

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ
الْحٰمِدُ لِلّٰهِ الْعَظِيْمِ



دانشگاه شهید بهشتی
دانشکده علوم ریاضی
رشته علوم کامپیوتر
گرایش داده کاوی

موضوع:

پیش‌بینی سری‌های زمانی در بازار مالی با استفاده از شبکه عصبی حافظه کوتاه مدت بلند مدت

دانشجو: مونا رستگار

استاد: جناب آقای دکتر علی کتان فروش

نیم سال اول ۱۴۰۱

فهرست مطالب

۴	مقدمه
۴	۱ بیان مسئله
۸	۲ تعریف داده
۱۸	۳ معناری شبکه عصبی مصنوعی:
۳۸	۴ نتایج و ارزیابی مدل:

فهرست تصاویر

۹	۱.۲ داده های مقاله مورد تایید
۱۱	۲.۲ نمودار خطی داده های مقاله
۱۳	۲.۲ روند تغییر مقادیر ستون های (open,close,high,adj close,vloum)
۱۳	۴.۲ نمودار روند تغییرات مقدار بسته شده قیمت بیت کوین از سال ۲۰۱۴ تا سال ۲۰۲۰
۱۴	۵.۲ نمودار روند تغییرات مقدار باز شده قیمت بیت کوین از سال ۲۰۱۴ تا سال ۲۰۲۰
۱۴	۶.۲ نمودار روند تغییرات با بیشترین قیمت بیت کوین از سال ۲۰۱۴ تا سال ۲۰۲۰
۱۵	۷.۲ نمودار روند تغییرات با کمترین قیمت بیت کوین از سال ۲۰۱۴ تا سال ۲۰۲۰
۱۵	۸.۲ نمودار اسکلت پلات بین دو ویژگی کم ترین حجم معاملات و بیشترین مقدار حجم معامله
۱۶	۹.۲ نمودار گرید توصیف حجم معاملات بیت کوین در سری زمانی سال ۲۰۱۴ تا سال ۲۰۲۰
۱۷	۱۰.۲ توصیف آماری داده های تحقیق
۱۷	۱۱.۲ نمودار اسکلت پلات بین ویژگی حجم و قیمت باز شده
۲۰	۱۳. شماتیکی از یک شبکه عصبی ساده
۲۲	۲.۳ نحوه عملکرد یک شبکه عصبی بازگشتی را به نمایش می گذارد
۲۲	۳.۳
۲۳	۴.۳ نحوه عملکرد لایه dropout را بیان می کند
۲۴	۵.۳ جایگاه و نحوه قرار گیری لایه batch normalization در شبکه را نشان می دهد
۲۵	۶.۳ در شکل ۶.۳ ساختار یک نورن در شبکه ال اس تی ام نمایش داده شده است
۲۵	۷.۳ شکل ۷.۳ سلول حالت در شبکه lstm را نشان می دهد
۲۶	۸.۳ شکل ۹.۳ سلول حالت در شبکه lstm را نشان می دهد
۳۶	۹.۳ معناری شبکه اول را نشان می دهد
۳۶	۱۰.۳
۳۸	۱۱.۳ نمودار روند تغییر مقدار تابع هزینه برای شبکه اول می باشد
۳۸	۱۲.۳ نمودار روند تغییر مقدار تابع هزینه برای شبکه دوم می باشد
۴۲	۱.۴ اختلاف بین مقدار پیش بینی شده توسط شبکه اول از مقادیر واقعی را نشان می دهد (خط نارنجی نشان دهنده مقدار مورد انتظار و خط آبی مقداری که مدل پیش بینی کرده است)
۴۲	۲.۴ اختلاف بین مقدار پیش بینی شده توسط شبکه دوم از مقادیر واقعی را نشان می دهد (خط بنفش نشان دهنده مقدار مورد انتظار و خط قرمز مقداری که مدل پیش بینی کرده است)

مقدمه:

۱ بیان مسئله

امروزه دانشمندان داده و مهندس‌ها با استفاده از ابزار یادگیری ماشین و هوش مصنوعی می‌توانند برای پیش‌بینی آینده از آن‌ها استفاده کنند. در میان ابزارهای یادگیری ماشین، ابزارهایی هستند مناسب برای داده‌های زمانی هستند. داده‌های زمانی؛ داده‌هایی هستند که تغییرات را در واحد زمان ثبت و نمایش می‌دهند. داده‌های زمانی هستند که به صورت طبیعی با آن‌ها در زندگی روزمره آشنا هستیم مانند فیلم و موزیک. ما از یک الگوریتم خاص از خانواده‌ی شبکه‌های عصبی بازگشتی^۱ استفاده می‌کنیم با نام ال‌اس‌تی‌ام^۲ است. عدم قطعیت در بازار سرمایه به معنای تفاوت مقادیر مورد انتظار و مقادیری است که در واقعیت اتفاق می‌افتد. طراحی روش‌های تحلیل و پیش‌بینی مختلف در بازار سرمایه نیز به دلیل بالا بودن این مقدار و نیاز به دانستن قیمت‌ها در آینده با قطعیت بیشتر یا عدم قطعیت کمتر بوده است. برای کسب سود در بازار سرمایه، سرمایه‌گذاران همواره به دنبال پیدا کردن سهم مناسب جهت سرمایه‌گذاری و قیمت مناسب برای خرید و فروش بوده‌اند و لذا تمام مدل‌های پیش‌بینی مطرح شده همواره به دنبال پاسخ دادن به سه سوال اساسی بوده‌اند؛ چه سهمی، در چه محدوده زمانی و در چه قیمتی خریداری شود و یا به فروش برسد. قبل از بررسی پاسخ‌های داده شده به این سوالات، باید به سوال جدی تری پاسخ داد. از جمله اینکه آیا پیش‌بینی بازارهای مالی ممکن است؟! در پیش‌بینی تمام ابعاد بازار سرمایه را در سه دسته کلی متدهای تکنیکال، متدهای بنیادین و متدهای ریاضی، شامل متدهای کلاسیک سری زمانی و رگرسیون و متدهای هوش مصنوعی قرار داد. در این پژوهش، با بررسی تمام موارد بالا و امکان‌سنجی تلفیق متدهای به کارگرفته شده جهت پیش‌بینی قیمت، به سوالات مطرح شده پاسخ داده خواهد شد و برای اولین بار، به پیش‌بینی دو قیمت برای دوره‌های جلوتر پرداخته می‌شود؛ قیمت بالا و قیمت پایین سهام. به این وسیله، سفته بازان می‌توانند با به کارگیری این متد، با دقت قابل قبولی به پیش‌بینی قیمت پرداخته و از طریق نوسان‌گیری، کسب سود کنند. بیت‌کوین مستقیماً بین طرفین معامله انتقال می‌یابد و تراکنش‌های آن نیازمند گذر از یک موسسه مالی (نهاد مرکزی واسط) نیست. بیت‌کوین معروف‌ترین رمز ارز است که از فناوری زنجیره بلاکی استفاده می‌کند. این واحد

¹Recurrent Neural Network ²LSTM

پولی نوپا در سال ۲۰۰۹ توسط ساتوشی ناکاموتو ابداع شد که به نظر می‌رسد نامی مستعار باشد. البته این امکان هم وجود دارد که ایشان عضو‌گروهی 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 تحریر 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. این در حالی است که طبق آمار، تا تاریخ ۱۰ می ۲۰۲۰ تعداد ۱۲۵,۳۷۲,۱۸۰ بیت‌کوین استخراج شده 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. بیت‌کوین به عنوان یک رمز ارز، قادر 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 (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. علاوه بر این، بیت‌کوین به طرز شگفت‌آوری غیرقابل پیش‌بینی تر از استانداردهای مختلف پولی مانند دلار امریکا است. همچنین، بیت‌کوین به عنوان 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 بیت‌کوین، 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. پیش‌بینی ارزش بیت‌کوین از جنس مسائل چند متغیره است. روش‌های قدیمی پیش‌بینی رمزارزها، مانند مدل‌های هموارسازی نمایی هولت وینترس، به فرضیات خطی متکی هستند و برای موثر بودن، نیازمند داده‌هایی هستند که بتوان آنها را به روند، فصلی و نویز 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، راه حلی جالب و فناورانه را ایجاد می‌کند.

تکنیک‌های مختلفی برای پیش‌بینی روندهای بازار سهام به کار گرفته شده است. با این حال، به دلیل پیچیدگی بازار سهام، نتایج کاملاً رضایت‌بخش نیستند. بسیاری از رویکردها یا فاقد تعریف واضح و معقول از روند هستند یا از منحصر به فرد بودن ویژگی زمان در داده‌های سهام غفلت می‌کنند و با آن‌ها مانند سایر ویژگی‌ها رفتار می‌کنند. از آنجا که بورس یک سیستم پویا، غیرخطی و با ابعاد بالا است که عوامل مختلفی مانند اقتصاد، سیاست‌ها، روان‌شناسی و غیره را درگیر می‌کند که پیش‌بینی آن را دشوار می‌کند. برخی از تئوری‌ها، به عنوان مثال، نظریه بازار کارآمد بیان می‌کند که دارایی‌ها، در یک بازار با جریان آزاد اطلاعات، همیشه به ارزش منصفانه خود معامله می‌شوند و همه اطلاعات موجود را منعکس می‌کنند، به این معنی که قیمت سهام فقط تحت تأثیر اطلاعات جدید است. بنابراین، عملکرد بهتر از بازار با انتخاب سهام یا زمان بندی بازار غیرممکن به نظر می‌رسد. با این حال محققان دریافت‌هایند که اگرچه پیش‌بینی حرکت سهام در کوتاه‌مدت دشوار است، اما در بلندمدت هنوز قابلیت پیش‌بینی وجود دارد. هوش مصنوعی در آینده نزدیک تمام جنبه‌های زندگی بشر را متحول می‌سازد، از آنجا که حوزه معاملات اقتصادی و بازار بورس یکی از حوزه‌های پیچیده می‌باشد که نیازمند تحلیل‌های دقیق، هوشمندانه و منطقی است به همین سبب استفاده از هوش مصنوعی می‌تواند درهای جدیدی به روی دنیای اقتصاد باز کند. با وجود چنین سیستم‌های ریسک معاملات کاهاش پیدا می‌کند، درنتیجه معامله گر با اطمینان بهتری معاملات خود را انجام دهد. هوش مصنوعی در زمینه معاملات اقتصادی و بازار بورس می‌تواند این کاربردها را داشته باشد: کمک به داشتن فرآیندی اتوماتیک و دقیق‌تر برای معاملات، پیش‌بینی روند پیش رو به کمک تجزیه و تحلیل داده‌های گذشته، با نظارت بر بازار و تجربه و تحلیل آن این سیستم‌های می‌توانن فرصت‌های معاملاتی برای داشتن خرید و فروش‌های کارآمد برای معامله گران ایجاد کنند.

استفاده از تکنولوژی و فناوری‌های به روز همیشه می‌تواند ما را یک قدم جلوتر از دیگران قرار دهد، به ویژه وقتی صحبت از زمینه‌هایی مثل بورس و سرمایه‌گذاری باشد. دو نمونه از بهترین تکنولوژی‌های امروز در بازار بورس هوش مصنوعی و یادگیری ماشین نام دارند که به تازگی پا به این عرصه گذاشته‌اند. با استفاده از این دو عامل و پلتفرم‌های

وابسته به آن می‌توان بازدهی معاملات را افزایش داد، در هزینه و وقت صرفه جویی کرد و نهایتاً درصد ریسک را تا حد قابل توجهی کاهش داد. در سال‌های اخیر استفاده از هوش مصنوعی و یادگیری ماشین یکی از امیدوارکننده‌ترین ابزارها برای تجزیه و تحلیل و پیش‌بینی داده‌های سری زمانی که شامل داده‌های بازارهای مالی می‌شود بوده است، پیش‌بینی بازار سهام عملیاتی است که تلاش می‌کند ارزش آینده سهام یک شرکت یا سایر نمادهای مالی معامله شده در بازارهای مالی را تعیین کند (پیش‌بینی کند). پیش‌بینی موفقیت‌آمیز قیمت آینده سهام، سود سرمایه‌گذار را به حداقل می‌رساند. در عصری زندگی می‌کنیم که تکنولوژی تا ریزترین بخش‌های زندگی فردی و اجتماعی انسان وارد شده است و اجتناب از آن امکان‌پذیر نیست. از جمله بازارهایی که چند سالی می‌شود به اجتناب‌ناپذیر بودن این حقیقت پی برده‌اند بازارهای مالی هستند. ورود بازار سرمایه به عصر تکنولوژی با معاملات الگوریتمی اتفاق افتاد، همه افراد از متضرر شدن در بازار بیزارند. آگاهی از بستر ایمن و مطمئن سرمایه‌گذاری برای افراد ضروری است. الگوریتم‌های یادگیری ماشین با استفاده از داده‌های بسیار زیادی که متخصصین در اختیارش قرار می‌دهند آموزش می‌بینند. سپس با توجه به یک چهارچوب منطقی به شما می‌گوید که به عنوان مثال این شرکت سابقه‌ی خوبی دارد یا نه. این مورد حتی در جهت شناسایی کلاهبرداران نیز مورد استفاده قرار می‌گیرد که شما را از معرض مورد تقلب قرار گرفتن ایمن می‌سازد. یکی از مهم‌ترین ویژگی‌های هوش مصنوعی و یادگیری ماشین این است که نیازی به مداخله انسانی ندارد. این موضوع سرعت ما را در زندگی روزمره به شدت افزایش می‌دهد، مدیریت زمان ما را معنادارتر و دیدگاه ما را بزرگ‌تر می‌کند. به بیان بهتر به کارگیری هوش مصنوعی و یادگیری ماشین مانند استخدام کارمندی است که کاملاً دقیق بوده، به صورت مستمر درحال یادگیری است، خسته نمی‌شود، قدرت پردازش خوبی دارد و می‌تواند سفارشی‌سازی شود. آیا واقعاً انسانی به این شکل وجود دارد؟ مسلماً خیر. سیستم‌های مبتنی بر هوش مصنوعی و یادگیری ماشین کاملاً بر اساس منطق تصمیم‌گیری می‌کند و فاقد هرگونه احساسی می‌باشند از این رو بدون درنظر گرفتن خرافات و احساسات به ارائه نتایج از داده‌ها زمان دار که به دقت آن می‌افزاید، اقدام می‌کند، استفاده از هوش مصنوعی سبب کاهش ریسک و صرفه جویی در زمان می‌شود.

استفاده از روش‌های غیر کلاسیک در شناسایی مدل و پیش‌بینی رفتار سیستم‌های پیچیده، مدبهاست در محاذ علمی و حتی حرفه‌ای متناول و معمول شده است. در بسیاری از سیستم‌های پیچیده و خصوصاً غیر خطی که

مدل سازی و به دنبال آن پیش بینی و کنترل آنها از طریق روش‌های کلاسیک و تحلیلی امری بسیار دشوار و حتی بعضاً غیر ممکن می‌نماید، از روش‌های غیر کلاسیک که از ویژگی‌هایی همچون هوشمندی، مبتنی بر معرفت و خبرگی برخوردار هستند، استفاده می‌شود. شبکه‌های عصبی، یکی از این روش‌های بدیع و در حال تحول است که در موضوعات متنوعی از قبیل الگوسازی، شناخت الگو، خوش‌بندی و پیش بینی به کار رفته و نتایج مفیدی داشته است. در این مقاله، از شبکه‌های عصبی در پیش بینی سری‌های زمانی داده‌های اقتصادی استفاده کرده ایم. مسئله‌ای که می‌خواهیم به حل آن بپردازیم، مسئله‌ی پیش بینی قیمت بیت‌کوین بر پایه‌ی اطلاعات ۲۴ ساعت گذشته از قیمت آن است.

۲ تعریف داده

داده‌ها از پلتفرم‌های بیت استمپ^۳ و کوین بیس^۴ که تغییرات قیمت بیت‌کوین را نشان می‌دهند و همچنین امکان خرید و فروش بیت‌کوین را برای کاربران فراهم می‌کنند تهیه شده است. همینطور از دیتابست کوین بیس استفاده کردیم که از سال ۲۰۱۷ به بعد بود. یا هو فاینس استفاده شده است. نتایج برای دیتابست کوین بیس بهتر بود. اما چیزی که جالب بود آن بود که مقدار ستون قیمت بسته برای هر دو دیتابست یکسان بود. خب این قضیه که کوین بیس بهتر شده این پیشنهاد را می‌دهد که داشتن داده‌ی بیشتر کارایی مدل را پایین می‌آورد. خب برای همین ما آزمایش دیگری ترتیب دادیم که در آن از داده‌های دیتابست بیت استمپ و فقط از سال ۲۰۱۹ به بعد استفاده می‌کنیم. بنابراین فرضیه ما در مورد میزان داده‌ها درست نبود. در یک نگاه عمیقتر به دیتابست بیت استمپ نشان می‌داد که داده‌های آن تا زمان ۲۲ آپریل سال ۲۰۲۰ و داده‌های دیتابست کوین بیس تا زمان ۱۹ ژانویه ۲۰۱۹ است. آن داده‌ی اضافی مربوط به ماههایی است که بحران ویروس کرونا به وقوع پیوست و به تبع آن نوسان‌های غیر عادی در قیمت بازار برای بیت‌کوین به وجود آمد که منجر به کاهش چشمگیر قیمت بیت‌کوین و همچنین قیمت نفت خام گردید. مدل قادر به پیش‌بینی این نوسانات نبود و بنابراین بر کارایی مدل تاثیر منفی گذاشت.

^۴کوین بیس

^۳Bitstamp

شاید یک نفر این گونه بحث کند که توزیع داده‌ها برای دیتاست‌های کوین بیس و بیت استمپ برای هر کدام از دیتاست‌های آموزش^۵، ویلیدیشن^۶ و تست^۷ بهم بسیار نزدیک هستند. یادمان باشد که برای داده‌های تست دیتاست کوین بیس ما بازه‌ی زمانی ۲ ماهه را داریم (داده‌ها از اکتبر ۲۰۱۸ تا اواسط ژانویه ۲۰۱۹ هستند) در حالی که داده‌های تست برای بیت استمپ برابر با ۴ ماه است (داده‌ها از پایان ماه دسامبر ۲۰۱۹ تا سومین هفته از ماه آپریل ۲۰۲۰ هستند). برای این قضیه باید گفت که خیلی ساده‌تر است که در پایان سال ۲۰۱۸ پیش بینی انجام شود تا در اول سال ۲۰۲۰ به خاطر همان بحران ویروس کرونا.

	Timestamp	Open	High	...	Volume_(BTC)	Volume_(Currency)	Weighted_Price
0	1417411980	300.0	300.0	...	0.01	3.0	300.0
1	1417412040	NaN	NaN	...	NaN	NaN	NaN
2	1417412100	NaN	NaN	...	NaN	NaN	NaN
3	1417412160	NaN	NaN	...	NaN	NaN	NaN
4	1417412220	NaN	NaN	...	NaN	NaN	NaN

شکل ۱۰.۲: داده‌های مقاله مورد تایید

کتابخانه‌ی پاندا عملیات کار با دیتا و آنالیز داده را با استفاده از اشیا دیتا فریم^۸ راحت می‌کند. بعد از خواندن داده می‌توانیم نمایی کلی از دیتاست را با استفاده از دستور `raw_data.head()` داشته باشیم. این نمایی کلی از داده را برای ما فراهم می‌کند. نام ستون‌ها و ۶ تا از سطرهای داده را برای ما نمایش می‌دهد. این دیتاست دارای ۸ ستون و به صورتی قالب‌بندی شده که یک پنجره‌ی زمانی به مدت ۶۰ ثانیه را برای ما نمایش می‌دهد. ستون‌ها به صورت زیر هستند:

۱. زمان شروع پنجره‌ی زمانی که به فرمت زمانی یونیکس است (TimeStamp).
۲. قیمتی که در هنگام شروع پنجره‌ی زمانی به دلار مشخص اسن (Open).
۳. بالاترین قیمت بیت‌کوین در داخل پنجره‌ی زمانی (High).
۴. کمترین قیمت بیت‌کوین در داخل پنجره‌ی زمانی (Low).
۵. قیمت بیت‌کوین در زمان بستن پنجره‌ی زمانی (Close).

⁵Train ⁶validation ⁷test ⁸data frame

۶. مقدار حجمی از بیتکوین در پنجره‌ی زمانی معامله شده است (Volume_(BTC))

۷. مقدار حجم دلارکه در پنجره‌ی زمانی معامله شده است (Volume_(Currency)).

۸. مقدار وزن میانگین قیمت به دلار برای پنجره زمانی (Weighted_Price).

فرمت زمانی یونیکس^۹ به این صورت است که میزان ثانیه‌هایی که از تاریخ ۱ ژانویه ۱۹۷۰ گذشته است را محاسبه می‌کند این ستون به صورت یک دیتا ایندکس عمل می‌کند.

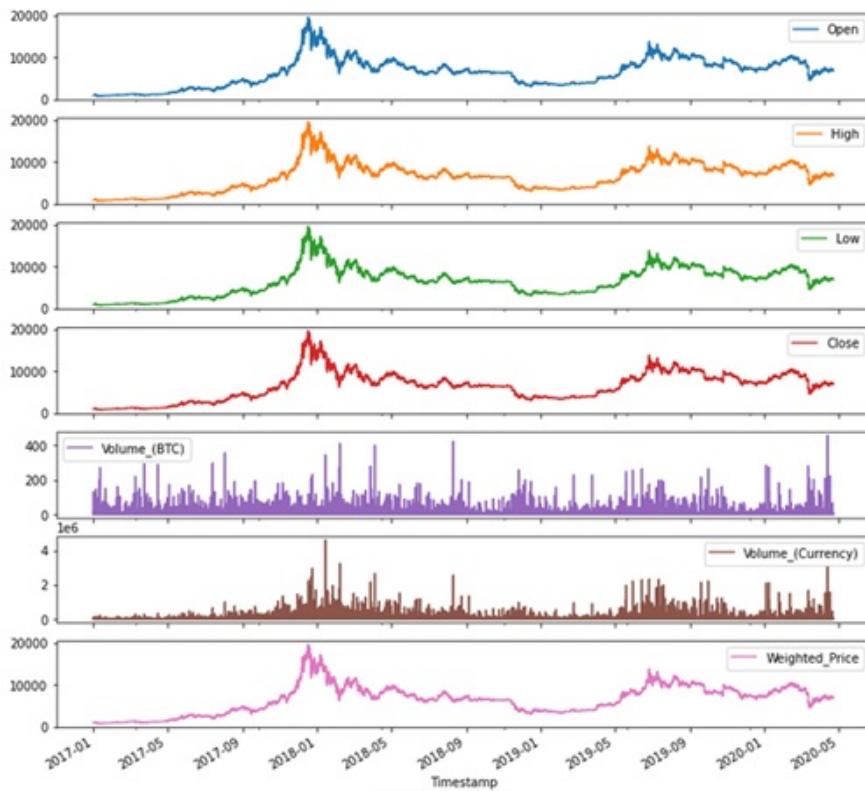
سلول‌های خالی با علامت Nan مشخص شده‌اند. این سلوول‌ها بیشتر به خاطر زمانی که بیتکوین هنوز خیلی معروف نشده بود، وجود دارند. در کتابخانه‌ی پانداز با استفاده از متدها ffill() می‌توان این سلوول‌های فاقد مقدار را با مقدارهای سطرهای قبلی جایگزین کرد.

برای اینکه از شرداده‌های Nan راحت شویم. داده‌های قبل از سال ۲۰۱۷ را نادیده می‌گیریم. از آنجا که ما می‌خواهیم پایه‌ی پیش‌بینی ما بر اساس مقادیر قیمت بیتکوین در هر ساعت باشد و هر سطر میزان قیمت را در ۶۰ ثانیه توصیف می‌کند با عمل subsampling، مقدار هر ۶۰ سطر را ترکیب کرده تا مقدار بیتکوین در یک ساعت را برای ما نمایش دهد.

ما فقط از مقدار ستون Close استفاده می‌کنیم و بقیه‌ی ستون‌ها را drope می‌کنیم. در بعضی تحلیل‌ها می‌شود از داده‌های همه‌ی ستون‌ها استفاده کرد. یا می‌توان از ترکیب ستون‌ها با یکدیگر ستون دیگری را تولید کرد و از آن برای پیش‌بینی استفاده کرد.

ما برای اینکه از کدوم دیتاست استفاده کنیم، تصمیم‌گیری کنیم. استفاده از کوین بیس یا بیت استمپ، ما همین‌طور می‌توانیم دیتاست‌های این دو منبع را باهم ترکیب کنیم. حال برای اینکه یک نمای کلی از داده‌های ستون‌های دیتاست داشته باشیم آنها را بر روی نمودار نمایش می‌دهیم:

⁹ Unix



شکل ۲.۲: نمودار خطی داده‌های مقاله

با یک نگاه کلی متوجه می‌شویم که داده‌های ستون‌های Open، High، Low، Close و Weighted_Price به هم‌دیگر هستند. داده‌های ستون‌های Volume_(BTC) و Volume_(Currency) نیز خیلی بسیار شبیه به هم‌دیگر هستند. همانطور که گفته شد ما بر روی ستون Close بهم نزدیک هستند هرچند از نظر Scale با یکدیگر تفاوت دارند. همانطور که گفته شد ما بر روی ستون Weighted_Price یعنی قیمت در آخر پنجره زمانی برای حل مسئله استفاده می‌کنیم.

در این پژوهش ما با کمک کتابخانه یاوه فایننس^{۱۰} که یک رابط بین پایتون و سایت یاوه فایننس می‌باشد داده‌های قیمت بیت‌کوین از تاریخ ۲۰۱۴/۰۹/۱۷ تا ۲۰۲۰/۰۸/۰۷ جمع‌آوری کردۀ‌ایم با توجه به آنکه هر زمان کد را اجرا کنیم قیمت بیت‌کوین را تا آن روز برای ما جمع‌آوری می‌کند. یاوه فایننس بخشی از شرکت یاوه می‌باشد که اخبار مالی، داده‌ها و تفسیر از جمله قیمت سهام، بیانیه‌های مطبوعاتی، گزارش‌های مالی و ... را ارائه می‌دهد. هم چنین برخی از ابزارهای آنلاین را برای مدیریت مالی شخصی ارائه می‌دهد. ما در این پژوهش سعی کردۀ‌ایم

¹⁰Yahoo Finance

شیکه‌های را آموزش دهیم که قیمت بسته شدن سهام در هر روز را پیش‌بینی کنند داده‌ها به صورت داده‌های فایل که به صورت فایل متنی هستند که داده‌ها با کاما از هم جدا شده‌اند.

ما برای خواندن داده‌ها از ورودی از کتابخانه‌ی پانداز^{۱۱} پایتون استفاده می‌کنیم:

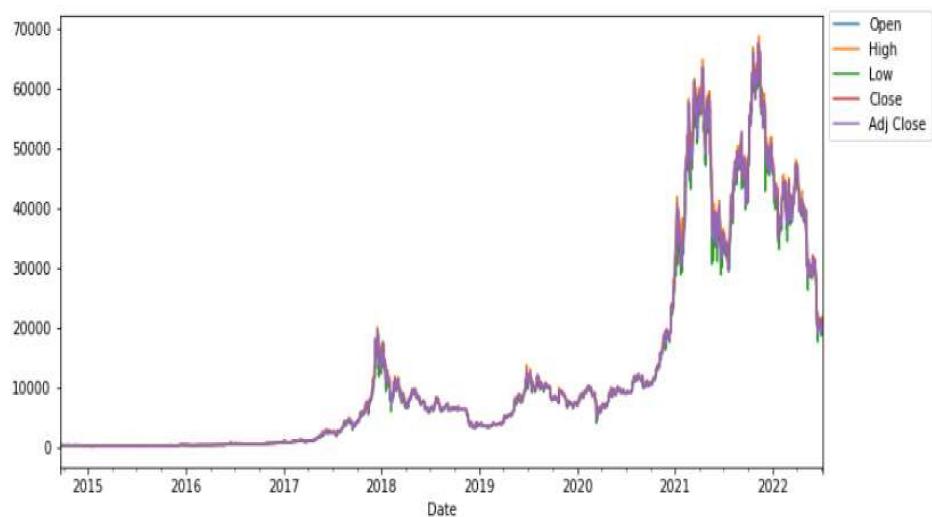
برای شرح داده‌های تحقیق و کسب اطلاعات از دیتاست از تابع `(info)`, `(head)` و توابع همبستگی داده استفاده کردایم. نمودارهای برای توصیف ستون‌های داده‌ها با استفاده از کتابخانه‌های مت‌پلات لیب^{۱۲} و سیبورن^{۱۳} و پلاتی^{۱۴} رسم کردایم در ادامه هر کدام از این نمودارها آورده شده‌اند و شرح مختصری از آنچه که به نمایش گذاشته‌اند بیان شده است.

شرح ستون‌های مختلف داده‌ها:

نام ستون	توضیحات
Index	تاریخ روزی که اطلاعات اوراق بهادر در آن ثبت شده‌اند
Open	قیمت باز شدن سهام در آن روز
Low	کمترین مقداری که سهام در آن روز باز شده است
High	بیشترین مقداری که سهام در آن روز باز شده است
Close	قیمت باز شدن اورق در آن روز
Adj close	قیمت بسته شدن تعديل شده در آن روز
Vloum	میزان حجم معاملات در آن روز

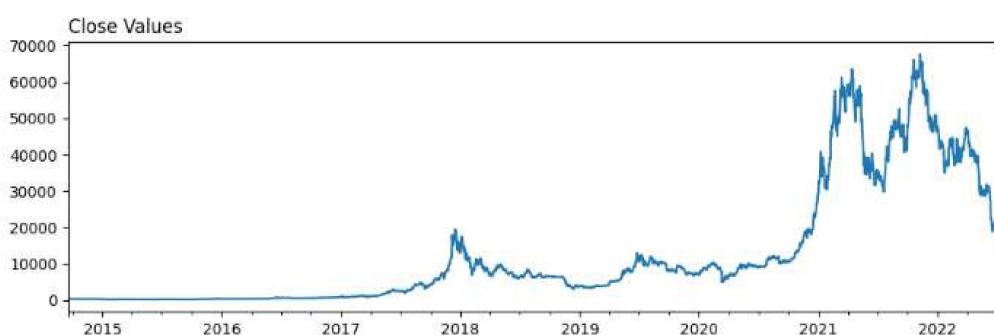
برای به خط کشیدن روند داده‌های تحقیق از نمودار زیر استفاده شده است. که مقادیر ویژگی‌های مهم بیت‌کوین از انتهای سال ۲۰۱۴ تا اواسط سال ۲۰۲۰ را نشان می‌دهد. که روند تغییرات تقریباً مشابه هم بوده است.

¹¹Pandas ¹²Matplotlib ¹³Seaborn ¹⁴plotly

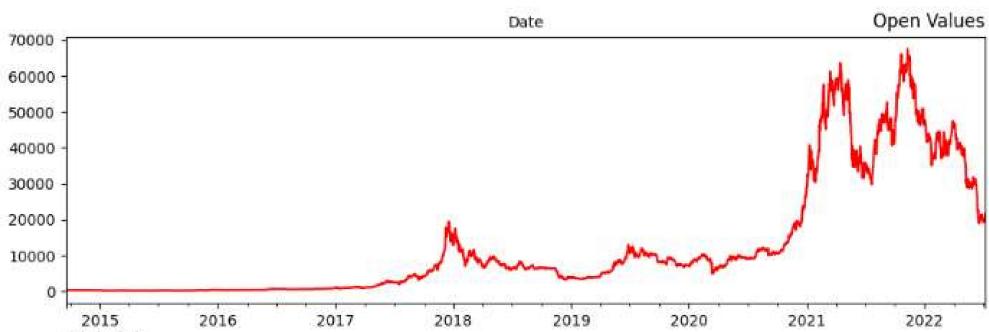


شکل ۳.۲: روند تغییر مقادیر ستون‌های (open,close,high,adj close,volume)

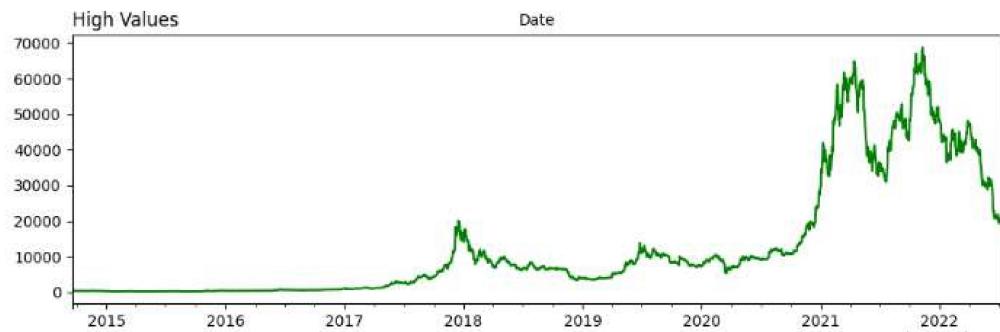
در نمودارهای زیر روند تغییر مقادیر ستون‌های مختلف دیتابست در طی زمان را مشاهد می‌کنیم:



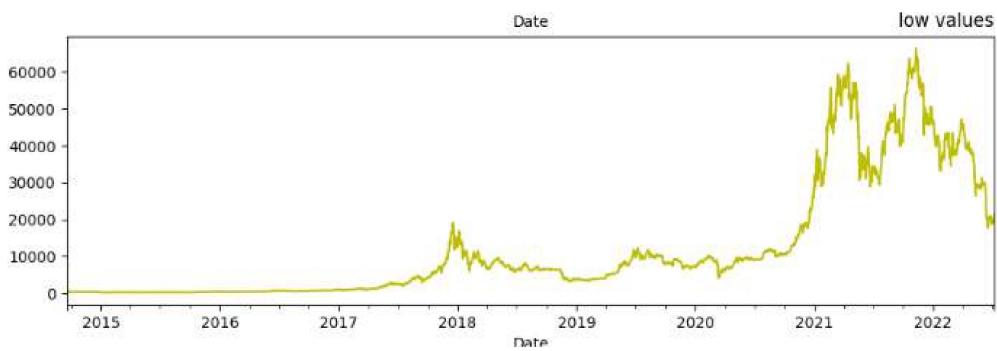
شکل ۴.۲: نمودار روند تغییرات مقدار بسته شده قیت بیتکوین از سال ۱۴۲۰ تا سال ۲۰۲۰



شکل ۵.۲: نمودار روند تغییرات مقدار باز شده قیمت بیتکوین از سال ۱۴۰۰ تا سال ۱۴۲۰

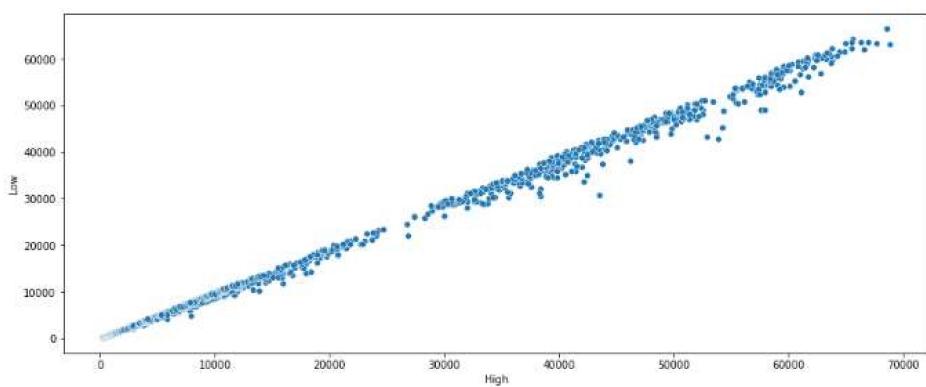


شکل ۶.۲: نمودار روند تغییرات با بیشترین قیمت بیتکوین از سال ۱۴۰۰ تا سال ۱۴۲۰

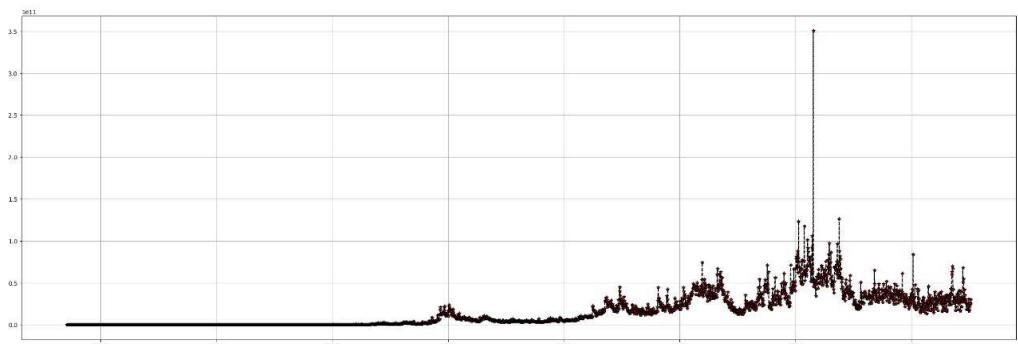


شکل ۷.۲: نمودار روند تغییرات با کمترین قیمت بیتکوین از سال ۱۴۰۰ تا سال ۱۴۰۹

توصیف وابستگی بین کمترین مقدار قیمت در حجم معاملات و بیشترین مقدار قیمت در حجم معاملات بیتکوین را نشان می‌دهد.

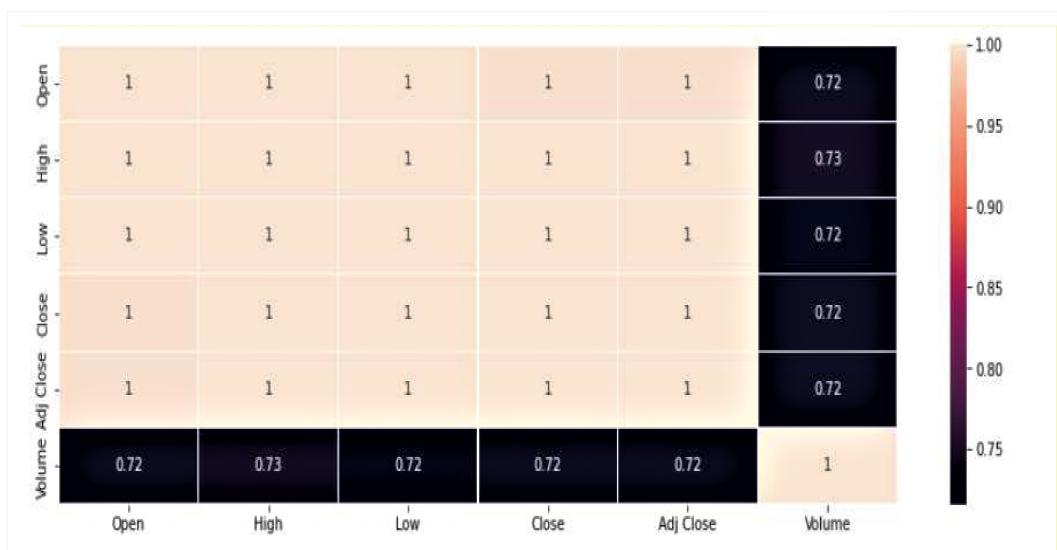


شکل ۸.۲: نمودار اسکلت پلات بین دو ویژگی کمترین حجم معاملات و بیشترین مقدار حجم معامله



شکل ۹.۲: نمودار گرید توصیف حجم معاملات بیتکوین در سری زمانی سال ۱۴ ۲۰ تا سال ۲۰ ۲۰

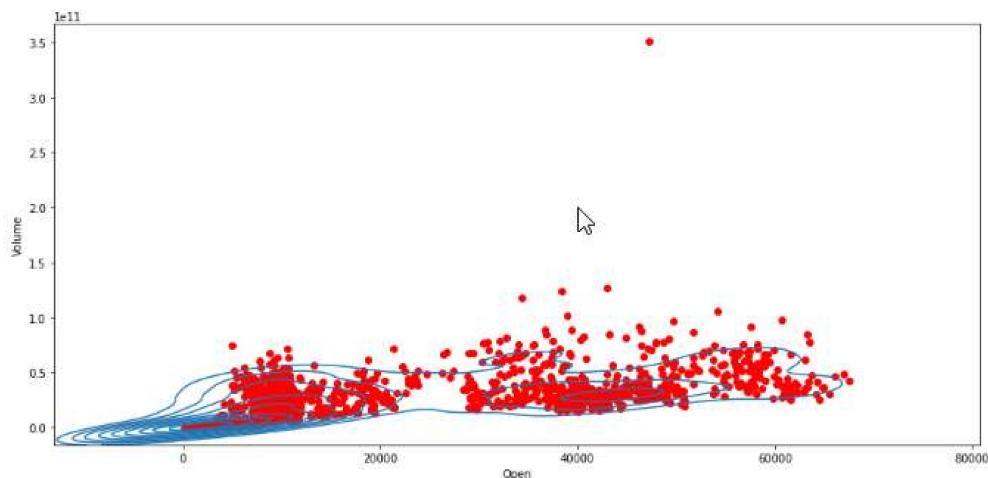
نمودار زیر همبستگی بین ستون‌ها س مختلف دیتاست را به ما نشان می‌دهد:



این جدول در هر سطر به ترتیب تعداد داده‌های موجود، میانگین، انحراف معیار، کمترین مقدار، چارک اول، چارک دوم، چارک سوم و بیشترین مقدار هر ستون را در خود گنجانیده است جدول زیر توصیف آماری کاملی از تمام ستون‌ها به نمایش می‌گذارد:

	Open	High	Low	Close	Adj Close	Volume
count	2852.000000	2852.000000	2852.000000	2852.000000	2852.000000	2.852000e+03
mean	12438.445471	12762.284354	12074.918764	12444.167710	12444.167710	1.543296e+10
std	16581.359304	17009.706219	16084.298465	16576.975235	16576.975235	1.991359e+10
min	176.897003	211.731003	171.509995	178.102997	178.102997	5.914570e+06
25%	635.306519	641.894257	625.045227	636.637512	636.637512	8.639670e+07
50%	6566.449951	6672.479980	6452.399902	6572.459961	6572.459961	5.853965e+09
75%	11529.242188	11816.213135	11286.059326	11545.715820	11545.715820	2.627581e+10
max	67549.734375	68789.625000	66382.062500	67566.828125	67566.828125	3.509679e+11

شکل ۱۰.۲: توصیف آماری داده‌ای تحقیق



شکل ۱۱.۲: نمودار اسکتر پلات بین ویژگی حجم و قیمت باز شده

و اسکتر پلات بین قیمت بسته شده و قیمت باز شده بیت‌کوین بررسی شده است. و هر گونه تحلیل آماری گه شناخت از نوع داده را به محقق داده باشد در این کد استفاده شده است. دیده گاه اماریست و نوع رسم نمودار ممکن است متفاوت باشد. در این تحقیق بیشتر از نمودار اسکتر بین ویژگی‌ها برای شناخت از ویژگی‌ها استفاده شده است.

۳ معماری شبکه عصبی مصنوعی:

در ابتدا به برخی از تعاریف تحقیق و ادبیات تحقیق خواهیم پرداخت.

تعاریف تحقیق

هوش مصنوعی^{۱۵}: هوش مصنوعی که گاهی اوقات هوش ماشینی نیز نامیده می‌شود شبیه‌سازی فرایندهای هوش طبیعی^{۱۶} توسط ماشین‌ها به ویژه سیستم‌های رایانه‌ای است. به عبارت دیگر، هوش مصنوعی به سامانه‌هایی گفته می‌شود که می‌توانند واکنش‌های مشابه رفتارهای هوشمند انسانی از جمله، درک شرایط پیچیده، شبیه‌سازی فرایندهای تفکری و شیوه‌های استدلالی انسانی و پاسخ موفق به آن‌ها، یادگیری و توانایی کسب دانش و استدلال برای حل مسائل را داشته باشند، بطور خلاصه هوش مصنوعی را دانش ساخت و طراحی عامل هوشمند تعریف کرده‌اند. این علم کاربردهای فراوانی در علوم رایانه، علوم مهندسی، تجارت، پزشکی و بسیاری از علوم دیگر دارد بعنوان مثال: در پزشکی تجزیه و تحلیل صدا قلب، ربات‌های پرستار، ارائه مشاوره و پیش‌بینی احتمال مرگ بیمار برای هر روش جراحی ...، در امور مالی و تجارت تجزیه و تحلیل بازارهای مالی، پیش‌بینی قیمت سهام‌ها، معاملات الگوریتمی، مدیریت دارایی و ... از کاربردهای هوش مصنوعی در این علوم هستند. هوش مصنوعی، موضوعی بسیار گستردۀ است که شاخه‌های متعددی دارد. شاخه‌های هوش مصنوعی عبارتند از: یادگیری ماشینی^{۱۷}، شبکه‌های عصبی^{۱۸} سیستم‌های خبره^{۱۹}، پردازش زبان طبیعی^{۲۰}، تشخیص گفتار^{۲۱} و بینایی ماشین^{۲۲} و - رباتیک^{۲۳} و منطق فازی^{۲۴} است.

یادگیری ماشین: یادگیری ماشینی شاخه‌ای از هوش مصنوعی^{۲۵} و علوم کامپیوتر^{۲۶} است که بر استفاده از داده‌ها و الگوریتم‌ها برای تقلید از روشی که انسان‌ها یاد می‌گیرند تمرکز دارد و به تدریج دقیق‌تر آن را بهبود می‌بخشد. یادگیری ماشین به عنوان بخشی از هوش مصنوعی در نظر گرفته می‌شود که به مطالعه الگوریتم‌های کامپیوتری می‌پردازد که می‌تواند به طور خودکار از طریق تجربه و با استفاده از داده‌ها بهبود یابد. الگوریتم‌های یادگیری ماشین مدلی را بر اساس داده‌های نمونه می‌سازند که به داده‌های آموزشی معروف است تا پیش‌بینی‌ها یا تصمیم‌گیری‌ها

¹⁵ Artificial Intelligence ¹⁶ Natural Intelligence ¹⁷ machine learning ¹⁸ Neural network
¹⁹ Expert Systems ²⁰ Natural language processing ²¹ speech recognition ²² Machine vision
²³ robotic ²⁴ Fuzzy logic ²⁵ artificial intelligence ²⁶ computer science

را بدون برنامه‌ریزی صریح انجام دهنده. الگوریتم‌های یادگیری ماشین در کاربردهای متنوعی مانند پزشکی، فیلتر کردن ایمیل، تشخیص گفتار و بینایی کامپیوتری استفاده می‌شوند. انواع الگوریتم‌های یادگیری ماشین عبارت‌اند از: درخت تصمیم^{۲۷}، لاجستیک رگرسیون^{۲۸}، ماشین بردار پشتیبان^{۲۹}، دسته بند بیز^{۳۰}، نزدیک‌ترین همسایه^{۳۱} و ... می‌توان اشاره کرد.

یادگیری عمیق^{۳۲}: یادگیری عمیق که در زبان فارسی به یادگیری ژرف نیز ترجمه شده است بخشی از خانواده یادگیری ماشین می‌باشد که بر روش‌های تمرکز دارد که مبتنی بر الگوریتم‌های شبکه عصبی مصنوعی^{۳۳} هستند. این الگوریتم‌ها دلالت‌اند که مغز انسان را شبیه‌سازی کنند. به طور خلاصه در یادگیری عمیق شبکه‌های عصبی مصنوعی و الگوریتم‌های مشابه مغز بشر از مجموعه‌های عظیم داده مهارت‌های مورد نظر را فرا می‌گیرند. همانطور که ما از طریق تجربه چیزهای جدید یاد می‌گیریم الگوریتم یادگیری عمیق نیز با هر بار تکرار یک کار مهارت خود را نسبت به دفعات قبلی بهبود می‌بخشد. دلیل استفاده از عبارت یادگیری عمیق این است که شبکه‌های عصبی لایه‌های مختلف یا عمیقی دارند که یادگیری را ممکن می‌سازد.

داده کاوی^{۳۴}: داده کاوی فرایندی برای تبدیل داده‌های خام به اطلاعات مفید می‌باشد، داده کاوی فرآیند استخراج و کشف الگوها در مجموعه داده‌های بزرگ است که شامل روش‌هایی در محل تلاقی یادگیری ماشین، آمار و سیستم‌های پایگاه داده است. به عبارت دیگر داده کاوی یک زیرشاخه بین رشته‌ای علوم کامپیوتر و آمار با هدف کلی استخراج اطلاعات (با روش‌های هوشمند) از مجموعه داده و تبدیل اطلاعات به یک ساختار قابل درک برای استفاده بیشتر است.

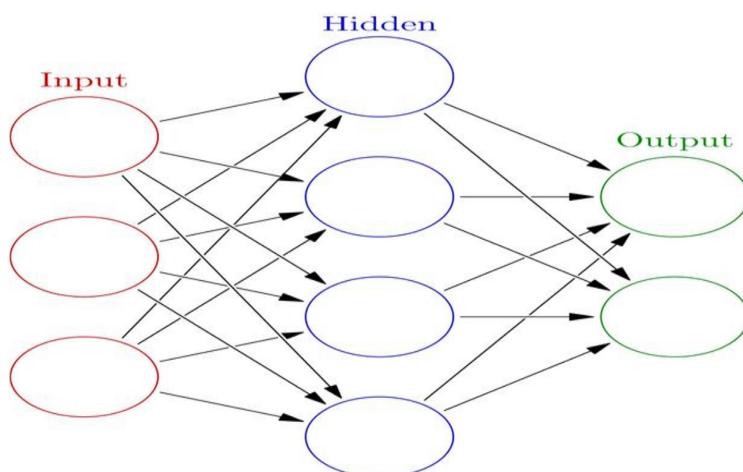
الگوریتم‌های معاملاتی: استفاده از برنامه‌های کامپیوتری برای ورود به سفارش‌های معاملاتی بدون دخالت انسان. این الگوریتم‌ها که می‌توانند بیش از یکی باشند، برای انجام معاملات بررسی‌های لازم را از جنبه‌های گوناگونی مانند زمان‌بندی، قیمت و حجم روی سفارشات و بازار انجام داده و تصمیم می‌گیرند. این امر کمک می‌کند تا بازار سرمایه به روشی اصولی‌تر و به دور از دخالت احساسات انسانی پیش رود که یکی از نتایج آن بالارفتن نقدینگی در بازار است.

²⁷Decision tree ²⁸Logistic Regression ²⁹Support vector machine ³⁰Naive Bayes Classifiers
³¹KNN ³²Deep learning ³³Artificial neural network ³⁴Data mining

اکنون وقت آن است تا به صورت مختصر به شرح برخی از انواع شبکه‌های عصبی و کاربرد آنها بپردازیم:

شبکه‌های عصبی مصنوعی^{۳۵}:

شبکه‌های عصبی مصنوعی که معمولاً به سادگی شبکه‌های عصبی نامیده می‌شوند، سیستم‌های محاسباتی هستند که از شبکه‌های عصبی بیولوژیکی الهام گرفته شده‌اند که مغز حیوانات را تشکیل می‌دهند. یک شبکه عصبی مصنوعی مبتنی بر مجموعه‌ای از واحداً یا گره‌های متصل به هم به نام نورون‌های مصنوعی است که نورون‌ها را در یک مغز بیولوژیکی مدل‌سازی می‌کنند. هر اتصال، مانند سیناپس‌های یک مغز بیولوژیکی، می‌تواند سیگنالی را به نورون‌های دیگر منتقل کند. یک نورون مصنوعی سیگنال‌ها را دریافت می‌کند و سپس آنها را پردازش می‌کند و می‌تواند به نورون‌های متصل به آن سیگنال دهد. "سیگنال" در یک اتصال یک عدد واقعی است و خروجی هر نورون با مقداری تابع غیرخطی از مجموع ورودی‌های آن محاسبه می‌شود. اتصالات لبه نامیده می‌شوند. نورون‌ها و لبه‌ها معمولاً وزنی دارند که با ادامه یادگیری تنظیم می‌شود. وزن باعث افزایش یا کاهش قدرت سیگنال در یک اتصال می‌شود. نورون‌ها ممکن است آستانه‌ای داشته باشند که سیگنال تنها در صورتی ارسال شود که سیگنال مجموع از آن آستانه عبور کند.



شکل ۱.۳: شماتیکی از یک شبکه عصبی ساده

³⁵Artificial neural network

نحوه یادگیری شبکه‌های عصبی مصنوعی:

شبکه‌های عصبی با پردازش مثال‌ها، که هر کدام حاوی یک «ورودی» و «نتیجه» شناخته شده هستند، یاد می‌گیرند (یا آموزش می‌بینند) و ارتباط‌های احتمالی را بین این دو تشکیل می‌دهند که در ساختار داده‌های خود شبکه ذخیره می‌شوند. آموزش یک شبکه عصبی از یک مثال مشخص معمولاً با تعیین تفاوت بین خروجی پردازش شده شبکه و خروجی هدف انجام می‌شود. این تفاوت همان خطأ است. سپس شبکه وزن‌های خود را بر اساس یک قانون یادگیری و با استفاده از این مقدار خطأ تنظیم می‌کند، تنظیمات پیاپی باعث می‌شود که شبکه خروجی را تولید کند که به طور فزاینده مشابه خروجی هدف باشد پس از تعداد کافی از این تنظیمات، آموزش را می‌توان بر اساس معیارهای خاصی خاتمه داد.

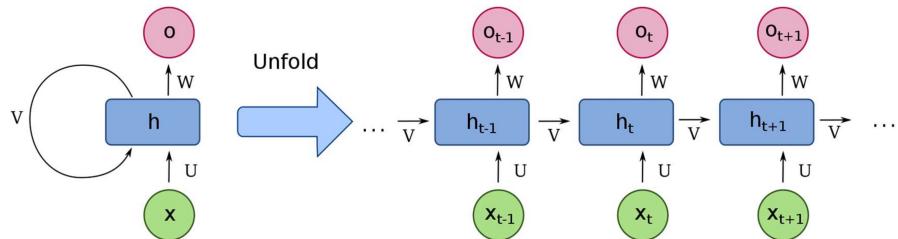
کاربرد شبکه‌های عصبی مصنوعی:

شبکه‌های عصبی مصنوعی به دلیل توانایی خود در بازتولید و مدل‌سازی فرآیندهای غیرخطی، در بسیاری از رشته‌ها کاربرد پیدا کرده‌اند. از جمله این کاربردها می‌توان به موارد زیر اشاره کرد: حوزه‌های کاربردی شامل شناسایی و کنترل سیستم (کنترل وسیله نقلیه، پیش‌بینی مسیر، کنترل فرآیند، مدیریت منابع طبیعی)، شیمی کوانتومی، تشخیص الگو (سیستم‌های رadar، شناسایی چهره، طبقه‌بندی سیگنال، بازسازی سه‌بعدی، تشخیص اشیا و موارد دیگر)، تجزیه و تحلیل داده‌های حسگر است و ...

شبکه‌های عصبی بازگشتی^{۳۶}:

شبکه عصبی بازگشتی یک کلاس از شبکه‌های عصبی مصنوعی است که در آن اتصالات بین گره‌ها یک گراف جهت‌دار یا غیر جهت‌دار را در امتداد یک دنباله زمانی تشکیل می‌دهند. این مسئله به آن اجازه می‌دهد تا رفتار پویا زمانی را نشان دهد. شبکه‌های عصبی بازگشتی که از شبکه‌های عصبی پیشخور^{۳۷} مشتق شده‌اند، می‌توانند از حالت داخلی (حافظه) خود برای پردازش توالی با طول متغیر ورودی‌ها استفاده کنند. شبکه‌های عصبی مکرر از نظر تئوری تورینگ کامل هستند و می‌توانند برنامه‌های دلخواه را برای پردازش توالی دلخواه ورودی‌ها اجرا کنند.

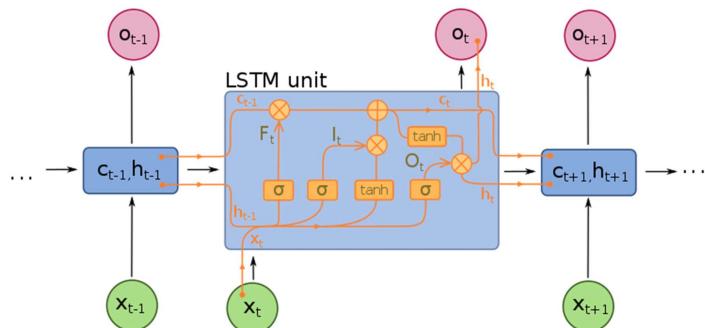
³⁶recurrent neural network ³⁷feedforward neural networks



شکل ۲.۳: نحوه عملکرد یک شبکه عصبی بازگشتی را به نمایش می‌گذارد

در ادامه به معرفی مختصری از رایج‌ترین، قدرتمندترین و پرکاربردترین شبکه عصبی بازگشتی حال حاضر می‌پردازیم که در حال حاضر در ابعاد گسترده از آن استفاده می‌شود:

حافظه کوتاه مدت بلند مدت^{۳۸}:



شکل ۳.۳:

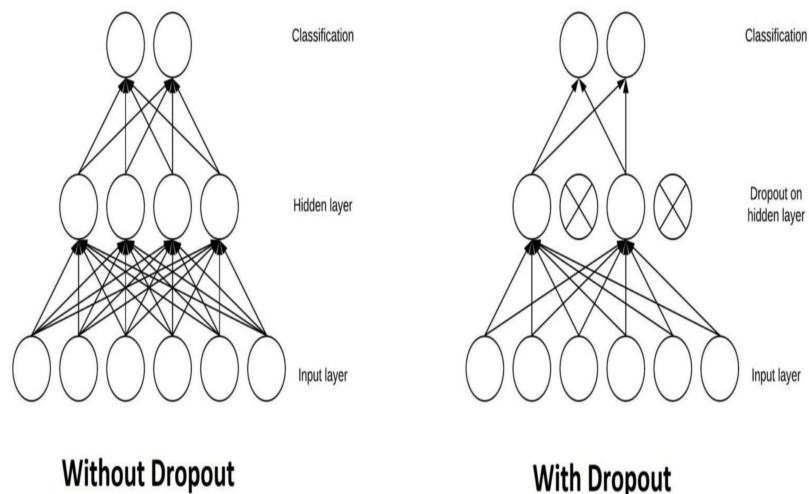
در شکل ۳.۳ شماتیکی از یک شبکه حافظه کوتاه مدت بلند مدت نمایش داده شده است. حافظه کوتاه مدت یک سیستم یادگیری عمیق است که از مشکل ناپدید شدن گرادیان جلوگیری می‌کند. حافظه کوتاه مدت بلند مدت معمولاً با گیت‌های تکراری به نام «دروازه‌های فراموش»^{۳۹} تقویت می‌شود، حافظه کوتاه مدت از ناپدید شدن یا منفجر شدن خطاهای پس از انتشار جلوگیری می‌کند. این شبکه‌ها می‌توانند

³⁸Long short-term memory ³⁹forget gate

وظایفی را بیاموزند که به خاطراتی از رویدادهای نیاز دارند که هزاران یا حتی میلیون‌ها گام زمانی گستته پیشتر اتفاق افتاده‌اند.

شرح لایه‌های مختلف در شبکه‌های عصبی:

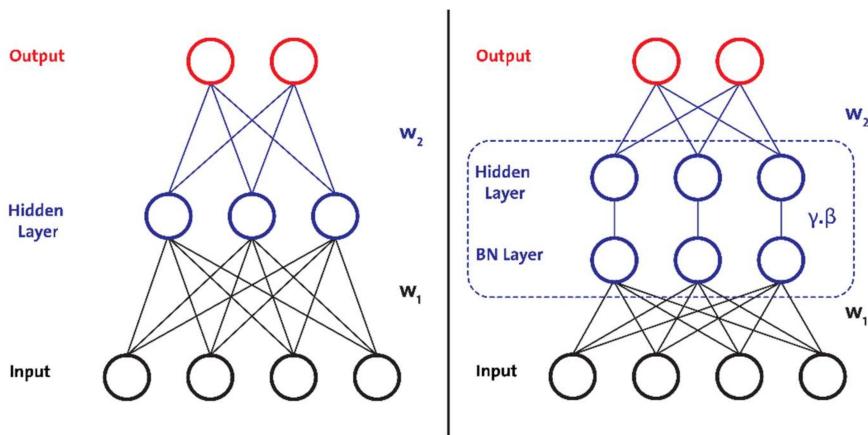
در اپ اوت^{۴۰}: در هر گام از آموزش به صورت تصادفی برخی از نورون‌ها را غیرفعال می‌کند این مسئله باعث آموزش مستقل از نورون شبکه می‌شود و همچنین از بیش برآش (اورفیت) شدن شبکه جلوگیری می‌کند.



شکل ۴.۳: نحوه عملکرد لایه dropout را بیان می‌کند

بتچ نرمالیز^{۴۱}: بعد از اعمال وزن‌ها و قبل از تابع فعالسازی به صورت یک لایه در شبکه قرار می‌گیرد و داده‌ها را نرمالیز می‌کند. باعث می‌شود سرعت یادگیری افزایش پیدا کند.

⁴⁰Dropout ⁴¹Batch normalize



شکل ۵.۳: جایگاه و نحوه قرارگیری لایه batch normalize در شبکه را نشان می‌دهد

لایه دنس (متراکم)^{۴۲}: در هر شبکه عصبی، یک لایه متراکم به لایه‌ای گفته می‌شود که عمیقاً به لایه قبلی خود متصل است، به این معنی که نورون‌های لایه به هر نورون لایه قبلی خود متصل هستند. این لایه پرکاربردترین لایه در شبکه‌های عصبی مصنوعی است. نورون لایه متراکم در یک مدل از هر نورون لایه قبلی خود خروجی دریافت می‌کند، جایی که نورون‌های لایه متراکم ضرب ماتریس-بردار را انجام می‌دهند.

در زیر فرمول کلی ضرب ماتریس-بردار آورده شده:

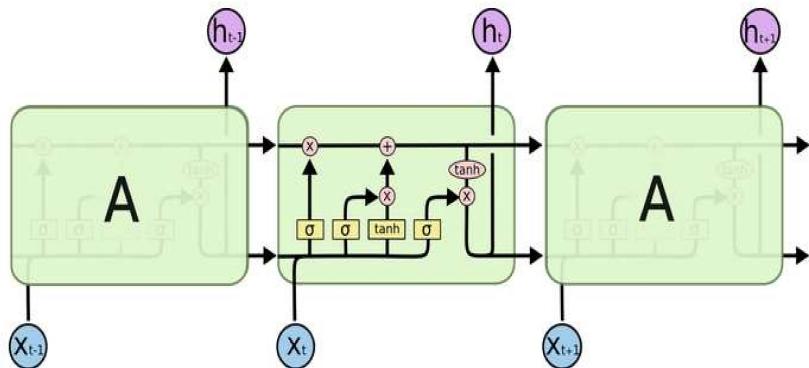
$$Ax = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n \end{bmatrix}$$

که در آن A یک ماتریس $(m \times n)$ و x یک ماتریس $n \times 1$ است. مقادیر درون ماتریس پارامترهای آموزش دیده لایه‌های قبلی هستند و همچنین می‌توانند توسط پس انتشار به روز شوند. پس انتشار متداول ترین الگوریتم مورد استفاده برای آموزش شبکه‌های عصبی پیشخور است. به طور کلی، پس انتشار در یک شبکه عصبی، گرادیان تابع هزینه را با توجه به وزن‌های شبکه برای ورودی یا خروجی واحد محاسبه می‌کند.

لایه شبکه‌های حافظه کوتاه مدت: شبکه‌های حافظه کوتاه مدت به طور گسترده‌ای برای حل کارهای متواالی مختلف مورد استفاده قرار گرفته‌اند. بیایید دریابیم که این شبکه‌ها چگونه کار می‌کنند. این شبکه‌ها ساختار کمی پیچیده‌تر

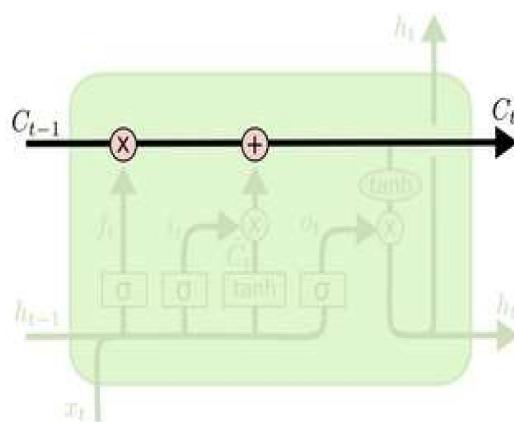
⁴²Dense

نسبت به شبکه‌های بازگشتی ساده دارند در ادامه به شرح کامل این شبکه می‌پردازیم:



شکل ۶.۳: در شکل ۶.۳ ساختار یک نورن در شبکه ال‌اس‌تی‌ام نمایش داده شده است.

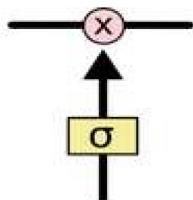
کلید ال‌اس‌تی‌ام‌ها وضعیت سلولی است، خط افقی که از بالای نمودار عبور می‌کند.



شکل ۷.۳: شکل ۷.۳ سلول حالت در شبکه LSTM را نشان می‌دهد

سلول حالت به نوعی شبیه تسمه نقاهه است. این به طور مستقیم در کل زنجیره اجرا می‌شود، تنها با برخی فعل و انفعالات خطی جزئی. بسیار آسان است که اطلاعات بدون تغییر در امتداد آن جریان یابد. ال‌اس‌تی‌ام توانایی

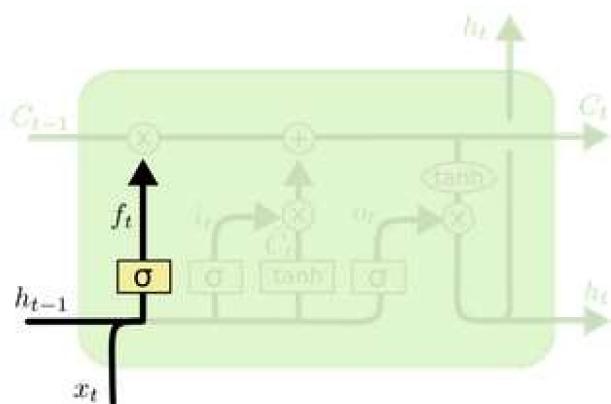
حذف یا اضافه کردن اطلاعاتی را به وسیله که با دقت توسط ساختارهایی به نام گیت تنظیم می‌شود به سلول حالت را دارد، گیتها راهی هستند که به صورت اختیاری اطلاعات را از خود عبور می‌دهند. آنها از یک لایه شبکه عصبی با تابع فعالسازی سیگموئید و یک عملیات ضرب نقطه‌ای تشکیل شده‌اند.



شکل ۸.۳: شکل ۹.۳ سلول حالت در شبکه LSTM را نشان می‌دهد

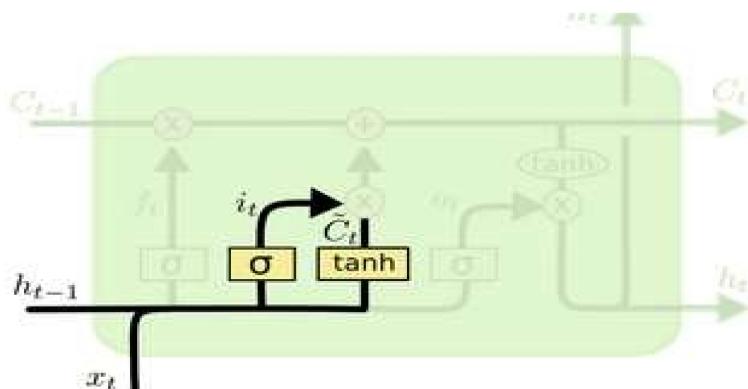
لایه سیگموئید اعدادی بین صفر و یک را خروجی می‌دهد و توضیح می‌دهد که چه مقدار از هر جزء باید عبور کند. مقدار صفر به معنای "نگذارید هیچ چیز عبور کند"، در حالی که مقدار یک به معنای "اجازه دهید همه چیز از بین بروند" می‌باشد.

یک LSTM دارای سه مورد از این دروازه‌ها برای محافظت و کنترل وضعیت سلول است. نحوه کار آن این رو قدم به قدم بررسی کنیم: اولین قدم در LSTM ما این است که تصمیم بگیرید چه اطلاعاتی را از سلول حالت دور ببریزد. این تصمیم توسط یک لایه سیگموئید به نام "دوازه فراموش" گرفته می‌شود. به $1 - ht$ و xt نگاه می‌کند و برای هر عدد در سلول حالت $1 - Ct$ عددی بین ۰ و ۱ خروجی می‌دهد. عدد ۱ نشان‌دهنده "کاملاً این را حفظ کنید" در حالی که نشان‌دهنده "به طور کامل حذف شود".



$$f_t = \sigma (W_f \cdot [h_{t-1}, x_t] + b_f)$$

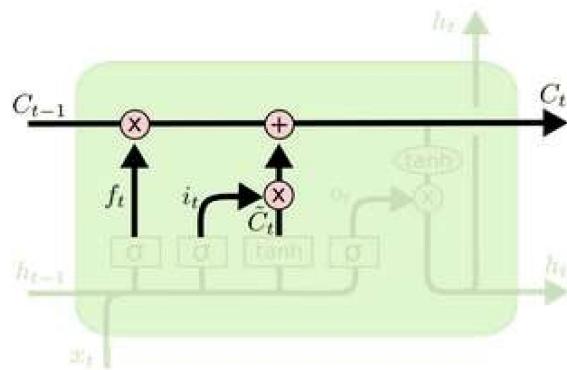
گام بعدی این است که تصمیم بگیریم چه اطلاعات جدیدی را در سلول حالت ذخیره کنیم. این گام دو بخش دارد. ابتدا، یک لایه سیگموئید به نام "لایه دروازه ورودی" تصمیم می‌گیرد که کدام مقادیر را به روز کنیم. سپس، یک لایه \tanh یک بردار از مقادیر کاندید جدید، \tilde{C}_t ایجاد می‌کند که می‌تواند به سلول حالت اضافه شود. در مرحله بعدی، این دو را با هم ترکیب می‌کنیم تا یک بهروزرسانی برای سلول حالت ایجاد کنیم.



$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i)$$

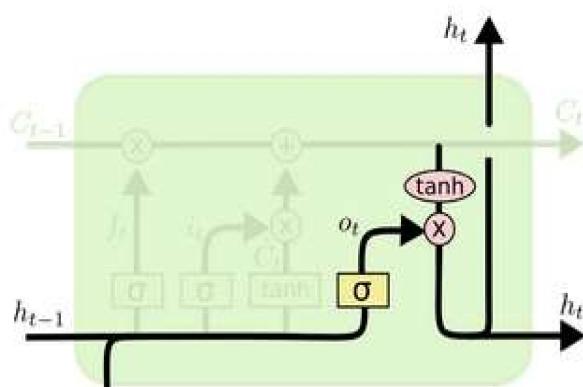
$$\tilde{C}_t = \tanh (W_C \cdot [h_{t-1}, x_t] + b_X)$$

اکنون زمان آن فرا رسیده است که سلول حالت قدیمی، C_{t-1} ، به سلول حالت جدید C_t بهروزرسانی شود. برای اینکار ایتدا سلول حالت را در f_t ضرب می‌کنیم با این کار مواردی را که می‌خواستیم فراموش کنیم از سلول حالت حذف می‌کنیم و سپس عبارت را با ($i_t \times \tilde{C}_t$) جمع می‌کنیم اکنون سلول حالت جدید بدست آمده است.



$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t$$

در نهایت، ما باید تصمیم بگیریم که چه چیزی را خروجی خواهیم داشت. این خروجی بر اساس سلول حالت ما خواهد بود، اما یک نسخه فیلتر شده خواهد بود. ابتدا یک لایه سیگموئید اجرا می‌کنیم که تصمیم می‌گیرد چه قسمت‌هایی از سلول حالت را خروجی بگیریم. سپس، سلول حالت را از طریق \tanh قرار می‌دهیم (برای فشار دادن مقادیر بین -1 و 1) و آن را در خروجی دروازه سیگموئید ضرب می‌کنیم، به طوری که فقط قسمت‌هایی را که تصمیم گرفتیم خروجی بگیریم.



$$o_t = \sigma(W_o[h_{t-1}, x_t], b_o)$$

$$h_t = o_t \times \tanh(C_t)$$

- معماری مدل شبکه عصبی الاس‌تی‌ام

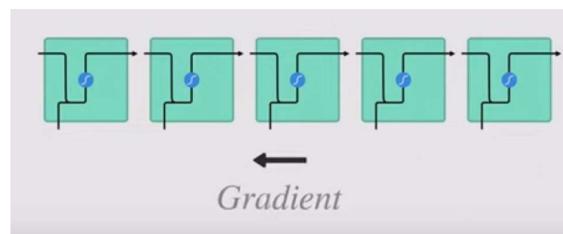
مدل LSTM یک نوع از خانواده‌ی شبکه‌های عصبی بازگشتی^{۴۳} است. شبکه‌های عصبی بازگشتی تلاش می‌کنند تا زمان یا یک رفتاری که به ترتیب وابسته است را مدل بکنند. و ترفند این شبکه به همین منظور آن است که خروجی لایه‌ی شبکه عصبی را در هر قدم به عنوان ورودی همان لایه‌ی شبکه عصبی برای استفاده در قدم بعدی، داده شود.

مدل شبکه‌های عصبی بازگشتی و مدل GRU بیشترین کاربرد را در مدل‌های به اصطلاح state-of-the-art را امروزه دارند. از این مدل‌ها برای تشخیص گفتار^{۴۴}، تولید گفتار^{۴۵} و تولید متن (Text Generation) کاربرد دارد. به عنوان مثال در ایجاد زیرنویس برای یک فیلم می‌شود از آن استفاده کرد.

شبکه عصبی بازگشتی یک کلاس از شبکه‌های عصبی مصنوعی است که در آن اتصالات بین گره‌ها یک گراف جهت‌دار یا غیر جهت‌دار را در امتداد یک دنباله زمانی تشکیل می‌دهند. این مسئله به آن اجازه می‌دهد تا رفتار پویا زمانی را نشان دهند. شبکه‌های عصبی بازگشتی که از شبکه‌های عصبی پیشخور^{۴۶} مشتق شده‌اند، می‌توانند از حالت داخلی (حافظه) خود برای پردازش توالی با طول متغیر ورودی‌ها استفاده کنند.

ایجاد مدل شبکه‌های عصبی بازگشتی برای حل مشکلاتی بود که مدل‌های اولیه RNN داشتند که به یک مورد اشاره می‌شود:

مشکل به صفر رسیدن این مشکل به این دلیل ایجاد می‌شود که زمانی که مدل در فاز یادگیری است و عمل backpropagation انجام می‌شود تا لایه‌ها به ترتیب از آخرین لایه تا اولین لایه‌ها آموختن بیینند. رفته رفته مقدار gradient کاهش پیدا می‌کند. با کاهش مقدار gradient، لایه‌های اولیه نمی‌توانند یادگیری موثری نسبت به لایه‌های اولیه داشته باشند. برای حل این مشکل از مدل LSTM استفاده شده است.

⁴³RNN⁴⁴Speech Recognition⁴⁵Speech Synthesis⁴⁶feedforward neural networks

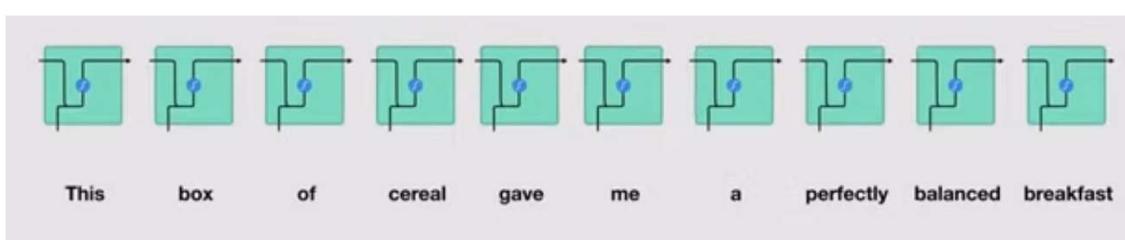
از نتایج منفی به صفر رسیدن مقدار gradient، این است که مدل دنباله‌های کوچکی را می‌تواند یاد بگیرد و قادر نخواهد بود با متن‌هایی که شامل مقادیر زیادی ورودی کاراکتر به عنوان یک دنباله هستند، کار کند.

نحوه‌ی کار شبکه‌های عصبی بازگشته:

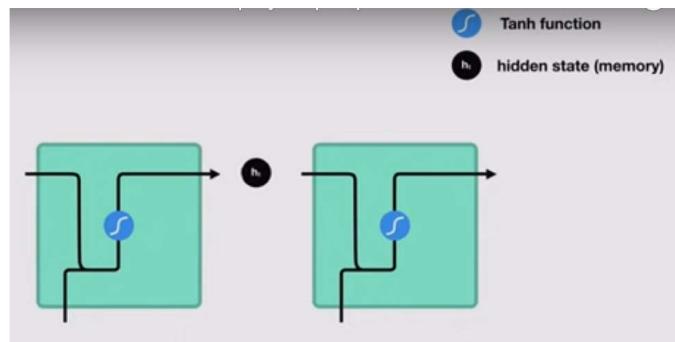
فرض کنید که می‌خواهیم بر مبنای متنی که کاربری در مورد محصولی نوشته است بفهمیم که آیا کاربر از خرید این کالا راضی است یا خیر؟

کاری که شبکه‌های عصبی بازگشته انجام می‌دهد مانند کاری است که انسان انجام می‌دهد. انسان با به خاطر سپردن کلمه‌های کلیدی، مانند: این کالا خیلی خوب بود، قطعاً دوباره آن را خواهم خرید و کلمه‌های کلیدی دیگر و حذف موارد اضافی که اطلاعاتی در مورد میزان رضایت مشتری ندارند؛ به ذهن می‌سپارد که آیا کاربر از خرید محصول راضی است یا خیر.

حال برای آنکه بفهمیم معماری شبکه‌های عصبی بازگشته چگونه کار می‌کند اول یک مدل ساده از RNN را بررسی می‌کنیم.

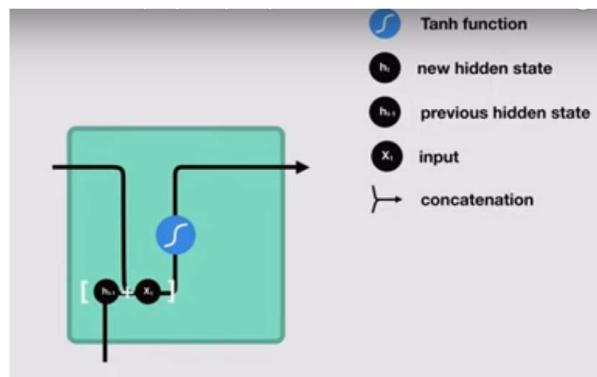


ابتدا سلول‌های RNN هر کدام یکی از کلمات جمله را به عنوان ورودی دریافت می‌کنند. این نکته باید ذکر شود که هر کلمه به صورت یک وکتور عددی به عنوان ورودی داده می‌شود.



خروجی هر سلول RNN به عنوان وضعیت پنهان (Hidden State) به سلول بعدی داده می‌شود. و شامل اطلاعاتی است از کلمه‌های قبلی که توسط مدل دیده شده است. در اینجا کلمه‌هایی که تا رسیدن به این سلول خوانده شده است به عنوان یک وضعیت پنهان استفاده می‌شود.

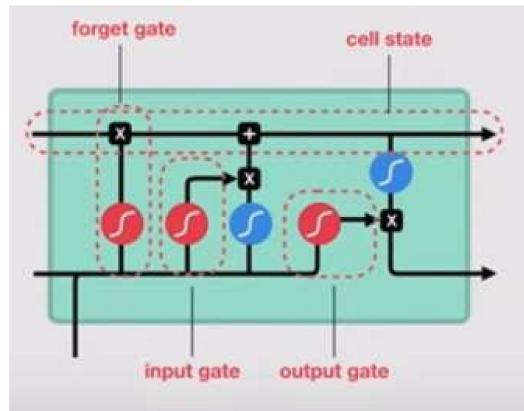
اینکه چگونه وضعیت پنهان محاسبه می‌شود. به شکل زیر دقت شود:



ابتدا داده‌ی ورودی مثلاً کلمه‌ی بارانی به ورودی به عنوان x داده می‌شود سپس این کلمه با وضعیت پنهان قبلی که از دیده شدن کلمه‌های اینجا هوا ترکیب شده و یک وکتور جدید می‌سازد. سپس این وکتور بعد از عبور از تابع تانژانت هایپربولیک مقداری بین -1 و 1 پیدا می‌کند. سپس این مقدار به عنوان وضعیت پنهان^{۴۷} جدید به سلول دیگر RNN داده می‌شود.

حال برای درک مدل LSTM به شکل زیر نگاه شود:

⁴⁷Hidden State



سلول شبکه‌های عصبی بازگشتی از ۴ بخش مهم تشکیل شده است:

۱. گیت فراموشی

۲. گیت ورودی

۳. گیت خروجی

۴. وضعیت پنهانی

که در ادامه به توضیح هر گیت می‌پردازیم:

گیت فراموشی^{۴۸}:

گیت فراموشی به عنوان ورودی خود، ورودی یک کلمه مثل باران را با وضعیت پنهان از کلمه‌های قبلی دیده شده ترکیب می‌کند. سپس این ترکیب را به تابع سیگموید خود می‌دهد. تابع سیگموید ورودی خود را به اعدادی بین صفر و یک نگاشت می‌دهد. عدد صفر یعنی من این اطلاعات را می‌خواهم فراموشی کنم و عدد یک یعنی می‌خواهم این اطلاعات حذف شود. در ادامه مشاهده می‌شود که خروجی گیت فراموشی در وضعیت سلول اطلاعات حذف شود. این همان ضربی است که صفر خروجی گیت فراموشی باعث حذف اطلاعات سلول وضعیت قبلی و ضرب شدن یک باعث ماندن آن اطلاعات می‌شود.

⁴⁸Forget State

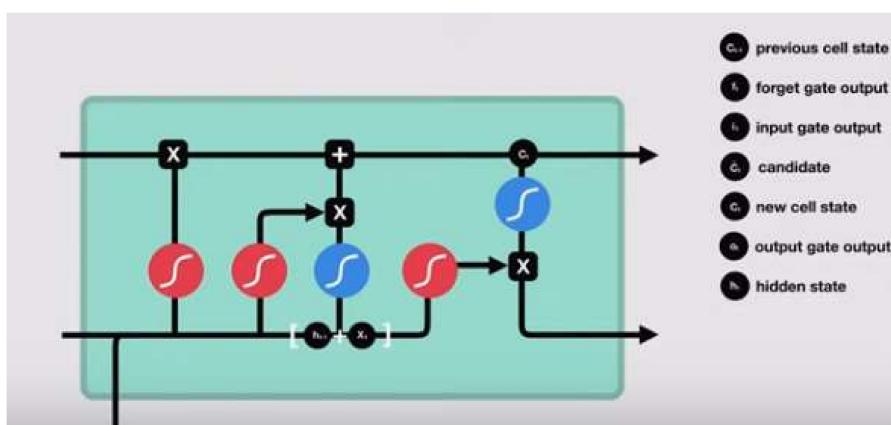
گیت ورودی^{۴۹}:

گیت ورودی شامل دو ورودی است به توابع آبی و قرمز رنگ در شکل بالا. ورودی به تابع سیگموید مثل گیت فراموشی اطلاعاتی که باید حذف شوند را مشخص می‌کند. تابع آبی رنگ، تابع تانژانت‌هاپربولیک است که باعث می‌شود ورودی بین ۱ و -۱ قرار بگیرد. این تابع باعث می‌شود تا یادگیری به درستی انجام شود. سپس خروجی هر دو تابع قرمز و آبی در هم ضرب می‌شود تا فقط اطلاعات مهم بماند. این اطلاعات که حالا اطلاعات مهم ترکیب شده از ورودی و خلاصه‌ای از اطلاعات وضعیت مخفی کلمه‌های قبلی متن است با وضعیت سلول قبلی ترکیب می‌شود.

گیت خروجی:

وظیفه‌ی گیت خروجی؛ محاسبه‌ی وضعیت پنهان جدید در خروجی خود است. که این کا را از حذف اطلاعات بی‌ارزش ترکیب ورودی و وضعیت پنهانی قبلی با استفاده از تابع سیگموید و ضرب کردن آن در وضعیت سلول قبلی بدست می‌آورد.

نتیجه‌ی همکاری این گیت‌ها با یکدیگر باعث می‌شود تا تنها اطلاعات مهم باقی بمانند و سلول وضعیت قبلی بتواند مانند حافظه اطلاعاتی که از خواندن یک متن بدست آورده را در خود ذخیره کند.



ما در این مثال از یک شبکه عصبی LSTM ساده که شامل ۱۶ واحد (unit) لایه‌ی LSTM می‌شود استفاده کردیم. تعداد سلول‌های LSTM در هر واحد یا یونیت استفاده کردیم به صورت خودکار از روی ابعاد ورودی داده

⁴⁹Input State

به مدل مشخص می‌شود. چون در مثال ما پنجره‌ی زمانی ۲۴ است در هر یک از ۱۶ واحد، ۲۴ سلول LSTM به وجود دارد.

در نهایت ما یک لایه‌ی از نوع dense layer به منظور انجام عمل پیش‌بینی به معماری خود اضافه می‌کنیم. که شامل یک واحد است و این واحد یک پیش‌بینی را تولید می‌کند که می‌توان آن را با یک برجسب مقایسه کرد. ما همچنین نیاز داریم که مقدار `return_sequence` را برابر `True` قرار دهیم. کاری که این پارامتر می‌کند آن است که سلول‌های موجود در هر واحد یا یونیت که آن‌ها نیز در لایه‌ی LSTM هستند. همه‌ی خروجی‌های خود را به یونیت‌های دیگر بدهند.

تابع `loss`، تابع بهینه ساز و متريک‌های مورد استفاده در مدل: تابع `loss` ما `MSE` یا `Mean Squared Error` است که به معنی به توان ۲ رساندن اختلاف بين پیش‌بینی مدل و برجسب است. برای تابع بهینه ساز `adam` استفاده می‌کنیم.

متريک ما برای اندازه‌گیری اختلاف بين پیش‌بینی و برجسب `absolute error` یا همان تفاضل رياضي بين پیش‌بینی و برجسب می‌باشد.

تقسيم داده و نرمالياز کردن داده‌ها:

برای مدل، ما چندین نوع دیتاست را در نظر بگيريم: `train`, `validation` و `test`. برای همين ما دیتاست را به چند بخش به ميزان ۷۰٪ و ۲۰٪ تقسيم می‌کنیم. همينطور برای اينکه فاز يادگيری ما بهينه‌تر انجام شود، نیاز داریم تا ديتا خود را به اصطلاح نرمالياز کنیم. برای اين منظور نیاز داریم تا از ميانگين و انحراف معivar دیتاست خود استفاده کنیم. تا اينجا ما به سوال‌های زير پاسخ داده‌ایم: در اين جاست که ما سه دیتاست آماده کرده‌ایم. داده‌های `Nan` را پر کرده‌ایم. عمل `subsampling` را انجام داده‌ایم و داده‌ها را نرمالياز کرده‌ایم، همچنین داده‌ها تميز شده‌اند. حالا تقریباً داده‌ها برای آن که به مدل خورانده بشوند آماده هستند.

تا اينجا داده‌های ما به صورت اشيای دیتافريم کتابخانه‌ی `tf.datasets` قبول می‌کنند. در ادامه نحوه ایجاد آن را بررسی می‌کنیم.

هدف ما این است که مقدار ارزش بیتکوین را بر اساس میزان قیمت ۲۴ ساعت گذشته‌ی آن پیش‌بینی کنیم. مدل ما به ورودی‌ها و برچسب‌های مربوط به این پنجره‌ی زمانی نیاز دارد. داکیومنت‌های فریم ورک Tesorflow توصیف مناسبی از متدان مربوط به پنجره‌ی زمانی time series کرده است. ما برای همین یک شی windowGenerator از کلاس آن به همین منظور ایجاد می‌کنیم.

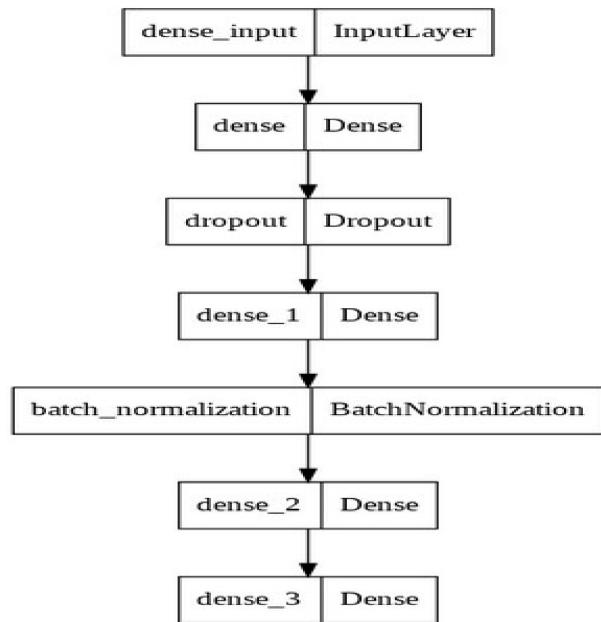
این کلاس یک کانتسراکتور و یک متد make_dataset و یک متد split_window دارد. در کانسراکتور ما ویژکی‌های پنجره زمانی را مشخص می‌کنیم و به آن اشاری به نام برش (Slice) می‌دهیم. ما از این اشاری برش بعدا برای متد plot استفاده خواهیم کرد.

متد make_dataset، داده‌های خام ورودی که بر اساس فرمت کتابخانه پانداز به صورت دیتا فریم هستند را به صورت آرایه‌های کتابخانه numpy تبدیل می‌کند و سپس آن‌ها را به صورت فرمت tf.datasets به صورت کاملاً مناسب برای مدل‌های سری زمانی تبدیل می‌کند.

حال متد split_window، این tf.datasets را از ورودی دریافت کرده و سپس آن را به دو دسته‌ی داده‌ی ورودی و داده‌ی برچسب داری که توسط مدل LSTM پذیرفته می‌شود تبدیل می‌کند.

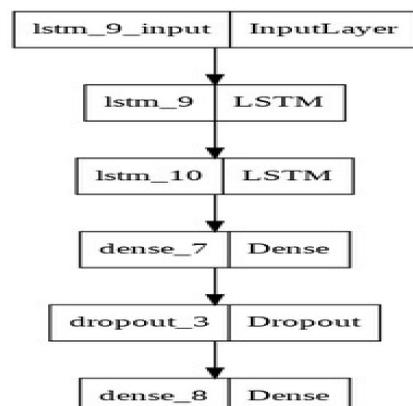
روش تحقیق حاضر که بر پایه این مفاهیم کار شده است دو شبکه عصبی data neural network(ann) و pس از آن از شبکه عصبی LSTM استفاده شده است. که از ۱۰۰ دوره نسل آموزش یا epochs را داریم و با نمودار history روند تابع loss را بررسی کرده‌ایم که این روند باید نزولی باشد.

در این پروژه دو شبکه عصبی با کمک کتابخانه کراس برای پیش‌بینی قیمت بست شدن بیتکوین در هر دور آموزش داده و سپس دقت آن‌ها را مورد ارزیابی قرار داده‌ایم که ساختار هر کدام از این شبکه‌ها در ادامه آورده شده است: شبکه اول: شامل ۵ لایه متراکم (dense) می‌باشد و ۱ لایه دارپ اوت (dropout)، یک لایه بتچ نرمالیز (batch normalization) می‌باشد.



شکل ۹.۳: معماری شبکه اول را نشان می‌دهد.

شبکه دوم: سه لایه ال‌اس‌تی‌ام (lstm) و یک لایه دارپ‌او (dense) و دو لایه متراکم (dropout) را شامل می‌شود.



شکل ۱۰.۳ :

شکل ۱۰.۳ معماری شبکه دوم را نشان می‌دهد.

آدام^{۵۰}:

یک الگوریتم بهینه‌سازی است که می‌تواند به جای روش کلاسیک گرادیان نزولی تصادفی برای بهروزرسانی وزن‌های شبکه بر اساس داده‌های آموزشی استفاده شود. آدام توسط Diederik Kingma و Jimmy Ba از OpenAI از دانشگاه تورنتو در مقاله (پوستر) ICLR 2015 خود با عنوان "Adam: A Method for Stochastic Optimization" ارائه شد، مزایای جذاب استفاده از Adam در مسائل بهینه‌سازی غیر محدب توسعه دهنده‌گان آن به شرح زیر بیان شده است:

۱. پیاده‌سازی ساده

۲. کارآمد برای محاسبات

۳. به حافظه کمی نیاز دارد

۴. برای اهداف غیر ثابت مناسب است.

۵. فرآیندهای تفسیر بصری دارند و معمولاً^{۵۱} نیاز به تنظیم کمی دارند

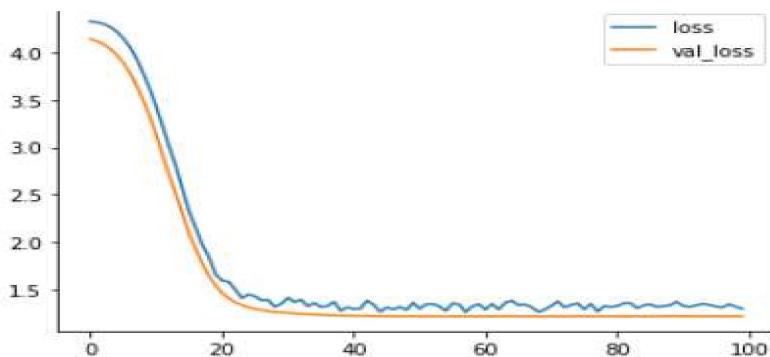
خطای میانگین مربعات^{۵۱}:

در ریاضیات و آمار، خطای میانگین مربعات روشی برای برآورد میزان خطاست که در واقع تفاوت بین مقادیر تخمینی و آنچه تخمین زده شده، است.

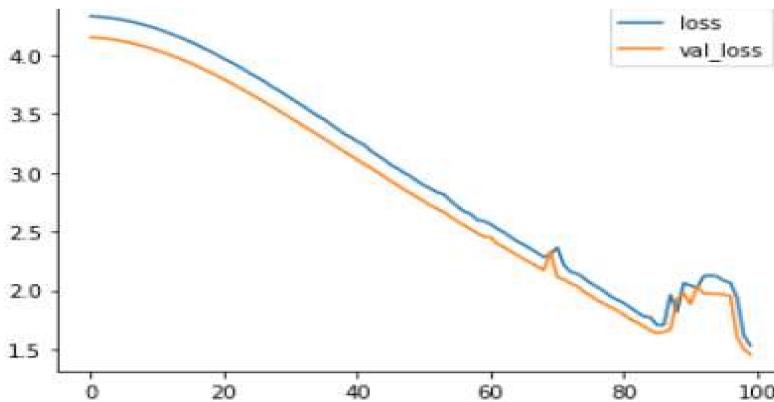
$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2,$$

ما در این تحقیق از تابع هزینه MSE و تابع بهینه‌سازی adam برای هر دو شبکه عصبی که توسعه داده‌ایم استفاده کردیم سپس هر شبکه در ۱۰۰ ایپاک آموزش دیده که در ادامه نمودار روند تغییر مقدار تابع هزینه برای هر دو شبکه نمایش داده شده:

⁵⁰adam ⁵¹Mean squared error



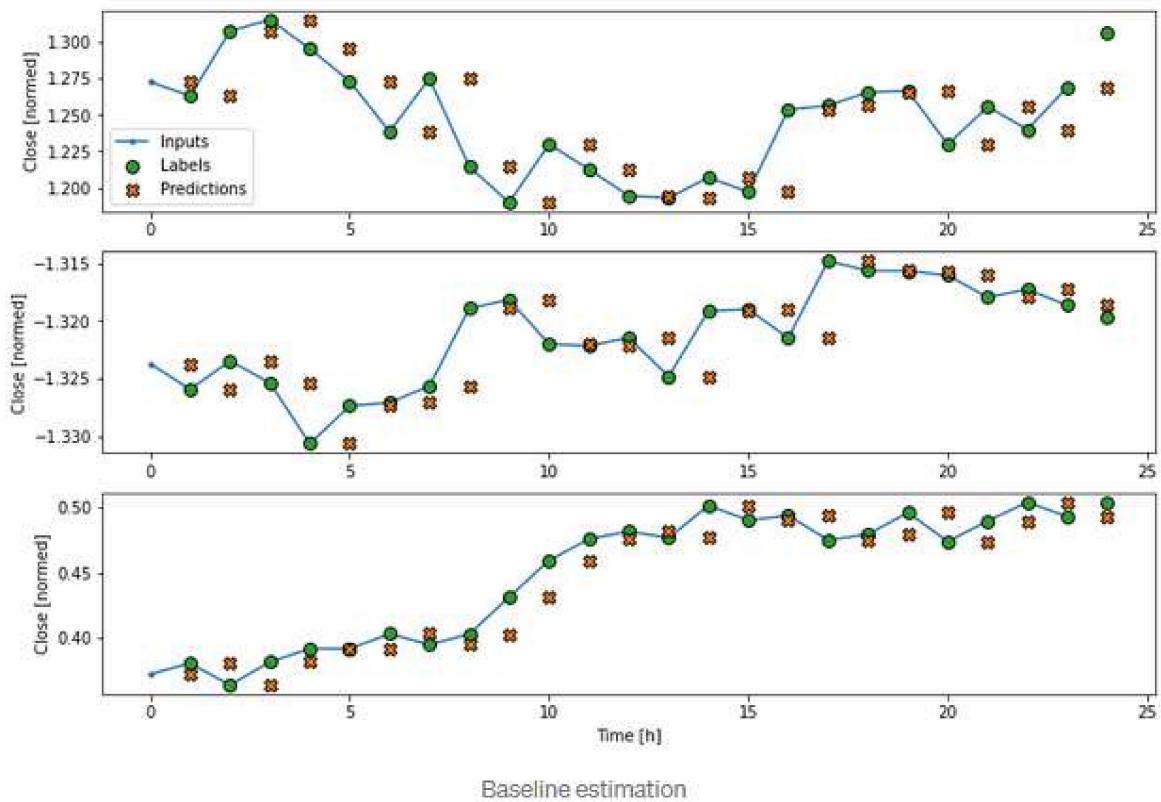
شکل ۱۱.۳: نمودار روند تغییر مقدار تابع هزینه برای شبکه اول می باشد



شکل ۱۲.۳: نمودار روند تغییر مقدار تابع هزینه برای شبکه دوم می باشد

۴ نتایج و ارزیابی مدل:

ما برای مقایسه از یک پیش‌بینی کننده استفاده می‌کنیم که برابر است با آخرین مقداری که قبل از آن وجود دارد، ما اسم این نوع پیش‌بینی کننده را baseline می‌گذاریم. ماننده یک تابع step translation است. که مقدار قبلی را به عنوان پیش‌بینی نقطه‌ی بعدی استفاده می‌کند.



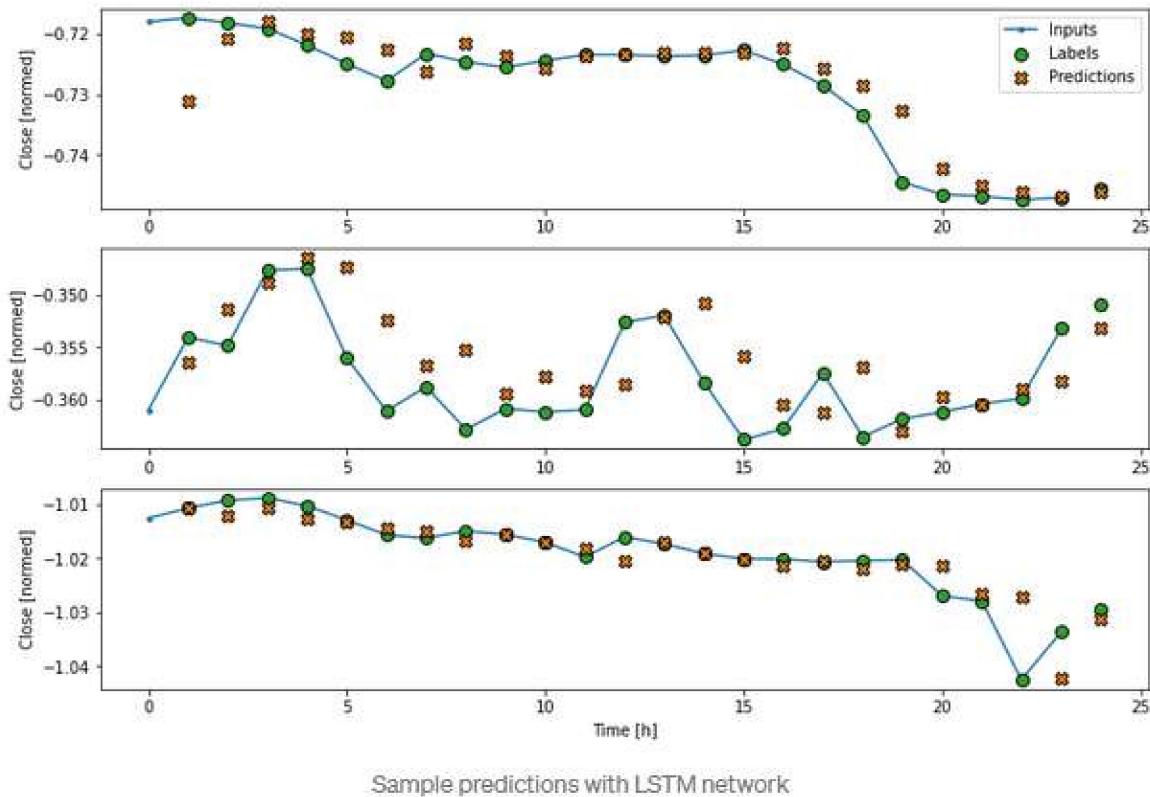
این پیش‌بینی کننده خطایی برابر با 113% دارد. به یاد داشته باشید که داده‌های ما نرمالایز شده‌اند به این معنی که میانگین برابر با صفر و انحراف معیار برابر با یک است. برای اینکه مقدار پیش‌بینی را بر واحد دلار داشته باشیم باید عمل نرمالایز را به صورت معکوس انجام دهیم به این منظور مقدار پیش‌بینی را ضرب در $36140 / 3500 = 10.26$ کنیم. که نتیجه می‌شود 39.55 دلار.

باعث سوپرایز نمی‌شده که چرا روش baseline خوب است. چیزی که نشون میده اینه که مقدار ارزش close خیلی به صورت ساعت به ساعت عجیب تغییر نمی‌کنه. خب اینکه پیش‌بینی کنیم که ثابت می‌مونه بد نیست.

مدل اصلی:

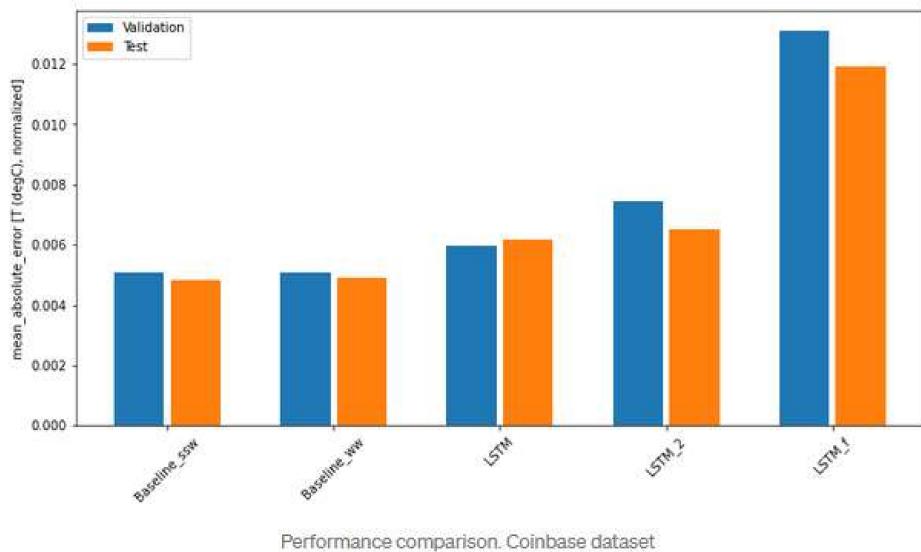
خطای مطلق یا absolute error برای مدل LSTM ما برابر با 126% است. مدل ما نسبت به مدل baseline برتری ندارد. چیزی که از نتایج به دست می‌آید مدل ما خوب می‌تواند چیزی که ترند شده را دنبال کند. اما زمانی که آن ترند یا الگو شکسته شود مدل شکست می‌خورد. به بیان دیگر. اگر در نمودار سه ساعت باشد که

ارزش بیتکوین به صورت سعودی است. مدل فکر می‌کند که این روند سعودی ادامه خواهد داشد. اما اگر در عوض نمودار سقوط کند. مدل شکست می‌خورد. این قضیه برای حالت نزولی نیز صدق می‌کند.



مقایسه مدل‌های مختلف:

ما دو تغییر اساسی را مورد استفاده قرار دادیم. یکی این بود که یک لایه‌ی دیگر LSTM به مدل اضافه کنیم و تغییر اساسی دیگر آن بود که مقدار `return_sequences` را برابر `True` قرار دهیم که این به این معنی است که لایه‌ی `dense layer` فقط از آخرین سلول LSTM به عنوان ورودی به خود استفاده می‌کند. با توجه به نمودار ما می‌فهمیم که بهترین نتیجه رو زمانی که از ساده‌ترین نوع معماری استفاده می‌کنیم بدست می‌آوریم.



نتیجه‌گیری:

پیش‌بینی زمانی یک روش خیلی قدرتمند ولی با محدودیت‌هایی نیز همراه است. همیشه گذشته اطلاعات کافی برای پیش‌بینی آینده به ما نمی‌دهد. نتیاج خطای مطلق (absolute error) برابر با $183^{\circ}/\circ$ و $192^{\circ}/\circ$ به ترتیب برای مدل LSTM و مدل Baseline بود.

در ادامه به شرح معیارهای ارزیابی تحقیق که در این پژوهش مورد استفاده قرار داده‌ایم می‌پردازیم و مقدار هر کدام از این معیارها را برای هر شبکه به تفصیل بیان خواهیم کرد:

:R2_score

یکی از معیارهای ارزیابی عملکرد برای مدل‌های یادگیری ماشین مبتنی بر رگرسیون است. به ضریب تعیین نیز معروف است. این معیار با اندازه‌گیری میزان واریانس در پیش‌بینی‌های توضیح داده شده توسط مجموعه داده کار می‌کند. به بیان ساده، تفاوت بین نمونه‌های موجود در مجموعه داده و پیش‌بینی‌های انجام شده توسط مدل است.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

:Mean absolute error

در آمار، میانگین خطای مطلق معیاری از خطاهای بین مشاهدات زوجی است که یک پدیده را بیان می‌کنند.

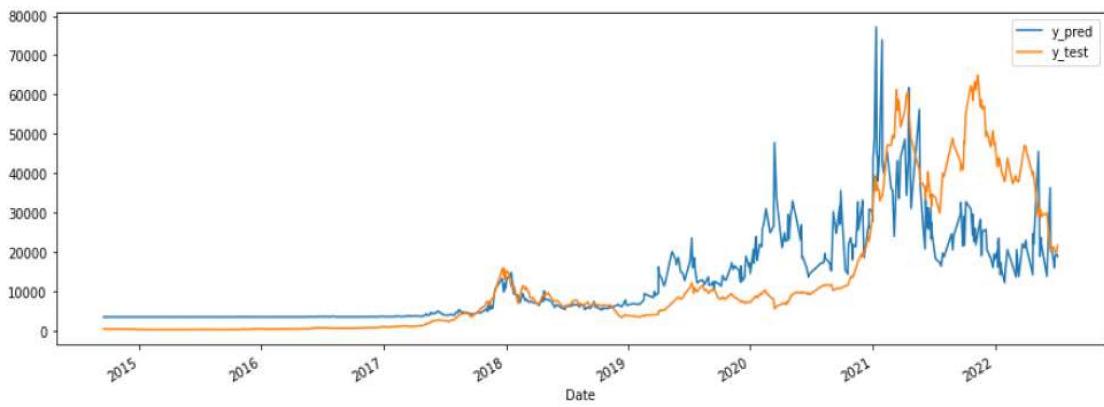
$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n}$$

و معیار mean squared error مقدار هر معیار برای شبکه اول به شرح زیر می‌باشد:

R^2 score: 0.543

Mean squared error: 1.861

Mean absolute error: 7175.270



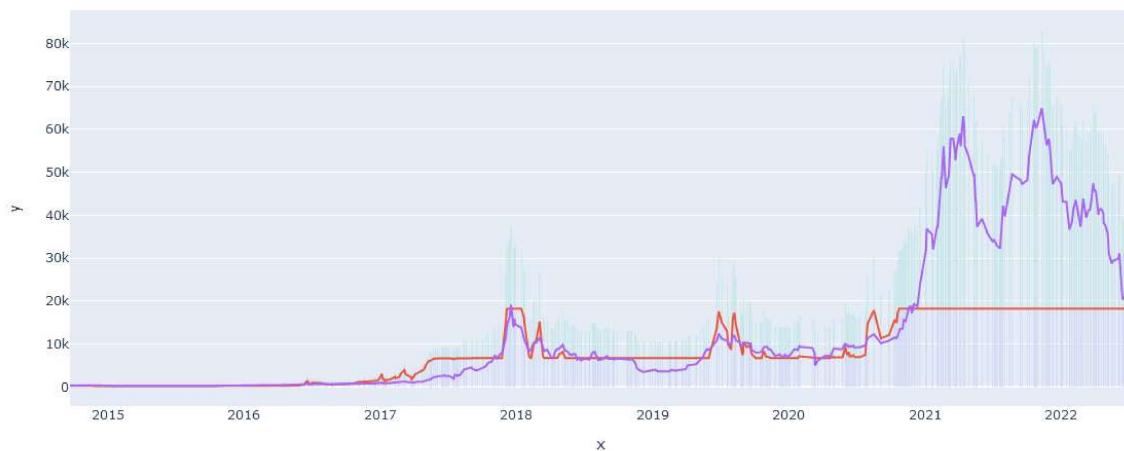
شکل ۱.۴: اختلاف بین مقدار پیش‌بینی شده توسط شبکه اول از مقادیر واقعی را نشان می‌دهد (خط نارنجی نشان‌دهنده مقدار مورد انتظار و خط آبی مقداری که مدل پیش‌بینی کرده است)

مقدار هر معیار برای شبکه دوم:

R^2 score : 0.458

Mean squared error: 0.312

Mean absolute error: 5973.618



شکل ۲.۴: اختلاف بین مقدار پیش‌بینی شده توسط شبکه دوم از مقادیر واقعی را نشان می‌دهد (خط بنفش نشان‌دهنده مقدار مورد انتظار و خط قرمز مقداری که مدل پیش‌بینی کرده است)

پیشنهادات آتی تحقیق:

حقوقان می‌توانند در آینده از مدل سری‌های زمانی آریما برای پیش‌بینی استفاده کنند. روند آن را با پیش‌بینی تحقیق حاضر مقایسه کنند.

مراجع

- [1] Dimson, E., & Mussavian, M. (1998). A brief history of market efficiency. *European financial management*, 4(1), **91-103**.
- [2] F. E. T. Burton and S. N. Shah, "Efficient market hypothesis," CMT Level I 2017: An Introduction to Technical Analysis, 2017.
- [3] Zhao, Z., Rao, R., Tu, S., & Shi, J. (2017, November). Time-weighted LSTM model with redefined labeling for stock trend prediction. In *2017 IEEE 29th international conference on tools with artificial intelligence (ICTAI)* (pp. **1210-1217**). IEEE.
- [4] Hardesty, L. (2017). Explained: neural networks. *MIT News*, 14.
- [5] Dupond, Samuel (2019). "A thorough review on the current advance of neural network structures". *Annual Reviews in Control*. 14: **200–230**.
- [6] Abiodun, Oludare Isaac; Jantan, Aman; Omolara, Abiodun Esther; Dada, Kemi Victoria; Mohamed, Nachaat Abdelatif; Arshad, Humaira (2018-11-01). "State-of-the-art in artificial neural network applications: A survey". *Heliyon*. 4 (11): e00938.
- [7] Tealab, Ahmed (2018-12-01). "Time series forecasting using artificial neural networks methodologies: A systematic review". *Future Computing and Informatics Journal*. 3 (2): **334–340**.
- [8] Hytyniemi, Heikki (1996). "Turing machines are recurrent neural networks". *Proceedings of STeP '96/Publications of the Finnish Artificial Intelligence Society*: **13–24**.
- [9] Gers, Felix A.; Schraudolph, Nicol N.; Schmidhuber, Jürgen (2002). "Learning Precise Timing with LSTM Recurrent Networks" (PDF). *Journal of Machine Learning Research*. 3: **115–143**. Retrieved 2017-06-13.
- [10] Hochreiter, Sepp (1991), Untersuchungen zu dynamischen neuronalen Netzen, Diploma thesis, Institut f. Informatik, Technische Univ. Munich, Advisor Jürgen Schmidhuber
- [11] Schmidhuber, Jürgen (January 2015). "Deep Learning in Neural Networks: An Overview". *Neural Networks*. 61: **85–117**.