



موسسه آموزش عالی زند شیراز

دانشکده مهندسی - نرم افزار کامپیوتر

طرح پیشنهادی پایان نامه کارشناسی ارشد


ترکیب تحلیل احساسات و رویکرد مبتنی بر شبکه های عصبی بازگشتی برای
تخمین و تحلیل رمزارزها

*The combination of sentiment analysis and an approach
based on recurrent neural networks for the estimation and
analysis of cryptocurrencies*

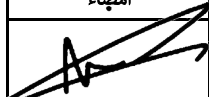
توسط :

علی منصورآبادی ۴۰۰۴۳۰۳۰۷

استاد / اساتید راهنما:

نام و نام خانوادگی	دانشکده	رتبه علمی	رشته - گرایش	امضاء
امیرحسین راسخ	مهندسی	استادیار	مهندسی کامپیوتر - نرم افزار	

استاد / اساتید مشاور:

نام و نام خانوادگی	دانشکده	رتبه علمی	رشته - گرایش	امضاء
علی قرمزبان	مهندسی	استادیار	علوم اعصاب شناختی - رایانش و هوش مصنوعی	

تاریخ:

۱۴۰۳/۰۲/۲۳

ارزهای رمزنگاری شده^۱، ارزهای مجازی^۲ هستند که بر اساس فناوری زنجیره بلوکی^۳ کار می‌کنند. از زمان معرفی بیت‌کوین^۴ در سال ۲۰۰۹، ارزهای دیجیتال به صورت گسترده‌ای در بازارهای مالی دیده می‌شوند. افراد در زمینه‌های متفاوتی از جمله تجارت ارزهای رمزنگاری و سرمایه‌گذاری در صندوق‌ها و شرکت‌های مربوطه فعالیت می‌کنند؛ طبق گزارش کوین‌مارکت‌کپ^۵، ارزش بازار جهانی ارزهای رمزنگاری شده در سپتامبر ۲۰۲۲، به مقدار تخمینی ۹۳۲.۴۹ میلیارد دلار رسیده است. اگرچه سرمایه‌گذاری‌ها به بازدهی پرسود انجامیده‌اند، اما نوسانات قیمتی پراکنده در اکثر ارزهای رمزنگاری، چالش‌ها و ریسک‌هایی را برای این بازار ایجاد کرده‌اند. به عنوان مثال، قیمت بیت‌کوین از زمان عرضه در بازار بسیار نوسانات داشته است و در سال‌های ۲۰۱۶ و ۲۰۱۷ اوج‌های ۱۲۲٪ و ۱۳۶۰٪ را به دست آورده است. اتریوم^۶، ایکس‌آرپی^۷ و لایت‌کوین^۸ نیز در سال ۲۰۱۷ نوسانات مشابهی را تجربه کرده‌اند [۱].

یک نکته مهم درباره ارزهای رمزنگاری شده این است که آن‌ها مستقل از بانک‌های سنتی هستند و توسط هیچ مرکزی کنترل نمی‌شوند، که این امر آن‌ها را از ارزهای سنتی متمایز می‌کند. از آنجایی که وجود زنجیره بلوکی برای ارزهای رمزنگاری لازم است، ادعا می‌شود که این ارزها همه ویژگی‌های زنجیره بلوکی را به ارث می‌برند. به عنوان مثال، بیت‌کوین به مردم راهی امن برای انجام معاملات دیجیتال به طور نیمه‌ناشناس^۹ فراهم می‌کند. فناوری زنجیره بلوکی توجه دولت‌های سراسر جهان را به خود جلب کرده است و انگیزه پشت این توجه، از نگرانی‌های مرتبط با جرائم نشات می‌گیرد. با این حال، شبکه بیت‌کوین بر روی الگوریتم‌های اثبات کار^{۱۰} و اثبات سهم^{۱۱} کار می‌کند که مصرف انرژی بالایی در فرآیندهای محاسباتی خود برای امن‌سازی شبکه می‌طلبند [۲].

اثبات کار الگوریتم توافقی است که در برخی از سیستم‌های زنجیره بلوکی مانند بیت‌کوین استفاده می‌شود که از کاربران می‌خواهد مقداری مشخص از کار محاسباتی را انجام دهند تا معاملات را تأیید کرده و آن‌ها را به زنجیره بلوکی اضافه کنند. این کار، که به عنوان استخراج شناخته می‌شود، معمولاً با استفاده از سخت‌افزارهای ویژه مانند مدارهای یکپارچه خاص برنامه‌ریزی شده^{۱۲} انجام می‌شود و مقدار قابل توجهی از انرژی را مصرف می‌کند. یک نقطه ضعف این سیستم‌ها این است که غالباً به خاطر مصرف انرژی بالا و متمرکز کردن قدرت

¹ Cryptocurrencies

² Virtual Currencies

³ Blockchain

⁴ Bitcoin (BTC)

⁵ Coinmarketcap

⁶ Ethereum (ETH)

⁷ Ripple (XRP)

⁸ Litecoin (LTC)

⁹ pseudo-anonymously

¹⁰ Proof of Work (PoW)

¹¹ Proof of Stake (PoS)

¹² Application Specific Integrated Circuits (ASICs)

استخراج مورد انتقاد قرار می‌گیرند. مکانیزم‌های جایگزین مانند اثبات‌سهم نیز هستند که نیازی به انجام کار محاسباتی برای تأیید معاملات ندارند و به جای این کار، کاربران مجبور هستند مقدار معینی از دارایی مورد نظر را نگه دارند. مزیت این‌گونه سیستم‌ها این است که با مصرف انرژی کمتر نسبت به سیستم‌های اثبات‌کار، گزینه موثرتری خواهند بود [۳].

سرمایه‌گذاران به یک رویکرد پیش‌بینی نیاز دارند تا بتوانند نوسانات قیمت ارزهای رمزنگاری شده را به‌طور موثری تحلیل کرده و ریسک را کاهش و سود را افزایش دهند. علاوه بر این، ممکن است از پیش‌بینی‌های نوسانات برای تخمین تغییرات قیمت آنها استفاده شود که برای توسعه و تحلیل استراتژی‌های مالی مفید است. با این حال، همانند پیش‌بینی قیمت سهام، که بازار آن پویا^{۱۳} و پیچیده است، پیش‌بینی قیمت ارزهای رمزنگاری شده نیز در حال حاضر به عنوان یکی از دشوارترین وظایف پیش‌بینی در حوزه مالی محسوب می‌شود [۴].

اکثر پژوهشگران، مسئله پیش‌بینی ارزهای دیجیتال را به عنوان یک مثال از پیش‌بینی سری زمانی^{۱۴} مطرح می‌کنند، زیرا که ایده اصلی این است که از داده‌های قیمت تاریخی^{۱۵} استفاده کرد تا قیمت‌های آینده را در یک دوره زمانی خاص در آینده پیش‌بینی کرد. تحلیل سری زمانی همچنین در پیش‌بینی هواشناسی و پیش‌بینی تقاضای خرده‌فروشی و تامین نیز استفاده شده است [۱].

رویکردهای یادگیری ماشین^{۱۶} می‌توانند الگوهای غیرخطی را استخراج کنند که در پیش‌بینی قیمت ارزهای دیجیتال بسیار مفید خواهد بود. با این حال، روش‌های سنتی یادگیری ماشین^{۱۷}، مانند شبکه‌های عصبی چندلایه^{۱۸} [۵] و ماشین‌های بردار پشتیبان^{۱۹} [۶] از برخی مشکلات مانند حساسیت به بیش‌برازش^{۲۰} رنج می‌برند و در استخراج الگوهای پنهان سطح بالا^{۲۱} از داده‌های توالی ارزهای رمزنگاری خوب عمل نمی‌کنند. برای رفع این مشکلات، مدل‌های پیش‌بینی مبتنی بر یادگیری عمیق^{۲۲} استفاده شده‌اند که عملکرد بهتری نسبت به روش‌های سنتی یادگیری ماشین را دارند [۷]. در کارهای اخیر نیز این موضوع اثبات شده است که شبکه‌های عصبی حافظه طولانی کوتاه مدت^{۲۳} و واحدهای بازگشتی دروازه‌ای^{۲۴} عملکرد به نسبت بهتری در پیش‌بینی قیمت ارزهای دیجیتال ارائه می‌دهند [۱].

¹³ Dynamic

¹⁴ Timeseries

¹⁵ Historical Data

¹⁶ Machine Learning (ML)

¹⁷ Traditional Machine Learning

¹⁸ Multi-layer perceptron (MLP)

¹⁹ Support Vector Machine (SVM)

²⁰ Overfit

²¹ High Level

²² Deep Learning (DL)

²³ Long Short-Term Memory (LSTM)

²⁴ Gated Recurrent Unit (GRU)

در تحقیقات دیگر بر روی روش‌های یادگیری گروهی^{۲۵} برای کاهش واریانس^{۲۶} و انحراف، با ترکیب مجموعه مدل‌های یادگیری ضعیف^{۲۷} مختلف تمرکز داشته‌اند [۸]. همچنین محققان به این نتیجه رسیده‌اند که مدل‌های یادگیری گروهی، عملکرد بهتری نسبت به مدل‌های یادگیری ماشین به صورت تکی دارند و اینکه مدل بوستینگ گرادیان^{۲۸} دقت بهتری نسبت به روش جنگل تصادفی^{۲۹} دارد [۹].

۲- مروری بر کارهای انجام شده:

اسلم و همکارانش در سال ۲۰۲۲ در [۱۰] برای تشخیص، تجزیه و تحلیل احساسات درباره توییت‌های مرتبط با رمزارزها، تحقیقات خود را انجام داده‌اند. تجزیه و تحلیل احساسات در رمزارزها اهمیت زیادی دارد زیرا به طور گسترده برای پیش‌بینی قیمت بازار رمزارز استفاده می‌شود و نیازمند طبقه‌بندی با دقتی بالا هستند. برای تحلیل احساسات، توییت‌ها از تویتر استخراج می‌شوند و مجموعه داده با استفاده از دو ابزار تکست‌بلاک^{۳۰} و تکست‌توایموشن^{۳۱} برچسب‌گذاری می‌شود. همچنین چندین مدل یادگیری ماشین و یادگیری عمیق برای طبقه‌بندی احساسات استفاده شده‌اند. این مطالعه از حافظه طولانی کوتاه مدت و واحد بازگشتی دروازه‌ای برای ایجاد یک مدل استفاده می‌کند تا عملکرد طبقه‌بندی را بهبود بخشد. یافته‌ها نشان می‌دهد که ترکیب حافظه طولانی کوتاه مدت و واحد بازگشتی دروازه‌ای در مقایسه با سایر مدل‌ها از نظر پیش‌بینی، عملکرد بهتری دارد.

چن در سال ۲۰۲۳ در مقاله [۱۱] برای پیش‌بینی قیمت بیت‌کوین روز بعد، متغیرهای قیمت بیت‌کوین، ویژگی‌های فنی خاص^{۳۲} بیت‌کوین، رمزارزهای دیگر، کالاها^{۳۳}، شاخص بازار^{۳۴}، ارز خارجی^{۳۵} و توجه عموم^{۳۶} را به عنوان ورودی استفاده کرده‌است. طبق این مطالعه، رگرسیون جنگل تصادفی^{۳۷} دقت پیش‌بینی بهتری نسبت به حافظه طولانی کوتاه مدت دارد. در تحقیقات قبلی، حافظه طولانی کوتاه مدت به عنوان یک الگوریتم با دقت بالا در پیش‌بینی قیمت بیت‌کوین شناخته و به صورت گسترده‌ای استفاده شده بود. در اصل این مقاله از الگوریتم یادگیری ماشین رگرسیون جنگل تصادفی استفاده می‌کند، که تاکنون به طور گسترده توسط سایر پژوهشگران در ادبیات پیشین استفاده نشده است و به این نتیجه رسیده شده که این مدل عملکرد به نسبت بهتری از خود نشان می‌دهد.

²⁵ Ensemble Learning

²⁶ Variance

²⁷ Weak Learners

²⁸ Gradient Boosting

²⁹ Random Forest

³⁰ TextBlob

³¹ TextToEmotion

³² Technical Features

³³ Commodities

³⁴ Market Index

³⁵ Foreign Exchange

³⁶ Public Attention

³⁷ Random Forest Regressor

سیبه و همکارانش در سال ۲۰۲۳ طبق مطالعه [۲]، از سه نوع تکنیک یادگیری عمیق به نام‌های حافظه طولانی کوتاه مدت، واحد بازگشتی دروازه‌ای و حافظه طولانی کوتاه مدت دوطرفه^{۳۸} برای پیش‌بینی قیمت‌های سه ارز دیجیتال بیت‌کوین، اتریوم و لایت‌کوین استفاده کرده‌اند. عملکرد مدل‌ها با استفاده از دو معیار اندازه‌گیری ریشه میانگین مربعات خطا^{۳۹} و درصد مطلق خطای متوسط^{۴۰} ارزیابی شده‌است. نتایج مطالعه نشان داد که مدل حافظه طولانی کوتاه مدت دو طرفه پیش‌بینی‌های دقیق‌تری را برای هر سه ارز ارائه می‌دهد و پس از آن مدل واحد بازگشتی دروازه‌ای قرار دارد. ترکیب جریان‌های رو به جلو و رو به عقب در مدل دوطرفه، عملکرد و پیش‌بینی در داده‌های سری زمانی را بهبود می‌بخشد. استنتاج نهایی این مطالعه این است که الگوریتم‌های یادگیری عمیق در پیش‌بینی قیمت رمزارزها موثر هستند و در مطالعات آینده، تأثیر توییت‌ها^{۴۱} و احساسات^{۴۲} بر قیمت‌های رمزارزها با استفاده از تکنیک‌های یادگیری ماشین بررسی خواهد شد.

شمشاد و همکارانش در سال ۲۰۲۳ در مقاله [۱۲] یک سیستم پیش‌بینی ارز دیجیتال ارائه داده‌اند که شش تکنیک پیشرفته یادگیری ماشین و یادگیری عمیق شامل ماشین‌های بردار پشتیبان، میانگین متحرک خودهمبسته یکپارچه^{۴۳}، پرافت فیس‌بوک^{۴۴}، حافظه طولانی کوتاه مدت، حافظه طولانی کوتاه مدت دو طرفه و حافظه طولانی کوتاه مدت به صورت پشت‌پشتی^{۴۵} پیاده‌سازی شده‌اند. برای ارزیابی کامل هر الگوریتم، از سه مجموعه داده ارز دیجیتال کاردانو^{۴۶}، بی‌ان‌بی^{۴۷} و اتریوم استفاده شده است. هر مدل بر روی هر مجموعه داده آموزش داده شده است تا قیمت ۲۴ ساعته ده روز آینده را پیش‌بینی کند. برای بهبود ارزیابی، نمونه‌های آموزش^{۴۸} و آزمون^{۴۹} هر ارز برای همه مدل‌ها ثابت نگه داشته شده است و عملکرد آن‌ها با استفاده از تکنیک‌های ارزیابی مختلف مانند میانگین مطلق خطا^{۵۰}، ریشه میانگین مربعات خطا، درصد مطلق خطای متوسط، ضریب تعیین^{۵۱} و میانگین مربعات خطا^{۵۲} ارزیابی شده است. نتایج این مطالعه نشان می‌دهد که مدل میانگین متحرک خودهمبسته یکپارچه در پیش‌بینی ارزها برای ده روز آینده، عملکرد بهتری نسبت به مدل‌های دیگر ارائه می‌دهد. از طرف دیگر، پرافت فیس‌بوک در تغییرات و نواسانات کم، عملکرد نسبتاً خوبی ارائه داده‌است. این مدل با دقت خوبی روندهای کاردانو و بی‌ان‌بی را پیش‌بینی کرده، اما روندهای اتریوم را به‌طور نادرستی پیش‌بینی کرده است.

³⁸ Bidirectional Long Short-Term Memory (Bi-LSTM)

³⁹ Root Mean Squared Error (RMSE)

⁴⁰ Mean Absolute Percentage Error (MAPE)

⁴¹ Tweets

⁴² Sentiment

⁴³ Autoregressive Integrated Moving Average (ARIMA)

⁴⁴ FB Prophet

⁴⁵ Stacked Long Short-Term Memory (Stacked LSTM)

⁴⁶ ADA

⁴⁷ BNB

⁴⁸ Train Set

⁴⁹ Test Set

⁵⁰ Mean Absolute Error (MAE)

⁵¹ R squared - Coefficient of determination

⁵² Mean Squared Error (MSE)

اویدله و همکارانش در سال ۲۰۲۳ در مطالعه [۱۳] مدل‌های یادگیری عمیق و روش‌های مبتنی بر درخت افزایشی^{۵۳} را مقایسه کرده و با استفاده از پیکربندی بهینه مدل‌ها، قیمت‌های پایانی روزانه شش رمزارز مختلف را پیش‌بینی کرده‌اند. براساس یافته‌های حاصل از این مطالعه، می‌توان به این نتیجه رسید که تکنیک‌های یادگیری عمیق نسبتاً درصد بالاتری از واریانس^{۵۴} کلی در داده‌ها را دریافت می‌کنند (بین ۸۸٪ و ۹۸٪). با این حال روش‌های مبتنی بر درخت افزایشی، دچار مشکلاتی در پیش‌بینی قیمت‌های پایانی روزانه برای رمزارزهای اتریوم، بی‌ان‌بی و دوج^{۵۵} هستند. مدل شبکه عصبی پیچشی^{۵۶} نتایج دقیق‌تری نسبت به سایر تکنیک‌های عمیق یادگیری ارائه می‌دهد و بعد از این مدل، واحد بازگشتی دروازه‌ای و سپس ماشین افزایشی گرادیان^{۵۷} قرار دارند.

مشارکت اصلی راثو و همکارانش در سال ۲۰۲۳ طبق مطالعه [۱۴]، بررسی استفاده از روش‌های یادگیری عمیق نوین برای پیش‌بینی ارزش و رفتار آینده رمزارزها است. در این مطالعه، مدل‌های یادگیری عمیق ترکیبی مبتنی بر لایه‌های حافظه طولانی کوتاه مدت، حافظه طولانی کوتاه مدت دو طرفه و شبکه عصبی پیچشی استفاده شده‌اند. بر اساس یافته‌های این مقاله می‌توان به این نتیجه رسید که ترکیب لایه‌های پیچشی و حافظه طولانی کوتاه مدت دوطرفه توانسته بهترین عملکرد را از خودش نشان دهد. در آینده قصد نویسندگان مقاله، ایجاد یک سیستم پشتیبان تصمیم‌گیری^{۵۸} برای پیش‌بینی رمزارزها است که دقیق و قابل اعتماد باشد. همچنین این ایده نیز بررسی شده است که می‌توان یک سیستم هوش مصنوعی مبتنی بر چارچوب تشخیص ناهنجاری^{۵۹} ساخت تا از الگوریتم‌های بدون نظارت^{۶۰} برای یافتن نقاط ناهنجار یا سیگنال‌های نادر استفاده کند.

موری و همکارانش در سال ۲۰۲۳ در مقاله [۱] مقایسه‌ای بین مدل‌های یادگیری عمیق، یادگیری ماشین و آماری برای پیش‌بینی قیمت‌های روزانه ارزهای رمزنگاری شده ارائه می‌دهند. نتایج نشان می‌دهد که به طور کلی رویکردهای عمیق بازگشتی^{۶۱}، بهترین مدل‌ها برای این کار یعنی پیش‌بینی ارزهای دیجیتال هستند. به طور خاص، حافظه طولانی کوتاه مدت بهترین عملکرد را دارد. همچنین لازم به ذکر است که مدل k نزدیک ترین همسایه^{۶۲} نیز تعادل خوبی بین دقت و هزینه محاسباتی ارائه می‌دهد. در آینده نیز نویسندگان مقاله قصد دارند مدل‌هایی را ارائه دهند که قیمت را به صورت ساعتی پیش‌بینی کند و همچنین احساسات بازار و نوسانات رمزارزهای دیگر را نیز مورد بررسی قرار دهند.

⁵³ Boosted Tree-Based Techniques

⁵⁴ Variance

⁵⁵ Dogecoin

⁵⁶ Convolutional Neural Network (CNN)

⁵⁷ Gradient Boosting Machine (GBM)

⁵⁸ Decision Support System

⁵⁹ Anomaly Detection

⁶⁰ Unsupervised Algorithms

⁶¹ Recurrent Deep Learning Approaches

⁶² K-Nearest Neighbors (KNN)

تریپاتی و شارما در سال ۲۰۲۳ در مقاله [۱۵] معماری‌های شبکه عصبی عمیق را برای پیش‌بینی معامله کوتاه‌مدت بیت‌کوین ارائه می‌دهند. رویکرد موردنظر از ورودی یک متغیره یعنی قیمت بسته‌شده^{۶۳} گذشته و ورودی‌های چند متغیره یعنی استفاده از شاخص‌های مالی و فنی استفاده می‌کند. مدل‌های شبکه عصبی مصنوعی عمیق^{۶۴}، حافظه طولانی کوتاه مدت، حافظه طولانی کوتاه مدت دوطرفه و ترکیب شبکه عصبی پیچشی و حافظه طولانی کوتاه مدت در این مقاله پیاده‌سازی شده‌اند. در این مقاله داده‌های پرت ابتدا کامل مدیریت شده‌اند و سپس داده‌ها وارد مدل‌ها شده‌اند. همچنین، تأثیر طول پنجره^{۶۵} نیز بررسی شده‌است تا معامله‌ای بهینه انتخاب شود. لازم به ذکر است که در این مقاله از بهینه‌سازی بیزین^{۶۶} برای یافتن هایپرپارامترهای^{۶۷} بهینه برای طراحی معماری شبکه‌های عصبی استفاده شده‌است. طبق نتایج گرفته شده، شبکه عصبی مصنوعی عمیق در مقایسه با سایر مدل‌ها عملکرد بهتری دارد. یافته‌ها نشان می‌دهد که هیچ شاخص فنی نمی‌تواند پیش‌بینی یکسانی را در تمام مراحل بازار داشته باشد. در آینده، نویسندگان مقاله قصد دارند که روش‌های بهتری برای شناسایی و مدیریت داده‌های پرت پیاده‌سازی نمایند؛ به طور مثال می‌توان از تکنیک‌های مقاوم‌تری مانند شبکه‌های عصبی رمزگذار خودکار^{۶۸} استفاده کرد.

بوتسکا و همکارانش در سال ۲۰۲۴ در مطالعه [۱۶]، مدل‌های یادگیری گروهی و یادگیری عمیق برای پیش‌بینی قیمت ارزهای رمزنگاری شده ارائه داده‌اند. آزمایش‌های زیادی با استفاده از داده‌های سری زمانی تاریخی از چهار ارز رمزنگاری شده انجام شده‌است. نتایج نشان می‌دهد که عملکرد بالایی برای روش‌های پیچیده یادگیری ماشین در هر چهار ارز رمزنگاری شده وجود دارد. مدل ماشین افزایش گرادیان سبک^{۶۹} ممکن است استراتژی‌های معاملاتی بسیار سودآوری را برای سرمایه‌گذاران در بازارهای بیت‌کوین، اتریوم و لایت‌کوین فراهم کند. برای بازار ریپل، مدل شبکه عصبی بازگشتی ساده^{۷۰} به عنوان بهترین مدل پیش‌بینی برای سرمایه‌گذاران توصیه می‌شود. به طور کلی، این یافته‌ها نشان می‌دهند که مدل‌های یادگیری گروهی و یادگیری عمیق می‌توانند به طور موثر سرمایه‌گذاران را در تصمیمات معاملاتی خود راهنمایی کنند. در آینده، قصد نویسندگان این مقاله این است که داده‌های زمان واقعی^{۷۱} بیشتری را بررسی کنند تا از مزایای مدل‌های یادگیری عمیق بیشتر بهره‌مند شوند.

⁶³ Close Price

⁶⁴ Deep Artificial Neural Network

⁶⁵ Window Size

⁶⁶ Bayesian Optimization

⁶⁷ Hyperparameters

⁶⁸ Autoencoder

⁶⁹ Light Gradient Boosting Machine (LightGBM)

⁷⁰ Simple Recurrent Neural Network (Simple RNN)

⁷¹ Real Time

همانطور که بیان شد، ارزشهای دیجیتال یکی از محبوبترین بازارهای مالی حال حاضر هستند که روزبهروز به تعداد کاربران آن افزوده می شود. به خاطر ذات و نوسانات بالا، برای تسلط به این بازار به مهارت، تجربه و زمان زیادی نیاز است. هدف این تحقیق ارائه روشی مبتنی بر یادگیری ماشین و به صورت دقیق تر یادگیری عمیق است تا در امر پیش بینی ارزشهای دیجیتال به معامله گران کمک کند. لزوم انتخاب رویکرد های یادگیری ماشین این است که در حله اول، داده موجود باشد و از کیفیت لازم برخوردار باشد. بعد از بررسی وجود داده، باید در داده، الگوهایی نیز وجود داشته باشد زیرا که الگوریتم های یادگیری ماشین در داده ها به دنبال الگوهایی برای یادگیری هستند که بتوانند از یادگیری خود، در داده های آینده استفاده کنند. به عنوان شرط آخر، الگوریتم های یادگیری ماشین در مسائلی باید استفاده شوند که نتوان آن ها را به صورت ریاضی فرموله کرد.

طبق موارد گفته شده می توان از الگوریتم های مبتنی بر یادگیری ماشین در ارزشهای دیجیتال استفاده کرد، زیرا که در این بازار هر سه شرط وجود دارد. برای شرط اول، داده های مورد نیاز برای این تحقیق را می توان از سایت های زیادی نظیر کوین مارکت کپ^{۷۲}، یاهو فایننس^{۷۳}، کگل^{۷۴} و ... بدست آورد. در داده های ارزشهای دیجیتال همواره الگو های گذشته تکرار می شوند، پس در قیمت ها و حرکت بازار الگو وجود دارد. همچنین به دلیل ماهیت بازار و نوسانات ذکر شده، حرکت این بازار را نمی توان به صورت ریاضیاتی فرموله کرد. طبق تحقیقات انجام شده، به دلیل پیچیدگی حرکت و الگوهای ارزشهای دیجیتال، بهتر است از الگوریتم های یادگیری عمیق استفاده شود و از متدهای یادگیری ماشین سنتی استفاده نشود؛ زیرا که روش های سنتی یادگیری ماشین قادر به کشف الگوهای پنهان و پیچیده ارزشهای دیجیتال نیستند.

برای این تحقیق، داده ها از سایت کوین مارکت کپ دریافت شده اند که شامل ستون های قیمت بسته شدن، قیمت باز شدن^{۷۵}، بیشترین قیمت^{۷۶}، کمترین قیمت^{۷۷}، حجم معاملات^{۷۸}، ارزش کلی بازار^{۷۹}، تاریخ^{۸۰}، تاریخ باز شدن، قیمت بسته شدن، تاریخ کمترین قیمت و تاریخ بیشترین قیمت می شود. از میان این ویژگی ها و ستون ها، در این تحقیق صرفا از سه ستون قیمت بسته شدن، تاریخ و حجم معاملات استفاده شده است. داده ها از اول ماه دوم میلادی سال ۲۰۱۸ تا آخر ماه سوم میلادی سال ۲۰۲۴ جمع آوری شده اند. لازم به ذکر است که داده ها برای این تحقیق به صورت روزانه^{۸۱} جمع آوری شده اند و هدف تحقیق، پیش بینی قیمت بسته شدن روز آینده با داشتن اطلاعات روز گذشته است.

⁷² Coinmarketcap

⁷³ Yahoo Finance

⁷⁴ Kaggle

⁷⁵ Open Price

⁷⁶ High

⁷⁷ Low

⁷⁸ Volume

⁷⁹ Marketcap

⁸⁰ Timestamp

⁸¹ Daily

پیش‌پردازش داده‌ها یکی از مهم‌ترین قدم‌ها در پیش‌بینی ارزهای دیجیتال است. یکی از عوامل مهم تاثیرگذار در دقت مدل، مدیریت کردن داده‌های پرت^{۸۲} است که در این تحقیق از روش‌هایی نظیر دامنه میان چارکی^{۸۳} استفاده شده است تا داده‌های پرت با حفظ ترتیب و توالی، مدیریت شوند. به دلیل متفاوت بودن محدوده ویژگی‌ها باید از روش‌های نرمال‌سازی^{۸۴} استفاده شود تا مدل به یک ویژگی اهمیت بالاتری ندهد. از روش‌های متداول استفاده شده در تحقیقات گذشته برای این امر می‌توان به نرمال‌سازی حداقل-حداکثر^{۸۵} طبق فرمول ۱ و نرمال‌سازی استاندارد^{۸۶} طبق فرمول ۲ اشاره کرد. داده‌های گم‌شده^{۸۷} نیز باید کنترل شوند و بهتر است حذف نشوند تا توالی حفظ شود. برای مدیریت این داده‌ها می‌توان به صورت دستی مقدار اصلی را پیدا کرد و یا میانگین داده قبلی و بعدی را به جای داده گم شده فعلی قرار داد.

فرمول ۱:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

فرمول ۲:

$$z = \frac{x - \mu}{\sigma}$$

در فرمول ۱، منظور از x داده فعلی است، x_{min} و x_{max} به ترتیب کوچک‌ترین و بزرگ‌ترین داده هستند و x_{scaled} در اصل مقدار نرمال‌شده و یا همان خروجی است. در فرمول ۲، منظور از x داده فعلی است، μ میانگین، σ انحراف معیار و z خروجی است.

همانطور که گفته شد اخبار و احساسات مردم در مورد ارزها، روی قیمت آن‌ها تاثیر زیادی دارد؛ از این‌رو داده‌های اخبار نیز باید جمع‌آوری شوند که می‌توان از منابعی مانند گوگل‌نیوز^{۸۸}، کوین‌دسک^{۸۹}، کریپتونیز^{۹۰} و .. استفاده کرد. داده‌های دریافت‌شده باید پیش‌پردازش شوند و سپس وارد الگوریتم‌های مربوطه برای تحلیل احساسات شوند. احساسات بدست‌آمده از اخبار باید با داده‌های تاریخی مربوطه اقدام شوند به‌طوری که احساسات هرروز محاسبه شود. از مدل‌ها و متدهای گوناگونی برای این امر استفاده می‌شود که از جمله آن‌ها می‌توان به برت^{۹۱}، فاین‌برت^{۹۲}، کریپتوبرت^{۹۳} و ... اشاره کرد.

⁸² Outliers

⁸³ Interquartile range (IQR)

⁸⁴ Normalization Methods

⁸⁵ Min-Max Normalization

⁸⁶ Standard Scaler

⁸⁷ Missing Data

⁸⁸ Googlenews

⁸⁹ Coindesk

⁹⁰ Cryptonews

⁹¹ Bidirectional Encoder Representations from Transformers (BERT)

⁹² Financial Bidirectional Encoder Representations from Transformers (FinBERT)

⁹³ Cryptocurrency Bidirectional Encoder Representations from Transformers (CryptoBERT)

از آن جایی که داده‌های ارز دیجیتال به صورت سری زمانی هستند، باید از الگوریتم‌هایی استفاده شوند که توانایی استخراج ویژگی‌های زمانی^{۹۴} به بهترین نحو را داشته باشند. شبکه‌های عصبی بازگشتی^{۹۵} از دسته الگوریتم‌هایی هستند که برای داده‌های سری زمانی، به دلیل داشتن حافظه، گزینه بسیار مناسبی هستند. مدل‌های حافظه طولانی کوتاه مدت و واحد بازگشتی دروازه‌ای از جمله این الگوریتم‌ها هستند که بر اساس تحقیقات گذشته، در حوزه رمزآموزها عملکرد بسیار خوبی ارائه داده‌اند. بنابراین در این تحقیق، تصمیم بر این شده که از الگوریتم‌های بازگشتی عمیق برای پیش‌بینی ارزهای دیجیتال استفاده شود با این امید که دقت بالایی کسب شود.

طبق تحقیقات گذشته، برای ارزیابی خطای مسائل رگرسیونی مانند ارزهای دیجیتال، از معیارهای میانگین مطلق خطا طبق فرمول ۳، ریشه میانگین مربعات خطا طبق فرمول ۴، درصد مطلق خطای متوسط طبق فرمول ۵ و میانگین مربعات خطا طبق فرمول ۶ استفاده خواهد شد.

فرمول ۳:

$$MAE = \frac{\sum |y_i - \hat{y}_i|}{n}$$

فرمول ۴:

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}}$$

فرمول ۵:

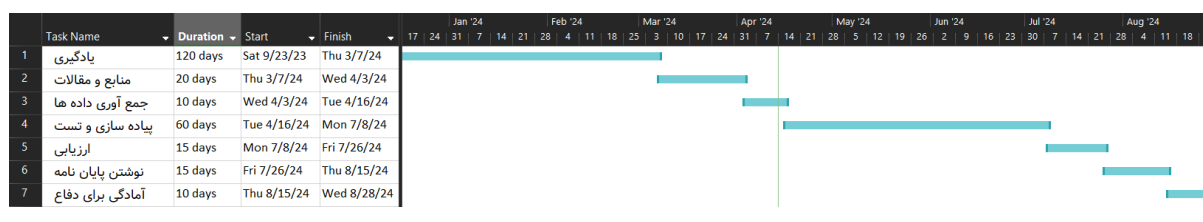
$$MAPE = \frac{1}{n} * \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100$$

فرمول ۶:

$$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n}$$

در هر چهار فرمول بالا، y_i مقدار واقعی، \hat{y}_i مقدار پیش‌بینی شده و n تعداد داده‌ها است.

۴- زمان بندی:



⁹⁴ Temporal Features

⁹⁵ Recurrent Neural Networks (RNN)

- [١] K. Murray, A. Rossi, D. Carraro, and A. Visentin, "On Forecasting Cryptocurrency Prices: A Comparison of Machine Learning, Deep Learning, and Ensembles," *Forecasting*, vol. 5, no. 1, pp. 196-209, 2023. [Online]. Available: <https://www.mdpi.com/2571-939.١٠/١/٥/٤>
- [٢] P. L. Seabe, C. R. B. Moutsinga, and E. Pindza, "Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach," *Fractal and Fractional*, vol. 7, no. 2, p. 203, 2023. [Online]. Available: <https://www.mdpi.com/2504-3110/7/2/203>.
- [٣] H. Vranken, "Sustainability of bitcoin and blockchains," *Current Opinion in Environmental Sustainability*, vol. 28, pp. 1-9, 2017/10/01/ 2017, doi: <https://doi.org/10.1016/j.cosust.2017.04.011>.
- [٤] I. E. Livieris, N. Kiriakidou ,S. Stavroyiannis, and P. Pintelas, "An Advanced CNN-LSTM Model for Cryptocurrency Forecasting," *Electronics*, vol. 10, no. 3, p. 287, 2021. [Online]. Available: <https://www.mdpi.com/2079-9292/10/3/287>.
- [٥] W. Kristjanpoller and M. Minutolo, "A hybrid volatility forecasting framework integrating GARCH, Artificial Neural network, Technical Analysis and Principal Components Analysis," *Expert Systems with Applications*, vol. 109, 05/01 2018, doi: 10.1016/j.eswa.2018.05.011.
- [٦] P. Hájek, L. Hikkerova, and J.-M. Sahut, "How well do investor sentiment and ensemble learning predict Bitcoin prices?," *Research in International Business and Finance*, vol. 64, p. 101836, 12/01 2022, doi: 10.1016/j.ribaf.2022.101836.
- [٧] W. Chen, H. Xu, L. Jia, and Y. Gao, "Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants," *International Journal of Forecasting*, vol. 37, 04/01 2020, doi: 10.1016/j.ijforecast.2020.02.008.
- [٨] D. Aggarwal, S. Chandrasekaran, and B. Annamalai, "A complete empirical ensemble mode decomposition and support vector machine-based approach to predict Bitcoin prices," *Journal of Behavioral and Experimental Finance*, vol. 27, p. 100335, 2020/09/01/ 2020, doi: <https://doi.org/10.1016/j.jbef.2020.100335>.
- [٩] X. Sun, M. Liu, and Z. Sima, "A novel cryptocurrency price trend forecasting model based on LightGBM," *Finance Research Letters*, vol. 32, p. 101084, 2020/01/01/ 2020, doi: <https://doi.org/10.1016/j.frl.2018.12.032>.
- [١٠] N. Aslam, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, "Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model," *IEEE Access*, vol. 10, pp. 39313-39324, 2022, doi: 10.1109/ACCESS.2022.3165621.
- [١١] J. Chen, "Analysis of Bitcoin Price Prediction Using Machine Learning," *Journal of Risk and Financial Management*, vol. 16, no. 1, p. 51, 2023. [Online]. Available: <https://www.mdpi.com/1911-8074/16/1/51>.
- [١٢] H. Shamshad, F. Ullah, A. Ullah, V. R. Kebande, S. Ullah, and A. Al-Dhaqm, "Forecasting and Trading of the Stable Cryptocurrencies With Machine Learning and Deep Learning Algorithms for Market Conditions," *IEEE Access*, vol. 11, pp. 122205-122220, 2023, doi: 10.1109/ACCESS.2023.3327440.
- [١٣] A. A. Oyedele, A. O. Ajayi, L. O. Oyedele, S. A. Bello, and K. O. Jimoh, "Performance evaluation of deep learning and boosted trees for cryptocurrency closing price prediction," *Expert Systems with Applications*, vol. 213, p. 119233, 2023.
- [١٤] K. R. Rao, M. L. Prasad, G. R. Kumar, R. Natchadalingam, M. M .Hussain, and P. C. S. Reddy, "Time-Series Cryptocurrency Forecasting Using Ensemble Deep Learning,"

in *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)*, 10-11 Aug. 2023, pp. 1446-1451, doi: 10.1109/ICCPCT58313.2023.10240083.

- [15] B. Tripathi and R. K. Sharma, "Modeling bitcoin prices using signal processing methods, bayesian optimization, and deep neural networks," *Computational Economics*, vol. 62, no. 4, pp. 1919-1945, 2023.
- [16] A. Bouteska, M. Z. Abedin, P. Hajek, and K. Yuan, "Cryptocurrency price forecasting – A comparative analysis of ensemble learning and deep learning methods," *International Review of Financial Analysis*, vol. 92, p. 103055, 2024/03/01/ 2024, doi: <https://doi.org/10.1016/j.irfa.2023.103055>.

به نام خدا

تعهدنامه

این جانب علی منصورآبادی به شماره‌ی دانشجویی ۴۰۰۴۳۰۳۰۷ دانشجوی مقطع کارشناسی ارشد تأیید می‌کنم که این پایان‌نامه حاصل پژوهش خودم است و در مواردی که از منابع دیگران استفاده شده، نشانی دقیق و مشخصات کامل آن را نوشته‌ام. همچنین اظهار می‌نمایم که تحقیق و موضوع پایان‌نامه‌ام تکراری نیست و موارد زیر را نیز تعهد می‌کنم:

۱- بدون کسب مجوز موسسه آموزش عالی زند و اجازه از استاد راهنما، تمام یا قسمتی از دستاوردهای پایان‌نامه‌ی خود را در مجامع و رسانه‌های علمی اعم از همایش‌ها و مجلات داخلی و خارجی به صورت مکتوب یا غیرمکتوب منتشر ننمایم.

۲- اسامی افراد خارج از کمیته‌ی پایان‌نامه را بدون اجازه‌ی استادراهنما به جمع نویسندگان مقاله‌های مستخرج از پایان‌نامه نامه اضافه نکنم.

۳- از درج نشانی یا وابستگی کاری (affiliation) نویسندگان سازمان‌های دیگر (غیر از موسسه آموزش عالی زند) در مقاله‌های مستخرج از پایان‌نامه بدون تأیید استاد راهنما اجتناب نمایم.

همه‌ی حقوق مادی و معنوی این اثر مطابق با آیین‌نامه‌ی مالکیت فکری، متعلق به موسسه آموزش عالی زند است. چنانچه مبادرت به عملی خلاف این تعهدنامه محرز گردد، موسسه زند در هر زمان و به هر نحو مقتضی حق هرگونه اقدام قانونی را در استیفای حقوق خود دارد.

علی منصورآبادی

امضا و تاریخ:

۱۴۰۳/۰۲/۲۳

