JCWR2022
May 11-12, 2022; Tehran, Iran
8th International Conference on Web Research

A Deep Learning Method to Predict Stock Prices

Faezeh Memarzadeh*[1], Hadi Khosravi Farsani[2], Taghi Javdani Gandomani[3]

[1] MSc Student of Computer Engineering, Faculty of Graduate Studies, Shahrekord University, Shahrekord, Iran
Fmemarzadeh7@gmail.com

[2] Assistant Professor, Department of Computer Engineering, Shahrekord University, Shahrekord, Iran
khosravi@sku.ac.ir

[3] Assistant Professor, Department of Computer Science, Shahrekord University, Shahrekord, Iran
tjavdani@yahoo.com

Abstract

Today, with the development of the world of communication and the emergence of social networks, people's lives and decisions have undergone many changes. Social networks with the aim of informing, have been able to play a major role in various aspects of life and decision making. In the meantime, people use social networks to make various decisions such as investment and type of stock.

In this paper, a method based on deep learning is designed and developed to predict stock value movement. The data was collected from the social networks Twitter and Yahoo Finance for three months. Twitter data is first tagged with semantic analysis and then tagged using CNN. Stock market value data was calculated along with the emotion index and analyzed by different modes of the proposed LSTM model to predict the trend of stock market value. The results show that emotion indices and calculation of HLPCT and PCTchange criteria have been effective in predicting the trend of stock market value with the least error. The results also show the superiority of the proposed method compared to the existing methods in terms of price prediction.

Keywords: Stock Price Prediction, Stock Market, Emotion Index, Twitter, Yahoo Finance, Deep Learning Network

ICWR2022
May 11-12, 2022; Tehran, Iran
8th International Conference on Web Research

ارائه یک روش مبتنی بر یادگیری عمیق جهت پیش‌بینی قیمت سهام

سیده فائزه معمارزاده*[۱]، هادی خسروی فارسانی[۲]، تقی جاودانی گندمانی[۳]

[۱] دانشجوی کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه شهرکرد، شهرکرد، ایران
Fmemarzadeh7@gmail.com

[۲] استادیار گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه شهرکرد، شهرکرد، ایران
khosravi@eng.sku.ac.ir

[۳] استادیار گروه علوم کامپیوتر، دانشکده علوم ریاضی، دانشگاه شهرکرد، شهرکرد، ایران
tjavdani@yahoo.com

چکیده

امروزه با پیشرفت دنیای ارتباطات و ظهور شبکه‌های اجتماعی، زندگی و تصمیم‌گیری افراد دستخوش تغییرات زیادی شده است. شبکه‌های اجتماعی با هدف اطلاع‌رسانی، توانسته‌اند نقش بزرگی را در جنبه های مختلف زندگی و تصمیم‌گیری‌ها بعمل آورند. در این میان، افراد برای تصمیم‌گیری مختلف همچون سرمایه‌گذاری و نوع سهام از شبکه‌های اجتماعی استفاده می‌کنند.

در این مقاله برای پیش‌بینی حرکت ارزش سهام روشی مبتنی بر یادگیری عمیق طراحی و توسعه داده می‌شود. داده‌ها از شبکه‌ی اجتماعی توییتر و یاهو مالی به مدت سه ماه، جمع‌آوری شده‌اند. ابتدا برچسب‌گذاری داده‌های توییتر با اعمال تحلیل معنایی صورت می‌گیرد و سپس با به‌کارگیری شبکه‌ی CNN، طبقه‌بندی برچسب‌ها ارزیابی می‌شوند. داده‌های ارزش بازار سهام همراه با شاخص احساسات محاسبه شده و توسط حالت‌های مختلف مدل پیشنهادی LSTM برای پیش‌بینی روند حرکت ارزش بازار سهام، بررسی می‌شوند. نتایج حاکی از آن است که شاخص‌های احساسات و محاسبه دو معیار HLPCT و PCTchange در پیش‌بینی روند حرکت ارزش بازار سهام با کمترین خطا تاثیرگذار بوده است. نتایج مقایسه با روش‌های قبلی نشان‌دهنده برتری روش پیشنهادی در میزان پیش‌بینی قیمت‌ها است.

کلمات کلیدی

پیش‌بینی قیمت سهام، بازار سهام، شاخص احساسات، توییتر، یاهو مالی، شبکه‌ی یادگیری عمیق

۱- مقدمه

بازارهای مالی در اقتصاد به‌عنوان نهادهای سرمایه‌گذاری و واسط میان سرمایه‌گذاران و اقتصاد هستند. از مهم‌ترین بازارهای مالی که می‌توان آن را نشانگر اقتصاد آن کشور به‌حساب آورد؛ بازار سهام آن کشور است. از آنجا که رکود و رونق بازار سهام علاوه بر اقتصاد ملی، اقتصاد جهانی را تحت تاثیر خود قرار می‌دهد؛ سرمایه گذاران علاقمند به کسب اطلاعات بیشتر در خصوص آینده این بازار و پیش‌بینی آن هستند. پیشرفت‌های اخیر و استفاده از رایانه در امور مالی، سرمایه‌گذاران و محققان را به استفاده از این امکانات جذب کرده است تا به دنبال طراحی سیستم‌های هوشمند که قادر به پیش‌بینی بازده و بهینه‌سازی یک سبد شوند [۱].

امروزه عوامل زیادی بر رکود و صعود بازار سهام و بورس اثر دارند که لازم است سرمایه‌گذاران پیش از هر تصمیم‌گیری، آنها را مدنظر خود قرار دهند. شبکه‌های اجتماعی، دیدگاه مذهب [۲]، مباحث سیاست [۳] و قیمت نفت [۴] از برجسته‌ترین عوامل تاثیرگذار بر تغییرات قیمت بازار سهام می‌باشند.

شبکه‌های اجتماعی با هدف اطلاع‌رسانی، می‌توانند نقش بزرگی در تغییرات قیمت بازار سهام داشته باشند. شبکه‌های اجتماعی معروف مانند توییتر، اینستاگرام، فیسبوک، لینکداین و فلیکر توانستند میلیون‌ها کاربر را

جهت کسب اطلاعات جذب کنند [۵]. اطلاعاتی که توسط شبکه‌های اجتماعی منتشر می‌شوند؛ می‌توانند مدل‌هایی را در بازار مالی، برای پیش‌بینی بلندمدت تولید کنند. هنگامی که اطلاعات جدیدی به بازارها می‌رسد، سرمایه‌گذاران ممکن است عقلانی یا غیرعقلانی واکنش نشان دهند. آنها ممکن است نظرات خود را در رسانه‌های اجتماعی ابراز کنند که بعداً می‌تواند به روندهای بازار سهام تبدیل شود.

فارغ از نقش کلیدی شبکه‌های اجتماعی بر سبک زندگی، می‌توان عملکرد شبکه‌های اجتماعی را در پیش‌بینی قیمت بازار سهام و بهبود در تصمیم‌گیری سرمایه‌گذاران بررسی نمود. طی سالیان اخیر، با وجود پیچیدگی‌های تغییرات بازار سهام و نبود الگویی واضح در روند قیمت ها، پیش‌بینی بازار سهام به مسئله‌ی تحقیقاتی چالش برانگیز تبدیل شده است. علیرغم تحقیقات متعدد انجام شده و ارائه مدلهای پیش بینی تغییرات بازار سهام، کماکان این مساله به دلیل ماهیت خود نیازمند بهبود می‌باشد. به همین دلیل، این تحقیق درصدد ارائه راهکاری در این خصوص انجام شده است.

ساختار این مقاله بدین گونه‌ای است که ابتدا پیشینه کارهای انجام شده در حوزه‌ی پیش‌بینی بازار سهام بررسی می‌شود. سپس در بخش ۳، راهکار پیشنهادی و نحوه جمع‌آوری داده‌ها را شرح می‌دهیم. در بخش ۴، نتایج ارزیابی مدل‌ها ارائه شده و در انتها نتیجه‌گیری را مشخص می‌کنیم.

۲- کلیات و پیشینه تحقیق

یکی از مهمترین مباحث موجود در حوزه اقتصاد، پیش‌بینی بازار سهام بوده که از دیر باز توجه سرمایه‌گذاران را به خود جلب کرده است. پژوهشگران همواره به دنبال طراحی سیستمی با قابلیت پیش‌بینی نزدیک به واقعیت برای بازار سهام هستند.

در بازارهای بورس و سرمایه‌گذاری، افراد زیادی با تخصص‌های گوناگون، از شیوه‌های خاصی برای تجزیه و تحلیل این بازارها استفاده کرده و اینگونه فعالیت خود را با اقدام به خرید و فروش سهام آغاز می‌کنند. طی تحقیقاتی که در بازارهای مختلف مالی صورت گرفته‌است؛ استفاده از انواع روش‌های تحلیلی، در تصمیم‌گیری سهامدارها و بهینه‌سازی انتخاب آنها مؤثر بوده است. به نظر می‌رسد دیدگاه های متفاوتی در خصوص تجزیه و تحلیل بازار سهام مطرح بوده است که مهمترین آنها در ادامه مورد اشاره قرار می‌گیرند.

۲-۱- تحلیل بازارهای مالی

استفاده مکمل از دو تحلیل بنیادی [۶] و تکنیکی [۷] در تصمیم‌گیری، به سرمایه‌گذاران سهام می‌تواند دید خوبی را ارائه کند. از تحلیل بنیادی، جهت یافتن ارزش ذاتی یک سهم و از تحلیل تکنیکی، برای یافتن زمان‌های مناسب جهت خرید و فروش، می‌توان استفاده کرد.

۲-۲- تحلیل بر اساس روش‌های مبتنی بر علم کامپیوتر

تکنیک‌های داده‌کاوی، عقیده‌کاوی و هوش مصنوعی از جمله روش‌های مطرح برای پیش‌بینی بازار سهام به‌حساب می‌آیند.

از جمله پژوهش‌هایی که از تکنیک داده‌کاوی بهره گرفته‌اند، پژوهش [۸] بوده که در آن گفته شده می‌توان با تجزیه و تحلیل قیمت سهام و با استفاده از تکنیک‌های داده‌کاوی، الگوهای سازنده‌ای برای تصمیم‌گیری تولید کرد. تکنیک‌های داده‌کاوی می‌تواند برای ارزیابی قیمت سهام گذشته استفاده شده و با برآورد شاخص‌های مالی مناسب، اطلاعات ارزشمندی بدست آورند.

عقیده‌کاوی یا تحلیل احساسات، استخراج دانش از متن با استفاده از روش‌های داده‌کاوی بوده که به‌عنوان شاخه‌ای از متن‌کاوی، روی استخراج عقاید متمرکز می‌باشد. در پژوهش [۹] اشاره شده است که دستیابی به مدلی که توانایی پیش‌بینی قیمت سهام را داشته باشد؛ کار دشواری است. رشد شبکه‌های اجتماعی موجب شده که با حجم وسیعی از نظرهای کاربرها روبرو شویم. در این پژوهش، عقیده‌کاوی مبتنی بر مدل را با معرفی یک مدل گرافیکی که بخشی از گفتار برای استخراج نظرهای کاربرها است، و آن را در دو مجموعه داده‌ی مختلف به زبان انگلیسی و فارسی تست می‌کنند. مجموعه داده‌های فارسی در این مقاله از شبکه اجتماعی سهام بورس ایران (سهام یاب) و مجموعه داده‌های انگلیسی هم از یاهو مالی جمع‌آوری شده است. در پیش-بینی بازار سهام توسط این مدل، به دقتی خوبی دست می‌یابند.

مفهوم جدیدی از هوش مصنوعی به نام یادگیری عمیق، عملکرد پیش-بینی را نسبت به مدل‌های یادگیری سنتی، بهبود می‌دهد. قابلیت پردازش حجم زیاد اطلاعات و در عین حال وجود قدرت محاسباتی بالا، از ویژگی‌های مهم یادگیری عمیق بوده که برتری آن را نسبت به برخی روش‌ها اعلام می‌کند [۱۰]. هو و همکارانش در زمینه یادگیری عمیق، پژوهشی را انجام دادند که از یک الگوریتم سینوسی کوسینوسی بهبود یافته[1] استفاده کرده که در آن پارامتر اضافی را در الگوریتم کوسینوس سینوسی معرفی می‌کند و برای بهینه‌سازی وزن و مبتنی بر شبکه‌های عصبی برگشت به عقب[2] است. الگوریتم سینوسی کوسینوسی بهبود یافته و شبکه‌های عصبی برگشت به عقب برای ایجاد یک شبکه‌ی جدید، که همان ISCA-BPNN است، برای پیش‌بینی قیمت سهام شاخص‌های صنعتی S & P 500 و Dow Jones، ترکیب شدند. علاوه بر این، داده‌های گوگل ترندز برای بهبود پیش‌بینی سهام مورد توجه قرار گرفته‌اند. در این پژوهش، دو نوع پیش‌بینی، یک بار بدون داده‌های گوگل ترندز و بار دیگر با داده‌های گوگل ترندز تحلیل شده است. نتایج حاکی از آن بود که نسبت آمار برای مدل ISCA-BPNN با داده‌های گوگل ترندز به ۸۶.۸۱٪ برای شاخص S & P 500 و به ۸۸.۹۸٪ برای شاخص صنعتی شاخص داو جونز رسیده است. در نهایت به تاثیر گوگل ترندز در بهبود پیش‌بینی روند شاخص بازار سهام اشاره می‌کند [۱۱].

در پژوهشی دیگر [۱۲] از داده‌های سری زمانی بازار سهام برای پیش‌بینی قیمت سهام استفاده شده و شش مدل پیشنهادی متفاوت برای یافتن بهترین عملکرد در پیش‌بینی، مورد ارزیابی قرار گرفته‌اند. مجموعه داده‌ها، هشت ویژگی از جمله قیمت در لحظه‌ی باز شدن، بالاترین قیمت، پایین‌ترین قیمت، قیمت در لحظه‌ی بسته‌شدن، حجم، گردش مالی، فراز و نشیب و تغییرات را دارند. طبق نتایجی که در این پژوهش ارائه شده، مدل پیشنهادی CNN-LSTM، با کمترین مقدار خطا نسبت به سایر مدل‌های مطرح شده، دارای عملکرد بهتری است.

۲-۳- تحلیل بر اساس منابع مورد استفاده

داده‌های تولید شده در منابع مختلف و تحلیل آنها، به عنوان ابزاری برای پیش‌بینی بازار سهام کاربرد دارند. در همین راستا یو و همکارانش از داده‌های شبکه‌های اجتماعی، بلاگ‌ها و اخبار منتشر شده برای پیش‌بینی بازار سهام، استفاده کردند. آنها در شبکه‌های اجتماعی مانند توییتر و فیسبوک به‌مدت ۳۰ روز و پیام‌ها را به دو دسته مثبت و منفی تقسیم کردند. آنها برای پیش‌بینی بازار سهام از روش‌های داده‌کاوی استفاده کردند و به دقت ۷۹٪ هم دست یافتند. نتایج نشان داد که رابطه‌ی قوی بین نوسانات بازار سهام و اخبار منتشر شده وجود دارد [۱۳].

علاوه بر موارد ذکرشده، پژوهش دیگری نیز وجود دارد که اظهار می‌کند گوگل و ویکیپدیا به پایگاه‌های اطلاعاتی مهم برای سرمایه‌گذاران تبدیل شده‌اند. در این پژوهش، با ترکیب داده‌های متنوع منابع آنلاین در دوره‌های زمانی و شاخص‌های فنی برای سهام می‌توان یک سیستم روزانه کارآمدتر و هوشمندتر را ایجاد کرد. مدل‌های یادگیری ماشین، درخت‌های تصمیم‌گیری، شبکه‌های عصبی و ماشین‌های بردار پشتیبانی را به‌عنوان الگوریتم‌هایی روی سهام اپل[۳] ارائه می‌دهند و سیستم ایجاد شده به دقت ۸۵٪ در پیش‌بینی حرکت روزانه سهام اپل می‌رسد [۱۴].

۳- راهکار پیشنهادی

هدف اصلی این پژوهش، ارائه‌ی مدلی مبتنی بر یادگیری عمیق برای پیش‌بینی روند حرکت ارزش بازار سهام و تخمین قیمت در لحظه‌ی بسته شدن است. در همین راستا از شبکه‌ی اجتماعی توییتر، به‌عنوان ابزار کمکی در پیش‌بینی روند حرکت ارزش بازار سهام و تخمین قیمت در لحظه‌ی بسته شدن، استفاده می‌گردد. همچنین مدل‌پیشنهادی در این پژوهش به دو قسمت تقسیم می‌شود. ابتدا از شبکه‌ی CNN برای ارزیابی طبقه‌بندی، سپس از شبکه‌ی LSTM در جهت پیش‌بینی روند حرکت ارزش بازار سهام استفاده می‌شود.

۱-۳- طراحی کلی مدل

شمای کلی مدل در شکل (۱) نمایان است که به شرح زیر می‌باشند:

- دریافت داده‌های توییتر و پیش‌پردازش: داده‌های توییتر حول یک کلمه‌ی کلیدی در توییتر مورد جستجو قرار گرفته و پس از ذخیره‌سازی داده‌ها، عملیات پیش‌پردازش روی آنها انجام می‌شود.
- تحلیل معنایی: در این مرحله با توجه به کتابخانه‌ی موجود نظیر Spacy، TEXTBLOB، VADER و غیره به هر توییت امتیازی تخصیص داده‌می‌شود که با توجه به این امتیاز، به هر توییت برچسب ۰ یا ۱ تعلق می‌گیرد.
- ارزیابی طبقه‌بندی توییت‌ها با استفاده از شبکه‌ی عصبی CNN: نتایج حاصل از برچسب‌گذاری را ارزیابی کرده و صحت آن سنجیده می‌شود.
- دریافت داده‌های ارزش بازار سهام: این داده‌ها و از یاهو مالی و در بازه‌ی زمانی تعیین شده دریافت و ذخیره می‌شوند.
- محاسبه شاخص احساسات: توییت‌های برچسب‌گذاری شده، به‌صورت روزانه ذخیره می‌شوند؛ سپس شاخص احساسات که در ادامه توصیف خواهند شد، محاسبه می‌شوند. در نهایت داده‌های ارزش بازار سهام و شاخص احساسات در قالب یک فایل ذخیره می‌شوند.
- پیش‌بینی روند حرکت ارزش بازار سهام و ارزیابی مدل: با بهره‌گیری از شبکه‌ی LSTM به‌عنوان ساختار مورد انتظار در پیش‌بینی، روند حرکت ارزش بازار سهام در بازه زمانی مشخص شده، تخمین زده می‌شود.

۲-۳- جمع‌آوری داده‌ها

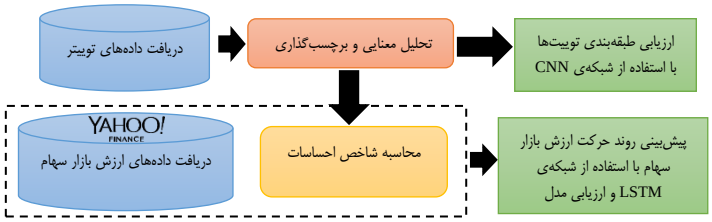
۱-۲-۳- داده‌های توییتر

شبکه‌های اجتماعی آنلاین، توجه زیادی را از کاربران و محققان در زمینه‌های مختلف جلب می‌کنند. بسیاری از محققان بر این باورند که احساسات عمومی در رسانه‌های اجتماعی به بازارهای مالی مربوط می‌شود [۱۵]. توییتر یکی از میکروبلاگ‌های مطرح در جهان است که کاربرها می‌توانند در محیط آن، به اشتراک‌گذاری پیام‌های متنی که نمایانگر افکار، علاقه و نظر آنها است، بپردازند.

در پژوهش حاضر، توییت‌ها براساس کلمه کلیدی در بازه‌ی زمانی مشخصی از ۲۰ اوت ۲۰۲۱ لغایت ۲۰ اکتبر ۲۰۲۱ برای دو شرکت Cisco و Intel از توییتر دریافت و در فرمت text ذخیره می‌شوند. عبارت های مورد جستجو برای دو شرکت مذکور در جدول (۱) نشان داده شده است.

۲-۲-۳- داده‌های ارزش بازار سهام

در این پژوهش، داده‌های دو شرکت Cisco و Intel از بازار DJIA، انتخاب و مورد ارزیابی قرار گرفتند. داده‌های ارزش بازار سهام این دو شرکت، اطلاعاتی از نوسانات بازار سهام را شامل می‌شوند که از ۲۰ اوت ۲۰۲۱ لغایت ۲۰ اکتبر ۲۰۲۱ در بازه‌ی زمانی روزانه از یاهو مالی استخراج شده‌اند. از آنجا که بازار سهام در روزهای آخر هفته و تعطیلات رسمی فعالیتی ندارند، این بازه‌ی زمانی شامل ۴۳ روز فعالیت و ۱۹ روز تعطیل رسمی که یک روز آن تعطیل رسمی مناسبتی (۶ سپتامبر ۲۰۲۱ آغاز سال نو یهودیان) می‌باشد. داده‌های ارزش بازار سهام، اطلاعات زیر را شامل می‌شوند:



شکل (۱) : شمای کلی مدل

جدول (۱) : عبارت‌های مورد جستجو

ردیف	نام شرکت	عبارت موردجستجو
۱	Cisco Systems	#CSCO, $CSCO, #Cisco, Cisco
۲	Intel	#INTC, $INTC, #Intel

- قیمت سهام در لحظه‌ی باز شدن
- قیمت سهام در لحظه‌ی بسته‌شدن
- حجم میزان سهام مبادله شده
- بیشترین قیمت سهام
- کمترین قیمت سهام
- نماد شرکت

۳-۳- پیش‌پردازش، تحلیل معنایی و برچسب گذاری

پردازش جملات محاوره‌ای پیام‌های توییتر، نیازمند تکنیک پیش‌پردازش است. در واقع، پیش‌پردازش، نقش مهمی در روند پردازش داده‌ها دارد [۱۶]. هدف از پیش‌پردازش، آماده‌سازی داده‌ها برای عملیات اصلی پردازش است. عملیات پیش‌پردازش برای داده‌های توییتر، با اعمال فرآیندهای تبدیل متن به حروف کوچک، ریشه‌یاب کلمات[۴]، توکن‌سازی[۵] و با رعایت مواردی اعم از حذف کردن توییت‌های تکراری، کلمات توقف[۶]، لینک‌ها، نشانه ها و علامت‌ها[۷] و فضاهای خالی[۸] موجود در توییت‌ها و غیره اجرا می‌شوند. همچنین به دلیل وجود تک حرف‌ها و اعداد که تاثیری در بار معنایی پیام‌ها ندارند، تکرار حروف که برای نمایش میزان شدت احساسات در پیام به‌کار می‌روند، نیز لازم است که آنها را حذف کنیم.

پس از آماده‌سازی داده‌ها توسط عملیات پیش‌پردازش، با کتابخانه‌ی TextBlob که دارای یک تابع تجزیه و تحلیل احساسات است؛ امتیازی تحت عنوان قطبیت[۹] به هر توییت اختصاص داده می‌شود. بر اساس مقدار قطبیت، امتیاز احساسات، مشخص شده و محاسبه به این‌صورت بوده که اگر امتیاز کمتر از ۰ باشد، احساسات منفی و اگر بیشتر از ۰ باشد، احساسات مثبت برگردانده می‌شود [۱۷]. پس از تحلیل معنایی، برچسب‌گذاری توییت‌ها و تقسیم آنها به دو دسته توییت‌های مثبت (برچسب ۱) و توییت‌های منفی (برچسب ۰) صورت می‌گیرد.

۴-۳- محاسبه‌ی شاخص احساسات

در این پژوهش از شاخص احساسات به‌عنوان ورودی‌های شبکه‌ی عصبی LSTM جهت پیش‌بینی روند حرکت ارزش بازار سهام استفاده می‌گردد. در ادامه این شاخص‌ها توصیف خواهند شد.

۱-۴-۳- شاخص کلی احساسات

در پژوهش [۱۸] طبق فرمول (۱) از یک شاخص کلی احساسات روزانه (BI_t) استفاده شده که $M^{bullish}$ بیانگر تعداد نظرات صعودی و $M^{bearish}$ تعداد نظرات نزولی روزانه است. گرایش کلی عاطفی سرمایه گذار توسط این شاخص نشان داده می‌شود.

$$BI_t = \ln \frac{1 + M_t^{bullish}}{1 + M_t^{bearish}} \quad [۱۸](۱)$$

در همین راستا، شاخص کلی احساسات روزانه را در پژوهش حاضر، برای توییت‌های منتشر شده در هر روز محاسبه کرده و همراه با داده‌های ارزش بازار سهام ذخیره می‌شوند.

۲-۴-۳- شاخص احساسات مثبت روزانه و شاخص احساسات منفی روزانه

برای توییت‌های منتشر شده در هر روز، دو شاخص احساسات روزانه با عنوان شاخص احساسات مثبت روزانه و شاخص احساسات منفی روزانه محاسبه می‌شوند که به‌ترتیب دو فرمول (۲) و (۳) می‌باشند؛ سپس نتایج همراه با داده‌های ارزش بازار سهام ذخیره می‌شوند.

$$sentiment_index_positive = \frac{positivetweets}{Totaltweets} \quad (۲)$$

$$sentiment_index_negative = \frac{negativetweets}{Totaltweets} \quad (۳)$$

در این فرمول‌ها، positivetweets تعداد توییت‌های مثبت هر روز و negativetweets تعداد توییت‌های منفی هر روز است. همچنین تعداد کل توییت‌ها در هر روز با Totaltweets نشان داده شده است.

۵-۳- محاسبه دو معیار اساسی HLPCT و PCT change

در پژوهشی جان کردونیس و همکارانش از دو معیار اساسی HLPCT و PCT change برای پیش‌بینی قیمت سهام از طریق تحلیل احساسات در توییتر استفاده کردند. این دو معیار برای یافتن ارتباط بین توییت‌ها و بازار سهام کاربرد دارند. این دو معیار در فرمول (۴) و (۵) تعریف شده‌اند [۱۹]. در پژوهش حاضر نیز از این دو معیار به عنوان ورودی های شبکه‌ی LSTM در جهت پیش‌بینی روند حرکت ارزش بازار سهام استفاده شده است.

$$HLPCT = \frac{High - Low}{Low} \quad (۴)$$

$$PCTchange = \frac{Close - Open}{Open} \quad (۵)$$

در این فرمول‌ها High بیانگر بیشترین قیمت سهام، Low کمترین قیمت سهام، Open قیمت سهام در لحظه‌ی باز شدن و Close قیمت در لحظه‌ی بسته شدن است.

۶-۳- استفاده از شبکه‌های یادگیری عمیق

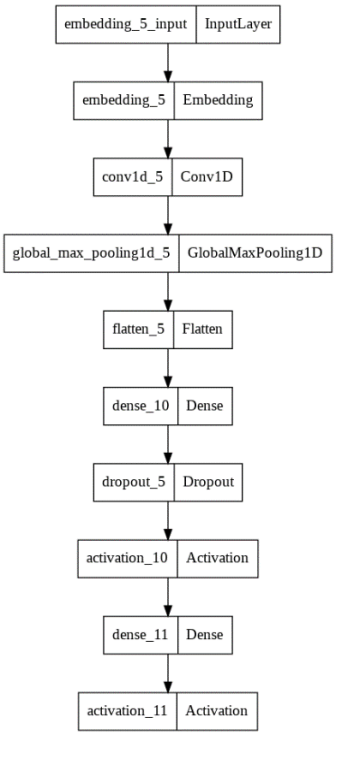
۱-۶-۳- طبقه‌بندی توییت‌ها با استفاده از شبکه‌ی CNN

مدل word2vec بخش بزرگی از متن ها می‌گیرد و یک بردار متناظر در فضا تولید می‌کند که هدف از ایجاد چنین مدل برداری، تولید یک بردار اعشاری به ازای هر کلمه در مجموعه داده است [۲۰]. پس از پاکسازی داده‌های توییتر، با استفاده از کتابخانه‌ی genism مدل برداری کلمات ایجاد می‌شود. سپس بردارها به مدل CNN وارد می‌شوند. در شبکه‌ی CNN، ابتدا لایه تعبیه‌ساز[۱۰] قرار گرفته که روش word2vec برای تبدیل کلمات به بردار عددی در قالب لایه تعبیه‌ساز استفاده شده است. سپس لایه conv و در انتها طبقه‌بندی داده‌ها در دو دسته صورت می‌گیرد. طرح گرافیکی لایه‌ها در مدل CNN، در شکل (۲) و (۳) به‌ترتیب مربوط به شرکت Cisco و Intel هستند.

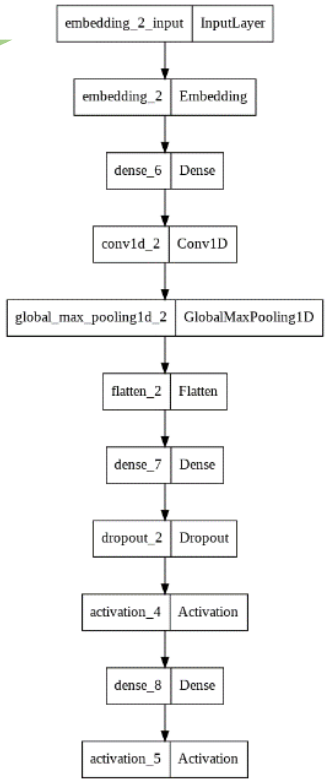
۳-۶-۲- پیش‌بینی با استفاده از شبکه‌ی LSTM

مدل LSTM ابتدا با یک حالت پایه، سپس با چهار حالت مختلف برای پیش‌بینی روند حرکت ارزش بازار سهام و تخمین قیمت در لحظه‌ی بسته شدن بررسی می‌شود. در این مدل یک تابع window_data() به کار رفته است که قیمت در لحظه‌ی بسته‌شدن مربوط به روزهای گذشته را ذخیره می‌کند تا برای پیش‌بینی قیمت روز بعد استفاده کند که مقدار آن برای مدل شرکت Cisco و Intel به‌ترتیب ۴ و ۲ قرار داده شده است. حالات مختلف مدل پیشنهادی به تفکیک شرح و نشان داده شده است.

- حالت پایه: در شکل (۴)، ویژگی‌ها شامل پنج ویژگی موجود در داده‌های ارزش بازار سهام است.
- حالت اول: در شکل (۵)، ویژگی‌ها شامل پنج ویژگی موجود در داده‌های ارزش بازار سهام و ویژگی ششم که همان شاخص کل احساسات محاسبه شده از داده‌های روزانه توییتر است.
- حالت دوم: در شکل (۶)، ویژگی‌ها شامل پنج ویژگی موجود در داده‌های ارزش بازار سهام و ویژگی ششم و هفتم که شاخص مثبت احساسات و شاخص منفی احساسات محاسبه شده از داده‌های روزانه توییتر است.
- حالت سوم: در شکل (۷)، ویژگی‌ها شامل HLPCT و PCTchange و ویژگی سوم نیز شاخص کل احساسات محاسبه شده از داده‌های روزانه توییتر است.

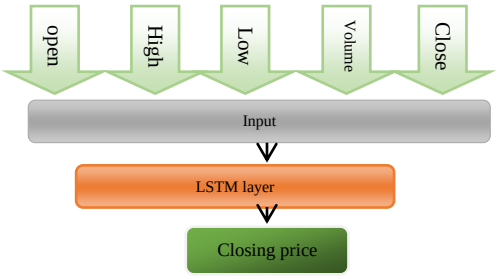


شکل (۳) : طرح گرافیکی لایه‌ها شرکت Intel

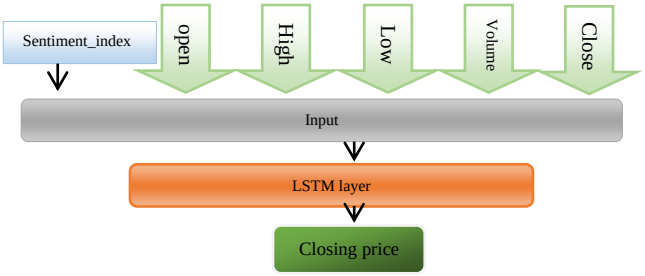


شکل (۲) : طرح گرافیکی لایه‌ها شرکت Cisco

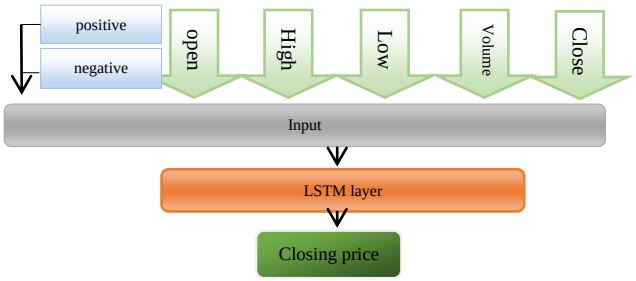
- حالت چهارم: در شکل (۸)، ویژگی‌ها شامل HLPCT و PCTchange و ویژگی سوم و چهارم که شاخص مثبت احساسات و شاخص منفی احساسات محاسبه شده از داده‌های روزانه توییتر است.



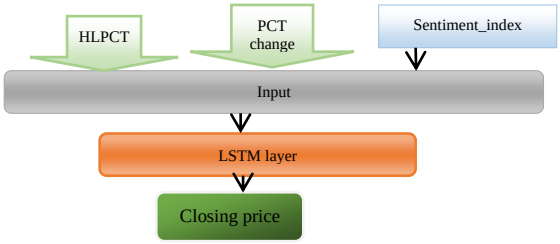
شکل (٤) : شمای حالت پایه مدل پیشنهادی



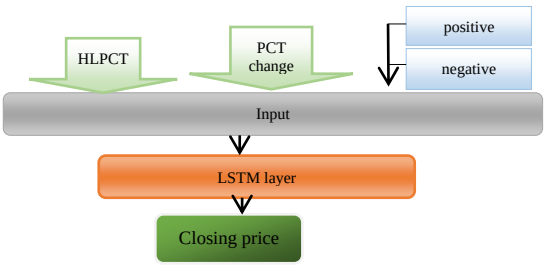
شکل (٥) : شمای حالت اول مدل پیشنهادی



شکل (٦) : شمای حالت دوم مدل پیشنهادی



شکل (۷) : شمای حالت سوم مدل پیشنهادی



شکل (۸) : شمای حالت چهارم مدل پیشنهادی

در این پژوهش از شاخص احساسات مثبت و منفی روزانه، همچنین ترکیب این دو شاخص با معیارهای HLPCT و PCT change در پیش‌بینی، به عنوان نوآوری‌های روش پیشنهادی استفاده شده است.

۳-۷- عملکرد معیارهای ارزیابی

هر مدل پس از نهایی شدن، بررسی شده و قابلیت تعمیم آن ارزیابی می‌شود.

۳-۷-۱- معیار ارزیابی در طبقه‌بندی

برای اعتبارسنجی دسته‌بندها، از دو معیار صحت[۱۱] و زیان[۱۲] که در ارزیابی طبقه‌بندی‌ها بیشترین کاربرد را دارند؛ استفاده می‌شود. طبق فرمول (۶) معیار صحت، نسبت بین تعداد پیش‌بینی‌های صحیح به تعداد کل پیش‌بینی‌ها است.

$$Accuracy = \frac{Number_of_correct_prediction_s}{Total_number_of_prediction_s} \quad (۶)$$

۳-۷-۲- معیار ارزیابی در پیش‌بینی

برخی از معیارهای ارزیابی که بیشتر در پیش‌بینی‌ها کاربرد دارند براساس فرمول‌های (۷) و (۸) و (۹) شامل خطای جذر میانگین مربعات[۱۳]، میانگین خطای مطلق[۱۴] و ضریب تعیین[۱۵] یا R^2 می‌باشند.

$$RMSE = \sqrt{\frac{1}{n}\sum(actualvalue - forecastvalue)^2} \quad (۷)$$

$$MAE = \frac{1}{n}\sum|actualvalue - forecastvalue| \quad (۸)$$

$$R^2 = 1 - \frac{\frac{1}{n}\left(\sum(actualvalue - forecastvalue)^2\right)}{\frac{1}{n}\left(\sum(\overline{y_i} - forecastvalue)^2\right)} \quad (۹)$$

متغییر actualvalue بیانگر مقدار واقعی و متغییر forecastvalue معرف مقدار پیش‌بینی شده است. متغییر y_i نشان دهنده میانگین مقادیر واقعی مجموعه داده‌های آزمایش است.

باید توجه داشت که در ارزیابی نهایی برای انتخاب بهترین مدل مناسب، مدلی انتخاب شود که دارای کمترین مقدار RMSE و MAE و بیشترین مقدار R^2 است.

۴- نتایج

۴-۱- نتایج ارزیابی طبقه‌بندی

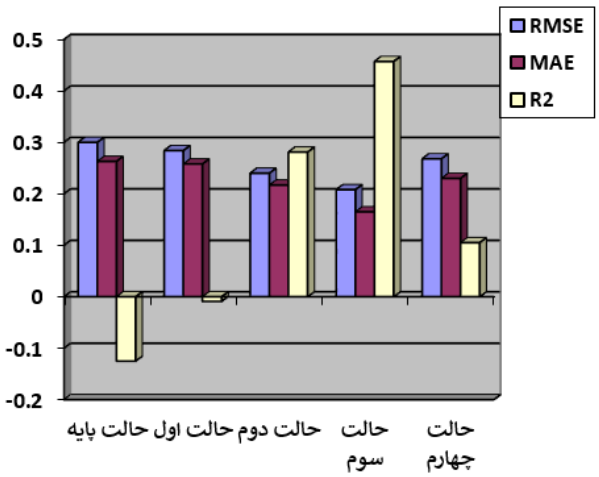
در مدل‌های شبکه‌ی یادگیری عمیق، قضاوت شخصی و تجربه در تعیین پارامترها برای ایجاد مدل بهینه، اهمیت دارد؛ همچنین بهترین مدل در حوزه‌ی یادگیری عمیق، فقط با آزمون و خطا بدست می‌آید. با پیاده‌سازی مدل CNN، بیشترین میزان صحت و کمترین میزان زیان در طبقه‌بندی برای داده‌های شرکت Cisco به‌ترتیب معادل ۹۸/۸۹٪ و ۵/۱۵٪ و برای داده‌های شرکت Intel به‌ترتیب معادل ۹۶/۹۴٪ و ۱۰/۵۰٪ بدست آمده است.

۴-۲-نتایج ارزیابی مدل پیش‌بینی

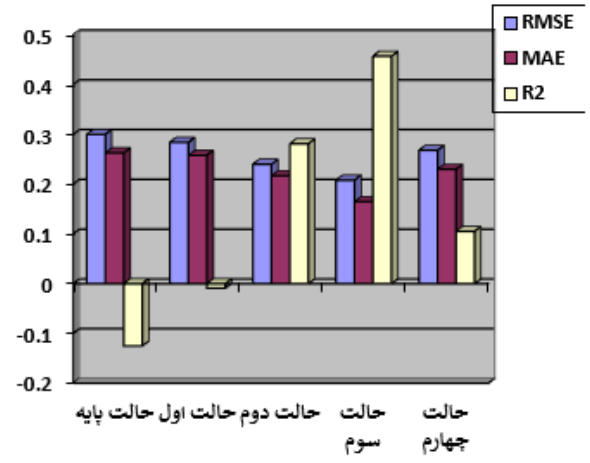
در مدل شبکه‌ی LSTM، به‌منظور جلوگیری از overfitting شدن، پارامتر dropout در مدل شرکت Cisco با اندازه ۰/۵ و برای مدل شرکت Intel با اندازه ۰/۲ تعریف شده است. با تکیه بر نتایج بدست آمده، برترین حالت در پیش‌بینی روند حرکت ارزش بازار سهام شرکت Cisco از مقایسه حالت پایه با چهار حالت پیاده‌سازی شده، حالت چهارم دارای خطای جذر میانگین مربعات، میانگین خطای مطلق و ضریب تعیین به‌ترتیب معادل ۰/۱۱، ۰/۰۹ و ۰/۸۰ است. همچنین برترین حالت برای شرکت Intel، حالت سوم با خطای جذر میانگین مربعات، میانگین خطای مطلق و ضریب تعیین به‌ترتیب معادل ۰/۲۰، ۰/۱۶ و ۰/۴۵ است. در شکل (۹) و (۱۰) نتایج تمامی حالت‌ها برای شرکت‌های Cisco و Intel نشان داده شده است.

در جدول (۲) نتایج حالت پایه و حالت اول و دوم پژوهش حاضر و نتایج دو مدل از پژوهشی دیگر [۱۸] که برای پیش‌بینی بازار سهام مبتنی بر یادگیری عمیق و احساسات سرمایه‌گذاران ارائه دادند، مشاهده می‌کنید. در پژوهش [۱۸]، یک مدل LSTM به‌عنوان مدل پایه و یک مدل S_LSTM را تعریف کردند که در آن، شاخص احساسات به مدل پایه افزوده شده است.

با مقایسه تمامی این نتایج، به بهبود نتایج RMSE و MAE در پژوهش حاضر می‌رسیم که تاثیر وجود شاخص کلی احساسات، شاخص احساسات مثبت و منفی روزانه در پیش‌بینی بازار سهام دو شرکت بوده که یعنی کمترین



شکل :(۹) نتایج شبکه‌ی LSTM برای شرکت Cisco



شکل :(۱۰) نتایج شبکه‌ی LSTM برای شرکت Intel

جدول (۲) : نتایج برتر پژوهش حاضر و پژوهشی دیگر [۱۸]

مدل	RMSE	MAE
LSTM[۱۸]	۸/۷۱	۷/۰۳
حالت پایه شرکت Cisco	۰/۲۳	۰/۱۹
حالت پایه شرکت Intel	۰/۳۰	۰/۲۶
S_LSTM[۱۸]	۴/۴۸	۳/۳۲
حالت اول شرکت Cisco	۰/۲۲	۰/۱۹
حالت اول شرکت Intel	۰/۲۸	۰/۲۵
حالت دوم شرکت Cisco	۰/۲۲	۰/۱۹
حالت دوم شرکت Intel	۰/۲۴	۰/۲۱

میزان خطا در پیش‌بینی است. بصورت کلی، نقش اساسی شاخص احساسات در عملکرد بهینه مدل پیشنهادی پژوهش حاضر در پیش‌بینی بازار سهام قابل توجه است.

۴-۳- نتیجه‌گیری

در این پژوهش برای پیش‌بینی سهام دو شرکت Cisco و Intel، نتایج تجربی نشان می‌دهند که وجود شاخص‌های احساسات بر پیش‌بینی روند حرکت ارزش بازار سهام تاثیر می‌گذارد. همچنین محاسبه دو معیار HLPCT و PCTchange، در پیش‌بینی بازار سهام با استفاده از شبکه‌ی یادگیری عمیق LSTM سودمند است.

مراجع

[1] R. Ramezanian, A. Peymanfar, and S. B. Ebrahimi, "An integrated framework of genetic network programming and multi-layer perceptron neural network for prediction of daily stock return: An application in Tehran stock exchange market," *Applied soft computing*, vol. 82, p. 105551, 2019.
[2] P. K. Narayan, "Profitability of technology-investing Islamic and non-Islamic stock markets," *Pacific-Basin Finance Journal*, vol. 52, pp. 70-81, 2018.
[3] S. A. Al-Thaqeb, "Do international markets overreact? Event study: International market reaction to US local news events," *Research in International Business and Finance*, vol. 44, pp. 369-385, 2018.
[4] T. P. Wisniewski, "Is there a link between politics and stock returns? A literature survey," *International Review of Financial Analysis*, vol. 47, pp. 15-23, 2016.
[5] X. Zhang, J. Shi, D. Wang, and B. Fang, "Exploiting investors social network for stock prediction in China's market," *Journal of computational science*, vol. 28, pp. 294-303, 2018.
[6] A. S. Wafi, H. Hassan, and A. Mabrouk, "Fundamental analysis models in financial markets–review study," *Procedia economics and finance*, vol. 30, pp. 939-947, 2015.
[7] R. T. F. Nazário, J. L. e Silva, V. A. Sobreiro, and H. Kimura, "A literature review of technical analysis on stock markets," *The Quarterly Review of Economics and Finance*, vol. 66, pp. 115-126, 2017.
[8] S. Prasanna, D. J. I. J. o. C. S. Ezhilmaran, and E. Technology, "An analysis on stock market prediction using data mining techniques," vol. 4, no. 3, pp. 49-51, 2013.
[9] A. Derakhshan and H. J. E. A. o. A. I. Beigy, "Sentiment analysis on stock social media for stock price movement prediction," vol. 85, pp. 569-578, 2019.
[10] M. Kraus, S. Feuerriegel, and A. J. E. J. o. O. R. Oztekin, "Deep learning in business analytics and operations research: Models, applications and managerial implications," 2019.
[11] H. Hu, L. Tang, S. Zhang, and H. J. N. Wang, "Predicting the direction of stock markets using optimized neural networks with Google Trends," vol. 285, pp. 188-195, 2018.
[12] W. Lu, J. Li, Y. Li, A. Sun, and J. Wang, "A CNN-LSTM-based model to forecast stock prices," *Complexity*, vol. 2020, 2020.
[13] Y. Yu, W. Duan, and Q. Cao, "The impact of social and conventional media on firm equity value: A sentiment analysis approach," *Decision support systems*, vol. 55, no. 4, pp. 919-926, 2013.
[14] B. Weng, M. A. Ahmed, and F. M. Megahed, "Stock market one-day ahead movement prediction using disparate data sources," *Expert Systems with Applications*, vol. 79, pp. 153-163, 2017.
[15] Y. Ruan, A. Durresi, and L. Alfantoukh, "Using Twitter trust network for stock market analysis," *Knowledge-Based Systems*, vol. 145, pp. 207-218, 2018.
[16] D. Ramachandran and R. Parvathi, "Analysis of twitter specific preprocessing technique for tweets," *Procedia Computer Science*, vol. 165, pp. 245-251, 2019.
[17] A. S. Neogi, K. A. Garg, R. K. Mishra, and Y. K. Dwivedi, "Sentiment analysis and classification of Indian farmers' protest using twitter data," *International Journal of Information Management Data Insights*, vol. 1, no. 2, p. 100019, 2021.
[18] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Computing and Applications*, vol. 32, no. 13, pp. 9713-9729, 2020.
[19] J. Kordonis, S. Symeonidis, and A. Arampatzis, "Stock price forecasting via sentiment analysis on Twitter," in *Proceedings of the 20th Pan-Hellenic Conference on Informatics*, 2016, pp. 1-6.
[20] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013, pp. 3111-3119.

زیرنویس‌ها

[1] ISCA
[2] BPNN
[3] APPL
[4] stemming
[5] tokenization[6] stopwords
[7] punctuations
[8] whitespaces
[9] polarity
[10] Embedding layer
[11] Accuracy
[12] Loss
[13] RMSE
[14] MAE

[15] R_squared