurrency trading using machine learning”. *Journal of Risk and Financial Management, 13*(8), 178, 2020

[14] G. Attanasio, L. Cagliero, P. Garza and E. Baralis, “Quantitative cryptocurrency trading: Exploring the use of machine learning techniques”. In *In Proceedings of the 5th*

Workshop on Data Science for Macro-Modeling with Financial and Economic Datasets (pp. 1–6). 2019.
[15] E. Akyildirim, A. Goncu and A. Sensoy, "Prediction of cryptocurrency returns using machine learning". *Annals of Operations Research, 297*(1), 3–36, 2021
[16] Z. Dongna, H. Xihui, L. Chi Keung and X. Bing, "Implications of cryptocurrency energy usage on climate change", Technological Forecasting and Social Change", Volume 187, 2023,
[17] M. Kumbure, C. Lohrmann, P. Luukka and J. Porras, "Machine learning techniques and data for stock market forecasting: A literature review", Expert Systems with Applications", Volume 197, 2022

پی‌نوشت

1- Cryptocurrency
2- Time Frame
3- Efficient Market Hypothesis
4- Random Walk
5- Adaptive Market Hypothesis
6- Auto-Regressive Conditional Heteroscedasticity
7- Auto-Regressive Moving Average
8- Auto-Regressive Integrated Moving Average
9- Long-Short-Term-Memory
10- Open-Close-Low-High-Volume (OCLHV)
11- https://api.nobitex.ir/