پیش‌بینی تغییرات قیمت بیتکوین با استفاده از تحلیل احساسات در شبکه‌های اجتماعی و سلبریتی‌ها به همراه رویکرد داده‌کاوی

علی قربانی

دانشجوی کارشناسی ارشد، گروه مهندسی صنایع، دانشکده مهندسی، دانشگاه سمنان

ali_qorbani@semnan.ac.ir

یوسف ربانی (نویسنده مسئول)

استادیار، گروه مهندسی صنایع، دانشکده مهندسی، دانشگاه سمنان

Rabbani@semnan.a.ir

رضا کامران راد

استادیار، گروه مهندسی صنایع، دانشکده مهندسی، دانشگاه سمنان

r.kamranrad@semnan.ac.ir

پیمان فلسفی

دانشجوی کارشناسی ارشد، گروه مهندسی صنایع، دانشکده مهندسی، دانشگاه سمنان

peyman.falsafi@semnan.ac.ir

نوع مقاله: علمی- پژوهشی تاریخ دریافت: ۱۴۰۱/۰۸/۲۷ تاریخ پذیرش: ۱۴۰۱/۱۰/۱۱

چکیده:

تحقیقات زیادی در مورد پیش‌بینی تغییرات قیمتی رمز ارزها انجام شده است که پژوهشگران از الگوریتم‌های یادگیری ماشین در این زمینه استفاده نمودند. هدف این مقاله ارائه رویکردی برای پیش‌بینی تغییرات قیمتی بیتکوین با استفاده از الگوریتم‌های یادگیری ماشین همراه با استفاده از داده‌های تحلیل احساسات مردم نسبت به افراد مشهور مانند ایلان ماسک در شبکه‌های اجتماعی توییتر، ردیت و تلگرام است. داده‌های جمع‌آوری شده برای این تحقیق نیز از سال ۲۰۲۱ و ۲۰۲۲ می‌باشد. وجه تمایز این تحقیق در استفاده از شاخص‌های تکنیکال، اقتصادی و تحلیل احساسات به صورت همزمان برای پیش‌بینی روند قیمت است. از روش حذف ویژگی بازگشتی برای انتخاب ویژگی استفاده گردیده و در مرحله بعد، با آزمایش ۸ الگوریتم طبقه‌بندی، نتایج مقایسه شده‌اند. در این تحقیق الگوریتم xgboost با رکورد دقت بدست آمده ۸۸ درصدی در پیش‌بینی روند تغییرات قیمت بیتکوین عملکرد عالی را نشان داد.

طبقه بندی JEL: C50 ،E42 ،G13

کلید واژه‌ها: بیتکوین، ارز دیجیتال، تحلیل احساسات، الگوریتم های یادگیری ماشین

۱. مقدمه

امروزه ارزهای دیجیتال به یک موضوع محبوب در حوزه‌های اقتصاد و سرمایه گذاری تبدیل شده‌اند . آنچه به عنوان یک ارز دیجیتال یا مجازی در نظر گرفته می‌شود که از فناوری رمزگذاری برای امنیت خود استفاده می‌کند. بنابراین کلاهبرداری و جعل بسیار مشکل بوده و حتی در صورت وجود نیز به راحتی قابل شناسایی و ردیابی می‌باشد. بیتکوین یکی از انواع ارزهای رمزنگاری شده است که به‌دلیل سادگی، شفافیت و محبوبیت روزافزون توجه بسیاری را به خود جلب کرده است. بیتکوین اولین بار توسط ناکاموتو[1](۲۰۰۸) طراحی و معرفی شد، سپس در سال ۲۰۰۹ وارد بازار آنلاین شد. بیتکوین اولین ارز دیجیتال جهان بود که در سال‌های اولیه مورد استقبال قرار نگرفت. اما در سال‌های اخیر توجه بسیاری از سرمایه گذاران، سیاست‌گذاران و رسانه‌ها را به خود جلب کرده است.

محققان در زمینه داده کاوی و یادگیری ماشینی علاقه‌مند به ارائه راه‌های جدید برای پیش‌بینی قیمت بیتکوین و نوسانات آن هستند. روش‌های پیش بینی سنتی و مرسوم برای مواردی مانند قیمت فروش تحت تاثیر عوامل فصلی بسیار مناسب است. گائو و همکاران (۲۰۱۸) همچنین پی بردند استفاده از روش‌های سنتی به دلیل عدم وجود قابلیت فصلی در بازار ارزهای رمزپایه و نوسانات زیاد آنها چندان مؤثر نیست. الگوریتم‌های جدیدی برای پیش بینی به همراه پیشرفت‌های اخیر در قدرت محاسباتی کامپیوتر و مهمتر از آن توسعه الگوریتم های پیشرفته یادگیری ماشین و رویکردهایی مانند یادگیری عمیق طراحی شده اند. مطالعات قبلی قیمت بیتکوین را به دو صورت پیش‌بینی کرده اند: تجزیه و تحلیل تجربی و تجزیه و تحلیل الگوریتم های یادگیری ماشینی. الگوریتم های یادگیری ماشین به طور گسترده ای برای پیش بینی های دقیق در بسیاری از زمینه ها از جمله امور مالی و برای توسعه سیستم های کمک به تصمیم گیری در معاملات بررسی شده است. ژنگ و جین[2](۲۰۱۷) از رگرسیون لجستیک، وانگ و لئو[3] (۱۹۹۶) از میانگین متحرک خودهمبسته یکپارچه[4]، اولاه[5] و همکاران (۲۰۲۱) از درخت تصمیم، چن[6] و همکاران (۲۰۱۷) از شبکه عصبی حافظه کوتاه مدت طولانی و چن و همکاران (۲۰۰۳)

---

1. Nakamoto
2. Zheng & Jin
3. Wang & Leu
4. Autoregressive integrated moving average
5. Ullah
6. Chen

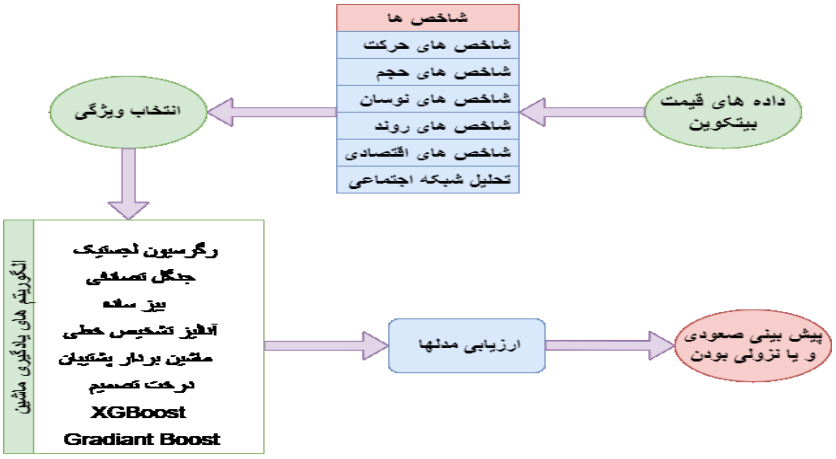
از شبکه عصبی استفاده کردند. با یادگیری جزئیات نمونه‌های گذشته، می‌توان برنامه‌ها و مدل‌های یادگیری ماشینی را توسعه داد که بر اساس داده‌های آموزشی پیش‌بینی می‌کنند. یک سوال که ممکن است هنگام پیش‌بینی قیمت بیت‌کوین با استفاده از الگوریتم‌های یادگیری ماشینی مطرح شود این است که چه ویژگی‌هایی باید در نظر گرفته شود؟ اگرچه روش‌های بیشتری در مورد انتخاب ویژگی وجود دارد که ابوعلیگاه و هنانده[1] (۲۰۱۵) و ابوعلیگاه و همکاران (۲۰۱۸) به آن پرداختند. ما این مساله را با ادغام نتایج کارهای تجربی اخیر توسط محققان با درک عمیق عوامل موثر بر قیمت بیتکوین حل می کنیم. به طور خاص، ما از ۸۰ اندیکاتور که بر روی داده‌های موجود بیتکوین اعمال شده است استفاده می کنیم و همچنین به بررسی حجم واکنش‌ها در شبکه‌های اجتماعی توییتر ، تلگرام و ردیت نیز می‌پردازیم که این داده‌ها نشان‌دهنده بازخورد فعالیت افراد مشهور در این زمینه است. از جمله معروفترین و تاثیرگذارترین افراد میتوان به آقای ایلان ماسک رییس شرکت تسلا و آقای پاول رییس فدرال رزرو و آقای چانگ پنگ مدیرعامل صرافی بایننس نام برد. بنابراین نیاز به یافتن روشی وجود دارد که بتواند به‌طور دقیق از الگوریتم‌های یادگیری ماشین برای پیش‌بینی قیمت بیتکوین استفاده کند. از آنجایی‌که استفاده بیش از ۱۰۰ ویژگی برای آموزش مدل‌های یادگیری ماشین باعث برازش بیش از حد ماشین می شود، ما درنظر گرفتیم تا از روش انتخاب ویژگی حذف بازگشتی، برای کاهش ابعاد مساله استفاده کنیم. بنابراین، الگوریتم‌های مختلف یادگیری ماشینی، از جمله ماشین‌های بردار پشتیبان، مدل‌های جنگل تصادفی و شبکه عصبی در مطالعات قبلی پیاده‌سازی شده‌اند. با این حال، کارهای قبلی به سادگی داده ها را بدون تمایز ویژگی های داده یا اندازه نمونه در مدل‌ها قرار می‌دهند.

در این پژوهش، با استفاده از الگوریتم ماشین بردار پشتیبان، به مهندسی ابعاد مساله می‌پردازیم. بعلاوه از رویکرد حذف ویژگی بازگشتی برای انتخاب ویژگی‌های مناسب استفاده کرده و با استفاده از ویژگی‌های منتخب، این مساله را حل می‌کنیم. ابتدا مجموعه داده‌ای که شامل اعمال شاخص‌های تکنیکال و بنیادین بر روی داده های روزانه بیتکوین است را بدست می‌آوریم و همچنین نتایج تحلیل حجم احساسات در شبکه‌های اجتماعی مختلف که ناشی از اخبار حاصل از قیمت بیتکوین و افراد مشهور است را به داده‌ها اضافه می‌کنیم. سپس، مهندسی ویژگی‌ها را انجام می‌دهیم. آنگاه، مدل‌های یادگیری ماشین، از جمله رگرسیون لجستیک و تحلیل تفکیک خطی و مدل‌های پیچیده‌تر یادگیری ماشین

[1]. Abualigah & Hanandeh

مانند جنگل تصادفی، XGBoost، ماشین بردار پشتیبان و درخت تصمیم را نیز پیاده می‌کنیم. در انتها، به مقایسه نتایج حاصل الگوریتم ها می‌پردازیم. تا جایی که ما مدانیم، این رویکرد با در نظر گرفتن ویژگی‌های ابعاد بالا برای مساله پیش‌بینی روند روزانه بیتکوین با ادغام بیش از ۱۰۰ شاخص تکنیکال و بنیادین تبلیغات رسانه‌ای و بازخورد اخبار افراد مشهور در این زمینه جدیدترین می‌باشد. در شکل شماره۱ رویکرد کلی در نظر گرفته شده را برای این تحقیق نشان داده‌ایم.

شکل شماره (۱): رویکرد کلی تحقیق



ماخذ: یافته‌های پژوهش

۲. پیشینه تحقیق

در اکثر تحقیقات، پژوهشگران با ایجاد رویکردهای متفاوت و مطرح کردن مسائل گوناگون در بازار سرمایه، به پیش‌بینی تغییرات قیمت پرداخته‌اند، اما به دلیل نوساناتی که در بازار سرمایه اتفاق می‌افتد، اغلب آن‌ها به نتایج قابل قبولی دست نیافته‌اند. اکثر تحقیقاتی که تاکنون در این زمینه انجام شده، دارای دو بخش هستند: ۱) پیش‌بینی قیمت و ۲) مدیریت سبد. در این پژوهش‌ها، از الگوریتم‌های یادگیری ماشین به خصوص یادگیری نظارت شده برای ایجاد ارتباط میان داده‌های ورودی و قیمت نهایی استفاده شده است. معمولا برای تحلیل داده‌های مربوط به بازارهای مالی یا ارزهای دیجیتال از اندیکاتورها استفاده می‌شود که سه تا از محبوب ترین شاخص‌های فنی که در تحلیل تکنیکال استفاده

می‌شوند عبارتنداز: واگرایی میانگین متحرک همگرایی[۱] که باز[۲] و همکاران (۲۰۱۵) به آن پرداختند و همچنین شاخص قدرت نسبی[۳] که وایلدر[۴] (۱۹۷۸) به آن پرداخت و میانگین متحرک و میانگین متحرک نمایی که هانتر[۵](۱۹۸۶) به استفاده از ان روی آورد. همچنین هوچریتر و اسمیدهوبر[۶] (۱۹۹۷) راه‌حل‌هایی از طریق کاهش ابعاد یادگیری عمیق، از طریق شبکه عصبی مصنوعی، انتخاب ویژگی و استخراج ویژگی از طریق شبکه‌های عصبی مصنوعی به کار گرفته اند و با استفاده از این رویکردها برخی از نویز موجود در داده‌های بازار را حذف می‌کنند. کرائوسا و فیوریگال[۷](۲۰۱۷) پیشنهاد می کنند که می توانند اخبار مالی را از طریق یادگیری عمیق در تصمیم گیری در بازارهای مالی استفاده کنند. اخبار بازارهای مالی ، پیام‌های مالی و روند احساسات در بازار را نشان می‌دهد. داده‌های کلان اقتصادی نیز در تحلیل‌های بازار مورد استفاده قرار می‌گیرد که این نوع داده‌ها شامل متغیرهای کلان اقتصادی است. ممکن است بر عناصر بازار مالی مانند نرخ ارز، نرخ بهره و ذخایر ارزی ناخالص بانک مرکزی تأثیر بگذارد. برای تحلیل اخبار بازارهای مالی نیز تحقیقاتی انجام‌شده است که اولاح و همکاران(۲۰۲۲) در تحقیق خود به این مساله پی بردند که توییت های مثبت افراد مشهور و احساسات مثبت دولت نسبت به بیتکوین به طور قابل توجهی با تغییرات مثبت در قیمت بیتکوین مرتبط است. برای تحقیقات بیشتر می‌توان به سویم[۸] و همکاران(۲۰۱۴)، بائو[۹] و همکاران(۲۰۱۷) و کیم و وون[۱۰](۲۰۱۸) رجوع کرد.

کشوری و ملایی(۱۳۹۸) در تحقیق خود با استفاده از رویکرد تحلیل استواری به پیش‌بینی قیمت بیت کوین پرداختند و با آموزش شبکه عصبی با رویکرد استواری ارتباط مشخصی بین ورودی‌ها و خروجی‌ها به دست آوردند به طوری‌که کمترین میزان خطا در رابطه پیشنهادی وجود داشته باشد. همچنین آنها در تحقیق خود از الگوریتم کرم شب تاب و الگوریتم خفاش استفاده کردند. همچنین بشیری و پاریاب(۲۰۲۰) در مقاله خود با استفاده از مجموعه داده ده رمز ارز به پیش‌بینی قیمت بیتکوین پرداختند که الگوریتم‌های

---

1. Moving Average Convergence Divergence
2. Baz
3. Relative Strength Index
4. Wilder
5. Hunter
6. Hochreiter & Schmidhuber
7. Kraus & Feuerriegel
8. Sevim
9. Bao
10. Kim & Won

ماشین بردار پشتیبان، جنگل تصادفی، گرادیان تقویتی و شبکه عصبی مقدار صحت ۵۲,۱۶۷۵ درصد را ثبت کردند. کورباسز[1] (۲۰۱۹) از شبکه‌های عصبی پیشخور با یک لایه پنهان برای پیش‌بینی قیمت بیتکوین استفاده کردند. یافته‌های آنها نشان می‌دهد که این شبکه عصبی دارای دقت بالایی (۶۰,۰۵ درصد) برای طبقه‌بندی حرکت قیمت در مقایسه با مطالعات دیگر مانند مک نالی[2](۲۰۱۸) است. برای پیش‌بینی قیمت بیتکوین، چن و همکاران(۲۰۱۹) مدل اریما را با شبکه عصبی پس انتشار[3] با در نظر گرفتن روابط خطی و غیرخطی روی مجموعه داده مقایسه کردند. هیو[4] و همکاران (۲۰۱۸) داده‌های قیمت ارزهای دیجیتال را از طریق رابط برنامه‌نویسی کاربردی جمع آوری و پردازش کردند. سپس از یک مدل یادگیری ماشینی به نام تقویت گرادیان[5] استفاده کردند. این روش به مدل اجازه می‌دهد تا تغییرات در داده‌های قیمت را یاد بگیرد. آنها دریافتند که حداکثر دقت مدل حدود ۶۳ درصد است. چن و همکاران (۲۰۲۰) از دو روش آماری LR و تجزیه و تحلیل تشخیص خطی و پنج مدل یادگیری ماشینی شامل جنگل تصادفی، XGBoost، تجزیه و تحلیل متمایز درجه دوم، ماشین بردار پشتیبان و شبکه عصبی برای پیش‌بینی قیمت بیتکوین استفاده کردند. نتایج آنها نشان می‌دهد که میانگین دقت روش‌های آماری (۶۵ ٪) بالاتر از میانگین دقت مدل های یادگیری ماشینی (۵۵/۳ ٪) برای قیمت روزانه بیتکوین است. همچنین برای قیمت بازه ۵ دقیقه‌ای بیتکوین، مدل های یادگیری ماشینی دقت بهتری نسبت به روش‌های آماری دارند که به ترتیب ۶۲/۲ و ۵۳ درصد بوده و شبکه عصبی حافظه کوتاه‌مدت بلندمدت بهترین دقت (۶۷٪/۲) را نسبت به سایر روش‌ها کسب کرده است.

تحقیقات پیشین بیشتر به استفاده از الگوریتم های یادگیری ماشین بر روی داده های قیمت بیتکوین پرداخته اند. اما این الگوریتم ها در بعضی از اوقات در مواجهه با داده ها، دچار بیش پردازش شده‌اند. همچنین از تحلیل احساسات کاربران توییتر برای پیش‌بینی قیمت بیتکوین استفاده شده است. ولی تا به حال به پیش بینی روند قیمتی بیتکوین با استفاده از تعداد بالایی از اندیکاتورها و همچنین تحلیل شبکه‌های اجتماعی توییتر، تلگرام و ردیت با تاثیر گذاری سلبریتی ها پرداخته نشده است.

---

1. Kurbucz
2. McNally
3. .backpropagation network
4. Heo
5. gradient boosting

۳. روش تحقیق

پیش‌بینی قیمت بیتکوین با استفاده از الگوریتم‌های یادگیری ماشینی یک مساله کاربردی است. بسیاری از تحقیقات موجود به سادگی بر دقت بالاتر بدون در نظر گرفتن ابعاد نمونه اشاره می‌کنند بدون این که به بیش پردازش الگوریتم توجه کنند. ما در این مطالعه به مهندسی ابعاد این مساله توجه داریم و سپس از الگوریتم‌های طبقه‌بندی کننده استفاده می‌کنیم تا به دقت بالایی دست یابیم.

۳-۱. ویژگی‌ها

همانطور که در قسمت‌های قبل شرح داده شد، رویکرد اصلی این تحقیق انتخاب ویژگی‌های مناسب برای الگوریتم‌های یادگیری ماشین با هدف پیش‌بینی قیمت بیتکوین است. ما به طور جداگانه مجموعه ویژگی‌های مناسب را برای پیش‌بینی قیمت روزانه بیتکوین از میان ۱۱۰ ویژگی با استفاده از مهندسی ویژگی انتخاب می‌کنیم. چهار نوع ویژگی برای پیش‌بینی قیمت روزانه وجود دارد. اولین ویژگی، داده‌های شبکه و دارایی اصلی بیتکوین، مانند قیمت باز شدن، بسته شدن، بیشترین قیمت، کمترین قیمت و حجم معاملات در روز است که می‌تواند در پیش‌بینی قیمت روزانه مفید باشد. نوع دوم، مربوط به شاخص‌ها می‌باشد که بر روی داده‌های نوع اول اعمال شده است. نوع سوم، توجه رسانه‌ها، سرمایه‌گذاران، و افراد مشهور است. در بازارهای مالی، احساسات نقش مهمی را برای پیش بینی قیمت ایفا می‌کند. بنابراین با به دست آوردن حجم احساسات در رسانه‌های توییتر، تلگرام و ردیت به پیش بینی روند بازار می‌پردازیم. در ادامه به توضیح ویژگی آخر می‌پردازیم.

۳-۱-۱. تحلیل احساسات

تحلیل احساسات در بازارهای مالی همان درک احساسات آن بازار و تصمیم‌گیری بر اساس آن می‌باشدکه حاصل ظهور رسانه‌های اجتماعی است. امروزه می‌توان با بهره‌گیری از محتوای رسانه‌های اجتماعی به تجزیه و تحلیل احساسات هر چیزی پرداخت که از اوایل سال ۲۰۰۰، تجزیه و تحلیل احساسات به یکی از فعال ترین حوزه‌های تحقیقاتی در پردازش زبان طبیعی[1] تبدیل شده است(مدهت[2] و همکاران ،۲۰۱۴). از شاخص‌هایی که در این تحقیق از آنها استفاده شده می‌توان به موارد زیر اشاره کرد:

---

[1]. natural language processing
[2]. Medhat

۱) شاخص ترس و طمع بیتکوین: شاخص ترس و طمع برای بازار ارز دیجیتال جزو محبوب ا است. این شاخص بطور روزانه به روز شده و باید در بازه زمانی روزانه از آن استفاده کرد که از این شاخص به عنوان معیاری برای سنجش احساسات سرمایه‌گذار و نوسانات بیت‌کوین در سطوح مختلف استفاده می‌شود(موکنی[1] و همکاران،۲۰۲۲). در جدول شماره ۱ نیز این شاخص ها به نمایش درآمده‌اند.

جدول شماره (۱) : انواع شاخص‌های تاثیر گذار در شناسایی ترس و طمع

طمع	ترس	
کم	زیاد	نوسانات
زیاد	کم	حجم معاملات
مثبت	منفی	تحریک احساسات رسانه‌های اجتماعی
کم	زیاد	تسلط بیتکوین
زیاد	کم	روند جستجوی گوگل

ماخذ: https://www.investopedia.com/terms/f/fear-and-greed-index.asp

۲) شاخص بازار گاوی و خرسی بیتکوین: این شاخص مبتنی بر احساسات رسانه‌های اجتماعی بیتکوین است که شامل پلتفرم های رسانه های اجتماعی می شود.

۳) تحلیل احساسات شبکه‌های اجتماعی و اخبار پایه‌ای: با گسترش استفاده از شبکه های اجتماعی و سهولت کاربرد آنها، افراد مشهور می‌توانند تاثیرات فراوانی روی بازار های مالی بگذارند. از جمله این افراد می‌توان به افراد سیاسی مشهور، اقتصاددانان، سرمایه داران بازارهای مالی و مدیران شرکت‌های مالی و صرافی‌ها اشاره کرد. همچنین مبانی اخبار پایه‌ای شامل موضوعاتی است که به طور کلی بر اقتصاد تأثیر مگذارد، از جمله آمارهای مربوط به بیکاری، عرضه و تقاضا، رشد و تورم، و همچنین ملاحظاتی برای سیاست‌های پولی یا مالی و تجارت بین المللی می‌باشد. همچنین می‌توان در شکل شماره ۲ حجم شبکه های اجتماعی برای ایلان ماسک مالک شرکت تسلا و از افراد تاثیرگذار در بازار ارزهای دیجیتال را در مقایسه با قیمت بیتکوین از تاریخ ۲۰۲۱/۱/۵ الی ۲۰۲۲/۵/۲۰ را مشاهده نمود.

[1] Mokni

شکل شماره (۲): حجم شبکه‌های اجتماعی برای ایلان ماسک

ماخذ: https://app.santiment.net

در جدول شماره ۲ نیز به شرح شاخص‌های مورد استفاده برای این مساله پرداخته‌ایم که با استفاده از کتابخانه ta در پایتون این شاخص‌ها بدست آمده‌اند. این کتابخانه برای تجزیه و تحلیل تکنیکال و انجام مهندسی ویژگی بر روی مجموعه داده‌های مالی مورد استفاده محققین این حوزه قرار می‌گیرد. با استفاده از ابزارهای این کتابخانه به جمع‌آوری داده های مربوط به شاخص‌های تکنیکالی قیمت بیتکوین می‌پردازیم.

۳-۲. مهندسی ویژگی‌ها

امکان تجزیه و تحلیل همه ابعاد در سطح کوچک در داده‌های با ابعاد زیاد، وجود ندارد. ممکن است روزها یا ماه‌ها طول بکشد تا تجزیه و تحلیل معناداری انجام شود که به هزینه زیادی نیاز خواهد داشت. استفاده از داده‌هایی با ابعاد بالا مشکلاتی را به دنبال خواهد داشت، مانند: با افزایش ابعاد، فضای مورد نیاز برای ذخیره اطلاعات افزایش می‌یابد، با افزایش ابعاد، امکان بیش برازش مدل نیز افزایش می‌یابد و با ابعاد بیشتر، پیچیدگی زمانی بیشتری نیز در آموزش مدل وجود دارد(برونلی[1]، ۲۰۲۰). ما در این مساله از روش حذف ویژگی بازگشتی برای کاهش تعداد ویژگی‌های مساله استفاده کرده‌ایم. به این صورت که ابتدا، برآوردگر روی مجموعه اولیه ویژگی‌ها آموزش می‌بیند و اهمیت هر ویژگی به دست می‌آید. سپس، کم اهمیت‌ترین ویژگی‌ها از مجموعه ویژگی‌های فعلی حذف می‌شوند. این روش به صورت بازگشتی بر روی مجموعه هرس شده تکرار می‌شود تا در نهایت به تعداد

[1] Brownlee

مورد نظر ویژگی برای انتخاب برسد. در این مدل از الگوریتم ماشین بردار پشتیبان به عنوان حل کننده برای انتخاب ۱۰ ویژگی برتر استفاده شده است.

۳-۳. دیتاست

مجموعه داده استفاده شده در این پژوهش مربوط به داده‌های روزانه قیمت بیتکوین برای سال‌های ۲۰۲۱ و ۲۰۲۲ است که از ۲۰۲۱/۱/۱ الی ۲۰۲۲/۵/۲۰ جمع آوری شده است. از تاریخ ۲۰۲۱/۱/۱ تا ۲۰۲۱/۱۲/۳۱ برای آموزش الگوریتم‌ها، و از تاریخ ۲۰۲۲/۱/۱ تا ۲۰۲۲/۵/۲۰ برای آزمایش استفاده کرده‌ایم. این داده‌ها از سایت Yahoo finance دریافت شده است. با استفاده از کتابخانه ta در پایتون به جمع‌آوری شاخص‌های مدنظر که در جدول شماره ۲ به آن‌ها اشاره شده است، پرداختیم. همچنین برای دریافت داده های تحلیل احساسات مربوط به سلبریتی‌ها در شبکه‌های اجتماعی توییتر، تلگرام و ردیت از سایت https://app.santiment.net استفاده نمودیم.

جدول شماره(۲): شاخص ها

نام شاخص‌ها	تعریف	انواع شاخص‌ها
Percentage Price Oscillator _ Percentage Volume Oscillator _ Rate of Chang _ Relative Strength Index _ Stochastic RSI _ True strength index _ Awesome Oscillator _ Ultimate Oscillator _ Williams %R _ Stochastic RSI %d _ Stochastic RSI %k _ Stochastic Oscillator Signal , ...	شاخص‌های حرکت ابزارهای تحلیل تکنیکال هستند که برای تعیین قدرت یا ضعف قیمت در بازارهای مالی استفاده می‌شوند. مومنتوم میزان افزایش یا کاهش قیمت ارزهای دیجیتال را اندازه گیری می‌کند. شاخص‌های متداول حرکت شامل شاخص قدرت نسبی و واگرایی میانگین متحرک[1] است.	شاخص‌های حرکت[2]
Accumulation/Distribution Index (ADI) _ Chaikin Money Flow _ Force Index _ Money Flow Index _ Negative Volume Index _ On-balance volume _ Volume-price trend _ Volume Weighted Average Price _ Ease of movement , ...	شاخص‌های حجم ابزارهای فنی برای تشخیص و ارزیابی قدرت بازارهای صعودی و نزولی هستند. رویکرد این شاخص ها بیشتر به این صورت است که به فشار خرید در مقابل فروش نگاه می‌کنند تا مشخص کنند کدام طرف کنترل قیمت را در دست دارد.	شاخص‌های حجم[3]

[1]. moving average convergence divergence (MACD)
[2]. Momentum Indicators
[3]. Volume Indicators

Average True Range _ Bollinger Bands _ Bollinger Channel High Band _ Bollinger Channel Low Band _ Bollinger Channel Middle Band _ Bollinger Channel Percentage Band _ Bollinger Channel Band Width _ Donchian Channel _ Donchian Channel High Band _ Keltner Channel Indicator Crossing High Band , ...	کاربرد شاخص نوسانات در این است که میزان فاصله از میانگین قیمت خود را اندازه گیری می‌کند که می‌تواند بالاتر و پایین تر از حد مورد نظر باشد. پراکندگی بازده ها را در طول زمان در قالبی بصری محاسبه می‌کند که محققین از آن برای سنجش اینکه آیا این ورودی ریاضی در حال افزایش یا کاهش است استفاده می‌کنند.	شاخص‌های نوسان[1]
Average Directional Movement Index _ Aroon Indicator _ Commodity Channel Index _ Detrended Price Oscillator _ Exponential Moving Average _ Ichimoku Kinkō Hyō _ KST Oscillator _ Moving Average Convergence Divergence _ Mass Index _ Parabolic SAR _ Simple Moving Average , ...	روند، جهت کلی بازار را نشان می‌دهد. در تحلیل تکنیکال، روندها با خطوط روند یا عملکرد قیمت مشخص می‌شوند که وقتی قیمت در حال ایجاد اوج‌های نوسانی بالاتر و نوسان پایین‌تر برای یک روند صعودی، یا نوسان پایین‌تر و نوسان پایین‌تر برای یک روند نزولی است، نشان داده می‌شوند.	شاخص‌های روند[2]
Real gdp _ real gdp per capita _ treasury yield _ federal funds rate _ cpi _ inflation _ inflation expectation _ consumer sentiment _ retail sales _ durables _ unemployment _ nonfarm payroll	یک شاخص اقتصادی یک اندازه گیری کلان اقتصادی است که توسط تحلیلگران برای درک فعالیت ها و فرصت‌های اقتصادی فعلی و آتی استفاده می‌شود. پرکاربردترین شاخص‌های اقتصادی مربوط به داده‌های اقتصادی ایالات متحده آمریکا می‌باشد که تاثیر زیادی بر اقتصاد دیگر کشورها می‌گذارد.	شاخص‌های اقتصادی[3]

ماخذ: یافته های تحقیق

۳-۴. الگوریتم‌های یادگیری ماشین

در این پژوهش از معروف ترین الگوریتم‌های یادگیری ماشین استفاده شده که در زیر به تشریح برخی از آنها پرداخته‌ایم.

الف) جنگل تصادفی[4]

در الگوریتمِ جنگل تصادفی از چندین درختِ تصمیم استفاده می‌شود. در واقع مجموعه‌ای از درخت‌های تصمیم، با هم یک جنگل را تولید می‌کنند و این جنگل می‌تواند تصمیم‌های بهتری را (نسبت به یک درخت) اتخاذ کند. در این روش مجموعه‌ای از درخت‌های تصمیم وجود دارند که به هر کدام از آن‌ها یک زیر مجموعه‌ای از داده‌ها تزریق می‌شود. هر کدام از الگوریتم‌ها عملیاتِ یادگیری را انجام می‌دهند. در نهایت، می‌تواند با استفاده از

[1]. Volatility Indicators
[2]. Trend Indicators
[3]. Economic Indicators
[4]. Random Forest

رای‌گیری، آن طبقه‌ای را که بیشترین رای را آورده است انتخاب کرده و به عنوانِ طبقه‌ی نهایی جهت انجامِ عملیات طبقه‌بندی قرار دهد (بریمن[1]،۲۰۰۱).

ب) الگوریتم طبقه بندی بیزی[2]

ریش[3](۲۰۰۱) در تحقیق خود به بیز ساده به‌عنوان یک الگوریتم طبقه‌بندی ساده اما مؤثر و متداول یادگیری ماشین اشاره کرد که در دسته‌ی یادگیری با ناظر[4]جای می‌گیرد. بیز ساده الگوریتمی احتمالی است که براساس نظریه‌ی بیز برای طبقه‌بندی استفاده می‌شود. روش طبقه‌بندی مبتنی بر قضیه بیز بر فرض استقلال در بین پیش‌بینی کننده‌ها استوار است.

ج) الگوریتم رگرسیون لجستیک[5]

در واقع مدل رگرسیون لجستیک با استفاده از تبدیل لگاریتم طبیعی به صورت نمایی بیان می‌شود. رگرسیون لجستیک خود به سه دسته دودویی ، اسمی و ترتیبی تقسیم بندی می‌شود که در رگرسیون باینری، تعداد این طبقه‌بندی کننده دو تا است زیرا متغیر پاسخ پیوسته نبوده و دو مقدار صفر و یک را می گیرد(رایت[6]، ۱۹۹۵). تابع لجستیک نیز در رابطه (۱) نشان داده شده است که به صورت زیر می باشد که در آن Y متغیر پاسخ می‌باشد .

$$P(Y=1|X_1, X_2, \dots, X_m) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\cdots+\beta_m X_m}} \quad (۱)$$

مدل رگرسیون لجستیک با گرفتن لگاریتم از نسبت شانس حاصل می‌شود که نسبت شانس و متعاقبا رگرسیون لجستیک چند متغیره در روابط (۲) و (۳) نشان داده شده است که نسبت شانس ، به صورت نسبت احتمال آن‌که یک پاسخ حاصل شود، در مقابل آن‌که سایر پاسخ ها اتفاق بیفتد، تعریف می‌شود.

$$Odds\ (Y=1) = \frac{P}{1-P} \quad (۲)$$

$$Log(Odds) = \beta_0 + \beta_1 X_1 + \cdots + \beta_m X_m \quad (۳)$$

---

1. Breiman
2. Naive Bayes Classifiers
3. Rish
4. Supervised Learning
5. Logistic Regression
6. Wright

د) الگوریتم آنالیز تشخیصی خطی[1]

«آنالیز تشخیصی خطی» یک روش آماری برای کاهش ابعاد یک مسئله و تشخیص دسته‌ها بوسیله بیشینه‌سازی نسبت «پراکندگی بین گروه‌ها» به «درون گروه‌ها» است. هدف این الگوریتم آن است که توزیع داده ها را براساس متغیرهای پیش بینی کننده در هر دسته مدل کند (با فرض معلوم بود Y) و سپس احتمال شرطی$P(Y=1 \mid X=x)$ را محاسبه کند(ایزنمن[2]،۲۰۱۳).

ه) الگوریتم ماشین بردار پشتیبان[3]

بردارهای پشتیبان به زبان ساده، مجموعه ای از نقاط در فضای n بعدی داده ها هستند که مرز دسته‌ها را مشخص می‌کنند و مرزبندی و دسته‌بندی داده‌ها براساس آن‌ها انجام می شود و با جابجایی یکی از آنها، خروجی دسته‌بندی ممکن است تغییر کند. سپس داده‌ها را به ابعاد بالا منتقل می‌کند و از رویکرد طبقه بندی‌کننده خود برای پیدا کردن صفحه‌ای استفاده می‌کند که بتواند دسته ها را جدا کند و برای ارتقاء ابعاد داده می‌توان از توابع مختلف مانند چندجمله ای درجه n استفاده کرد (نوبل[4]،۲۰۰۶).

و) XGBOOST

دسته ای از الگوریتم‌های توان افزایی[5] است که یادگیرندگانِ ضعیف را به یادگیرندگان قوی تبدیل می‌کند. الگوریتم توان افزایی با آموزشِ یک یادگیرنده‌یِ پایه شروع می‌شود و سپس توزیع نمونه‌های آموزشی را با توجه به نتیجه‌یِ یادگیرنده‌یِ پایه تنظیم می‌کنند تا نمونه های دسته بندیِ نادرست مورد توجه بیشتر یادگیرندگانِ پایه یِ بعدی قرار گیرد. پس از آموزشِ اولین یادگیرنده‌یِ پایه، یادگیرنده‌یِ پایه دوم با نمونه های آموزشی تنظیم شده آموزش می‌بیند و از نتیجه برای تنظیم مجدد توزیع نمونه آموزشی استفاده می‌شود. چنین فرآیندی تکرار میشود تا زمانی که تعداد یادگیرندگانِ پایه به مقدار از پیش تعریف شده T برسد و در نهایت این یادگیرندگانِ پایه وزن و ترکیب شوند. الگوریتم XGBoost یک پیاده سازی از تقویت گرادیان درخت تصمیم گیری است که برای سرعت و کارایی بالا طراحی شده است (چن و گوسترین[6]،۲۰۱۶).

---

[1]. Linear Discriminant Analysis
[2]. Izenman
[3]. Support Vector Machine Algorithm
[4]. Noble
[5]. boosting
[6]. Chen & Guestrin

۳-۵. شاخص ارزیابی

ما در تحلیل نتایج این الگوریتم‌ها، بعد از بدست آوردن ماتریس در هم ریختگی مربوط به هر کدام از آن‌ها، به محاسبه چهار معیار دقت ، صحت ، حساسیت و معیار f1 پرداخته-ایم که در آن TP تعداد مثبت‌های واقعی، TN تعداد منفی‌های واقعی، FP تعداد مثبت کاذب، و FN تعداد منفی کاذب است. از مدل‌های یادگیری ماشین، ماتریس سردرگمی می سازیم و نتایج را به چهار دسته تقسیم می‌کنیم. این ماتریس حاوی نتایج آماری در مورد داده‌های طبقه بندی واقعی و پیش بینی های انجام شده توسط الگوریتم های یادگیری ماشین است(لوکیو[1] و همکاران،۲۰۱۹). در در جدول شماره ۳ این ماتریس نشان داده شده است.

جدول شماره (۳): ماتریس سردرگمی

		پیش بینی	
		۱	۰
واقعی	۱	TP	TN
	۰	FP	FN

ماخذ: (لوکیو و همکاران، ۲۰۱۹)

معیار دقت[2] معیاری است که به ما نشان می‌دهد چند درصدِ «بله» های الگوریتم درست بوده و از طریق رابطه(٤) محاسبه می گردد:

$$Precision = TP / TP + FP \quad (۴)$$

پارامتر صحت[3]، متداول‌ترین، اساسی‌ترین و ساده‌ترین معیار اندازه‌گیری کیفیت یک دسته‌بندی است و عبارت است از میزان تشخیص صحیح دسته‌بندی در مجموع دو دسته. این پارامتر در واقع نشان‌گر میزان الگوهایی است که درست تشخیص داده شده‌اند و از طریق رابطه (٥) تعریف می‌شود:

$$Accuracy = (TP+TN) / (TP+FN+FP+TN) \quad (۵)$$

معیار دیگر، حساسیت[4] یا نرخ پاسخ‌های مثبت درست می‌باشد. حساسیت به معنی نسبتی از موارد مثبت است که آزمایش آن‌ها را به درستی به عنوان نمونه مثبت تشخیص داده است و از طریق رابطه(۶) محاسبه می‌شود :

---

1. Luque
2. .Precision
3. Accuracy
4. Recall

$$recall\ (TPR) = TP / (TP+FN) \quad (۶)$$

معیار مهم دیگری به نام «معیار اف[1]» وجود دارد که برای ارزیابی عملکرد دسته‌بندی‌ها بسیار مورد استفاده قرار می‌گیرد و از ترکیب دو پارامتر حساسیت و دقت حاصل می‌شود. «معیار اف» به صورت رابطه (۷) تعریف می‌شود:

$$F\text{-measure} = 2 * (recall * Precision) / (recall + Precision) \quad (۷)$$

۴. برآورد مدل و آزمون فرضیه‌ها

در جدول ۴ عملکرد تمام مدل های یادگیری ماشین را در مورد قیمت روزانه بیتکوین نشان داده‌ایم. از نتایج، می‌توان متوجه شد که الگوریتم XGBoost عملکرد بهتری نسبت به سایر مدل‌ها دارد. معیار صحت و دقت این الگوریتم به ترتیب برابر ۸۸ و ۸۹ درصد می باشد و بعد از آن مدل Gradiant boost با اختلاف بسیار کمی بهتر می‌باشد که معیار صحت و دقت آن به ترتیب برابر با ۸۵ و ۸۷ درصد می باشد. اما از نظر حساسیت مدل آنالیز تشخیص خطی بهتر است زیرا معیار آن برابر با ۸۵ درصد بوده و از سایر مدل ها بیشتر است. معیار f1 نیز برای دو مدل XGBoost و Gradiant boost برابر می باشد. پوتاتوندا و راما[2] (۲۰۱۸) در مقاله خود به پارامترهای بهینه مورد استفاده در الگوریتم XGBoost پرداختند که ما در جدول شماره ۵ به برخی از این هایپرپارامترهای بهینه که در این پژوهش استفاده شده است، پرداختیم.

جدول شماره(۴): نتایج ارزیابی الگوریتم های یادگیری ماشین

الگوریتم ها	Accuracy	Precision	recall	F1
رگرسیون لجستیک	۷۹	۸۴	۷۲	۷۷
جنگل تصادفی	۸۳	۸۷	۷۸	۸۲
بیز ساده	۵۰/۳۷	۰	۰	۰
آنالیز تشخیصی خطی	۸۴	۸۳	۸۵	۸۴
ماشین بردار پشتیبان	۸۱	۸۲	۸۱	۸۱
Gradiant Boost	۸۵	۸۷	۸۲	۸۵
XGBoost	۸۸	۸۹	۸۲	۸۵
درخت تصمیم	۸۱	۸۴	۷۸	۸۱

ماخذ: یافته‌های تحقیق

[1]. F-measure
[2]. Putatunda & Rama

جدول شماره (۵): پارامترهای مدل XGBoost

gamma	learning_rate	max_depth	n_estimators	n_jobs
۰	۰,۳	٦	۱۰۰	٤

ماخذ: یافته‌های تحقیق

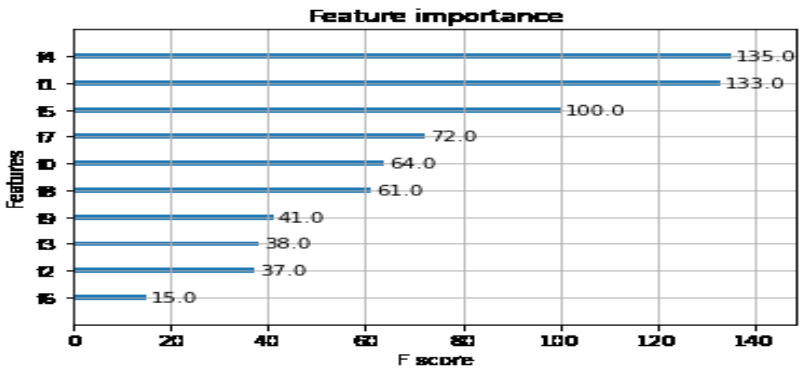
همچنین ۱۰ ویژگی منتخب که از روش حذف ویژگی بازگشتی بدست آمده است در جدول شماره ۶ نشان داده شده است که در شکل شماره ۳ نیز اهمیت هر کدام از ویژگی‌ها از منظر الگوریتم XGBoost به نمایش درآمده است.

جدول شماره (۶): ویژگی های استفاده شده در الگوریتم های یادگیری ماشین

شماره ویژگی	۱	۲	۳	٤	٥
نام ویژگی	Rsi9	Rsi3	ROC	sma10	Stochastic RSI
شماره ویژگی	٦	۷	۸	۹	۱۰
نام ویژگی	On-Balance Volume	Donchian Channel Middle Band	Aroon down Indicator	Aroon Up Indicator	Commodity Channel Index

ماخذ: یافته‌های تحقیق

شکل شماره (۳): اهمیت هر ویژگی با استفاده از الگوریتم XGBoost



ماخذ: یافته‌های تحقیق

همانطور که در شکل شماره ۳ نشان داده شده است، متغیرهای شاخص قدرت نسبی، شاخص قدرت نسبی تصادفی و شاخص تعادل حجم جزء مهمترین شاخص ها در پیش‌بینی روند توسط الگوریتم XGBoost شناخته شدند.

۶. نتیجه‌گیری و پیشنهادها

در این مطالعه، از الگوریتم‌های یادگیری ماشین بر اساس شاخص‌های متعدد برای پیش‌بینی صعودی و یا نزولی بودن قیمت بیت‌کوین استفاده شده است. همچنین نشان داده شد که می‌توان با استخراج ویژگی‌های بیشتر و انتخاب ویژگی‌های کارآمدتر، توسط روش انتخاب ویژگی بازگشتی، به دقت قابل قبولی از این پیش بینی رسید. در این پژوهش از شاخص‌های اقتصاد کلان و شبکه‌های اجتماعی با تاثیرگذاری افراد مشهور روی قیمت استفاده شده است. همچنین از ۱۰ ویژگی منتخب استفاده شده است که اسامی این ویژگی‌ها در جدول شماره ۶ نشان داده شده است که نشان‌دهنده تاثیر گذاری شاخص های تکنیکال در پیش بینی روند بیتکوین می‌باشد. برای پیش‌بینی صعودی یا نزولی بودن قیمت از روش‌های یادگیری ماشین استفاده شده است که به ترتیب الگوریتم‌های XGBoost، Gradiant Boosting و آنالیز تشخیص خطی، برای پیش‌بینی قیمت روزانه بیتکوین، عملکرد بهتری را نسبت به سایر الگوریتم‌ها نشان دادند. در این تحقیق از داده های سال ۲۰۲۱ برای آموزش و از داده های ۵ ماهه اول سال ۲۰۲۲ برای سنجش پیش بینی روند استفاده شد. دقت الگوریتم XGBoost برای این مسئله ۸۸ درصد بود. برای مطالعه جامع‌تر پیش‌بینی قیمت بیت‌کوین در آینده، جمع‌آوری داده‌های قیمت با جزئیات و ویژگی‌های مختلف با ابعاد بیشتر ضروری است. بعلاوه برای بهبود این مطالعه، می‌توان از مدل‌های یادگیری عمیق نیز استفاده کرد.

منابع و ماخذ:

بشیری، میثم و پاریاب، سید حسین (۱۳۹۸)، پیش‌بینی قیمت بیت کوین با استفاده از الگوریتم‌های یادگیری ماشین، اقتصاد کاربردی، ۱۰(۳۴ و ۳۵): ۱۳-۱.

کشوری، محمدمهدی و ملایی، احسان (۱۳۹۸)، پیش‌بینی قیمت بیت کوین با رویکرد تلفیقی شبکه عصبی و تحلیل استواری. اولین کنفرانس بین المللی مدیریت دانش، بلاکچین و اقتصاد، بیست و سوم آبان ۱۳۹۸، انجمن مدیریت دانش ایران، تهران.

Abualigah, L. M., Khader, A. T. & Hanandeh, E. S. (2018), A combination of objective functions and hybrid krill herd algorithm for text document clustering analysis, Engineering Applications of Artificial Intelligence, 73: 111-125 .

Abualigah, L. M. Q. &. Hanandeh, E. S. (2015), Applying genetic algorithms to information retrieval using vector space model, International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol, 5 .

Bao, W., Yue, J. & Rao, Y. (2017), A deep learning framework for financial time series using stacked autoencoders and long-short term memory, PloS one, 12(7): e0180944 .

Baz, J., Granger, N., Harvey, C. R., Le Roux, N. & Rattray, S. (2015), Dissecting investment strategies in the cross section and time series, Available at SSRN 2695101 .

Breiman, L. (2001), Random forests, Machine learning, 45(1): 5-32 .

Brownlee, J. (2020), Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python: Machine Learning Mastery.

Chen, A. S., Leung, M. T. & Daouk, H. (2003), Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index, Computers & Operations Research, 30(6): 901-923 .

Chen, C. C., Chang, J. H., Lin, F. C., Hung, J. C., Lin ,C. S. & Wang, Y. H. (2019), Comparison of forcasting ability between backpropagation network and ARIMA in the prediction of bitcoin price, Paper presented at the 2019 international symposium on intelligent signal processing and communication systems (ISPACS.(

Chen, G., Chen, Y. & Fushimi, T. (2017), Application of deep learning to algorithmic trading, Retrieved from

Chen, T. & Guestrin, C. (2016), Xgboost: A scalable tree boosting system, Paper presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.

Chen, Z., Li, C. & Sun, W. (2020), Bitcoin price prediction using machine learning: An approach to sample dimension engineering, Journal of Computational and Applied Mathematics, 365: 112395 .

Guo, T., Bifet, A. & Antulov-Fantulin, N. (2018), Bitcoin volatility forecasting with a glimpse into buy and sell orders. Paper presented at the 2018 IEEE international conference on data mining (ICDM).

Heo, J. S., Kwon, D. H., Kim, J. B., Han, Y. & An, C. H. (2018), Prediction of cryptocurrency price trend using gradient boosting, KIPS transactions on software and data engineering, 7(10): 387-396 .

Hochreiter, S. & Schmidhuber, J. (1997), Long short-term memory, Neural computation, 9(8): 1735-1780 .

Hunter, J. S. (1986), The exponentially weighted moving average, Journal of quality technology, 18(4): 203-210 .

Izenman, A. J. (2013), Linear discriminant analysis. In Modern multivariate statistical techniques (pp. 237-280): Springer.

Kim, H. Y. & Won, C. H. (2018), Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models, Expert Systems with Applications, 103: 25-37 .

Kraus, M. & Feuerriegel, S. (2017), Decision support from financial disclosures with deep neural networks and transfer learning, Decision Support Systems, 104: 38-48 .

Kurbucz, M. T. (2019), Predicting the price of bitcoin by the most frequent edges of its transaction network, Economics letters, 184: 108655 .

Luque, A., Carrasco ,A., Martín, A. & de Las Heras, A. (2019), The impact of class imbalance in classification performance metrics based on the binary confusion matrix, Pattern Recognition, 91: 216-231 .

McNally, S., Roche, J. & Caton, S. (2018), Predicting the price of bitcoin using machine learning, Paper presented at the 2018 26th euromicro international conference on parallel, distributed and network-based processing (PDP).

Medhat, W., Hassan, A. & Korashy, H. (2014), Sentiment analysis algorithms and applications: A survey, Ain Shams engineering journal, 5(4): 1093-1113 .

Mokni, K., Bouteska, A. & Nakhli, M. S. (2022),Investor sentiment and Bitcoin relationship: A quantile-based analysis, The North American Journal of Economics and Finance, 60:101657 .

Nakamoto ,S. (2008), Bitcoin: A peer-to-peer electronic cash system, Decentralized Business Review, 21260 .

Noble, W. S. (2006), What is a support vector machine?, Nature biotechnology, 24(12): 1565-1567 .

Putatunda, S. & Rama, K. (2018), A comparative analysis of hyperopt as against other approaches for hyper-parameter optimization of XGBoost. Paper presented at the Proceedings of the 2018 International Conference on Signal Processing and Machine Learning.

Rish, I. (2001), An empirical study of the naive Bayes classifier. Paper presented at the IJCAI 2001 workshop on empirical methods in artificial intelligence.

Sevim, C., Oztekin, A., Bali, O., Gumus, S. & Guresen, E. (2014), Developing an early warning system to predict currency crises, European Journal of Operational Research, 237(3): 1095-1104 .

Ullah, A., Imtiaz, F., Ihsan, M. U. M., Alam, M., Rabiul, G. & Majumdar, M. (2021), Combining machine learning classifiers for stock trading with effective feature extraction, arXiv preprint arXiv: 2107.13148 .

Ullah ,S., Attah-Boakye, R., Adams, K. & Zaefarian, G. (2022), Assessing the influence of celebrity and government endorsements on bitcoin's price volatility, Journal of Business Research, 145: 228-239 .

Wang, J.-H. & Leu, J.-Y. (1996), Stock market trend prediction using ARIMA-based neural networks. Paper presented at the Proceedings of International Conference on Neural Networks (ICNN'96).

Wilder, J. W. (1978), New concepts in technical trading systems: Trend Research.

Wright, R. E. (1995). Logistic regression .

Zheng, A. & Jin, J. (2017), Using ai to make predictions on stock market. cs229. stanford. edu .