



پردیس علوم  
دانشکده ریاضی، آمار و علوم کامپیوتر

# بررسی میزان دستکاری در بازار سرمایه با استفاده از روش‌های یادگیری ماشین

نگارنده

مهیار محمدی متین

استاد راهنما: سمانه افتخاری مهابادی

پایان‌نامه برای دریافت درجه کارشناسی

در رشته علوم کامپیوتر

بهار ۱۴۰۲

## چکیده

بازار سرمایه جزء ضروری اقتصادهای مدرن محسوب می‌شود. رونق بازارهای مالی و معاملات پربسامد باعث افزایش و پیچیده‌تر شدن تخلف‌های معاملاتی شده است [۱]. دستکاری بازارهای مالی می‌تواند منجر به انحراف قیمت‌ها یا افزایش نوسانات شده و بر سرمایه‌گذاران و کل بازار تأثیر بگذارد [۲]. علی‌رغم تلاش‌های نهادهای نظارتی برای کنترل دستکاری بازار مالی، هنوز هم به اشکال مختلف مانند معاملات چرخشی، معاملات همروند و روش‌های دیگر، فعالیت متقلبانه در بازار وجود دارد. از طرفی یادگیری ماشین به ابزاری محبوب در بازارهای مالی برای شناسایی الگوها و ناهنجاری‌ها تبدیل شده است [۳] [۴].

در این پژوهش ابتدا ۳ الگوریتم ریسک محور برای تشخیص معاملات مشکوک به تقلب معرفی می‌کنیم که در بازه شش ماه و روی بیش از ششصد نماد فرابورسی پیاده‌سازی می‌شود. سپس اثرات تقلب‌ها بر کلیت بازار و سهم را می‌سنجیم. برای این سنجش از رگرسیون و مدل‌های طولی استفاده خواهیم کرد.

کلیدواژه: کشف تقلب در بازار سرمایه، بازار سهام، نوسانات بازده و قیمت، تحلیل آماری، یادگیری ماشین

## تقدیم به

این اثر از صمیم قلب تقدیم می‌شود، به دنبال کنندگان بی‌امان حقیقت در داخل و خارج از مرزهای ایران که به طور خستگی‌ناپذیری برای کشف حقایق و روشن کردن زوایای تاریک جامعه تلاش می‌کنند. این افراد چه روزنامه نگار باشند، چه بازرسان، یا شهروندان نگران، باید در شبکه پیچیده‌ای از منافع سیاسی و مالی حرکت کنند تا حقیقت را افشا کنند. تعهد همیشگی شما برای روشن کردن حقیقت، حتی در مبهم‌ترین شرایط، به عنوان چراغ امید برای کسانی است که به دنبال عدالت هستند.

تلاش شما برای سلامت ملت و رفاه مردم آن حیاتی است و من مفتخرم که این پژوهش را به شما تقدیم کنم. باشد که این کار به تلاش شما برای تقویت جامعه‌ای شفاف‌تر، عادلانه‌تر و پاسخگوتر کمک کند. انصاف، شجاعت و جستجوی حقیقت، نیروی محرکه و تلاش جمعی ما برای فردایی بهتر است.

## سپاسگزاری

صمیمانه از سرکار خانم دکتر افتخاری، عضو هیئت علمی گروه آمار دانشکده ریاضی دانشگاه تهران و جناب آقای دکتر مهدی نوری، عضو هیئت علمی دانشکده اقتصاد دانشگاه تهران به خاطر کمک‌ها و حمایت‌های بی‌دریغشان در طول تدوین این پروژه تشکر می‌کنم. صبر و حوصله ایشان در پاسخگویی به سوالات اینجانب و ایده‌های خلاقانه ایشان نقش بسزایی در پیشرفت این پژوهش ایفا کرده است.

در پایان از خانواده‌ام به خاطر صبوری و درک آن‌ها در مواقعی که مجبور شدم تحصیل را بر گذراندن وقت با آن‌ها اولویت دهم تشکر می‌کنم. پشتیبانی همیشگی شما به من اعتماد به نفس داده است تا رویاهایم را دنبال کنم و بر هر مانعی که بر سر راهم قرار می‌گیرد غلبه کنم.

## پیشگفتار

اعتماد عمومی به سلامت بازارها، قدرت نقدشوندگی و کارایی آن‌ها را افزایش می‌دهد و این امر در بازار سرمایه و به ویژه بورس اوراق بهادار، اهمیت ویژه‌ای دارد. میزان اعتماد عمومی به بازار به اندازه ریسک موجود در این بازارها، به عنوان یکی از عوامل اصلی تأثیرگذار بر تصمیم‌گیری سرمایه‌گذاران محسوب می‌گردد.

دستکاری قیمت‌های بازار به کشف قیمت واقعی زیان می‌رساند و قیمت منصفانه اوراق بهادار را نادرست جلوه می‌دهد. قیمت‌های تحریف شده، سرمایه‌گذار را مجبور به انتقال به بازارهای کارآمدتر می‌کند. این امر به شرکت‌ها انگیزه اضافی می‌دهد تا اوراق بهادارشان را در بازارهای دیگری که مقررات بهتر و نظارت مؤثرتری دارند منتشر کنند. از این رو، روش‌های شناخت و شیوه‌های برطرف کردن دستکاری قیمت سهام منافع بزرگی را برای پژوهشگران، تنظیم‌کننده‌ها و مبادله‌گران بازار حاصل می‌کند.

وقایع سال ۱۳۹۹ در بازار سرمایه نگاه همگانی مردم و مسئولین را به بازار سرمایه و نتیجتاً بازارهای مالی تغییر داد. در کنار کشف پتانسیل‌ها، خاطرات تلخ و تجاربی باقی ماند که بخشی از آن تجارب معطوف به ناظرین بازار بورس بوده است. این فشار بر دپارتمان‌های نظارتی بازار سرمایه، لزوم توسعه در برخی ابعاد را به همگان نشان داد. برای مثال توسعه نرم افزارهای نظارتی یکی از حوزه‌هایی است که امروز به خوبی دغدغه تمامی مسئولان بازار سرمایه است. امروزه با حرکت به سمت فناوری‌های نظارتی مبتنی بر ریسک به جای نظارت مبتنی بر هشدارها و قواعد ساده، می‌توان نگاه وسیع‌تر و جامع‌تری به حوزه نظارت داشت. توسعه ابزارها و تکنیک‌های به‌روز و پرسرعت می‌تواند نظارت بر بازارها را تسهیل کرده و از سوءاستفاده از بازارها ممانعت کند. یافته‌های این مطالعه برای

سیاست‌گذاران، تنظیم‌کننده‌ها و سرمایه‌گذاران در تصمیم‌گیری آگاهانه در مورد فعالیت‌های بازار مالی مفید خواهد بود.

تشخیص دستکاری قیمت‌نهایی توسط میرفیض فلاح و همکاران در پاییز ۹۴ [۵] و محمدحسین پوست فروش و همکاران در بهار ۱۳۹۵ [۶] و کارول کامرتون<sup>۱</sup> در سال ۲۰۰۹ [۷] و سنجیده شد. تاثیرات دستکاری قیمت بر کیفیت بازار نیز توسط هوانگ<sup>۲</sup> و همکاران در سال ۲۰۰۵ اندازه‌گیری شد [۸].

در این پژوهش، ابتدا به بررسی تعاریف و انواع معاملات غیرقانونی در بازار سرمایه می‌پردازیم تا خواننده با اصطلاحات حوزه مدیریت مالی و اصطلاحات حوزه نظارتی و برخی اصطلاحات حوزه علوم کامپیوتر و آمار آشنا شود.

سپس سه نوع الگوریتم خلاقانه پیشنهاد و پیاده‌سازی می‌کنیم تا بتوانیم هر کدام از سه تقلب متداول در بازارهای سرمایه را تشخیص دهیم. در این الگوریتم‌ها از برخی کاربردهای جبرخطی و محاسبات ماتریسی، نظریه گراف، خوشه‌بندی و... استفاده شده است.

یک روش مبتنی بر گراف برای تشخیص معاملات چرخشی در حوزه مالیات ستانی توسط متئوس<sup>۳</sup> و همکاران در سال ۲۰۱۷ مطرح شد [۹]. همچنین پالشیکار<sup>۴</sup> در سال ۲۰۰۸ مقاله محبوبی در حوزه تشخیص مجموعه تبانی با استفاده از خوشه‌بندی گراف مطرح کرده است [۱۰].

در حوزه معاملات نهانی نیز پژوهشی در بازار سهام کره جنوبی توسط پارک<sup>۵</sup> در سال ۲۰۱۰ منتشر شده است [۱۱]. پیش‌تر نیز دونوهو<sup>۶</sup> و همکاران در سال ۲۰۰۴ معاملات نهان را در بازار اختیار معامله بررسی کرده اند [۱۲].

اخیرا نیز موریک<sup>۷</sup> و همکاران مقاله ای در حوزه تشخیص معاملات همروند به منظور دستکاری قیمت در سال ۲۰۲۱ منتشر کرده اند [۱۳]. همچنین پژوهشی از ویکتور<sup>۸</sup> در سال ۲۰۱۹ بر روی

---

<sup>1</sup>Carole Comerton

<sup>2</sup>Yu Chuan Huang

<sup>3</sup>Jithin Mathews

<sup>4</sup>Girish Keshav Palshikar

<sup>5</sup>Young S. Park

<sup>6</sup>Steve Donoho

<sup>7</sup>Goran Murić

<sup>8</sup>Friedhelm Victor

دستکاری قیمت بازار رمزارزها انجام شده است [۱۴].

در ادامه، از نتایج بدست آمده از میزان دستکاری و تقلب در نمادهای بورسی مختلف در بازه زمانی مشخص، به دنبال رابطه‌ای معنادار بین تعداد تکرارهای معاملات متقلبانه و دیگر متغیرهای بازار می‌گردیم. این متغیرها می‌توانند میزان نوسانات بازده سهام یا شاخص کل، ناهنجاری در بتای سهام، میزان حباب نماد بورسی، و... باشند. برای این سنجش نیز از مدل‌های آماری، آزمون‌های فرض و تحلیل سری زمانی استفاده می‌کنیم.

# فهرست مطالب

۱	مفاهیم مقدماتی	۱
۱	۱.۱ تعاریف حوزه بازارهای مالی و اقتصاد	۱
۱	۱.۱.۱ بورس و اوراق بهادار	۱
۴	۲.۱.۱ سازوکار معاملات	۴
۷	۳.۱.۱ فاصله قیمتی و نقدشوندگی	۷
۸	۴.۱.۱ حجم مبنا، آخرین قیمت و قیمت پایانی	۸
۱۰	۵.۱.۱ ریسک و بازده	۱۰
۱۲	۲.۱ تعاریف تخصصی ریاضیات و علوم کامپیوتر	۱۲
۱۲	۱.۲.۱ بردار و ماتریس	۱۲
۱۴	۲.۲.۱ مفاهیم حوزه گراف	۱۴
۱۶	۳.۲.۱ سایر تعاریف	۱۶
۱۸	۳.۱ کتابخانه های استفاده شده	۱۸
۲۰	۲ الگوریتم های یافتن معاملات متقلبانه	۲۰
۲۱	۱.۲ معاملات چرخشی	۲۱
۲۱	۱.۱.۲ تعریف	۲۱
۲۳	۲.۱.۲ الگوریتم شناسایی	۲۳
۲۷	۲.۲ معاملات همروند	۲۷



۲۷	تعریف	۱.۲.۲
۲۸	الگوریتم شناسایی	۲.۲.۲
۳۲	معاملات با سود مشکوک	۳.۲
۳۲	تعریف	۱.۳.۲
۳۲	الگوریتم شناسایی	۲.۳.۲
۳۵	سنجش اثر معاملات مشکوک بر سهم و بازار	۳
۳۵	گرفتن داده و محاسبات اولیه	۱.۳
۳۷	آمار توصیفی	۲.۳
۴۰	نرمال سازی	۳.۳
۴۰	بررسی توصیفی	۱.۳.۳
۴۱	آزمون شپيرو	۲.۳.۳
۴۲	تبدیل باکس-کاکس	۳.۳.۳
۴۲	بررسی همبستگی و شیب خط رگرسیون	۴.۳
۴۴	معرفی مدل های جایگزین سری زمانی	۵.۳
۴۴	داده های مقطعی	۱.۵.۳
۴۵	داده های طولی و پنلی	۲.۵.۳
۴۶	اجرای مدل دادگان پنلی	۳.۵.۳
۴۷	نتایج و خروجی	۴
۴۷	نتایج الگوریتم معاملات چرخشی	۱.۴
۴۸	نتایج الگوریتم سود مشکوک	۲.۴
۴۸	نتایج الگوریتم معاملات همروند	۳.۴
۴۹	نتایج مدل مقطعی	۴.۴
۵۱	نتایج مدل پنلی	۵.۴

# فصل ۱

## مفاهیم مقدماتی

### ۱.۱ تعاریف حوزه بازارهای مالی و اقتصاد

#### ۱.۱.۱ بورس و اوراق بهادار

##### تعاریف اولیه

بورس یکی از انواع بازارهای مالی است. این بازار محل مبادله انواع اوراق بهادار (سهام، حق تقدم سهام، اوراق خزانه، انواع گواهی‌های سپرده کالایی، اوراق مشتقه مانند اختیار معامله، انواع صندوق‌های سرمایه‌گذاری و ...) است. این بازار بزرگ به چهار زیر مجموعه «بورس اوراق بهادار تهران، فرابورس ایران، بورس کالا و بورس انرژی» تقسیم شده که تحت نظارت سازمان بورس و اوراق بهادار فعالیت می‌کنند.

فلسفه کلیدی شکل‌گیری بورس در جهان، ایجاد بازاری منسجم، قانونمند، پیوسته و شفاف برای تجمع سرمایه‌های خرد و کلان سرمایه‌گذاران به منظور تأمین مالی شرکت‌ها و پروژه‌های آنها است. در حقیقت مبنای اساسی تشکیل بازار بورس، همان مفهوم اولیه شراکت و تقسیم سود است. به این معنی که اگر فردی سهام شرکتی را خریداری کند، به نسبت تعداد سهامی که در اختیار دارد، مالک شرکت شده و ضمن بهره‌مندی از سودهای کسب‌شده شرکت، می‌تواند از نوسان قیمت سهام آن نیز

منتفع شود. همچنین این شخص قادر است با حضور در مجامع عمومی شرکت، در تصمیم‌گیری‌های آن مشارکت داشته باشد.

### **تعریف سهم و ارزش‌گذاری آن**

سهم یکی از انواع اوراق بهادار است. در اصل شما با پرداخت مبلغی بخشی از سهام یک شرکت را خریداری می‌کنید، به طوری که بسته به میزان سهامی که خریداری کرده‌اید در آن شرکت مالکیت دارید یا به عبارتی سهامدار آن می‌شوید.

در واقع اگر تمامی سرمایه شرکتی را به قسمت‌های مساوی تقسیم کنیم به هر قسمت یک سهم می‌گویند.

هر سهم دارای دو ارزش اسمی و بازاریست. ارزش اسمی در ایران ۱۰۰ تومان است. برای مثال شرکتی که به تازگی تاسیس و وارد بورس اوراق بهادار می‌شود برای این که تعداد سهام را مشخص کند و هر سهم ارزش‌گذاری شود باید میزان سرمایه را تقسیم بر ۱۰۰ تومان کند و بعد از مشخص شدن تعداد سهام کل، قیمت یا ارزش بازاری سهم براساس عرضه و تقاضا تعیین می‌شود. پس قیمتی که ما روزانه در هر سهم می‌بینیم همان قیمت بازاری یا واقعی آن سهم است.

### **حقوقی و حقوقی**

به طور کلی دو گروه سهام‌دار در بازار بورس فعالیت می‌کنند، افراد حقیقی و حقوقی. افراد حقیقی که به آن‌ها سهام‌داران خرد نیز گفته می‌شود، تمام کسانی هستند که به طور مستقل در بازار بورس فعالیت می‌کنند.

حقوقی‌ها یا همان سهام‌داران عمده مجموعه‌ای از سهام‌داران حقیقی هستند که در قالب یک شرکت سبگردانی، شرکت سرمایه‌گذاری، صندوق سرمایه‌گذاری، کارگزاری و غیره در بازار سهام فعالیت می‌کنند. یعنی هر فرد نمی‌تواند به صورت مستقل معامله کند بلکه همه مجموعه تحت عنوان صندوق یا شرکتی که در آن عضو هستند، فعالیت می‌کنند. در این حالت سازمان مربوطه یک سهام‌دار حقوقی نامیده می‌شود و شامل مجموعه‌ای از افراد (مدیر عاملان، کارمندان و غیره) و همچنین نام تجاری، کد اقتصادی و اطلاعات ثبت شرکت می‌باشد. پس منظور از یک سهام‌دار حقوقی، یک نهاد فعال

در بازار بورس است.

## فرا بورس

مهم‌ترین وظیفه شرکت فرا بورس ایران، ساماندهی و هدایت بخشی از بازار سرمایه است که شرایط ورود به بورس اوراق بهادار را پیدا نکرده و یا تمایل به ورود سریع‌تر به بازار را دارند. در همین راستا رویه پذیرش شرکت‌ها به گونه‌ای است که با احراز حداقل شرایط و در سریع‌ترین زمان ممکن، امکان ورود به بازار را داشته باشند. نام این بازار در ایران از معادل انگلیسی آن<sup>۱</sup> گرفته شده است. به همین دلیل به این بازار اختصاراً OTC نیز گفته می‌شود. لازم به ذکر است که سازمان بورس اوراق بهادار به عنوان نهاد ناظر، بر تمام فعالیت‌های مالی و عملیاتی شرکت‌های بورس، فرا بورس و بورس کالا نظارت دارد.

در این پژوهش تمامی نماد های حاضر در بازار فرا بورس بررسی شده اند. که معمولاً احتمال رخ دادن کلاهبرداری در آن بیشتر است. زیرا حجم شرکت و معاملات آن کمتر از بازار بورس بوده و دستکرای قیمت به سهولت انجام می‌شود.

**بازارگردان** برای حل مشکلاتی از قبیل تشکیل مداوم صف‌های خرید و فروش شرکت‌ها در بازار سرمایه، نهادی به نام بازارگردان<sup>۲</sup> و صندوق‌های بازارگردانی تشکیل شده‌اند. بازارگردان‌ها موظف‌اند روزانه تعداد مشخصی از سهام یک شرکت را معامله کنند. به عبارتی در زمانی که بورس درگیر صف‌های خرید است بازارگردان در نقش فروشنده ظاهر می‌شود و با فروش سهام، نقدشوندگی آن را افزایش می‌دهد. بازارگردان از ایجاد هیجان در معاملات جلوگیری می‌کند و زمانی که بورس به دلیل اخبار و حواشی نامناسب درگیر صف‌های فروش می‌شود با خرید سهام شرکت به صورت روزانه، با افزایش نقدشوندگی سهم، هیجان منفی کاذب را کاهش می‌دهد.

به عبارتی می‌توان گفت بازارگردان به نهاد یا شرکتی گفته می‌شود که با کسب مجوز لازم و با تعهد به افزایش نقدشوندگی، عرضه و تقاضای متوازن اوراق بهادار و تعیین دامنه نوسان قیمت، اقدام به

<sup>1</sup>Over The Counter

<sup>2</sup>Market Maker

خرید و فروش سهام می‌کند و عرضه و تقاضای بازار را کنترل می‌کند.

## سبدگردان

سبدگردانی یا مدیریت پرتفوی<sup>۳</sup> خدمتی است که توسط شرکت‌های سبدگردان انجام شده و به سرمایه‌گذاران برای مدیریت و چیدمان سبد دارایی‌شان کمک می‌کند. سبدگردانی در واقع به معنای سپردن دارایی‌ها به یک شخص حقوقی مجاز به منظور تصمیم‌گیری بهتر برای خرید، فروش یا نگهداری اوراق بهادار است. سبدگردان این اقدامات را به جای سرمایه‌گذار انجام می‌دهد. سبدگردان به پشتوانه تخصص و تجربه خود و در راستای کسب سود برای سرمایه‌گذار، دارایی‌ها را برای قرار دادن در سبد سرمایه‌گذاری انتخاب می‌کند. سبدگردان در ازای ارائه خدمات، کارمزد دریافت می‌کند. پرداخت دستمزد و کارمزد سبدگردان صرفاً بر اساس صورت‌حساب صادره انجام می‌شود. رابطه بین سبدگردان و مشتری رابطه بین وکیل و موکل است. سبدگردان می‌تواند به وکالت از مشتری خود در مجامع عمومی شرکت‌ها حضور داشته باشد و رای خود را اعلام کند. سود سهام، سود سایر اوراق بهادار، دریافت سهام جایزه، دریافت حق تقدم خرید سهام و شرکت در افزایش سرمایه را به نیابت از مشتری خود انجام دهد.

## ۲.۱.۱ سازوکار معاملات

### هسته معاملات

منظور از هسته، مرکز کنترل و به عبارتی قلب تپنده هر سیستم است که تمامی پل‌های ارتباطی در سیستم‌های مختلف را به هم دیگر مرتبط می‌کند. در واقع به هرآنچه که کنترل‌کننده امورات یک سیستم باشد، هسته گفته می‌شود. این تعریف برای هسته معاملات بورس نیز دقیقاً صادق است؛ به این معنا که تمامی اقدامات انجام شده در بورس به وسیله این هسته، کنترل و همچنین سازمان‌دهی می‌شوند.

توجه داشته باشید که انجام معاملات بورسی نیازمند یک سیستم پردازشگر بسیار قدرتمند و پرسرعت است که بتواند در مدت‌زمان کوتاهی عملیات موردنظر کاربر را انجام داده و آن را سامان‌دهی کند.

<sup>3</sup>Portfolio Management

هسته معاملات بورس قدرت پردازش اطلاعات مختلف را دارد که به منظور تقویت کارایی خود نیاز به ماژول‌های اساسی دارد؛ این کارایی هسته باید به گونه‌ای باشد که قادر به دریافت حجم بسیار بالایی از اطلاعات کاربران مختلف به صورت همزمان و بدون وجود اختلال باشد. در رابطه با نحوه کارکرد هسته معاملات بورس می‌توان گفت که در زمانی که کاربران بورسی سفارشی را درخواست می‌کنند، درخواست آن‌ها در کارگزاری به ثبت می‌رسد؛ حال پس از ثبت سفارش در کارگزاری نوبت به انتقال اطلاعات به هسته معاملات بورس می‌رسد تا پردازش‌های لازم بر روی آن انجام گیرد. پردازش بر روی اطلاعات دریافتی توسط هسته در نهایت منجر به انجام خرید و فروش در بازار سهام می‌شود.

### کارگزاری

شرکت کارگزاری بورس به شخص حقوقی گفته می‌شود که اوراق بهادار را برای دیگران و به حساب آن‌ها معامله می‌کند. به بیانی ساده‌تر شرکت کارگزاری به نهاد مالی گفته می‌شود که از سازمان بورس و اوراق بهادار مجوز دریافت کرده و به فعالیت قانونی در بازار بورس می‌پردازد.

### تطابق قیمت

می‌دانیم که برای انجام یک معامله در بورس یا سایر بسترهای اقتصادی، به خریدار و فروشنده نیاز داریم. این دو عامل سفارش‌های خودشان را در سیستم ثبت می‌کنند. در نتیجه درخواست‌هایی که مشابه هم هستند با هم جفت و جور می‌شوند و معامله انجام می‌شود. عمل منطبق شدن سفارش‌های خرید و فروش در بازار بورس با نام تطابق قیمتی<sup>۴</sup> شناخته می‌شود که برای کشف قیمت انجام می‌شود.

وقتی سفارش‌های خرید و فروش با هم جفت می‌شوند، بورس اجازه‌ی انجام معامله را می‌دهد و درخواست‌های خریدار و فروشنده اجرایی می‌شود. مهم‌ترین کارکرد جفت شدن، کشف قیمت است. در بیشتر بورس‌های اوراق بهادار در دنیا، جفت شدن قیمت به وسیله‌ی الگوریتم‌های کامپیوتری انجام می‌شود تا دقت لازم را داشته باشد. البته در ابتدا کارگزاری‌ها سفارش‌ها را به صورت

---

<sup>4</sup>Matching

حضور جفت می کردند. در سال ۲۰۱۰ سیستم تغییر پیدا کرد و تمام بازارهای اصلی به ابزارهای الکترونیک روی آوردند.

نکته‌ی مهمی که در مورد جفت شدن قیمتی وجود دارد این است که این عمل باید با سرعت و دقت بالایی انجام شود. زمانی که سیستم جفت کننده کند باشد، در امر معاملات اختلال ایجاد می کند و سود سرمایه گذاران را به خطر می اندازد. چرا که ممکن است این کندی سرعت باعث معامله در قیمتی پایین تر از سطوح ایده آل شود.

برخی از پروتکل های جفت شدن قیمتی به نفع خریداران است و برخی دیگر به نفع فروشندگان. می شود از این شیوه ها بهره مند شد و معاملات فرکانس بالا را نیز به خوبی انجام داد. در واقع بورس های مختلف به دنبال آن هستند که با اولویت بندی معاملات، فرصتی برای خریداران و فروشندگان ایجاد کنند که آن ها بتوانند به شکلی برابر از این بستر بهره مند شوند. در پی این مسئله، حجم معاملات و سفارش ها هم زیاد خواهد شد که عاملی مهم و مثبت برای بورس ها به شمار می رود.

### انواع سفارشات

سفارش روز: سفارشی است که اعتبار آن در پایان روز معاملاتی به اتمام می رسد و به صورت خودکار از سامانه معاملات حذف می شود.

سفارش جلسه: سفارشی است که اعتبار آن در پایان جلسه رسمی معاملاتی به اتمام می رسد و به صورت خودکار از سامانه معاملاتی حذف می شود.

سفارش بدون محدودیت زمانی: این سفارش تا زمانی که توسط کارگزار حذف نشود در سامانه معاملاتی باقی مانده و معتبر می باشد.

سفارش مدت دار: سفارشی است که اعتبار آن در پایان تاریخی که در زمان ورود سفارش مشخص شده به اتمام می رسد و پس از آن به صورت خودکار از سامانه معاملات حذف می شود.

سفارش زمانی: کارگزار در زمان ورود سفارش، تعداد روزهای مشخصی را جهت اعتبار سفارش مشخص می کند. این سفارش تا تاریخی که معادل تاریخ ورود سفارش به علاوه تعداد روزهای تعیین شده است، معتبر بوده و سپس به صورت خودکار از سامانه معاملات حذف می شود.

## عرضه و تقاضا

یک مفهوم اساسی در علم اقتصاد است که می تواند بر قیمت سهام و سایر دارایی های مالی در بازار سهام تأثیر بگذارد. به طور کلی، زمانی که تقاضا برای یک سهم از عرضه بیشتر شود، قیمت تمایل به بالا رفتن دارد، در حالی که زمانی که عرضه بیش از تقاضا باشد، قیمت تمایل به کاهش دارد.

دلیل این امر این است که قیمت سهام نشان دهنده توافق بازار در مورد ارزش شرکت و چشم انداز آینده آن است. وقتی تعداد خریداران برای یک سهم خاص بیشتر از فروشندگان باشد، این نشان می دهد که سرمایه گذاران تقاضای بیشتری برای سهام دارند که می تواند به دلیل اخبار مثبت در مورد شرکت، عملکرد مالی قوی یا سایر عوامل مطلوب باشد. در نتیجه، با رقابت خریداران برای خرید سهام، قیمت تمایل به افزایش دارد و فروشندگان ممکن است بتوانند قیمت بالاتری را به دست آورند.

برعکس، وقتی تعداد فروشندگان بیشتر از خریداران برای یک سهم خاص باشد، این نشان می دهد که سرمایه گذاران تقاضای کمتری برای سهام دارند که می تواند به دلیل اخبار منفی در مورد شرکت، عملکرد مالی ضعیف یا سایر عوامل نامطلوب باشد. در نتیجه، با رقابت فروشندگان برای فروش سهام خود، قیمت کاهش می یابد و خریداران ممکن است بتوانند سهام را با قیمت کمتری خریداری کنند.

به طور کلی، تأثیر متقابل بین عرضه و تقاضا، محرک اصلی قیمت سهام است و سرمایه گذاران باید هنگام تصمیم گیری برای سرمایه گذاری، عوامل اساسی را که می توانند بر عرضه و تقاضا تأثیر بگذارند، به دقت در نظر بگیرند.

## ۳.۱.۱ فاصله قیمتی و نقدشوندگی

### فاصله قیمتی عرضه و تقاضا

به تفاوت بین قیمت خریدار و قیمت فروشنده اسپرد قیمت خرید و فروش<sup>۵</sup> می گویند. قیمت مربوطه می تواند قیمت اوراق بهادار، اوراق مشارکت، کالا یا هر دارایی دیگری باشد. شخصی که

<sup>۵</sup>Bid-Ask Spread



قصد فروش دارایی دارد، قیمت پیشنهادی خریدار را دریافت می‌کند و شخصی که قصد خرید دارد، باید قیمت درخواست شده توسط فروشنده را ببیند. در تعریف دقیق‌تر، فاصله بهترین پیشنهاد خرید (بالاترین قیمت) و بهترین پیشنهاد فروش (پایین‌ترین قیمت) را اسپرد گویند. یکی از کاربردهای اسپرد، سنجش نقدشوندگی در بازار و محاسبه هزینه معاملات است. هر چقدر اسپرد کمتر باشد، نشان‌دهنده نقدشوندگی بیشتر و هزینه معاملات کمتر است. هزینه معاملات کمتر از این لحاظ که اگر خریدار پس از خرید بخواهد فوراً سهام خود را بفروشد، زیان بالایی بابت تفاوت قیمت خرید و فروش متحمل نمی‌شود.

### نقد شوندگی

نقدشوندگی<sup>۶</sup> به این معنا است که یک دارایی را در سریع‌ترین زمان ممکن و به راحتی بتوان نقد کرد، به عبارت دیگر شرایط معامله آن دارایی ساده باشد. به همین شکل نقدشوندگی سهام یعنی یک سهام را چقدر سریع می‌توان به ارزش واقعی خود فروخت. اگر بگوییم نقدشوندگی یک سهم بیشتر است یعنی افراد زیادی روزانه آن را خرید و فروش می‌کنند و حجم معاملات برای آن بالا است. زمانی که وارد دنیای سرمایه‌گذاری می‌شوید باید به موضوع نقدشوندگی دقت کنید، چون اگر به هر علتی تصمیم به تبدیل دارایی خود به پول نقد گرفتید، این کار را بتوانید در سریع‌ترین زمان ممکن انجام دهید. از جمله دارایی‌های با نقدشوندگی بالا می‌توانیم به ارز، طلا، سهام شرکت‌های بورسی و اوراق مشارکت اشاره کنیم.

## ۴.۱.۱ حجم مبنا، آخرین قیمت و قیمت پایانی

### حجم مبنا

طبق تعریف تارنمای شرکت بورس اوراق بهادار تهران به معنای تعداد مشخصی از یک نوع اوراق بهادار است که باید روزانه معامله شود تا تمام درصد تغییرات مجاز در آن تاریخ در مشخص کردن قیمت روز آینده در نظر گرفته شود.

---

<sup>۶</sup>Liquidity

هدف از تعیین این شاخص به دست آوردن اطمینان سرمایه‌گذاران از قیمت‌های واقعی و نه کاذب است. به عبارت دیگر آن‌ها مطمئن هستند نوسانات قیمت سهام به علت خرید و فروش تعداد مشخصی از سهام شرکت است.

### آخرین قیمت و قیمت پایانی

قیمت آخرین معامله، آخرین قیمتی است که در آن لحظه سهم در آن خرید و فروش شده است. قیمت پایانی، نشان‌دهنده میانگین موزون قیمت معاملات یک سهم است که با توجه به حجم مبنا و حجم معاملات سهم و همچنین قیمت پایانی روز قبل، محاسبه می‌شود.

باید توجه داشت قیمت آخرین معامله در اغلب موارد به میزان زیادی تغییر می‌کند. در همین راستا معمولاً برای اشاره به قیمت یک سهم، از قیمت پایانی استفاده می‌شود. با این وجود، قیمت آخرین معامله برای سفارش‌گذاری در بورس اهمیت زیادی دارد. اگر فردی که بخواهد معامله را به سرعت انجام دهد، می‌تواند قیمت سفارش خود را نزدیک به آخرین قیمت یا معادل آن در نظر بگیرد تا احتمال انجام سفارش بیشتر شود.

قیمت پایانی هر سهم در پایان جلسه معاملاتی، مبنای قیمت معاملات روز بعد می‌باشد. به عبارت دیگر بازه مجاز قیمتی روز بعد، بر اساس قیمت پایانی تعیین می‌شود. در صورتی که حجم معاملات انجام شده سهم طی روز مساوی یا بیشتر از حجم مبنای آن باشد، قیمت پایانی سهم برابر با میانگین موزون قیمت معاملات انجام‌شده نماد طی جلسه معاملاتی خواهد بود، در غیر این صورت، قیمت پایانی به شکل زیر محاسبه می‌شود:

$$P_1 = P_0 + \left[ \frac{N}{M} (VWAP - P_0) \right]$$

$P_1$  = قیمت پایانی روز جاری

$P_0$  = قیمت پایانی روز قبل

$VWAP$  = میانگین موزون قیمت سهام معامله شده

$N$  = حجم معاملات انجام‌شده در بازار عادی

$$M = \text{حجم مینا}$$

## ۵.۱.۱ ریسک و بازده

### بازده

بازده سهام<sup>۷</sup> معیار با اهمیتی در تصمیم‌گیری مالی و به معنای میزان سود و تغییرات قیمتی یک سهام است. در محاسبه بازده سهام تغییرات قیمتی که از مرحله افزایش سرمایه و یا از سود سهمی به وجود می‌آیند، در فرمول وارد نمی‌شوند. بازده سهام به شکل درصد بیان می‌شود. بنابراین، بازده سودی است که به صورت مستقیم از محل سرمایه‌گذاری به دست می‌آورد و یا سودی است که از تغییرات قیمت ناشی می‌شود.

سرمایه‌گذار باید بداند که چقدر از سرمایه خود را درگیر سرمایه‌گذاری کرده و چقدر هزینه برای کسب سود از سرمایه‌گذاری پرداخت نموده است. این هزینه‌ها باید با سود به دست آمده مقایسه شوند تا ارزشمندی سرمایه‌گذاری محاسبه شود. در این بین، بازه زمانی درگیری سرمایه نیز بسیار با اهمیت است. اطلاع از بازده سهام یا هر سرمایه‌گذاری دیگر، نوعی ابزار انگیزشی محسوب می‌شود.

### قیمت تعدیل شده

ابزاری است که با استفاده از آن فاصله‌های قیمتی<sup>۸</sup> پوشش داده می‌شود. این فاصله قیمتی در بورس، ناشی از افزایش سرمایه و توزیع سود نقدی است. علت اصلی تغییرات قیمت سهم در زمان بازگشایی، ثابت ماندن ثروت سهامداران پس از افزایش سرمایه و سود نقدی می‌باشد. موضوع افزایش سرمایه در مجمع عمومی فوق‌العاده و موضوع سود نقدی در مجمع عمومی عادی سالیانه بررسی می‌شود. مجمع عمومی عادی سالیانه با هدف رسیدگی به صورت‌های مالی، رسیدگی به گزارش مدیران و بازرسان شرکت و همچنین تصویب میزان سود نقدی تقسیمی بین سهام‌داران تشکیل می‌شود. مجمع عمومی فوق‌العاده، با موضوعاتی مانند تغییرات سرمایه، تغییر اساسنامه و انحلال شرکت برگزار می‌شود. در نهایت می‌توان بیان کرد که قیمت هر سهم پس از افزایش سرمایه و یا توزیع سود نقدی

<sup>7</sup>Stock returns

<sup>8</sup>price gap

در زمان برگزاری هر یک از مجامع کاهش می‌یابد که با استفاده از این ابزار، قیمت پیش از مجمع تعدیل خواهد شد.

استفاده از قیمت های تعدیل شده به جای قیمت های خام سهام از آن جهت حائز اهمیت است، که گپ قیمتی ایجاد شده به علت تقسیم سود یا مجمع می‌تواند باعث افزایش واریانس و ریسک بازده های سهم شود و سهام کم ریسک را پر ریسک‌تر جلوه بدهد.

### انواع ریسک

ریسک به احتمال از دست دادن پول یا عدم دستیابی به بازده مورد انتظار در اثر عوامل مختلف اشاره دارد. برخی از ریسک ها ذاتا مشتق شده از خود بازار هستند، در حالی که برخی دیگر مختص اوراق بهادار یا استراتژی های سرمایه گذاری هستند.

انواع مختلفی از ریسک در بازارهای مالی وجود دارد که عبارتند از:

ریسک بازار<sup>۹</sup> : به احتمال زیان ناشی از تغییرات در کل بازار مانند نوسانات نرخ بهره، تورم یا رویدادهای ژئوپلیتیکی اشاره دارد.

ریسک اعتباری<sup>۱۰</sup> : این ریسک به احتمال زیان ناشی از نکول توسط وام گیرنده یا طرف مقابل، مانند ناشر اوراق قرضه یا طرف مقابل مشتقات اشاره دارد.

ریسک نقدینگی<sup>۱۱</sup> : به احتمال زیان زمانی اشاره دارد که سرمایه گذار به دلیل کمبود خریدار یا فروشنده در بازار نتواند دارایی را با سرعت کافی یا با قیمت منصفانه بفروشد.

ریسک عملیاتی<sup>۱۲</sup> : به احتمال زیان ناشی از خطا یا خرابی در فرآیندهای داخلی، سیستم‌ها یا پرسنل اشاره دارد.

ریسک شهرت<sup>۱۳</sup> : به احتمال زیان ناشی از آسیب به اعتبار یک شرکت اشاره دارد که می‌تواند بر توانایی آن در جذب مشتری یا سرمایه گذار تأثیر بگذارد.

---

<sup>9</sup>Market risk

<sup>10</sup>Credit risk

<sup>11</sup>liquidity risk

<sup>12</sup>Operational risk

<sup>13</sup>Reputational risk

ریسک سیستمی<sup>۱۴</sup> : به احتمال زیان ناشی از فروپاشی گسترده سیستم مالی، مانند زمان بحران مالی اشاره دارد.

از سوی دیگر، ریسک پورتفولیو به ریسک کلی یک سبد سرمایه گذاری با در نظر گرفتن انواع مختلف ریسک مرتبط با هر سرمایه گذاری فردی اشاره دارد. با تنوع بخشیدن به پرتفوی در بین طبقات مختلف دارایی و استراتژی های سرمایه گذاری، سرمایه گذاران می توانند ریسک پرتفوی را کاهش دهند و به طور بالقوه به بازده تعدیل شده با ریسک بهتری دست یابند.

در این پژوهش منظور از ریسک، انحراف معیار<sup>۱۵</sup> بازده سهام است. ریسک یک سهام اغلب با انحراف استاندارد آن اندازه گیری می شود، که معیاری آماری از میزان تغییرات یا پراکندگی مجموعه ای از نقاط داده از میانگین آن است. در زمینه سهام، انحراف استاندارد معمولاً برای اندازه گیری نوسان یا تغییرپذیری بازده سهام در یک دوره زمانی معین استفاده می شود. انحراف استاندارد بالاتر نشان می دهد که بازده سهام بی ثبات تر بوده یا به طور گسترده تری از میانگین متفاوت بوده است، که نشان می دهد سهام دارای سطح ریسک بالاتری است. از سوی دیگر، انحراف معیار کمتر نشان می دهد که بازده سهام پایدارتر بوده یا کمتر از میانگین متفاوت بوده است، که نشان می دهد سهام دارای سطح ریسک پایین تری است.

## ۲.۱ تعاریف تخصصی ریاضیات و علوم کامپیوتر

### ۱.۲.۱ بردار و ماتریس

بردار لیست مرتبی از اعداد (یا دیگر اشیاء ریاضی) است که می تواند برای نمایش کمیتی استفاده شود که هم اندازه و هم جهت دارد. در ریاضیات، بردارها اغلب با حروف پررنگ، مانند  $\mathbf{v}$  نشان داده می شوند.

---

<sup>14</sup>Systemic risk

<sup>15</sup>standard deviation

ماتریس یک آرایه دو بعدی از اعداد (یا دیگر اشیاء ریاضی) است که می‌تواند برای نمایش مجموعه ای از بردارها یا تبدیل خطی بین فضاهای برداری استفاده شود. در ریاضیات، ماتریس‌ها اغلب با حروف بزرگ پررنگ، مانند **A** نشان داده می‌شوند و معمولاً به شکل  $A = [a_{ij}]$  نوشته می‌شوند، که در آن  $a_{ij}$  نشان‌دهنده مقدار سطر  $i$  و ستون  $j$  است.

بردارها و ماتریس‌ها در محاسبات کارت گرافیک<sup>۱۶</sup> (واحد پردازش گرافیک) مهم هستند زیرا محاسبات سریع‌تر و کارآمدتری را برای انواع خاصی از الگوریتم‌ها، به ویژه آنهایی که شامل مقادیر زیادی از موازی‌سازی داده‌ها هستند، امکان‌پذیر می‌کنند.

پردازنده‌های گرافیکی برای انجام محاسبات موازی روی مجموعه داده‌های بزرگ، مانند مواردی که در رندر گرافیک سه‌بعدی، شبیه‌سازی‌های علمی و برنامه‌های یادگیری ماشین یافت می‌شوند، طراحی شده‌اند. برای دستیابی به این هدف، پردازنده‌های گرافیکی مجهز به هزاران هسته پردازشی هستند که می‌توانند به صورت موازی کار کنند، و از معماری‌های حافظه تخصصی استفاده می‌کنند که امکان دسترسی سریع به حجم زیادی از داده‌ها را فراهم می‌کند.

بردارها و ماتریس‌ها به ویژه برای پردازنده‌های گرافیکی مناسب هستند زیرا می‌توان آنها را به راحتی موازی کرد و به صورت دسته‌ای پردازش کرد که می‌تواند انواع خاصی از محاسبات را تا حد زیادی تسریع کند. به عنوان مثال، در ضرب ماتریس، عناصر جداگانه در هر سطر و ستون را می‌توان به طور مستقل محاسبه کرد، که آن را به یک عملیات ایده آل برای پردازش موازی در یک GPU تبدیل می‌کند. به طور مشابه، در الگوریتم‌های یادگیری ماشین مانند شبکه‌های عصبی، محاسبات شامل ماتریس‌های بزرگی از وزن‌ها و فعال‌سازی‌ها می‌شود که می‌توانند به طور موازی با استفاده از GPU پردازش شوند.

## پشته و صف

در علوم کامپیوتر، پشته<sup>۱۷</sup> و صف<sup>۱۸</sup> دو ساختار داده رایج هستند که برای ذخیره و سازماندهی

<sup>16</sup>GPU

<sup>17</sup>Stack

<sup>18</sup>Queue

مجموعه‌ای از عناصر استفاده می‌شوند.

پشته یک ساختار داده خطی است که از اصل آخرین ورودی، اولین خروجی<sup>۱۹</sup> پیروی می‌کند، به این معنی که آخرین عنصر اضافه شده به پشته اولین عنصری است که حذف می‌شود. عناصر را می‌توان از بالای پشته اضافه یا حذف کرد.

از طرف دیگر، صف یک ساختار داده خطی است که از اصل اولین ورودی، اولین خروجی<sup>۲۰</sup> پیروی می‌کند، به این معنی که اولین عنصر اضافه شده به صف اولین عنصری است که حذف می‌شود. عناصر را می‌توان به پشت صف اضافه کرد و از جلوی صف حذف کرد.

## ۲.۲.۱ مفاهیم حوزه گراف

در علوم کامپیوتر، گراف<sup>۲۱</sup> مجموعه‌ای از گره‌ها (همچنین به عنوان رئوس شناخته می‌شوند)<sup>۲۲</sup> و یال‌هایی<sup>۲۳</sup> است که جفت گره‌ها را به هم متصل می‌کنند. یک گراف می‌تواند برای نمایش انواع ساختارهای دنیای واقعی مانند شبکه‌های اجتماعی، سیستم‌های حمل و نقل و شبکه‌های کامپیوتری استفاده شود.

در یک گراف، هر راس نشان دهنده یک موجودیت یا مفهوم متمایز است، مانند یک شخص، شهر یا کامپیوتر. در حالی که هر یال نشان دهنده یک رابطه یا ارتباط بین دو راس است. بسته به اینکه رابطه بین راس‌ها یک طرفه یا دو طرفه باشد یال‌ها می‌توانند جهت دار یا بدون جهت باشند. یال‌ها همچنین می‌توانند وزن دار باشند، به این معنی که دارای یک مقدار عددی هستند که قدرت یا فاصله رابطه بین گره‌ها را نشان می‌دهد.

گراف‌ها اغلب با استفاده از انواع ساختارهای داده مانند ماتریس مجاورت یا لیست مجاورت نشان داده می‌شوند. این ساختارهای داده دسترسی کارآمد به راس‌ها و یال‌های گراف را امکان پذیر می‌کنند.

---

<sup>19</sup>Last-In-First-Out (LIFO)

<sup>20</sup>First-In-First-Out (FIFO)

<sup>21</sup>Graph

<sup>22</sup>vertex

<sup>23</sup>edge

ماتریس مجاورت یک ماتریس مربعی است که نشان دهنده ارتباط بین راس‌ها در یک گراف است. ماتریس دارای یک سطر و ستون برای هر راس در نمودار است و مقدار در سطر  $i$  و ستون  $j$  نشان دهنده یال بین راس  $i$  و راس  $j$  است. اگر یک یال بین دو راس وجود داشته باشد، ورودی معمولاً روی ۱ تنظیم می‌شود، در حالی که اگر یال وجود نداشته باشد، ورودی روی ۰ تنظیم می‌شود. ماتریس مجاورت یک ساختار داده مفید برای نمایش گراف‌ها است زیرا امکان جستجو و دستکاری کارآمد یال‌ها و راس‌ها را فراهم می‌کند. علاوه بر این، ماتریس مجاورت می‌تواند برای محاسبه تنک بودن <sup>۲۴</sup> یک گراف استفاده شود. تنک بودن گراف، معیاری است که نشان می‌دهد چگالی یا پراکندگی گراف چقدر است و به عنوان نسبت تعداد یال‌های نمودار به حداکثر تعداد یال‌های ممکن تعریف می‌شود. تنک بودن یک گراف را می‌توان به راحتی از روی ماتریس مجاورت آن با شمارش تعداد مقادیر غیر صفر (یعنی تعداد یال‌ها) و تقسیم بر تعداد کل مقادیر ماتریس محاسبه کرد.

### جستجوی اول عمق

جستجوی اول عمق <sup>۲۵</sup> الگوریتم پیمایش گراف است که برای کاوش و جستجو از طریق یک نمودار یا ساختار داده درختی استفاده می‌شود. در DFS، الگوریتم از یک گره خاص (گره ریشه) <sup>۲۶</sup> شروع می‌شود، تا آنجا که ممکن است در امتداد هر شاخه قبل از عقب نشینی کاوش می‌کند، و سپس به کاوش در شاخه بعدی ادامه می‌دهد تا زمانی که همه گره‌ها بازدید شوند. الگوریتم DFS را می‌توان به صورت بازگشتی یا تکراری پیاده‌سازی کرد. هنگامی که به صورت بازگشتی پیاده‌سازی می‌شود، الگوریتم از گره ریشه شروع می‌شود، از رأس بازدید نشده مجاور بازدید می‌کند، آن را به عنوان بازدید شده علامت‌گذاری می‌کند، و سپس به صورت بازگشتی رأس مجاور را تا زمانی که همه راس‌ها بازدید شوند بررسی می‌کند.

### خوشه بندی

<sup>24</sup>Sparsity

<sup>25</sup>Depth-First Search(DFS)

<sup>26</sup>Root



خوشه‌بندی تکنیکی در یادگیری ماشین و تجزیه و تحلیل داده است که شامل گروه‌بندی نقاط مشابه بر اساس ویژگی‌های آنها می‌شود. هدف از خوشه‌بندی شناسایی الگوها و ساختارهایی در داده‌ها است که می‌تواند به ما در درک بهتر آن‌ها و پیش‌بینی‌هایی در مورد نقاط داده جدید کمک کند. خوشه‌بندی گراف که به تشخیص جامعه<sup>۲۷</sup> نیز شناخته می‌شود، نوعی از خوشه‌بندی است که به طور خاص برای تجزیه و تحلیل گراف‌ها یا شبکه‌ها استفاده می‌شود. در خوشه‌بندی گراف، هدف شناسایی گروه‌هایی از گره‌ها (یا رئوس) در یک گراف است که به شدت با یکدیگر مرتبط هستند، در حالی که نسبتاً از گروه‌های دیگر در گراف جدا هستند. این گروه‌ها جوامع نامیده می‌شوند و می‌توانند بینشی در مورد ساختار و عملکرد شبکه ارائه دهند.

### ۳.۲.۱ سایر تعاریف

#### پیچیدگی زمانی

پیچیدگی زمانی<sup>۲۸</sup> معیاری است که نشان می‌دهد با افزایش اندازه داده‌های ورودی، یک الگوریتم چقدر زمان می‌برد تا یک مسئله را حل کند. معمولاً به عنوان تابعی از اندازه ورودی بیان می‌شود که با  $n$  نشان داده می‌شود و بدترین سناریو را برای عملکرد الگوریتم نشان می‌دهد. پیچیدگی زمانی مهم است زیرا به ما کمک می‌کند کارایی الگوریتم‌های مختلف را برای حل یک مسئله ارزیابی کنیم. الگوریتمی با پیچیدگی زمانی کمتر معمولاً سریعتر و کارآمدتر از الگوریتمی با پیچیدگی زمانی بالاتر است، به خصوص برای اندازه‌های ورودی بزرگ. به عنوان مثال، یک الگوریتم با پیچیدگی زمانی  $O(n)$  به این معنی است که زمان اجرای الگوریتم به صورت خطی با اندازه ورودی رشد می‌کند، در حالی که الگوریتم با پیچیدگی زمانی  $O(n^2)$  به این معنی است که زمان اجرا به صورت درجه دوم رشد می‌کند.

#### مثبت و منفی کاذب

<sup>27</sup>Community Detection

<sup>28</sup>Time Complexity

در یک مسئله طبقه‌بندی باینری<sup>۲۹</sup>، هدف این است که پیش‌بینی کنیم که آیا یک مشاهده به دسته خاصی تعلق دارد یا خیر. دو نتیجه احتمالی خواهیم داشت.

مثبت: پیش‌بینی می‌شود که مشاهده متعلق به دسته باشد.

منفی: پیش‌بینی می‌شود که مشاهده به دسته تعلق نداشته باشد.

بر اساس این دو نتیجه، چهار سناریو ممکن می‌تواند رخ دهد.

مثبت صحیح<sup>۳۰</sup>: مشاهده متعلق به دسته است و به درستی پیش‌بینی می‌شود که به دسته تعلق دارد.

مثبت کاذب<sup>۳۱</sup>: مشاهده به دسته تعلق ندارد، اما به اشتباه پیش‌بینی می‌شود که به دسته تعلق دارد.

منفی صحیح<sup>۳۲</sup>: مشاهده به دسته تعلق ندارد و به درستی پیش‌بینی شده است که به دسته تعلق ندارد.

منفی کاذب<sup>۳۳</sup>: مشاهده متعلق به دسته است، اما به اشتباه پیش‌بینی شده است که به دسته تعلق ندارد.

به عبارت دیگر، مثبت کاذب زمانی رخ می‌دهد که یک تست یا مدل به اشتباه یک مورد منفی را مثبت تشخیص دهد، در حالی که منفی صحیح زمانی رخ می‌دهد که یک مورد منفی به درستی به عنوان منفی شناسایی شود. معمولاً مثبت‌های کاذب در شرایطی عواقب منفی بیشتری دارند مانند آزمایش‌های پزشکی. در مورد تشخیص تقلب نیز این مورد صدق می‌کند و هدف ما کاهش مثبت‌های کاذب است.

## همبستگی

همبستگی پرکاربردترین معیار آماری برای ارزیابی روابط بین متغیرها است. این معیار، میزان ارتباط خطی دو متغیر را بیان می‌کند (به این معنی که آنها با یک نرخ ثابت تغییر می‌کنند).

ضریب همبستگی، که با  $r$  نشان داده می‌شود، از ۱- تا ۱ متغیر است. ضریب همبستگی ۱ نشان دهنده همبستگی مثبت کامل است، به این معنی که دو متغیر با هم در یک جهت حرکت می‌کنند.

<sup>29</sup>Binary Classification

<sup>30</sup>True Positive

<sup>31</sup>False Positive

<sup>32</sup>True Negative

<sup>33</sup>False Negative

ضریب همبستگی ۱ - نشان دهنده همبستگی منفی کامل است، به این معنی که دو متغیر با هم در جهت مخالف حرکت می‌کنند. ضریب همبستگی صفر نشان دهنده عدم همبستگی است، به این معنی که متغیرها به هم مرتبط نیستند. همبستگی پیرسون برای اندازه‌گیری قدرت و جهت یک رابطه خطی بین دو متغیر استفاده می‌شود:

$$r(xy) = \frac{cov(x, y)}{s_x * s_y}$$

### ۳.۱ کتابخانه‌های استفاده شده

#### **networkx**

یک کتابخانه محبوب پایتون برای کار با شبکه‌های پیچیده مانند شبکه‌های اجتماعی، شبکه‌های بیولوژیکی و شبکه‌های حمل و نقل است. این ابزارها را برای ایجاد، دستکاری و تجزیه و تحلیل گراف‌ها و شبکه‌ها با تمرکز بر سهولت استفاده و انعطاف‌پذیری فراهم می‌کند. برخی از ویژگی‌های کلیدی NetworkX عبارتند از:

پشتیبانی از ایجاد و دستکاری گراف‌ها و شبکه‌ها با گره‌ها و یال‌ها. پشتیبانی از انواع مختلف گراف‌ها، از جمله گراف‌های جهت دار و غیر جهت دار، گراف‌های وزن دار و بدون وزن، و گراف چندگانه (گراف‌هایی با یال‌های متعدد بین گره‌ها). طیف وسیعی از الگوریتم‌ها برای تجزیه و تحلیل و تجسم گراف‌ها، از جمله تشخیص جامعه، خوشه‌بندی، الگوریتم‌های کوتاه‌ترین مسیر و موارد دیگر.

#### **scipy.sparse**

هنگامی که صحبت از ماتریس‌های تنک به میان می‌آید، نیازمند الگوریتم‌های ویژه‌ای هستیم که متفاوت از ماتریس‌های عادی با این ماتریس‌ها رفتار کند. زیرا معمولاً این ماتریس‌ها با وجود تنک بودن اندازه بزرگی دارند و عملیات‌های ماتریس‌های عادی روی آنها زمانبر خواهد بود. این

کتابخانه پیاده‌سازی‌های کارآمدی از ماتریس‌های تنک و عملیات مرتبط را ارائه می‌دهد. علاوه بر آن این کتابخانه طیف وسیعی از توابع را برای انجام عملیات ماتریس رایج بر روی ماتریس‌های تنک، مانند ضرب ماتریس، ضرب ماتریس-بردار، وارونگی ماتریس، و تجزیه ارزش ویژه ارائه می‌دهد. این توابع برای ماتریس‌های پراکنده بهینه‌سازی شده‌اند و می‌توانند عملکرد قابل توجهی را نسبت به عملیات ماتریس متراکم در هنگام برخورد با ماتریس‌های تنک و بزرگ ارائه دهند.

## **numpy & pandas**

NumPy و Pandas دو کتابخانه محبوب پایتون هستند که برای تجزیه و تحلیل داده‌ها استفاده می‌شوند. NumPy کتابخانه‌ای است که از آرایه‌ها و ماتریس‌های بزرگ و چند بعدی به همراه مجموعه بزرگی از توابع ریاضی برای کار بر روی این آرایه‌ها پشتیبانی می‌کند. اغلب برای محاسبات علمی و وظایف تجزیه و تحلیل داده‌ها که به عملیات عددی با کارایی بالا نیاز دارند استفاده می‌شود. از سوی دیگر، Pandas کتابخانه‌ای است که بر روی NumPy ساخته شده است که ساختارها و عملکردهای داده سطح بالایی را برای دستکاری و تجزیه و تحلیل داده‌ها فراهم می‌کند. اغلب برای تمیز کردن داده‌ها، آماده سازی و تجزیه و تحلیل وظایف و همچنین برای تصویرسازی داده‌ها استفاده می‌شود.

## فصل ۲

### الگوریتم های یافتن معاملات متقلبانه

نرم افزارهای نظارتی از گذشته وجود داشته اند. این نرم افزارها به صورت مستقل یا به عنوان بخشی از سایر نرم افزارهای جامع به کاربران ارائه می شده اند. دسته بندی بر اساس نوع نگاه و منطق ورود به حوزه نظارت به دو دسته زیر تقسیم بندی شده اند:

۱. قاعده محور<sup>۱</sup>: به زبان ساده، سیستم های قاعده محور با پیاده سازی یک سری قواعد روی جریان اطلاعات موجود در سیستم، هشدارهایی را فعال می کنند. این سیستم ها به صورت سنتی کاملاً ثابت بوده و در بسیاری موارد انبوهی از نتایج را تولید می کنند. مشکل بزرگ این مدل، تعداد نتایج و کثرت مثبت های کاذب است. تعداد بالای نتایج نه تنها دید منسجم نسبت به تخلفات نمی دهد بلکه با توان پیگیری سازمان نظارت نیز متناسب نیست.

۲. ریسک محور<sup>۲</sup>: به صورت خلاصه در سیستم های ریسک محور، تلاش بر این است که فرآیند نظارت و تطبیق به ترتیب از موجودیت هایی آغاز شود که ریسک بیشتری را دارند و سپس به سایرین تسری پیدا کند. کلیدی ترین روش دستیابی به یک سیستم کارای ریسک محور، تخصیص

---

<sup>۱</sup>Rule Based

<sup>۲</sup>Risk Based

ریسک دقیق و متناسب می‌باشد. به بیان دیگر، اگر ریسک به شکل صحیحی تخصیص یابد، می‌توان بزرگترین تخلفات را شناسایی کرده و از ادامه زنجیره جلوگیری نمود. ابزار تخصیص ریسک می‌تواند از تکنیک‌های پیشرفته یادگیری ماشین و هوش مصنوعی یا تکنیک‌های ساده و خلاقانه بهره گیرد.

در مورد تشخیص تقلب، هدف ما در گام اول پیاده‌سازی الگوریتم‌های ریسک محور به منظور تشخیص سه نوع از متداول‌ترین تقلب‌های بازار سرمایه است. بدین منظور از داده‌های بازه ۶ ماهه ابتدای سال ۱۳۹۹ تا انتهای شهریور همان سال در بازار فرابورس ایران استفاده شده‌است. این دادگان شامل اطلاعات تمام ریزمعاملات صورت گرفته در بیش از ۶۰۰ نماد فعال در فرابورس است. این اطلاعات کاملاً محرمانه بوده و به صورت کدگذاری شده و تحت نظارت سازمان فرابورس ایران در اختیار پژوهشگر قرار داده شده است.

این دادگان شامل پایگاه داده عظیمی از جزئیات معاملات شامل کد خریدار، کد فروشنده، تاریخ و ساعت معامله، کد کارگزاری‌های خریدار و فروشنده، حجم و قیمت معامله و باقی اطلاعات جزئی دیگر است.

در ادامه جزئیات الگوریتم‌های طراحی شده برای تشخیص هرکدام از انواع تخلفات آمده است.

## ۱.۲ معاملات چرخشی

### ۱.۱.۲ تعریف

یکی از تخلف‌های متداول، استفاده از روش معاملات چرخشی<sup>۳</sup> است. در این روش یک حلقه بسته از افراد حداقل دو نفره تشکیل شده و سهم بین آنها دست به دست می‌شود و نهایتاً به نفر اول که صاحب سهم بود باز می‌گردد. در این نوع معامله، وابسته به روش انجام آن حجم ثابت و یا ارزش ثابتی از سهم دست به دست می‌شود و به نفر اول باز می‌گردد. این تخلف عموماً با هزینه ناچیز قابل انجام است. بازیگران معمولاً از فاصله بین سرخط خرید و فروش استفاده کرده تا معاملات بین

<sup>۳</sup>Circular Trading

یکدیگر را انجام دهند. بدین صورت که در هنگامی افزایش اسپرد سهم، یک معامله خرید و یک معامله فروش بزرگ را به طور همزمان و با حجم نزدیک به هم در وسط حدفاصل سرخط خرید و فروش متبادل کنن تا کمترین ریسک را از جهت اجرای معامله بدست باقی معامله گران داشته باشند. این معاملات اهداف زیر را دنبال می کنند:

#### ۱. افزایش حجم معاملات

معمولا برخی سهم های بازار حجم مبادله پایینی دارند. افزایش حجم معاملات حتی اگر بدون دستکاری قیمت انجام بگیرد، باعث تبادر سیگنال مثبت در مورد آن سهم خاص است در حالی که این حجم انجام شده توسط گروه بسیار کوچکی از افراد انجام شده و غیرواقعی و گمراه کننده است. معامله گران خرد مخصوصا با این سیگنال گمراه شده و تصور می کنند سهم با حجم معاملات بالا نقدشوندگی قابل قبولی دارد و وارد سهم می شوند.

#### ۲. رساندن سهم به حجم مبنای آن و تغییر قیمت نهایی سهم

همانطور که توضیح داده شد، حجم مبنا به عنوان ابزاری برای محاسبه قیمت نهایی سهم معرفی شد تا در نمادهای کم معامله که بازارگردان هم ندارند، تصمیم عده ای اندکی از افراد باعث تغییر قیمت نماد نشود. در این حالت از معاملات چرخشی به منظور دور زدن قانون حجم مبنا استفاده می شود و با هزینه اندک (کارمزد کارگزاری) ابتدا قیمت را به دلخواه خود تغییر داده و سپس حجم مبنا را پر می کنند.

#### ۳. داغ تر کردن فضای بازار.

در تابلوی نمای بازار که در سایت مدیریت فناوری بورس تهران<sup>۴</sup> قرار دارد، نمادها بر اساس حجم معامله شان در آن صفحه نمایش داده می شوند. نمادهایی که در روز های کم معامله بازار ناگهان پرمعامله باشند بین نماد های دیگر به چشم می آیند. مخصوصا وقتی بین هم گروهی های خود بزرگ باشند. برای مثال حجم معاملات گروه رایانه به نسبت گروه خودرویی و پتروشیمی بسیار

<sup>4</sup>tsetmc.com

کمتر است. اما نقش مهمی در متنوع‌سازی پرتفو سهامداران ایفا می‌کنند و اکثر سهامداران سعی می‌کنند یکی از نماد های گروه رایانه را هرچند با مقدار اندک در سبد سهامشان داشته باشند. در این صورت معامله‌گرانی که از معامله‌چرخشی استفاده می‌کنند، نمادی از گروه رایانه را که خودشان صلاح بدانند افزایش حجم معامله می‌دهند تا بزرگ بودن آن به نسبت باقی نمادها به چشم بیاید و سهامداران همان سهم را برای سرمایه‌گذاری انتخاب کنند.

۴. انتقال منافع بین کد حقوقی یک شرکت و مدیرعامل یا سهامدار با کد حقیقی. به انتقال حجم قابل توجهی از یک سهم بین دو کد بورسی (چه حقیقی و چه حقوقی) کد به کد گویند. بدین صورت که بین دو شخص حقیقی یا حقوقی قراردادی منعقد شده است که طی آن به جای پول نقد، دو طرف سهم مبادله می‌کنند. هنگامی که حجم این مبادلات پایین باشند، کد به کد کردن فاقد اشکال است و دو طرف می‌توانند با گذاشتن سفارش خرید و فروش همزمان در بین اسپرد قیمتی سهم را جا به جا کنند. اما هنگامی که این انتقال ارزش بیش از اندازه باشد به طوری که درصد قابل توجهی از حجم معاملات آن روز نماد باشد، این افراد باید از معاملات بلوکی برای تبادل پیشنهاد استفاده کنند. معاملات بلوکی مختص کد به کد در حجم های بالا است و در تابلوی بورس نمایش داده نمی‌شود و در خارج از زمان بازار صورت می‌گیرد تا باقی معامله‌گران به اشتباه نیفتند. انجام یک معامله بلوکی با حجم بالا در تایم بازار و در عنوان کد به کد مصداق معامله چرخشی است.

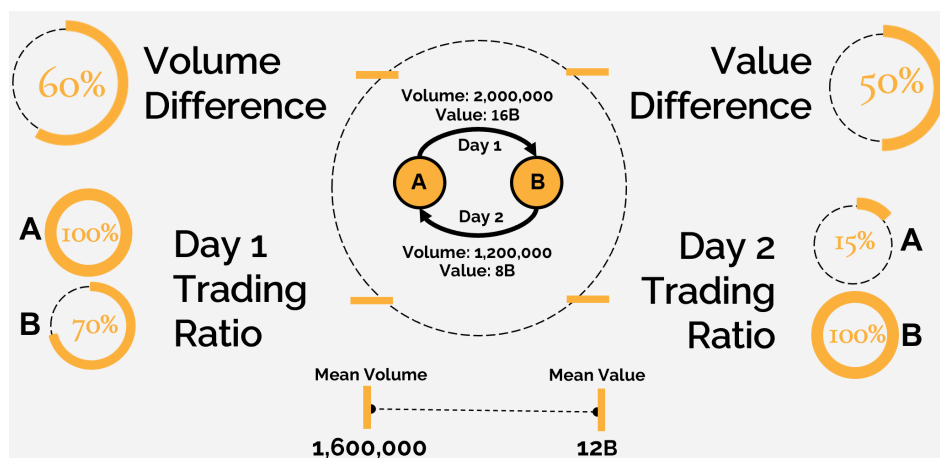
## ۲۰۱.۲ الگوریتم شناسایی

در تشخیص معاملات چرخشی دو ورودی از کاربر گرفته می‌شود و براساس آنها یک معامله چرخشی محسوب می‌شود:

اولین ورودی حداقل درصد حجم رفت به برگشت و یا حداقل درصد ارزش رفت به برگشت است که هر دو نوع معامله چرخشی گفته شده را تشخیص می‌دهد. این پارامتر از آن نظر قابل اهمیت است، که لزوماً چرخش یک سهم بین یکسری افراد نشان از معامله چرخشی ندارد. بلکه میزان آن از نظر حجم و یا ارزش باید نزدیک بهم باشد تا معامله چرخشی در نظر گرفته شود. دومین پارامتر،



حداقل ارزش سرمایه در گردش و یا حجم سهام در گردش است. ناظر می‌تواند با این فیلتر، معاملات چرخشی که ارزش در گردش آنها بالاتر است را در نظر بگیرند و مابقی را حذف کند.



شکل ۱.۲: ورودی‌های معاملات چرخشی

در مثال شکل ۱.۲ در روز اول B در جهت خرید و A در جهت فروش ظاهر شده‌اند و معامله با حجم ۲ میلیون واحد و ارزش ۱۶ میلیارد ریال انجام داده‌اند. در روز دوم B نصف سهم خریداری شده (از نظر ارزش) را به A بازگردانده است. در این معامله درصد تغییر ارزش ۵۰ درصد و درصد تغییر دارایی (حجم سهام) ۶۰ درصد است. میانگین حجم مبادله شده ۱۶۰۰۰۰۰ واحد و میانگین ارزش معامله شده ۱۲ میلیارد ریال است. از طرفی در معاملات روز اول معامله گر A، تمام معاملات او با B انجام شده است. اما در مورد B تنها ۷۰ درصد معاملاتش با A بوده است. در روز دوم این نسبت ۱۵ و ۱۰۰ است. میزان این عدد می‌تواند به ما کمک کند که تشخیص دهیم این معاملات اتفاقی بوده‌اند و یا واقعا از پیش هماهنگ شده‌اند.

تشخیص حلقه‌های با دو معامله کننده با استفاده از برداری سازی دادگان و استفاده از کتابخانه‌های numpy و pandas ساده است. دو معامله گر A و B را فرض کنید. اگر معامله از A به B با حجم مناسب انجام شده باشد (حجم معامله از آستانه ورودی داده شده توسط کاربر زیادتر باشد) ما مقدار AB را به بردار معاملات اضافه می‌کنیم. در نهایت برداری از تمام معاملات بدست خواهد آمد. حال کافیست تمام معاملات را برعکس کنیم و به جای مثلا معامله AB مقدار BA قرار دهیم.

حال بردار جدید را با بردار قبلی ترکیب می‌کنیم و مقادیر تکراری را گزارش می‌کنیم. در صورتی هم معامله AB و هم BA در بردار وجود داشته باشند، این معامله به عنوان یک معامله چرخشی با دو معامله گر گزارش می‌شود.

تشخیص حلقه های با سه معامله گر پیچیدگی مسئله را بیشتر می‌کند. چرخش سهم بین سه نفر لزوماً نشان دهنده چرخشی بودن نیست زیرا باید زمان هم لحاظ شود. فرض کنیم معاملات بین سه معامله گر A و B و C وجود داشته باشد. معاملات به ترتیب

$$A \rightarrow B, B \rightarrow C, C \rightarrow A$$

را معامله چرخشی محسوب می‌کنیم اما در حالت

$$A \rightarrow B, C \rightarrow A, B \rightarrow C$$

دیگر معامله چرخشی با سه معامله گر نداریم. پس درکل مسئله به صورت برداری شده قابل حل نیست و ناچار به بررسی غیر برداری با هزینه زمانی بیشتر هستیم. ایده اولیه برخورد با چنین مسئله‌ای، تبدیل آن به یک گراف جهت‌دار است. بدین صورت که هر راس نماینده یک معامله گر و هر یال گراف نماینده یک معامله با حجم حداقل آستانه تعیین شده توسط ناظر است. هر یال همچنین دارای زمان انجام معامله است. حال می‌خواهیم حلقه‌های گراف را پیدا کنیم. تعداد حداکثر حلقه‌های موجود در یک گراف جهت‌دار، تعداد آن در گراف کامل است که برابر با  $2^n$  است. (که در اینجا  $n$  تعداد رئوس گراف است) زیرا هرکدام از راس‌ها می‌توانند در حلقه باشند یا نباشند. پس پیدا کردن تمامی حلقه‌ها در گراف جهت‌دار از پیچیدگی زمانی  $O(2^n)$  است.

با این حال مسئله ما پیدا کردن حلقه‌های حداکثر با سه راس است. در این صورت می‌توانیم انتظار داشته باشیم که مسئله در زمان کمتری قابل حل باشد. معروف‌ترین روش پیدا کردن حلقه در گراف، استفاده از الگوریتم DFS با عمق ۳ است. در این الگوریتم، ما عمق حرکت در هر جهت را تا ۳ راس محدود می‌کنیم. اما DFS تضمین نمی‌کند که تمام حلقه‌ها را در گراف جهت‌دار پیدا کند. زیرا ممکن است از برخی از یال‌هایی که به رئوس بازدید شده قبلی منتهی می‌شوند بازدید نکند، زیرا دنبال کردن این یال‌ها یک حلقه ایجاد می‌کند. در واقع، DFS زمانی خاتمه می‌یابد که تمام

راس‌هایی را که از رأس شروع قابل دسترسی هستند بازدید کند، که ممکن است همه یال‌های گراف را شامل نشود.

به همین دلیل، مجبوریم الگوریتم DFS را به نوعی تغییر دهیم. ایده کلی بدین صورت است که DFS را بدون حافظه فراخوانی کنیم. بدین صورت که راس‌های بازدید شده را به خاطر نسپاریم تا با استفاده از مسیرهای دیگر باز به آنها دسترسی داشته باشیم. تفاوت دیگر آن است که هنگام جلو رفتن در یک مسیر از راس‌ها، باید در زمان هم جلو برویم. یعنی اگر یال AB در زمان ۹:۴۴ صبح اتفاق افتاده است، یال BC نباید قبل از ۹:۴۴ اتفاق افتاده باشد. در این صورت وارد این مسیر نشده و برگشت می‌زنیم. این الگوریتم را برای تمامی راس‌های گراف صدا می‌زنیم. بدین صورت در هر بار فراخوانی این الگوریتم برای هر راس، تمامی حلقه‌های شامل آن راس شناسایی می‌شود. همین تغییر کوچک باعث می‌شود که پیچیدگی زمانی ما به  $O(n^3)$  برسد. زیرا برای هر راس، ما تمام همسایه‌های آن و همسایه همسایه‌های آن را بازدید می‌کنیم که از پیچیدگی  $O(n^2)$  است. چون  $n$  راس داریم در نهایت پیچیدگی زمانی  $O(n^3)$  خواهد بود. این پیچیدگی زمانی اما خیلی مبالغه‌آمیز به نظر می‌رسد. به ۲ دلیل این پیچیدگی زمانی در واقعیت اتفاق نمی‌افتد:

۱. اولین دلیل آن است که در محاسبه فرض شده است که گراف کامل است. در حالی که ماتریس مجاورت گراف معاملات اتفاقاً تنک است. زیرا هر کاربر با تمام کاربران دیگر معامله با حجم بالا نداشته است.

۲. دلیل دوم وجود زمان در یال‌ها است. این باعث می‌شود که نیازی به بررسی تمامی همسایه همسایه‌ها نباشد.

در نتیجه در مورد اردر زمانی تجربی الگوریتم‌ها می‌توان گفت که بسیار سریع‌تر از خوانش داده از پایگاه داده (زمان انتقال داده از حافظه اولیه<sup>۵</sup> به ثانویه<sup>۶</sup>) عمل می‌کند که نشانگر سرعت کافی برای پردازش است.

---

<sup>۵</sup>Hard Disk

<sup>۶</sup>RAM

## ۲.۲ معاملات همروند

### ۱.۲.۲ تعریف

نوعی از معاملات که جزو تخلفات نیز هستند، معاملات هماهنگ شده یا همروند است. این نوع تخلف یکی از روش های دستکاری قیمت<sup>۷</sup> نیز به شمار می‌رود که به معنای کاهش یا افزایش قیمت توسط یک شخص یا گروه و به هدف دستکاری قیمت است. در این مدل معامله، جمعی از بازیگران به صورت همزمان اقدام به ورود به یک سهم می‌کنند و پس از مدتی به طور همزمان از سهم خارج می‌شوند. معمولاً این افراد پرتفوی مشابه همدیگر دارند. این روش معمولاً به هدف دستکاری قیمت و یا پنهان کردن معاملات بزرگ صورت می‌گیرد. مخصوصاً در نمادهای با حجم معاملات کوچکتر که هزینه تغییر قیمت کمتر باشد و به مدت چند روز و پشت هم اتفاق می‌افتد که حجم زیادی از نماد خریداری می‌شود تا معامله‌گران خیال کنند که خبر خوبی در سهم اتفاق افتاده است و آنها هم سهم را خریداری کنند و با افزایش تقاضا سهم به صورت دومینو وار افزایش قیمت می‌خورد و با قیمت ذاتی آن فاصله می‌گیرد. در این حالت می‌گوییم سهم حباب پیدا کرده است. به این معنی که همه می‌دانند که ارزش واقعی سهم این نیست، اما همچنان سهم را خریداری می‌کنند زیرا پیش‌بینی می‌کنند که سهم قرار است افزایش قیمت داشته باشد.

برعکس فرایند بالا نیز محتمل است که سهامداران حقوقی که حجم زیادی از سهام را دارند شروع به فروش سنگین سهم کنند و قیمت را کاهش دهند. در این حالت سهامداران انتظار دارند که خبر بدی از شرکت منتشر شده باشد که باعث کاهش سوددهی آن شده است و حاضر می‌شوند حتی با ضرر و یا زیر ارزش ذاتی سهم را به فروش برسانند. این موقعیت به خصوص در زمانی که کل بازار و شاخص کل کاهشی باشد رخ می‌دهد که سهامداران به آینده بازار خوش‌بین نیستند و راحت‌تر حاضرند سهام را ارزان به فروش برسانند. نهایتاً پس از کاهش دومینو وار قیمت سهم، معامله‌گران همروند دست به خرید در قیمت بسیار پایین زده و از خریدشان در قیمت پایین سود بالایی را کسب می‌کنند.

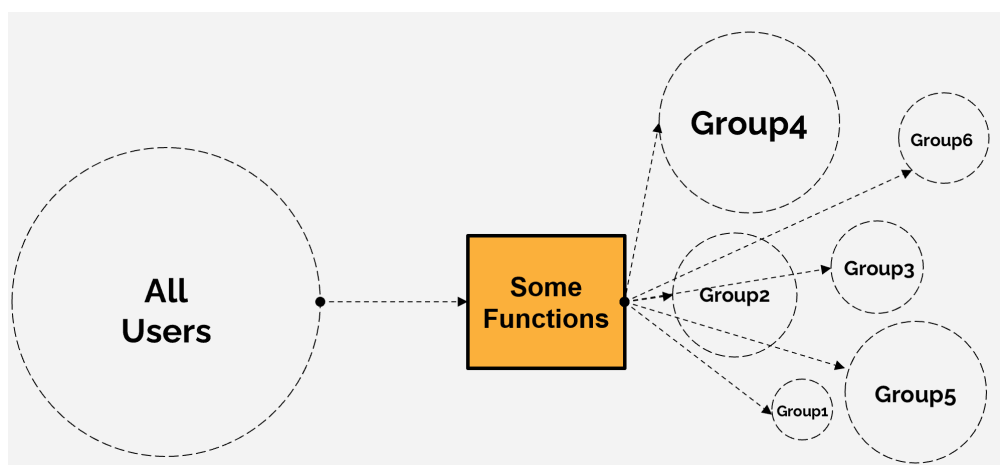
استثنا این نوع تخلف‌ها سبدگردان‌ها هستند که معمولاً با کد بورسی های مختلف اقدام به خرید و

---

<sup>7</sup>pump & dump

فروش همزمان سهم های مختلف می کنند. در این مورد، اقدامات سببگردان ها نیز توسط ناظر بازار بررسی خواهد شد.

## ۲.۲.۲ الگوریتم شناسایی



شکل ۲.۲: هدف کلی الگوریتم معاملات همروند

هدف ما خوشه بندی کردن کاربران به گروه های با رفتار مشابه است. (شکل ۲.۲) در ابتدا باید مجموعه عظیم معاملات را فیلتر کنیم. بدین منظور نمادهای کم معامله و یا افرادی که به میزان خیلی کمی معامله کرده اند توسط ناظر حذف می شوند. همچنین فرض ما از معاملات همروند معاملاتی است که افراد دخیل، با حجم بالا سعی در دستکاری آن دارند. پس برای هر معامله گر تنها معاملاتی که بیشترین حجم هارا در آن معامله کرده اند لحاظ شده و باقی معاملاتش پاک می شود. سپس الگوریتم اجرا شده و گروه های همروندی بدست می آیند.

ابتدا هر خرید هر معامله گر را تبدیل به یک بردار دودویی می کنیم. بدین صورت که اگر حجم معاملات سمت خرید شخص در یک روز خاص و در یک نماد خاص از یک آستانه مشخص شده بیشتر بود، یعنی در آن روز و نماد، معامله خرید داشته است و عدد متناظر آن یک می شود و در غیر این صورت صفر باقی می ماند.

همانطور که در شکل ۳.۲ مشخص است، هر مولفه بردارمان یک نماد در یک روز خاص است.

User buy bool vector -> (inst1,inst2,inst3,inst4,...,inst1,inst2,insts3,..., inst1,inst2,insts3,...)

Day1

Day2

Day30

InstrumentID	116	174	176	178	179	181	184	185	186	187	...	609	610	614	615	619	620	632	653	656	663
user																					
8	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
23	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
45	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
47	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
62	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1428795	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
1428804	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
1428818	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
1428839	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
1428840	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False

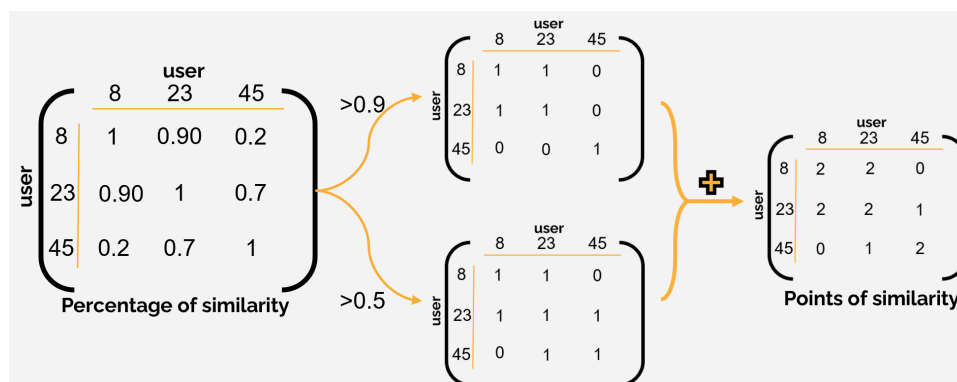
14296 rows x 1282 columns

شکل ۳.۲: تبدیل معاملات شخص به بردار دودویی

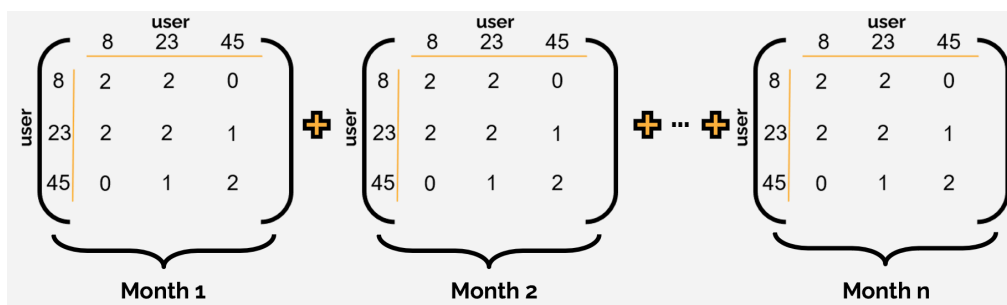
در نهایت هر معامله گر تبدیل به یک بردار باینری می شود و از کنار هم گذاشتن این معامله گران یک ماتریس به شدت تنک به وجود می آید. برای ذخیره سازی و عملیات ماتریسی روی ماتریس های تنک حتما باید از کتابخانه های مخصوص این نوع ماتریس ها مثل `scipy.sparse` استفاده شود. از ضرب این ماتریس در ترانواده آن، می توان شباهت دو به دو معامله گران را پیدا کرد. بدین صورت که هر معامله مشابه از یک نماد و در یک روز اگر مشترک باشد، حاصل آن یک و در غیر این صورت صفر خواهد شد. در نهایت درصد شباهت دو بردار دو معامله گر، نسبت تعداد یک های حاصل ضرب آنها، به ماکسیمم یک های دو بردار است.

در مرحله بعد با آستانه های ۰/۹ و ۰/۵ میزان شباهت را تبدیل به امتیاز شباهت می کنیم (شکل ۴.۲) ماتریس سمت راست امتیاز شباهت معاملات خرید هر دو کاربر دلخواه است که محاسبه آن به این صورت است که شباهت بیش از ۹۰ درصد ۲ امتیاز و شباهت بیشتر از ۵۰ درصد ۱ امتیاز و شباهت کمتر از ۵۰ درصد فاقد امتیاز خواهد بود.

برای یکسری سیستم ها با قدرت سخت افزاری پایین تر فرض شده است که داده به صورت ماه به ماه بررسی شده و نتایج جمع شود که این مورد توسط ناظر و با توجه به سیستم سخت افزاری قابل ورودی دادن است (شکل ۵.۲) ماتریس بدست آمده میزان شباهت معاملات خرید دو به دوی تمام



شکل ۴.۲: محاسبه ماتریس شباهت در معاملات



شکل ۵.۲: ادغام شباهت ها در هر ماه

معامله‌گران است. در نهایت همین ماتریس را برای معاملات فروش نیز محاسبه می‌کنیم و حاصل را در هم ضرب می‌کنیم تا میزان شباهت معاملات خرید و فروش دو به دو کاربران پیدا شود (شکل ۶.۲)

ویژگی مهمی که از ضرب این دو ماتریس حاصل می‌شود آن است که اگر بین معاملات خرید دو نفر شباهت هرچند زیادی وجود داشته باشد اما در معاملات فروش آنها شباهتی دیده نشود، ماتریس شباهت آنها صفر خواهد شد. این ویژگی از این جهت حائز اهمیت است که معامله‌گران هماهنگ صرفاً در یک طرف مشابه عمل نمی‌کنند. زیرا در هنگام مثبت بودن بازار همواره همه شروع به خرید سهم‌های مشخصی می‌کنند و اگر ما مبنا را یک طرف خرید یا فروش در نظر بگیریم همه اینها همروند می‌شوند در حالی که اینطور نیست. تعریف همروندی برای ما ورود و خروج همزمان

$$\begin{array}{c}
 \begin{array}{c} \text{user} \\ \left( \begin{array}{c|ccc} & 8 & 23 & 45 \\ \hline 8 & 2n & z & x \\ 23 & z & 2n & y \\ 45 & x & y & 2n \end{array} \right) \\ \text{Buy} \\ \text{Points of similarity}
 \end{array}
 \otimes
 \begin{array}{c}
 \begin{array}{c} \text{user} \\ \left( \begin{array}{c|ccc} & 8 & 23 & 45 \\ \hline 8 & 2n & z' & x' \\ 23 & z' & 2n & y' \\ 45 & x' & y' & 2n \end{array} \right) \\ \text{Sell} \\ \text{Points of similarity}
 \end{array}
 =
 \begin{array}{c}
 \begin{array}{c} \text{user} \\ \left( \begin{array}{c|ccc} & 8 & 23 & 45 \\ \hline 8 & 4n^2 & z'' & x'' \\ 23 & z'' & 4n^2 & y'' \\ 45 & x'' & y'' & 4n^2 \end{array} \right) \\ \text{Graph}
 \end{array}
 \end{array}$$

شکل ۶.۲: محاسبه ماتریس مجاورت گراف شباهت

به یک یا چند سهم است.

در نهایت ماتریس مشابهت معاملات همچنان ماتریس تنک است. حال این ماتریس را به صورت ماتریس مجاورت یک گراف وزن دار بدون جهت در نظر بگیرید. اگر این گراف را رسم کنیم، افراد بهم متصل یا مولفه های همبندی گراف، همان شبکه افراد مشکوک به معاملات همروند خواهند بود.

برای کاهش درصد مثبت های کاذب لازم است که فیلترهایی برای انتخاب نهایی مولفه های همبندی لحاظ شود. برای مثال لزوما ورود و خروج همزمان به فقط یک سهم خاص نشان دهنده همروند بودن نیست. زیرا خیلی از مردم رفتار گله ای داشته و هنگامی که اوضاع سهم خوب باشد با هم وارد و هنگامی که شرایط بد باشد با هم خارج می شوند و این نمونه ای از مثبت کاذب تشخیص داده شده توسط الگوریتم است. برای رفع این مشکل می توانیم شبکه افراد را روی تعداد سهم های همروند فیلتر کنیم. برای مثال یک گروه همروند هستند، اگر حداقل در ۳ نماد متفاوت همروند بوده باشند. همچنین روی ارزش معاملات همروند می توان آستانه تعیین کرد که معاملات خرد و کوچک که تاثیری روی روند معاملات نداشتند لحاظ نشوند. فیلتر دیگر نیز میزان پراکندگی میزان سرمایه گذاری افراد است. اگر در گروه همروندی تمامی افراد به میزان نزدیک به همدیگر سرمایه گذاری کرده بودند و واریانس سرمایه گذاریشان پایین بود، احتمال همروند بودن افزایش خواهد یافت.



## ۳.۲ معاملات با سود مشکوک

### ۱.۳.۲ تعریف

تشخیص سود مشکوک برای معامله‌گران از آن جهت قابل اهمیت است، که برخی سودهای مشکوک به علت دستیابی فرد به اطلاعات نهان بازار<sup>۸</sup> و یا هماهنگی با فرد حقوقی<sup>۹</sup> است. در این نوع تقلب، معامله‌گر با اطلاع قبلی از تغییر قیمت یک نماد، دست به خرید آن می‌زند (در بازارهای دو طرفه معکوس این روند در مورد اطلاع از کاهش قیمت نیز صادق است).

معاملات با سود مشکوک لزوماً بر پایه اطلاعات نهان شرکت نیستند. ممکن است فردی در کارگزاری که مسئول اجرای سفارش بزرگی از شخصی حقوقی است، پیش از انجام سفارش، خودش سفارش مشابهی با حجم کوچکتر انجام دهد و پس از تغییر قیمت موقعیت معکوس بگیرد. برای مثال، کارمند کارگزاری‌ای را در نظر بگیرید که به او گفته می‌شود حجم زیادی از سهام یک شرکت را خریداری کند. آن فرد نیز از موقعیت سو استفاده کرده و ابتدا خرید را خودش انجام می‌دهد و سپس سفارش بزرگ را اجرا می‌کند. در این صورت از افزایش قیمت ایجاد شده استفاده کرده و در قیمت بالاتر سهم خودش را می‌فروشد.

این نوع سودها معمولاً از سودهای عادی افراد قابل متمایز شدن است. تجربه اندک فرد در یک نماد (مبتدی بودن معامله‌گر)، سود کوتاه مدت و با درصد بالا در بازه زمانی کوتاه، خرید پیش از یک خرید حقوقی بزرگ، خرید پیش از یک تغییر قیمت لحظه‌ای در بازار و فروش در قیمت بالا و ناپایدار و... نشانه‌های انجام یک معامله غیرقانونی است.

### ۲.۳.۲ الگوریتم شناسایی

در ابتدا هدف الگوریتم پیدا کردن معاملات مشکوک و سود مشکوک صرف نظر از فرد است. (شکل ۷.۲) اینکه کدام معاملات و سودها مشکوک هستند وابسته به نظر ناظر سهم است. پس ناظر سهم ابتدا بازه زمانی که از نظرش بازه زمانی کوتاه مدت است را مشخص می‌کند. برای مثال ممکن

<sup>۸</sup>Insider Trading

<sup>۹</sup>Front Running



شکل ۷.۲: هدف کلی الگوریتم تشخیص سود مشکوک

است سود طی روز مدنظر باشد که ورودی را یک می‌دهد و یا سود هفتگی مدنظر باشد. همچنین می‌تواند آستانه سود مشکوک از نظر درصد سود و ارزش سود را مشخص کند.

مهم‌ترین چالش ما محاسبه سود معاملات است که روش دقیق اما بهینه‌ای در مقیاس بزرگ وجود ندارد. پس ابتدا نیاز به فیلتر کردن معامله‌گران و معاملات داریم. پس معاملات معامله‌گران در هر نماد بررسی می‌شود و در صورتی که حجم معاملاتشان در هر یک از سمت خرید یا فروش کمتر از آستانه مشخص شده توسط ناظر بود، تمام معاملات آن شخص در آن نماد حذف خواهد شد. با همین روش وابسته به آستانه مشخص شده بین ۹۵ الی ۹۹/۷ درصد معاملات حذف خواهند شد.

در مرحله بعد از بین معاملات و اشخاص باقی مانده، محاسبه سود به ازای هر شخص و هر نماد انجام خواهد شد. روش محاسبه سود استفاده از پشته زمان‌دار است. روش محاسبه بدین صورت است که معاملات سمت خرید شخص در پشته قرار می‌گیرند. اما معاملاتی که از نظر اختلاف زمانی با معامله سرپشته بیش از آستانه سود کوتاه مدت باشد، از پشته حذف می‌شوند. با این روش تضمین می‌کند که سود های محاسبه شده کوتاه مدت هستند و سود های بلند مدت لحاظ نمی‌شوند. به محض انجام شدن یک معامله سمت فروش، آن معامله با اولین معامله حاضر در پشته جفت شده و سود آن محاسبه می‌شود. در صورتی که حجم فروش کمتر از خرید بود، از حجم خرید به همان میزان کم می‌شود و در صورتی که بیشتر از آن بود، به اندازه اختلاف حجم معاملات، همین روند برای باقی پشته تکرار می‌شود تا زمانی که سفارش خریدی در پشته باقی نماند.

با این روش، جدول معاملات به جدول سود-ضرر تبدیل می‌شود که به ازای هر معامله فروش که در بازه زمانی کوتاه مدت قبل از آن، سفارش خرید متقابلی موجود باشد یک سطر سود-ضرر خواهیم داشت. (شکل ۸.۲) حال وابسته به ورودی داده شده توسط ناظر، سودهای مشکوک از باقی سودها جدا می‌شوند و به لیست سودهای مشکوک می‌رسیم. (شکل ۹.۲) در نهایت معامله‌گرانی که

	Instrument	User	buyDay	sellDay	invest	Profit	DayDiff	ProfPercent	ProfPercentDay
0	202	252779	20190612	20190616	3924000000	136000000	2.0	3.0	1.0
1	202	252779	20190709	20190710	2677000000	156066928	1.0	6.0	3.0
2	202	252779	20190727	20190728	441600000	21400000	1.0	5.0	2.5
3	202	252779	20190814	20190817	5082000000	254000000	1.0	5.0	2.5
4	202	253375	20190904	20190916	11198174720	-128254120	5.5	-1.0	-0.0
...	...	...	...	...	...	...	...	...	...
3654	250	1423744	20190716	20190727	150350000	34450000	7.0	23.0	3.0
3655	250	1423744	20190730	20190807	524400000	15600000	6.0	3.0	0.5
3656	250	1423744	20190811	20190811	184600000	400000	0.0	0.0	0.0
3657	250	1423744	20190831	20190908	9468190000	407170000	6.0	4.0	0.5
3658	250	1424133	20190731	20190810	34040746863	4701112040	4.0	14.0	3.0

شکل ۸.۲: جدول معاملات مشکوک

User	Instrument	buyDay	sellDay	Profit	invest	DayDiff	ProfPercent	ProfPercentDay	Repeated	TotalUserProf
1179846.0	4.0	2019-09-14	2019-09-15	-133750000.0	1.250000e+09	1	-11.0	-5.500000	1	-133750000.0
59245.0	97.0	2019-04-23	2019-04-23	172050603.0	3.902870e+09	0	4.0	4.000000	2	440250603.0
777969.0	97.0	2019-07-28	2019-07-28	106274330.0	2.382917e+09	0	4.0	4.000000	1	106274330.0
777748.0	103.0	2019-06-24	2019-06-25	61676032.0	1.207832e+09	1	5.0	2.500000	2	-158343828.0
777748.0	103.0	2019-06-25	2019-06-25	-220019860.0	6.394004e+09	0	-4.0	-4.000000	2	-158343828.0

شکل ۹.۲: جدول کاربران مشکوک در هر سهم

سودهای مشکوکشان از نظر تعداد یا ارزش بالا باشد، به عنوان معامله گر مشکوک (و یا معامله گر موفق) گزارش می شوند.

## فصل ۳

# سنجش اثر معاملات مشکوک بر سهم و بازار

### ۱.۳ گرفتن داده و محاسبات اولیه

در ابتدا سعی شد که به صورت کیفی رابطه همبستگی بین میزان تقلب‌ها و باقی شاخصه های سهم پیدا شود.

بدین منظور لازم بود که در بازه شش ماه مورد بررسی، قیمت سهام از سایت مدیریت فناوری بورس توسط یک خزنده وب<sup>۱</sup> دریافت شود. بدین صورت می‌توانیم بازده سهم را به صورت روزانه و ماهانه داشته باشیم. همچنین واریانس و انحراف معیار بازده‌های سهم که به عنوان ریسک سهم شناخته می‌شود را در این بازه زمانی حساب می‌کنیم.

شاخص مهم دیگری که مورد بررسی قرار می‌گیرد، بتای برگرفته از مدل ارزش‌گذاری دارایی سرمایه‌ای<sup>۲</sup> است. بتا معیاری از ریسک سیستماتیک یک دارایی یا یک سبد نسبت به کل بازار است. با حرف یونانی بتا ( $\beta$ ) نشان داده می‌شود. این پارامتر تخمینی از بازده مورد انتظار یک دارایی یا یک سبد را ارائه می‌دهد.

مدل ارزش‌گذاری دارایی سرمایه‌ای یک مدل پرکاربرد در امور مالی است که رابطه بین بازده مورد

---

<sup>1</sup>web crawler

<sup>2</sup>Beta CAPM(Capital Asset Pricing Model)

انتظار یک دارایی یا پرتفوی و ریسک آن را توصیف می‌کند. این مدل فرض می‌کند که سرمایه گذاران منطقی و ریسک گریز هستند و برای پذیرش سطوح بالاتر ریسک به بازده مورد انتظار بالاتری نیاز دارند. این مدل همچنین فرض می‌کند که بازار کارآمد است و همه سرمایه‌گذاران به اطلاعات یکسان دسترسی دارند. فرمول محاسبه بتا در ادامه آمده است.

$$\beta = \frac{cov(Er_{stock}, Er_{Market})}{var(Er_{Market})}$$

در این فرمول،

$Er_{Market}$ : بازده مورد انتظار بازار است.

$Er_{stock}$ : بازده مورد انتظار سهم است.

بتا معیاری برای سنجش حساسیت دارایی یا پرتفوی به حرکات بازار است. بتای ۱ به این معنی است که دارایی یا پرتفوی مطابق با بازار حرکت می‌کند، در حالی که بتای بزرگتر از ۱ به این معنی است که دارایی یا پرتفوی نسبت به حرکات بازار حساس‌تر است و بتای کمتر از ۱ به این معنی است که حساسیت آن کمتر است. بتای منفی به این معنی است که دارایی یا پرتفوی در جهت مخالف بازار حرکت می‌کند.

بتا با استفاده از داده‌های تاریخی در مورد دارایی یا پورتفوی و بازار تخمین زده می‌شود. این مدل فرض می‌کند که بتا در طول زمان ثابت می‌ماند.

به منظور محاسبه بتا ما نیازمند تغییرات شاخص کل در بازه زمانی مورد بررسی نیز هستیم. پس داده شاخص کل را نیز از سایت مدیریت فناوری بورس تهران گرفته و بتا هر سهم را به صورت ماهانه بدست می‌آوریم. در آخر نتیجه جدول شکل ۱.۳ خواهد بود.

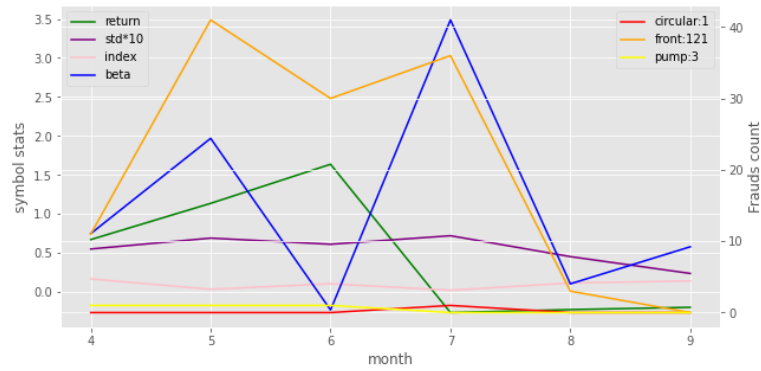
	return	std	index	beta	circular	front	pump
4	0.410866	0.051063	0.169261	2.594352	1.0	7.0	1.0
5	0.077016	0.057964	0.036165	1.378314	0.0	4.0	1.0
6	-0.003352	0.052600	0.107166	2.128115	0.0	1.0	0.0
7	0.062780	0.034928	0.023790	1.416434	0.0	6.0	2.0
8	0.966245	0.074397	0.116699	0.960581	0.0	41.0	1.0
9	-0.258047	0.019591	0.142771	0.042997	0.0	1.0	0.0

شکل ۱.۳: آمار شش ماهه یکی از نمادها

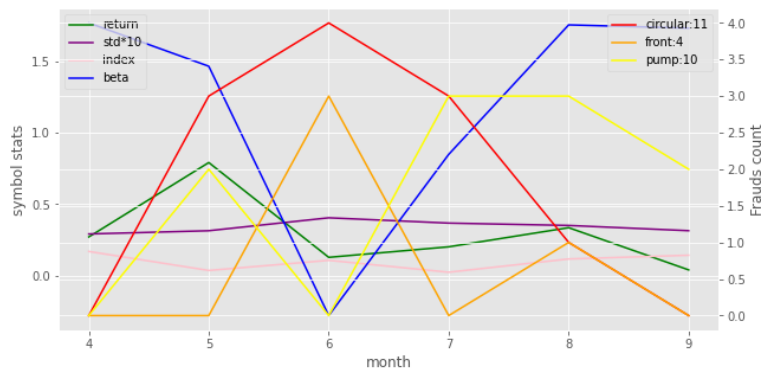
## ۲.۳ آمار توصیفی

برای آن که دید بهتری به دادگان داشته باشیم برای ۲۸ نماد که بیشترین میزان تقلب را داشتند، شاخص کل، بازده سهم، انحراف معیار و بتای سهم را در کنار تعداد تقلبها رسم می‌کنیم. برخی از این جدولها در ادامه رسم شده‌اند.

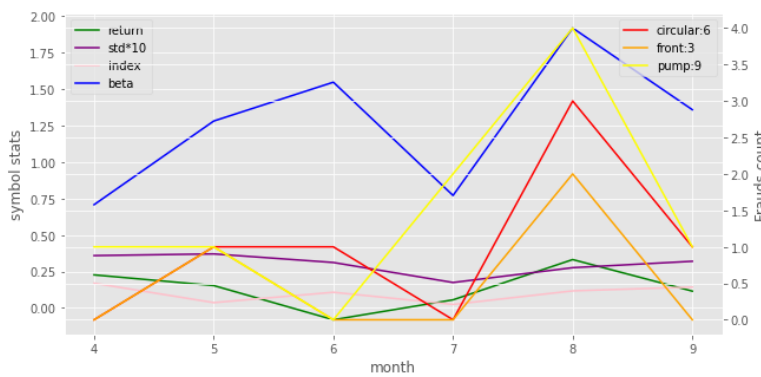
در نمودار شکل ۲.۳ رابطه همروندی معناداری بین بتا و تعداد سودهای مشکوک وجود دارد. در نمودار شکل ۳.۳ رابطه معکوس بین میزان معاملات چرخشی و بتا در سهم خنجر وجود دارد. در حالی که در سهام سرچشمه (شکل ۴.۳) رابطه مستقیم است. در نمودار شکل ۵.۳ رابطه مستقیم بین تعداد سودهای مشکوک و بازده و ریسک سهم وجود دارد. در نمودار شکل ۶.۳ در سهم اوان، رابطه معکوس معناداری بین تعداد معاملات همروند و بتا سهم وجود دارد.



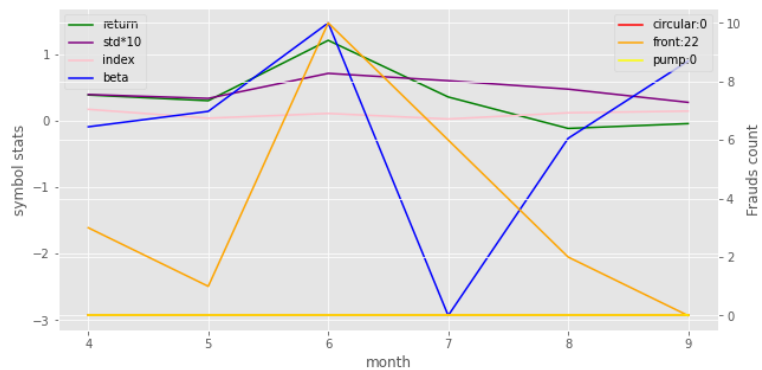
شکل ۲.۳: نمودار توصیفی سهام شمواد



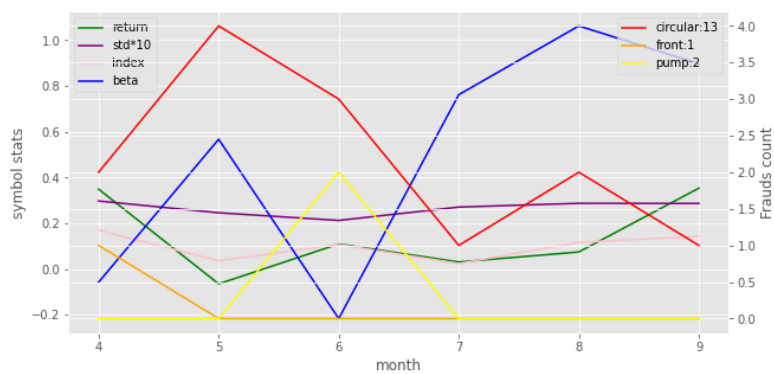
شکل ۳.۳: نمودار توصیفی سهام خنزر



شکل ۴.۳: نمودار توصیفی سهام سرچشمه



شکل ۵.۳: نمودار توصیفی سهام نژاد



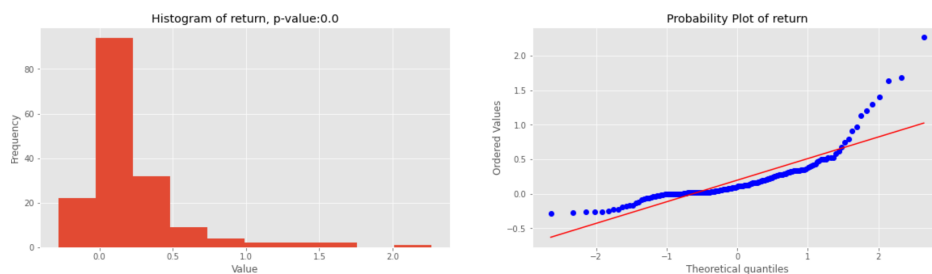
شکل ۶.۳: نمودار توصیفی سهام اوان



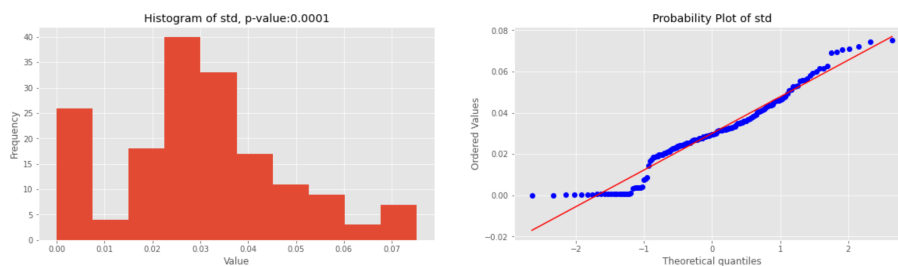
## ۳.۳ نرمال سازی

### ۱.۳.۳ بررسی توصیفی

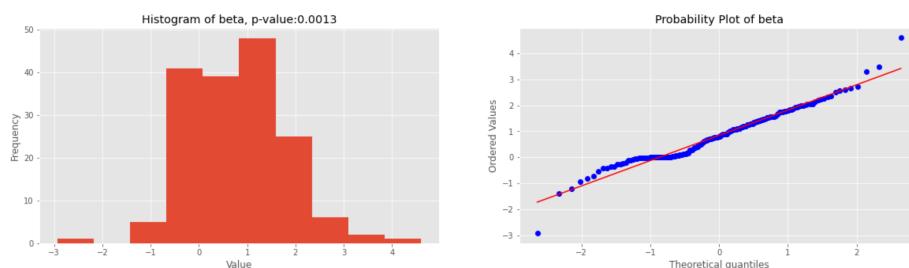
یکی از چالش های مهم این پژوهش، بازه زمانی کوتاه مورد بررسی دادگان است. به این معنا که بازه بررسی تنها شش ماه است و تنها شش مقدار برای ورودی دادن به مدل های سری زمانی خواهیم داشت. پس استفاده از مدل های سری زمانی توصیه نمی شود. پیش از رفتن سراغ مدل ها، بر روی اجماع دادگان بررسی هایی انجام می دهیم تا نرمال بودنشان بررسی شود. بدین منظور نمودار توزیع و نمودار احتمالات را برای هر پارامتر رسم می کنیم. (شکل ۷.۳ الی ۹.۳)



شکل ۷.۳: نمودار سنجش توصیفی نرمال بودن بازده



شکل ۸.۳: نمودار سنجش توصیفی نرمال بودن ریسک



شکل ۹.۳: نمودار سنجش توصیفی نرمال بودن بتا

### ۲.۳.۳ آزمون شپیرو

علاوه بر شکل توصیفی توزیع هر کدام از دادگان، آزمون شپیرو<sup>۳</sup> نیز بر روی آن‌ها اعمال شده است. شپیرو یک آزمون آماری است که برای تعیین اینکه آیا نمونه‌ای از داده‌ها از توزیع نرمال آمده است یا خیر استفاده می‌شود.

آزمون شپیرو با مقایسه توزیع مشاهده شده نمونه با توزیع نرمال مورد انتظار کار می‌کند. آماره آزمون با گرفتن تفاوت بین توزیع مشاهده شده و توزیع نرمال مورد انتظار و تقسیم بر انحراف استاندارد توزیع نرمال مورد انتظار محاسبه می‌شود. آماره آزمون به دست آمده با یک مقدار بحرانی بر اساس حجم نمونه و سطح معناداری مورد نظر مقایسه می‌شود.

فرضیه صفر آزمون شپیرو این است که نمونه از توزیع نرمال بدست می‌آید. اگر مقدار  $p$  کمتر از سطح معناداری انتخاب شده (اغلب ۰.۰۵) باشد، فرضیه صفر رد می‌شود و می‌توان نتیجه گرفت که نمونه از توزیع نرمال به دست نمی‌آید. اگر مقدار  $p$  بیشتر از سطح معنی‌داری باشد، نمی‌توان فرضیه صفر را رد کرد و می‌توان فرض کرد که نمونه از توزیع نرمال است. پس از سنجش این آزمون و مطمئن شدن از نرمال نبودن دادگان، سراغ نرمال سازی دادگان می‌رویم.

<sup>۳</sup>Shapiro-Wilk

### ۳.۳.۳ تبدیل باکس-کاکس

تبدیل باکس-کاکس<sup>۴</sup> یک تکنیک آماری است که برای تبدیل متغیرهای وابسته غیر نرمال به توزیع نرمال استفاده می‌شود. این تبدیل توسط پارامتر لامبدا  $\lambda$  تعریف می‌شود که نوع تبدیل اعمال شده را تعیین می‌کند. این تبدیل می‌تواند طیف وسیعی از تبدیل‌ها، از جمله تبدیل‌های لگاریتمی، ریشه دوم و تبدیل‌های متقابل را به عنوان موارد خاص انجام دهد. هنگامی که فرض نرمال بودن نقض می‌شود، تبدیل باکس-کاکس اغلب در تحلیل و مدل سازی آماری استفاده می‌شود. با تبدیل داده‌ها، می‌توان به بهبود اعتبار آزمون‌های آماری، کاهش ناهمسانی و خطی‌تر کردن رابطه بین متغیرها کمک کرد.

### ۴.۳ بررسی همبستگی و شیب خط رگرسیون

ابتدا روی هر سهم که دیتای نرمال شده نیز دارند، همبستگی پیرسون و همبستگی اسپیرمن<sup>۵</sup> و همچنین شیب معادله خط رگرس شده بر دادگان را بدست می‌آوریم. همبستگی رتبه اسپیرمن یک معیار آماری است که برای ارزیابی قدرت و جهت رابطه یکنواخت بین دو متغیر استفاده می‌شود. این روش توسط چارلز اسپیرمن توسعه داده شد و یک روش ناپارامتریک است، به این معنی که بر فرض توزیع خاصی برای متغیرها متکی نیست. ضریب همبستگی رتبه اسپیرمن با اختصاص رتبه به مقادیر هر متغیر و سپس محاسبه ضریب همبستگی پیرسون بر روی رتبه‌ها محاسبه می‌شود. ضریب همبستگی رتبه اسپیرمن اغلب زمانی استفاده می‌شود که متغیرهای مورد تجزیه و تحلیل ترتیبی هستند، به این معنی که مقادیر آنها را می‌توان رتبه بندی کرد. همچنین زمانی مفید است که رابطه بین متغیرها خطی نباشد یا زمانی که نقاط پرت وجود دارد که می‌تواند بر نتایج یک تحلیل همبستگی پارامتریک مانند همبستگی پیرسون تأثیر بگذارد. همچنین برای شیب خط رگرسیون در دادگان را محاسبه می‌کنیم تا دید بهتری از نتایج داشته باشیم. برای بررسی نتایج، ما سهم‌های شمواد، خزر، سرچشمه، ثنظام و اوان که در بخش قبل به صورت

<sup>4</sup>Box-Cox

<sup>5</sup>Spearman rank correlation

	circular	front	pump
return	(-0.44, -0.65, -0.03)	(0.52, 0.26, 11.25)	(0.93, 0.88, 0.43)
std	(0.49, 0.65, 1.3)	(0.9, 0.94, 889.79)	(0.45, 0.29, 9.58)
index	(-0.64, -0.65, -0.53)	(-0.83, -0.77, -252.44)	(0.09, 0.1, 0.6)
beta	(0.84, 0.65, 0.03)	(0.6, 0.54, 7.67)	(-0.22, -0.1, -0.06)

شکل ۱۰.۳: همبستگی پیرسون، اسپیرمن و شیب خط رگرسیون (نماد شمواد)

	circular	front	pump
return	(0.26, 0.09, 1.67)	(-0.21, -0.1, -0.17)	(0.18, 0.24, 0.94)
std	(0.75, 0.71, 307.31)	(0.74, 0.68, 37.26)	(0.03, 0.24, 8.54)
index	(-0.73, -0.79, -21.81)	(0.16, -0.03, 0.58)	(-0.54, -0.48, -12.7)
beta	(-0.89, -0.88, -0.42)	(-0.41, -0.34, -0.02)	(0.22, 0.0, 0.08)

شکل ۱۱.۳: همبستگی پیرسون، اسپیرمن و شیب خط رگرسیون (نماد خنزر)

	circular	front	pump
return	(0.0, nan, 0.0)	(0.89, 0.77, 6.99)	(0.0, nan, 0.0)
std	(0.0, nan, 0.0)	(0.96, 0.94, 215.36)	(0.0, nan, 0.0)
index	(0.0, nan, 0.0)	(-0.18, -0.31, -11.47)	(0.0, nan, 0.0)
beta	(0.0, nan, 0.0)	(-0.02, -0.09, -0.06)	(0.0, nan, 0.0)

شکل ۱۲.۳: همبستگی پیرسون، اسپیرمن و شیب خط رگرسیون (نماد نثظام)

	circular	front	pump
return	(-0.54, -0.44, -3.69)	(0.59, 0.39, 0.16)	(-0.09, 0.13, -0.04)
std	(-0.64, -0.44, -232.99)	(0.47, 0.65, 6.97)	(-0.83, -0.65, -19.55)
index	(-0.25, -0.15, -5.13)	(0.59, 0.65, 0.49)	(0.07, -0.13, 0.09)
beta	(-0.41, -0.53, -0.91)	(-0.52, -0.39, -0.05)	(-0.67, -0.65, -0.1)

شکل ۱۳.۳: همبستگی پیرسون، اسپیرمن و شیب خط رگرسیون (نماد اوان)

	circular	front	pump
return	(0.49, 0.31, 3.78)	(0.62, 0.68, 0.82)	(0.75, 0.58, 7.25)
std	(0.05, -0.03, 7.43)	(0.2, 0.07, 5.19)	(-0.44, -0.52, -84.86)
index	(0.13, 0.06, 2.4)	(-0.27, -0.14, -0.86)	(-0.08, -0.15, -1.85)
beta	(0.93, 0.93, 2.2)	(0.6, 0.51, 0.24)	(0.37, 0.09, 1.11)

شکل ۱۴.۳: همبستگی پیرسون، اسپیرمن و شیب خط رگرسیون (نماد سرچشمه)

توصیفی بررسی شدند را دوباره مورد بررسی قرار می‌دهیم. همانطور که مشخص است نتایج بخش توصیفی تایید شد و روابط قابل توجهی بین تقلب‌ها و سایر پارامترها برقرار شد. با این حال این اثرات در هر نماد متفاوت و حتی در برخی نمادها در تعارض با یکدیگر است. برای بررسی دقیق‌تر باید از مدل‌های طولی یا مقطعی استفاده کنیم تا برآیند اثرات را در تمامی نمادها بسنجیم.

## ۵.۳ معرفی مدل‌های جایگزین سری زمانی

همانطور که ذکر کردیم، بازه زمانی کوتاه مورد بررسی باعث ایجاد محدودیت‌هایی می‌شود که نمی‌توانیم از مدل‌های سری زمانی استفاده کنیم زیرا در مقطع کوچکی از زمان دادگان را داریم. در عوض می‌توانیم دادگان را با دو مدل دیگر تحلیل کنیم.

### ۱.۵.۳ داده‌های مقطعی

داده‌های مقطعی<sup>۶</sup>: نوعی از داده‌ها هستند که در یک نقطه از زمان، معمولاً از نمونه‌ای از افراد، سازمان‌ها یا سایر واحدهای تجزیه و تحلیل جمع‌آوری می‌شوند. در این پژوهش می‌توانیم با ادغام دادگان و حذف بعد زمان، دادگان مقطعی داشته باشیم. در داده‌های مقطعی، هر مشاهده واحد متفاوتی را در یک نقطه زمانی خاص نشان می‌دهد و داده‌ها ممکن است شامل اطلاعات متغیرها یا ویژگی‌های مختلف واحدها باشد.

به عنوان مثال، یک مجموعه داده مقطعی از افراد ممکن است شامل اطلاعاتی در مورد سن، جنسیت، سطح تحصیلات، درآمد، و سایر ویژگی‌های آنها باشد که در یک مقطع زمانی، مانند یک نظرسنجی یا سرشماری جمع‌آوری شده است. مجموعه داده‌های مقطعی از شرکت‌ها ممکن است شامل اطلاعات مربوط به صنعت، اندازه، درآمد، سود و سایر متغیرهای جمع‌آوری شده از صورت‌های مالی یا سایر منابع در یک مقطع زمانی خاص باشد.

داده‌های مقطعی را می‌توان برای توصیف ویژگی‌های یک جمعیت در یک مقطع زمانی خاص، برای

<sup>6</sup>Cross-sectional data

مقایسه گروه‌ها یا دسته‌های مختلف واحدها، یا برای شناسایی روابط بین متغیرها استفاده کرد. با این حال، داده‌های مقطعی به تنهایی نمی‌توانند برای بررسی تغییرات در طول زمان یا ایجاد روابط علی بین متغیرها استفاده شوند، زیرا اطلاعاتی در مورد چگونگی تغییر متغیرها در طول زمان یا ارتباط آنها با یکدیگر در طول زمان ارائه نمی‌دهند. برای این کار، داده‌های طولی یا سری زمانی مورد نیاز است.

از این جهت سراغ داده‌های طولی نیز خواهیم رفت.

### ۲.۵.۳ داده‌های طولی و پنلی

داده‌های طولی<sup>۷</sup> به داده‌هایی اطلاق می‌شود که از افراد یا واحدهای یکسان در چندین نقطه از زمان جمع‌آوری می‌شوند. در داده‌های طولی، هر مشاهده نشان دهنده واحد مشاهده شده در نقاط مختلف زمان است. برای مثال، یک مجموعه داده طولی از افراد ممکن است شامل اطلاعاتی در مورد سن، جنسیت، سطح تحصیلات، درآمد و سایر ویژگی‌های جمع‌آوری شده در مقاطع زمانی مختلف در طی چندین سال باشد. از داده‌های طولی می‌توان برای بررسی تغییرات در طول زمان، شناسایی الگوها یا روندها و ایجاد روابط علی بین متغیرها استفاده کرد.

از سوی دیگر، داده‌های پنلی<sup>۸</sup>: نوع خاصی از داده‌های طولی هستند که در آن مجموعه‌ای از افراد یا واحدها در چندین نقطه از زمان مشاهده می‌شوند. در داده‌های پنلی، هر مشاهده نشان‌دهنده واحد مشاهده شده در مقاطع مختلف زمانی است و داده‌ها ممکن است شامل اطلاعات متغیرها یا ویژگی‌های واحدها باشد. برای مثال، مجموعه داده‌های پنلی شرکت‌ها ممکن است شامل اطلاعاتی در مورد اندازه، درآمد، سود و سایر متغیرها باشد که در مقاطع زمانی مختلف طی چندین سال جمع‌آوری شده‌اند. داده‌های پنلی را می‌توان برای بررسی تغییرات در طول زمان، شناسایی الگوها یا روندها، ایجاد روابط علی بین متغیرها و کنترل اثرات خاص یا خاص زمان استفاده کرد.

تفاوت اصلی بین داده‌های طولی و داده‌های پنلی این است که داده‌های طولی می‌توانند افراد یا واحدهای مختلف مشاهده شده در طول زمان را شامل شوند، در حالی که داده‌های پنلی شامل

<sup>۷</sup>Longitudinal data

<sup>۸</sup>Panel data

مجموعه‌ای از افراد یا واحدهای مشاهده شده در طول زمان است. داده‌های پنلی اغلب در اقتصاد سنجی و علوم اجتماعی ترجیح داده می‌شوند، زیرا امکان کنترل ناهمگونی مشاهده‌نشده و تأثیرات خاص فرد را در زمان تغییر نمی‌دهند. داده‌های طولی اغلب در تحقیقات پزشکی و سایر زمینه‌ها استفاده می‌شود که در آن افراد یکسان باید در طول زمان ردیابی شوند، اما در جایی که ممکن است لزوماً روی تأثیرات خاص فرد نباشد.

### ۳.۵.۳ اجرای مدل دادگان پنلی

حال می‌خواهیم بررسی کنیم که سه متغیر بتا، بازده و ریسک تا چه اندازه می‌توانند تخمین‌گر میزان تقلب‌های رخ داده در بازار باشند. با این روش می‌توانیم با استفاده از پارامترهای هر نماد، پیش‌بینی کنیم که کدام سهام پتانسیل انجام تقلب بیشتری دارند. بدین منظور از مدل پنل حداقل مربعات<sup>۹</sup> استفاده می‌کنیم. این مدل، یک روش آماری است که برای تخمین مدل‌های رگرسیون خطی برای داده‌های پنلی استفاده می‌شود. این مدل مشابه رگرسیون حداقل مربعات است، اما این واقعیت را در نظر می‌گیرد که افراد یا واحدهای یکسان چندین بار در طول زمان مشاهده می‌شوند. این قابلیت باعث می‌شود تا برای مدل‌سازی اثرات فردی و زمانی خاص، و همچنین تخمین روابط کلی بین متغیرها استفاده شود. همچنین می‌تواند پنل‌های نامتعادل را که همه افراد یا واحدها تعداد مشاهدات یکسانی ندارند را کنترل کند. در ادامه و در بخش نتیجه‌گیری، خروجی‌های مربوط به داده‌های پنلی و داده‌های مقطعی که بر روی کل دادگان تمام سهام‌ها اعمال شده‌است، آمده است.

---

<sup>۹</sup>PanelOLS (Panel Ordinary Least Squares)

## فصل ۴

# نتایج و خروجی

### ۱.۴ نتایج الگوریتم معاملات چرخشی

به عنوان خروجی ما تمام معاملات چرخشی در بازه ورودی و در نمادهای ورودی را خواهیم داشت که در شرایطی که کاربر ذکر کرده بود صدق کنند. در مثال شکل ۱.۴ معاملات چرخشی نماد کگهر در یک بازه ۶ ماهه خروجی داده شده است. جدول بالایی شکل ۱.۴ مربوط به معاملات چرخشی بین دو نفر و جدول پایینی مربوط به معاملات چرخشی بین سه نفر است. ناظر بازار می تواند با توجه به اطلاعات معاملات چرخشی تشخیص دهد که کدام یک از موارد واقعا معامله چرخشی بوده اند. دقت کنید که روش ما مبتنی بر ریسک است و این معامله ها از نظر الگوریتم مشکوک و با ریسک تقلب بالا تشخیص داده شده اند و لزوما ممکن است تقلب نباشند. در نتایج خروجی کارگزاری ها و شناسه معامله گر سمت هسته نیز آمده است که در صورت مطابقت داشتن احتمال چرخشی بودن را افزایش می دهد. از طرفی امکان مشاهده روز به روز معاملات هر شخص وجود دارد (شکل ۲.۴) برای مثال دو روز معاملاتی که رفت و برگشت بین این دو نفر (شکل ۲.۴) انجام شده است نشان می دهد که اکثر معاملاتشان مشترک بوده است. نمودار آبی رنگ مختص کاربر اول و نمودار قرمز رنگ مختص کاربر دوم است. همانطور که در شکل مشخص است، در روز اول دو معامله از ۴ معامله کاربر آبی با قرمز بوده و ۲ معامله از ۵ معامله کاربر قرمز رنگ با آبی بوده که



	person1	person2	Volume Diff	Value Diff	Volume Mean	Value Mean	Percents	dates	traders	kargs
1	940691	547313	76.086129	81.438866	115715.0	1.663B	(6,65), (5,100)	20190622, 20190709	(840/840) , (438/438)	(64/64) , (33/33)
2	1253617	940411	70.017714	94.208581	100778.0	1.197B	(79,47), (8,83)	20190518, 20190608	(1080/1080) , (855/855)	(79/79) , (64/64)
	ids	Trades Value	Trades Time	Vol Percent	traders	kargs				
1	(552133, 949978, 1234156)	(7.575B, 1.506B, 1.233B)	(20190622, 20190625, 20190702)	((11, 85), (100, 18), (6, 100))	(463/463) , (877/877) , (1074/1075)	(34/34) , (66/66) , (78/78)				

شکل ۱.۴: خروجی الگوریتم برای هر نماد

نشان می‌دهد نزدیک به نصف معاملاتشان با یکدیگر بوده است و بدین صورت احتمال چرخشی بودن افزایش پیدا می‌کند.

## ۲.۴ نتایج الگوریتم سود مشکوک

در خروجی لیست کاربران و سودهای مشکوک آنها داده می‌شود. سپس ناظر می‌تواند کاربران را برحسب تعداد و اندازه سود های مشکوکی که داشتند مرتب کرده و بررسی کند. (شکل ۳.۴) از بین معامله گران مشکوک در لیست بالا، ناظر موارد مشکوک را انتخاب کرده و جزئیات سود های مشکوک آنها نمایش داده خواهد شد. (شکل ۴.۴) برای مثال برخی از معاملات مشکوک یکی از کاربران مشکوک در جدول شکل ۴.۴ آمده است.

## ۳.۴ نتایج الگوریتم معاملات همروند

در نهایت ابتدا نمودار پراکندگی سرمایه‌گذاری افراد همروند نمایش داده می‌شود. معمولاً افرادی که معامله همروند انجام داده اند، سرمایه گذاری های نزدیک بهمی انجام داده اند. سپس آمار خرید و فروش در سهم های مشترکشان در یک نمودار نقطه ای نمایش داده می‌شود. در این نمودار در صورتی که فقط یک فرد خرید یا فروش یک نماد را در تاریخ خاصی انجام دهد نقطه سبز ثبت می‌شود. هرچه تعداد بیشتری از گروه همروندی در آن روز همین معامله مشابه با او را انجام داده



باشند، دایره به سمت قرمز شدن پیش می‌رود. در نتیجه اگر تمامی افراد گروه همروندی در یک روز خرید یک نماد را انجام دهند، دایره صد درصد قرمز نمایش داده خواهد شد. (شکل ۵.۴)

## ۴.۴ نتایج مدل مقطعی

جدول ۶.۴، حاصل از بین بردن بعد زمان و ادغام دادگان بدون توجه به زمان آنها است. بدین صورت، همانند داده‌های مقطعی با آنها برخورد خواهیم کرد.

در این بررسی به این نتیجه خواهیم رسید که معاملات چرخشی، روندی معکوس با ریسک سهم دارد. این رابطه به طور ضعیف‌تری با بتا نیز دیده می‌شود که دلیل آن رابطه مثبت ریسک و بتا است. این همبستگی معکوس را بدین صورت تفسیر می‌کنیم که معاملات چرخشی باعث کاهش

	User	TotSusInv	TotSusProf	SusProfitPer	SusSucRate	TotInv	TotProf	TotSucRate	Sus/All_Prof	Sus/All_Inv	Repeated	Start	End
0	1423213	169.4B	10.3B	6	86	402.8B	18.1B	88	57	42	32	2019-04-07	2019-08-21
1	1402293	104.6B	9.0B	8	84	288.4B	15.7B	85	57	36	74	2019-04-13	2019-09-04
2	1313408	50.0B	8.0B	16	97	97.5B	8.9B	92	89	51	16	2019-05-15	2019-09-08
3	830072	73.8B	7.8B	10	86	274.3B	11.0B	79	70	26	35	2019-03-25	2019-09-16
4	1424133	132.3B	7.4B	5	85	215.8B	11.7B	86	63	61	17	2019-05-05	2019-08-21
5	602475	54.2B	7.2B	13	96	181.3B	10.2B	92	70	29	40	2019-04-10	2019-08-25
6	927325	91.8B	7.0B	7	81	307.5B	4.4B	65	157	29	13	2019-05-25	2019-08-28
7	1177064	58.5B	6.7B	11	98	266.7B	16.1B	98	41	21	31	2019-04-15	2019-09-17

شکل ۳.۴: جدول کاربران مشکوک (تجمعی بر روی سهم)

	User	buyDay	sellDay	DayDiff	InstCode	invest	Profit	ProfPercent	ProfPercentDay
0	1313408	20190515	20190520	5	266	1.0B	256.0M	25.0%	4
1	1313408	20190519	20190520	1	550	1.3B	131.0M	9.0%	4
2	1313408	20190525	20190529	4	315	1.1B	227.0M	19.0%	3
3	1313408	20190525	20190601	7	462	2.8B	621.0M	21.0%	2
4	1313408	20190601	20190603	2	266	2.7B	279.0M	10.0%	3
5	1313408	20190609	20190610	1	246	1.4B	114.0M	8.0%	4
6	1313408	20190717	20190722	5	550	3.4B	774.0M	22.0%	3
7	1313408	20190730	20190731	1	587	2.8B	286.0M	10.0%	5
8	1313408	20190730	20190731	1	614	4.6B	553.0M	11.0%	5

شکل ۴.۴: جزئیات معاملات مشکوک یک کاربر به منظور صحت سنجی

ریسک سهم شده است و اثر معناداری با باقی پارامترها ندارد. دلیل این اتفاق آن است که معاملات چرخشی برای گرم کردن سهم استفاده می‌شوند و معمولاً زمانی اتفاق می‌افتد که حجم معاملات پایین است تا ریسک شکست این نوع از تقلب را کاهش دهد. همچنین معاملات چرخشی تثبیت کننده قیمت پایانی هستند بدین معنا که قیمت پایانی را با انحراف معیار کمتر روی عدد مورد معامله ثابت می‌کنند. در حالی که بدون این نوع معاملات در نمادهای کوچک‌تر با معاملات اندکی واریانس زیادی بدست می‌آید.

همچنین معاملات با اطلاعات نهان، در زمانی رخ می‌دهند که بازده‌های بالا در آن ماه به وجود آمده است. چرا که بازده بالا به نوعی پیش‌نیاز رخ دادن این نوع تقلب است. همچنین این تقلب باعث



شکل ۵.۴: خروجی مربوط به نتیجه و صحت سنجی

افزایش ریسک معاملات شده است.

به نظر می‌رسد رابطه معناداری بین معاملات همروند و باقی شاخص‌ها پیدا نمی‌شود زیرا کمتر در بازار بورس تهران رخ داده است و اثرات آن به جز همبستگی پایین با بازده قابل چشم‌پوشی است.

## ۵.۴ نتایج مدل پنلی

در جدول شکل ۷.۴ نتیجه قابل توجهی بدست آمده است. حدود ۴۰ درصد از تغییرات متغیر وابسته توسط متغیرهای پیشگو توضیح داده شده است.<sup>۱</sup>

آماره F در این جدول معیاری برای اهمیت کلی مدل است. به طور خاص، این فرضیه صفر را

<sup>1</sup>R-squared: 0.403

	circular	front	pump
<b>return</b>	(-0.03, -0.06, -0.2)	(0.41, 0.44, 1.92)	(0.22, 0.24, 0.68)
<b>std</b>	(-0.51, -0.43, -20.65)	(0.73, 0.77, 20.64)	(0.16, 0.15, 3.05)
<b>index</b>	(-0.2, -0.21, -1.52)	(-0.21, -0.23, -1.1)	(-0.05, -0.05, -0.16)
<b>beta</b>	(-0.28, -0.31, -0.19)	(0.15, 0.15, 0.07)	(0.1, 0.14, 0.03)

شکل ۶.۴: همبستگی پیرسون، اسپیرمن و شیب خط رگرسیون (همه نمادها)

آزمایش می‌کند که همه ضرایب موجود در مدل برابر با صفر هستند، که نشان می‌دهد متغیرهای مستقل به طور قابل توجهی تغییر در متغیر وابسته را توضیح نمی‌دهند. در این مورد، آماره  $F$  برابر با  $۳۷/۱۶۷$  است که مقدار  $p$  مربوط به آن ۰ است. این بدان معناست که مدل از نظر آماری معنادار است و حداقل یکی از متغیرهای مستقل به طور معناداری با متغیر وابسته مرتبط است.

در آخر ضریب مربوط به هر کدام از متغیرهای پیشگو آمده است که در آن برای مثال ضریب بازده برابر با  $۸/۱$  است بدین معنا که هر واحد افزایش در بازده باعث رخ دادن ۸ تقلب بیشتر در حوزه اطلاعات نهانی شده است.

همچنین آماره  $t$  برای هر ضریب محاسبه شده است که نسبت ضریب برآورد شده به خطای استاندارد آن است. قدر مطلق بزرگتر آماره  $t$  نشان دهنده شواهد قوی‌تر در برابر رد فرضیه صفر است. ستون مربوط به مقدار  $p$ ، احتمال مشاهده یک آماره  $t$  را به صورت افراطی یا شدیدتر از آنچه محاسبه شده نشان می‌دهد، با فرض صحت فرضیه صفر.

به عنوان مثال، مقدار  $p$  برای برآورد ضریب بازده  $۰/۰۰۰۳$  است. این بدان معنی است که اگر ضریب واقعی در واقع صفر بود، تنها  $۰/۰۳$  درصد احتمال مشاهده آماره  $t$  شدیدتر از آنچه محاسبه شده است وجود دارد. این شواهد قوی علیه فرضیه صفر ارائه می‌دهد و نشان می‌دهد که ضریب بازده به طور قابل توجهی با صفر در سطوح معمولی از اهمیت آماری (مانند سطح معنی داری ۵٪ یا ۱٪) متفاوت است.

در مقابل، مقدار  $p$  برای برآورد ضریب بتا  $۰/۴۷۱۸$  است. این نشان می‌دهد که شواهد کافی برای

رد فرضیه صفر وجود ندارد که ضریب واقعی بتا صفر است، زیرا مقدار  $p$  بزرگتر از سطح معنی داری معمولی ۰/۰۵ است. این لزوماً به این معنا نیست که بتا یک پیش بینی کننده مفید برای متغیر وابسته نیست، بلکه به این معناست که شواهد موجود در داده‌ها به اندازه کافی قوی نیست تا نتیجه‌گیری شود که رابطه آماری معنی داری بین بتا و متغیر وابسته وجود دارد. پس در نهایت ریسک و بازده متغیرهای پیشگوی مناسبی برای تخمین میزان تقلب های انجام شده با اطلاعات نهان هستند.

PanelOLS Estimation Summary						
Dep. Variable:	front	R-squared:		0.4033		
Estimator:	PanelOLS	R-squared (Between):		0.4805		
No. Observations:	168	R-squared (Within):		0.3210		
Date:	Thu, Aug 03 2023	R-squared (Overall):		0.4033		
Time:	11:10:39	Log-likelihood		-606.32		
Cov. Estimator:	Unadjusted					
		F-statistic:		37.167		
Entities:	28	P-value		0.0000		
Avg Obs:	6.0000	Distribution:		F(3,165)		
Min Obs:	6.0000					
Max Obs:	6.0000	F-statistic (robust):		37.167		
		P-value		0.0000		
Time periods:	6	Distribution:		F(3,165)		
Avg Obs:	28.000					
Min Obs:	28.000					
Max Obs:	28.000					
Parameter Estimates						
	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
beta	-0.5460	0.7571	-0.7211	0.4718	-2.0409	0.9489
ret	8.1931	2.2001	3.7240	0.0003	3.8492	12.537
std	149.45	33.585	4.4498	0.0000	83.136	215.76

شکل ۷.۴: نتایج مدل پنلی برای پیش بینی معاملات با اطلاعات نهان

در جدول شکل ۸.۴ نیز نتایج مشابه این بار برای معاملات همروند آمده است. همانطور که مشخص است رابطه ضعیف‌تر بوده و تنها ۲۱ درصد معاملات همروند پیش بینی شده است که اکثراً وابسته به پیشگویی ریسک بوده است. زیرا فرضیه صفر برای بتا و بازده رد نشده است. با این حال آماره  $F$  نشان می‌دهد که رابطه معنادار است.

در جدول شکل ۹.۴ طبق آماره  $F$  فرض صفر رد نمی‌شود و در نتیجه رابطه معناداری برای پیش بینی معاملات چرخشی بدست نیامد.

PanelOLS Estimation Summary			
Dep. Variable:	pump	R-squared:	0.2100
Estimator:	PanelOLS	R-squared (Between):	0.3230
No. Observations:	168	R-squared (Within):	0.0042
Date:	Thu, Aug 03 2023	R-squared (Overall):	0.2100
Time:	11:10:39	Log-likelihood	-346.90
Cov. Estimator:	Unadjusted		
		F-statistic:	14.623
Entities:	28	P-value	0.0000
Avg Obs:	6.0000	Distribution:	F(3,165)
Min Obs:	6.0000		
Max Obs:	6.0000	F-statistic (robust):	14.623
		P-value	0.0000
Time periods:	6	Distribution:	F(3,165)
Avg Obs:	28.000		
Min Obs:	28.000		
Max Obs:	28.000		

Parameter Estimates						
	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
beta	0.1781	0.1616	1.1020	0.2721	-0.1410	0.4972
ret	0.2619	0.4697	0.5576	0.5779	-0.6655	1.1892
std	21.224	7.1698	2.9603	0.0035	7.0681	35.381

شکل ۸.۴: نتایج مدل پنلی برای پیش بینی معاملات همروند

PanelOLS Estimation Summary			
Dep. Variable:	circular	R-squared:	0.0058
Estimator:	PanelOLS	R-squared (Between):	0.0089
No. Observations:	168	R-squared (Within):	-0.0004
Date:	Thu, Aug 03 2023	R-squared (Overall):	0.0058
Time:	11:10:39	Log-likelihood	-844.47
Cov. Estimator:	Unadjusted		
		F-statistic:	0.3204
Entities:	28	P-value	0.8106
Avg Obs:	6.0000	Distribution:	F(3,165)
Min Obs:	6.0000		
Max Obs:	6.0000	F-statistic (robust):	0.3204
		P-value	0.8106
Time periods:	6	Distribution:	F(3,165)
Avg Obs:	28.000		
Min Obs:	28.000		
Max Obs:	28.000		

Parameter Estimates						
	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
beta	-0.2851	3.1246	-0.0912	0.9274	-6.4543	5.8842
ret	0.5000	9.0795	0.0551	0.9561	-17.427	18.427
std	83.807	138.60	0.6047	0.5462	-189.86	357.47

شکل ۹.۴: نتایج مدل پنلی برای پیش بینی معاملات چرخشی

## واژه‌نامه فارسی به انگلیسی

threshold	آستانه
stock market	بازار سرمایه
market maker	بازارگردان
return	بازده
stock market	بورس
anomaly detection	تشخیص ناهنجاری
matching	تطابق قیمتی
fraud	تقلب
sparse	تنک
depth-first search (DFS)	جستجوی اول عمق
cycle	حلقه (گراف)
clustering	خوشه‌بندی
price manipulation	دستکاری قیمت
vertex	راس (گراف)
portfolio manager	سبدگردان
time series	سری زمانی
stock	سهم
supply and demand	عرضه و تقاضا



bid-ask spread.....	فاصله سر خط خرید و فروش در لیست سفارشات
OTC.....	فرا بورس
order-book.....	لیست سفارشات
false positive.....	مثبت کاذب
finance.....	مدیریت مالی
insider trading.....	معاملات با استفاده از اطلاعات نهان
circular trading.....	معاملات چرخشی
front running.....	معامله پیش از معاملات بزرگ بازار
institutional trader.....	معامله گر حقوقی
retail trader.....	معامله گر حقیقی
concurrent trading.....	معاملات همروند
false negative.....	منفی کاذب
surveillance.....	نظارت
liquidity.....	نقدشوندگی
correlation.....	همبستگی
GPU.....	پردازنده گرافیکی
time complexity.....	پیچیدگی زمانی
machine learning.....	یادگیری ماشین
edge.....	یال (گراف)

## واژه‌نامه انگلیسی به فارسی

anomaly detection	تشخیص ناهنجاری
bid-ask spread	فاصله سر خط خرید و فروش در لیست سفارشات
circular trading	معاملات چرخشی
clustering	خوشه‌بندی
concurrent trading	معاملات همروند
correlation	همبستگی
cycle	حلقه (گراف)
depth-first search (DFS)	جستجوی اول عمق
edge	یال (گراف)
false negative	منفی کاذب
false positive	مثبت کاذب
finance	مدیریت مالی
fraud	تقلب
front running	معامله پیش از معاملات بزرگ بازار
GPU	پردازنده گرافیکی
insider trading	معاملات با استفاده از اطلاعات نهان
institutional trader	معامله‌گر حقوقی
liquidity	نقدشوندگی

machine learning	یادگیری ماشین
market maker	بازارگردان
matching	تطابق قیمتی
order-book	لیست سفارشات
OTC	فرا بورس
portfolio manager	سبدگردان
price manipulation	دستکاری قیمت
retail trader	معامله گر حقیقی
return	بازده
sparse	تنک
stock market	بازار سرمایه
stock market	بورس
stock	سهم
supply and demand	عرضه و تقاضا
surveillance	نظارت
threshold	آستانه
time complexity	پیچیدگی زمانی
time series	سری زمانی
vertex	راس (گراف)

## کتابنامه

- [۱] ریحانه ربیعی، محمد ندری، مسلم پیمانی، و علی جابری زاده، (۱۳۹۷) “بررسی تأثیر دستکاری قیمت بر کارایی بازار در بورس اوراق بهادار تهران”، بورس اوراق بهادار، vol. ۱۱، no. ۴۳، pp. ۱۱۳-۱۳۱.
- [2] Gerace, D., Chew, C., Whittaker, C. and Mazzola, P. (2014) ‘Stock Market Manipulation on the Hong Kong Stock Exchange’, Australasian Accounting, Business and Finance Journal, 8, pp. 105-140.
- [3] James, R., Leung, H. and Prokhorov, A. (2020) ‘A Machine Learning Attack on Illegal Trading’, SSRN Electronic Journal, doi: 10.2139/ssrn.3722391.
- [4] Maxim, M. and Asm, A. (2017) ‘A new method of measuring stock market manipulation through structural equation modeling (SEM)’, Investment Management and Financial Innovations, 14, pp. 1-10.
- [۵] میرفیض فلاح شمس لیاستانی، مهرزاد مینویی، ایوب بادپا، (۱۳۹۱) طراحی مدلی برای شناسایی احتمال وقوع دستکاری قیمت پایانی در بورس اوراق بهادار تهران.

- [۶] پوست فروش محمدحسین، ناصر صدرآبادی علیرضا، معین الدین محمود، (۱۳۹۵) کشف دستکاری قیمت سهام با استفاده از تحلیل ممیزی خطی و تحلیل ممیزی درجه دوم. نشریه: دانش مالی تحلیل اوراق بهادار (مطالعات مالی)، دوره: ۹، شماره: ۲۹. صفحات: ۱۷-۳۹
- [7] Comerton-Forde, C. and Putniņš, T.J. (2011) 'Measuring closing price manipulation', *Journal of Financial Intermediation*, 20(2), pp. 135-158.
- [8] Huang, Y.-C., Chen, R. and Cheng, Y. (2005) 'Stock manipulation and its impact on market quality'.
- [9] Priya, J.M., Kumar, K.S., Babu, C.S. and Rao, S.K.V. (2017) 'A graph theoretical approach for identifying fraudulent transactions in circular trading', *DATA ANALYTICS 2017*, pp. 36-42.
- [10] Palshikar, G.K. and Apte, M.M. (2008) 'Collusion set detection using graph clustering', *Data Mining and Knowledge Discovery*, 16, pp. 135-164.
- [11] Park, Y.S. and Lee, J. (2010) 'Detecting insider trading: The theory and validation in Korea Exchange', *Journal of Banking & Finance*, 34(9), pp. 2110-2120.
- [12] Donoho, S. (2004) 'Early detection of insider trading in option markets', in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 420-429.
- [13] Nghiem, H., Muric, G., Morstatter, F. and Ferrara, E. (2021) 'Detecting cryptocurrency pump-and-dump frauds using market and social signals', *Expert Systems with Applications*, 182, p. 115284.

- [14] Victor, F. and Hagemann, T. (2019) ‘Cryptocurrency pump and dump schemes: Quantification and detection’, in 2019 International Conference on Data Mining Workshops (ICDMW), pp. 244-251.

## **Abstract**

Capital market is an essential part of modern economies. The prosperity of financial markets and high-frequency transactions have caused the increase and complexity of trading violations. Manipulation of financial markets can lead to price deviations or increase volatility and affect investors and the entire market. Despite the efforts of regulatory bodies to control financial market manipulation, there is still fraudulent activity in the market in various forms such as circular trading, pump and dump, insider trading and other methods. On the other hand, machine learning has become a popular tool in financial markets to identify patterns and anomalies.

In this research, we first introduce 3 risk-based algorithms to detect suspected fraudulent transactions, which are implemented in six months and on more than six hundred OTC symbols. Then we measure the effects of frauds on the market and share.



College of Science

School of Mathematics, Statistics, and Computer Science

# Investigating the amount of manipulation in the capital market using machine learning methods

**Mahyar Mohammadi**

Supervisor: Samane Eftekhari

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

2023