پردیس علوم
دانشکده ریاضی، آمار و علوم کامپیوتر

روانشناسی بازارهای مالی بر مبنای داده‌های شبکه‌های اجتماعی

نگارنده

زهرا خطیبی

استاد راهنما: سمانه افتخاری مهابادی

پایان‌نامه برای دریافت درجه کارشناسی
در رشته علوم کامپیوتر

تابستان ۱۴۰۲

چکیده

هدف اصلی این پژوهش، بررسی تاثیر احساسات کاربران شبکه‌های اجتماعی بر روی بازارهای مالی است. نتایج نشان می‌دهد که احساسات کاربران شبکه‌های اجتماعی می‌تواند به عنوان یکی از عوامل تأثیرگذار در حرکت بازارهای مالی در نظر گرفته شود. به این منظور، ابتدا به معرفی مفاهیم مقدماتی از حوزه مالی و حوزه پردازش زبان طبیعی می‌پردازیم. سپس به مقدمه‌ای از تحلیل احساسات متنی، کاربردها و روش‌های متفاوت آن پرداخته می‌شود. در بخش‌های اصلی به معرفی و شرح کامل روش تحلیل احساسات استفاده شده در این پژوهش خواهیم پرداخت و در نهایت تأثیر احساسات کاربران بر روی بازار بورس اوراق بهادار بررسی می‌کنیم. در این بخش، با استفاده از روش تحلیل احساسات متنی معرفی شده، احساسات کاربران درباره بازار بورس اوراق بهادار تحلیل می‌شود. سپس، با استفاده از روش‌ها و مدل‌های آماری، رابطه بین احساسات کاربران و حرکت بازار بورس اوراق بهادار مورد بررسی قرار می‌گیرد.

کلمات کلیدی: بازارهای مالی، پردازش زبان طبیعی، تحلیل احساسات، برچسب‌گذاری احساسی، بازار بورس اوراق بهادار

سپاسگزاری

با احترام و تقدیر، سپاسگزارم از اساتید راهنمای عزیز، سرکار خانم دکتر سمانه افتخاری مهابادی، عضو محترم هیئت علمی گروه آمار دانشکده ریاضی، آمار و علوم کامپیوتر دانشگاه تهران و جناب آقای دکتر مهدی نوری، عضو محترم هیئت علمی دانشکده اقتصاد دانشگاه تهران که به لطف دانش، تجربه و راهنمایی‌های ارزشمند این دو بزرگوار، انجام این پژوهش بر من هموار شد.
همچنین از خانواده عزیز و مهربانم نهایت قدردانی را دارم که همواره در تمام عرصه‌های زندگی، پشتیبان و مشوّق من بوده‌اند. هرچه تا کنون داشته و در آینده خواهم داشت، متعلق به آن‌هاست.

پیشگفتار

بازارهای مالی از دیدگاه روانشناسی یکی از مهم‌ترین و پیچیده‌ترین حوزه‌های مطالعاتی هستند که همواره مورد توجه پژوهشگران و محققان قرار گرفته‌اند. در سال‌های اخیر با ظهور شبکه‌های اجتماعی، این حوزه به یک بعد جدید از پژوهش‌ها و مطالعات در حوزه روانشناسی بازارهای مالی تبدیل شده است.

بازارهای مالی سیستم‌های پیچیده‌ای هستند که تحت تأثیر طیف وسیعی از عوامل مختلف از جمله شاخص‌های اقتصادی، رویدادهای سیاسی و احساسات سرمایه‌گذاران هستند. احساسات سرمایه گذاران، محرک اصلی رفتار بازار است، زیرا منعکس کننده باورها، نگرش و احساسات شرکت کنندگان در بازارهای مالی است. اگر چه در مدل‌سازی‌ها، رفتار شرکت کنندگان بازارهای مالی، رفتاری عقلایی در نظر گرفته می‌شود، اما واقعیت این است که شرکت‌کنندگان همواره عقلایی نیستند و احساسات و باورها در تصمیم‌گیری آن‌ها موثر است. پژوهش‌های مالی بسیاری در زمینه اقتصاد مالی‌رفتاری، مهر تاییدی بر این موضوع است [۱].

با توجه به روند روزافزون استفاده از شبکه‌های اجتماعی، ارتباط بین فعالیت کاربران در این شبکه‌ها و بازارهای مالی بیشتر به چشم می‌خورد. در حالی که در گذشته، اطلاعات مربوط به بازارهای مالی بیشتر از طریق رسانه‌های سنتی مانند رادیو، تلویزیون یا روزنامه‌ها منتشر می‌شد، امروزه با استفاده از شبکه‌های اجتماعی، کاربران می‌توانند به راحتی و به سرعت اخبار و تحلیل‌های مربوط به بازارهای مالی را دریافت و منتشر کنند. این پلتفرم‌ها به کانال‌های محبوبی برای سرمایه‌گذاران تبدیل شده‌اند تا نظرات و باورهای خود را در مورد بازارهای مختلف بیان کنند. رفتار سرمایه‌گذاران که از عوامل مختلفی از جمله ادراک و حس آنان سرچشمه می‌گیرد، بر فرآیند تصمیم‌گیری آنان تأثیر زیادی می‌گذارد. به عنوان مثال، اگر کاربران در شبکه‌های اجتماعی در مورد یک شرکت یا صنعت

خاصی با احساسات مثبت صحبت کنند، این می‌تواند نشانه‌ای از روند رو به رشد بازارهای مربوط به آن شرکت یا صنعت باشد. به طور مشابه، اگر کاربران در شبکه‌های اجتماعی در مورد یک شرکت یا صنعت با احساسات منفی صحبت کنند، این می‌تواند نشانه‌ای از روند کاهشی بازارهای مربوط به آن شرکت یا صنعت باشد.

هدف این پژوهش شناسایی الگوهایی در احساسات سرمایه گذاران و تاثیر آن‌ها بر بازارهای مالی است. شناخت و شناسایی الگوها در احساسات سرمایه‌گذاران، نقش مهمی در پیش‌بینی بازارهای مالی دارد و این مهم می‌تواند به کسب سود بیشتر و شناخت روند تغییرات بازارهای مالی کمک کند [۲].

پیشرفت تکنیک‌های هوش مصنوعی مانند شبکه‌های عصبی پیچیده و پردازش زبان طبیعی، نقش بسزایی در شناسایی احساسات انسان‌ها دارند. محققان بسیاری در تلاش‌اند با کمک این تکنیک‌ها، احساسات و نگرش افراد را از روی متن، صدا و... شناسایی کنند [۳].

با بررسی نظرات کاربران در شبکه‌های اجتماعی، می‌توان به بررسی تمایلات و نگرش‌های آنان نسبت به شرکت‌ها و صنایع مختلف پرداخت. این اطلاعات می‌تواند برای پیش‌بینی روند بازارهای مالی و تصمیم‌گیری‌های سرمایه‌گذاری مفید باشد. به طور مشابه، با استفاده از مدل‌های یادگیری عمیق، می‌توان روند بازارهای مالی را با دقت بیشتری پیش‌بینی کرد [۴]. بنابراین نتایج این مطالعه می‌تواند برای سرمایه‌گذاران، تحلیل‌گران و سایر افرادی که به بازارهای مالی علاقه‌مند هستند، سودمند باشد.

فهرست مطالب

فصل ۱

مفاهیم مقدماتی

۱.۱ تعاریف مقدماتی

بازار مالی

به مجموعه‌ای از ابزارهای مالی و مکانیزم‌هایی گفته می‌شود که برای خرید و فروش دارایی‌های مالی مورد استفاده قرار می‌گیرد، بازار مالی گفته می‌شود. این بازارها شامل مجموعه‌ای از ابزارهای مالی مانند سهام، اوراق بهادار، ارز، کالاها و دیگر ابزارهای مالی هستند که به عنوان ابزاری برای جذب سرمایه و ارتباط بین سرمایه‌گذاران و شرکت‌ها و دولت‌ها به کار می‌روند.

شبکه‌های اجتماعی

شبکه‌های اجتماعی، پلتفرم‌هایی هستند که برای برقراری ارتباط و اشتراک گذاری اطلاعات و تجربیات فردی به کار می‌روند. این شبکه‌ها به عنوان یکی از ابزارهای ارتباطی اصلی در دنیای امروز، از طریق اینترنت و تکنولوژی‌های مرتبط با آن ارائه می‌شوند. برخی از شبکه‌های اجتماعی

معروف شامل فیسبوک [1]، توییتر [2]، اینستاگرام [3]، لینکدین [4] و یوتیوب [5] هستند.

احساسات

تعریف احساسات به عنوان یک تجربه شخصی و شناختی از یک وضعیت خاص صورت می‌گیرد. این تجربه‌ها شامل واکنش‌های شناختی، فیزیولوژیکی و رفتاری به محرک‌های مختلف هستند. احساسات انواع مختلف و طیف وسیعی دارند، از جمله شادی، ناراحتی، ترس، عصبانیت، عشق، خشم، تعجب. این تجربه‌ها ممکن است به صورت ناخودآگاه و یا طبیعی به وجود بیایند و یا با تمرین و آموزش قابل کنترل و مدیریت باشند. احساسات به عنوان بخشی از تجربه زندگی ما، در ارتباط با دیگران، تصمیم‌گیری‌های ما و رفتارهای ما نقش مهمی ایفا می‌کنند. همچنین در علوم رفتاری، احساسات به عنوان موضوعی مورد بررسی قرار می‌گیرند و در مطالعات روان‌شناسی، علوم شناختی و علوم اعصاب مورد توجه خاصی قرار دارد.

۲.۱ تعاریف حوزه مالی

در حوزه مالی، اصطلاحات مختلفی به کار می‌رود. این مفاهیم در درک بهتر فرایندهای مالی کمک می‌کنند. در زیر تعاریف برخی از مفاهیم حوزه مالی آمده است.

روانشناسی بازار مالی

به بررسی و تحلیل مکانیزم‌ها و الگوهای رفتاری معامله‌گران و فعالان بازارهای مالی، روانشناسی بازار مالی گفته می‌شود.

[1] Facebook
[2] Twitter
[3] Instagram
[4] LinkedIn
[5] YouTube

ابزار مالی

به هر نوع دارایی قابل معامله، اعم از پول نقد، ملک، سند رسمی یا قانونی مانند برگه سهام، اوراق بهادار و اوراق قرضه که ارزش مبادله‌ای داشته باشند، ابزار مالی گفته می‌شود.

اوراق بهادار

یک ابزار مالی دارای ارزش پولی است که قابلیت تبدیل به پول نقد را دارد. اوراق بهادار می‌تواند نمایانگر ایفای مالکیت در یک شرکت سهامی عام (از طریق سهام) یا رابطه طلبکاری با یک نهاد دولتی یا یک شرکت (از طریق اوراق قرضه) یا حقوق مالکیت دیگری باشد.

سرمایه‌گذاری

سرمایه‌گذاری به معنی اختصاص دادن پول برای چیزی با انتظار سود و منافع از آن در آینده است. به‌طور دقیق‌تر سرمایه‌گذاری تعهد پول یا سرمایه برای خرید وسایل یا دارایی‌های دیگر، به منظور به دست آوردن منافع از آن است. در علم اقتصاد، سرمایه‌گذاری یعنی خرید کالایی که اکنون مصرف نمی‌شود اما در آینده فرد به آن نیاز پیدا خواهد کرد و آن کالا برای او سودآور خواهد بود. در علم مالی، سرمایه‌گذاری به این معنی است که فرد یک دارایی مالی نظیر سهام را می‌خرد و پیش‌بینی می‌کند که آن دارایی مالی در آینده سودآور خواهد بود و قیمتش افزایش خواهد یافت، لذا با فروش به قیمت بالاتر سود به دست خواهد آورد.

بازار بورس

بورس یک بازار مالی است که در آن سهام، اوراق بهادار و دیگر ابزارهای مالی توسط خریداران و فروشندگان معامله می‌شود.

نوسانات بازار

نوسانات بازار به تغییرات قیمت ابزارهای مالی در بازار مالی اشاره دارد که می‌تواند تحت تأثیر عوامل مختلفی مانند شرایط اقتصادی، سیاسی، فرهنگی و احساسات شرکت کنندگان باشد.

ریسک

ریسک به میزان عدم قطعیت و تأثیر پذیری سرمایه‌گذاری در بازار مالی اشاره دارد. ریسک می‌تواند ناشی از عوامل مختلفی مانند تغییرات نوسانات بازار، شرایط اقتصادی و سیاسی باشد.

مدیریت ریسک

به مجموعه روش‌ها و استراتژی‌هایی گفته می‌شود که برای کاهش ریسک سرمایه‌گذاری در بازار مالی به‌کار می‌روند.

بازده

بازده به میزان سود و زیانی اشاره دارد که از سرمایه‌گذاری در بازار مالی به دست می‌آید. بازده می‌تواند به صورت سالانه و یا برای دوره‌های زمانی مختلفی از جمله روزانه، هفتگی، ماهانه و فصلی محاسبه شود.

ارزش

ارزش به میزان مالی ابزارهای مالی اشاره دارد. ارزش می‌تواند از طریق نرخ سودآوری، نرخ بازده، نرخ نقدشوندگی و دیگر شاخص‌های مالی محاسبه شود.

تحلیل بنیادی

تحلیل بنیادی به روش‌هایی گفته می‌شود که برای بررسی مؤلفه‌های اساسی اقتصادی، مالی و سیاسی که تأثیر مستقیم و غیرمستقیمی بر نوسانات بازار دارند، استفاده می‌شود.

تحلیل تکنیکال

تحلیل فنی به روش‌هایی گفته می‌شود که برای تحلیل نمودارهای قیمت ابزارهای مالی و پیش‌بینی نوسانات بازار استفاده می‌شود.

۳.۱ تعاریف حوزه پردازش زبان طبیعی

در حوزه پردازش زبان طبیعی اصطلاحات مختلفی به کار می‌رود. این مفاهیم در درک بهتر مفاهیم و فرایندهای پردازش زبان کمک می‌کنند. در زیر تعاریف برخی از مفاهیم حوزه پردازش زبان طبیعی آمده است.

پردازش زبان طبیعی

پردازش زبان طبیعی به کاربرد تکنولوژی برای پردازش و تحلیل زبان‌های طبیعی، مانند زبان انگلیسی، فارسی، عربی و سایر زبان‌ها، به منظور استخراج اطلاعات مفید و استفاده از آن‌ها در فرایندهای مختلف مانند تحلیل متن، ترجمه ماشینی و سیستم‌های گفتاری اشاره دارد.

متن

متن به مجموعه‌ای از کلمات و جملات مرتبط با یکدیگر گفته می‌شود که به منظور انتقال اطلاعات در زبان طبیعی به کار می‌روند.

بردار

به مجموعه‌ای از اعداد یا واژه‌هایی گفته می‌شود که به منظور نمایش مجموعه‌ای از ویژگی‌های یک متن استفاده می‌شود. بردارها برای استفاده در الگوریتم‌های یادگیری ماشینی و تحلیل متن به کار می‌روند.

تحلیل متن

به فرایند تحلیل و استخراج اطلاعات مفید از متن‌های طبیعی با استفاده از الگوریتم‌های پردازش زبان طبیعی می‌گویند. این فرایند شامل استخراج اطلاعات مانند دسته‌بندی متن، استخراج ارتباطات و ارزیابی احساسات متن است.

استخراج ارتباطات

به فرایند استخراج ارتباطات موجود در متون طبیعی با استفاده از الگوریتم‌های پردازش زبان طبیعی اشاره دارد. این فرایند شامل تشخیص و استخراج ارتباطات بین اجزای مختلف متن مانند کلمات، جملات و پاراگراف‌ها است.

تحلیل احساسات

به فرایند تحلیل و استخراج اطلاعات مربوط به احساسات موجود در متون طبیعی اشاره دارد. این فرایند شامل تشخیص و استخراج احساسات مثبت، منفی و خنثی موجود در متون است و می‌تواند برای تحلیل احساسات کاربران شبکه‌های اجتماعی و ارتباط آن با بازارهای مالی استفاده شود.

ترجمه ماشینی

ترجمه ماشینی به شیوه‌ای از ترجمه متن با استفاده از الگوریتم‌های پردازش زبان طبیعی با استفاده از کامپیوتر اشاره دارد. این فرایند شامل ترجمه متون از یک زبان به زبان دیگر است و می‌تواند برای ترجمه متون موجود در شبکه‌های اجتماعی به کار رود.

۴.۱ معرفی کتابخانه‌ها

۱.۴.۱ کتابخانه Pandas

کتابخانه Pandas یکی از قدرتمندترین و محبوب‌ترین کتابخانه‌های پایتون برای کار با داده‌های برداری و ماتریسی است. این کتابخانه به کاربر این امکان را می‌دهد تا با داده‌های بزرگ و پیچیده کار کند و با سرعت بسیار بالایی آن‌ها را تحلیل کند.
کتابخانه Pandas با استفاده از توابع متعددی مانند توابع انتخاب، ترکیب، تفکیک و تغییر نوع داده، به کاربر اجازه می‌دهد تا داده‌های خود را به آسانی تحلیل کند.
یکی از دلایل برتری Pandas، سرعت بالای آن است. این کتابخانه با استفاده از آرایه‌های نام‌گذاری شده [۶] و بهینه‌سازی‌هایی مانند استفاده از مفهوم «Broadcasting» به کاربر امکان تحلیل سریع داده‌های خود را می‌دهد.
Broadcasting یکی از ویژگی‌های مهم کتابخانه Pandas است که به کاربر اجازه می‌دهد تا عملیات مختلف را بر روی داده‌های با ابعاد مختلف بدون نیاز به تغییر اندازه آرایه‌ها، انجام دهد و به سرعت به جواب‌های دلخواه خود برسد.
بنابراین، با توجه به قابلیت‌های متنوع و سرعت بالای کتابخانه پانداز، این کتابخانه به یکی از بهترین ابزارهای تحلیل داده تبدیل شده است.

۲.۴.۱ کتابخانه Numpy

Numpy یکی از کتابخانه‌های پرکاربرد در زمینه برنامه‌نویسی علمی و محاسباتی است که در زبان پایتون قابل استفاده است. این کتابخانه به کاربر اجازه می‌دهد تا با داده‌های بزرگ عددی و ماتریسی به صورت سریع و کارآمد کار کند. Numpy به کاربر امکان می‌دهد تا برای کار با داده‌های بزرگ، از آرایه‌های Numpy استفاده کند که قابلیت‌های بسیاری را برای کاربر فراهم می‌کند.
یکی از دلایل برتری و سرعت بالای این کتابخانه، استفاده از آرایه‌های چند بعدی است. با استفاده از آرایه‌های Numpy، می‌توان به راحتی با داده‌های چند بعدی کار کرد و تمام عملیات‌های ریاضی

[6] labeled arrays

مانند جمع، ضرب، ترانهاده را روی آن‌ها اعمال کرد.
همچنین، Numpy دارای توابع بهینه‌سازی شده برای انجام عملیات ریاضی است. این کتابخانه از الگوریتم‌های بهینه برای انجام عملیات ریاضی استفاده می‌کند و به کاربر امکان می‌دهد تا با سرعت بالایی عملیات‌های پیچیده ریاضی را انجام دهد.

۳.۴.۱ کتابخانه Hazm

هضم [۷]، یک کتابخانه پردازش زبان فارسی برای زبان پایتون می‌باشد. با استفاده از این کتابخانه، قابلیت نرمال‌سازی متن، استخراج جملات و واژه‌ها، پیدا کردن ریشه کلمات، تحلیل صرفی و نحوی جملات و شناسایی وابستگی‌های دستوری متن فارسی فراهم می‌شود.

۴.۴.۱ کتابخانه FastText

fastText یک کتابخانه متن‌کاوی قدرتمند است که برای حل مسائلی مانند شناسایی زبان، تشخیص احساسات، برچسب‌گذاری موضوع و ترجمه ماشینی طراحی شده است. این کتابخانه توسط گروه تحقیقاتی Facebook AI Research (FAIR) ارائه شده است و در زبان برنامه‌نویسی پایتون قابل استفاده می‌باشد.
موتور fastText از الگوریتم‌های شبکه‌های عصبی بازگشتی (RNN) [۸] و شبکه‌های عصبی پیچشی (CNN) [۹] برای آموزش مدل‌های پردازش زبان طبیعی استفاده می‌کند. با استفاده از این الگوریتم‌ها، می‌توان مدل‌هایی برای پردازش متن طراحی کرد که قابلیت پیش‌بینی برچسب‌های متنی را دارند. با استفاده از fastText می‌توان یک مدل آموزش داد که با داده‌های ورودی، پیش‌بینی احساسی یک متن را انجام دهد.
از ویژگی‌های fastText می‌توان به امکان ساخت بردارهای تعبیه شده [۱۰] برای کلمات اشاره کرد. این بردارها، به صورت بردارهای عددی هستند که هر کلمه را با یک بردار در فضای nـبعدی

[7] www.roshan-ai.ir/hazm/
[8] Recurrent Neural Network
[9] Convolutional Neural Network
[10] embeddings

نشان می‌دهند. این بردارهای تعبیه شده، در بسیاری از مسائل پردازش زبان مانند مدل‌سازی زبان طبیعی و تشخیص احساسات، بسیار مفید هستند.
از دیگر ویژگی‌های fastText می‌توان به سرعت بالا و قابلیت استفاده در داده‌های بزرگ اشاره کرد. همچنین، این کتابخانه به راحتی در سیستم‌های توزیع‌شده قابل استفاده است و اجازه می‌دهد تا مدل‌های پردازش زبان طبیعی را بر روی چندین سرور آموزش داده و برای پردازش داده‌های بزرگ استفاده کنیم.
در کل، fastText یک کتابخانه پرقدرت و کارآمد برای پردازش زبان است که برای حل مسائل متنی در صنایع مختلف مورد استفاده قرار می‌گیرد.

۵.۴.۱ کتابخانه Selenium

کتابخانه Selenium یکی از محبوب‌ترین کتابخانه‌ها در زمینه خودکارسازی وب است. این کتابخانه به کاربر اجازه می‌دهد تا برنامه‌هایی را برای کنترل مرورگر وب و انجام عملیات مختلف بر روی صفحات وب ایجاد کند.
استفاده از کتابخانه Selenium برای مطالعه داده‌های متنی بسیار مفید است، زیرا به کاربر اجازه می‌دهد تا اطلاعات مورد نظر خود را از صفحات وب به صورت خودکاراستخراج کنید.
همچنین، کتابخانه Selenium به کاربر امکان می‌دهد تا به راحتی با مرورگرهای مختلفی مانند Google Chrome کار کند و از ویژگی‌های مختلف آن‌ها بهره ببرد. این کتابخانه به کاربر امکان این را می‌دهد تا عملیات مختلفی مانند کلیک کردن بر روی دکمه‌ها، پر کردن فرم‌ها، وارد کردن متن و استخراج اطلاعات را از صفحات وب انجام دهد.
استفاده از کتابخانه Selenium بسیار ساده است و از طریق زبان‌های برنامه‌نویسی متنوعی مانند پایتون [11]، جاوا [12] و سی شارپ [13] امکان دسترسی به این کتابخانه وجود دارد. با استفاده از این کتابخانه، می‌توان به سرعت و به صورت خودکار داده‌های متنی را از صفحات وب استخراج کرد و از آن‌ها برای تحلیل و یا استفاده در سایر کاربردها استفاده نمود.

[11] Python
[12] Java
[13] C#

۵.۱ معرفی آزمون‌های آماری

۱.۵.۱ آزمون Shapiro

آزمون Shapiro یکی از آزمون‌های آماری است که برای ارزیابی توزیع نمونه‌ای یک متغیر پیوسته به کار می‌رود. این آزمون برای تعیین اینکه آیا داده‌های نمونه از یک توزیع نرمال پیروی می‌کنند یا نه، استفاده می‌شود.

در آزمون Shapiro، با استفاده از مقدار p-value، تصمیم گیری می‌شود که آیا داده‌های نمونه از یک توزیع نرمال پیروی می‌کنند یا خیر. اگر مقدار p-value کمتر از سطح معناداری معین شده باشد، ممکن است نمونه داده‌ها از یک توزیع نرمال پیروی نکند. استفاده از آزمون Shapiro در تحقیقات آماری بسیار رایج است و به کاربر اجازه می‌دهد تا با استفاده از آن، توزیع نمونه‌های خود را بررسی کند و در صورت لزوم، از آزمون‌های آماری دیگر برای تحلیل داده‌های خود استفاده کند.

۲.۵.۱ آزمون Levene

آزمون Levene یکی از آزمون‌های آماری است که برای تعیین اینکه آیا واریانس دو یا چند گروه داده‌ای متفاوت است یا نه، به کار می‌رود. در واقع، این آزمون برای بررسی همبستگی بین دو یا چند متغیر استفاده می‌شود.

طریقه انجام آزمون Levene به این صورت است که در ابتدا، داده‌های نمونه از گروه‌های مختلف جمع‌آوری می‌شوند. در مرحله بعد، با استفاده از آزمون Levene و با استفاده از مقدار p-value، تصمیم‌گیری می‌شود که آیا واریانس گروه‌های داده‌ای مختلف از یکدیگر متفاوت هستند یا خیر. اگر مقدار p-value کمتر از سطح معناداری معین شده باشد، ممکن است واریانس گروه‌ها از یکدیگر متفاوت باشد.

استفاده از آزمون Levene در تحقیقات آماری بسیار رایج است و به کاربر امکان این را می‌دهد تا با استفاده از آن، واریانس داده‌های گروه‌های مختلف خود را بررسی کند و در صورت لزوم، از آزمون‌های آماری دیگر برای تحلیل داده‌های خود استفاده کند.

۳.۵.۱ آزمون Anova

آزمون Anova یکی از آزمون‌های آماری است که برای تعیین اینکه آیا میانگین گروه‌های داده‌ای مختلف از یکدیگر متفاوت هستند یا خیر، به کار می‌رود. این آزمون برای تحلیل داده‌های چند گروهی استفاده می‌شود.

طریقه انجام آزمون Anova به این صورت است که در ابتدا، داده‌های نمونه از گروه‌های مختلف جمع‌آوری می‌شوند. در مرحله بعد، با استفاده از این آزمون و با استفاده از مقدار p-value، تصمیم گیری می‌شود که آیا میانگین گروه‌های داده‌ای مختلف از یکدیگر متفاوت هستند یا خیر. اگر مقدار p-value کمتر از سطح معناداری معین شده باشد، ممکن است میانگین گروه‌ها از یکدیگر متفاوت باشد.

استفاده از این آزمون در تحقیقات آماری بسیار رایج است و به کاربر اجازه می‌دهد تا با استفاده از آن، تفاوت میانگین داده‌های گروه‌های مختلف خود را بررسی کند و در صورت لزوم، از آزمون‌های آماری دیگر برای تحلیل داده‌های خود استفاده کند.

۶.۱ معرفی معیارهای ارزیابی عملکرد مدل‌های پیش‌بینی

معیارهای ارزیابی عملکرد مدل‌های پیش‌بینی، ابزارهایی هستند که برای ارزیابی دقت و کارایی یک مدل پیش‌بینی مورد استفاده قرار می‌گیرند. این معیارها با استفاده از مقایسه پیش‌بینی‌های مدل با واقعیت، به ما اطلاعاتی درباره عملکرد مدل در پیش‌بینی داده‌ها ارائه می‌کنند.

۱.۶.۱ ماتریس درهم ریختگی

ماتریس درهم‌ریختی [۱۴] یکی از ابزارهای مهم برای ارزیابی عملکرد مدل‌های پیش‌بینی است. این ماتریس، در واقع یک جدول دوبعدی است که نشان می‌دهد که چه میزان از پیش‌بینی‌های مدل صحیح و چه میزان از آن‌ها نادرست بوده‌اند. ماتریس درهم‌ریختی شامل چهار خانه اصلی است که در ادامه معرفی می‌شوند.

[14] Confusion Matrix

- True Positive (TP): تعداد پیش‌بینی‌های درست مثبت
- False Positive (FP): تعداد پیش‌بینی‌های نادرست مثبت
- True Negative (TN): تعداد پیش‌بینی‌های درست منفی
- False Negative (FN): تعداد پیش‌بینی‌های نادرست منفی

ماتریس درهم ریختگی		برچسب پیش‌بینی شده	
		مثبت	منفی
برچسب واقعی	مثبت	TP	FN
	منفی	FP	TN

شکل ۱.۱: ماتریس درهم ریختگی

۲.۶.۱ معیار دقت

دقت [۱۵] معیاری است که نشان می‌دهد چه میزان از پیش‌بینی‌های مدل درست بوده است. برای محاسبه دقت، نسبت تعداد پیش‌بینی‌های درست بر تعداد کل پیش‌بینی‌ها (درست و نادرست) محاسبه می‌شود. دقت مقیاسی از صحت کلی مدل است. مزیت اصلی دقت این است که از تمام پیش‌بینی‌های مدل استفاده می‌کند و نسبتاً ساده است.

$$\text{دقت} = \frac{TP + TN}{TP + FN + FP + TN}$$

[15] Accuracy

۳.۶.۱ معیار صحت

معیار صحت [۱۶] نشان می‌دهد که چه میزان از پیش‌بینی‌های مثبت مدل درست بوده‌اند. برای محاسبه معیار صحت، نسبت تعداد پیش‌بینی‌های مثبت درست بر تعداد کل پیش‌بینی‌های مثبت مدل در نظر گرفته می‌شود. این معیار در مواردی که تشخیص دادن پیش‌بینی‌های نادرست مثبت ممکن است منجر به خطرات جدی شود، مانند تشخیص افراد مبتلا به بیماری‌های خطرناک، بسیار مهم است.

$$\text{صحت} = \frac{TP}{TP + FP}$$

۴.۶.۱ معیار پوشش

معیار پوشش [۱۷] نشان می‌دهد که مدل چه میزان از وضعیت‌های واقعی را به درستی تشخیص داده است. برای محاسبه معیار پوشش، نسبت تعداد پیش‌بینی‌های درست منفی بر تعداد کل پیش‌بینی‌های منفی مدل در نظر گرفته می‌شود. این معیار در مواردی که در نتیجه عدم تشخیص وضعیت‌های واقعی، خطرات احتمالی برای افراد وجود دارد، بسیار مهم است.

$$\text{پوشش} = \frac{TP}{TP + FN}$$

۵.۶.۱ معیار F1 Score

این معیار نشان می‌دهد که چه میزان از پیش‌بینی‌های درست مدل، واقعیت را پوشش داده است. معیار F1-score ترکیبی از معیار صحت و معیار پوشش است که با استفاده از رابطه زیر محاسبه می‌شود.

$$\text{F1 Score} = \frac{\text{۲} \times \text{معیار پوشش} \times \text{معیار دقت}}{\text{معیار پوشش} + \text{معیار دقت}}$$

[16] Precision

[17] Recall

فصل ۲

مقدمه‌ای بر تحلیل احساسات

تحلیل احساسات به معنی تشخیص و بررسی احساسات، عواطف و نگرانی‌های انسان‌ها در قبال یک موضوع خاص است. در بازارهای مالی نیز، احساسات و رفتارهای سرمایه‌گذاران می‌توانند بر روی قیمت‌ها و حجم معاملات تأثیر بگذارند. به همین دلیل، تحلیل احساسات می‌تواند برای پیش‌بینی رفتار بازارهای مالی و تصمیم‌گیری‌های سرمایه‌گذاران بسیار مفید باشد [۵].

۱.۲ کاربردهای تحلیل احساسات در بازارهای مالی

تحلیل احساسات در بازارهای مالی دارای کاربردهای متعددی است. در زیر به برخی از کاربردها و تحقیقات انجام شده در هر یک از این زمینه‌ها اشاره می‌شود:

پیش‌بینی رفتار بازار

احساسات سرمایه‌گذاران نسبت به یک موضوع خاص، به ما کمک می‌کند تا بتوانیم رفتار بازار را پیش‌بینی کنیم. همانطور که اشاره شد، پیش‌بینی رفتار بازار از اهمیت زیادی برخوردار است؛ این موضوع در سال‌های اخیر مورد توجه بسیاری از پژوهشگران در حوزه علوم مالی قرار گرفته است [۶].

کشف تأثیر خبرها

تاثیر اخبار و اتفاقات مختلف بر رفتار بازار مالی از جمله موضوعات مهمی است که توسط سرمایه‌گذاران و تحلیل‌گران بازار بررسی می‌شود. اخبار به‌طور مستقیم و غیرمستقیم می‌توانند بر رفتار بازار مالی تاثیرگذار باشند، بنابراین با بررسی احساسات سرمایه‌گذاران نسبت به یک خبر خاص، می‌توانیم به تأثیر آن در بازار پی ببریم [۷].

بررسی تأثیر تبلیغات

تحلیل احساسات می‌تواند به بررسی تأثیر تبلیغات و دیگر اقدامات تبلیغاتی در بازار کمک کند. با بررسی احساسات سرمایه‌گذاران نسبت به یک تبلیغ خاص، می‌توانیم به تأثیر آن در بازار پی ببریم.

پیش‌بینی تغییرات قیمت

با بررسی احساسات سرمایه‌گذاران می‌توانیم به پیش‌بینی تغییرات قیمت بازار پی ببریم [۸].

۲.۲ روش‌های تحلیل احساسات

در حوزه تحلیل احساسات، روش‌های مختلفی برای تشخیص احساسات در متون وجود دارند. این روش‌ها می‌توانند به دو دسته روش‌های مبتنی بر یادگیری ماشین و روش‌های مبتنی بر لغت‌نامه تقسیم شوند.

۱.۲.۲ روش مبتنی بر یادگیری ماشین

در روش‌های مبتنی بر یادگیری ماشین، از الگوریتم‌های یادگیری ماشین برای تشخیص احساسات در متن استفاده می‌شود. این الگوریتم‌ها، با استفاده از داده‌های آموزشی، به یادگیری نحوه تشخیص احساسات در متن می‌پردازند. در این روش، دو نوع داده ورودی به الگوریتم می‌دهیم: متن و برچسب احساسات مثبت، منفی و خنثی. سپس با استفاده از الگوریتم‌های یادگیری ماشین، احساسات مثبت،

منفی و خنثی متن بررسی می‌شوند. به‌عنوان مثال، در این روش، از الگوریتم‌های SVM [1]، Naive Bayes و شبکه‌های عصبی برای تحلیل احساسات استفاده می‌شود.
به عنوان مثال، Rani and Kumar برای تحلیل احساسات داده‌های بررسی فیلم‌های هندی، از شبکه‌های عصبی عمیق و الگوریتم‌های CNN [2] استفاده کرده‌اند. نتایج مدل پیشنهادی آن‌ها با دقت ٪۹۵، عملکرد بهتری نسبت به الگوریتم‌های یادگیری ماشین کلاسیک دارد [۹].

۲.۲.۲ روش مبتنی بر لغت‌نامه

روش‌های مبتنی بر لغت‌نامه، به دلیل سادگی و سرعت بالایی که دارند، از محبوبیت بالایی برخوردارند. در این روش، از یک لغت‌نامه احساساتی برای تعیین احساسات در متن استفاده می‌شود. این لغت‌نامه، شامل لغاتی است که با احساسات مثبت، منفی و خنثی مرتبط هستند. در این روش، با تعیین احساسات مثبت، منفی و خنثی لغات موجود در متن، احساس کلی متن تعیین می‌شود. به‌عنوان مثال، در این روش، از لغت‌نامه‌های احساساتی مختلف مانند SentiWordNet، AFINN و Loughran-McDonald برای تحلیل احساسات استفاده می‌شود.
بیشتر افرادی که در حوزه تحلیل احساسات فعالیت می‌کنند، به روش‌های مبتنی بر یادگیری ماشین تمرکز کرده‌اند و تنها تعداد کمی از آن‌ها به روش‌های مبتنی بر واژگان توجه دارند. تا کنون تعداد زیادی لغت‌نامه برای تحلیل احساسات متون به زبان انگلیسی تولید شده است، اما تولید لغت‌نامه برای تحلیل احساسات به زبان فارسی به اندازه کافی مورد توجه قرار نگرفته است.
بررسی‌های بسیاری در حوزه تحلیل احساسات به زبان فارسی نشان می‌دهد که استفاده از ترجمه مستقیم لغات و عبارات فارسی به انگلیسی و سپس استفاده از مدل‌های زبان انگلیسی، بسیار محدود و ناموثر می‌باشد. این محدودیت برای تحلیل‌گران احساسات مشکل ساز است؛ به دلیل وجود اختلافات زبانی و فرهنگی بین زبان انگلیسی و فارسی، ترجمه مستقیم بعضی از لغات و عبارات باعث از دست رفتن معنای اصلی و فرضیاتی که در تحلیل احساسات مطرح می‌شوند، می‌شود. در زبان فارسی برخی از واژگان و عباراتی وجود دارند که به مفهوم خاصی می‌انجامند که با ترجمه مستقیم به زبان انگلیسی، این مفاهیم به درستی منتقل نمی‌شوند. برای رفع این مشکل، می‌یاست

[1] Support Vector Machines
[2] Convolutional Neural Network

لغت‌نامه‌هایی خاص زبان فارسی داشته باشیم.

از جمله روش‌های مبتنی بر لغت‌نامه برای تحلیل احساسات به زبان فارسی، می‌توان به لغت‌نامه‌هایی برای تشخیص قطبیت کلمات و عبارات به زبان فارسی تولید شده است، اشاره کرد؛ مانند لغت‌نامه LexiPers که شامل بیش از ۶۰۰۰ کلمه است [۱۰]. همچنین، در تحلیل احساسات به زبان فارسی، لغت‌نامه CNRC نیز ارائه شده که با استفاده از لغت‌نامه NRC که برای زبان انگلیسی می‌باشد، ایجاد شده است. این لغت‌نامه برای تحلیل احساسات در سطح جمله به کار می‌رود [۱۱].

برخلاف روش مبتنی بر لغت‌نامه که به صورت دستی و بر اساس ارتباط لغات با احساسات مشخص شده، روش مبتنی بر یادگیری ماشین به صورت خودکار و با استفاده از داده‌های آموزشی به یادگیری الگوهای تشخیص احساسات در متن می‌پردازد. بنابراین، روش مبتنی بر یادگیری ماشین معمولاً دقت بیشتری در تحلیل احساسات دارد. اما برای استفاده از این روش، نیاز به داده‌های آموزشی و برچسب‌گذاری شده وجود دارد که این موضوع می‌تواند مشکلاتی در جمع‌آوری داده‌ها و تحلیل احساسات برای زبان‌های کمتر استفاده شده از جمله زبان فارسی ایجاد کند [۱۲].

در مقابل، روش مبتنی بر لغت‌نامه به دلیل عدم نیاز به داده‌های آموزشی و سرعت بالایی که دارد، مناسب برای پردازش متون بزرگ است. اما این روش نسبت به روش‌های مبتنی بر یادگیری ماشین، دقت پایین‌تری در تحلیل احساسات دارد و ممکن است برای متن‌های با ساختار پیچیده و بدون قید و شرط کارایی کمتری داشته باشد [۱۳].

در فصل بعد، نحوه تحلیل و تشخیص احساسات کاربران شبکه‌های اجتماعی توضیح داده می‌شود.

فصل ۳

تحلیل احساسات

به منظور تحلیل احساسات کاربران، ابتدا نیاز است که داده‌های متنی جمع‌آوری شوند. سپس نرمال‌سازی‌ها و اصلاح دادگان صورت می‌گیرد و در نهایت دادگان به جهت شناسایی به مدل تحلیل احساسات ورودی داده می‌شوند تا برچسب گذاری شوند.

۱.۳ جمع‌آوری داده

در حوزه تحلیل احساسات متنی، جمع‌آوری داده‌های مناسب از اهمیت فراوانی برخوردار است. در این پژوهش، داده‌های مرتبط با نظرات کاربران فضای مجازی درباره سهام‌های بازار بورس اوراق بهادار در نظر گرفته شده است. این داده‌ها با استفاده از کتابخانه Selenium از سایت سهام‌یاب [1] جمع‌آوری شده‌اند.

سهام‌یاب، شبکه هوشمند سرمایه‌گذاری و اطلاع‌رسانی بورس ایران است که تلاش نموده تا در کنار اطلاع‌رسانی آخرین وضعیت بازار بورس، آخرین معاملات شرکت‌ها، اخبار و اطلاعیه‌های آن‌ها، محیطی یکپارچه برای به اشتراک گذاری ایده‌ها و نظرات کاربران گرامی را ایجاد کند و از این طریق امکانی فراهم کند تا اطلاعات تالار گفتگو، سایت‌های اطلاع‌رسانی و خبرگزاری‌ها پیرامون بازار سرمایه ایران در دسترس کاربران قرار دهد.

[1] www.sahamyab.com

در حوزه شبکه‌های اجتماعی و جمع‌آوری داده‌های مربوط به بازار بورس، سهام‌یاب به عنوان یکی از شبکه‌های مهم و موثر شناخته شده است. یکی از دلایل برتری سهام‌یاب نسبت به سایر شبکه‌های اجتماعی، عدم امکان حذف پیام‌های کاربران توسط آن‌ها می‌باشد. همچنین، ویرایش پیام‌ها نیز تنها تا نیم ساعت بعد از ارسال نظر امکان پذیر است. این ویژگی‌های سهام‌یاب، باعث می‌شود که افراد با دقت بیشتری نظر خود را ثبت کنند. این موضوع در تحلیل احساسات متنی بسیار مهم است زیرا نظر و احساسات کاربران به طور دقیق ثبت شده است و تحت عوامل مختلف، کاربر امکان تغییر آن را به نفع خودش نداشته است.

پس از استخراج داده‌های متنی از سایت سهام‌یاب، این داده‌ها برای تحلیل احساسات متنی مورد استفاده قرار می‌گیرند. داده‌ها شامل نظرات، پیش‌بینی‌ها و احساسات کاربران در مورد سهام‌های بازار بورس اوراق بهادار می‌باشند. بر اساس این داده‌ها و با استفاده از روش‌های پیش‌پردازش مناسب، می‌توان به تحلیل احساسات مثبت، منفی و یا خنثی متون مرتبط با سهام‌های مختلف پرداخت.

۲.۳ نرمال‌سازی داده

برای نرمال‌سازی داده‌ها، از ابزاری به نام هضم [۲] استفاده شده است. مشکلات نوشتاری در زبان فارسی بسیار رایج‌تر از زبان انگلیسی هستند. در ادامه به برخی از این مشکلات می‌پردازیم.

۱.۲.۳ رفع مشکل نیم‌فاصله

برخی از کلمات فارسی از چند قسمت تشکیل شده‌اند. این کلمات معمولاً با یک فاصله نیم‌فاصله از هم جدا می‌شوند. اما معمولاً افراد در توییت‌هایی که در شبکه‌های اجتماعی منتشر می‌کنند، این نکته را رعایت نمی‌کنند. به عنوان مثال، کلمه «آن‌ها» به طور نادرست به شکل «آن ها» نوشته می‌شود.

[2] www.roshan-ai.ir/hazm/

۲.۲.۳ حذف حروف تکراری

در برخی موارد، به منظور تاکید بیشتر بر احساسات یا بیان شدت بیشتری در بیان، افراد از تکرار حروف در کلمات استفاده می‌کنند. به عنوان مثال، به جای نوشتن «نه» ، از «نههههههه» استفاده می‌شود. با این حال، تکرار حروف در کلمات، تاثیری در نتیجه تحلیل احساسات ندارد و فقط برای تاکید بیشتر در بیان استفاده می‌شود. به همین دلیل، با استفاده از ابزار هضم، می‌توان حروف اضافه کلمات را حذف کرده و به شکل استاندارد آن‌ها را نمایش داد. این کار می‌تواند به بهبود دقت در تحلیل احساسات کمک کند.

۳.۲.۳ یکسان‌سازی حروف مشابه

یکی دیگر از عواملی که ممکن است منجر به مشکل در کدگذاری کاراکترها شود، تفاوت در کدگذاری حروف در استاندارد یونیکد است. به عنوان مثال، حرف «ی» را می‌توان به صورت‌های مختلف «ی» ، «ي» و «ئ» نوشت که دارای کدگذاری متفاوتی در استاندارد یونیکد هستند.
به منظور حل این مشکل، با استفاده از ابزار هضم، تمامی کلمات به صورت نرمال سازی می‌شوند و حروف استفاده شده در کلمات، یکسان می‌شوند. به این ترتیب، تفاوت در کدگذاری کاراکترها برطرف شده و مشکلات مربوط به نمایش کاراکترها در سیستم‌ها که از استاندارد یونیکد پشتیبانی نمی‌کنند، رفع می‌شود.

۳.۳ اصلاح داده

در تحلیل احساسات دادگان متنی به زبان فارسی، یکی از مشکلات اساسی و مهم، وجود کلمات محاوره‌ای و غیررسمی در متون است. تفاوت در گفتار و نوشتار بعضی از کلمات، باعث می‌شود برخی افراد نظرات خود را به صورت محاوره‌ای ثبت کنند. این مشکل باعث می‌شود که تحلیل احساسات دادگان به صورت درست و دقیق صورت نگیرد و نتایج نادرستی به دست آید.
برای حل این مشکل، از ابزار هضم استفاده شده است. ابزار هضم توانایی تبدیل کلمات محاوره‌ای به کلمات رسمی و استاندارد را داراست. این ابزار با تحلیل کلمات و متون، قادر به شناسایی

کلمات محاوره‌ای و غیررسمی است و با استفاده از لغت‌نامه‌های معتبر، آن‌ها را به کلمات رسمی و استاندارد تبدیل می‌کند. به این ترتیب، امکان تحلیل دقیق‌تر احساسات و نظرات ارائه شده در دادگان متنی به زبان فارسی، فراهم می‌شود. در جدول ۱.۳ برخی از این اصلاحات آمده است.

جدول ۱.۳: اصلاح دادگان محاوره‌ای

واژگان محاوره‌ای قبل از اصلاح	واژگان بعد از اصلاح
منفیه	منفی است
قیمتا	قیمت‌ها
میریزه	می‌ریزد
مثبت نمیشن	مثبت نمی‌شوند
مجمع نمیره	مجمع نمی‌رود

۴.۳ استفاده از مدل‌های پیش‌آماده به منظور تحلیل احساسات

پس از انجام فرآیند نرمال سازی و اصلاح دادگان، داده‌های مورد نیاز برای تحلیل احساسات و برچسب‌زنی آماده می‌شوند.
استفاده از مدل‌های پیش‌آموزش‌دیده در تحلیل احساسات، توانایی بالایی در شناسایی نظرات و ارزیابی احساسات و موضوعات مختلف دارد. این مدل‌ها با استفاده از شبکه‌های عصبی عمیق، به صورت خودکار و با دقت بالا، اطلاعات مفیدی از داده‌ها استخراج می‌کنند و برای تحلیل و برچسب‌زنی از آن‌ها استفاده می‌شود. در این مدل پیش‌آماده از کتابخانه fastText استفاده می‌شود.

۱.۴.۳ شیوه برچسب گذاری

برای پیاده‌سازی یک مدل پردازش زبانی با استفاده از fastText ، ابتدا باید داده‌های آموزشی را جمع آوری کنیم. همان‌طور که بیان شد، داده‌های آموزشی از سایت سهام یاب جمع‌آوری شده است

و برای هر پیام، یک برچسب احساسی (مثبت، منفی یا خنثی) تعیین شده است.
بعد از جمع‌آوری، نرمال‌سازی و اصلاح داده‌ها، باید آن‌ها را به نحوی که fastText توانایی خواندن آن‌ها را داشته باشد، آماده کنیم. به این منظور، می‌توانیم از کتابخانه pandas استفاده کنیم.
بعد از آماده‌سازی داده‌ها، مدل را با داده‌های آموزشی برچسب‌دار آموزش می‌دهیم. برای این کار، از تابع train supervised کتابخانه fastText استفاده می‌کنیم. این تابع، مدل را با داده‌های آموزشی برچسب‌دار آموزش می‌دهد و مدل آموزش‌دیده را برمی‌گرداند.
بعد از آموزش مدل، می‌توانیم داده‌های آزمایشی را به مدل بدهیم تا تحلیل احساسات پیام‌ها را استخراج کنیم. برای این کار، از تابع predict کتابخانه fastText استفاده می‌کنیم. این تابع، برای هر پیام، یک بردار فازی شامل احساسات مثبت و منفی مشخص می‌کند و پس از آن یک برچسب احساسی (مثبت، منفی یا خنثی) را بر اساس وزن احساسات تعیین شده، مشخص می‌کند.
در جدول ۲.۳ مثال‌هایی از دادگان برچسب گذاری شده توسط مدل آمده است.

جدول ۲.۳: مثال از دادگان برچسب گذاری شده

نظر کاربران	برچسب احساسی
#گدنا فردا اگر در صف فروش قفل نشود لااقل ۵ درصد قیمت پایانی‌اش افت می‌کند	منفی
#خزامیا فردا سهم باز می‌شود. اگر پول سنگین دارید مثبت‌های بالا اردر بذارید	مثبت
#شپنا سلام بر همگی. دوستان کسی از مجمع خبری ندارد؟ خبر خوش، خبر بد؟	خنثی
#دانا رشد خوبی خواهد کرد نگران نباشید	مثبت
اسمش مجمع بود ولی در حقیقت کابوس بود	منفی

۵.۳ بررسی دقت مدل تحلیل احساسات

جهت ارزیابی صحت و دقت مدل تحلیل احساسات، از مجموعه داده‌های برچسب‌دار استفاده نموده‌ایم. نتایج نشان می‌دهد که این مدل عملکرد بسیار قابل قبولی دارد و با دقت ۸۶٫۹ درصد، قادر به تحلیل و برچسب‌گذاری احساسات مثبت، منفی و خنثی در داده‌ها است.

به منظور ارزیابی تعمیم‌پذیری مدل، از داده‌های تست بدون برچسب نیز استفاده شده است. نتایج این دادگان به صورت دستی بررسی شدند. اکثر مغایرت‌های موجود در برچسب‌گذاری، به علت پیجیدگی زبان فارسی است. مدل تحلیل احساسات در زبان فارسی به دلیل پیچیدگی ساختاری و خصوصیات خاص آن، به چالش‌های فراوانی برخورد می‌کند. زبان فارسی به دلیل وجود جملات اصطلاحی و استفاده از اصطلاحات خاص، می‌تواند منجر به تفسیرهای نادرستی در تحلیل احساسات شود. به همین دلیل، در برخی موارد، برچسب‌گذاری‌ها ممکن است به طور اشتباهی انجام شود چرا که فهم دقیق معنایی آن‌ها برای مدل‌های تحلیل احساسات تقریبا ناممکن است. همچنین، استفاده از عبارات ابهام‌زا در برخی موارد نیز می‌تواند به برچسب‌گذاری نادرست منجر شود. در جدول ۳.۳ مثال‌هایی از این اشتباهات آمده است.

جدول ۳.۳: نمونه‌هایی از اشتباهات مدل تحلیل احساسات

نظر کاربران	برچسب	توضیحات
#خودرو داریم به نفت می‌رسیم امیدوارم شاخص از ۱۸۰۰ برگرده	مثبت	منظور کاربر از "به نفت رسیدن"، شدت زیاد کاهش قیمت است. اما مدل تحلیل احساسات، "به نفت رسیدن" را از لحاظ احساسی مثبت برداشت می‌کند، چرا که نفت را به عنوان ماده‌ای ارزشمند شناسایی می‌کند.
#دی چقدر خوبه. دو روز رنج مثبت زده، الان صف فروش شده	مثبت	منظور کاربر وجود نوسانات زیاد در سهم دی و وجود ناپایداری در قیمت است که احساسی منفی به دنبال دارد. اما به دلیل اصطلاحاتی همچون "دی چقدر خوبه" و "دو روز رنج مثبت زده"، مدل این پیام را با احساسات مثبت شناسایی کرده است.
#خساپا سلام. سه روز پیش تحلیل شاخص کل رو منتشر کردم و پیشنهاد می‌کنم امروز در زمان ریزش سهم از صف فروش خریدانجام بدید	منفی	احساس کاربر به وضوح مثبت است، اما به دلیل استفاده از "ریزش سهم" و "صف فروش"، برچسب‌گذاری منفی شده است.

با توجه به موارد فوق، مشخص است که تحلیل احساسات در زبان فارسی دارای چالش‌های قابل توجهی است و ممکن است باعث ایجاد اشتباه شود. یکی از روش‌های مفید برای رفع این مشکل، استفاده از داده‌های آموزشی بسیار متنوع است.

در فصل بعد، با توجه به این نتایج بدست آمده، تاثیر احساسات بر رفتار بازارهای مالی را بررسی خواهد شد. به این منظور، با استفاده از روش‌های آماری و مدل‌سازی، تلاش می‌شود تا تاثیر احساسات مثبت و منفی بر بازارهای مالی بررسی شود.

فصل ۴

تاثیر احساسات کاربران بر روی بازارهای مالی

در فصل قبل، با نحوه عملکرد مدل تحلیل احساسات در این پژوهش آشنا شدیم که ابزار مفیدی برای بررسی رفتار کاربران و نظراتشان درباره سهم‌های مختلف بازار بورس اوراق بهادار می‌باشد. برای بررسی تاثیر احساسات کاربران بر روی بازار بورس اوراق بهادار، ابتدا از ۱۱ صنعت و صندوق فعال در بازار بورس ایران، یک سهم انتخاب شده است؛ در این راستا، تلاش شد سهام شرکت‌هایی انتخاب شوند که در یک روز معاملاتی، ارزش معاملاتشان زیاد باشند و علاقه‌مندان زیادی در بازار سرمایه داشته باشند. در ادامه لیست این سهم‌ها آمده است.

۱. سهم فولاد، از صنعت فلزات اساسی
۲. سهم وبملت، از صنعت بانکداری
۳. سهم خودرو، از صنعت خودرو سازی
۴. سهم اخابر، از صنعت مخابرات
۵. سهم زملارد، از صنعت زراعت

۶. سهم حکشتی، از صنعت حمل و نقل

۷. سهم گدنا، از صنعت هتل و گردشگری

۸. سهم دانا، از صنعت بیمه

۹. سهم خزامیا، از صنعت قطعات خودرو

۱۰. سهم اطلس، از صندوق‌های ETF

۱۱. سهم طلا، از صندوق طلا

۱.۴ اجتماع احساسات روزانه

جهت بررسی تأثیر احساسات بر روی بازار بورس اوراق بهادار، به دو مرحله اصلی پرداخته‌ایم. در مرحله اول، نظرات کاربران در خصوص سهم‌های معرفی شده تحلیل احساسات شده‌اند. در این مرحله، با استفاده از مدل تحلیل احساسات که در فصل قبل به آن پرداختیم، نظرات کاربران را مورد بررسی و تحلیل قرار دادیم و برچسب گذاری احساسی شدند. پس از آن، با استفاده از روشی مبتنی بر روش‌های فازی، برآیند احساسات روزانه هر سهم را محاسبه کردیم.

از فصل قبل می‌دانیم در زمان تحلیل احساسات، قبل از برچسب‌زنی احساسی، نسبتی برای هر نظر کاربران محاسبه می‌شود که نشان دهنده میزان مثبت یا منفی بودن احساس متن است. به عنوان مثال، نظر کاربری که به صورت ۳۰ درصد منفی و ۷۰ درصد مثبت است، در نهایت با برچسب مثبت شناخته می‌شود. حال در این مرحله از روش برآیندگیری احساسات روزانه، بین میزان قطبیت مثبت و منفی نظرات یک روز برای هر سهم میانگین گرفته می‌شود تا برآیند نهایی برای هر سهم به دست آید. بنابراین احساسات روزانه هر سهم در بازار سرمایه به دست آمده‌اند. حال نکته قابل توجه این است که احساسات روزهای گذشته در احساسات افراد تأثیرگذار هستند و نمی‌توان آن‌ها را نادیده گرفت. به همین دلیل،از روش میانگین وزنی با ضرایب متفاوت برای احساسات گذشته و حال استفاده شده است.

به طور خاص در این روش محاسباتی، با توجه به اینکه احساسات دو روز قبل نیز تأثیرگذار می‌باشند،

از ضریب‌های مختلفی برای محاسبه احساس نهایی استفاده می‌شود. به این صورت که بین احساسات همان روز با ضریب ۰/۷ ، احساسات روز قبل با ضریب ۰/۲ و احساسات دو روز قبل با ضریب ۰/۱ میانگین وزنی گرفته می‌شود. با این روش، شدت احساسات مثبت و منفی روزانه هر سهم با دقت بیشتری محاسبه می‌شود. در ادامه، مثال ساده‌ای از این فرآیند آورده شده است.

فرض کنید جدول ۱.۴ ، تمام نظرات کاربران در خصوص سهم حکشتی از صنعت کشتیرانی در روز ۹ مرداد ۱۴۰۲ باشد.

جدول ۱.۴: نمونه نظرات کاربران در مورد سهم حکشتی

نظرات کاربران	میزان قطبیّت مثبت	میزان قطبیّت منفی	برچسب احساسی
#حکشتی روند صعودی تثبیت شد. دوستان نگران نباشید ۱۵ درصد ضررتون هم نهایتا سه روزه جبران می‌شود. فقط نگهداری سهام پیشنهاد می‌شود.	۶۳ درصد	۳۷ درصد	مثبت
#حکشتی تبریک به تمام سهامداران صبور حکشتی الحمدالله گزارشی که میگفتیم نشست روی کدال.علی الحساب ۴ تا از شرکتهای اصلیش حدود ۳۶۰ تومن سود ساختن که حدود ۳۲۰ تومنش رو تقسیم کردن توی مجمع.	۹۲ درصد	۸ درصد	مثبت
#حکشتی زیاد جدی نگیر. با نیم درصد مثبت خوردن چطور امکان داره بگید روند صعودی سهم تثبیت شده!	۴۷ درصد	۵۳ درصد	منفی
#حکشتی خب خداروشکر پول خرید کشنده ها هم در اومد. امروز فردا هم یه نوسان مثبت خوب داریم.	۹۱ درصد	۹ درصد	مثبت
#حکشتی حکشتی بیش از ۱۰۰۰ تومان اصلاح قیمتی داشته. صعودش بزودی آغاز می‌شود.	۸۳ درصد	۱۷ درصد	مثبت

حال با داشتن تحلیل احساسات نظرات کاربران، به محاسبه احساس این روز می‌پردازیم. ۴ نظر با برچسب احساسی مثبت و ۱ نظر با برچسب احساسی منفی مشاهده می‌شود. طبق توضیحات داده شده، بین میزان قطبیت مثبت و منفی نظرات میانگین می‌گیریم.
میانگین قطبیّت مثبت نظرات:

$$\frac{۶۳ + ۹۲ + ۴۷ + ۹۱ + ۸۳}{۵} = ۷۵/۲$$

میانگین قطبیّت منفی نظرات:

$$\frac{۳۷ + ۸ + ۵۳ + ۹ + ۱۷}{۵} = ۲۴/۸$$

بنابراین با قطبیّت ۷۵/۲ درصد، احساسات کاربران در مورد سهم حکشتی مثبت بوده است. حال فرض کنید به همین روش، برای روز قبل و دو روز قبل نیز برایند احساسی محاسبه شده است و نتایج در جدول ۲.۴ آمده است:

جدول ۲.۴: برآیند احساسی نظرات کاربران برای سه روز متوالی سهم حکشتی

تاریخ	برآیند احساسی
۷ مرداد ۱۴۰۲	۲۳/۳ درصد مثبت، ۷۶/۷ درصد منفی
۸ مرداد ۱۴۰۲	۳۷/۹ درصد مثبت، ۶۲/۱ درصد منفی
۹ مرداد ۱۴۰۲	۷۵/۲ درصد مثبت، ۲۴/۸درصد منفی

در نهایت، با توجه به الگوریتم بیان شده، شدت احساسی را برای روز ۹ مرداد ۱۴۰۲، برای سهم حکشتی به صورت زیر محاسبه می‌کنیم.
میانگین وزنی قطبیّت مثبت:

$$(۰/۱ \times ۲۳/۳) + (۰/۲ \times ۳۷/۹) + (۰/۷ \times ۷۵/۲) = ۶۲/۵۵$$

میانگین وزنی قطبیّت منفی:

$$(۰/۱ \times ۷۶/۷) + (۰/۲ \times ۶۲/۱) + (۰/۷ \times ۲۴/۸) = ۳۷/۴۵$$

بنابراین به عنوان یک نمونه، شدت احساسی روز ۹ مرداد ۱۴۰۲ برای سهم حکشتی محاسبه کردیم. با انجام مراحل مذکور، توانستیم شدت روزانه احساسات هر سهم را به دست آوریم. در این روش، تأثیر احساسات روزهای گذشته نیز به کار گرفته شد که باعث بهبود و دقت پیشگو می‌شود. استفاده از شدت احساسات روزانه به عنوان پیشگو در مدل‌های آماری، امکان پیش‌بینی تغییرات آینده قیمت سهام را با دقت بالاتری فراهم می‌کند. بنابراین، با توجه به دقت و صحت محاسبات و نتایج بدست آمده، شدت احساسات روزانه برای همه روزها و تمام سهم‌های معرفی شده، قابل استفاده در مدل‌های آماری مختلف می‌باشد.

۲.۴ بررسی وجود ارتباط احساسات کاربران و بازده سهم

ابتدا به بررسی ارتباط بین شدت احساسات مثبت و منفی روزانه کاربران هر سهم با بازده روزانه همان سهم پرداخته شده است.

به این منظور، ابتدا نرمال بودن داده‌ها با استفاده از آزمون شپیرو [۱] بررسی شد. در مرحله بعد با استفاده از آزمون همگنی واریانس [۲] ، برابری واریانس‌ها سنجیده شد. در نهایت از آزمون آنووا [۳] استفاده شده است. در این آزمون، متغیر وابسته بازده سهم و متغیر مستقل شدت احساسات مثبت و منفی محاسبه شده است. برای انجام آزمون آنالیز واریانس، ابتدا داده‌ها بر اساس مقادیر شدت احساسات مثبت و منفی تفکیک شده‌اند. تفکیک داده‌ها بر اساس میزان شدت مثبت و منفی احساسات صورت گرفته است. به طوری که اگر شدت احساسات مثبت بیشتر از ۵۰ درصد باشد، دادگان که همان بازده روزانه سهم‌های مختلف هستند، در دسته احساسات مثبت قرار گرفته و اگر شدت احساسات مثبت کمتر از ۵۰ درصد باشد، دادگان در دسته احساسات منفی قرار گرفته و در

[1] Shapiro test
[2] Levene test
[3] Anova test

غیر این صورت، در دسته احساسات خنثی قرار می‌گیرند.
بعد از تفکیک داده‌ها، برای هر دسته از داده‌ها، میانگین بازده سهم‌ها و واریانس بازده سهم‌ها محاسبه شده است. سپس با استفاده از آماره آزمون وجود ارتباط معنادار بین شدت احساسات مثبت و منفی و بازده سهم بررسی شده است.
نتایج آماری نشان داد که ارتباط معناداری بین شدت احساسات مثبت و بازده سهم و همچنین بین شدت احساسات منفی و بازده سهم‌ها وجود دارد و میانگین بازده‌ها در گروه‌های مختلف احساسی با یکدیگر اختلاف معناداری دارند. این نتایج نشان می دهد که شدت احساسات مثبت و منفی کاربران بر روی عملکرد بازیگران بازار بورس اوراق بهادار تاثیرگذار است و می‌تواند به عنوان یکی از عوامل موثر بر بازده سهم در بورس در نظر گرفته شود.

۳.۴ یافتن ارتباط میان احساسات کاربران و بازده سهم

برای یافتن ارتباط بین شدت احساسات و جهت بازده سهام، از مدل لجستیک [۴] استفاده شده است. منظور از جهت بازده سهام، مثبت یا منفی بودن بازده در یک روز معاملاتی است. در ادامه به شرح مراحل می‌پردازیم.

۱.۳.۴ معرفی مدل لجستیک

مدل لجستیک یکی از مدل‌های رگرسیون است که برای مدل‌سازی رابطه بین یک یا چند متغیر پیش‌گو و یک متغیر پاسخ باینری استفاده می‌شود. در این مدل، متغیر پاسخ دو مقدار ۰ و ۱ را به خود می‌گیرد که به ترتیب نشان‌دهنده عدم وجود یا وجود یک ویژگی یا رویداد هستند. به عنوان مثال، پاسخ ۰ به معنی بازده منفی سهم و پاسخ ۱ به معنی بازده مثبت سهم در نظر گرفته می‌شود.
در مدل لجستیک، احتمال وقوع یک رویداد برای مقادیر مختلف متغیرهای پیش‌بین استخراج می‌شود. این احتمال به عنوان متغیر پاسخ مدل در نظر گرفته می‌شود. برای محاسبه این احتمال، از تابع لجستیک استفاده می‌شود که مقدار خروجی آن بین ۰ و ۱ قرار دارد. در واقع، تابع لجستیک

[4] Logistic Model

نوعی تابع سیگموئید [5] است که برای تبدیل مقادیر پیش‌بین به مقادیر پاسخ استفاده می‌شود. این مدل در مدل‌سازی احتمالاتی و تحلیل داده‌ها به عنوان یک تابع فعال‌سازی برای تبدیل مقادیر ورودی به مقادیر خروجی بین ۰ و ۱ استفاده می‌شود.
مدل لجستیک به دلیل سادگی و قابلیت استفاده در مسائل مختلف، از محبوبیت بالایی برخوردار است و در زمینه‌های مختلفی از جمله علوم اجتماعی، علوم سیاسی، علوم پزشکی، علوم رفتاری و علوم اقتصادی استفاده می‌شود.

۲.۳.۴ شرح دادگان

برای هر سهم، ساختار داده شامل سه ستون در نظر گرفته شده است. ستون اول شامل تاریخ، ستون دوم شامل شدت احساسات روزانه مثبت سهم برحسب درصد و ستون سوم جهت بازده متناظر روزانه سهم است. واضح است که شدت احساسات روزانه منفی، از مقدار شدت احساسات روزانه مثبت قابل محاسبه است، بنابراین نیازی نیست آن را به صورت مجزا در دادگان وارد کنیم.

۳.۳.۴ آموزش مدل

دادگان به دو دسته آموزشی و آزمایشی تقسیم شدند. از مدل لجستیک برای آموزش دادگان شد. شدت احساسات مثبت به عنوان متغیر مستقل (ورودی) و جهت بازده سهام را به عنوان متغیر وابسته (خروجی) در نظر گرفته شده است. با استفاده از داده‌های آموزشی، مدل آموزش داده شد.

۴.۳.۴ نتایج مدل

جهت ارزیابی عملکرد مدل، مدل بر روی دادگان آزمایش اجرا شده و از معیارهای متفاوتی همچون دقت [6]، معیار صحت [7]، معیار پوشش [8]، و معیار ۱ F [9] استفاده شده است. در جدول ۳.۴ نتایج

[5] Sigmoid Function
[6] Accuracy
[7] Precision
[8] Recall
[9] F1 score

آمده است.

جدول ۳.۴: نتایج مدل لجستیک بر روی دادگان آزمایشی بر حسب درصد

نام سهم	معیار دقت	معیار صحت	معیار پوشش	معیار F1 Score
فولاد	۷۴٫۲	۶۶٫۸	۷۱٫۵	۶۹٫۰
وبملت	۵۸٫۹	۴۱٫۲	۶۳٫۷	۵۰٫۳
خودرو	۶۹٫۸	۶۲٫۱	۶۶٫۴	۶۴٫۱
اخابر	۴۳٫۷	۳۸٫۹	۴۷٫۲	۴۲٫۶
زملارد	۵۱٫۲	۴۰٫۱	۵۵٫۷	۴۶٫۸
حکشتی	۵۴٫۶	۴۳٫۸	۵۸٫۲	۵۰٫۳
گدنا	۶۲٫۵	۵۰٫۱	۵۹٫۸	۵۴٫۵
دانا	۷۶٫۱	۶۸٫۵	۷۳٫۹	۷۰٫۹
خزامیا	۷۳٫۷	۷۰٫۵	۷۱٫۶	۷۱٫۰
اطلس	۷۸٫۳	۷۲٫۹	۷۶٫۵	۷۴٫۶
طلا	۴۱٫۵	۳۴٫۷	۴۲٫۶	۳۸٫۲

با توجه به نتایج مدل، می‌توان نتیجه گرفت که تعداد نظرات کاربران در یک دوره زمانی، بر دقت مدل لجستیک در پیش‌بینی جهت بازار بر اساس شدت احساسات، تأثیرگذار است. به عبارت دیگر، با افزایش تعداد نظرات کاربران، دقت مدل در پیش‌بینی جهت بازار بالاتر می‌رود.

در این زمینه، مشاهده می‌شود که دقت مدل برای سهامی که بخش زیادی از بازار سرمایه را تشکیل می‌دهند و از مخاطبان بیشتری برخوردارند، بهتر است. به علاوه، کیفیت نظرات کاربران نیز بر دقت مدل تأثیرگذار است. به عنوان مثال، سهم اطلس نسبت به سایر سهم‌ها، از تعداد نظرات کمتری در

یک بازه زمانی مشخص برخوردار است. با این حال، به دلیل این که این سهم، متاثر از شاخص کل بازار سرمایه است، نظرات کاربران درباره آن، اکثراً به تحلیل تکنیکال و به دور از حاشیه پرداخته می‌شود. این موضوع سبب بالا رفتن دقت مدل در پیش‌بینی جهت بازار شده است.
به علاوه، شدت احساسات کاربران سهم طلا کم‌ترین دقت را در پیش‌بینی جهت بازده طلا به خود اختصاص داده است. دلیل این موضوع آن است که جهت بازدهی سهم طلا، به عوامل زیادی از جمله قیمت جهانی طلا، قیمت ارز و روابط سیاسی داخلی و خارجی وابسته است. به همین دلیل، تأثیر احساسات بر جهت بازدهی سهم طلا، کاهش می‌یابد و بنابراین دقت مدل در پیش‌بینی جهت بازده سهم طلا از مابقی سهم‌ها کمتر است.

۴.۴ بهبود عملکرد مدل

به منظور بهبود عملکرد، شاخص کل بازار سرمایه به عنوان یک متغیر تأثیرگذار در مدل لجستیک مورد استفاده قرار گرفته شده است. شاخص کل بازار سرمایه، نمایانگر عملکرد کلی بازار سهام است. شاخص کل به عنوان یک عامل مهم اقتصادی بررسی شده است چرا که می‌تواند جهت بازده سهام را به شدت تحت تأثیر خود قرار دهد.
به این منظور، یک شرط آستانه در خصوص شاخص کل تعریف شده است تا در صورتی که شاخص کل بازار از این آستانه بزرگ‌تر یا کوچک‌تر باشد، روز مربوطه را در مدل لجستیک لحاظ نکنیم. ریشه‌های زیادی برای افزایش یا کاهش شاخص کل بازار سرمایه در یک روز خاص وجود دارد. عواملی از جمله اعلام خبری خاص در خصوص شرایط اقتصادی، اجتماعی و فرهنگی می‌تواند بر روی شاخص کل تاثیرگذار باشد که ارتباط مستقیمی با سهم مشخصی ندارد و مشمول تمامی سهم‌های بازار می‌شود. بنابراین با در نظر گرفتن آستانه به عنوان یک معیار برای کنار گذاشتن روزهایی استفاده می‌شود که بازده سهام به طور قابل توجهی تحت تأثیر شاخص کل بازار قرار می‌گیرند.
با امتحان کردن چند مقدار آستانه متفاوت برای شاخص کل، نتایج نشان می‌دهد آستانه ۱/۵ درصد برای بهبود این مدل بسیار مناسب است. به طوری‌که اگر شاخص کل بیشتر از ۱/۵ درصد و کمتر از

۱/۵ – درصد باشد، آن روزها در مدل مورد بررسی قرار داده نمی‌شوند. با اعمال این شرط، شاخص کل بازار سرمایه به عنوان یک متغیر ورودی مهم در مدل لجستیک مورد استفاده قرار می‌گیرد.
بنا به نتایج، اضافه کردن حد آستانه شاخص کل بازار سرمایه به مدل لجستیک، بهبود خوبی در پیش‌بینی جهت بازده سهام و افزایش دقت مدل ارائه می‌دهد. جدول ۴.۴ نتایج مدل لجستیک را پس از حذف روزهای خارج از آستانه نشان می‌دهد که در مقایسه با نتایج جدول ۳.۴ از عملکرد بهتری برخوردار است.

جدول ۴.۴: نتایج مدل لجستیک برحسب درصد با حذف روزهای خارج از آستانه

نام سهم	معیار دقت	معیار صحت	معیار پوشش	معیار F1 Score
فولاد	۷۵/۵	۶۸/۱	۷۳/۲	۷۰/۵
وبملت	۶۰/۷	۴۳/۰	۶۶/۲	۵۲/۴
خودرو	۷۱/۱	۶۳/۲	۶۸/۰	۶۵/۵
اخابر	۴۶/۵	۴۱/۴	۵۰/۴	۴۵/۷
زملارد	۵۳/۴	۴۲/۰	۵۸/۳	۴۹/۲
حکشتی	۵۶/۸	۴۵/۵	۶۰/۶	۵۲/۴
گدنا	۶۴/۳	۵۲/۰	۶۲/۱	۵۶/۸
دانا	۷۷/۹	۷۰/۱	۷۵/۵	۷۲/۴
خزامیا	۷۳/۲	۷۰/۹	۷۱/۱	۷۰/۹
اطلس	۷۸/۸	۷۲/۹	۷۷/۰	۷۴/۸
طلا	۴۱/۸	۳۵/۴	۴۳/۱	۳۸/۸

تغییر اعمال شده در سبب بهبود در عملکرد مدل برای اکثر سهم‌ها شده است. اما باید توجه داشت که تاثیر زیادی در بهبود عملکرد سهم‌های اطلس و طلا دیده نمی‌شود.

سهم اطلس رفتاری مشابه با شاخص کل دارد و به همین دلیل احساسات کاربران درباره این سهم همواره مشابه با احساساتی است که درباره شاخص کل دارند. با حذف روزهای خارج از آستانه، دقت مدل در پیش‌بینی سهم اطلس تغییر چندانی نمی‌کند.

همچنین، همان‌طور که اشاره شد، سهم طلا تحت تأثیر عوامل متعددی قرار دارد و به طور کلی تحت تأثیر شاخص کل قرار نمی‌گیرد. بنابراین، تغییرات اعمال شده در دادگان در ارتباط با شاخص کل، تأثیر چندانی در بهبود عملکرد پیش‌بینی سهم طلا نیز ندارد.

فصل ۵

نتیجه گیری

در این پژوهش، تأثیر احساسات کاربران شبکه‌های اجتماعی بر بازار بورس اوراق بهادار بررسی شد. با استفاده از مدل‌های تحلیل احساسات مبتنی بر یادگیری ماشین، داده‌های متنی جمع‌آوری شده و برچسب گذاری احساسی شدند. سپس با استفاده از مدل لجستیک و تجزیه و تحلیل داده‌های جمع‌آوری شده و بررسی روابط شدت احساسات کاربران با جهت بازده سهام، توانستیم با دقت قابل توجهی، جهت بازار را برای سهام‌های مختلف براساس احساسات روزانه کاربران آن‌ها پیش‌بینی کنیم.

علاوه بر این، با کنار گذاشتن روزهایی که شاخص کل بازار سرمایه به طور قابل توجهی افزایش یا کاهش یافته بود، توانستیم دقت مدل را بهبود بخشیم. این نکته نشان می‌دهد که ردیابی و تحلیل دقیق تغییرات شاخص کل، نقش مهمی در بهبود پیش‌بینی‌های مدل دارد و روزهایی که شاخص کل تغییر زیادی داشته باشد، بازده سهم‌ها بیش از احساسات، از روند بازار الگو می‌گیرند.

نتایج نشان داد که شدت احساسات، تأثیر مثبت و معناداری بر عملکرد بازار مالی دارد. این امر به ما امکان می‌دهد تصمیم‌گیری‌های هوشمندانه‌تر در خصوص سرمایه‌گذاری در بازارهای مالی انجام دهیم.

بنابراین استفاده از تحلیل احساسات به عنوان یک ابزار کاربردی در تحلیل و پیش‌بینی بازارهای مالی پیشنهاد می‌شود. این روش می‌تواند به سرمایه‌گذاران و متخصصان مالی کمک کند تا تصمیمات بهتری در خصوص سرمایه‌گذاری و مدیریت ریسک خود اتخاذ کنند. لازم به ذکر است که تحلیل

احساسات تنها یکی از عوامل مؤثر بر بازارهای مالی است و باید با سایر شاخص‌ها و عوامل اقتصادی و مالی مورد بررسی قرار گیرد.

فصل ۶

واژه‌نامه فارسی به انگلیسی

ابزار مالی ..... Financial instrument
ارزش ..... Value
آستانه ..... threshold
اقتصاد مالی‌رفتاری ..... Behavioral finance
اوراق بهادار ..... Security
بازار بورس ..... Exchange (organized market)
بازار مالی ..... Financial market
بازده ..... Return
بردار ..... Vector
بردار تعبیه شده ..... Embedding vector
پوشش ..... Recall
پویش وب ..... Web crawling
تحلیل احساسات ..... Sentiment analysis
تحلیل بنیادی ..... Fundamental analysis
تحلیل تکنیکال ..... Technical analysis
داده آموزشی ..... Train data

داده آزمایشی ..... Test data
داده برچسب‌دار ..... Labeled data
دقت ..... Accuracy
رفتار عقلایی ..... Rational behavior
روش‌های آماری ..... Statistical methods
ریسک ..... Risk
ساختار داده ..... Dataframe
سرمایه گذاری ..... Investment
سهام ..... Stock
شاخص اقتصادی ..... Economics index
شاخص کل ..... TEPIX
شبکه اجتماعی ..... Social network
شبکه عصبی بازگشتی ..... Recurrent Neural Network
شبکه عصبی پیچیده ..... Convolutional Neural Network
شرکت سهامی عام ..... Public company
صحت ..... Precision
علوم اعصاب ..... Neuroscience
علوم رفتاری ..... Behavioural sciences
ماتریس درهم ریختگی ..... Confusion Matrix
ماشین بردار پشتیبان ..... Support Vector Machine
مدل‌های پیش‌بینی ..... Prediction models
مدیریت ریسک ..... Risk management
نوسانات بازار ..... Market fluctuations
نیم‌فاصله ..... Half space
پردازش زبان طبیعی ..... Natural language processing
یادگیری عمیق ..... Deep learning

فصل ۷

واژه‌نامه انگلیسی به فارسی

دقت ..... Accuracy
اقتصاد مالی‌رفتاری ..... Behavioral finance
علوم رفتاری ..... Behavioural sciences
ماتریس درهم ریختگی ..... Confusion Matrix
شبکه عصبی پیچیده ..... Convolutional Neural Network
ساختار داده ..... Dataframe
یادگیری عمیق ..... Deep learning
شاخص اقتصادی ..... Economics index
بردار تعبیه شده ..... Embedding vector
بازار بورس ..... Exchange (organized market)
ابزار مالی ..... Financial instrument
بازار مالی ..... Financial market
تحلیل بنیادی ..... Fundamental analysis
نیم‌فاصله ..... Half space
سرمایه گذاری ..... Investment
داده برچسب‌دار ..... Labeled data

یادگیری ماشین ..... Machine learning
نوسانات بازار ..... Market fluctuations
پردازش زبان طبیعی ..... Natural language processing
علوم اعصاب ..... Neuroscience
صحت ..... Precision
مدل‌های پیش‌بینی ..... Prediction models
شرکت سهامی عام ..... Public company
رفتار عقلایی ..... Rational behavior
پوشش ..... Recall
شبکه عصبی بازگشتی ..... Recurrent Neural Network
بازده ..... Return
مدیریت ریسک ..... Risk management
ریسک ..... Risk
اوراق بهادار ..... Security
تحلیل احساسات ..... Sentiment analysis
شبکه اجتماعی ..... Social network
روش‌های آماری ..... Statistical methods
سهام ..... Stock
ماشین بردار پشتیبان ..... Support Vector Machine
شاخص کل ..... TEPIX
تحلیل تکنیکال ..... Technical analysis
داده آزمایشی ..... Test data
آستانه ..... Threshold
داده آموزشی ..... Train data
یونی‌کُد ..... Unicode
ارزش ..... Value

بردار ..... Vector

پویش وب ..... Web crawling

کتاب‌نامه

[1] Reed, M. (2016). A study of social network effects on the stock market. *Journal of Behavioral Finance*, 17(4), 342-351.

[2] Rao, T., & Srivastava, S. (2012). Analyzing stock market movements using twitter sentiment analysis.

[3] Cambria, E., Olsher, D., & Rajagopal, D. (2014). Senticnet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. In *Twenty-eighth AAAI conference on artificial intelligence* (pp. 8928).

[4] Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, 83, 187-205.

[5] Carosia, A. E. O., Coelho, G. P., & Silva, A. E. A. (2020). Analyzing the Brazilian financial market through Portuguese sentiment analysis in social media. *Applied Artificial Intelligence*, 34(1), 1-19.

[6] Guo, X., & Li, J. (2019). A novel twitter sentiment analysis model with baseline correlation for financial market prediction with im-

proved efficiency. In *Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)* (pp. 472-477). IEEE.

[7] Chiong, R., Fan, Z., Hu, Z., Adam, M. T., Lutz, B., & Neumann, D. (2018). A sentiment analysis-based machine learning approach for financial market prediction via news disclosures. In *Proceedings of the genetic and evolutionary computation conference companion* (pp. 278-279).

[8] Wei, P., & Wang, N. (2016). Wikipedia and stock return: Wikipedia usage pattern helps to predict the individual stock movement. In *Proceedings of the 25th International Conference Companion on World Wide Web* (pp. 591-594). International World Wide Web Conferences Steering Committee.

[9] Rani, S., & Kumar, P. (2019). Deep learning based sentiment analysis using convolution neural network. *Arabian Journal for Science and Engineering*, 44(4), 3305-3314. doi: 10.1007/s13369-018-3500-z

[10] Sabeti, B., Hosseini, P., Ghassem-Sani, G., & Mirroshandel, S. A. (2019). Lexipers: An ontology based sentiment lexicon for Persian. *arXiv preprint arXiv:1911.05263*.

[11] Basiri, M. E., & Kabiri, A. (2017). Translation is not enough: Comparing lexicon-based methods for sentiment analysis in Persian. In *2017 International Symposium on Computer Science and Software Engineering Conference (CSSE)* (pp. 36-41). IEEE. doi: 10.1109/CSICSSE.2017.8320114

[12] Dashtipour, K., Gogate, M., Adeel, A., Ieracitano, C., Larijani, H., & Hussain, A. (2018). Exploiting deep learning for Persian sentiment analysis. In *International conference on brain inspired cognitive systems* (pp. 597-604). Springer. doi: 10.1007/978-3-030-00563-4_58

[13] Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*.

[14] Sul, H., Dennis, A. R., & Yuan, L. I. (2014). Trading on twitter: The financial information content of emotion in social media. In *2014 47th Hawaii International Conference on System Sciences* (pp. 806-815). IEEE.

[15] Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., & Trajanov, D. (2020). Evaluation of sentiment analysis in finance: From lexicons to transformers. *IEEE Access*, 8, 131662-131682.

[16] Gupta, R., & Chen, M. (2020, August). Sentiment analysis for stock price prediction. In *2020 IEEE conference on multimedia information processing and retrieval (MIPR)* (pp. 213-218). IEEE.

[17] Guijarro, F., Moya-Clemente, I., & Saleemi, J. (2019). Liquidity risk and investors' mood: Linking the financial market liquidity to sentiment analysis through Twitter in the S&P500 index. *Sustainability*, 11(24), 7048.

[18] Daniel, M., Neves, R. F., & Horta, N. (2017). Company event popularity for financial markets using Twitter and sentiment analysis. *Expert Systems with Applications*, 71, 111-124.

[19] Basiri, M. E., Naghsh-Nilchi, A. R., & Ghassem-Aghaee, N. (2014). A framework for sentiment analysis in Persian. *Open Transactions on Information Processing*, 1(3), 1-14. doi: 10.15764/OTIP.2014.03001

[20] Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). doi: 10.1145/1014052.1014073

[21] Deng, L., & Wiebe, J. (2015). Mpqa 3.0: An entity/event-level sentiment corpus. In *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human Language Technologies* (pp. 1323-1328). doi: 10.3115/v1/N15-1146

[22] Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1), 163-173. doi: 10.1002/asi.21662

[23] Piñeiro-Chousa, J., Vizcaíno-González, M., & Pérez-Pico, A. M. (2017). Influence of social media over the stock market. *Psychology & Marketing*, 34(1), 101-108.

[24] Narayanan, V., Arora, I., & Bhatia, A. (2013). Fast and accurate sentiment classification using an enhanced naive bayes model. In

International Conference on Intelligent Data Engineering and Automated Learning (pp. 194-201). Springer. doi: 10.1007/978-3-642-41278-3_24

[25] Wüthrich, B., Cho, V., Leung, S., Permunetilleke, D., Sankaran, K., & Zhang, J. (1998). Daily stock market forecast from textual Web data. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics* (pp. 2720-2725).

Abstract

The main objective of this research is to investigate the impact of social media users' emotions on financial markets. The results show that social media users' emotions can be considered as one of the influential factors in the movement of financial markets. To this end, we first introduce the basic concepts of the financial and natural language processing domains. Then, an introduction to the various applications and methods of sentiment analysis is presented. In the main sections, we will introduce and fully describe the sentiment analysis method used in this research. Finally, the influence of social media users' emotions on the stock market is examined. In this section, using the introduced sentiment analysis method, social media users' emotions about the stock market are analyzed. Then, the relationship between social media users' emotions and the movement of the stock market is examined using statistical methods.

Keywords: Financial markets, Natural language processing, Sentiment analysis, Emotional tagging, Stock market

College of Science

School of Mathematics, Statistics, and Computer Science

Psychology of Financial Markets based on Social Networks Data

Zahra Khatibi

Supervisor: Samaneh Eftekhari Mahabadi

A thesis submitted in partial fulfillment of the requirements for the degree of B.Sc. in Computer Science

Summer 2023