Archive of SID

فصلنامه اقتصاد کاربردی
دوره ۱۰، شماره ۳۴ و ۳۵، پاییز و
زمستان ۱۳۹۹

پیش‌بینی قیمت بیت‌کوین با استفاده از الگوریتم‌های یادگیری ماشین

میثم بشیری [۱]، سید حسین پاریاب [۲]

تاریخ دریافت: ۹۹/۰۸/۱۵ تاریخ پذیرش: ۱۴۰۰/۰۵/۱۰

چکیده:

بیت‌کوین معروف‌ترین رمز ارز است که از فناوری زنجیره بلوکی استفاده می‌کند. در این پژوهش، مجموعه داده‌های مربوط به ده رمزارز مورد استفاده قرار گرفته و یک مجموعه داده جدید، با در نظر گرفتن قیمت نهایی هر رمز ارز و برای دست‌یابی به هدف تحقیق و تعیین این‌که چگونه جهت و صحت قیمت بیت‌کوین را می‌توان با استفاده از تکنیک‌های داده کاوی پیش‌بینی کرد، تشکیل شده است. مهندسی ویژگی مشخص کرد که هر ده رمز ارز به شدت با یکدیگر ارتباط دارند. این کار با اجرای روش یادگیری نظارت شده انجام شده است که در آن از جنگل تصادفی، طبقه‌بندی بردار پشتیبان، گرادیان تقویتی، و شبکه عصبی در گروه طبقه‌بندی و از رگرسیون خطی، شبکه عصبی بازگشتی و رگرسیون گرادیان تقویتی استفاده شده است. در این پژوهش الگوریتم‌های ماشین بردار پشتیبان، جنگل تصادفی، گرادیان تقویتی و شبکه عصبی مقدار صحت ۵۲،۱۶۷۵ درصد را ثبت کردند.

کلید واژه: پیش‌بینی، رمزارز، بیت‌کوین، یادگیری ماشین.
طبقه‌بندی JEL: C50، E42، G13.

[۱] کارشناسی ارشد مدیریت اجرایی، پژوهشگر موسسه مطالعات و پژوهش‌های بارگانی (نویسنده مسئول). ایمیل: Maysam.Bashiri@gmail.com

[۲] دانشجوی دکترا تجارت الکترونیکی، پژوهشگر موسسه مطالعات و پژوهش‌های بارگانی. ایمیل: h.paryab@gmail.com

www.SID.ir

Archive of SID

مقدمه

امروزه رفاه جوامع مدرن بستگی به وضعیت اقتصادی بازار دارد. در قلب هر اقتصاد بازار، بازارهای مالی با تعادل میان عرضه و تقاضا قرار دارد. بنابراین مطالعه بازار و یادگیری در مورد حرکات آن بسیار مهم است. درک حرکات و تحولات بازار در درجه اول توانایی پیش‌بینی تحولات آینده را تسهیل می‌کند. توانایی پیش‌بینی وضعیت اقتصادی بازار برابر با توانایی تولید ثروت با جلوگیری از ضرر و زیان مالی و ایجاد سود مالی است. هر چند ماهیت بازارها به‌گونه‌ای است که پیش‌بینی آن‌ها دشوار است، لذا پژوهش‌های بسیاری در زمینه پیش‌بینی بازار با استفاده از روش‌های مختلف صورت گرفته است. پیش‌بینی بازار سهام و ارزش رمز ارزها جز مواردی است که برای کل صنعت از اهمیت فوق العاده‌ای برخوردار است. قیمت سهام با رفتار سرمایه‌گذاران انسانی تعیین می‌شود و سرمایه‌گذاران با استفاده از اطلاعات موجود در دسترس عموم، قیمت سهام و رمز ارزها را برای پیش‌بینی نحوه عملکرد یا واکنش بازار تعیین می‌کنند. فاکتورهای زیادی برای با اهمیت جلوه دادن پیش‌بینی بازار سهام و رمز ارزها وجود دارد اما شاید بتوان گفت که به دلیل ماهیت گذرا بودن روندها و سرعت بالای معاملات و تغییرات در روندها داشتن پیش‌بینی نزدیک به واقعیت می‌تواند کمک زیادی به بالا بردن سود مادی شرکت‌های سرمایه‌گذاری و البته سرمایه‌گذارانی که به‌صورت شخصی وارد بازار می‌شوند کرد. اما در این میان نظرات مخالفی وجود دارد مبنی بر این‌که بازار سهام و رمز ارزها قابل پیش‌بینی نیست و دلیل آن نیز تعدد پارامترهای تاثیرگذار است. اما روش‌های نوین یادگیری ماشین مانند استفاده از شبکه‌های عصبی و داده کاوی در سال‌های اخیر کمک زیادی به حل این مشکل کرده است. ضمنا با توجه به در نظر گرفتن این مساله مهم که تمام پارامترهای تاثیرگذار در یک پارامتر به نام قیمت خلاصه می‌شوند می‌توان تا حد زیادی مساله را ساده‌سازی کرد (آقاخانی، کریمی، ۱۳۹۳).

بیت‌کوین یک نسخه کاملا همتا به همتا از پول نقد الکترونیک است که به وسیله آن پرداخت‌های آنلاین میسر می‌شود. بیت‌کوین مستقیماً بین طرفین معامله انتقال می‌یابد و تراکنش‌های آن نیازمند گذر از یک موسسه مالی (نهاد مرکزی واسط) نیست (ناکاموتو[1]، ۲۰۰۸). بیت‌کوین معروف‌ترین رمز ارز است که از فناوری زنجیره بلوکی استفاده می‌کند. این واحد پولی نوپا در سال ۲۰۰۹ توسط ساتوشی ناکاموتو[2] ابداع شد که به نظر می‌رسد نامی مستعار باشد. البته این امکان هم وجود دارد که ایشان عضو گروهی بزرگ‌تر از مبتکران ایده بیت‌کوین باشند. در حال حاضر بیت‌کوین با مبادلات دنیای واقعی تطابق و سازگاری بیشتری پیدا کرده است و بعضاً معاملات مهمی مثل خرید خانه یا ماشین نیز با بیت‌کوین انجام می‌شود. حجم معاملات با بیت‌کوین در آغاز سال ۲۰۱۹ بسیار بالا بوده و شروعی رویایی داشته است. این مسئله باعث تحیر افراد فعال در بازار سرمایه شده است. به همین دلیل تاجران و سرمایه‌گذاران تلاش می‌کنند تا بیشترین سود را از این وضعیت ببرند. هر چه حجم مبادلات بیت‌کوین افزایش پیدا کند، بالطبع قیمت آن نیز افزایش پیدا خواهد کرد. علاوه بر این، تعداد بیت‌کوین بسیار محدود بوده و تنها ۲۱ میلیون بیت‌کوین را می‌توان استخراج کرد (ناکاموتو، ۲۰۰۸). این در حالی است که طبق آمار، تا تاریخ ۱۰ می ۲۰۲۰ تعداد ۱۸،۳۷۲،۱۲۵ بیت‌کوین استخراج شده (کوین مارکت کپ[3]، ۲۰۲۰) و تنها ۲،۶۲۷،۸۷۵ بیت‌کوین دیگر را می‌توان استخراج کرد و این یعنی احتمالاً بعد از استخراج این مقدار، ارزش بیت‌کوین افزایش پیدا خواهد کرد. بیت کوین به عنوان یک رمز ارز، فاقد شکل فیزیکی است به طوری که انجام تحلیل آن به نحوی غیرممکن شده است. در نتیجه، بسیاری از سرمایه‌گذاران، پیوسته با استفاده از شاخص‌های تجاری تحلیل فنی[4] (الگوهای هندسی ایجاد شده از قیمت‌های پیشین و حجم معاملات) برای درک و پیش‌بینی روند آتی قیمت بیت‌کوین تلاش می‌کنند. هدف از این تحقیق تعیین چگونگی پیش‌بینی صحت و دقت قیمت بیت‌کوین با استفاده از الگوریتم‌های یادگیری ماشین است (مک نالی و همکاران[5]، ۲۰۱۸). علاوه بر این، بیت کوین به طرز شگفت‌آوری غیرقابل پیش‌بینی‌تر از استانداردهای مختلف پولی مانند دلار امریکا است. هم‌چنین، بیت کوین به‌عنوان یک رمز ارز اصلی، پیش‌رو بوده که انتظار می‌رود که رشد بیشتری داشته باشد و این امر الهام بخش تحقیقات در این زمینه است. پیش‌بینی قیمت بیت کوین، یک فرصت را برای کسب سود از طریق خرید و فروش رمز ارز ایجاد می‌کند.

قیمت بیت‌کوین بر خلاف بازار سهام، به وقایع تجاری یا مداخله دولت بستگی ندارد. پیش بینی ارزش بیت‌کوین از

www.SID.ir

Archive of SID

جنس مسائل چند متغیره است (مک‌نالی و همکاران، ۲۰۱۸). روش‌های قدیمی پیش‌بینی رمزارزها، مانند مدل‌های هموارسازی نمایی هولت وینترس[۶]، به فرضیات خطی متکی هستند و برای موثر بودن، نیازمند داده‌هایی هستندکه بتوان آنها را به روند، فصلی و نویز[۷] تقسیم کرد. این روش برای کارهایی مانند پیش‌بینی فروش محصولات، مناسب است که اثرات فصلی در آن وجود دارد. به دلیل فقدان چنین ویژگی در بازار بیت‌کوین(یا حداقل ندرت آن) و نوسانات زیاد قیمت بیت‌کوین، این روش‌ها برای این کار چندان مؤثر نیستند. با توجه به پیچیدگی کار، داده کاوی، راه‌حلی جالب و فناورانه را ایجاد می‌کند (چتفیلد و ام.یار[۸]، ۱۹۸۸).

مروری بر کارهای پیشین

در عصر انفجار اطلاعات شرکت‌های فردی هر روزه حجم زیادی از داده را تولید و جمع‌آوری خواهندکرد. استخراج اطلاعات مفید از پایگاه داده و تبدیل کردن اطلاعات به نتایج عملی چالش اصلی است که شرکت‌ها با آن روبرو هستند. با توجه به پیشرفت کشور در زمینه فناوری اطلاعات و نگاه‌های ویژه به دولت الکترونیکی و نفوذ استفاده از سیستم‌های رایانه‌ای در صنعت و ایجاد بانک‌های اطلاعاتی بزرگ توسط ادارات دولتی، در بانک‌ها و بخش‌های خصوصی نیاز به استفاده از داده کاوی به طور عمیق احساس می‌شود.

بیت‌کوین یک رمزارز و نظام پرداخت جهانی با کارکردهای مشابه پول بی‌پشتوانه است، ولی از نظر حقوقی هنوز هیچ کشوری آن را به عنوان پول قانونی به رسمیت نشناخته‌است. از نظر فنی بیت‌کوین نخستین پول دیجیتال غیرمتمرکز است. چرا که بدون بانک مرکزی یا مسئول مرکزی کار می‌کند. این شبکه همتا به همتا است و تراکنش‌ها، مستقیماً و بدون واسطه بین کاربران انجام می‌شود (شیرزور علی‌ابادی، رمضان‌زاده، ۱۳۹۹). ارزش بیت کوین درست مانند سهام، البته با برخی تفاوت‌ها نسبت به آن است. در حال حاضر تعدادی الگوریتم در بازار بورس برای پیش‌بینی قیمت‌ها استفاده می‌شود. با این حال، مولفه‌های مؤثر بر قیمت بیت‌کوین متفاوت هستند. بنابراین لازم است ارزش بیت‌کوین پیش‌بینی گردد تا تصمیمات صحیح برای سرمایه‌گذاری بر روی آن اتخاذ شود (ولانکار و همکاران، ۲۰۱۸).

پیش‌بینی قیمت رمزارزها از اهمیت خاصی برای معامله‌گران برخوردار است تا بیشترین سود را کسب کنند؛ لذا همواره به دنبال راهکارهای منطقی و دقیق جهت پیش‌بینی بوده‌اند. تکنیک‌های داده کاوی علاوه بر جمع‌آوری و مدیریت داده‌ها، آنالیز و پیش‌بینی‌هایی را نیز شامل می‌شود و از طریق کشف الگوهای موجود و روابط ناشناخته میان داده‌ها به معامله‌گران در امر پیش‌بینی یاری می‌رساند. اگر چه در سال‌های اخیر بیشتر بازارهای سرمایه جهان به سمت کارایی پیش می‌روند، اما اعتقاد به پیش‌بینی رفتار بازار کاهش نیافته است. فرضیه سیستم پیچیده و سازگارشونده که به بیان واقعیات می‌پردازد و با مفروضات ساده واقع‌بینانه و با کنار گذاشتن قطعیت و قطعیت‌گرایی و جایگزینی تفکر غیرخطی، بجای نگرش رابطه علت و معلولی و ایجاد مدل‌ها و راه‌حل‌های غیرخطی و سیستم‌های یادگیرنده، به بازار می‌نگرد و حتی آن را قابل پیش‌بینی می‌داند. ابزارهایی در این محیط غیرخطی و پویای پرآشوب، متناسب خواهند بود که برای چنین محیط‌هایی طراحی شده و یا همانند انسان در چنین محیط‌هایی تصمیم‌گیری و عمل نمایند. هوش مصنوعی و یادگیری ماشین یکی از این ابزارهایی است که از مغز انسان الگوبرداری شده و قابلیت تشخیص الگوهای پیچیده، یادگیری و پیش‌بینی را دارد. به این دلیل فرضیه «بازار سهام با توجه به رویدادهای برون‌زاد ودرون‌زاد، که قابلیت مدل شدن و پیش‌بینی را دارد» در این تحقیق شکل گرفت (عبدی، دل‌آرا، دانشجو، ۱۳۹۸).

یکی از راه‌های پیش‌بینی قیمت رمزارزها استفاده از هوش مصنوعی و تکنیک‌های داده کاوی است. داده‌کاوی یکی از پیشرفت‌های اخیر در راستای فناوری‌های مدیریت داده‌هاست. داده کاوی مجموعه‌ای از فنون است که به شخص امکان می‌دهد تا ورای داده‌پردازی معمولی حرکت کند و به استخراج اطلاعاتی که در انبوه داده‌ها مخفی و یا پنهان است کمک می‌کند. انگیزه برای گسترش داده کاوی به طور عمده از دنیای تجارت در دهه ۱۹۹۰ پدید آمد. مثلا داده کاوی در حوزه بازاریابی، بدلیل پیوستگی غیرقابل انتظاری که بین پروفایل یک مشتری و الگوی خرید او ایجاد می‌کند اهمیتی خاص دارد؛ داده کاوی ارتباط نزدیکی با آمار محاسباتی دارد (و اغلب با آن هم‌پوشانی دارد)، تمرکز این شاخه نیز پیش‌بینی کردن و تحلیل توسط رایانه است و پیوند محکمی با بهینه‌سازی ریاضی دارد، که آن هم روش‌ها، تئوری‌ها و

www.SID.ir

Archive of SID

کاربردهایی را وارد میدان می‌کند. یادگیری ماشین گاهی اوقات با داده کاوی ادغام می‌شود؛ تمرکز این زیرشاخه بر تحلیل اکتشافی داده‌ها است و با عنوان یادگیری بی نظارت شناخته می‌شود (ویتن و فرانک[۹]، ۲۰۰۲).

طبق سیکل هایپ[۱۰] کمپانی گارتنر، یادگیری ماشین اکنون در مرحله اوج توقعات زیاد قرار دارد. روش‌های یادگیری ماشین، سیستم‌های مختلف را قادر می‌سازند که یاد بگیرند، بررسی کنند و پیشنهادهای کاربردی ارائه دهند. این سیستم‌ها به مرور زمان که با داده‌ها، شبکه‌ها و افراد تعامل دارند، باهوش‌تر می‌شوند. با استفاده از رویکردهای یادگیری ماشین و هوش مصنوعی[۱۱]، این سیستم‌ها قادر هستند افراد را در حل مسائل مهم، کاربردی و روزمره یاری دهند. غالباً این کار با استفاده از داده‌هایی انجام می‌شود که به دلیل حجم زیاد یا ماهیت نامفهوم، برای انسان‌ها چندان قابل استفاده نیست. ابزارها و روش‌های مبتنی بر یادگیری ماشین، بر خلاف سایر ابداعات و اختراعات بشر، برای رفع محدودیت‌ها و نیازهای فیزیکی نیستند، بلکه هدف آن‌ها ساختن سیستم‌هایی است که به جای انسان بیاندیشند، یاد بگیرند و یاد بدهند (گارتنر، ۲۰۱۶). در طی یک دهه آینده، به نظر می‌رسد که شاهد استفاده هر چه بیشتر یادگیری ماشین در طراحی سیستم‌های دارای تعامل با انسان خواهیم بود. یادگیری ماشین به این معنا است که ماشین بتواند برنامه، ساختار یا داده‌هایش را براساس ورودی‌ها یا در پاسخ به اطلاعات خارجی، به نحوی تغییر دهد که رفتارش به آن چه از او انتظار می‌رود نزدیکتر شود، به‌عبارت دیگر می‌توان گفت یعنی قدرت تجزیه و تحلیل داشته باشد (آبسونگ و همکاران، ۲۰۲۰).

برگسترا و بنجیو[۱۲] (۲۰۱۲) تنظیم مدل حذف تصادفی را برای کاهش شانس بیش‌برازش[۱۳] از مدل مورد برای پیش‌بینی استفاده قرار دادند. این تکنیک به‌عنوان یک طرح حذف نویز که در آن، با دستکاری داده‌های آموخته شده، بیش‌برازش کنترل می‌شود، پذیرفته شده است. بنابراین، این مدل قابلیت تعمیم‌پذیری خود را از لحاظ عملکرد بهتر روی داده‌های نامرئی افزایش می‌دهد. یک نکته مهم دیگر که باید در هنگام پیش‌بینی داده‌های مربوط به سری‌های زمانی مورد توجه قرار گیرد این است که برخی ویژگی‌ها ممکن است شامل اطلاعات اساسی باشند.

چن و همکاران[۱۴] (۲۰۱۳) در مطالعات خود یک مدل مرجع بالقوه را برای پیش‌بینی قیمت بیت‌کوین طراحی و پیاده‌سازی کرده‌اند. این مدل در مدت ۵۰ روز به طرز شگفت‌آوری ۸۹ درصد شاخص سودآوری با نسبت شارپ[۱۵] ۴,۱ ارائه داد. عملکرد سرمایه‌گذاری را می‌توان با نسبت شارپ مورد بررسی قرار داد، در حالی که ریسک آن را تعدیل می‌کند. پس از انتخاب بازه آزمون، استراتژی خرید و حفظ توسط کاربر اتخاذ شد و رشد چشم‌گیر ۳۳ درصدی را به همراه داشت. این مطالعه اطلاعاتی را بدست می‌دهد که نشان‌دهنده وجود الگوهایی در خوشه‌هاست که توسط نویسنده اصلی به صورت دستی استفاده شده است؛ درست مانند آنچه در کتاب‌های معاملاتی به عنوان الگوی "سر و شانه‌ها[۱۶]" نام برده می‌شود. بنابراین، ریسک کار در هنگام انتخاب داده، جهت دست‌یابی به نتایج عالی در نظر گرفته شده است. لیکن پژوهش مذکور از نظر قابلیت مقیاس‌پذیری ریسک بالایی دارد.

شاه و ژانگ[۱۷] (۲۰۱۴) در تحقیق خود بیان داشته‌اند که روش‌های یادگیری ماشینی برای پیش‌بینی قیمت بیت‌کوین جهت اهداف تحقیقاتی چندان استفاده نشده است. پژوهش‌های بسیاری در رابطه با بیت‌کوین وجود دارد و در میان آنها تنها ۱۹ تحقیق از الگوریتم‌های یادگیری ماشین برای پیش‌بینی قیمت بیت‌کوین استفاده می‌کنند. بنابراین، این حوزه به دلیل داشتن ادبیات محدود، فرصتی مناسب برای بررسی و پژوهش ایجاد می‌کند.

گئورگولا و همکاران[۱۸] (۲۰۱۵)، در مطالعات خود از ماشین بردار پشتیبان بر محور عقیده‌کاوی[۱۹] استفاده کرده و از دیگر سو، عوامل تعیین‌کننده قیمت بیت‌کوین را بررسی نموده و یک تناوب همبستگی مثبت در هر دو مدل بدست آورده است. در این پژوهش از یک مدل اصلاح خطای برداری، برای بررسی وجود ارتباط بلندمدت، میان متغیرهای مرتبط استفاده شده است. طبق نتایج این پژوهش، تناوب نرخ هش شبکه[۲۰] و تعداد مشاهده مطالب مربوطه در ویکی‌پدیا، به شدت با قیمت بیت‌کوین در ارتباط است. افزون بر این، در تحقیقات دیگر، از آمار مشاهدات مطالب و عقاید و نظرات در مورد آن، به عنوان پیش‌بینی کننده قیمت بیت‌کوین استفاده شده است. لیکن در این پژوهش از مجموعه داده محدود تاریخچه قیمت بیت‌کوین استفاده شده و هم‌چنین صرفاً به مجموعه داده شبکه اجتماعی توئیتر، اکتفا شده است. ضمناً صحت نتایج براساس شاخص‌های ارزیابی مدل‌های داده‌کاوی، ارزیابی نشده است.

www.SID.ir

Archive of SID

ماتا و همکاران[۲۱] (۲۰۱۵) در پژوهش‌های خود روابط بین قیمت بیت‌کوین، آمار جستجو در مورد بیت‌کوین در گوگل‌ترندز و توئیت‌ها بررسی کرده است و خروجی حاصل از مدل، همبستگی قوی را نشان نمی‌دهد، اما رابطه‌ی ضعیف تا متوسطی بین قیمت بیت‌کوین و توئیت‌های مثبت و گوگل‌ترندز وجود دارد. لیکن در مدل مذکور محدودیت اندازه نمونه وجود دارد که شامل داده‌های ۶۰ روز است. در این‌جا از نظرات و عقاید به عنوان متغیر و از کانال‌های شبکه‌های اجتماعی مبتنی بر وب، به عنوان مثال، ردیت[۲۲] یا توئیتر برای پخش اطلاعات نادرست استفاده می‌شود. به عنوان یک نتیجه از ترفند "بالا ببر و بفروش[۲۳]"، سرمایه‌گذار قرار است از این اطلاعات نادرست که بر روی رسانه‌های اجتماعی پخش شده است، بهره ببرد و این اطلاعات کمک می‌کند تا خرید در قیمت بسیار پایین و فروش در قیمت بسیار بالا انجام شود.

ماتا و همکاران[۲۴] (۲۰۱۵)، در مطالعات خود به جای پیش‌بینی قیمت بیت‌کوین، حجم معاملات را پیش‌بینی کرده است. این تحقیق، نشان می‌دهد که قیمت بیت‌کوین با گوگل‌ترندز، به شدت ارتباط دارد. نمونه داده‌ها، مربوط به دوره‌ای کمتر از یک سال قبل از این که منابع داده برای اجرا در نظر گرفته شود، است. لیکن به دلیل نقصان بازه زمانی، داده‌های نمونه فوق از گوگل ناکافی است.

گرویس و همکاران[۲۵] (۲۰۱۵)، از شبکه عصبی مصنوعی[۲۶] و ماشین بردار پشتیبان برای پیش‌بینی قیمت بیت‌کوین با بررسی زنجیره بلوکی بیت‌کوین استفاده نموده است. مدل از صحت ۵۵ درصد برخوردار بوده و نتیجه گرفته است که داده‌هایی که از زنجیره بلوکی جمع‌آوری می‌شوند، در جایی که قیمت‌ها توسط کنترل صرافی‌هایی است و رفتار آن‌ها خارج از محدوده زنجیره بلوکی است، دارای ثبات غیرطبیعی هستند. معایب این پژوهش، استفاده از صرفاً یک مجموعه داده (قیمت‌کنونی بیت‌کوین) و ارزیابی مدل فقط با شاخص صحت در مدل دسته‌بندی می‌باشند.

مادن و همکاران[۲۷] (۲۰۱۵) قیمت بیت‌کوین را با استفاده از داده‌های حاصل از زنجیره بلوکی و بکارگیری ماشین بردار پشتیبان، جنگل تصادفی، مدل خطی تعمیم یافته دو جمله‌ای، به عنوان الگوریتم‌های یادگیری ماشین، پیش‌بینی کرده است. صحت مدل بدون اعتبارسنجی متقابل و لحاظ قابلیت تعمیم آن ۹۸,۷ درصد بدست آمد. تا آنجا که به مقوله یادگیری ماشین مربوط می‌شود، پیش‌بینی قیمت بیت‌کوین را می‌توان مانند دیگر پیش‌بینی‌های مربوط به سری‌های زمانی، به عنوان مثال سهام و فارکس که زیرمجموعه حوزه مالی است، در نظر گرفت. استفاده از شبکه عصبی مصنوعی، ایده جدیدی نیست. این مطالعه نشان می‌دهد که تکنیک پس انتشار در این زمینه زیاد اجرا می‌شود (رملهارت و همکاران[۲۸]) و فرض بر این است که شبکه عصبی مصنوعی برای پیش‌بینی و مدل‌سازی بر مبنای مسئله سری‌های زمانی غیرخطی مناسب است (تانگ و همکاران[۲۹]، ۱۹۹۱و ویجند و همکاران[۳۰]، ۱۹۹۰). پرسپترون چند لایه[۳۱]، توسط پژوهشگران متعددی، برای پیش‌بینی قیمت سهام اجرا شده است (یون و همکاران[۳۲]، ۱۹۹۱). پیش‌بینی قیمت سهام آی‌بی‌ام با استفاده از پرسپترون چند لایه، به دلیل فقدان اطلاعات کافی در مدلی که توسط (وایت[۳۳]، ۱۹۸۸) طراحی شد، ارزش محدودی را بدست آورد. به‌علاوه، روال‌های جستجوی شبکه خطا و پارامترهای آزمایش، برای این پژوهش مورد استفاده قرار گرفتند و محدودیت این روش در یافتن مقدار بیشینه مطلق، مشاهده شد. در این مطالعه مشخص شد که روش جستجوی تصادفی در مقایسه با جستجوی شبکه بسیار مفید است زیرا از جستجوی تصادفی برای یافتن مدل‌های بهتر در یک زمان مشخص استفاده می‌شود.

مک نالی و همکاران[۳۴] (۲۰۱۸) در مطالعات خود مجموعه داده‌های جمع‌آوری‌شده از شاخص قیمت بیت‌کوین را برای پیش‌بینی قیمت آن، براساس الگوریتم‌های شبکه عصبی بازگشتی بیزی بهینه[۳۵] و حافظه طولانی‌مدت-کوتاه‌مدت استفاده کرده است. در این تحقیق، نویسنده صحت طبقه‌بندی را با استفاده از حافظه طولانی کوتاه مدت ۵۲ درصد و خطای جذر میانگین مربعات ۸ درصد بدست آورده و مدل میانگین متحرک خودهمبسته یکپارچه[۳۶] را در برابر مدل‌های یادگیری عمیق پیاده‌سازی کرده است. در این پژوهش، ارزیابی مدل‌ها از لحاظ معیارهای صحت، دقت، حساسیت و هم‌چنین با استفاده از معیار خطای میانگین جذر مربعات به عنوان ابزار مقایسه خطاهای پیش‌بینی صورت گرفته است. این مطالعه نشان می‌دهد که مدل مذکور نسبت به روش‌های یادگیری عمیق غیرخطی عملکرد بهتری ندارد.

روش پیشنهادی

در این پژوهش از روش CRISP-DM (فرآیندهای

www.SID.ir

Archive of SID

استاندارد صنایع متقابل برای داده‌کاوی به طور مصطلح کریسپ) استفاده خواهد شد. این فرآیند استاندارد، با تلاش کنسرسیومی که در ابتدا با دایملر-کرایسلر، SPSS و NCR تشکیل شده بود، ساخته شد. کریسپ مدلی فرآیندی استاندارد باز است که رویکردهای عمومی متخصصان داده‌کاوی را تشریح می‌کند، این روش‌شناسی پرکاربردترین مدل تحلیلی می‌باشد. فرآیندکریسپ چابک بوده و عموماً در پروژه‌های داده کاوی مورد استفاده قرار می‌گیرد، زیرا روال کار را در گذار از فازهای مختلف مانند تبدیل داده و مدل‌سازی داده‌ها در چرخه عمر ساده می‌کند (آزیودو و سانتوس[۳۷]، ۲۰۰۶). گام‌هایی که در روش کریسپ انجام می‌شوند به ترتیب عبارتند از:

فهم تجاری[۳۸]: شامل گردآوری موارد مورد نیاز و تعیین اهداف داده کاوی است.

مجموعه داده‌های مورد استفاده برای این پژوهش از بانک اطلاعاتی Kaggle و تاریخچه پایگاه داده CoinMarketCap جمع‌آوری شده است. در این تحقیق از ۱۰ مجموعه داده رمز ارز استفاده شده است. هر مجموعه داده از ۷ ویژگی تشکیل شده و روش انتخاب ویژگی‌های اصلی برای پیش‌بینی قیمت بیت‌کوین، با مشاهده روند هر رمز ارز در مقایسه با توزیع قیمت پایانی بیت‌کوین است.

درک داده[۳۹]: نگاه نزدیک و بررسی دسترسی به داده‌ها برای فرایند داده‌کاوی که شامل گردآوری، توصیف، کشف و تغییر کیفیت داده‌ها می‌شود.

داده‌های مورد استفاده در این پژوهش شامل تغییرات قیمتی ده رمز ارز اول این بازار از نظر میزان سرمایه است که فهرست آن در این پژوهش طبق جدول (۱) می‌باشد.

جدول ۱- فهرست ۹ رمز ارز مورد استفاده به عنوان متغیرهای وابسته و مستقل پژوهش

ردیف	عنوان متغیر	نوع متغیر			کمی		تعریف متغیر	نحوه اندازه‌گیری	مقیاس	بازه زمانی	تناوب بروزرسانی
		زمینه‌ای	مستقل	وابسته	پیوسته	گسسته					
۱	**BTC**			✓		✓	قیمت بیت‌کوین برحسب دلار امریکا	شمارش	عددی		یک روز
۲	**ETH**		✓			✓	قیمت اتریوم[۴۰] برحسب دلار امریکا	شمارش	عددی		یک روز
۳	**XRP**		✓			✓	قیمت ریپل[۴۱] برحسب دلار امریکا	شمارش	عددی		یک روز
۴	**BCH**		✓			✓	قیمت بیت‌کوین‌کش[۴۲] برحسب دلار امریکا	شمارش	عددی		یک روز
۵	**BSV**		✓			✓	قیمت بیتکوین اس وی[۴۳] برحسب دلار امریکا	شمارش	عددی		یک روز
۶	**LTC**		✓			✓	قیمت لایت کوین[۴۴] برحسب دلار امریکا	شمارش	عددی		یک روز
۷	**BNB**		✓			✓	قیمت بایننس کوین[۴۵] برحسب دلار امریکا	شمارش	عددی		یک روز
۸	**CRO**		✓			✓	قیمت کریپتوکوین[۴۶] برحسب دلار امریکا	شمارش	عددی		یک روز
۹	**EOS**		✓			✓	قیمت ایاس[۴۷] برحسب دلار امریکا	شمارش	عددی		یک روز

- فرمت دریافت و نگهداری اطلاعات در قابل فایل csv است.

- منابع گردآوری داده‌های پژوهش، پایگاه داده kaggle

www.SID.ir

Archive of SID

و تاریخچه قیمتی سایت coinmarketcap می‌باشد.

آماده‌سازی داده: این مرحله شامل انتخاب، پاک‌سازی، ساختاربندی و ادغام داده‌ها می‌شود. این مرحله شامل تمام فعالیت‌های مربوط به نهایی کردن مجموعه داده‌ها است که برای پاک‌سازی داده‌ها و تبدیل آن مورد استفاده قرار می‌گیرد و این فعالیت قابل گنجاندن در یک چارچوب زمانی نمی‌باشد، زیرا این مرحله مکرراً در پژوهش حاضر مورد استفاده گرفته است، تا زمانی که داده‌ها کاملا آماده شده‌اند. آزمایش اولیه داده‌های موجود با هدف یافتن درک اولیه نسبت به آن، نشان داد که متغیر تاریخ، برای هر کدام از رمز ارزها، از مقدار بخصوصی شروع شده است که این تاریخ هم‌زمان با عرضه رسمی آن رمزارز به بازار می‌باشد. اولین قیمت برای بیت‌کوین در تاریخ ۲۹ آوریل ۲۰۱۳، برای اتریوم ۷ آگوست ۲۰۱۵ و برای سایر رمز ارزها نیز هر کدام از تاریخ بخصوصی شروع شده است. در نتیجه به‌عنوان یک پیش پردازش[۴۸] بر روی مجموعه داده، اشتراک بازه زمانی همه رمزارزها، به عنوان بازه مورد بررسی پژوهش در نظر گرفته شد و مابقی زمان‌ها حذف گردید.

در کل ده مجموعه داده برای این پژوهش استفاده شده است که هرکدام دارای ویژگی‌های جدول (۲) هستند.

جدول ۲- ویژگی‌های مجموعه داده‌های مورد استفاده در پژوهش

ردیف	عنوان فارسی ویژگی	اصطلاح رایج انگلیسی
۱	قیمت بازشدن رمز ارز	Open
۲	بیشینه قیمت	High
۳	کمینه قیمت	Low
۴	قیمت بسته شدن	Close
۵	حجم مبادلات	Volume
۶	ارزش بازار	Market Capital

مجموعه داده مربوط به بیت‌کوین از آوریل ۲۰۱۳ تا جولای ۲۰۲۰ بسط داده شد که توزیع قیمت بسته شدن آن در شکل ۱ ترسیم شده است.

شکل ۱- نمودار تغییرات قیمت بیت‌کوین از آوریل ۲۰۱۳ تا جولای ۲۰۲۰

Source: https://coinmarketcap.com/currencies/bitcoin

می‌توان مشاهده کرد که چگونه قیمت در طول زمان تغییرکرده است. با توجه به تعدد ویژگی‌ها، قیمت بسته شدن هر رمز ارز به عنوان ویژگی اصلی آن در نظر گرفته شد. بنابراین مجموعه داده جدید متشکل از ویژگی‌های جدول (۳) ایجاد شد.

www.SID.ir

Archive of SID

جدول ۳- مجموعه داده جدید مورد استفاده در پژوهش متشکل از قیمت بسته شدن روزانه هر رمز ارز

ردیف	عنوان ویژگی	نماد	ردیف	عنوان ویژگی	نماد
۱	قیمت بسته شدن ایاس	EOS	۷	قیمت بسته شدن بیت‌کوین‌کش	BCH
۲	تاریخ	Date	۸	قیمت بسته شدن بیت‌کوین‌اس‌وی	BSV
۳	قیمت بسته شدن بیت‌کوین	BTC	۹	قیمت بسته شدن لایت‌کوین	LTC
۴	قیمت بسته شدن اتریوم	ETH	۱۰	قیمت بسته شدن بایننس‌کوین	BNB
۵	قیمت بسته شدن ریپل	XRP	۱۱	قیمت بسته شدن کریپتوکوین	CRO

مدل‌سازی: در این مرحله از داده‌های پردازش شده برای آموزش مدل استفاده می‌شود. در این روش ۶۶ درصد داده ها به عنوان داده‌های آموزشی و ۳۳ درصد داده‌ها به عنوان داده‌های آزمایشی در نظر گرفته می‌شود.

ارزیابی: در این مرحله نتایج ارزیابی شده، فرایند انجام کار بازبینی و مراحل بعدی انجام می‌شوند. برای ارزیابی روش پیشنهادی از میانگین خطای پیش‌بینی استفاده می‌شود.

توسعه: نتایج به‌دست آمده توسعه یافته و برای بهبود عملکرد به کار گرفته می‌شوند. در این مرحله عملکرد روش پیشنهادی مورد ارزیابی قرار گرفته می‌شود و با توجه به نتایج خروجی عملکرد مورد بررسی قرار می‌گیرد.

تحلیل و ارزیابی نتایج

- صحت، دقت، بازیابی، امتیاز F1

شاخص‌های ارزیابی صحت، دقت، بازیابی و معیار F1، شاخص‌های مهم ارزیابی عملکردی هستند که مشخص کننده بهترین مدل برای پیش‌بینی قیمت هستند (جاشی و همکاران[۴۹]، ۲۰۱۶ و برونلی[۵۰]، ۲۰۱۴).

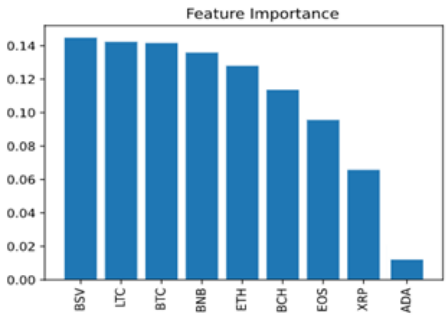
مقدار شاخص صحت، برابر با نسبت تعداد پیش‌بینی‌های صحیح به کل پیش‌بینی‌ها است (فرمول ۱)- به عبارت ساده‌تر، اگر مدل به صحت بالایی دست یابد، به این معنی است که مدل بهتر است. در این تحقیق مسئله به عنوان یک طبقه‌بندی دو کلاسه در نظر گرفته می‌شود. در واقع با توجه به پیش‌بینی درست یا غلط در مورد نتیجه‌گیری تصمیم‌گیری می‌شود.

فرمول (۱):
$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FN+FP+TN}}$$

علاوه بر این، همان‌طور که در جدول (۴) قابل مشاهده است، الگوریتم گرادیان تقویتی از گروه دسته‌بندی، بیشتر مقدار صحت را بدست آورده است. این در حالی است که الگوریتم جنگل تصادفی، کمترین مقدار صحت را بدست آورده است. همان‌طور که در شکل (۲) دیده می‌شود، رمز ارزها به شدت با یکدیگر همبستگی دارند.

شکل ۲ نشان می‌دهد که ETH و BCH با بیت‌کوین همبستگی متوسطی دارند، در حالی که BSV و LTC بیشترین همبستگی را دارند. بنابراین، فرضیه ما این است که BSV و LTC بهترین پیش‌بینی‌ها را تولید می‌کنند.

الگوریتم‌های گروه دسته‌بندی، از نظر شاخص‌های عملکردی که معیار صحت در آنها مهم‌ترین نقش را دارد، به مقادیر جدول (۴) دست یافتند.



شکل ۲- همبستگی ویژگی‌ها با ارزش بیت کوین در روش دسته‌بندی

www.SID.ir

Archive of SID

جدول ۴- شاخص‌های عملکرد الگوریتم‌های گروه دسته‌بندی

معیار / الگوریتم	جنگل تصادفی	بردار پشتیبان	گرادیان تقویتی	شبکه عصبی چند لایه پرسپترون
صحت (درصد)	۵۰	۵۱,۱۱	۵۶,۷	۵۰,۸۶
بازیابی	۰,۵۰	۰,۵۱	۰,۵۴	۰,۵۱
امتیاز (F1)	۰,۴۵	۰,۵۴	۰,۵۴	۰,۴۶
دقت	۰,۵۶	۰,۶۳	۰,۵۵	۰,۳۷

الگوریتم گرادیان تقویتی، با عدد ۵۶,۷ درصد، بیشترین مقدار را برای معیار صحت حاصل کرد. الگوریتم جنگل تصادفی به مقدار صحت ۵۰ درصد، ماشین بردار پشتیبان ۵۱,۱۱ و شبکه عصبی به عدد ۵۰,۸۶ درصد دست یافت.

سایر معیارهای ارزیابی عملکرد در گروه دسته‌بندی، شامل دقت، بازیابی و امتیاز F1 هستند. دقت برابر است با نسبت موارد صحیح طبقه‌بندی‌شده توسط الگوریتم از یک کلاس مشخص، به کل تعداد مواردی که الگوریتم چه به صورت صحیح و چه به صورت غلط، در آن کلاس طبقه‌بندی کرده است (فرمول ۲).

فرمول (۲): $\text{Precision} = \frac{TP}{TP+FP}$

معیار بازیابی، برابر نسبت موارد صحیح طبقه‌بندی شده توسط الگوریتم از یک کلاس، به تعداد موارد حاضر در کلاس مذکور که مطابق فرمول۳ محاسبه می‌شود.

فرمول (۳): $\text{Recall} = \text{Sensitivity} = (TPR) = \frac{TP}{TP+FN}$

امتیاز F1 در واقع یک نوع میانگین بین معیار دقت و یادآوری است که مطابق فرمول ۴ محاسبه می‌شود.

فرمول (۴): $F-1 = 2 * \frac{Precision*Recall}{Precision+Recall}$

الگوریتم بردار پشتیبان، به مقدار ۰,۶۳ برای معیار دقت دست یافت که بدین معنی است که بردار پشتیبان، دارای دقت بالاتری است و بیانگر نرخ پایین‌تر مثبت کاذب[۵۱] آن است. بنابراین مقدار حاصل از الگوریتم ماشین بردار پشتیبان مطلوب‌تر است. الگوریتم بردار پشتیبان به عدد ۰,۵۱ برای معیار بازیابی دست یافت که این مقدار نسبتاً مطلوب است زیرا بیشتر از ۰,۵ است. مقدار حاصل برای امتیاز F1 عدد ۰,۵۴ بود. در مورد الگوریتم گرادیان تقویتی، مقدار ۰,۵۵ برای معیار دقت، ۰,۵۴ برای معیار بازیابی و ۰,۵۴ برای امتیاز F1 حاصل شد. درخصوص الگوریتم شبکه عصبی، روش یادگیری پرسپترون چند لایه (شبکه عصبی سه لایه یا یک لایه مخفی و تعداد ۲۰ نرون در لایه مخفی)، استفاده می‌شود. هم‌چنین در شبکه عصبی از تابع یادگیری ReLU استفاده می‌شود. در واقع در شبکه‌های جدید ترجیح می‌دهند به جای sigmoid، از توابع فعال‌سازی برای لایه‌های مخفی استفاده کنند. این تابع به صورت زیر تعریف می‌شود:

$f(x) = \max(x,0)$

در این تابع، اگر مقدار X بزرگتر از صفر باشد، خروجی Xاست، و اگر مقدار X کوچکتر یا مساوی صفر باشد، خروجی صفر است.

مقدار ۰,۳۷ برای معیار دقت، ۰,۵۱ برای معیار بازیابی و ۰,۴۶ برای امتیاز F1 بدست آمد. الگوریتم جنگل تصادفی به مقدار ۰,۵۶ برای معیار دقت، ۰,۵۰ برای بازیابی، ۰,۴۵ برای امتیاز F1 دست یافت.

همبستگی‌هایی که برای دست‌یابی به صحت، بسیار مهم هستند (میزان اهمیت ویژگی‌ها[۵۲])، به ترتیب امتیاز عبارتند از: پر واضح است که BSV با امتیاز ۰,۱۴۴۸۰۶۰۷ بیشترین همبستگی را داشته و پس از آن به ترتیب LTC امتیاز ۰,۱۴۲۳۱۵۶، BNB امتیاز ۰,۱۳۵۹۰۷۸۸ و ETH امتیاز ۰,۱۲۷۹۹۵۸۱ را بدست آورده‌اند. هم‌چنین BCH، EOS و XRP به ترتیب امتیازهای ۰,۱۱۳۶۲۰۶۲، ۰,۰۹۵۶۰۶۵ و ۰,۰۶۵۷۶۳۸۱ را بدست آورده‌اند. ویژگی‌های USDT و ADA امتیاز ۰,۰۲۰۲۵۴۸۲ و ۰,۰۱۲۰۷۶۸۱ را بدست آورده‌اند. درکل BSV بیشترین امتیاز را میان سایر رمز ارزها بدست آورد.

ضریب تشخیص پیرسون

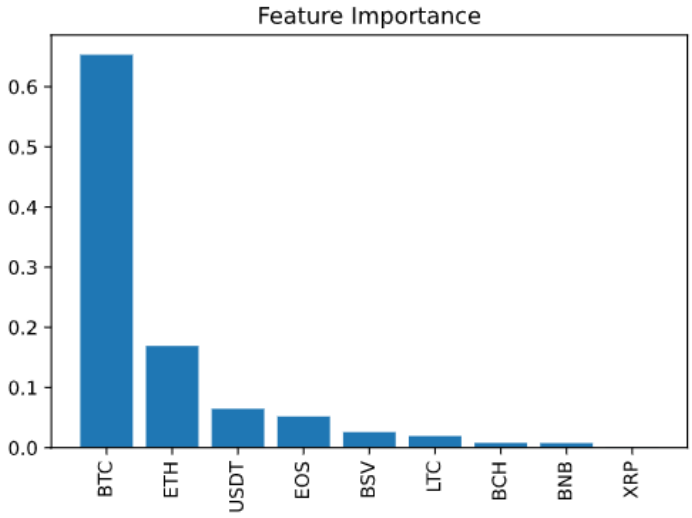
در این تحقیق علاوه بر استفاده از طبقه‌بندی از رگرسیون

www.SID.ir

Archive of SID

نیز برای ارزیابی عملکرد استفاده می‌شود. در این مرحله برای ارزیابی روش رگرسیون از از ضریب تشخیص R^2 پیرسون استفاده می‌کنیم. مقدار R^2 معادل نسبت واریانس در مجموعه داده‌ای است که در مدل مشاهده می‌شود که مقدار آن در بازه ۰ و ۱ است. هر چه R^2 بزرگتر باشد، یعنی واریانس بیشتری توسط مدل تشریح می‌شود. به بیان دیگر میزان تطابق و کارآمدی پیش‌بینی‌های یک مدل رگرسیون را با داده‌های ورودی آن نشان می‌دهد. در این پژوهش، از ضریب تشخیص R^2 پیرسون به‌عنوان ابزاری جهت مقایسه بین مدل‌های مختلف کرده‌ایم. شکل ۳ ویژگی‌هایی که برای پیش‌بینی قیمت بیت‌کوین، بیشترین اهمیت را دارند، نشان می‌دهد.

کاملاً روشن است که BTC (قیمت روز بیت‌کوین) بالاترین اهمیت را در تعیین قیمت روز بعد آن دارد و مطابق شکل ۳، BTC بیشترین امتیاز را معادل ۰٫۶۵۳۱۳۶۵۶۳ بدست آورده است. در حالی که ETH امتیاز ۰٫۱۶۹۰۶۳۱۲۵، USDT امتیاز ۰٫۰۶۴۶۱۴۷۷، EOS امتیاز ۰٫۰۵۱۸۸۰۸۳ و BSV امتیاز ۰٫۰۲۵۹۴۶۷۰ را بدست آورده است. هم‌چنین LTC، BCH و BNB، به ترتیب امتیازهای ۰٫۰۱۹۳۴۰۲۲ و ۰٫۰۰۷۹۲۲۲۱ و ۰٫۰۰۷۶۰۸۹۴ را بدست آورده‌اند. رمز ارز XRP نیز امتیاز ۰٫۰۰۰۴۸۶۶۴ را بدست آورد. در مجموع BTC بیشترین امتیاز را میان سایر ویژگی‌ها بدست آورد.



شکل ۳- همبستگی رمز ارزها با بیت کوین در روش رگرسیون

در گروه رگرسیون، سه الگوریتم پیاده‌سازی شدند و گرادیان تقویتی بیشترین مقدار را برای R^2 برابر با ۰٫۹۸۶۸ بدست آورد. الگوریتم رگرسیون خطی مقدار ۰٫۹۸۶۵ و شبکه عصبی بازگشتی مقدار ۰٫۹۶ را بدست آوردند.

الگوریتم جنگل تصادفی به این دلیل انتخاب شد که از میانگین‌گیری برای افزایش صحت و بررسی بیش برازش[۵۳] استفاده می‌کند، زیرا متناسب با چندین درخت تصمیم مختلف بر روی نمونه‌های مختلف است. الگوریتم جنگل تصادفی، به مقدار ۵۰ درصد برای شاخص صحت دست یافت.

الگوریتم بردار پشتیبان حتی زمانی که ویژگی‌ها بیشتر از نمونه‌ها باشند، عملکرد بهتری دارد. بنابراین بردار پشتیبان، قوی‌تر عمل می‌کند. محدودیت این الگوریتم در انتخاب هسته است و با نمونه‌های زیاد، نتایج خوبی بدست نمی‌دهد. این الگوریتم به ۵۱٫۱۱ درصد برای معیار صحت، دست یافت. الگوریتم گرادیان تقویتی، برای هر دو گروه مسائل دسته‌بندی و رگرسیون، تکنیکی موثر و دقیق است. محدودیت این الگوریتم در مقیاس‌پذیری آن است. الگوریتم گرادیان تقویتی در گروه دسته‌بندی به مقدار صحت ۵۶٫۰۷ درصد دست یافت.

شبکه عصبی از تابع فعال‌ساز مثل ReLU[۵۴] برای لایه پنهان و تغییر خروجی براساس لایه ورودی بهره می‌برد. یکی از محدودیت‌های این الگوریتم، دشواری در تنظیم دقیق آن و در نتیجه، دشواری در اجراست. شبکه عصبی به صحت ۵۰٫۸۶ درصد دست یافت.

هم‌چنین در این پژوهش، الگوریتم جنگل تصادفی به عدد ۰٫۵۶ برای معیار دقت، عدد ۰٫۵۰ برای معیار بازیابی و عدد ۰٫۴۵ برای امتیاز F1 دست یافت. امتیاز mF1 ثبت کاذب و

www.SID.ir

Archive of SID

منفی کاذب را نیز لحاظ می‌کند. الگوریتم ماشین بردار پشتیبان به عدد ۰,۶۳ برای معیار دقت، ۰,۵۱ برای معیار بازیابی و ۰,۵۴ برای امتیاز F1 دست یافت. الگوریتم گرادیان تقویتی به نتایج ۰,۵۵ برای معیار دقت، ۰,۵۴ برای معیار بازیابی و ۰,۵۴ برای امتیاز F1 دست یافت. شبکه عصبی نیز به نتایج ۰,۳۷ برای معیار دقت، ۰,۵۱ برای بازیابی و ۰,۴۶ برای امتیاز F1 دست یافت. در نتیجه، الگوریتم گرایان تقویتی در مقایسه با سایر الگوریتم‌های گروه دسته‌بندی، به بیشترین مقدار برای معیار صحت دست یافت که طبعاً فرصتی برای تحقیق و بررسی‌های آتی در این زمینه و با استفاده از الگوریتم‌های مختلف و ایده‌های مشابه این پژوهش در زمینه مسائل دسته‌بندی ایجاد می‌کند.

از سوی دیگر، الگوریتم‌های شبکه عصبی بازگشتی، رگرسیون خطی، رگرسیون گرادیان تقویتی، در گروه رگرسیون استفاده شدند. در این بخش، قیمت بیت‌کوین با استفاده از شبکه عصبی بازگشتی برای یک روز پیش‌بینی شد که به مقدار ۰,۹۶ برای ضریب تشخیص R^2 پیرسون دست یافت. الگوریتم رگرسیون خطی به مقدار ۰,۹۸۶۵ برای ضریب تشخیص R^2 پیرسون دست یافت. رگرسیون گرادیان تقویتی بیشترین مقدار را برای ضریب R^2 بدست آورد که برابر با ۰,۹۸۶۸ بود.

نتیجه‌گیری و پیشنهادها

بیت‌کوین ارز رمزنگاری شده‌ای است که مبتنی بر فناوری بلاکچین کار می‌کند و بدلیل رشد سریع قیمت، سرمایه‌گذاران زیادی را به خود جذب کرده است. در بازار مبادلات بیت‌کوین، دو عامل سرعت و دقت در رسیدن به سود و یا ضرر سرمایه‌گذاران بسیار با اهمیت است. از این رو در این پژوهش نوعی از تکنیک‌های داده کاوی و روش‌های هوشمند به منظور پیش‌بینی قیمت بیت‌کوین استفاده کرده‌ایم. الگوریتم‌های داده کاوی، برای حل مسائل جهان واقعی، بسیار کاربردی هستند. در این پژوهش، هدف اصلی این بود که چگونه می‌توان جهت و صحت پیش‌بینی قیمت بیت‌کوین را براساس الگوریتم‌های هوشمند و داده کاوی تعیین کرد. مجموعه داده مرتبط با ده رمز ارز برای شکل دادن یک مجموعه داده جدید، انتخاب و جهت نیل به هدف پژوهش استفاده شدند. بدین ترتیب، روش یادگیری با نظارت، که در آن الگوریتم‌های جنگل تصادفی، ماشین بردار پشتیبان، گرادیان تقویتی و شبکه عصبی پیاده‌سازی شدند، برای یافت بهترین میزان صحت مدل، استفاده شد.

راه‌های متعددی برای بهبود روش استفاده شده در این پژوهش از جمله روش‌های مختلف یادگیری عمیق وجود دارند. داده‌های مربوط به بیت‌کوین نقش مهمی را در این میان بازی می‌کنند. استفاده از مجموعه داده صرافی‌های متعدد مثل بلاک‌چین، بایننس[۵۵]، بیترکس[۵۶] و پولونیکس[۵۷]، لحاظ کردن میانگین قیمت بسته شدن رمز ارز، که از این صرافی‌ها جمع‌آوری شده است، می‌تواند بینش متفاوتی در مورد مدل استفاده شده، حاصل کند. روش‌های یادگیری ماشین بدون نظارت هم می‌توانند زاویه نگرش متفاوتی در خصوص داده‌ها ایجاد کنند و دریچه جدیدی را برای پژوهش در این زمینه برای محققان آینده بگشایند. به علاوه پیشنهاد می‌شود که کسب دانش کافی در زمینه تکنولوژی پشت پرده بیت‌کوین، در دستور کار پژوهشگران آتی لحاظ گردد.

منابع

آقاخانی، کیارش و کریمی، عباس (۱۳۹۳). بررسی روش‌های پیش‌بینی قیمت سهام در بازار بورس و معرفی روش بهینه، همایش ملی الکترونیکی دستاوردهای نوین در علوم مهندسی و پایه، اردبیل.

شیرزور علی‌ابادی، زهرا، رمضان‌زاده، حمید (۱۳۹۹). بیت‌کوین و پیش‌بینی آینده آن، ششمین کنفرانس بین‌المللی مهندسی صنایع و سیستم‌ها، مشهد.

عبدی، نسرین، دل‌آرا، چنگیز، دانشجو، پریسا (۱۳۹۸). پیش‌بینی قیمت بیت‌کوین با استفاده از شبکه عصبی LSTM، نهمین کنفرانس ملی علوم و مهندسی کامپیوتر و فناوری اطلاعات، بابل.

Azevedo, A.I.R.L and M.F. Santos (2008). KDD, SEMMA and CRISP-DM: a parallel overview. IADS-DM.

Bergstra, J and Y. Bengio (2013). Random search for hyper-parameter optimization, Journal of machine learning research.

Brownlee, J (2014). Classification accuracy is not enough: More performance measures you can use, Machine Learning Mastery.

Chatfield, C and M. Yar (1988). Holt-Winters forecasting: some practical issues, Journal of the

www.SID.ir

Archive of SID

Rumelhart, D.E, G.E. Hinton and R.J. Williams (1986). Learning representations by back-propagating errors.

Samuel, A.L (1959). Some studies in machine learning using the game of checkers, IBM Journal of research and development.

Shah, D and K. Zhang (2014). Bayesian regression and Bitcoin, 52nd annual Allerton conference on communication, control, and computing (Allerton).

Tang, Z, C. De Almeida, and P.A Fishwick (1991). Time series forecasting using neural networks vs. Box-Jenkins methodology, Simulation.

Velankar, S. S. Valecha and S. Maji (2018). Bitcoin price prediction using machine learning, 20th International Conference on Advanced Communication Technology (ICACT).

Wager, S. S. Wang and P.S. Liang (2013). Dropout training as adaptive regularization, in Advances in neural information processing systems.

Weigend, A.S, B.A. Huberman and D.E. Rumelhart (1990). Predicting the future: A connectionist approach, International journal of neural systems.

White, H (1988). Economic prediction using neural networks: The case of IBM daily stock returns.

Witten, I.H and E. Frank (2002). Data mining: practical machine learning tools and techniques with Java implementations, Acm Sigmod Record.

Yoon, Y and G. Swales (1991). Predicting stock price performance: A neural network approach, in Proceedings of the twenty-fourth annual Hawaii international conference on system sciences.

یادداشت‌ها

[1] *Nakamoto, S. and A, (2008*)
[2] *Satoshi Nakamoto*
[3] *CoinMarketCap, (2020)*
[4] *Technical Analysis*
[5] *McNally, S., J. Roche, and S. Caton,(2018)*
[6] *Holt-Winters exponential smoothing models*
[7] *Noise*
[8] *Chatfield, C. and M. Yar, 1988.*

Royal Statistical Society: Series D (The Statistician.

Chen, G.H, S. Nikolov and D. Shah (2013). A latent source model for nonparametric time series classification, in Advances in Neural Information Processing Systems.

CoinMarketCap (2020). Bitcoin Statistics.

Gartner, I, Gartner's (2016). Hype Cycle for Emerging Technologies, August.

Georgoula, I, et al (2015). Using time-series and sentiment analysis to detect the determinants of bitcoin prices, Available at SSRN 2607167, Conference: 9th Mediterranean Conference on Information SystemsAt: Samos.

Greaves, A and B. Au (2015). Using the bitcoin transaction graph to predict the price of bitcoin, No Data. Joshi, R, Accuracy, precision (2016). recall & f1 score: Interpretation of performance measures, Retrieved April.

Madan, I, S. Saluja and A. Zhao (2015). Automated bitcoin trading via machine learning algorithms, URL. Matta, M, I. Lunesu and M. Marchesi (2015). Bitcoin Spread Prediction Using Social and Web Search Media, in UMAP Workshops.

Matta, M, I. Lunesu and M. Marchesi (2015). The predictor impact of Web search media on Bitcoin trading volumes, 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K).

McNally, S, J. Roche, and S. Caton (2018). Predicting the price of bitcoin using machine learning, in 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP). Nakamoto, S and A (2008). Bitcoin, A peer-to-peer electronic cash system.

Obthong, M, et al (2020). A survey on machine learning for stock price prediction: algorithms and techniques.

[9] *Witten, I.H. and E. Frank, (2002).*
[10] *Hype Cycle*
[11] *Artificial Intelligence*
[12] *Bergstra, J. and Y. Bengio, (2012)*
[13] *Overfitting*
[14] *Chen, G.H., S. Nikolov, and D. Shah, (2013)*
[15] *Sharpe Ratio*
[16] *Head and Shoulders*

www.SID.ir

Archive of SID

[17] *Shah, D. and K. Zhang, (2014)*
[18] *Georgoula, I., et al, (2015)*
[19] *Sentiment Analysis*
[20] *Network Hash Rate*
[21] *Matta, M., I. Lunesu, and M. Marchesi, (2015a)*
[22] *Reddit*
[23] *Pump and Dump*
[24] *Matta, M., I. Lunesu, and M. Marchesi, (2015b)*
[25] *Greaves, A. and B. Au, (2015)*
[26] *Artificial Neural Network*
[27] *Madan, I., S. Saluja, and A. Zhao, (2015)*
[28] *Rumelhart, D.E., G.E. Hinton, and R.J. Williams, (1986).*
[29] *Tang, Z., C. De Almeida, and P.A. Fishwick, (1991).*
[30] *Weigend, A.S, B.A. Huberman, and D.E. Rumelhart, (1990).*
[31] *Multilayer Perceptron*
[32] *Yoon, Y. and G. Swales, (1991).*
[33] *White, H.,(1988).*
[34] *McNally, S., J. Roche, and S. Caton, (2018)*
[35] *Bayesian Optimized RNN*
[36] *Autoregressive integrated moving average (ARIMA)*
[37] *Azevedo, A.I.R.L. and M.F. Santos, (2008).*
[38] *Business Understanding*
[39] *Data Understanding*
[40] *Ethereum*
[41] *Ripple*
[42] *Bitcoin cash*
[43] *Bitcoin SV*
[44] *Litecoin*
[45] *Binance coin*
[46] *Crypto-coin*
[47] *EOS*
[48] *Preprocessing*
49 *Joshi, R., Accuracy, precision, (2016).*
50 *Brownlee, J.,(2014).*
[51] *False Positive*
[52] *Feature Importance*
[53] *Over-fitting*
[54] *Rectified Linear Unit*
[55] *Binance.com*
[56] *Bittrex.com*
[57] *Poloniex.com*

www.SID.ir