

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده مهندسی صنایع و مدیریت
پایان نامه کارشناسی ارشد مدیریت صنعتی

خوشه بندی سهام با توجه به شاخص های نقدشوندگی براساس ضریب تشابه تطابق نسبی

نگارنده: نگین السادات احمدی

استاد راهنما

دکتر رضا شیخ

استاد مشاور

دکتر مجید عامری

بهمن ۱۳۹۹

شماره: ۱۴-۹۹-۶۹۳۶
تاریخ: ۹۹/۱۲/۱۳

باسمه تعالی



مدیریت تحصیلات تکمیلی

فرم شماره (۳) صورتجلسه نهایی دفاع از پایان نامه دوره کارشناسی ارشد

با نام و یاد خداوند متعال، ارزبایی جلسه دفاع از پایان نامه کارشناسی ارشد خانم / آقای احمدی نگین السادات با شماره دانشجویی ۹۷۰۱۹۴۴ رشته مدیریت صنعتی-تولید و عملیات تحت عنوان خوشه بندی سهام با توجه به شاخص های نقد شوندگی بر اساس ضریب تشابه تطابق نسبی که در تاریخ ۲۸/۱۱/۱۳۹۹ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح ذیل اعلام می گردد:

قبول (با درجه:): مردود
نوع تحقیق: نظری عملی

| عضو هیأت داوران | نام و نام خانوادگی | مرتبه علمی | امضاء |
|---------------------------|----------------------------------|------------|-------|
| ۱- استاد راهنمای اول | دکتر رضا شیخ | | |
| ۲- استاد راهنمای دوم | | | |
| ۳- استاد مشاور | دکتر مجید عامری | | |
| ۴- نماینده تحصیلات تکمیلی | دکتر محمد رستمی | | |
| ۵- استاد ممتحن اول | دکتر عبدالمجید عبدالباقی | استادی | |
| ۶- استاد ممتحن دوم | عطا آبادی دکتر علی اکبر حسینی | | |

نام و نام خانوادگی رئیس دانشکده:

تاریخ و امضاء و مهر دانشکده:



تصوه: در صورتی که کسی مردود شود حداکثر یکبار دیگر (در مدت مجاز تحصیل) می تواند از پایان نامه خود دفاع نماید (دفاع مجدد نباید زودتر از ۴ ماه برگزار شود).

تقدیم اثر

این پایان‌نامه را ضمن تشکر و سپاس بیکران و در کمال افتخار و امتنان تقدیم می‌نمایم به محضر ارزشمند پدر و مادر عزیزم به خاطر همه‌ی تلاش‌های محبت‌آمیزشان و همچنین استادان فرزانه و فرهیخته‌ای که در راه کسب علم و معرفت مرا یاری نمودند. از پروردگار مهربان تقاضا دارم به همه ما توفیق خدمتی سرشار از شور و نشاط و همراه و همسو با علم و دانش و پژوهش جهت رشد و شکوفایی ایران سربلند عنایت فرماید.

شکر و قدردانی

از استاد فاضل و اندیشمند جناب آقای دکتر رضا شیخ به عنوان استاد راهنما که همواره اینجانب را مورد لطف و محبت خود قرار داده، کمال تشکر را دارم. همچنین از جناب آقای مجید عامری که با راهنمایی‌هایشان به این جانب در اتمام این پایان‌نامه یاری رساندند، سپاس گزارم.

تعمدنامه

اینجانب نگین السادات احمدی دانشجوی دوره کارشناسی ارشد رشته مدیریت صنعتی دانشکده مهندسی صنایع و مدیریت دانشگاه صنعتی شاهرود نویسنده پایان نامه کارشناسی ارشد تحت راهنمایی دکتر رضا شیخ متعهد می شوم.

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهش های محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « **Shahrood University of Technology** » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود . استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

حکیده

پیش‌بینی بازار سهام با استفاده از روش‌های داده‌کاوی یکی از موضوعات مهم در سرمایه‌گذاری و در تحقیقات بازارهای مالی است. برای پیش‌بینی این بازار تلاش‌های بسیاری با استفاده از روش‌های سنتی صورت گرفته، اما با توجه به افزایش روزافزون حجم اطلاعات، این روش‌ها برای تحلیل این میزان از اطلاعات مناسب نیستند. داده‌کاوی قادر به کشف الگوهای پنهان و پیش‌بینی روندهای آتی در بازار سهام است. بازار سهام را می‌توان یک مسئله داده‌کاوی در نظر گرفت و خوشه‌بندی یکی از روش‌های داده‌کاوی بوده که یک استراتژی مناسب جهت پیش‌بینی و هدایت تصمیم‌های سرمایه‌گذاران است. نقدینگی سهام به معنای قابلیت خرید و فروش سهام در کوتاه‌ترین زمان و با حداقل هزینه می‌باشد. لذا سرمایه‌گذاران شرکت‌هایی را که دارای نقدشوندگی بالایی باشند، انتخاب می‌کنند. در این پژوهش خوشه‌بندی شرکت‌های فعال در بازار بورس با توجه به شاخص‌های نقدشوندگی براساس ضریب تشابه تطابق نسبی انجام شده است. به این منظور ۴۲ شرکت پذیرفته شده در بورس اوراق بهادار طی سال‌های ۱۳۹۷-۱۳۹۸ بر مبنای شاخص‌های نقدشوندگی منتخب شامل گردش سهام، اختلاف قیمت پیشنهادی خرید و فروش، معیار ژانگ، اهرم مالی، عملکرد شرکت، ضریب انعطاف‌پذیری تجارت، اختلاف قیمتی موثر، تعدیل تعداد روزهای بدون معامله براساس گردش، تعداد معاملات، حجم معاملات، درصد روزهای معاملاتی و اندازه شرکت براساس الگوریتم پیوند کامل و ضریب تشابه تطابق نسبی خوشه‌بندی شدند. نتایج این بررسی نشان می‌دهد می‌توان از شاخص‌های تعداد معاملات و حجم معاملات برای پیش‌بینی نقدشوندگی استفاده کرد.

کلمات کلیدی: خوشه‌بندی، ضریب تشابه، نقدشوندگی، بورس اوراق بهادار تهران

فهرست مطالب

| | |
|----|----------------------------------|
| ۱ | فصل ۱: کلیات |
| ۲ | ۱-۱ مقدمه |
| ۲ | ۲-۱ بیان موضوع تحقیق |
| ۴ | ۳-۱ ضرورت و اهمیت پژوهش |
| ۵ | ۴-۱ نوآوری پژوهش |
| ۶ | ۵-۱ سوال پژوهش |
| ۶ | ۶-۱ هدف پژوهش |
| ۶ | ۷-۱ نوع تحقیق |
| ۶ | ۸-۱ فرضیه‌های تحقیق |
| ۷ | ۹-۱ روش پژوهش |
| ۷ | ۱۰-۱ قلمرو تحقیق |
| ۷ | ۱-۱۰-۱ قلمرو موضوعی |
| ۸ | ۲-۱۰-۱ قلمرو مکانی |
| ۸ | ۳-۱۰-۱ قلمرو زمانی |
| ۸ | ۱۱-۱ تعریف مختصر واژه‌های کلیدی |
| ۹ | ۱۲-۱ خلاصه و جمع‌بندی فصل |
| ۱۱ | فصل ۲: مبانی نظری و ادبیات تحقیق |
| ۱۲ | ۱-۲ مقدمه |
| ۱۳ | ۲-۲ مفاهیم بنیادی |

۱۳..... ۲-۲-۱ بازارهای مالی و طبقه‌بندی آنها

۱۴..... ۲-۲-۲ بازار مالی

۱۴..... ۳-۲-۲ نقش بازار مالی

۱۵..... ۴-۲-۲ تقسیم بندی بازارهای مالی

۱۵..... ۵-۲-۲ بازار سرمایه

۱۶..... ۶-۲-۲ بازار پول

۱۷..... ۷-۲-۲ بازار اولیه و ثانویه

۱۸..... ۸-۲-۲ بازارهای خارج از بورس و بازارهای سازمان یافته (بورس)

۱۹..... ۹-۲-۲ بازار بورس

۲۰..... ۱۰-۲-۲ تعریف سهام و انواع آن

۲۲..... ۱۱-۲-۲ نقدشوندگی

۲۳..... ۲-۲-۱۲ داده کاوی

۲۷..... ۱۳-۲-۲ طبقه‌بندی

۲۹..... ۲-۲-۱۴ خوشه‌بندی سلسله‌مراتبی

۳۱..... ۱۵-۲-۲ پیشینه تحقیق

۳۷..... ۱۶-۲-۲ خارجی

۴۳..... ۲-۲-۱۷ خلاصه و جمع‌بندی فصل

۴۵ فصل ۳ متدولوژی تحقیق

۴۶..... ۳-۱ مقدمه

۴۶..... ۳-۲ فرایند خوشه‌بندی

۷۷..... ۳-۳ خلاصه و جمع‌بندی فصل

فصل ۴: تجزیه و تحلیل داده‌ها

۷۹

۱-۴ مقدمه ۸۰

۲-۴ جامعه و نمونه آماری ۸۰

۱-۲-۴ روش گردآوری داده‌ها ۸۲

۲-۲-۴ الگوریتم پیاده‌سازی پژوهش ۸۳

۳-۲-۴ الگوریتم خوشه‌بندی ۸۷

۳-۴ ارزشیابی خوشه‌بندی ۹۱

۴-۴ خلاصه و جمع‌بندی فصل ۹۶

فصل ۵ نتیجه‌گیری و پیشنهادات آتی

۹۷

۱-۵ مقدمه ۹۸

۲-۵ خلاصه پژوهش ۹۸

۳-۵ نتایج پژوهش ۹۹

۴-۵ محدودیت‌های پژوهش ۱۰۰

۵-۵ پیشنهادهایی برای پژوهش‌های آتی ۱۰۰

۶-۵ خلاصه و جمع‌بندی فصل ۱۰۱

۱۰۲

مراجع

فهرست جداول

- جدول ۱-۳. شاخص‌های نقدشوندگی مورد استفاده در پژوهش ۵۹
- جدول ۲-۳. روابط باینری بین جفت اشیا ۶۱
- جدول ۳-۳. تجزیه و تحلیل ضرایب تشابه ویژگی‌ها S_{ij} مطابق با هر ماتریس ۶۳
- جدول ۴-۳. تحلیل واریانس ۷۵
- جدول ۱-۴. نمونه آماری پژوهش ۸۲
- جدول ۲-۴. نتیجه محاسبات شاخص‌های نقدشوندگی سال ۱۳۹۷ ۸۳
- جدول ۳-۴. نتیجه محاسبات شاخص‌های نقدشوندگی سال ۱۳۹۸ ۸۴
- جدول ۴-۴. نتیجه محاسبات شاخص‌های نقدشوندگی سال ۱۳۹۷ ۸۵
- جدول ۵-۴. نتیجه محاسبات شاخص‌های نقدشوندگی سال ۱۳۹۸ ۸۶
- جدول ۶-۴. ماتریس وقوع شاخص‌های سال ۱۳۹۷ ۸۸
- جدول ۷-۴. ماتریس وقوع شاخص‌های سال ۱۳۹۸ ۸۹
- جدول ۸-۴. نتیجه خوشه‌بندی نمونه آماری سال ۱۳۹۷ ۹۰
- جدول ۹-۴. نتیجه خوشه‌بندی نمونه آماری سال ۱۳۹۸ ۹۱
- جدول ۱۰-۴. نتایج ارزشیابی خوشه‌بندی براساس شاخص سیلوئت ۹۱
- جدول ۱۱-۴. درصد هر شاخص در هر خوشه سال ۱۳۹۷ ۹۲
- جدول ۱۲-۴. درصد هر شاخص در هر خوشه سال ۱۳۹۸ ۹۲
- جدول ۱۳-۴. نتایج تحلیل واریانس برای شاخص‌های نقدشوندگی ۹۲
- جدول ۱۴-۴. نتایج اختلاف میانگین شاخص‌های نقدشوندگی سال ۱۳۹۷ ۹۳
- جدول ۱۵-۴. نتایج اختلاف میانگین شاخص‌های نقدشوندگی سال ۱۳۹۸ ۹۴
- جدول ۱۶-۴. شاخص‌هایی که فرض صفر برای آنها رد شده است. ۹۵
- جدول ۱۷-۴. شاخص‌هایی که فرض صفر برای آنها رد شده است. ۹۵

فهرست اشکال

شکل ۳-۱. مراحل فرایند خوشه‌بندی ۴۸

شکل ۳-۲. افراز تغییرات به چند منبع تغییر ۷۲

فصل ۱ : کلیات پژوهش

۱-۱ مقدمه

بازارهای متعددی برای سرمایه‌گذاری وجود دارد. اکثر سرمایه‌گذاری‌ها در کشورهای توسعه یافته از طریق بازارهای مالی انجام می‌شود. بازار سهام یکی از محبوب‌ترین اشکال سرمایه‌گذاری است. گفته می‌شود نقدشوندگی مایه حیات بازار سهام است و پیامدهای برجسته‌ای برای معامله‌گران، تنظیم‌کنندگان و شرکت‌های فعال در بازار بورس دارد. نقدینگی مفهوم پیچیده‌ای است که آن را به یکی از عناصر اصلی‌ای تبدیل می‌کند که سرمایه‌گذاران تصمیم به سرمایه‌گذاری یا عدم سرمایه‌گذاری می‌گیرند.

با توجه به رشد روز افزون حجم اطلاعات بشر به فکر دستیابی به اطلاعات نهان در پایگاه داده‌های حجیم افتاد و چون روش‌های سنتی قادر به پاسخگویی به این نیاز نبودند، داده‌کاوی به عنوان فرایندی کارا مورد استفاده قرار گرفته است. داده‌کاوی با نشان دادن جزئیات بیشتر فضای حاکم بر کسب و کار را شفاف‌تر کرده و به افراد کمک می‌کند تا تصمیم‌گیری واقع‌بینانه داشته باشند. داده‌کاوی در بازارهای مالی و بورس که حجم عظیمی از داده‌ها به صورت روزانه تولید می‌شوند، کاربرد زیادی دارد. داده‌کاوی دارای تکنیک‌های مختلفی می‌باشد. یکی از این تکنیک‌ها، خوشه‌بندی است. می‌توان با خوشه‌بندی شرکت‌ها سرمایه‌گذاران در تصمیم‌گیری‌ها کمک کرد. لذا در این پژوهش سعی می‌شود شرکت‌های مورد مطالعه در بازار بورس سهام با توجه به شاخص‌های نقدشوندگی خوشه‌بندی شوند.

ساختار فصل اول پژوهش به این صورت است: ابتدا مسئله مورد نظر و ضرورت پرداختن به آن بیان می‌شود. پس از آن اهمیت، روش، قلمرو و نوآوری تحقیق و در ادامه تعاریف مختصری از واژه‌های کلیدی بیان خواهد شد.

۱-۲ بیان موضوع تحقیق

نقدینگی به معنای تسهیل، تسریع و کاهش هزینه در فرایند تبدیل دارایی مالی به وجه نقد و برعکس یعنی تبدیل وجه نقد به دارایی مالی است که یکی از کارکردهای مهم بازارهای مالی و به‌ویژه بورس اوراق

بهدار می‌باشد [۱]. میزان نقدشوندگی بالا نشان از شفاف‌سازی اطلاعات و نزدیکی قیمت اوراق بهادر به ارزش ذاتی آن‌ها دارد. هر مقدار نقدشوندگی سهام بالاتر باشد، ریسک نگهداری آن کمتر است، زیرا سرعت تبدیل آن به وجه نقد بیشتر است. بنابراین تعدادی از سرمایه‌گذاران به دلیل داشتن ماهیت ریسک‌گریزی سعی در انتخاب سهامی دارند که قابلیت نقدشوندگی بالایی داشته باشد تا در مواقع لزوم، فروش سهام را با سرعت بالا و در زمان کوتاه و با کمترین تغییر قیمت انجام دهند [۲]. ریسک عدم نقدشوندگی برای سرمایه‌گذار در زمان تصمیم به فروش دارایی کاهش می‌یابد، اگر وی اطمینان داشته باشد که همیشه طی مدت زمان کوتاهی قادر خواهد بود دارایی خود را به پول نقد تبدیل کند و بازار مناسبی برای فروش دارایی وجود دارد و این موجب اطمینان خاطر وی از سرمایه‌گذاری خواهد شد [۳].

با توجه به پیشرفت فناوری اطلاعات و شیوه‌های گوناگون برای تولید و جمع‌آوری داده‌ها، اطلاعات مربوط به مبادلات تجاری و کسب و کار سریع‌تر از هر زمان دیگری جمع‌آوری می‌شوند. مدیریت داده‌های خام و تبدیل داده‌های خارجی شرکت‌ها به دانش و اطلاعات با استفاده از روش‌های گوناگون، نقش اساسی و محوری دارد. یکی از این روش‌ها داده کاوی است که در دهه ۹۰ مطرح گردید و به عنوان یک شاخه تحقیقاتی جدید اهمیت آن روز به روز افزایش می‌یابد. داده‌کاوی، علم و فن‌آوری کاوش داده‌ها برای کشف الگوهای ناشناخته و یکی از مراحل فرایند کشف دانش در پایگاه داده‌ها است [۴].

در بازارهای سهام مقدار زیادی داده تولید می‌شود و باعث رویارویی سرمایه‌گذاران با حجم وسیعی از اطلاعات و داده‌ها می‌شود که به تنهایی به آنان در تصمیم‌گیری کمک نمی‌کند و باعث سردرگمی آنها خواهد شد

روش‌های مختلفی برای کمک به سرمایه‌گذاران در تصمیم‌گیری‌های مربوطه پیشنهاد شده است. خوشه‌بندی از تکنیک‌های داده‌کاوی می‌باشد که سرمایه‌گذاران می‌توانند در بهینه‌سازی سبد سهام از آن بهره ببرند. خوشه‌بندی یکی از انواع مدل‌سازی داده است که برای استخراج اطلاعات ارزشمند از

مجموعه‌ای از داده‌ها بکار می‌رود. در این پژوهش سعی بر آن است تا با در نظر گرفتن شاخص‌هایی برای نقدشوندگی سهام شرکت‌های مورد مطالعه در بازار بورس تهران، الگوریتم خوشه‌بندی پیاده‌سازی شود.

۱-۳ ضرورت و اهمیت پژوهش

بخش مهمی از اقتصاد کشور را سرمایه‌گذاری در بورس تشکیل می‌دهد و در سرتاسر جهان بیشترین میزان سرمایه از طریق بازارهای سهام مبادله می‌شود و اقتصاد ملی به شدت متأثر از عملکرد بازار بورس است. با توجه به اینکه بورس اوراق بهادار تهران، سازمانی در تجهیز پس‌اندازها و هدایت آن به سمت سرمایه‌گذاری مولد و مفید به حال جامعه و اقتصاد کشور است، مطالعه پیرامون موضوعات وابسته به این سازمان از اهمیت ویژه‌ای برخوردار است.

شواهد تجربی نشان می‌دهد در تصمیم‌گیری‌ها عامل نقدشوندگی می‌تواند نقش مهمی را ایفا کند. به عبارتی برخی سرمایه‌گذاران ممکن است به سرعت به منابع مالی سرمایه‌گذاری خود نیاز داشته باشند که در چنین مواردی قدرت نقدشوندگی می‌تواند اهمیت زیادی داشته باشد [۵]. بنابراین یکی از موضوعات مهم در ارتباط با سازمان بورس اوراق بهادار نقدشوندگی سهام شرکت‌های پذیرفته شده در این سازمان است.

با افزایش جهانی‌سازی اقتصاد و تکامل فناوری اطلاعات، داده‌های مالی با سرعت بی‌سابقه‌ای تولید و انباشته می‌شوند. در نتیجه نیاز جدی به رویکردهای خودکار برای استفاده موثر و کارآمد از انبوه داده‌های مالی برای حمایت از افراد و موسسه‌های مالی در برنامه‌ریزی استراتژیک و تصمیم‌گیری جهت سرمایه‌گذاری احساس می‌شود. برای این منظور از تکنیک‌های داده‌کاوی برای کشف الگوهای پنهان و پیش‌بینی روندها و رفتارهای آینده در بازارهای مالی استفاده می‌شود.

در این پژوهش خوشه‌بندی سهام شرکت‌های مورد مطالعه در بازار بورس اوراق بهادار تهران با توجه به شاخص‌های نقدشوندگی انجام می‌شود. خوشه‌بندی بازار سهام اطلاعات مفیدی را در مورد نقدشوندگی

سهام شرکت‌های مختلف در اختیار سرمایه‌گذاران و متخصصان مالی قرار می‌دهد و استفاده از این الگوریتم برای ساخت پرتفوی مالی می‌تواند پرتفولویی قابل اعتمادتر نسبت به زمانی که خوشه‌بندی انجام نمی‌شود ارائه دهد و همچنین خوشه‌بندی باعث کاهش زمانی می‌شود که سرمایه‌گذاران قرار است برای پرتفوهای کارآمد با توجه به سهام‌های مشابه‌ای که در خوشه‌ها گروه‌بندی شده‌اند، صرف کنند.

به طور کلی، یافتن شرکت‌هایی که در طول زمان دارای رفتاری مشابه در بازار سهام هستند برای تحلیل‌گران بسیار مفید است. استفاده از تکنیک خوشه‌بندی یک استراتژی مناسب برای این منظور است.

۱-۴ نوآوری پژوهش

بازار سهام یا بورس اوراق بهادار شامل مجموعه‌ای از خریداران و فروشندگان است که به مبادله سهام شرکت‌ها و دیگر اوراق بهادار می‌پردازند. در طول سال‌ها مشاهده شده است بسیاری از افرادی که در این بازار فعالیت دارند، متحمل خسارت زیادی شده‌اند. مطالعه پیرامون موضوع‌های وابسته به این سازمان از اهمیت ویژه‌ای برخوردار است. از این رو نیاز به یک سیستم پیش‌بینی قابل اعتماد احساس می‌شود. با این حال نوسانات بازار سهام باعث می‌شود که این نوع پیش‌بینی‌ها بسیار دشوار باشد. بازار سهام را می‌توان به عنوان یک مسئله داده‌کاوی جهت هدایت تصمیم سرمایه‌گذاران استفاده کرد.

تحقیقات متعددی در رابطه با نقدشوندگی در بازار سهام و کاربرد خوشه‌بندی در این بازار انجام شده است اما تاکنون پژوهشی در رابطه با خوشه‌بندی شرکت‌های فعال در بازار بورس با توجه به شاخص‌های نقدشوندگی و براساس ضریب تشابه نسبی انجام نشده است و لذا تحقیق حاضر از این حیث دارای نوآوری می‌باشد.

۱-۵ سوال پژوهش

چگونه می‌توان خوشه‌بندی سهام را با توجه به شاخص‌های نقدشوندگی و براساس ضریب تشابه تطابق نسبی انجام داد؟

۱-۶ هدف پژوهش

مهمترین هدف پژوهش خوشه‌بندی شرکت‌های فعال در بورس اوراق بهادار تهران با توجه به شاخص‌های نقدشوندگی و براساس ضریب تشابه نسبی است.

۱-۷ نوع تحقیق

با توجه به ماهیت موضوع و هدف آن، تحقیق پیش‌رو از نوع تحقیقات کاربردی می‌باشد، زیرا نتایج حاصل از آن در تصمیم‌گیری‌های تخصیص بهینه مالی سرمایه‌گذاران نقش مهمی ایفا می‌کند. با توجه به کمی‌سازی نقدشوندگی سهام با استفاده از شاخص‌های نقدشوندگی منتخب و انجام آزمون‌های آماری برای تحلیل داده‌ها، این پژوهش از نوع پژوهش‌های کمی می‌باشد.

۱-۸ فرضیه‌های تحقیق

فرضیه تحقیق حاضر به صورت زیر قابل بیان است:

شاخص‌های نقدشوندگی بین خوشه‌های ایجاد شده توسط تکنیک خوشه‌بندی پیوند کامل دارای تفاوت هستند.

۹-۱ روش پژوهش

مراحل تحقیق حاضر را می‌توان به صورت زیر خلاصه کرد:

الف) بررسی ادبیات تحقیق در مورد نقدشوندگی سهام، انواع شاخص‌های نقدشوندگی و انتخاب شاخص‌های نقدشوندگی

ب) بررسی ادبیات تحقیق در زمینه داده‌کاوی و خوشه‌بندی، انواع روش‌های خوشه‌بندی و نحوه اجرا و ارزیابی الگوریتم

ج) انتخاب نمونه آماری از شرکت‌های فعال در بازار بورس اوراق بهادار تهران

د) جمع‌آوری داده‌ها از پایگاه‌های داده مختلف، آماده‌سازی داده‌های گردآوری شده و طراحی جدول‌های داده با استفاده از نرم‌افزار Excel

ه) اجرای الگوریتم خوشه‌بندی، ارزیابی و تفسیر نتایج

۱۰-۱ قلمرو تحقیق

هر پژوهشی باید دارای قلمرو و دامنه مشخصی باشد تا پژوهشگر در همه مراحل پژوهش، بر موضوع اسط کافی داشته باشد و بتواند نتایج حاصل از نمونه انتخابی را به جامعه تعمیم دهد. این پژوهش نیز مانند سایر پژوهش‌ها دارای دامنه‌های زیر می‌باشد:

۱-۱۰-۱ قلمرو موضوعی

از لحاظ تکنیک به کار گرفته شده جهت تحلیل جزء روش‌های کمی و در حوزه تکنیک‌های داده‌کاوی است و همچنین در حوزه کاربردی مرتبط با مدیریت مالی و بازار سهام می‌باشد.

۱-۱۰-۲ قلمرو مکانی

قلمرو مکانی تحقیق حاضر بورس اوراق بهادار تهران و شرکت‌های فعال در آن است.

۱-۱۰-۳ قلمرو زمانی

محدوده زمانی مورد بررسی در این پژوهش اطلاعات بورس اوراق بهادار تهران طی سال‌های ۱۳۹۸-۱۳۹۷ می‌باشد.

۱-۱۱ تعریف مختصر واژه‌های کلیدی

نقدشوندگی: نقدشوندگی به صورت توانایی تبدیل دارایی مالی به وجه نقد تعریف می‌شود و اشاره به انجام معاملات به سرعت، با هزینه کم و بدون تأثیر قرار دادن شدید قیمت دارد.

خوشه‌بندی: خوشه‌بندی گروه‌بندی نمونه‌ها یا مشاهدات درون کلاس‌هایی از اشیاء مشابه است. یک خوشه مجموعه‌ای از نمونه‌هاست که شبیه به یکدیگر و بی شباهت از نمونه‌های خوشه‌های دیگر است.

ضریب تشابه: ضرایب تشابه برای تعیین میزان شباهت بین جفت اشیاء بکار می‌روند که اکثر ضرایب تشابه به صورت نسبت بیان می‌شوند. در میان ضرایب تشابه دو نوع جاکردی و غیر جاکردی وجود دارد. در قالب جاکردی، ضریب شباهت به عنوان معیار سطحی تطابق‌ها بیان می‌شود که در آن تعداد تطابق‌ها بر تعداد تطابق‌های مورد انتظار تقسیم می‌شود. در قالب غیر جاکردی یک اصطلاح اضافی نیز وجود دارد (تعداد خطاها)، که از صورت کسر کم شده و بر تعداد تطابق‌های مورد انتظار تقسیم می‌شود. در قالب جاکردی ضریب تشابه بین ۰ تا ۱ است درحالی‌که در غیر جاکردی بین ۱- تا ۱ می‌باشد.

۱-۱۲ خلاصه و جمع‌بندی فصل

در فصل حاضر به تشریح مسئله پژوهش و ضرورت آن که نتیجه انجام پژوهش به دنبال آن می‌باشد، پرداخته شد. سپس محدوده انجام پژوهش از لحاظ موضوعی، مکانی و زمانی مشخص گردید. سوال پژوهش تدوین و در ادامه اصطلاحات و واژگان کلیدی تعریف شد.

پایان نامه پیش رو شامل پنج فصل است که مرور کلی فصل‌های بعدی آن به شرح زیر است:

در فصل دوم با بیان مفاهیمی چون بازار، بورس، نقدشوندگی، خوشه‌بندی و انواع آن به ادبیات پیشینه پژوهش اشاره خواهد شد. در فصل سوم اطلاعات مربوط به روش انجام این پژوهش توضیح داده خواهد شد. در فصل چهارم به بیان تجزیه و تحلیل داده‌ها پرداخته خواهد شد. در نهایت در قالب فصل پنجم محدودیت‌هایی که پژوهشگر با آنها مواجه بود به همراه پیشنهادهایی که برای پژوهش‌های آتی میتواند مفید باشد، مطرح خواهد گردید.

فصل ۲ مبانی نظری و ادبیات تحقیق

۱-۲ مقدمه

در دنیای معاصر پایگاه داده حجم وسیعی از اطلاعات را تشکیل می‌دهد. بنابراین دسترسی به این اطلاعات، داده کاوی را به یک مسئله ضروری و برجسته تبدیل کرده است. داده کاوی یک علم و فناوری بررسی داده‌ها برای یافتن الگوهای کشف نشده را ایجاد کرده است. نهادهای مالی مانند بازارهای سهام، مجموعه داده‌های انبوهی ایجاد می‌کنند که نیاز به ابزارهای پویا و چندوجهی داده کاوی را به وجود می‌آورد. مطالعات مربوط به داده کاوی به خاطر اهمیت کاربرد آنها و تولید اطلاعات توجه زیادی را به خود اختصاص داده است. امروزه، تجزیه و تحلیل داده های مالی به طور فزاینده ای در بازار کسب و کار اهمیت پیدا کرده است. از آنجا که شرکت ها به صورت روز افزون از عملیات روزانه، داده جمع‌آوری می کنند، انتظار دارند که از اطلاعات جمع آوری شده موجود برای کمک به تصمیم گیری معقول دانش مفیدی بگیرند. موسسات بانکی و مالی روش های مختلف داده کاوی را برای ارتقا عملکرد تجاری خود اعمال کرده اند. در میان این تکنیک ها، خوشه بندی به عنوان یک روش قابل توجه برای تجزیه و تحلیل داده ها در نظر گرفته شده است.

این فصل از دو بخش اصلی تشکیل شده است. بخش اول مربوط به مفاهیم بنیادی می‌باشد که در آن انواع بازار و سهام، نقدشوندگی، داده کاوی و انواع روش‌های آن و در ادامه خوشه‌بندی و الگوریتم‌های مختلف آن توضیح داده خواهد شد. بخش دوم در قالب پیشینه پژوهش، پژوهش‌های صورت گرفته در زمینه خوشه‌بندی در داخل و خارج کشور بررسی می‌شود.

۲-۲ مفاهیم بنیادی

۱-۲-۲ بازارهای مالی و طبقه‌بندی آنها

به طور کلی بازارها در یک دسته‌بندی به دو بخش بازار مالی و بازار کالا و خدمات تقسیم می‌شوند.

بازار کالا و خدمات

این بازار در واقع نخستین بازاری می‌باشد که انسان‌ها برای رفع نیاز خود ابداع کرده‌اند. انسان‌ها آموختند تا محل ساختاریافته و مشخصی را برای مبادله کالاهای خود ایجاد کنند و یک سلسله مقررات و قوانین خاص بازار را تنظیم نموده و از رویه‌ها و هنجارهای به وجود آمده توسط افراد موجود در آن تبعیت نمایند.

ابتدا مبادله پایاپای مستقیم در این نوع بازار شکل گرفت، که در آن افراد، کالاها و خدمات گوناگون را با یکدیگر مبادله می‌کردند. سپس یک نوع خاص از کالا به عنوان مبنای ارزش‌گذاری مشخص شد و دوره پول کالایی آغاز گردید و به وسیله آن مبادلات تهاتری یا پایای غیرمستقیم شکل گرفت. این کالای خاص در جوامع مختلف می‌توانست متفاوت باشد. سپس واحد پولی مورد پذیرش، به عنوان تسهیل‌کننده معاملات مورد استفاده قرار گرفت و پس از آن دوره پول اعتباری آغاز شد. در این بازار، ارزش هر کالا یا خدمات با مبلغی پول تعریف شده و معمولاً معاملات درون این بازار، تحت سازوکار شناخته شده‌ای انجام می‌پذیرد. کالا و خدمات متنوع و نامشابه در این بازار عرضه می‌شود. در بازار کالا و خدمات، واحد پولی می‌تواند به عنوان واسطه معاملات در قسمت‌های گوناگون بازار به راحتی جابجا شود و محدودیتی برای آن وجود ندارد. در واقع ما می‌توان نقدینگی موجود را در بخش‌های مختلف بازار و تحت قوانین و اصول آن به گردش درآورد. دو رکن اصلی بازار کالا و خدمات تولیدکنندگان کالا و خدمات به عنوان عرضه‌کنندگان و مصرف‌کنندگان به عنوان تقاضاکنندگان می‌باشند. تعادل در این دو رکن، تعادل در کل بازار را به وجود آورده و در حالت معمول از رکود یا تورم جلوگیری می‌کند [۶].

۲-۲-۲ بازار مالی

بازار مالی، بازاری است که در آن دارایی‌های مالی از قبیل اوراق قرضه و سهام، مورد معامله قرار می‌گیرند. در بازارهای مالی زمانی که یک طرف معامله دارایی‌های مالی را خریداری می‌کند، وجوه انتقال می‌یابد. این بازار جریان انتقال وجوه را تسهیل کرده و بدین ترتیب امکان سرمایه‌گذاری خانواده‌ها، بنگاه‌های تجاری و نهادهای دولتی را فراهم می‌کند.

بازارهای مالی، منابع پولی را از بازارهایی که دارای وجوه اضافی هستند به بازارهای نیازمند به این منابع انتقال می‌دهند. مشارکت‌کنندگان اصلی در بازارهای مالی را سه دسته خانوارها، بنگاه‌های تجاری و نهادهای دولتی تشکیل می‌دهند. واحدهای دارای مازاد وجوه نقد مشارکت‌کنندگانی هستند که سرمایه بازارهای مالی را تأمین می‌کنند که خانوارها بخش اصلی این واحدها را تشکیل می‌دهند. در مقابل، واحدهای کسری وجوه نقد مشارکت‌کنندگانی هستند که برای تهیه پول وارد بازارهای مالی می‌شوند. بسیاری از واحدهای دارای کسری وجوه نقد به منظور تهیه پول اقدام به انتشار اوراق بهادار به واحدهای دارای مازاد وجوه نقد می‌کنند [۷].

۳-۲-۲ نقش بازار مالی

بازار مالی دارای سه کارکرد اقتصادی می‌باشد. اولین کارکرد، تعیین قیمت دارایی مبادله شده از طریق روابط متقابل خریدار و فروشنده در بازار مالی است. به عبارت دیگر، این روابط تعیین‌کننده بازده مورد انتظار از دارایی مالی می‌باشد. هم‌چنان که انگیزه بنگاه برای به دست آوردن وجوه به بازده مورد انتظاری که سرمایه‌گذار طلب می‌کند بستگی دارد، این ویژگی بازار مالی نیز نشان‌دهنده آن است که در هر اقتصاد، وجوه چگونه میان دارایی‌های مالی گوناگون تقسیم گردد. این فرایند قیمت‌یابی نامیده می‌شود. کارکرد دوم، بازارهای مالی، سازوکاری برای سرمایه‌گذار فراهم می‌کنند تا بتواند در هر زمان

دلخواه، دارایی مالی را به فروش رساند. این ویژگی بازار اصطلاحاً ایجاد نقدینگی نامیده می‌شود. این ویژگی برای سرمایه‌گذاران دارای اهمیت است، به ویژه آن زمان که شرایط وی او را وادار یا تشویق به فروش می‌کند. دارنده اوراق بهادار، در صورت نبود نقدینگی، ناگزیر به نگهداری آن تا زمان سررسید و یا حتی تا زمان انحلال شرکت نزد خود خواهد بود. آن چه سبب تفاوت بازارها می‌شود، میزان نقدینگی آنهاست زیرا همه بازارهای مالی به نوعی نقدینگی ایجاد می‌کنند. سومین کارکرد، بازارهای مالی هزینه معاملات را کاهش می‌دهند. معاملات دارای دو نوع هزینه جستجو و هزینه اطلاعات هستند. هزینه جستجو شامل هزینه های آشکار مانند هزینه درج آگهی، خرید یا فروش و نیز هزینه های پنهان مثل زمان صرف شده برای یافتن طرف معامله است. هزینه های جستجو در بازار مالی سازمان یافته کاهش می‌یابد. هزینه های مربوط به برآورد ویژگی های سرمایه گذاری یک دارایی مالی هزینه های اطلاعات است، به عبارت دیگر مقدار احتمال وقوع جریان نقدینگی که از یک دارایی مالی انتظار می‌رود [۶].

۲-۲-۴ تقسیم بندی بازارهای مالی

بازار مالی، با توجه به نوع نگرش به بازار دارای تقسیم بندی های مختلفی است. انواع بازارهای مالی را می‌توان بر اساس ساختار سررسید و ساختار معاملات اوراق بهادار طبقه بندی کرد: معمولاً بازارهای مالی که تسهیل کننده معاملات اوراق بدهی هستند بر اساس سررسید تقسیم بندی می‌شوند. بازارهای پولی بازارهایی هستند که جریان وجوه کوتاه مدت با سررسید کمتر از یکسال را تسهیل میکنند و بازارهایی که جریان وجوه بلندمدت را تسهیل می‌کنند، به عنوان بازارهای سرمایه شناخته می‌شوند.

۲-۲-۵ بازار سرمایه

بر اساس طبقه بندی بازار مالی با نگرش به سررسید دارایی ها، به بازار دادوستد ابزارهای مالی با سررسید بیشتر از یک سال و دارایی های بدون سررسید، بازار سرمایه گفته می‌شود. این بخش از بازار

مالی در گردآوری منابع پس اندازی و تأمین نیازهای سرمایه‌گذاری واحدهای تولیدی نقش مهمتری دارد. بازار سرمایه نسبت به بازار پول گسترده‌تر بوده و نیز دارای تنوع ابزاری بیشتری می‌باشد. بازار سرمایه پس انداز واحدهای اقتصادی دارای مزاد را به واحدهای سرمایه‌گذاری که به آن نیازمندند انتقال می‌دهد. بنابراین بازار سرمایه، سرمایه‌گذاران و واحدهای پس اندازی را با یکدیگر مرتبط می‌سازد. از طرف دیگر مکانیسم‌های تعبیه شده در این بازار از طریق رشد حجم پس‌انداز و سرمایه‌گذاری، پیشرفت اقتصادی را تسریع می‌کنند. تأمین مالی مخارج دولت از طریق بازار سرمایه نیز از قدیمی‌ترین و متداول‌ترین شکل مبادلات مالی است. در اغلب اقتصادهای مدرن، اتکای دولت‌ها به بازار سرمایه نه تنها از این جهت مهم است که دولت‌ها بخش عمده هزینه‌های خود را از طریق انتشار اوراق بدهی تأمین می‌کنند، بلکه استقراض دولت غالباً حجم انبوهی از کل وجوه عرضه شده در بازار توسط وام‌دهندگان را جذب می‌کند [۸].

بورس اوراق بهادار مهمترین رکن بازار سرمایه است و ایجاد زمینه‌ای برای تأمین مالی سرمایه پذیران مهمترین کارکرد آن می‌باشد.

۲-۲-۶ بازار پول

بازار پول در اقتصاد ایران که دارای بازار مالی بانک محور است، بازار مهمی محسوب می‌شود. بازار پول بازاری برای دادوستد پول و دیگر دارایی‌های مالی جانشین نزدیک پول است که سررسید کمتر از یک سال دارند. همچنین از بازار پول می‌توان به عنوان بازار ابزارهای مالی کوتاه مدت دارای ویژگی اندک بودن ریسک عدم پرداخت، نقدشوندگی و ارزش اسمی زیاد نام برد. تبخشی عمده فعالیت این بازار در کاربرد ابزارهایی است که به اشخاص و بنگاه‌های تجاری این امکان را می‌دهند که به سرعت نقدینگی خود را به میزان مطلوب در آورند [۹].

بازار پول از بانک‌ها، مؤسسات اعتباری غیربانکی و سایر مکان‌هایی که دادوستد ابزارهای مالی بازار پول در آن انجام می‌شود تشکیل می‌شود به عبارت دیگر این بازار یک بازار سازمان یافته نیست و محل جغرافیایی خاصی برای بازار پول در نظر گرفته نمی‌شود. افراد یا واحدهای دارای مازاد نقدینگی، مشارکت‌کنندگان در بازار پول هستند که به عنوان پس اندازکننده از طریق مستقیم یا غیر مستقیم منابع خود را با سررسیدهای کوتاه مدت در اختیار واحدهای نیازمند منابع مالی قرار می‌دهند. به عبارتی، مهمترین هدف بازار پول ایجاد تسهیلات برای واحدهای اقتصادی به منظور تأمین نقدینگی به صورت کوتاه مدت و نیز تأمین سرمایه در گردش است [۱۰].

۲-۲-۷ بازار اولیه و ثانویه

صرف نظر از اینکه اوراق بهادار متعلق به بازار سرمایه هستند یا بازار پول، تفکیک میان معاملات بازار اولیه و بازار ثانویه ضروری است.

اوراق بهادار شرکت‌ها در بازار اولیه برای اولین بار صادر شده و در معرض مبادله قرار می‌گیرد. به عبارت دیگر اوراقی که در این بازار عرضه می‌گردد تماماً دست اول بوده و به طور مستقیم از خود شرکت ناشر یا نماینده فروش آن خریداری می‌شود.

در بازار ثانویه اوراق بهاداری وجود دارند که قبلاً انتشار یافته است. به طور مثال اگر فردی بخواهد سهم یا ورقه قرضه ای را که قبلاً توسط دیگر خریداران از ناشر آن خریداری شده است خریداری نماید، باید به بازار ثانویه مراجعه کند. بورس اوراق بهادار مهمترین نهاد بازار ثانویه برای معاملات اوراق بهادار است. سهامی که قبلاً منتشر شده در بازار ثانویه خرید و فروش می‌شود. اوراق عرضه شده در این بازار از نوع دست دوم یا دست چندم است [۶].

بازارهای ثانویه، معامله اوراق بهادار و بازارهای اولیه انتشار اوراق بهادار جدید موجود را تسهیل می‌نمایند. وجوه مورد نیاز ناشر اوراق بهادار با معاملات در بازار اولیه تأمین می‌شود، درحالی‌که معاملات در

بازار ثانویه این‌گونه نیست. سهام جدید شرکت یا اوراق بهادار جدید خزانه‌داری در بازار اولیه معامله می‌شوند درحالی‌که بازار ثانویه از معامله سهام موجود شرکت یا اوراق بهادار خزانه‌داری توسط هر فرد یا هر واحد تجاری دارنده آن تشکیل می‌گردد. نقدشوندگی ویژگی مهم اوراق بهادار معامله شده در بازار ثانویه است که نشان می‌دهد تا چه اندازه می‌توان اوراق بهادار را بدون کاهش قیمت و راحت به فروش رساند. بعضی از اوراق بهادار، بازار ثانویه فعالی دارند به عبارت دیگر در هر زمان، تعداد زیادی خریدار و فروشنده مایل به معامله آن اوراق بهادار هستند. سرمایه‌گذاران، نقدشوندگی اوراق بهادار را ترجیح داده تا در هر زمان بتوانند بدون کاهش قیمت آنها را به فروش برسانند. اگر اوراق بهادار، قابلیت نقدشوندگی نداشته باشند، سرمایه‌گذاران ممکن است نتوانند خریدار مناسبی برای آن در بازار ثانویه بیابند و گاه مجبورند برای جذب مشتری، اوراق بهادار فاقد قابلیت نقدشوندگی را با کاهش قیمت زیاد به فروش برسانند [۷].

۲-۸ بازارهای خارج از بورس و بازارهای سازمان یافته (بورس)

در بازار خارج از بورس، معاملات اوراق بهادار در قالب شبکه ارتباط الکترونیک یا غیر الکترونیک برپایه مذاکره صورت می‌گیرد. این بازارها به صورت شبکه‌های مبتنی بر ارتباطات تلفنی و رایانه‌ای هستند که معامله‌گران را با یکدیگر مرتبط می‌سازند، درحالی‌که بورس‌ها سازمان یافته‌اند و از نظر فیزیکی مکان معینی دارند. در بازارهای خارج از بورس اکثر معاملات بین مؤسسات مالی یا بین یک مؤسسه مالی و مشتری آن صورت می‌گیرد. معمولاً معامله‌گران در بازارهای خارج از بورس، بسیار بیشتر از معامله‌گران در بورس می‌باشند. شرایط قرارداد در این بازار توسط یک مرجع تعیین نمی‌شود و معامله‌گران از طریق مذاکره مستقیم با یکدیگر می‌توانند نیازهای خود را به شکل مناسب رفع کنند. مهم‌ترین مسأله در بازارهای خارج از بورس، وجود ریسک اعتباری و احتمال عدم عمل به تعهدات از سوی طرفین معامله می‌باشد که برای آن تدابیری اندیشیده شده است. دستورالعمل فعالیت بازارهای خارج از بورس به تصویب

شورای عالی بورس و اوراق بهادار در سال ۱۳۸۶ رسید و موافقت این شورا با راه اندازی بازار خارج از بورس نیز اعلام گردید.

اوراق قرضه دولتی بدون ریسک تا اوراق قرضه شرکتها، انواع سهام شرکتها و ابزارهای مشتقه، اوراق بهاداری هستند که در بازارهای خارج از بورس مورد معامله قرار می‌گیرند. معاملات اوراق بهادار و ابزار مشتقه در بازارهای خارج از بورس حجم بیشتری نسبت به معاملات بورس به خود اختصاص داده است. حتی در برخی موارد معاملات سهام در بازار خارج از بورس از حجم معاملات انجام شده در بازارهای نظام یافته پیشه گرفته است [۱۰].

۲-۲-۹ بازار بورس

بورس کلمه‌ای فرانسوی به معنای کیف پول می‌باشد. بورس به عنوان یک نهاد عمده و اساسی در بازار سرمایه، نهاد سازمان یافته‌ای است که در کنار سایر مؤسسات و سازمانها، وظایف چندگانه‌ای را برعهده دارد. شفافیت اطلاعات، مدیریت انتقال ریسک و توزیع آن، کشف قیمت، ایجاد بازار رقابتی و جمع آوری سرمایه‌ها و پس اندازهای کوچک برای تأمین سرمایه مورد نیاز فعالیتهای اقتصادی از کارکردهای بورس است.

بازار بورس اوراق بهادار، بازاری متمرکز، منسجم و سازمان یافته است که از دارای دو ویژگی مهم است: اولاً همه معاملات اوراق پذیرفته شده در بازار بورس، درتالار بورس و توسط کارگزاران و معامله‌گرانی انجام می‌شود که عضو بازار بورس هستند. ثانیاً معاملات به صورت حراجی بوده و کارگزاران تلاش می‌کنند تا اوراق بهادار ارائه شده را با کمترین قیمت ممکن برای مشتریان خود خریداری نمایند، و در مقابل که کارگزاران یاد شده سعی دارند اوراق بهادار را به بالاترین قیمت پیشنهاد شده به فروش رسانند [۱۱].

۲-۲-۱۰ تعریف سهام و انواع آن

سهام بخشی از سرمایه شرکت سهامی است که تعیین کننده میزان مشارکت و تعهدات و منافع صاحب آن در شرکت سهامی می باشد. ورقه سهام، سند قابل معامله‌ای است که نماینده تعداد سهامی است که صاحب آن سهام، در شرکت سهامی دارد در واقع سهم به قسمت‌های مساوی حاصل از تقسیم سرمایه شرکت سهامی اطلاق می‌شود. هر فرد پس از خرید سهام، مالک جزئی از دارایی شرکت می‌گردد. با خرید سهام فعلیت در بازار سرمایه آغاز می‌شود.

تقسیم‌بندی سهام

سهام را می‌توان از جهات گوناگونی تقسیم‌بندی کرد.

سهام از نظر شکل به دو نوع بانام و بی‌نام تقسیم می‌شود.

سهام با نام: سهامی است که نام دارنده آن در ورقه سهام بیان شده یا در دفتر سهام شرکت ثبت گردد. این امر موجب مشخص شدن هویت صاحب سهام برای شرکت می‌شود. انتقال این سهام باید در دفتر ثبت سهام شرکت به ثبت رسیده و توسط انتقال‌دهنده یا وکیل یا نماینده قانونی این انتقال باید امضاء شود و در موارد زیر صدور سهام با نام الزامی است :

مادامی که تمام مبلغ اسمی هر سهم بی نام پرداخت نشده باشد .

سهامی که مدیران شرکت به عنوان وثیقه می‌سپارند و تا خاتمه حساب دوره تصدی مدیران غیرقابل انتقال می‌باشد.

سهام محجورین و صغار به منظور حمایت از حقوق آنها.

سهام بی‌نام : سهامی بی‌نام است که در برگه سهام، نام کسی به عنوان صاحب آن درج نشده و یا این برگه در دفتر سهام شرکت به نام کسی ثبت نشده باشد. این سهم به راحتی قابل انتقال بوده و در واقع

برای انتقال اعتبار این برگه به دیگری نیاز به هیچ‌گونه تشریفات و قرارداد انتقال نیست و ارزش سهام متعلق به کسی است که ورقه سهام را در اختیار دارد. این نوع سهام با سهام بانام دارای ارزش مساوی بوده و دارنده آن در سود و زیان سهام شریک است. از جمله مزایای آن سرعت زیاد در انتقال آن می‌باشد. از معایب آن می‌توان به دشواری اثبات مالکیت در صورت مفقود شدن و یا سرقت و خیانت در امانت اشاره کرد. هیچ‌گونه تشریفات برای انتقال اعتبار این نوع سهام ضرورت ندارد.

سهام به دو نوع عادی و ممتاز از نظر حقوق دارنده سهام تقسیم‌بندی می‌شود:

سهام عادی: دارنده سهام عادی، در واقع، صاحب کسری از مالکیت شرکت است و تمام ریسک موجود در فعالیت اقتصادی شرکت را به میزان سهامی که در اختیار دارد، متقبل شده به عبارت دیگر، سهام عادی همان سهامی است که شرکت‌ها عرضه می‌کنند و سهامداران نسبت به سهام خود، مالک شرکت می‌شوند. می‌توان سهام عادی را در هر زمان فروخت و در صورت شرایط مناسب بازار و اقتصاد با قیمت بالاتر و در غیر این صورت با قیمت پایین‌تر سهام فروخته می‌شود. در بورس اوراق بهادار تهران فقط سهام عادی معامله می‌شود.

سهام ممتاز: تساوی حقوق سهامداران یکی از اصول اساسی شرکت‌های سهامی است، این اصل در دو مورد نقض شده و منجر به صدور سهام ممتاز می‌گردد. مورد اول زمانی است که شرکت با مشکلات مالی مواجه باشد و نتواند به آسانی به منابع مالی دسترسی یابد در این صورت ناچار به افزایش امتیازات برای سهامداران جدید است. مورد دوم در هنگام افزایش سرمایه شرکت است که از ذخایر مالی خود برای نیل به این هدف استفاده می‌نماید که در این صورت باید حقوق بیشتری برای سهامداران سابق شرکت در نظر گرفته شود.

امتیازاتی که توسط سهام ممتاز به وجود می‌آید می‌تواند به صورت برداشت نفع ثابت از سود قابل تقسیم شرکت بوده و یا اینکه دارنده سهم ممتاز می‌تواند نسبت به صاحبان سهام عادی در برداشت ارزش اسمی ممتاز به هنگام انحلال، تصفیه و تقسیم دارائی شرکت اولویت داشته باشد. به هنگام انحلال

و تصفیه شرکت مطالبات معوقه صاحبان سهام ممتاز از دارائی شرکت پرداخت می‌گردد و نیز هنگام استهلاک سهام شرکت صاحبان سهام ممتاز دارای اولویت می‌باشند. در بورس تهران سهام ممتاز وجود ندارد.

سهام انتفاعی: این سهام زمانی به فرد تعلق می‌گیرد که بدون داشتن حقی نسبت به سرمایه شرکت، از منافع آن بهره‌مند می‌شود. در واقع هنگامی که شخص با اختیار قرارداد اطلاعات علمی و فنی به شرکت منجر به ایجاد منفعت برای آن می‌گردد، می‌تواند از منافع حاصل شده سهم ببرد.

سهام مؤسس: نوع دیگری از سهام بوده که به مؤسسين شرکت تعلق می‌گیرد. میزان سود آن از سهام عادی بیشتر بوده و در اساسنامه مشخص می‌شود. این نوع سهام در اختیار کسانی قرار می‌گیرد که عضو هیئت مؤسس شرکت محسوب می‌شوند.

۲-۲-۱۱ نقدشوندگی

نقدینگی بازار به عنوان توانایی امنیت بازار شناخته می‌شود [۱۲] و جز اساسی بازارهای سهام است. به طور کلی به سهولت تبادل اوراق بهادار با قیمت معین اشاره دارد. سطح ثابتی از نقدینگی بازار برای فعالان بازار، بنگاه‌ها و نهادهای نظارتی اهمیت دارد، زیرا از تداوم تجارت با قیمت‌های مطلوب اطمینان می‌دهد، هزینه‌های افزایش سرمایه را تنظیم می‌کند و عملکرد بدون اصطکاک بازار سهام را فراهم می‌کند. نقدینگی بازار به عنوان یک حوزه مورد تحقیق در زمینه مالی در سه دهه اخیر، به عنوان مفهومی مبهم در نظر گرفته می‌شود و توسط ادبیات موجود به روش‌های مختلفی توصیف شده است. پاناییدس و همکاران^۱ [۱۳] نقدینگی بازار را معامله امن با هزینه کمتر نسبت به ارزش واقعی آن تعریف می‌کند، در حالی که هاسبروک و شوارتز^۲ [۱۴] از آن به عنوان بی‌واسطگی در انجام تجارت یاد می‌کنند. از طرف

۱- Panayides Et Al

۲- Hasbrouck And Schwartz

دیگر، لیو^۱ [۱۵] نقدینگی بازار را توانایی انجام تجارت در مقادیر زیادی از امنیت بدون تأخیر در زمان و عدم تأثیر عمده بر قیمت آن می داند.

بلک^۲ [۱۶] ویژگی های متمایزی از نقدینگی بازار را ارائه داد. وی ذکر کرد که بازار نقدینگی بازاری است که تجارت مداوم هر مقدار از اوراق بهادار را با قیمت های نزدیک به قیمت فعلی بازار آنها در مدت زمان نسبتاً کوتاه تضمین می کند.

نقدینگی بازار به عنوان یک ویژگی مهم در استراتژی های معاملاتی، تنظیم فرم سهام و پیش بینی دقیق بازده پرتفوی عمل می کند [۱۷،۱۸]. همچنین، مقدار کافی نقدینگی بازار باعث افزایش جریان اطلاعات خاص امنیت در بین فعالان بازار می شود بنابراین موجب افزایش کارایی و ثبات در بازار سهام می شود [۱۹،۲۰].

کایل^۳ [۲۱] اشاره می کند که نقدینگی یک بازار را می توان از نظر سه جنبه درک کرد. مقدار اوراق بهادار قابل معامله (عمق)، توانایی قیمت های اوراق بهادار برای بازیابی سریع پس از شوک نقدینگی (انعطاف پذیری) و هزینه های انجام شده در امنیت تجارت (استحکام). بعلاوه، زمان لازم برای اجرای تجارت (فوریت) و شدت تأثیر حجم معاملات بر قیمت های امنیتی (عرض) نیز به عنوان جنبه های اضافی نقدینگی بازار در نظر گرفته می شود [۲۲].

۲-۲-۱۲ داده کاوی^۴

توسعه فناوری اطلاعات باعث ایجاد تعداد زیادی پایگاه داده و داده های عظیم در مناطق مختلف شده است. تحقیقات در پایگاه داده ها و فناوری اطلاعات باعث ایجاد رویکردی برای ذخیره سازی و استفاده

۱- Liu

۲-Black

۳- Kyle

۴- Data Mining

داده‌های گرانبها برای تصمیمات آینده شده است. داده‌کاوی فرایند استخراج اطلاعات مفید و الگوها از تعداد زیادی داده می‌باشد [۲۳].

یک تعریف ساده از داده کاوی جستجوی روابط و الگوهایی در پایگاه داده های بزرگی است که در تعدادی بی‌کران از داده‌ها پنهان شده است [۲۴]. به طور ساده داده کاوی، استخراج دانش از مقدار زیادی داده می‌باشد. بسیاری داده‌کاوی را معادل با کشف دانش می‌دانند درحالی‌که داده‌کاوی گامی مهم در کشف دانش است [۲۵]. اولین توضیحات متمایز برای این دو اصطلاح توسط فایاد و همکاران^۱ با اشاره به آن که کشف دانش یک فرایند کلی از کشف اطلاعات جدید از منابع داده است که داده کاوی مرحله‌ای که در آن قوانین و الگوریتم‌های خاصی اعمال می‌شود، بیان شد. این الگوریتم‌ها اساس کشف دانش را تشکیل می‌دهند که می‌توان از داده‌های داده شده الگوهای جالبی را استخراج کرد.

جدا از داده کاوی، کشف دانش شامل یک سری فعالیت‌ها مانند آماده سازی داده ها، انتخاب داده‌ها، پالایش داده‌ها و پیش‌پردازش، استخراج برای روابط جالب و ارائه و تجسم الگوهای به دست آمده است [۲۶].

داده‌کاوی به یک نوع از داده‌ها محدود نشده و معمولاً داده‌های مختلفی در داده‌کاوی وجود دارند. روش‌های متفاوتی برای انواع مختلف داده‌ها مناسب هستند و یافتن یک راهکار کلی، تلاشی بیهوده به نظر می‌رسد.

تکنیک‌های داده‌کاوی را می‌توان بر روی داده‌های غیر ساخت یافته مانند متون، نیمه ساخت یافته مانند اسناد و ساخت یافته مانند جداول اعمال کرد. در ادامه به طور خلاصه توضیح وجود دارد:

یکی از رایج‌ترین شکل‌های ورودی برای الگوریتم‌های داده‌کاوی جداول هستند. در جداول سطرها نماینده نمونه‌ها و ستون‌ها ویژگی و صفات خاص نمونه‌ها را تشکیل می‌دهند. اغلب روش‌های داده‌کاوی با

۱- Fayyad, Piatetsky-Shapiro, And Smyth

این شکل از داده‌ها دچار مشکل نمی‌شوند. حتی در برخی از کاربردها کاربران ابتدا داده‌های خود را به صورت جدول در می‌آورند.

اکثر روش‌های داده‌کاوی تمرکز بر روی داده‌های ساخت‌یافته مانند جداول دارند، حال آنکه حجم عظیمی از اطلاعات در دسترس در دنیای واقعی به صورت نیمه‌ساخت‌یافته و یا غیرساخت‌یافته ذخیره شده‌اند. این پایگاه داده شامل مجموعه بزرگی از مستندات متنی مانند مقالات، کتاب‌ها و صفحات وب می‌شوند. این موضوع اهمیت استفاده از روش‌های داده‌کاوی را برای این نوع از داده‌ها دو چندان کرده است. عموماً این داده‌ها نیمه ساخت‌یافته هستند. برای مثال یک مقاله شامل برخی از ویژگی‌های ساخت‌یافته مانند نویسنده، عنوان، تاریخ چاپ و... همچنین شامل واژه‌هایی است که از هیچ ساختاری (صرف‌نظر از ساختمان یک جمله) پیروی نمی‌کنند.

انبار داده‌ها به عنوان شکل دیگری از داده‌ها، مخزنی از اطلاعات جمع‌آوری شده از چندین منبع داده‌ای تحت یک ساختار واحد است که از آنها می‌توان به تنهایی نیز جهت تحلیل داده‌ها استفاده کرد. به دلیل آنکه این داده‌ها از منابع مختلفی جمع‌آوری می‌شوند، عملیاتی چون پالایش داده‌ها، حذف نویز و داده‌های ناقص و تبدیل داده‌ها به شکل‌های مناسب برای داده‌کاوی روی آن انجام می‌گیرد.

شکل دیگری از داده‌ها پایگاه داده تراکنشی^۱ است و حاوی مجموعه رکوردهایی است که هر یک از آن‌ها دلالت بر یک تراکنش واحد همراه با اطلاعات دیگر دارد. نمونه بارز از این نوع داده‌ها تحلیل سبد خرید مشتریان فروشگاه‌ها است.

صفت خاصه زمان یک ویژگی بسیار مهم برای مجموعه داده‌ها، صفت خاصه زمان است. پایگاه داده‌ای شامل صفت خاصه زمان اطلاعات مفیدتر و دقیق‌تری را تحت اختیار کاربران قرار می‌دهد. چنین پایگاه

۱- Transactional Database

داده‌ای که حاوی رخدادهایی است که با زمان تغییر می‌کند، پایگاه داده سری‌های زمانی نامیده می‌شود. رفتار محتویات تراکنش‌ها در رابطه با زمان می‌تواند توسط تکنیک‌های داده‌کاوی بررسی شود.

امروزه وب به عنوان یک مخزن داده پویا شمال انواع داده‌ها از جمله صدا، متن و تصویر است. وب‌کاوی، تکنیک‌های داده‌کاوی را با این مجموعه داده‌ها پیوند می‌دهد. جستجو در داده‌های چندرسانه‌ای نیز می‌تواند به وب‌کاوی کمک کند. یک سیستم مدیریت پایگاه داده چندرسانه‌ای مجموعه وسیعی از داده‌های چندرسانه‌ای را ذخیره و مدیریت می‌کند.

یک پایگاه داده مکان‌محور^۱ شامل مجموعه داده‌های زیادی در رابطه با مکان است. نقشه‌ها، تصاویر پزشکی نمونه‌ای از این داده‌ها بشمار می‌روند. امروزه داده‌کاوی این نوع از پایگاه داده به طور گسترده‌ای مورد استفاده کاربران قرار می‌گیرد.

عملیات داده‌کاوی را می‌توان در دو گروه تقسیم‌بندی کرد، که گروه اول به توصیف خصوصیات داده‌ها در پایگاه داده‌ها می‌پردازد و گروه دوم با مدل‌سازی داده‌های در دسترس سعی دارند تا به یک پیش‌بینی صحیح از داده‌های آتی و آزمایشی برسند. می‌توان از دو روش هنگامی که کاربر هیچ‌گونه نظری در مورد نوع الگوها ندارد، استفاده و نتایج را مشاهده و ارزیابی کند [۲۷].

داده‌کاوی شامل تکنیک‌های مختلفی است که انتخاب تکنیک‌های داده‌کاوی براساس ویژگی‌های داده‌ها و نیازهای کاری می‌باشد [۲۸].

روش‌های کشف قوانین انجمنی، طبقه‌بندی و خوشه‌بندی از عمده‌ترین تکنیک‌های داده‌کاوی می‌باشد.

قوانین انجمنی

طی سال‌های گذشته در میان تکنیک‌های داده‌کاوی توجه خاصی به الگوریتم‌های کشف الگوهای

۱- Spatial Database

مکرر وجود داشته است. همان طور که از نام این الگوریتمها مشخص است، به دنبال الگوهایی است که به دفعات در مجموعه داده‌ها دیده می‌شوند.

الگوریتم‌های کشف قوانین انجمنی، می‌توانند تعداد بسیار زیادی از قوانین را تولید کنند. حتی در صورت وجود تعداد اندک ارقام داده‌ها نیز حجم وسیعی از قوانین وجود دارد. حتی در صورت مفید فرض شدن کلیه الگوها، برای کاربر امکان‌پذیر نیست تا قضاوت مناسبی میان آنها داشته باشد. به همین دلیل نیاز به الگوریتم‌های مؤثر جهت محدود نمودن این فضای وسیع و همچنین معیارهایی جهت ارزیابی قوانین انجمنی به خوبی احساس می‌شود.

۲-۲-۱۳ طبقه‌بندی

پایگاه داده‌ها منبع بسیار غنی از اطلاعات پنهان است که می‌توان به کمک این اطلاعات تصمیمات هوشمندی را اتخاذ نمود. در این میان طبقه‌بندی و تخمین دو شکل از تحلیل داده‌ها محسوب می‌شوند که می‌توان به کمک آنها مدلی جهت توصیف داده‌ها استخراج کرد یا برای داده‌های بعدی جهتی متصور شد. بدین وسیله داده‌هایی با حجم بالا نیز بهتر فهمیده می‌شوند.

روش‌های نظارت شده‌ای مانند تخمین و طبقه‌بندی بدنبال کشف رابطه‌ای میان صفات خاص ورودی (که گاه متغیرهای مستقل نامیده می‌شوند) با یک یا چندین صفت خاصه هدف (که گاه متغیر وابسته نامیده می‌شوند) هستند. درنهایت این رابطه با یک ساختار به عنوان مدل نمایش داده می‌شود.

با کمک این مدل و به شرط داشتن صفات خاصه ورودی می‌توان مقدار صفت خاصه هدف را تخمین زد. به عبارت دیگر با کمک مدل می‌توان نمونه‌ها را به یکی از چندین طبقه تعریف شده منتسب و یا مقدار تعیین شده‌ای را برای صفت خاصه هدف تعیین کرد.

فرایند ساخت مدل یک فرایند دومرحله‌ای است، در مرحله اول که به عنوان مرحله یادگیری شناخته می‌شود با کمک مجموعه داده‌های آموزشی که برچسب کلاس تمام نمونه‌های آن مشخص است، مدل ساخته می‌شود. در مرحله دوم مدل بدست آمده با کمک مجموعه داده‌های آزمایشی با برچسب کلاس‌های نامعلوم، اعتبارسنجی می‌شود. در واقع ارزشیابی مدل با توجه به اینکه کلاس چه تعداد از نمونه داده‌های آزمایشی، درست تخمین زده شده است، محاسبه می‌شود.

خوشه‌بندی

فرایند گروه‌بندی مجموعه‌ای از داده‌ها و قرار دادن آنها در طبقاتی از نمونه‌های مشابه خوشه‌بندی نام دارد. یک خوشه مجموعه‌ای از داده‌هاست که نسبت به دیگر داده‌های همان خوشه شبیه بوده ولی متفاوت از نمونه‌های دیگر خوشه‌هاست.

تحلیل خوشه‌ها کاربردهای بسیارمتفاوتی از جمله تشخیص الگو، پردازش تصاویر، تحلیل داده‌ها و تحلیل‌های تجاری دارد.

می‌توان خوشه‌بندی را برای پیش‌پردازش داده‌ها و یا آماده‌سازی آن در مراحل کشف و استخراج دانش بکار برد. بسته‌های نرم‌افزاری بسیاری همچون SPSS و xlstat دارای روش‌های کلاسیکی از خوشه‌بندی هستند که این موضوع اهمیت مسئله را در تحلیل داده‌ها بیش از پیش مشخص می‌سازد.

یک الگوریتم خوشه‌بندی می‌تواند دارای مشخصات مطلوبی مانند قابلیت مقیاس‌پذیری، توانایی مواجهه با انواع داده‌ها، استخراج خوشه‌هایی به هر شکل دلخواه، توانایی مقابله با داده‌های نویز، عدم حساسیت به ترتیب ورود داده‌ها، پذیرش داده‌هایی با ابعاد بالا و همچنین قابل فهم بودن نتایج نهایی الگوریتم باشد.

موضوع اصلی در روش‌های خوشه‌بندی شباهت و عدم شباهت دو نمونه داده است. در هر خوشه نمونه‌هایی که تشابه بیشتر دارند قرار می‌گیرند. به عبارت نمونه‌های مشابه در یک خوشه و نمونه‌های غیرمشابه در خوشه‌های متفاوت گروه‌بندی می‌شوند [۲۷].

۲-۲-۱۴ خوشه‌بندی سلسله‌مراتبی^۱

در روش‌های خوشه‌بندی سلسله‌مراتبی، خوشه‌ها توسط الگوهای تکراری تقسیم و با استفاده از رویکردهای بالا به پایین و پایین به بالا تشکیل می‌شوند. دو دسته از روش‌های سلسله‌مراتبی به نام‌های تجمیعی^۲ و تقسیمی^۳ وجود دارد. روش تجمیعی از رویکرد پایین به بالا پیروی می‌کند. این رویکرد با قرار دادن همهٔ اشیاء در خوشه‌های مجزا آغاز می‌شود، سپس در هر گام دو خوشه را ادغام می‌کند تا زمانی که کلیهٔ اشیاء در یک خوشه قرار بگیرند یا شرایط خاتمه مورد رضایت ایجاد شده باشد. روش تقسیمی از رویکرد بالا به پایین پیروی می‌کند. در رویکرد تقسیمی همهٔ اشیاء در یک خوشه قرار دارند سپس در هر گام با کمک یک معیار تشابه در چند مرحله به صورت سلسله‌مراتبی این خوشه به خوشه‌های کوچکتر تقسیم می‌شود این روند تا زمانی که هر شی در یک خوشه قرار بگیرد یا شرایط خاتمه الگوریتم فراهم شود ادامه می‌یابد.

روش‌های خوشه‌بندی سلسله‌مراتبی را نیز می‌توان براساس اندازه‌گیری‌های شباهت یا پیوندها [۲۹] گروه‌بندی کرد:

^۱-Hierarchical Clustering (Hc) Methods
^۲-Agglomerative Hierarchical Clusterin
^۳-Divisive Hierarchical Clustering

۲-۲-۱۴-۱ خوشه‌بندی پیوندتکی^۱

خوشه‌بندی پیوند تکی، روش حداقل و نزدیک‌ترین همسایه نیز نامگذاری می‌شود. این روش توسط یک جفت عنصر که هر کدام متعلق به یک خوشه است و بهم بسیار نزدیک هستند، انجام می‌شود. در این خوشه‌بندی فاصله بین هر دو خوشه توسط نزدیک‌ترین فاصله بین هر عضو از یک خوشه با هر عضوی دیگر از خوشه دیگر مشخص می‌شود که این مطلب نیز به عنوان شباهت تعریف می‌شود. اگر مقادیر شباهت اندازه‌گیری شده باشد، در این صورت شباهت بین دو جفت خوشه برابر با بزرگترین مقدار شباهت بین هر عضو از یک خوشه با هر عضو از خوشه دیگر است [۳۰].

۲-۲-۱۴-۲ خوشه‌بندی پیوند کامل^۲

در این روش که روش حداکثر و روش دورترین همسایه نیز نامیده می‌شود، فاصله بین دو خوشه توسط دورترین فاصله بین هر عضو از یک خوشه و هر عضوی دیگر از خوشه دیگر تعیین می‌شود. در ابتدای فرایند، هر عنصر در خوشه ای از خود قرار دارد. سپس خوشه‌ها به صورت گروهی به گروه‌های بزرگتر تقسیم می‌شوند تا زمانی که تمام عناصر در خوشه قرار گیرند.

۲-۲-۱۴-۳ خوشه‌بندی پیوند میانگین^۳

در روش خوشه‌بندی پیوند میانگین که به عنوان روش حداقل واریانس نیز شناخته می‌شود، فاصله بین دو خوشه از طریق میانگین فاصله بین هر عضو یک خوشه با هر عضو خوشه دیگر تعیین می‌شود [۳۱].

۱- Single-linkage Clustering
۲- Complete-Linkage Clustering
۳- Average-Linkage Clustering

خوشه‌بندی و طبقه‌بندی، هر دو از وظایف اساسی در داده‌کاوی است. طبقه‌بندی بیشتر به عنوان یک روش یادگیری نظارت شده و خوشه‌بندی به عنوان یک روش یادگیری بدون نظارت استفاده می‌شود (برخی از روش‌های خوشه‌بندی برای هر دو مورد است). هدف خوشه‌بندی توصیفی است، در حالیکه هدف طبقه‌بندی پیش‌بینی است [۳۲].

تکنیک‌های طبقه‌بندی به منظور تشخیص گروه‌ها یا کلاس‌های هر نمونه روش مؤثری است اما در این‌گونه روش‌ها نمونه‌های ما دارای یک لیبل هستند تا الگوریتم بتواند با کمک آن مدلی را طراحی ارزیابی شباهت در گروه‌هایی قرار گرفته سپس به هر گروه لیبل به عنوان کلاس زده می‌شود [۲۳].

۲-۲-۱۵ پیشینه تحقیق

۲-۲-۱۵ داخلی

الگوریتم خوشه‌بندی کاربردهای فراوانی در زمینه تقسیم‌بندی بازار، درجه‌بندی اعتبار، پیش‌بینی ورشکستگی و... دارند. تا کنون در ایران پژوهش‌های زیادی در زمینه خوشه‌بندی سهام شرکت‌های پذیرفته شده در بورس با استفاده از روش‌های مختلف خوشه‌بندی انجام شده است در حالیکه تاکنون پژوهشی در زمینه خوشه‌بندی سهام با توجه به شاخص‌های نقدشوندگی با استفاده از ضریب تشابه تطابق نسبی ارائه نشده است، بنابراین در این قسمت به بیان پیشینه پژوهش‌های مشابه که با سایر روش‌ها انجام شده، پرداخته شده است که از این جمله می‌توان به پژوهش‌های زیر اشاره کرد:

فرید و پورحمیدی (۱۳۹۰) با استفاده از داده‌های صورت‌های مالی شرکت‌های پذیرفته شده در بورس اوراق بهادر تهران در دوره زمانی ۱۳۸۴ تا ۱۳۸۸ و روش خوشه‌بندی به بخش‌بندی سهام شرکت‌ها پرداختند، نتایج بررسی آنها نشان داد که بخش اعظم شرکت‌ها در سبد سهام ترکیبی قرار می‌گیرند، لیکن گرایش آنها به سبد سهام رشدی است [۳۳]. میرزایی و یغمایی مقدم (۱۳۹۰) به شناسایی کیفیت

سهام جهت کمک به سرمایه‌گذاران در انتخاب تعدادی سهام مناسب با استفاده از یک سیستم خبره پرداختند و همچنین خوشه‌بندی فازی سی- میانگین را برای مقایسه با روش معمولی به کار گرفتند و به این نتیجه رسیدند علاوه بر بازدهی پارامترهای دیگر در رتبه‌بندی نقش دارد [۳۴]. امیری و همکاران (۱۳۹۲) با جمع‌آوری اطلاعات ۱۸۱ شرکت بورسی در طی ۳ سال، ابتدا ۵۰ شرکت را با استفاده از روش تاپسیس رتبه‌بندی کرده، سپس این رتبه بندی را با استفاده از شبکه عصبی به چهار دسته تقسیم کردند. پس از این دسته‌بندی، نتایج شبکه عصبی با استفاده از روش‌های آماری مورد بررسی قرار گرفت که نتایج آماری دسته‌بندی شبکه عصبی را تأیید کرد. در نهایت همه شرکت‌ها توسط شبکه عصبی دسته‌بندی شدند و با نظر خبرگان مورد مقایسه قرار گرفتند که بررسی‌ها نشان‌دهنده این موضوع بود که نتایج شبکه عصبی بسیار نزدیک به نظر خبرگان بوده است [۳۵].

طاهری و خاوری‌نژاد (۱۳۹۲) با جمع‌آوری اطلاعات مورد نیاز از ۲۰۴ شرکت پذیرفته شده در بورس تهران در فاصله زمانی بین ۱۳۸۶ تا ۱۳۹۰، از طریق شاخص‌هایی مانند نسبت جاری، نسبت بدهی و... به رتبه‌بندی صنایع مختلف پرداخته‌اند. از بین ۳۲ طبقه فعالیتی در شرکت‌های بورسی، ۱۱ طبقه تولیدی شناسایی شدند. نتایج این پژوهش نشان داد که در سال ۱۳۸۹ صنعت ماشین‌آلات و تجهیزات در جایگاه اول قرار گرفت ولی در سال ۱۳۹۰ این صنعت در رتبه چهارم جای گرفت و صنعت خودروسازی و قطعات جایگاه اول را به خود اختصاص داد [۳۶].

پورزرنندی و کیخا (۱۳۹۳) سبد سهام بهینه را با استفاده از الگوریتم خوشه‌بندی k-means و الگوریتم ژنتیک انجام دادند. برای این منظور اطلاعات مربوط به سالهای ۱۳۸۸ تا ۱۳۹۰ را بررسی کردند. ابتدا طبقه‌بندی سهام‌های پرتفوی ابتدایی با روش K-Means به دسته‌های کوچکتر انجام دادند، سپس طبقه‌ای بهینه‌تری یعنی بایشتترین بازده و کمترین ریسک، را به عنوان ورودی الگوریتم خود انتخاب کردند و بعد از آن الگوریتم ژنتیک را برای این طبقه از سهام اجرا کرده و نتایج حاصل از این تحقیق نشان داد که بازده بدست آمده از اجرای الگوریتم ژنتیک به تنهایی در مقایسه با اجرای الگوریتم بر روی طبقه

بهینه، کمتر و ریسک آن بیشتر است و طبقه‌بندی داده‌ها و سپس اجرای الگوریتم ژنتیک روی طبقه بهینه باعث دستیابی به پرتفویی می‌شود که نسبت به پرتفوی حاصل از اجرای الگوریتم ژنتیک به تنهایی دارای ریسک کمتر و بازدهی بیشتر می‌باشد [۳۷].

شعبان پورالمشیری و فرسادنیا (۱۳۹۴) به بخش‌بندی سهام شرکت‌ها با استفاده از داده‌های صورت‌های مالی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران پرداختند. برای این منظور ۲۷۱ شرکت از میان شرکت‌های لیست شده در بورس اوراق بهادار تهران را برای دوره ۵ ساله (۱۳۸۷ تا ۱۳۹۱) انتخاب کردند. از روش خوشه‌بندی فازی و روش الگوریتم دو مرحله‌ای نگاشت ویژگی خودسامان استفاده نمودند. نتایج مطالعه آنها نشان داد که بخش اعظم شرکتها در سبد سهام ترکیبی قرار می‌گیرند، لیکن گرایش آنها به سبد سهام رشدی است [۳۸].

سلگی و همکاران (۱۳۹۴) به طبقه‌بندی سهام رشدی و ارزشی در بورس اوراق بهادار تهران پرداختند. به این منظور ۵۵ شرکت فعال در صنعت محصولات غذایی، آشامیدنی و قند و شکر از میان شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران را به عنوان نمونه در سالهای ۱۳۹۲-۱۳۹۳ انتخاب کردند. روش کار آنها استفاده از الگوریتم خوشه‌بندی فازی سی- میانگین بود. نتایج حاصل از تحقیق، طبقه‌بندی سه دسته سهام‌های رشدی (۰.۳۴.۵)، ارزشی (۰.۴۷.۳) و نیمه رشدی-نیمه ارزشی بود [۳۹].

اقبال نیا و همکاران (۱۳۹۴) در پژوهش خود به خوشه‌بندی شرکت‌های موجود در بورس اوراق بهادار براساس شباهت در روند حرکتی قیمت سهام شرکت‌ها با استفاده از روش خوشه‌بندی سه مرحله‌ای پرداختند. در مرحله اول، کاهش ابعاد داده‌ها و خوشه‌بندی تقریبی شرکت‌ها انجام شده است. سپس در مرحله دوم، خوشه‌های ایجاد شده در مرحله قبل، به زیرخوشه‌های با کیفیت و دقت بیشتر تقسیم شدند. در نهایت، زیر خوشه‌های ایجاد شده، در مرحله سوم ادغام و خوشه‌های نهایی ارائه شدند. مدل ارائه شده، با استفاده از روش مجموع مربعات خطا مورد ارزیابی قرار گرفته است. نتایج تحقیق نشان داد روش

خوشه‌بندی سه مرحله ای در مقایسه با سایر الگوریتم‌های خوشه بندی مرسوم از نظر اثربخشی و کیفیت، عملکرد بهتری دارد [۴۰].

صادقی و همکاران (۱۳۹۵) به خوشه‌بندی شرکت‌های پذیرفته شده در بورس بهادار تهران به منظور برآوردن نیازهای سرمایه‌گذاران در انتخاب سهام پرداختند. آنها ۳۷۴ شرکت پذیرفته شده در بورس اوراق بهادار تهران را با استفاده از روش k-means و براساس معیارهای مالی خوشه‌بندی کردند. نتیجه پژوهش آنها نماینگر آن است که خوشه‌بندی باعث کاهش زمانی می‌شود که سرمایه‌گذاران قرار است برای پرتفویهای کارآمد با توجه به سهام‌های مشابه‌ای که در خوشه‌ها طبقه‌بندی شده‌اند، صرف کنند [۴].

جلیوند و همکاران (۱۳۹۶) با بهره‌گیری از روش کتابخانه‌ای و مطالعه مروری به بررسی نقش تکنیک خوشه‌بندی بر بهبود مدیریت ریسک در شرکت‌های بیمه پرداخته و الگوریتم‌های خوشه‌بندی ریسک مالی، عملیاتی و مشتریان را مورد بررسی قرار دادند [۴۱].

پاک‌مرام و همکاران (۱۳۹۶) به انتخاب و بهینه‌سازی سهام با استفاده از سه الگوریتم ژنتیک، فرهنگی و ازدحام ذرات پرداختند و به این منظور ۱۰۶ شرکت پذیرفته شده بورس اوراق بهادار تهران، در طی دوره زمانی ۱۳۸۶ الی ۱۳۹۳ را مورد آزمون قرار دادند. آنها در پژوهش خود به بررسی تفاوت بین میانگین بازده سرمایه گذاری در سبدهای منتخب بر اساس سه روش پرداخته و آزمون‌های آماری مربوط به نتایج، حاکی از عدم وجود اختلاف معنادار بین سه الگوریتم بود. از طرفی به منظور مقایسه دو الگوریتم و بررسی برتری الگوریتم‌ها، این دو روش بهینه سازی از دو بعد تابع هدف و نسبت بازده و ریسک مورد مقایسه قرار گرفتند و از آنجایی که الگوریتم ژنتیک مقدار تابع هدف کمتری داشته یا به عبارتی با کمترین خطا به بهترین نتیجه رسیده است، نسبت به الگوریتم‌های دیگر بهتر عملکرد را داشته و نشان دهنده برتری نسبی این الگوریتم در انتخاب سبد سهام بهینه است [۴۲].

مهربان‌پور و حبیب‌زاده (۱۳۹۷) به خوشه‌بندی و پیش‌بینی سودآوری شرکت‌ها و تعیین عوامل مؤثر بر سودآوری شرکت‌های عضو بورس اوراق بهادار پرداختند. به این منظور ۸۸۸ شرکت در محدوده زمانی

۱۳۸۷-۱۳۹۵ انتخاب کردند. پس از پردازش اولیه داده‌ها و با استفاده از نرم‌افزار متلب و Clementine و با استفاده از معیار SSE و روش K-means شرکت‌ها را به ۳ خوشه تبدیل شد و نتایج این خوشه‌بندی‌ها را بوسیله معیار سنجش کیفیت، مورد سنجش قرار دادند. در ادامه با استفاده از درخت تصمیم C5 خوشه‌ها را تحلیل و متغیرهای تاثیرگذار بر سودآوری، شناسایی کردند. از ۳۲ متغیر تحلیل شده تنها ۸ متغیر شامل: سودخالص به کل دارایی، فروش به کل دارایی، سودخالص به حقوق صاحبان سهام، سود عملیاتی به فروش خالص، سود و زیان انباشته به حقوق صاحبان سهام، سودخالص به فروش خالص، کل بدهی‌ها به کل دارایی‌ها و دارایی‌های جاری به کل دارایی‌ها بر سودآوری شرکت‌ها تاثیر می‌گذارند. در نهایت با در نظر گرفتن این متغیرها، پیش‌بینی سودآوری شرکت‌ها طبق هر خوشه انجام شد که دقت پیش‌بینی خوشه‌ها به ترتیب ۸۶/۳۴ درصد، ۸۸/۱۵ درصد و ۸۱/۶۸ درصد بدست آمد [۴۳].

مشرفی و بهنامیان (۱۳۹۸) در پژوهش خود به تجزیه و تحلیل مدل مبنایی بهینه‌سازی سبد سهام با در نظر گرفتن دانش مدیریت مالی و سرمایه‌گذاری جهت ارزیابی ریسک وبازده با توجه به محدودیت‌هایی از قبیل دارایی فرد خریدار برای خرید هر سهم، پرداختند. آنها مدلی جدید را در قالب برنامه‌ریزی خطی جهت بهینه‌سازی سبد سرمایه‌گذاری و با در نظر گرفتن نرخ بازده مورد انتظار و حداقل ریسک و دارایی فرد، طراحی کردند. بعد از مطرح کردن مدل مورد نظر در قالب برنامه‌ریزی خطی و بیان محدودیت‌های مربوط به آن، انواع مختلف سرمایه‌گذاری را بررسی کرده که یک سرمایه‌گذار می‌تواند جهت تشکیل سبد سرمایه‌گذاری خود، آنها را مورد بررسی قرار دهد. در نهایت، برای حل این مدل یک روش با استفاده از الگوریتم ژنتیک ارائه و در ارتباط با نمونه‌های واقعی اجرا و تحلیل کردند. براساس نتایج این تحقیق، مدل جدید، ریسک نامطلوب را به میزان بسیار زیادی در مقایسه با مدل‌های ارائه شده قبلی کاهش داده، به گونه‌ای که این روند با افزایش تعداد سهام مورد مطالعه به صورت پله‌ای و نزولی ادامه می‌یابد [۴۴].

انصاری و همکاران (۱۳۹۸) به ارائه مدلی دو مرحله‌ی براساس برنامه‌ریزی ترکیبی عدد صحیح پرداختند که عملکرد پرتفوی را نسبت به روش یک مرحله‌ای بهبود می‌بخشد. مرحله اول مربوط به کمینه‌سازی خطای ردیابی و مرحله دوم مربوط به بشینه‌سازی بازه تحت مقادیر تلورانس مجاز برای خطای ردیابی است. به این منظور ۵۰ شرکت فعال در بورس اوراق بهادار را در نظر گرفتند. یافته‌های پژوهش نشان‌گر عملکرد بهتر مدل دو مرحله‌ای نسبت به مدل یک مرحله‌ای است [۴۵]. صیادی و امید (۱۳۹۸) در پژوهش خود از یک مدل بهینه‌سازی پرتفوی مبتنی بر پیش‌بینی برای انتخاب پرتفوی سهام گروه‌های وابسته به نفت در بازار بورس تهران استفاده کردند. آنها ابتدا سهام فرآورده‌های نفتی و صنایع شیمیایی پیش‌بینی براساس داده‌های خوشه‌بندی شده بازار سهام و مبتنی بر رهیافت داده‌کاوی را پیش‌بینی کردند. سپس، با استفاده از عوامل مؤثر بر تغییرات شاخص هر گروه مانند قیمت نفت خام، نرخ ارز، نرخ بهره جهانی، قیمت جهانی طلا و شاخص S&P500، شاخص هر صنعت را با استفاده از الگوریتم‌های شبکه عصبی MLP و RBF تخمین زده و در نهایت با مقایسه عملکرد هر یک از الگوریتم‌ها، بهترین الگوریتم برای پیش‌بینی رفتار شاخص هر صنعت شناسایی کردند. در ادامه با استفاده از الگوریتم‌های خوشه‌بندی K-means، SOM و FCM شرکت‌های موجود در این دو صنعت از لحاظ نسبت‌های مالی خوشه‌بندی کرده و با بهترین الگوریتم سهام مناسب از هر گروه شناسایی کردند. نتایج تحقیق آنها بیانگر آن است که الگوریتم MLP از دقت بالاتری برخوردار است. همچنین الگوریتم FCM بهترین خوشه‌ها را تولید می‌کند. نتایج تجربی نیز نشان داد، سهام پتروشیمی سپاهان و خارگ در کوتاه‌مدت و پتروشیمی خارگ و فناوران و پالایشگاه نفت تهران بیشترین بازده را در پرتفوی در افق میان‌مدت و بلندمدت دارد [۴۶]. آذر و همکاران (۱۳۹۹) به بهینه‌سازی سبد سهام با رویکرد کمینه کردن ریسک در معرض خطر و بیشینه کردن بازده پرتفوی به صورت همزمان پرداختند. الگوریتم مورد استفاده آنها الگوریتمی پویا براساس الگوریتم ژنتیک و مفاهیم ارزش در معرض خطر بود. از الگوریتم مورد نظر برای بدست آوردن وزن‌های بهینه در سبد سهام و همچنین بهبود الگوریتم k-means برای طبقه‌بندی

داده‌ها استفاده کردند. نتایج تحقیق آنها نشان داد که طبقه‌بندی داده‌ها و سپس اجرای الگوریتم ژنتیک روی طبقه بهینه موجب دستیابی به پرتفویی می‌شود که نسبت به پرتفوی حاصل از اجرای الگوریتم ژنتیک کلاسیک به تنهایی و یا حتی همراه به الگوریتم k-means ساده جهت طبقه‌بندی، دارای ریسک کمتر و بازدهی بیشتر است [۴۷]. مقدم و همکاران (۱۳۹۹) در پژوهش خود با استفاده از کلان داده‌های ارزشمند تولید شده توسط بازار سهام وروش خوشه‌بندی افرازی و به کمک الگوریتم ن k-means به تعیین نقاط سیگنال معاملات سهام پرداختند. آنها در پژوهش خود داده‌های صنایع خودرو و فرآورده‌های نفتی را طی سال ۱۳۸۷ تا ۱۳۹۶ مورد استفاده قرار دادند. نتایج پژوهش نشان داد که مدل مورد استفاده در شناسایی و پیش بینی سیگنال‌های فروش صادره در نقاط حداکثری دارای عملکرد قابل توجهی بوده و با دقت قابل قبولی قابل پیش‌بینی می‌باشند. در واقع این سیگنال‌ها دارای خطای کمتری بوده و بهتر پیش‌بینی گردیده است [۴۸].

۲-۲-۱۶ خارجی

شین و سان^۱ (۲۰۰۴) مشتریان خود را با سه روش K-Means, SOM, FCM در سه گروه بخش‌بندی نمودند. نتایج پژوهش آنها نشان داد که روش تحلیل خوشه‌ای K-Means فازی تواناترین روش خوشه‌بندی مشتریان است [۴۹].

باسالتو و همکاران^۲ (۲۰۰۵) از الگوریتم خوشه‌بندی نقشه آشوب برای خوشه‌بندی شرکت‌های فهرست شده در شاخص جونز (DJ) استفاده کردند که شناسایی خوشه‌های شرکت‌های شاخص بازار سهام داده شده را می‌توان در راهبردهای بهینه‌سازی پرتفوی مورد استفاده قرار داد [۵۰].

۱- H. W. Shin And S. Y. Sohn

۲- N. Basalto, R. Bellotti, F. De Carlo, P. Facchi, And S. Pascazio

دوهرتی و همکاران^۱ (۲۰۰۵) از الگوریتم خوشه‌بندی سلسله‌مراتبی TreeGNG برای بخش‌بندی بازار سهام در بازه‌زمانی ده ساله استفاده کردند. از نظر آنها روش بکار برده شده ساختار بخش‌بندی را نسبت به خوشه‌بندی تجمعی آماری واضح‌تر مشخص می‌کند [۵۱].

لیو، لین و هو^۲ (۲۰۰۸) با استفاده از یک رویکرد داده‌کاوی دو مرحله‌ای، مسائل مربوط به سرمایه‌گذاری بازار سهام در بازار سهام تایوان بررسی کردند. در مرحله اول الگوریتم اپریوری^۳ که یک متدولوژی قواعد وابستگی است، برای پیش‌بینی سرمایه‌گذاری‌های احتمالی، سپس الگوریتم k-means را به منظور خوشه‌بندی سهام برای اطلاعات سرمایه‌گذاری اجرا کردند. با انجام این کار، این مقاله چندین گزینه جایگزین برای پرتفوی سهام بورس اوراق بهادار تایوان را در شرایط مختلف ارائه داد [۵۲].

ناندا و همکاران^۴ (۲۰۱۰) در پژوهش خود اقدام به خوشه‌بندی سهام بورس بمبئی برای سال ۲۰۰۷-۲۰۰۸ کردند و الگوریتم FCM، k-Means و SOM را برای تحقیق خود بکار گرفتند. از طبقه‌بندی سهام ایجاد شده خوشه‌هایی را برای تشکیل پرتفوی انتخاب نمودند. آنها فاکتورهای نسبت قیمت به سود هر سهم، نسبت قیمت به ارزش دفتری، نسبت قیمت به سود نقدی هر سهم انتخاب کردند. نتیجه تحقیق آنها نشان می‌داد که الگوریتم خوشه‌بندی K-Means، خوشه‌های جمع و جورتری برای طبقه‌بندی داده‌های سهام نسبت به الگوریتم FCM و SOM ایجاد می‌کند [۵۳].

لی و همکاران^۵ (۲۰۱۱) در پژوهش خود از مدل آر.اف.ام و الگوریتم خوشه‌بندی برای طبقه‌بندی مشتریان به منظور تعیین استراتژی‌های متمایز بازاریابی در حوزه کسب و کار ساخت و تولید استفاده کردند [۵۴].

۱- K. A. J. Doherty, R. G. Adams, And N. Davey
۲- Shu-Hsien Liao, Hsu-Hui Ho, Hui-Wen Lin
۳- Apriori Algorithm
۴- S.R. Nanda, B. Mahanty, M.K. Tiwa
۵- Der-Chiang Li, Wen-Li Dai, Wan-Ting Tseng

گزار و همکاران^۱ (۲۰۱۱) در پژوهش خود یک سیستم سه مرحله‌ای پیش‌بینی بازار سهام را ارائه دادند. در مرحله اول آنالیز رگرسیون چندگانه برای تعریف متغیرهای اقتصادی و مالی که ارتباط قوی با خروجی دارند، بکار بردند. در مرحله دوم خوشه‌بندی فازی را برای ایجاد یک مدل پیش‌بینی اجرا کردند. در مرحله آخر برای پیش‌بینی قیمت آینده سهام، شبکه عصبی را اجرا کردند. نتایج شبیه سازی شبکه نشان داد که مدل پیشنهادی تحقیق آنها بهتر از مدل‌های سنتی برای پیش‌بینی قیمت بورس است [۵۵].

مارتل و همکاران^۲ (۲۰۱۲) در پژوهش خود به بهبود الگوریتم ژنتیک برای بهینه‌سازی قوانین تجارت با استفاده از داده‌ای موجود در بورس ارواق بهادار مادرید پرداختند [۵۶].

واسیل و ماریکا^۳ (۲۰۱۴) در پژوهش خود یک الگوریتم خوشه‌بندی پویا به کار رفته در داده‌های سری زمانی مالی را ارائه دادند، که این الگوریتم الهام گرفته از روش خوشه‌بندی GK است. این الگوریتم راروی داده‌های روزانه بازده سهام با تعیین اینکه چقدر این تکنیک رویدادهای سیستمیک موجود در بازار را شناسایی می‌کند که معمولاً از طریق روش‌های آماری کلاسیک آنها مشاهده نمی‌شود، برای بکار بردند. نتایج تجربی تحقیق آنها نشان داد که راندمان خوشه‌بندی بستگی به طول سری زمان، روند کلی در بازار و همچنین خصوصیات کیفی دیگر دارد [۵۷].

وانگ و همکاران^۴ (۲۰۱۴) در پژوهش خود یک الگوریتم خوشه‌بندی مبتنی بر فازی را با یک ساختار تحلیل سلسله مراتبی برای خوشه‌بندی مشتریان ارائه دادند. خصوصیات مشتریان را با استفاده از متغیرهای زبانی تحت معیارهای مهم و جزئی ارایه کردند و سپس روش انتگرال فازی را برای ترسیم معیارهای فرعی در معیارهای سلسله مراتبی بالاتر براساس اعداد فازی دوزنقه‌ای بکار بردند. یک الگوریتم

۱- David Enkea, Manfred Grauer, Nijat Mehdiyev

۲- J. Straßburg, C. González-Martel, and V. Alexandrov

۳- Vasile-George Marica

۴- Y. Wang, X. Ma, Y. Lao, and Y. Wang

خوشه‌بندی فازی براساس مجموعه فازی Axiomatic برای گروه‌بندی مشتریان به چندین خوشه توسعه دادند. شاخص اعتبار خوشه‌بندی را برای ارزیابی اثربخشی الگوریتم پیشنهادی و پیدا کردن راه‌حل بهینه خوشه‌بندی طراحی کردند. نتایج حاصل از مطالعه موردی در Anshun، نشان داد که رویکرد پیشنهادی از سه الگوریتم غالب دیگر برای حل مساله خوشه‌بندی مشتری بهتر عمل کرده است. مناطق خوشه‌بندی آزمایشی را که توسط پنج تصمیم‌گیرنده در شهر Anshun تعیین شده بود، برای ارزیابی اثربخشی رویکرد پیشنهادی مورد استفاده قرار دادند. نتایج اعتبار سنجی پژوهش آنها نشان داد که نتایج خوشه‌بندی شده از روش پیشنهادی با مناطق خوشه‌ای واقعی از دنیای واقعی همخوانی دارد [۵۸].

بس و چن^۱ (۲۰۱۴) در پژوهش خود روشی برای گسترش الگوریتم استاندارد خوشه‌بندی FCM با استفاده از توابع عضویت برای تشخیص این که چگونه مشتریان در طول زمان بین خوشه‌ها حرکت می‌کنند، توسعه دادند. این مطالعه منجر به کشف استفاده جدید و الگوهای درآمدی برای مشتریان، شناسایی دو گروه از مشتریان شد که رفتار مهاجرتی را در طول زمان نشان می‌داد، و تعیین کاربرد خاص و ویژگی‌های درآمدی که بر مهاجرت مشتری تأثیر می‌گذارند. این یافته‌ها بینش‌هایی را برای ارائه‌کنندگان خدمات تلفن همراه درباره نحوه شناسایی تغییرات موقتی در رفتار مشتری فراهم کرده است [۵۹].

واخته و آقابرگی^۲ (۲۰۱۴) یک مدل خوشه‌بندی سه‌مرحله‌ای جدید برای دسته‌بندی شرکت‌ها براساس شباهت در شکل بازارهای سهام ارائه کردند. ابتدا، داده‌های سری زمانی با وضوح پایین برای طبقه‌بندی تقریبی شرکت‌ها استفاده شدند. سپس در مرحله دوم، شرکت‌های از پیش خوشه‌بندی شده به چند زیر خوشه خالص تقسیم کردند. در نهایت، زیرخوشه‌ها در فاز سوم ادغام شدند. آنها نشان دادند

۱- Bose and X. Chen

۲- S. Aghabozorgi and Y. W. Teh

که این روش در مقایسه با الگوریتم‌های خوشه‌بندی متداول موجود، عملکرد خوبی در کارایی و اثربخشی دارد [۶۰].

متیو و بینی^۱ (۲۰۱۶) در پژوهش خود از خوشه‌بندی و رگرسیون برای کمک به سرمایه‌گذاران در انتخاب شرکت‌هایی با سوددهی بیشتر استفاده کردند. شاخص اعتبارسنجی را برای تجزیه و تحلیل عملکرد روش‌های مختلف خوشه‌بندی مانند تکنیک سلسله‌مراتبی و تفکیکی و مبتنی بر چگالی استفاده کردند. در میان تکنیک‌های مختلف خوشه‌بندی، تکنیک مبتنی بر مدل مانند خوشه‌بندی EM و تکنیک تفکیکی مانند k-means عملکرد بالاتری داشتند. برای پیش‌بینی قیمت آینده سهام جهت کمک به خریداران و فروشندگان در انتخاب شرکت‌های موجود در بورس از تکنیک رگرسیون چندگانه استفاده کردند [۶۱].

ژانگ و انکه^۲ (۲۰۱۷) یک فرآیند داده کاوی را برای پیش‌بینی جهت روزانه شاخص P & S ۵۰۰ براساس ۶۰ ویژگی مالی و اقتصادی ارائه کردند. از روش FCM در ابتدا برای خوشه‌بندی داده‌های از قبل پردازش شده استفاده کردند. تحلیل اجزای اصلی در کنار کل مجموعه داده و هر یک از هفت خوشه اعمال کردند. شبکه‌های عصبی مصنوعی و مدل‌های رگرسیون منطقی را برای طبقه‌بندی به منظور پیش‌بینی جهت روزانه بازده بازار آتی مورد استفاده قرار دادند. مطابق با نتایج تحقیق آنها، دقت پیش‌بینی شبکه‌های عصبی مصنوعی بالاتر از طبق‌بندی رگرسیون منطقی بود [۶۲].

لی^۳ (۲۰۱۹) یک روش جدید مبتنی بر خوشه‌بندی سنتی K-means ارائه کرده که نتایج تجربی در مجموعه داده‌های مختلف نشان می‌دهد که این روش نسبت به روش‌های سنتی خوشه‌بندی سری‌های زمانی چند متغیره، برتری دارد [۶۳].

۱- Bini B.Sa ,Tessy Mathew

۲- Xiao Zhong , David Enke

۳- Hailin L

لاها و ماجومدار^۱ (۲۰۲۰) روش‌های جدیدی را برای طبقه‌بندی و خوشه‌بندی سری‌های زمانی پیشنهاد کردند. یک روش خوشه‌بندی جدید SOM - TDA و یک روش طبقه‌بندی جدید RF - TDA براساس TDA ارایه کردند. با استفاده از SOM - TDA به بررسی شباهت‌ها و تفاوت‌های توپولوژیکی برخی از مدل‌های سری زمانی معروف مورد استفاده در امور مالی پرداختند. همچنین از RF - TDA برای بررسی اینکه آیا ویژگی‌های توپولوژیکی می‌توانند برای تمایز بین مدل‌های سری زمانی با استفاده از داده‌های شبیه‌سازی شده مورد استفاده قرار گیرند، استفاده کردند. عملکرد RF - TDA در کار طبقه‌بندی با سه روش طبقه‌بندی دیگر مورد مقایسه قرار دادند. همچنین کاربرد RF - TDA را برای طبقه‌بندی سری‌های زمانی مالی با استفاده از داده‌های قیمت واقعی سهام متعلق به بخش‌های مختلف در نظر گرفتند. RF - TDA در دو آزمایش مبتنی بر داده‌های واقعی قیمت سهام به خوبی عمل کرد. این امر نشان می‌دهد که ویژگی‌های توپولوژیکی سری‌های زمانی قیمت‌های سهام در بخش‌های مختلف یک‌سان نیستند و ویژگی‌های متمایزی دارند که می‌توانند با استفاده از TDA تشخیص داده شوند [۶۴].

کومار و همکاران^۲ (۲۰۲۰) به بررسی تکنیک‌های آماری و یادگیری ماشین مورد استفاده برای پیش‌بینی بازار سهام در ۳۰ مقاله پرداختند. نتیجه تحقیق آنها استفاده بیشتر روش‌های شبکه‌های عصبی و شبکه‌های عصبی مصنوعی می‌باشد [۶۵].

همیچه و مغار^۳ (۲۰۲۰) در پژوهش خود به ساخت مدلی با استفاده از شبکه‌های عصبی بازایی شده (RNN) و (LSTM) برای پیش‌بینی ارزش بازار سهام آینده پرداختند [۶۶].

لو و ژانگ^۴ (۲۰۲۱) اعتبار و برتری مدل پیش‌بینی شبکه عصبی الگوریتم BP را در پژوهش خود نشان دادند. هدف از این کار استفاده از شبکه عصبی و الگوریتم BP در طبقه‌بندی و پیش‌بینی الگوهای

۱- Sourav Majumdar, Arnab Kumar Laha

۲- Deepak Kumar, Pradepta Kumar Sarangi, Rajit Verma

۳- Adil Moghara, Mhamed HAMICHE

۴- D. Zhang and S. Lou

قیمت سهام است. نتایج نشان داد که دقت پیش‌بینی قیمت سهام تحت پیش‌بینی الگوریتم فازی یادگیری عمیق ۶۲.۱۲٪ و تحت پیش‌بینی شبکه عصبی الگوریتم BP ۷۳.۲۹٪ است. بنابراین پیش‌بینی مدل پیش‌بینی روند قیمت سهام شبکه عصبی الگوریتم BP بهتر از مدل پیش‌بینی الگوریتم فازی یادگیری عمیق است [۶۷].

۱۷-۲-۲ خلاصه و جمع‌بندی فصل

در فصل حاضر در بخش اول به تفصیل به بررسی مفاهیم نظری استفاده شده در پژوهش پرداخته شد که درک درست آنها اهمیت کاربرد خوشه‌بندی در بازار سهام را آسان‌تر می‌نماید. در بخش دوم نیز پیشینه‌ای از پژوهش‌های انجام شده توسط پژوهشگران مختلف در داخل و خارج از کشور در رابطه با خوشه‌بندی بازار سهام ارائه شده است.

فصل ۳ متدولوژی تحقیق

۳-۱ مقدمه

انتخاب روش پژوهشی مناسب یکی از مهمترین مراحل یک پژوهش علمی برای انجام آن است. هدف از انتخاب روش تحقیق آن است تا محقق مشخص نماید چه شیوه و روشی را اتخاذ می‌کند تا او را هرچه آسان‌تر، دقیق‌تر و سریع‌تر در دستیابی به پاسخ‌هایی که برای پرسش تحقیق در نظر گرفته شده یاری کند. هر تحقیقی برای دستیابی به اهداف خاصی صورت می‌گیرد این هدف در واقع خود را در قالب مسئله تحقیق نمودار می‌سازد و از طریق بیان آن آشکار می‌شود [۶۸].

در راستای بیان روش علمی تحقیق حاضر، فرایند خوشه‌بندی و متغیرهای پژوهش را بیان خواهد شد. سپس الگوریتم خوشه‌بندی پیوند کامل را توضیح داده و دلایل انتخاب ضریب تشابه تطابق نسبی مطرح می‌گردد. همچنین روش‌های آماری مورد استفاده برای آزمون فرضیه تحقیق نیز معرفی و تشریح می‌شوند.

۳-۲ فرایند خوشه‌بندی

فرایند خوشه‌بندی با توجه به معیار خاصی که برای خوشه‌بندی انتخاب می‌شود منجر به تقسیم‌بندی متفاوتی از مجموعه داده‌ها می‌گردد. بنابراین، نیاز به یک پیش‌پردازش قبل از انجام خوشه‌بندی است. گام‌های اساسی برای انجام فرایند خوشه‌بندی به طور خلاصه می‌تواند به صورت زیر باشد:

انتخاب ویژگی

هدف آن است که به درستی ویژگی‌هایی که روی آنها خوشه‌بندی انجام می‌شود انتخاب شود، بنابراین قبل از انجام خوشه‌بندی شاید نیاز به پیش‌پردازش داده‌ها باشد.

الگوریتم خوشه‌بندی

این مرحله اشاره به انتخاب الگوریتم خوشه‌بندی دارد که منجر به تعریف یک خوشه‌بندی خوب از مجموعه داده می‌شود. به طور عمده اندازه‌گیری مجاورت و معیار خوشه‌بندی، یک الگوریتم خوشه‌بندی و کارایی آن برای انجام خوشه‌بندی مناسب روی مجموعه داده‌ها را مشخص می‌کند.

اندازه‌گیری مجاورت: معیاری است که توسط آن میزان شباهت بین دو نقطه مشخص می‌شود. در بیشتر موارد باید از انتخاب ویژگی‌هایی که به طور برابر در محاسبه اندازه مجاورت دخیل هستند، اطمینان پیدا شود.

معیار خوشه‌بندی

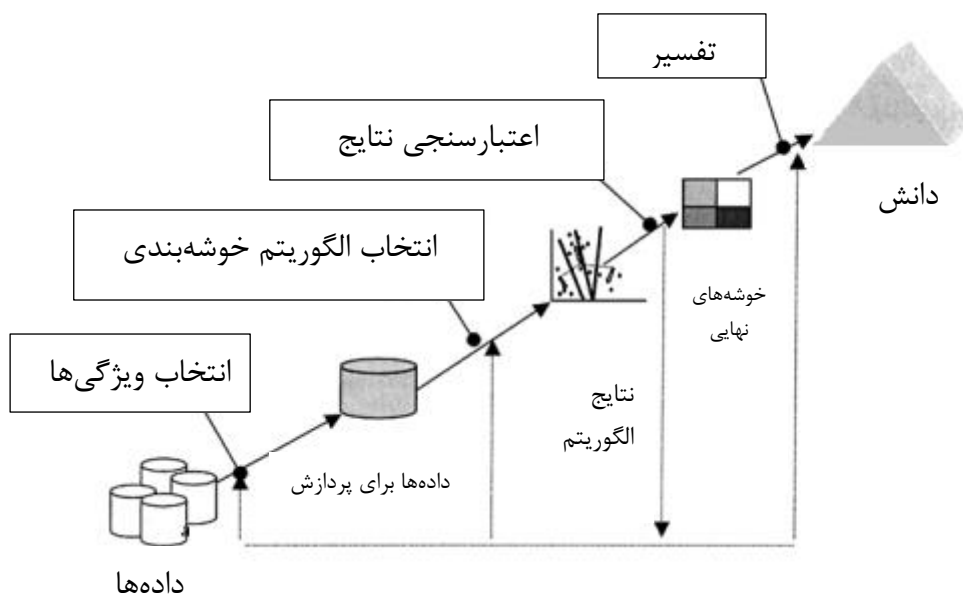
در این گام باید معیار خوشه‌بندی تعریف شود که می‌تواند به وسیله یک تابع هزینه یا دیگر قوانین بیان گردد. بنابراین باید یک معیار معقول خوشه‌بندی تعریف شده که منجر به تقسیم‌بندی مناسب مجموعه داده‌ها می‌شود.

اعتبار سنجی نتایج

درستی نتایج الگوریتم خوشه‌بندی با استفاده از معیار و تکنیک‌های مناسب بررسی می‌شود. از آنجاییکه الگوریتم‌های خوشه‌بندی فارغ از نوع روش‌های خوشه‌بندی، خوشه‌هایی را که قبلاً ناشناخته بوده‌اند تعریف می‌کند، بنابراین در بسیاری از کاربردها نیاز به ارزیابی تقسیم‌بندی داده‌ها است.

تفسیر نتایج

در بسیاری از موارد، متخصصان مجبور به تلفیق نتایج خوشه‌بندی با سایر شواهد تجربی و تجزیه و تحلیل آنها هستند تا منجر به یک نتیجه‌گیری صحیح شود [۶۹]. شکل ۳-۱ فرایند خوشه‌بندی را نشان می‌دهد.



شکل ۳-۱. مراحل فرایند خوشه‌بندی

۳-۱-۲-۱ روش خوشه‌بندی مورد استفاده در این پژوهش

تکنیک خوشه‌بندی مورد استفاده در این پژوهش، الگوریتم خوشه‌بندی سلسله مراتبی می‌باشد، که با پیوند اشیاء با بیشترین تشابه در هر گام شروع شده، به تدریج خوشه‌هایی از اشیاء را ایجاد می‌کند. چندین روش مختلف خوشه بندی سلسله مراتبی در دسترس است که در این میان خوشه بندی پیوند کامل به عنوان مناسب‌ترین روش برای مدل ما یافت شده است.

خوشه بندی پیوند کامل خوشه بندی مبتنی بر ضریب شباهت است. این الگوریتم با محاسبه ضرایب شباهت بین جفت های اشیاء شروع می شود. برای محاسبه ضریب تشابه بین گروه‌های اشیاء از پایین‌ترین سطح شباهت استفاده می‌گردد.

۳-۲-۱ الگوریتم خوشه‌بندی پیوند کامل

در الگوریتم خوشه‌بندی پیوند کامل ابتدا ضریب تشابه برای هر جفت شی اندازه‌گیری شده، سپس ماتریس شباهت تشکیل می‌شود. بعد از ایجاد ماتریس، الگوریتم، اشیایی با بالاترین ضریب تشابه را باهم گروه‌بندی می‌کند و این چرخه تا زمانی که حداکثر مقدار ضریب تشابه برای اشیایی که به یک خوشه اختصاص داده نشده‌اند، کمتر از آستانه از پیش تعیین شده باشد یا تعداد خوشه‌های از پیش تعریف شده بدست آید، تکرار می‌شود.

۱. در زیر الگوریتم گام به گام نشان داده شده است:
۲. تشکیل ماتریس شباهت با محاسبه ضریب شباهت برای هر جفت از اشیاء
۳. پیدا کردن اشیایی با بالاترین ضریب تشابه و گروه‌بندی آنها با یکدیگر
۴. ردیف‌هایی که با گروه‌های اشیاء که گروه‌بندی شده‌اند، مطابقت دارد، حذف کنید.
۵. یک ردیف جدید به ماتریس برای گروه اشیاء جدید اضافه کنید و ضرایب تشابه را با استفاده از رابطه (۳-۱) حساب کنید:

$$S_{tv} = \text{MIN}\{S_{mn}\} \quad m \in t \ \& \ n \in v \quad (1-3)$$

که t گروه اشیاء جدید و v اشاره به سایر گروه‌های اشیاء دارد.

از مزایای خوشه‌بندی پیوند کامل آن است که از ادغام دو خوشه باهم به دلیل بالا بودن سطح تشابه دو عضو باهم در حالیکه مابقی اعضاء شباهتی ندارند، جلوگیری می‌کند.

۳-۲-۱-۳ مهمترین اجزاء برای اکثر روش‌های خوشه‌بندی عبارتند از: اشیاء، ویژگی‌ها، ضریب

شباهت

اشیاء: از مدل خوشه‌بندی برای گروه‌بندی اشیا به روشی استفاده می‌شود که اشیاء با بیشترین شباهت در یک گروه قرار داشته که با اشیاء در یک گروه دیگر متفاوت هستند. در این پژوهش اشیاء، ۴۲ شرکت مورد مطالعه در بورس هستند.

ویژگی‌ها: برای مقایسه اشیا با یکدیگر و اندازه‌گیری شباهت آنها، به یک مجموعه از ویژگی‌ها نیاز است. در این پژوهش خوشه‌بندی جامعه آماری براساس شاخص‌های نقدشوندگی است.

شاخص‌های نقدشوندگی که در این پژوهش در نظر گرفته شده عبارتند از:

(۱) گردش سهام^۱ (TO): این شاخص فراوانی یا تعداد دفعات تکرار معاملات دارایی را اندازه‌گیری می‌کند و به بعد عمق و زمان نقدشوندگی اشاره دارد.

گردش سهام از نسبت سهام معامله شده به سهام منتشر شده بدست می‌آید [۷۰]:

$$TO_{i,t} = \frac{vol_{i,t}}{share_{i,t}} \quad (2-3)$$

در رابطه (۳-۲) $vol_{i,t}$ سهام i مبادله شده در ماه t ، $share_{i,t}$ سهام i منتشر شده در ماه t است. هر مقدار نسبت تعداد سهام معامله شده به تعداد سهام منتشر شده بزرگتر باشد، نقدشوندگی بازار نیز بیشتر است و برعکس. بنابراین بالاتر بودن نرخ گردش سهام بیانگر نقدشوندگی بیشتر آن است.

(۲) اختلاف قیمت پیشنهادی خرید و فروش نسبی سهام (PBA)^۲: این شاخص پر استفاده‌ترین معیار در ادبیات نقدشوندگی دارایی‌ها است قیمت پیشنهادی خرید، قیمت اوراق بهادری است که توسط بازارساز خریداری می‌شود، و قیمتی که با آن اوراق بهادر را می‌فروشند، قیمت پیشنهادی فروش نامیده

۱-Turn Over

۲- Proportional Bid-Ask Spread

می‌شود. اختلاف پیشنهادی خرید و فروش، اختلاف بین این دو قیمت خواهد بود. هنگامی که بیشتر از یک بازار ساز وجود دارد، تفاوت بین بالاترین قیمت پیشنهادی خرید و پایین‌ترین قیمت پیشنهادی فروش را شکاف قیمت پیشنهادی خرید و فروش می‌نامند و نقطه تعادلی بین قیمت پیشنهادی خرید و فروش قرار می‌گیرد [۷۱]. اختلاف قیمت پیشنهادی خرید و فروش مطلق یا واقعی سهام از تفاضل قیمت پیشنهادی فروش قیمت پیشنهادی خرید سهام بدست می‌آید. اختلاف قیمت پیشنهادی خرید و فروش نسبی سهام که در این پژوهش مورد استفاده قرار گرفته است با استفاده از رابطه (۳-۳) بدست می‌آید:

$$PBA_{i,t} = \frac{1}{D_{i,t}} \sum_{d=1}^{D_{i,t}} (AP_{i,t} - BP_{i,t}) / (0.5AP_{i,t} + 0.5BP_{i,t}) \quad (3-3)$$

که در آن $AP_{i,t}$ بهترین پیشنهاد قیمتی فروش و $BP_{i,t}$ بهترین قیمت پیشنهادی خرید سهام i در روز t و D تعداد روزهای معامله سهام i در ماه t است.

هرچه اختلاف قیمت پیشنهادی خرید و فروش سهام کمتر باشد، خریدار و فروشنده زودتر به قیمت تعادلی می‌رسند و حجم معاملات سهام افزایش می‌یابد. این امر منجر به افزایش نقدینگی می‌شود [۷۱]. اختلاف قیمت پیشنهادی سهام به دلیل سه نوع هزینه است: هزینه‌های انتخاب نادرست، هزینه‌های اجرای سفارش و هزینه‌های نگهداری موجودی. هزینه‌های انتخاب نادرست که به عنوان بخشی مهم در اختلاف قیمت پیشنهادی خرید و فروش سهام ناشی از عدم تقارن اطلاعات در مورد سهام شرکت هستند [۷۲].

اختلاف قیمت پیشنهادی خرید و فروش به عنوان معیار نقدشوندگی در پژوهش‌های نخستین استفاده می‌شد که براساس آن سهام‌هایی با قیمت پایین‌تر، به عنوان سهام‌هایی با قدرت نقدشوندگی بالاتر معرفی می‌شدند، چون به صورت طبیعی سهام‌های با قیمت پایین‌تر اختلاف قیمت خرید و فروش

پایین‌تری داشتند. این معیار نسبی بر این مشکل که اختلاف قیمت خرید و فروش یک تابع صعودی از قیمت سهام است غلبه می‌کند [۷۰].

۳) نسبت عدم نقدشوندگی^۱: ژانگ^۲ (۲۰۱۰) که در پژوهش خود به بررسی شاخص‌های نقدشوندگی در بازارهای نوپا پرداخته، معتقد است برخلاف معیارهای آمیهود^۳ و نسبت حجم صفر^۴، معیارهای آمیوست^۵ و گاما^۶، توسط سایر معیارها مغلوب می‌شوند و مناسب نیستند. فونگ و همکاران^۷ و همچنین کیم و لیو^۸ استفاده از دو معیار عدم نقدشوندگی آمیهود و نسبت حجم صفر را در بازارهای نوپا بلامانع و مناسب یافتند [۷۳، ۷۴، ۷۵].

آمیهود نسبت عدم نقدشوندگی را به صورت قدر مطلق بازده روزانه یک سهم تقسیم بر حجم معامله آن در همان روز محاسبه کرد. این مقدار که رابطه عکس با نقدشوندگی سهام دارد با استفاده از رابطه (۴-۳) بدست می‌آید:

$$\text{Illiquid}_{i,t} = T_y^{-1} \sum |r_{i,t,y}| / \text{vol}_{i,t,y} \quad (4-3)$$

که $|r_{i,t,y}|$ قدر مطلق بازده سهام i در روز t در سال y ، $\text{vol}_{i,t,y}$ حجم معامله سهام i در روز t در سال y است. مجموعه روزها برابر T_y است، یعنی تعداد روزهای سال y که برای آن $\text{vol}_{i,t,y}$ محاسبه می‌شود.

۱- Illiquidity Ratio

۲- Huiping Zhang

۳- Amihud

۴- Zero Vol

۵- Amivest

۶- Gamma

۷- Fongtkingsley, Holden, Caig W & Trzcinka, Charles A

۸- Sunho Kim & Kuan-Hui Lee

نسبت عدم نقدشوندگی حرکت قیمت در ارتباط با حجم معامله را نشان می‌دهد.

آمیهد نشان داد که نسبت عدم نقدشوندگی رابطه مسقیم با بازده سهام دارد [۷۶]. گوینکو و همکاران^۱ نشان دادند که معیار آمیهد تأثیرات قیمت را به خوبی نشان می‌دهد [۷۷]. برای روزهای معاملاتی با حجم صفر معیار آمیهد، غیر قابل تعریف است. بنابراین نیاز به حضور معیار دیگری برای پوشش این ضعف احساس می‌شود. لسموند و همکاران^۲ نسبت تعداد روزهای بازده صفر سهام در ماه، به تعداد کل روزهای معاملاتی آن سهام در همان ماه به عنوان معیار مبتنی بر بازده صفر پیشنهاد دادند [۷۸]. لسموند و همکاران بیان می‌کنند که بازده صفر زمانی اتفاق می‌افتد که معامله‌گران، معاملات خود را

به دلیل عدم اطلاعات کافی در مورد جبران هزینه‌های معاملات، کاهش داده و یا معامله نمی‌کنند. بنابراین هیچ تغییر قیمتی از نسبت به روز قبل وجود نخواهد داشت. به عبارت دیگر معامله‌گران در صورت هزینه‌های بالا معاملات، تصمیم به عدم انجام معامله گرفته و بنابراین قیمت بدون تغییر خواهد ماند. این معیار برای یک روز با حجم معاملاتی مثبت که به دلیل وجود هزینه‌های معاملاتی بالا بازده صفر کسب کرده است، مناسب نمی‌باشد. بکارت و همکاران^۳ به منظور رفع این نقص نسبت مبتنی بر صفر دیگری تحت عنوان حجم صفر ارائه کردند، که تعداد روزهای حجم صفر یک سهم در ماه به تعداد کل روزهای معاملاتی آن سهم در ماه است و به zero vol معروف می‌باشد [۷۹].

در سال ۲۰۱۴ ژانگ که یکی از محققان برجسته در زمینه نقدشوندگی می‌باشد، معیار جدیدی تحت عنوان معیار اصلاح شده عدم نقدشوندگی آمیهد^۴ که با نماد AdjILLIQ نشان داده می‌شود ارائه داد [۷۴]. وی در سال ۲۰۱۰ در پژوهش خود در سال به این نتیجه دست یافته بود که در بازارهای نوظهور معیارهای آمیهد و نسبت حجم صفر به منظور اندازه‌گیری نقدشوندگی، بهترین عملکرد را دارند،

۱- R. Y. Goyenko, C. W. Holden, and C. A. Trzcink

۲- Lesmond, Ogden, Trzcinka

۳- Bekaert, Harvey, Lundblad

۴- Modified Version Of The Amihud Illiquidity Measure

دست به تلفیق این دو معیار زد. معیار جدید که تلفیق معیار اصلی آمیهود و فراوانی حجم صفر می باشد برای انواع بازارهای نوظهور به خصوص بازارهایی که تعداد معاملات کم و سهام با نرخ گردش پایین به خوبی عمل کرده است و همبستگی بسیاری با اختلاف قیمت پیشنهادی خرید و فروش و تأثیر قیمت دارد، از رابطه (۳-۵) بدست می آید:

$$AdjILLQ_{i,m} = \left[\ln \left(\frac{1}{N_{i,m}} \sum_{t=1}^{N_{i,m}} \frac{|R_{i,t}|}{VOL_{i,t}} \right) \right] \times (1 + Zero\ Vol_{i,m}) \quad (5-3)$$

که در این رابطه $N_{i,m}$ تعداد روزهایی از ماه m است که سهام i مورد معامله قرار گرفته و حجم غیرصفر دارد $|R_{i,t}|$ مقدار بازده سهام i در روز t ، $vol_{i,t}$ حجم ریالی سهام i در روز t و $zero\ vol_{i,m}$ درصد روزهای حجم صفر سهام در ماه m است.

۴) تعدیل تعداد روزهای بدون معامله براساس گردش $(LM)^1$: این معیار با هدف بدست آوردن لبعاد گوناگون نقدشوندگی توسط لیو پیشنهاد شده و روی سرعت معاملات تمرکز ویژه ای دارد. این معیار به صورت رابطه (۳-۶) تعریف می شود:

$$LM_{i,t} = \left[NoZV_{i,t} + \frac{1}{480000} \frac{Turnover_{i,t}}{NoTD_t} \right] \times \frac{21}{NoTD_t} \quad (6-3)$$

که $NoZV_{i,t}$ تعداد روزهای بدون معامله سهام i در ماه t ، $turnover_{i,t}$ نرخ گردش سهام i در

ماه t ، $NoTD_{i,t}$ تعداد کل روزهای معامله در بازار در ماه t را نشان می دهد. ضرب در عامل $\frac{21}{NoTD_t}$

تعداد روزهای معامله در یک ماه را به عدد ۲۱ استاندارد کرده، بنابراین این معیار را در طول زمان

قابل مقایسه می کند. عدد ۴۸۰۰۰۰ در مخرج توسط لیو پیشنهاد شده تا مقدار عبارت $\frac{1}{480000} \frac{Turnover_{i,t}}{NoTD_t}$

۱- Turnover-Adjusted Number Of Zero Daily Volume

۲-Liu

بین صفر و یک قرار گیرد. NoZV خود به عنوان شاخص نقدشوندگی بکار می‌رود، به این معنی که هرچه تعداد روزهای بدون معامله کمتر باشد، معاملات سهام به صورت متناوب انجام شده و بنابراین سهام نقدشوندگی بیشتری دارد. این امر پیوستگی انجام معاملات را منعکس کرده و تأخیر بالقوه در اجرای یک معامله را نشان می‌دهد [۱۵].

از سوی دیگر نقش NoZV مشابه تعداد بازده‌های روزانه صفر در معیار ارائه شده توسط لسموند و همکاران بوده و از این رو بعد هزینه معاملات را نیز منعکس می‌کند. سهام‌هایی که طبق این معیار نقدشوندگی بالایی دارند، آن‌هایی هستند که به صورت متناوب معامله شده و در ماه مربوطه گردش بالایی دارند [۷۸].

۵) حجم معاملات (TV)^۱: این شاخص از نظر فلمینگ^۲ از معیارهای نقدشوندگی است و از منابع

اطلاعاتی برای معامله‌گران می‌باشد [۸۰]، حجم معاملات از حاصلضرب قیمت معامله سهم در حجم معاملات بدست می‌آید. حجم معاملات رابطه مستقیمی با نقدشوندگی سهام دارد [۸۱].

۶) تعداد معاملات: تعداد معاملات در واحد زمان با $N_{i,t}$ بیان می‌شود که مقادیر بیشتر آن، نقدشوندگی بهتر سهام را بیان می‌کند.

۷) درصد روزهای انجام معاملات^۳ (PTD): این شاخص از نسبت تعداد روزهایی که سهم مورد نظر حداقل یکبار داد و ستد شده است بر کل روزهایی که بورس در سال مورد نظر فعالیت معامله‌ای دارد، به دست می‌آید.

۱- Transaction volum

۲-Felming

۳- Percentage of transaction days

۸) شکاف قیمتی مؤثر (ES)^۱ از معیارهای عدم نقدشوندگی است و به صورت عکس تفسیر می‌گردد. از دو برابر قدر مطلق اختلاف آخرین قیمت معامله روز و قیمت متوسط تقسیم بر قیمت متوسط بدست می‌آید. این شاخص به صورت روزانه برای هر شرکت محاسبه و سپس برای هر سال از میانگین شاخص روزانه بدست می‌آید [۸۲]. رابطه (۷-۳) نحوه محاسبه این شاخص را نشان می‌دهد.

$$ES = 2 \times \frac{|Price - M|}{M} \quad (7-3)$$

۹) اندازه شرکت (SIZE): این شاخص حجم و گستردگی فعالیت یک شرکت را تعیین می‌کند. شاخص‌هایی مانند میزان دارایی‌ها، ارزش دفتری سرمایه شرکت، ارزش بازار، تعداد کارکنان، سرمایه شرکت به عنوان شاخص اندازه شرکت به کار می‌روند [۸۳]. از دیدگاه نظری رابطه مثبتی بین اندازه شرکت و نقدشوندگی وجود دارد. در این پژوهش این شاخص نظیر کیل و لون^۲ از طریق لگاریتم طبیعی جمع ارزش دفتری کل دارایی‌ها بدست می‌آید [۸۲].

۱۰) اهرم (LEV)^۳: اهرم مالی اشاره به افزایش سود سرمایه‌گذاری از طریق ابزارهای مالی یا بدهی دارد. یک شرکت یا سرمایه‌گذار می‌تواند به غیر از دارایی‌ها و سرمایه‌اش، بدون ایجاد سرمایه‌گذار جدید، از طریق قرض گرفتن (ایجاد بدهی)، سرمایه و فعالیت‌های مالی خود را افزایش دهد.

اهرم مالی سالیان زیادی است که در بخش‌های مختلف بازارهای مالی مورد استفاده قرار می‌گیرد. هرچه میزان وام و قرض دریافتی بیشتر باشد، شرکت می‌تواند سرمایه بیشتری را خریداری نماید در مقابل ریسک بیشتری را برای بازپرداخت وام و قرض متحمل خواهد شد. پرس^۴ از نظر تجربی و نظری نشان می‌دهد اهرم مالی بیشتر باعث افزایش درجه ریسک‌پذیر بودن شرکت می‌شود و سرمایه‌گذاران

۱- Effective Spread

۲- Kale, J.R., Loon, Y.C

۳- Leverage

۴- Press

ریسک‌گریز را از معامله بازمی‌دارد، بنابراین سبب کاهش نقدشوندگی سهام می‌شود [۸۴]. این شاخص از تقسیم کل بدهی شرکت، بر کل دارایی‌ها بدست می‌آید [۸۲].

۱۱) عملکرد شرکت (CP) اقداماتی در جهت حداکثرسازی ثروت سهامداران که با معیارهای متفاوتی از جمله بازده دارایی‌ها، بازده حقوق صاحبان سهام، ارزش افزوده اقتصادی و ارزش افزوده نقدی محاسبه می‌شود. محققان حسابداری و مدیریت مالی برای اندازه‌گیری عملکرد شرکت شاخص‌های متفاوتی را ارائه کرده‌اند. در سال ۱۹۶۹ شاخص Q-توبین که یکی از معتبرترین شاخص‌ها می‌باشد، توسط جیمز توبین^۲ معرفی گردید. معیاری که در این پژوهش نیز به عنوان معیار عملکرد در نظر گرفته شده است شاخص Q-توبین است. این شاخص از طریق رابطه (۸-۳) بدست می‌آید:

$$Q = \frac{\text{ارزش بازاری شرکت}}{\text{ارزش دفتری دارایی‌ها}} \quad (۸-۳)$$

که این رابطه برابر است با:

$$\frac{\text{COMVAL} + \text{SBOND} + \text{STDEB}}{\text{SRC}} \quad (۹-۳)$$

در رابطه (۳-۹) COMVAL نشانگر ارزش بازاری سهام عادی، SBOND ارزش دفتری بدهی‌های بلندمدت و STDEB به ترتیب ارزش دفتری بدهی‌های کوتاه مدت است. SRC نیز ارزش دفتری کل دارایی‌های شرکت است. برای محاسبه ارزش بازاری سهام، میانگین موزون تعداد سهام شرکت در طی سال محاسبه می‌گردد و سپس قیمت پایانی سهام در آخرین روز معاملاتی در پایان هر سال استخراج گردیده و سپس با ضرب تعداد سهام در قیمت پایانی، ارزش بازاری هر شرکت به صورت جداگانه محاسبه می‌شود. ارزش دفتری بدهی‌های بلندمدت و کوتاه مدت نیز از صورت‌های مالی حسابرسی شده شرکت‌ها استخراج خواهد شد [۸۵].

۱- Companys Performance

۲- James Tobin

فروغی، فرهمند و ابراهیمی (۱۳۹۰) رابطه نقدشوندگی سهام و عملکرد شرکت را طی دوره زمانی ۱۳۸۱ تا ۱۳۸۷ در بورس اوراق بهادار تهران بررسی کردند. در این پژوهش نقدشوندگی با استفاده از معیار اختلاف قیمت پیشنهادی خرید و فروش و عملکرد شرکت با شاخص Q-توبین محاسبه شده است. نتایج تحقیق آنها نشان از رابطه مثبت و معنادار بین نقدشوندگی سهام و عملکرد شرکت دارد [۸۶].

۱۲) ضریب انعطاف پذیری تجارت^۱: فوریت زمان اجرای مورد نیاز برای یک معامله را نشان می دهد، که این بستگی به تمایل طرفین معامله برای اجرای مقدار اعلام شده اوراق بهادار با قیمت تعیین شده و بدون تأخیر در زمان دارد. در این مطالعه از ضریب انعطاف پذیری تجارت همانطور که وانزالا^۲ پیشنهاد کرده است [۸۷] استفاده می شود که به درستی سرعت اجرای معامله را نشان می دهد:

$$CET = \frac{\% \Delta T}{\% \Delta P} \quad (۱۰-۳)$$

جایی که ΔT / بیانگر درصد تغییر در حجم معاملات روزانه سهام و ΔP درصد تغییر در قیمت پایانی روزانه را نشان می دهد. ضریب انعطاف پذیری بیشتر معاملات نشان دهنده فوریت بالاتر و در نتیجه نقدینگی بیشتر است.

در جدول ۱-۳ شاخص های مورد استفاده در پژوهش به طور خلاصه بیان شده است:

۱- Coefficient of Elasticity of Tradin

۲-R. W. Wanzala

جدول ۳-۱. شاخص‌های نقدشوندگی مورد استفاده در پژوهش

| شاخص نقدشوندگی | مخفف | روش محاسبه |
|---|--------------------|--|
| $TO_{i,t} = vol_{i,t} / share_{i,t}$ | TO | گردش سهام |
| $PBA_{i,t} = \frac{1}{D_{i,t}} \sum_{d=1}^{D_{i,t}} (AP_{i,t} - BP_{i,t}) / (0.5AP_{i,t} + 0.5BP_{i,t})$ | PBA _{i,t} | اختلاف قیمت پیشنهادی خرید و فروش |
| $AdjILLQ_{i,m} = [\ln(\frac{1}{N_{i,m}} \sum_{t=1}^{N_{i,m}} \frac{ R_{i,t} }{VOL_{i,t}})] \times (1 + Zero Vol_{i,m})$ | AdjILLQ | معیار ژانگ |
| $LM_{i,t} = [NoZV_{i,t} + \frac{1}{480000} \frac{Turnover_{i,t}}{NoTD_t}] \times \frac{21}{NoTD_t}$ | LM | تعدیل تعداد روزهای بدون معامله براساس گردش |
| حجم معاملات × قیمت معامله سهام | TV | حجم معاملات |
| تعداد معاملات در واحد زمان | N _{i,t} | تعداد معاملات |
| از تقسیم تعداد روزهایی که سهم مورد نظر حداقل یکبار داد و ستد شده است بر کل روزهایی که بورس در سال مورد نظر فعالیت معامله‌ای دارد. | PTD | درصد روزهای انجام معامله |
| $2 \times \frac{ price - M }{M}$ | ES | شکاف قیمتی مؤثر |
| از حاصلضرب قیمت سهام شرکت در پایان سال در تعداد سهام شرکت در پایان سال بدست می‌آید. | SIZE | اندازه شرکت |
| از کل بدهی شرکت تقسیم بر ارزش بازار سرمایه به علاوه ارزش دفتری کل بدست می‌آید. | LEV | اهرم |
| $\frac{COMVAL + SBOND + STDEB}{SRC}$ | CP | عملکرد شرکت |
| $\frac{\% \Delta T}{\% \Delta P}$ | CET | ضریب انعطاف‌پذیری تجارت |

ضریب تشابه: در تجزیه و تحلیل خوشه‌ای، شباهت دو شی یا خوشه‌ها در ضرایب تشابه محصور شده اند. ضرایب تشابه توابع ریاضی هستند که شباهت دو شی را بر اساس مقادیر ویژگی‌ها اندازه‌گیری می‌کنند. ضرایب شباهت بسیاری توسط محققان ارائه شده است. ضریب شباهت مورد استفاده در این پژوهش ضریب شباهت تطابق نسبی است.

۳-۲-۱-۴ دلایل انتخاب ضریب تشابه تطابق نسبی

برای نشان دادن میزان تشابه بین دوشی i و j می‌توان از ضریب تشابه (S_{ij}) استفاده کرد. محققان ضرایب تشابه متفاوتی را برای آنالیز خوشه بندی پیشنهاد کرده اند. اما این ضرایب در بیشتر موارد مناسب مشکلات خاص هستند. برای ارزیابی ضریب تشابه بارونی و بوسر^۱ (۱۹۷۶) مجموعه‌ای از خواص ضریب تشابه را تعریف کرده اند. همان‌طور که از ادبیات مشهود است تعداد زیادی ضریب تشابه برای آنالیز خوشه‌بندی وجود دارد و گروهی از این ضرایب تشابه توسط رومسبرگ^۲ (۱۹۸۴) و اسنیت و سوکال^۳ (۱۹۷۳) ارائه شده است.

در میان این ضرایب تشابه، ضریب تشابه جاکارد^۴ (۱۹۰۸) اولین مورد است. خصوصیت این ضریب آن است که وقتی دو شی دارای ویژگی‌های یکسانی هستند، حداکثر شباهت بین آنها یک است. این ضریب به جهت کدگذاری حساس است یعنی در صورت تعویض عناصر یک و صفر در ماتریس وقوع، موجب تغییر شباهت بین اشیاء می‌شود. برخی از ضرایب مانند ضریب هامن^۵ نسبت به جهت کدگذاری حساس نمی‌باشد. بیشتر ضرایب تشابه به صورت نسبت بیان می‌شوند. در میان ضرایب تشابه، دو نوع ضریب وجود دارد، که جاکاردی و غیر جاکاردی^۶ است. که در نوع جاکاردی، ضریب تشابه براساس سطوحی از تطابق بیان می‌شود، که تعداد تطابق‌ها تقسیم بر تعداد تطابق‌های مورد انتظار می‌شود. در نوع غیر جاکاردی یک مقدار دیگر تحت عنوان عدم تطابق از صورت کسر کم شده و تقسیم بر تعداد تطابق‌های مورد انتظار می‌شود. مقدار ضریب تشابه در نوع جاکاردی بین صفر و یک و در نوع غیر جاکاردی بین یک و منفی یک است. هر ضریب تشابه چه جاکاردی و چه غیر جاکاردی باید از یک مجموعه ویژگی‌ها برای تشکیل بهتر خوشه پیروی کند. برای تعریف این ویژگی‌ها، برخی مفاهیم در جدول ۲-۳ تعریف شده است.

۱- Baroni-Urbani and Buser

۲- Romesburg

۳- Sneath and Soka

۴- Jaccard coefficient

۵- Hamann coefficient

۶- Jaccardian and non-Jaccardian

ناتوان است. زیرا این ضریب تعداد تطابق‌های (۰-۰) را در نظر نمی‌گیرد، که همان متغیر d در صورت است. این توضیح نیز برای ضرایب سورنسون^۱، اوچیایی^۲، آر یوسل^۳ و آر آو^۴ صدق می‌کند.

اکنون ماتریس وقوع A_3 که مجموعه متغیرهای آن (2,4,4,2) است در نظر بگیرید:

$$A_3 = \begin{bmatrix} 011010110100 \\ 110100100011 \end{bmatrix} \begin{matrix} 1 \\ 2 \end{matrix}$$

مقدار ضریب تشابه آریوسل و آر آو بین شی ۱ و ۲ برای ماتریس وقوع A_3 ، ۰.۱۷ است. که منعکس کننده صحیح مقدار شباهت نیست زیرا تعداد عدم تطابق‌ها (۰-۱ و ۱-۰) در ماتریس A_3 بیشتر از ماتریس A_2 است.

تعداد تطابق‌های (۱-۱) در هر دو ماتریس A_2 و A_3 برابر است. بنابراین مقدار ضریب تشابه برای ماتریس A_3 باید کمتر از ماتریس A_2 باشد. دلیل این عدم تطابق در نظر نگرفتن متغیر d در صورت کسر ضریب RR است.

برای ماتریس وقوع A_1 مقدار شباهت بین شی i و j برای ضرایب تشابه اشاره شده در جدول (۳-۳) بین ۰.۳۳ تا ۰.۸ است. برای ماتریس وقوع A_2 این مقدار بین ۰.۱۷ تا ۰.۹۱ است. یک انسجام باید بین مقادیر تشابه وجود داشته باشد، یعنی برای ماتریس وقوع مشابه، مقادیر شباهت با اندازه‌گیری‌های مختلف باید بهم نزدیکتر باشد. هنگامی که ضرایب تشابه برای ماتریس‌های وقوع A_1 ، A_2 و A_3 ارزیابی شد، اکثر معیارهای تشابه مقادیر مختلف را از خود نشان دادند. و همچنین مشاهده شد که اکثر ضرایب در پیروی از ویژگی‌های ضرایب تشابه موفق نبوده‌اند [۸۹].

۱- Sorenson
 ۲- Ochiai
 ۳- Russell
 ۴- Rao

جدول ۳-۳. تجزیه و تحلیل ضرایب تشابه ویژگی‌ها S_{ij} مطابق با هر ماتریس

| A_1 | A_2 | A_3 | ۱ | ۲ | ۳ | ۴ | ۵ | فرمول | نام ضریب |
|-------|-------|-------|-----------------------------|------|------------------------|----------------|-----------------------------|---|-------------------------------|
| ۰.۵ | ۰.۵ | ۰.۲ | دارد | * | دارد | دارد | * | $\frac{a}{a+b+c}$ | Jaccard |
| ۰.۳۳ | ۰.۶۷ | ۰.۳۳ | دارد | دارد | $a=0$ $d \approx n$ | دارد | $a \leq n$ $d \approx n$ | $\frac{(a+d) - (b+c)}{(a+d) + (b+c)}$ | Hamann |
| ۰.۶۰ | ۰.۸۸ | ۰.۶۰ | $a \approx n$ $d=0$ | دارد | دارد | $a=n$ $d=0$ | $a \approx n$ $d=0$ | $\frac{ad-bc}{ad+bc}$ | Yule |
| ۰.۶۷ | ۰.۸۳ | ۰.۳۳ | $d=n$ | دارد | $a=0$ $d \approx n$ | دارد | $a \leq n$ $d \approx n$ | $\frac{a+d}{a+b+c+d}$ | Simple Matching |
| ۰.۵ | ۰.۷۱ | ۰.۲ | $d=n$ | دارد | $a=0$ $d \approx n$ | دارد | $a \leq n$ $d \approx n$ | $\frac{a+d}{a+2(b+c)+d}$ | Rogers&Tanimoto |
| ۰.۸ | ۰.۹۱ | ۰.۵ | $d=n$ | دارد | $a=0$ $d \approx n$ | دارد | $a \leq n$ $d \approx n$ | $\frac{2(a+d)}{2(a+d)+b+c}$ | Sneath&Sokal |
| ۰.۳۳ | ۰.۱۷ | ۰.۱۷ | دارد | * | دارد | دارد | * | $\frac{a}{a+b+c+d}$ | R Ussell& R Ao |
| ۰.۶۷ | ۰.۷۵ | ۰.۳۳ | $a \neq 0$ $d \approx n$ | دارد | دارد | دارد | $a \neq 0$ $d \neq n$ | $\frac{a+(ad)^{\frac{1}{2}}}{a+b+c+(ad)^{\frac{1}{2}}}$ | Baroni-Urban&Buser |
| ۰.۶۷ | ۰.۶۷ | ۰.۳۳ | دارد | * | دارد | دارد | * | $\frac{2a}{2a+b+c}$ | Sorenson |
| ۰.۶۷ | ۰.۶۷ | ۰.۳۳ | دارد | * | دارد | دارد | * | $\frac{a}{[(a+b)(a+c)]^{\frac{1}{2}}}$ | Ochiai |
| ۰.۵۰ | ۰.۳۸ | ۰.۲۹ | دارد | دارد | دارد | دارد | دارد | $\frac{a+(ad)^{\frac{1}{2}}}{a+b+c+d+(ad)^{\frac{1}{2}}}$ | Relative Matching Coefficient |

۳-۲-۱-۵ چگونگی ساخت ضریب تشابه تطابق نسبی

به طور کلی ضریب تشابه با چهار متغیر a, b, c, d بیان می‌شود و کیفیت هر ضریب تشابه بستگی به ارزیابی هر یک از این چهار متغیر اساسی دارد. مقدار هر ضریب تشابه یک است وقتی که هر دو متغیر a

و c صفر هستند (ویژگی ۱)، که متغیرهای b و c نباید در صورت کسر ضریب تشابه درج شوند. از ویژگی ۱ و ۲ نتیجه می‌شود که هر دو متغیر a و c باید در صورت کسر ضریب تشابه تعریف شده درج شوند. باز هم، از ویژگی ۴ بدیهی است که متغیر a باید به طور مستقل در صورت کسر ظاهر شود. برای ایجاد خواص ۲، ۳، ۴ و ۵ به طور همزمان، ad علاوه بر متغیر a در صورت کسر در نظر گرفته می‌شود. برای حفظ ترتیب ad مانند ترتیب a, b, c جذر ad در نظر گرفته می‌شود. بنابراین صورت کسر ضریب تشابه جدید به صورت $a + (ad)^{\frac{1}{2}}$ می‌باشد. مخرج معمولاً برای نرمالایز ضریب تشابه در نظر گرفته می‌شود. این پدیده در ضریب تشابه بارونی و همکاران^۱ مشهود است. برای مثال، در یک ماتریس وقوع اگر $b=c=0$ باشد، برای هر ترکیب مقادیر a و d ضریب تشابه بارونی مقدار یک را نتیجه خواهد داد (بیشترین تشابه). برای غلبه بر این مشکل مخرج ضریب تشابه جدید $a + b + c + d + (ad)^{\frac{1}{2}}$ در نظر گرفته شده است. این اندازه‌گیری شباهت ضریب تطابق نسبی نام‌گذاری شده و به صورت رابطه (۳-۱۱) تعریف شده است [۸۸]:

$$S_{ij} = \frac{a + (ad)^{\frac{1}{2}}}{a + b + c + d + (ad)^{\frac{1}{2}}} \quad (3-11)$$

۳-۲-۱-۶ ارزشیابی خوشه‌بندی

تعیین تعداد خوشه

تعیین تعداد خوشه درست در یک مجموعه داده حائز اهمیت است. نه فقط بخاطر آنکه بعضی از الگوریتم‌های خوشه‌بندی مانند K-Means نیازمند تعیین تعداد خوشه‌ها به عنوان یک پارامتر هستند، بلکه تعداد مناسب خوشه‌ها روند صحیح خوشه‌بندی را کنترل می‌کند، که می‌تواند به عنوان یافتن یک تعادل مناسب بین دقت^۲ و فشردگی^۳ در تجزیه و تحلیل خوشه‌بندی در نظر گرفته شود. اگر همه مجموعه داده به عنوان یک خوشه در نظر گرفته شود، فشردگی داده‌ها حداکثر خواهد بود، اما چنین

۱- Baroni-Urban&Buse
 ۲- Accuracy
 ۳-Compressibility

خوشه‌بندی ارزشمند نخواهد بود. از سوی دیگر اگر هر شی موجود در مجموعه داده را به صورت یک خوشه در نظر گرفته شود، بیشترین میزان دقت بدست می‌آید، اگرچه با داشتن هر شی در یک خوشه خلاصه‌سازی داده‌ها حاصل نخواهد شد. بنابراین باید از ترکیب دو معیار فوق استفاده کرد.

یافتن تعداد صحیح خوشه‌ها به سادگی نیست زیرا این تعداد صحیح مبهم است. یافتن تعداد صحیح خوشه‌بندی اغلب وابسته به شکل توزیع و مقیاس در مجموعه داده است. تعداد زیادی روش برای تخمین تعداد صحیح خوشه‌ها وجود دارد. که به طور مختصر چند روش مؤثر و محبوب را بیان شده است.

یک روش ساده برای تعیین تعداد خوشه‌ها مقدار تقریبی $\sqrt{\frac{n}{2}}$ برای یک مجموعه داده با n نقطه می‌باشد. که انتظار می‌رود هر خوشه دارای $\sqrt{2n}$ نقطه باشد.

روش البو^۱ به دنبال بهینه‌سازی یک معیار مخصوص مانند مجموع مربعات فواصل درون خوشه‌ای است. مجموع فواصل درون خوشه‌ای داده‌ها، میزان فشردگی خوشه‌بندی انجام شده را نشان می‌دهد و هدف حداقل‌سازی آن تا جای ممکن است. در نتیجه یک ایده ابتکاری برای انتخاب تعداد صحیح خوشه‌ها، استفاده از نقطه عطف در نمودار مجموع مربعات فواصل درون خوشه‌ای با توجه به تعداد خوشه‌ها است.

$k > 0$ را در نظر بگیرید. می‌توان k خوشه را روی مجموعه داده‌ها با توجه به الگوریتم انجام داد. سپس مجموعه مربعات فواصل داده‌ها را محاسبه کرد. آن‌گاه می‌توان منحنی مجموع مربعات را با توجه به k رسم نمود. اولین یا برجسته‌ترین نقطه عطف به عنوان تعداد صحیح خوشه‌ها توصیه می‌شود.

تعداد صحیح خوشه‌ها در یک مجموعه داده همچنین می‌تواند توسط اعتبارسنجی متقابل انجام شود، تکنیکی که بیشتر در طبقه‌بندی مورد استفاده قرار می‌گیرد. در ابتدا مجموعه داده شده D را به m قسمت تقسیم کنید. سپس از $m-1$ قسمت برای ساخت یک مدل خوشه‌بندی استفاده کنید و از قسمت باقیمانده برای آزمودن کیفیت خوشه‌بندی استفاده کنید. برای مثال، برای هر نقطه موجود در مجموعه

۱- elbow method

آزمون می‌توان نزدیکترین مرکز خوشه را یافت. بنابراین می‌توان از مجموعه مربعات فواصل بین نقطه‌ای و نزدیکترین مرکز خوشه برای اندازه‌گیری میزان تناسب خوشه‌بندی با مجموعه آزمون استفاده کرد. برای هر عدد صحیح $k > 0$ این فرایند را m بار تکرار می‌شود تا خوشه‌بندی از k خوشه با استفاده از هر قسمت در مجموعه آزمون بدست آورده شود. میانگین اندازه‌گیری کیفیت به عنوان اندازه‌گیری کیفیت کلی در نظر گرفته می‌شود. سپس می‌توان اندازه‌گیری کیفیت کلی را با توجه مقادیر مختلف k مقایسه کرده و بهترین تعداد خوشه‌ها که مناسب داده‌ها است را پیدا کرد [۸۹].

برای ارزیابی کیفیت خوشه‌بندی چندین روش وجود دارد. این روش‌ها براساس کلی به دو گروه با توجه به دسترسی به داده مرجع^۱ تقسیم می‌شوند. داده مرجع خوشه‌بندی ایده آلی است که توسط متخصصین ساخته می‌شود.

اگر داده مرجع موجود باشد، می‌توان از ارزیابی بیرونی استفاده کرد که خوشه‌بندی را با داده مرجع مقایسه می‌کند. در غیر این صورت می‌توان از ارزیابی درونی استفاده کرد.

ارزیابی خارجی

نتایج خوشه بندی در ارزیابی خارجی بر اساس داده‌هایی که برای خوشه بندی استفاده نشدند، مانند برجسب‌های کلاس شناخته شده و معیارهای خارجی ارزیابی می‌شود. متخصصین قبل از طبقه‌بندی چنین معیارهایی را تعیین میکنند. بنابراین مجموعه معیارها می‌تواند به عنوان یک استاندارد برای ارزیابی استفاده شود. این نوع روش‌های ارزیابی اینکه چقدر خوشه بندی به کلاس‌های معیاری پیش تعیین شده نزدیک است، را تعیین می‌کند. با این حال، اینکه آیا این داده‌های واقعی مناسب است یا فقط بر روی مجموعه داده‌های مصنوعی در داده مرجع است، مورد بحث قرار گرفته‌است، از آنجا که کلاس‌ها می‌توانند ساختار داخلی داشته باشند، ویژگی‌های موجود ممکن است اجازه جدا شدن خوشه‌ها یا کلاس‌ها را ندهند.

۱- Ground Truth

چندین اندازه گیری برای ارزیابی خارجی وجود دارد که در ادامه چند روش بیان شده است:

شاخص خلوص^۱: در سال ۲۰۰۲ زاهو^۲ و همکارانش شاخص خلوص را ارائه دادند. این شاخص برای خوشه‌هایی که دارای یک کلاس واحد هستند، اندازه گیری می‌شود. برای محاسبه آن، برای هر خوشه، تعداد نقاط داده از کلاس معمول در خوشه مورد نظر شمرده می‌شود، سپس تمام خوشه‌ها جمع شده و بر تعداد نقاط داده تقسیم می‌شود. با توجه به مجموعه ای از خوشه‌های M و مجموعه ای از کلاس‌های D ، هر دو پارامتر با N نقطه داده، می‌توان خلوص را به صورت رابطه (۳-۱۲) تعریف کرد:

$$\text{Purit} = \frac{1}{N} \sum_{m \in M} \max_{d \in D} |m \cap d| \quad (3-12)$$

نتیجه محاسبه این رابطه بین صفر و یک است و هرچه این مقدار به یک نزدیکتر باشد نشان از دقت خوشه‌بندی و تعداد خوشه بهینه دارد.

شاخص فولکز-مالوز^۳: به منظور استفاده از این شاخص باید چندین اصطلاح تعریف شود:

مثبت صحیح: اگر زوجی از مشاهدات که در یک دسته هستند، در یک خوشه نیز قرار بگیرند، نتیجه خوشه‌بندی را برای این زوج، مثبت صحیح (True Positive) نامیده می‌شود. مثبت صحیح را با TP نیز نشان می‌دهند.

منفی صحیح: اگر زوجی از مشاهدات که در دو دسته مجزا قرار دارند، در دو خوشه مجزا نیز جای گیرند، نتیجه خوشه‌بندی برای این زوج، منفی صحیح (True Negative) نامیده می‌شود. TN نشان‌دهنده منفی صحیح است.

مثبت کاذب: اگر زوجی از مشاهدات که در دو دسته مجزا قرار دارند، در یک خوشه جای گیرند، نتیجه خوشه‌بندی برای این زوج مثبت کاذب (False Positive) است. مثبت کاذب با FP نیز نشان داده می‌شود.

۱- Purity

۲-Zaho

۳- Fowlkes-Mallows

منفی کاذب: اگر زوجی از مشاهدات که در یک دسته قرار دارند به اشتباه در دو خوشه قرار گیرند، نتیجه خوشه‌بندی برای این زوج منفی کاذب (False Negative) است. منفی کاذب را با FN نشان می‌دهند.

حال اگر تعداد زوج‌های مثبت صحیح با |TP|، تعداد مثبت کاذب با |FP| و تعداد منفی کاذب با |FN| نشان داده شود، شاخص فولکز-مالوز که با FMI نشان داده می‌شود با استفاده از رابطه (۳-۱۳) بدست می‌آید:

$$FMI = \frac{|TP|}{\sqrt{(|TP|+|FP|)(|TP|+|FN|)}} \quad (3-13)$$

مقدار شاخص فولکز-مالوز بین صفر تا یک است. مطابقت بین دو شیوه برچسب‌گذاری بیشتر است زمانی که مقدار FMI به یک نزدیکتر باشد. زمانی که مقدار این شاخص به صفر برسد، خوشه‌بندی به صورت تصادفی انجام شده است.

شاخص رند^۱: در سال ۱۹۷۱ این شاخص توسط ویلیام رند^۲ ارائه شد. شاخص رند اینکه خوشه‌ها چه مقدار به عیار طبقه‌بندی شبیه‌اند را محاسبه می‌کند. همچنین می‌توان این شاخص را برای اندازه‌گیری درصد تصمیمات درست که توسط الگوریتم ساخته شده است، استفاده کرد. استفاده از مفاهیم مثبت صحیح، مثبت کاذب، منفی کاذب و منفی صحیح در محاسبه این شاخص برای ارزیابی نتایج خوشه‌بندی اهمیت زیادی دارد. می‌توان شاخص رند را با توجه به این مفاهیم به صورت رابطه (۳-۱۴) محاسبه کرد:

$$RI = \frac{TP+T}{TP+FN+FP+TN} \quad (14-3)$$

ارزیابی داخلی

۱-Rand Index
۲-William M.Rand

وقتی داده مرجع یک مجموعه داده در دسترس نباشد، برای بررسی کیفیت خوشه‌بندی از ارزیابی درونی استفاده می‌کنیم. به طور کلی روش‌های درونی خوشه‌بندی را با توجه به دو مفهوم تفکیک‌پذیری و فشردگی ارزیابی می‌کنند. تعدادی از روش‌های درونی در زیر مطرح شده است:

شاخص دان^۱: در سال ۱۹۷۴ خانواده شاخص‌های دان طی مقاله‌ای براساس مفاهیم فشردگی و تفکیک‌پذیری توسط دان معرفی شد. او با دو معیار فاصله^۲ و قطر^۳، میزان فشردگی و تفکیک‌پذیری را محاسبه کرد.

اگر فاصله بین دو خوشه C_i و C_j با $D(C_i, C_j)$ نشان داده شود، می‌توان میزان تفکیک‌پذیری در خوشه‌بندی را به صورت رابطه (۳-۱۵) محاسبه کرد:

$$D(C_i, C_j) = \min_{y \in C_j, x \in C_i} d(x, y) \quad (15-3)$$

همچنین به منظور اندازه‌گیری فشردگی خوشه‌ها، از قطر هر خوشه استفاده می‌شود. برای خوشه C_i مقدار قطر توسط رابطه (۳-۱۶) بدست می‌آید:

$$\text{diam}(C_i) = \max_{x, y \in C_i} d(x, y) \quad (16-3)$$

حال شاخص دان به صورت رابطه (۳-۱۷) تعریف می‌شود:

$$DI = \frac{\min_{1 \leq i, j \leq n} d(C_j, C_i)}{\max_{1 \leq i \leq n} \text{diam}(C_i)} \quad (17-3)$$

فاصله بین دو خوشه به عنوان معیاری برای تفکیک‌پذیری در صورت این کسر دیده می‌شود و قطر هر خوشه نیز در مخرج قرار دارد. نسبت این دو، مقیاسی برای سنجش فاصله بین دو خوشه خواهد بود. بنابراین کوچکترین مقدار این نسبت برای همه خوشه‌ها، می‌تواند شاخصی برای ارزیابی خوشه‌بندی

۱- Daunn Index
 ۲- Cluster Distance
 ۳- Diameter

باشد. مقدار بزرگتر این شاخص بیانگر تفکیک پذیری بهتر و در نتیجه خوشه‌بندی موثرتر است. بر همین اساس اگر نسبت میزان تفکیک پذیری به قطر خوشه‌ها نیز مقدار بزرگی باشد، خوشه‌بندی به خوبی انجام شده است.

شاخص نیم‌رخ^۱: در سال ۱۹۸۶ شاخص نیم‌رخ برای اولین بار توسط رسوو^۲ و همکارانش مطرح شد که یکی از متداول‌ترین روش‌های اعتبارسنجی خوشه‌بندی قطعی است. این معیار به پیوستگی درون خوشه‌ها و نیز به میزان تفکیک پذیری آن‌ها بستگی دارد. مقدار شاخص نیم‌رخ برای هر نقطه، میزان تعلق آن نقطه به خوشه‌اش در مقایسه با خوشه مجاور را اندازه می‌گیرد. فرض کنید نقطه‌ای مانند x_i در میان داده‌هایی که خوشه‌بندی کرده‌اید وجود دارد و در طی مراحل خوشه‌بندی نیز k خوشه (C_1, C_2, \dots, C_k) ایجاد شده است. برای محاسبه معیار نیم‌رخ احتیاج به آشنایی با دو مفهوم اصلی است:

میانگین فاصله یک نقطه از خوشه با نقاط دیگر آن خوشه، این مقدار را با $a(i)$ نشان داده و به صورت رابطه (۱۸-۳) محاسبه می‌شود:

$$a(i) = \frac{1}{n} \sum_{i=1}^n d(x_i, x_I) \quad (18-3)$$

این معیار را می‌توان ملاکی برای ارزیابی تعلق نقطه x_i در خوشه‌اش در نظر گرفت. مقدار کوچکتر $a(i)$ نشان‌دهنده میزان تعلق بیشتر این نقطه به خوشه‌اش است. همچنین این شاخص را می‌توان براساس توابع فاصله مانند فاصله منهتن و اقلیدسی نیز محاسبه کرد.

حداقل میانگین فاصله نقطه با خوشه‌های دیگر، فرض کنید نقطه x_i به خوشه C_j تعلق دارد. حال میانگین فاصله این نقطه را با نقاط خوشه‌های دیگر مانند C_k اندازه‌گیری می‌شود. خوشه‌ای که دارای کمترین میانگین فاصله برای نقطه x_i باشد، به عنوان خوشه مجاور با این نقطه نامیده می‌شود. مقدار میانگین فاصله نقطه x_i با نقاط خوشه مجاور را با $b(i)$ نشان داده می‌شود:

۱-Silhouette Index
۲-J Peter Rousseeuw

$$b(i) = \min_{1 \leq l \leq K} \frac{1}{n} \sum_{y_m \in C_l} d(x_i, y_m) \quad (3-19)$$

به این ترتیب شاخص نیمرخ برای نقطه x_i بوسیله رابطه (۳-۲۰) اندازه‌گیری می‌گردد:

$$S(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))} \quad (3-20)$$

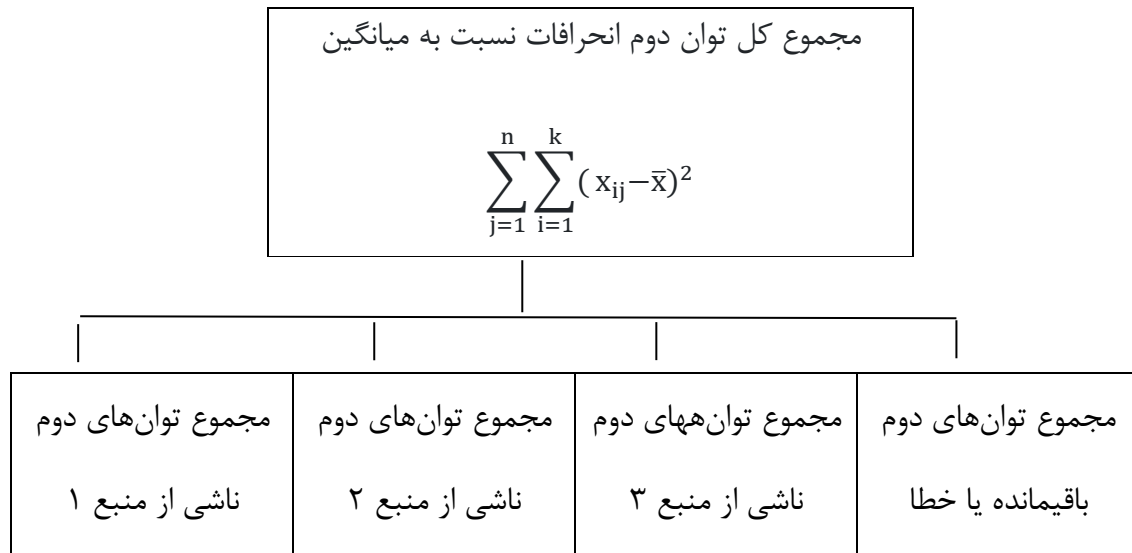
مقدار شاخص نیمرخ در صورتی که $a(i)$ کوچکتر از $b(i)$ باشد، مثبت می‌شود و برعکس اگر $b(i)$ کوچکتر از $a(i)$ باشد، مقدار شاخص نیمرخ منفی شده و نشانگر خوشه‌بندی ضعیف است زیرا نقطه x_i بیش از آنکه شبیه خوشه خودش باشد به خوشه مجاور شباهت دارد. مقدار این شاخص با توجه به رابطه بالابین -1 تا $+1$ است. مقدار نزدیک به یک بیانگر انطباق خوب بین نقطه و خوشه‌اش نسبت به خوشه مجاور است. عمل خوشه‌بندی زمانی درست انجام شده است که معیار نیمرخ برای همه نقاط درون خوشه‌ها نزدیک به یک باشد. در حالیکه کوچک بودن مقدار نیمرخ برای خوشه‌ها، بیانگر خوشه‌بندی ضعیف می‌باشد که ممکن است به علت انتخاب نامناسب تعداد خوشه‌ها (k) باشد. حال اگر میانگین مقدار نیمرخ برای نقاط هر خوشه محاسبه شود، معیاری برای ارزیابی هر خوشه بدست می‌آید. همچنین میانگین کل مقادیر نیمرخ معیاری برای ارزیابی عملیات خوشه‌بندی محسوب می‌شود [۹۰، ۸۹].

در پژوهش‌های کمی از شاخص‌ها و آزمون‌های آماری برای تحلیل داده‌ها و بررسی فرضیه‌های آماری استفاده می‌شود. به این منظور در این پژوهش از تحلیل واریانس برای بررسی فرضیه آماری و از آزمون کمترین تفاوت معنی‌دار برای بررسی فرضیه آماری استفاده شده است.

۳-۲-۱-۲ تحلیل واریانس

می‌توان میزان انحرافات کل در مجموعه داده‌ها را با استفاده از تحلیل واریانس به مؤلفه‌هایی افراز کرد. هر مؤلفه به دلیلی قابل تشخیص بوده، می‌تواند به یک منبع انحراف نسبت داده شود. همچنین، یک مؤلفه، انحراف حاصل از عامل‌های کنترل نشده و خطاهای تصادفی مربوط به اندازه‌های پاسخ‌ها را نشان

می‌دهد. فرض کنید داده‌های x_{ij} برای $i=1,2,\dots,k$ و $j=1,2,\dots,n$ بوده و میانگین کل آنها \bar{X} باشد. در این صورت تغییر نسبت به میانگین به صورت $\sum_{j=1}^n \sum_{i=1}^k (x_{ij} - \bar{X})^2$ مجموع توان‌های دوم انحرافات درمی‌آید که مجموع کل توان دوم انحرافات نامیده می‌شود. روش تحلیل واریانس این مقدار را به قسمت‌هایی تجزیه می‌کند که در شکل ۲-۳ سه منبع تغییر قابل شناسایی بعلاوه مؤلفه خطا مشخص شده است:



شکل ۲-۳. افراز تغییرات به چند منبع تغییر

تعداد منابع تغییر قابل شناسایی می‌باشد و فرمول مجموع‌های توان دوم مؤلفه‌ها به طرح آزمایشی که داده‌ها توسط آن جمع‌آوری شده و به مدل آماری‌ای که برای تحلیل مناسب تشخیص داده شده است بستگی دارد.

در حالت کلی، در چنین مسائلی k نمونه تصادفی به اندازه n از k جامعه وجود دارد. مقدار مشاهده‌ی Z_{ij} از جامعه i ام با x_{ij} نشان داده می‌شود. فرض می‌شود متغیرهای تصادفی متناظر که همه مستقلند، توزیع‌های نرمال با میانگین‌های μ_i و واریانس مشترک σ^2 دارند. با این فرض‌ها هر مشاهده‌ای را می‌توان به صورت $x_{ij} = \mu_i + e_{ij}$ به ازای $i=1,2,\dots,k$ و $j=1,2,\dots,n$ نشان داد که در آن e_{ij} ها مقادیر متغیر تصادفی مستقل نرمال با میانگین صفر و واریانس مشترک σ^2 هستند و برای امکان تعمیم آن به

وضعیت‌های پیچیده‌تر، معمولاً آن را به صورت $X_{ij} = \mu + \alpha_i + e_{ij}$ نمایش می‌دهند که به μ ، میانگین کل و به α_i ، اثر تیماری می‌گویند، به طویکه $\sum_{i=1}^k \alpha_i = 0$ است.

رابطه (۳-۲۱) فرض صفری که آزمون می‌شود را نشان می‌دهد

$$H_0 = \mu_1 = \mu_2 = \dots = \mu_k \quad (۲۱-۳)$$

$$H_0 = \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$$

به معنای آنکه تمام میانگین‌های جامعه باهم برابرند و یا همه اثرهای تیماری صفرند. فرض مقابل عبارت است از:

دست کم دو تا از میانگین‌ها برابر نیستند. $H_1 =$

دست کم یکی از اثرهای تیماری مخالف صفر است. $H_1 =$

آزمون بر تحلیل تغییرپذیری کل داده‌های تلفیق شده مبتنی است که با رابطه (۳-۲۲) نشان داده می‌شود:

$$\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x})^2 \quad (۲۲-۳)$$

که در آن \bar{x} از رابطه (۳-۲۳) بدست می‌آید.

$$\bar{x} = \frac{1}{nk} \sum_{i=1}^k \sum_{j=1}^n x_{ij} \quad (۲۳-۳)$$

در صورت درستی فرض صفر، همه این تغییرپذیری ناشی از شانس است، در غیر این صورت، بخشی از مجموع توان‌های دوم بالا ناشی از اختلاف‌های بین تیمارها خواهد بود. حال تغییرپذیری کل فوق به دو جزء انحرافات حاصل از تیمار و انحرافات باقیمانده تفکیک می‌شود:

مجموع کل توان دوم = مجموع توان دوم حاصل از تیمار + مجموع توان دوم باقیمانده

$$\sum_{j=1}^n \sum_{i=1}^k (x_{ij} - \bar{x})^2 = n \cdot \sum_{i=1}^k (\bar{x}_i - \bar{x})^2 + \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \quad (۲۴-۳)$$

که در آن میانگین مشاهدات جامعه m/i و \bar{X} میانگین همه nk مشاهده است. فرمول فوق به صورت $SST=SS(Tr) + SSE$ نشان داده می‌شود.

انحرافات کل داده‌های تلفیق شده به دو جزء تفکیک می‌شود: و جزء اول $SS(Tr)$ تغییر تصادفی را وقتی که فرض صفر درست باشد و جزء دوم $-SSE-$ تغییر شانس یعنی تغییرات داخل نمونه‌ها را اندازه می‌گیرد. همچنین جزء اول تغییر بین میانگین‌های جامعه‌ای را وقتی فرض صفر درست نباشد اندازه‌گیری می‌کند. درجه آزادی مجموع توان دوم تیمارها $k-1$ و درجه آزادی مجموع توان دوم خطا $k(n-1)$ است. در این صورت، میانگین توان دوم تیمارها و میانگین توان دوم خطا که به ترتیب با $MS(Tr)$ و MSE نشان داده می‌شوند، با رابطه (۳-۲۵) و (۳-۲۶) بدست می‌آیند.

$$MS(Tr) = \frac{SS(Tr)}{k-1} \quad (۲۵-۳)$$

$$MSE = \frac{SSE}{K(n-1)} \quad (۲۶-۳)$$

زمانی که میانگین توان دوم تیمارها نسبت به میانگین توان دوم خطا کم باشد، نتیجه می‌گردد که میانگین‌های جامعه‌ها با هم تفاوت ندارند و فرض H_0 پذیرفته می‌شود. چنانچه میانگین توان‌های دوم تیمارها نسبت به میانگین توان دوم خطا زیاد باشد، نتیجه می‌شود که میانگین‌های جامعه‌ها باهم تفاوت دارند و فرض H_0 رد می‌شود. نظریه آماری می‌گوید طبق فرض صفر نسبت:

$$\frac{MS(Tr)}{MSE} \quad (۲۷-۳)$$

که با F نشان داده می‌شود، به صورت رابطه (۳-۲۸) محاسبه می‌شود:

$$F = \frac{MS(Tr)}{MSE} = \frac{SS(Tr)/(k-1)}{SSE/K(n-1)} \quad (۲۸-۳)$$

که دارای توزیع F بوده و صورت دارای درجه آزادی $K-1$ و مخرج دارای درجه آزادی $K(n-1)$ می‌باشد.

معمولاً تجزیه مجموع توان‌های دوم و درجات آزادی به همراه میانگین‌های توان دوم به شکل جدولی به نام جدول تحلیل واریانس (جدول ۴-۳) ارائه می‌شود. این جدول ستون اضافی دیگری دارد که مقدار F را نیز حساب می‌کند.

جدول ۴-۳ تحلیل واریانس

| منبع تغییرات | مجموع توان‌های دوم | درجه آزادی | میانگین توان‌های دوم | F |
|--------------|--------------------|------------|-------------------------------|----------------------|
| تیمارها | SS(Tr) | k-1 | $\frac{SS(Tr)}{k-1} MS(Tr) =$ | $\frac{MS(Tr)}{MSE}$ |
| خطا | SSE | k(n-1) | $MSE = \frac{SSE}{K(n-1)}$ | |
| جمع | SST | Kn-1 | | |

LSD ۳-۲-۱-۸ روش

روش LSD^۱ یا کمترین تفاوت معنی‌دار مستلزم محاسبه کوچک‌ترین تفاوت معنی‌دار ممکن بین دو میانگین است. دو شرط برای استفاده از این روش وجود دارد:

تعداد محدودی از مقایسه‌های مستقل از قبل طراحی شده باشد.

همه میانگین‌های گروه‌های آزمایشی با یک گروه کنترل مقایسه شود.

در این روش برای هر مقایسه، یک فرض صفر به گونه‌ای زیر برای میانگین‌ها آزمون می‌شود:

$$H_0 = C = 0$$

$$H_1 = C \neq 0$$

^۱ - Least Significant Difference

که C عبارت است از تفاوت بین میانگین‌ها. در این روش باید برای هر مقایسه دوتایی مقدار C به دست آورده شود که $\dots \hat{c} = \bar{x}_i - \bar{x}_j$ است.

برای هر مقایسه دوتایی خطای استاندارد تفاوت بین میانگین‌ها، S_d ، باید حساب شود. در زمانی که دو گروه مستقل بوده و حجم نمونه آنها برابر باشد S_d از رابطه (۳-۲۹) بدست می‌آید:

$$S_d = \sqrt{\frac{2MSE}{n}} \quad (۲۹-۳)$$

مقدار بحرانی یا کمترین تفاوت معنی‌دار از رابطه (۳-۳۰) محاسبه می‌شود:

$$LSD = t(\alpha, df) \times S_d \quad (۳۰-۳)$$

اگر $LSD \gg |\bar{x}_i - \bar{x}_j|$ باشد، میانگین دو جامعه μ_i و μ_j باهم تفاوت دارند [۹۱].

۳-۳ خلاصه و جمع‌بندی فصل

در ابتدای این فصل فرایند خوشه‌بندی و الگوریتم منتخب توضیح داده شد. در ادامه دلایل انتخاب ضریب تشابه تطابق نسبی ارائه گردید. اجزای روش خوشه‌بندی شامل متغیرها، اشیا و ضریب تشابه مشخص شد. همچنین روش‌های ارزشیابی خوشه‌بندی مورد بررسی قرار گرفت. در انتها روش‌های آماری مورد استفاده در پژوهش مطرح شد.

فصل ۲ تجزیہ و تحلیل داده‌ها

۴-۱ مقدمه

در فصل سوم فرایند خوشه‌بندی، متغیرهای مورد استفاده در پژوهش و نحوه محاسبه آنها و چگونگی تجزیه و تحلیل داده‌ها تشریح شد. در پژوهش‌های علمی، تجزیه و تحلیل داده‌های آماری جمع‌آوری شده مرحله مهمی از تحقیق به شمار می‌رود که در طی آن با بهره‌گیری از روش‌های آماری متناسب با روش تحقیق، اطلاعات بدست آمده مورد آزمون قرار گرفته و براساس نتایج حاصل از آزمون به تأیید یا رد فرضیه‌های پژوهش و تفسیر آن‌ها پرداخته می‌شود.

در این فصل با توجه به تعریف خوشه‌بندی، الگوریتم خوشه‌بندی سلسله‌مراتبی و ضریب تشابه منتخب در فصل سوم و همچنین شاخص‌های نقدشوندگی تعریف شده، اطلاعات ۴۲ شرکت پذیرفته شده در بورس جمع‌آوری شده، سپس با توجه به شاخص‌های نقدشوندگی و محاسبه ضریب شباهت، خوشه‌بندی پیوند کامل انجام شده است. نتایج این خوشه‌بندی و تجزیه و تحلیل آن در این فصل بیان شده است.

۴-۲ جامعه و نمونه آماری

جامعه آماری این پژوهش کلیه شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران می‌باشد. در این پژوهش از روش نمونه‌گیری تصادفی بهره گرفته شده است که بدین منظور پس از بررسی‌های صورت گرفته اطلاعات یک سری از شرکت‌ها که در دسترس نبوده است از لیست مورد بررسی حذف گردیدند. نمونه آماری برای سال‌های ۱۳۹۷-۱۳۹۸ با توجه به شرایط زیر انتخاب شده است:

(۱) نام شرکت تا پایان سال ۱۳۹۸ در فهرست شرکت‌های پذیرفته شده در بورس درج شده باشد.

(۲) پایان سال مالی منتهی به اسفندماه بوده و همچنین طی دوره پژوهش شرکت سال مالی خود را تغییر نداده باشد.

۳) آمار و اطلاعات مربوط به متغیرها و یادداشت‌های توضیحی صورت‌های مالی قابل دسترسی باشد.

۴) شرکت‌های مورد نظر بیش از سه ماه توقف معاملاتی نداشته باشند.

۵) از سال ۱۳۸۸ در بورس حضور داشته باشند.

۶) پژوهش برای شرکت‌های غیرمالی انجام شده است. بنابراین بانک‌ها و کلیه شرکت‌های سرمایه‌گذاری، موسسات مالی و لیزینگ‌ها از نمونه حذف شدند. از جمله دلایل حذف شرکت‌ها و مؤسسات مالی:

الف - وجود تفاوت در تفسیر ریسک مالی (نسبتهای اهرمی بالا) در شرکت‌های مالی و غیرمالی، به طوری که این ریسک برای شرکت‌های مالی، عادی بوده و برای شرکت‌های غیرمالی ممکن است غیرعادی به نظر رسیده و منجر به اتخاذ تصمیمات نامناسب گردد.

ب- وجود تفاوت در روش‌ها و برآوردهای حسابداری در شرکت‌های مالی و غیرمالی که تداخل آنها با یکدیگر ممکن است به بروز نتایج اشتباه منتج گردد.

ج- عدم شفافیت طبقه‌بندی میان فعالیت‌های عملیاتی و تأمین مالی شرکت‌های مالی و غیرمالی

با توجه به محدودیت‌های پیش‌گفته تعداد ۴۲ شرکت طی دوره ۲ ساله به عنوان نمونه قابل دسترس انتخاب شدند که در جدول ۱-۴ قابل مشاهده می‌باشند.

جدول ۴-۱. نمونه آماری پژوهش

| نام شرکت | نام صنعت | نام شرکت | نام صنعت |
|----------------------------|------------------------------|--------------------------|----------------------------------|
| کاشی الوند | کاشی و سرامیک | ایران خودرو | ساخت خودرو و قطعات |
| تأمین ماسه ریخته‌گری | استخراج سایر معادن | سایپا | ساخت خودرو و قطعات |
| معدنی و صنعتی چادرملو | استخراج کانه‌های فلزی | محورسازان ایران خودرو | ساخت خودرو و قطعات |
| معدنی و صنعتی گل‌گوه‌ر | استخراج کانه‌های فلزی | کمک‌فنا ایندامین | ساخت خودرو و قطعات |
| معادن منگنز ایران | استخراج کانه‌های فلزی | تراکتورسازی ایران | ساخت خودرو و قطعات |
| شیشه و گاز | سایر محصولات کانی غیر فلزی | چرخشگر | ساخت خودرو و قطعات |
| سرامیک‌های صنعتی اردکان | سایر محصولات کانی غیر فلزی | رادیاتور ایران | ساخت خودرو و قطعات |
| کارخانجات تولیدی شیشه رازی | سایر محصولات کانی غیر فلزی | فولاد مبارکه اصفهان | فلزات اساسی |
| کارخانجات تولید قندی | ماشین‌آلات و دستگاه‌های برقی | ملی صنایع مس ایران | فلزات اساسی |
| ایران ترانسفر | ماشین‌آلات و دستگاه‌های برقی | نورد آلومینیوم | فلزات اساسی |
| لامپ پارس شهاب | ماشین‌آلات و دستگاه‌های برقی | گروه صنعتی سپاهان | فلزات اساسی |
| پتروشیمی جم | محصولات شیمیایی | به‌نوش ایران | محصولات غذایی |
| صنایع شیمیایی سینا | محصولات شیمیایی | بیسکویت گرجی | محصولات غذایی |
| کارتن ایران | محصولات کاغذی | شیرپاستوریزه پگاه اصفهان | محصولات غذایی |
| شرکت ارتباطات ایران | مخابرات | پالایش نفت تبریز | فراورده‌های نفتی |
| حفاری شمال حفاری شمال | استخراج نفت گاز | پالایش نفت اصفهان | فراورده‌های نفتی |
| ایران مریئوس | منسوجات | کارخانجات داروپخش | مواد و محصولات دارویی |
| مبین انرژی خلیج فارس | رضه برق، گاز، بخار و آب گرم | داروسازی کوثر | مواد و محصولات دارویی |
| گروه صنعتی بارز | لاستیک و پلاستیک | راورده‌های تزریقی ایران | مواد و محصولات دارویی |
| آبسال | ماشین‌آلات و تجهیزات | داده‌پردازی | رایانه و فعالیت‌های وابسته به آن |
| کشاورزی و دامپروری مگسال | زراعت و خدمات وابسته | پارس الکترونیک | ساخت دستگاه‌ها و وسایل ارتباطی |

۴-۲-۱ روش گردآوری داده‌ها

در این پژوهش جمع‌آوری داده‌های مورد نیاز بخش مبانی نظری و سوابق پژوهش، با استفاده از مطالعات کتابخانه‌ای، نشریات داخلی و بین‌المللی و جمع‌آوری سایر داده‌های مربوط به متغیرهای پژوهش از طریق شرکت مدیریت فناوری بورس تهران^۱، گزارش‌های ادواری سازمان بورس اوراق بهادار تهران^۲ و سامانه اطلاع‌رسانی ناشران کدال^۳ انجام شده است.

۱- www.tsetmc.com

۲- www.tse.ir

۳- www.codal

۴-۲ الگوریتم پیاده‌سازی پژوهش

همان‌طور که در فصل سوم ذکر شد، فرایند خوشه‌بندی شامل انتخاب ویژگی، الگوریتم خوشه‌بندی،

اعتبارسنجی نتایج و تفسیر نتایج می‌باشد.

انتخاب ویژگی‌ها

برای انجام الگوریتم خوشه‌بندی ویژگی‌هایی نیاز است که براساس آنها خوشه‌بندی انجام شود. در این

پژوهش خوشه‌بندی نمونه آماری با توجه به شاخص‌های نقدشوندگی تعریف شده در فصل سوم صورت

گرفته است و به این منظور داده‌های مورد نیاز جمع‌آوری شده و محاسبات با کمک نرم افزار Excel انجام

گرفت، نتایج محاسبات شاخص‌ها در جداول جدول ۴-۲ تا جدول ۴-۵ ارائه شده است:

جدول ۴-۲. نتیجه محاسبات شاخص‌های نقدشوندگی سال ۱۳۹۷

| | PBA _{i,t} | TO | ADJILLQ | LM | TV | N _{i,t} |
|-----------------------|--------------------|--------|---------|-------|----------|------------------|
| فولاد مبارکه اصفهان | 0.022 | 0.0201 | -29.617 | 1.064 | 6.02E+13 | 773680 |
| ایران خودرو | 0.042 | 0.0976 | -28.58 | 1.099 | 1.16E+13 | 369791 |
| معدنی و صنعتی چادرملو | 0.0362 | 0.0878 | -28.484 | 0.959 | 1.38E+13 | 244245 |
| پتروشیمی جم | 0.0364 | 0.0274 | -27.468 | 3.06 | 6.27E+12 | 58998 |
| داروپخش | 0.031 | 0.0527 | -21.333 | 2.513 | 2.18E+11 | 8304 |
| ملی صنایع مس ایران | 0.0314 | 0.0539 | -29.193 | 1.278 | 3.89E+13 | 654477 |
| شرکت سیار ایران | 0.0177 | 0.0221 | -29.165 | 0.704 | 7.38E+12 | 78860 |
| مبین انرژی خلیج فارس | 0.0229 | 0.0543 | -29.144 | 0.753 | 1.06E+13 | 127458 |
| پالایش نفت تبریز | 0.0322 | 0.2018 | -28.842 | 0.765 | 1.47E+13 | 112429 |
| سرامیک صنعتی اردکان | 0.0421 | 0.1536 | -26.986 | 1.301 | 4.14E+12 | 186827 |
| سایپا | 0.0417 | 0.0432 | -29.291 | 1.16 | 2.91E+13 | 1307298 |
| ایران ترانسفو | 0.0377 | 0.072 | -24.58 | 0.888 | 7.92E+12 | 106092 |
| گروه صنعتی بارز | 0.0302 | 0.1159 | -26.77 | 0.95 | 1.90E+12 | 60914 |
| بهنوش | 0.0393 | 0.0539 | -22.84 | 0.647 | 1.00E+11 | 7723 |
| پارس الکترونیک | 0.0403 | 0.1406 | -25.442 | 0.874 | 8.96E+11 | 36804 |
| تراکتورسازی ایران | 0.0364 | 0.0365 | -24.331 | 1.165 | 1.00E+12 | 19487 |
| حفاری شمال | 0.0306 | 0.0544 | -26.1 | 1.718 | 4.44E+12 | 142530 |
| داده پردازی | 0.0399 | 0.1596 | -25.625 | 0.861 | 8.62E+11 | 48809 |
| پالایش نفت اصفهان | 0.0345 | 0.0351 | -27.552 | 0.856 | 4.29E+13 | 559622 |
| شیشه رازی | 0.0402 | 0.0035 | -26.749 | 0.971 | 1.37E+12 | 84466 |
| کارخانجات تولیدی قندی | 0.0482 | 0.0137 | -26.009 | 3.968 | 1.32E+12 | 61098 |
| کارتن ایران | 0.0463 | 0.0265 | -26.425 | 3.908 | 4.99E+12 | 116403 |
| بیسکوویت گرجی | 0.0319 | 0.0414 | -24.368 | 7.544 | 5.60E+11 | 10173 |
| گروه صنعتی سپاهان | 0.0401 | 0.0679 | -24.539 | 1.176 | 7.07E+11 | 29586 |

| | | | | | | |
|-------------------------|--------|--------|---------|-------|-------------|--------|
| صنعتی و معدنی گل گوهر | 0.0299 | 0.0256 | -27.871 | 0.87 | 2.00E+13 | 210250 |
| ایران مرینوس | 0.0213 | 0.0065 | -24.437 | 1.82 | 29468569799 | 2368 |
| داروسازی کوثر | 0.0425 | 0.1044 | -25.402 | 1.125 | 6.42E+11 | 46727 |
| کشاورزی دامپروری مگسال | 0.0417 | 0.09 | -20.995 | 0.595 | 9.88E+11 | 50241 |
| تامین ماسه | 0.0407 | 0.104 | -26.598 | 1.099 | 2.62E+12 | 82625 |
| نورد آلومینیوم | 0.046 | 0.115 | -26.508 | 0.523 | 3.49E+12 | 177903 |
| پگاه اصفهان | 0.041 | 0.0059 | -21.904 | 1.691 | 1.04E+11 | 4305 |
| شیشه و گاز | 0.0407 | 0.0746 | -24.11 | 1.293 | 9.76E+11 | 39934 |
| منگنز ایران | 0.0415 | 0.1528 | -25.744 | 0.746 | 2.03E+12 | 88975 |
| آبسال | 0.0424 | 0.075 | -23.975 | 0.992 | 3.56E+11 | 18848 |
| کاشی الوند | 0.0407 | 0.1789 | -23.667 | 0.408 | 8.22E+11 | 40745 |
| لامپ پارس شهاب | 0.423 | 0.1088 | -21.733 | 1.293 | 2.32127E+11 | 12157 |
| چرخشگر | 0.0464 | 0.095 | -25.516 | 0.448 | 9.78E+11 | 68799 |
| رادیاتور ایران | 0.0412 | 0.0267 | -23.617 | 0.416 | 3.61612E+11 | 24940 |
| کمک فنر ایندامین | 0.0482 | 0.2863 | -22.348 | 0.642 | 6.24E+11 | 55086 |
| محورسازان ایران خودرو | 0.0444 | 0.078 | -25.912 | 1.575 | 1.01E+12 | 48339 |
| صنایع شیمیایی سینا | 0.0452 | 0.0218 | -21.623 | 0.718 | 3.22E+11 | 10910 |
| فراورده های تزریق ایران | 0.0229 | 0.0021 | -20.367 | 1.816 | 80097579524 | 3579 |

جدول ۳-۴. نتیجه محاسبات شاخص های نقدشوندگی سال ۱۳۹۸

| | PBA _{i,t} | TO | ADJILLQ | LM | TV | N _{i,t} |
|-----------------------|--------------------|--------|---------|-------|------------|------------------|
| فولاد مبارکه اصفهان | 0.0277 | 0.028 | -29.334 | 0.689 | 8.08E+13 | 1540159 |
| ایران خودرو | 0.0467 | 0.256 | -30.978 | 0.78 | 8.09E+13 | 1265018 |
| معدنی صنعتی چادرملو | 0.0333 | 0.109 | -31.018 | 1.635 | 2.18E+13 | 475718 |
| پتروشیمی جم | 0.0367 | 0.0709 | -29.203 | 0.651 | 1.55E+13 | 356667 |
| داروپخش | 0.0448 | 0.441 | -20.496 | 1.118 | 6.89E+12 | 174883 |
| ملی صنایع مس ایران | 0.0303 | 0.0524 | -31.497 | 0.869 | 5.85E+13 | 1029670 |
| شرکت سیار ایران | 0.0272 | 0.0682 | -26.981 | 1.506 | 1.19E+13 | 268311 |
| مبین انرژی خلیج فارس | 0.0384 | 0.476 | -30.814 | 1.023 | 1.31E+13 | 212438 |
| پالایش نفت تبریز | 0.0453 | 0.1304 | -29.989 | 1.473 | 1.38E+13 | 237597 |
| سرامیک صنعتی اردکان | 0.0426 | 0.3487 | -27.488 | 1.239 | 2.49E+13 | 547321 |
| سایپا | 0.0461 | 0.0767 | -31.389 | 0.75 | 9.41E+13 | 2506741 |
| ایران ترانسفر | 0.0431 | 0.3711 | -29.978 | 0.98 | 3.74E+13 | 742000 |
| گروه صنعتی بارز | 0.0458 | 0.2078 | -28.59 | 1.125 | 7.6749E+12 | 138717 |
| بهنوش | 0.0481 | 0.3139 | -26.415 | 1.225 | 3.07E+12 | 77188 |
| پارس الکترونیک | 0.047 | 0.2767 | -27.189 | 0.962 | 8.96E+11 | 36804 |
| تراکتورسازی ایران | 0.0792 | 0.1198 | -28.487 | 1.057 | 6.98E+12 | 151517 |
| حفاری شمال | 0.0755 | 0.4821 | -29.962 | 1.92 | 4.44E+12 | 142530 |
| داده پرداز | 0.09179 | 0.4147 | -27.624 | 0.739 | 8.62E+11 | 48809 |
| پالایش نفت اصفهان | 0.0325 | 0.057 | -31.264 | 0.874 | 4.297E+13 | 559622 |
| شیشه رازی | 0.0451 | 0.0094 | -29.173 | 1.838 | 1.37E+12 | 84466 |
| کارخانجات تولید بقندی | 0.04983 | 0.0306 | -27.954 | 0.811 | 1.327E+12 | 61098 |
| کارتن ایران | 0.0447 | 0.0227 | -28.857 | 1.064 | 1.81E+13 | 414759 |
| بیسکوئیت گرچی | 0.0466 | 0.1775 | -29.084 | 1.049 | 8.24E+12 | 248691 |

| | | | | | | |
|-------------------------|--------|--------|---------|-------|-----------|--------|
| گروه صنعتی سپاهان | 0.0502 | 0.2702 | -27.149 | 0.856 | 6.84E+12 | 146334 |
| صنعتی و معدنی گل گوهر | 0.312 | 0.0518 | -30.797 | 1.067 | 2.91E+13 | 503066 |
| ایران مریئوس | 0.0467 | 0.2378 | 25.797 | 0.637 | 1.87E+12 | 60476 |
| داروسازی کوثر | 0.0482 | 0.2535 | -28.115 | 1.328 | 6.42E+12 | 186263 |
| کشاورزی دامپروری مگسال | 0.05 | 0.2711 | -24.508 | 0.698 | 4.704E+12 | 152455 |
| تامین ماسه | 0.0501 | 0.2822 | -29.467 | 1.214 | 1.39E+13 | 217009 |
| نورد آلومینیوم | 0.0457 | 0.3853 | -28.207 | 0.829 | 7.72E+12 | 250023 |
| پگاه اصفهان | 0.0544 | 0.0464 | -26.592 | 0.658 | 3.26E+12 | 52755 |
| شیشه و گاز | 0.0463 | 0.214 | -28.434 | 0.805 | 1.01E+13 | 210031 |
| منگنز ایران | 0.0497 | 0.303 | -28.142 | 0.428 | 9.59E+12 | 150934 |
| آبسال | 0.043 | 0.2191 | -27.645 | 0.686 | 1.06E+13 | 213598 |
| کاشی الوند | 0.0436 | 0.2419 | -27.497 | 0.591 | 7.76E+12 | 202858 |
| لامپ پارس شهاب | 0.0541 | 0.3147 | -26.973 | 0.785 | 3.49E+12 | 100091 |
| چرخشگر | 0.053 | 0.3433 | -27.715 | 0.571 | 7.56E+12 | 235688 |
| رادیاتور ایران | 0.0468 | 0.1275 | -28.178 | 0.976 | 9.40E+12 | 159051 |
| کمک فبر اینداین | 0.0532 | 0.4947 | -27.575 | 0.932 | 5.65E+12 | 150166 |
| محورسازان ایران خودرو | 0.0475 | 0.2977 | -28.401 | 0.559 | 1.49E+13 | 473685 |
| صنایع شیمیایی سینا | 0.0543 | 0.0396 | -26.328 | 0.488 | 2.90E+12 | 80917 |
| فراورده های تزریق ایران | 0.0443 | 0.0348 | -25.374 | 0.262 | 2.61E+12 | 92535 |

جدول ۴-۴. نتیجه محاسبات شاخص های نقدشوندگی سال ۱۳۹۷

| | PTD | ES | SIZE | LEV | CP | CET |
|-----------------------|-------|--------|-------|-------|-------|-------|
| فولاد مبارکه اصفهان | 0.966 | 0.018 | 19.77 | 0.32 | 1.285 | 1.327 |
| ایران خودرو | 0.958 | 0.0269 | 19.27 | 1.321 | 1.676 | 1.454 |
| معدنی و صنعتی چادرملو | 0.97 | 0.026 | 18.4 | 0.263 | 2.269 | 1.496 |
| پتروشیمی جم | 0.917 | 0.021 | 17.31 | 0.389 | 1.689 | 1.041 |
| داروپخش | 0.908 | 0.022 | 14.72 | 0.718 | 1.444 | 2.608 |
| ملی صنایع مس ایران | 0.954 | 0.017 | 19.21 | 0.33 | 1.693 | 1.468 |
| شرکت سیار ایران | 0.979 | 0.01 | 19.23 | 0.528 | 1.385 | 1.395 |
| مبین انرژی خلیج فارس | 0.975 | 0.02 | 17.77 | 0.221 | 2.069 | 1.467 |
| پالایش نفت تبریز | 0.975 | 0.028 | 17.55 | 0.423 | 2.079 | 2.07 |
| سرامیک صنعتی اردکان | 0.954 | 0.026 | 14.03 | 0.286 | 2.03 | 2.278 |
| سایپا | 0.97 | 0.0278 | 19.36 | 1.551 | 1.666 | 1.515 |
| ایران ترانسفو | 0.983 | 0.0204 | 16.88 | 0.704 | 1.671 | 2.595 |
| گروه صنعتی بارز | 0.979 | 0.0222 | 16.92 | 0.795 | 1.222 | 3.526 |
| بهنوش | 0.987 | 0.0303 | 15.39 | 0.894 | 1.273 | 3.053 |
| پارس الکترونیک | 0.979 | 0.0297 | 13.82 | 0.078 | 1.888 | 2.124 |
| تراکتورسازی ایران | 0.966 | 0.0251 | 16.08 | 0.599 | 1.348 | 3.151 |
| حفاری شمال | 0.947 | 0.0255 | 16.05 | 0.686 | 1.587 | 3.4 |
| داده پردازی | 0.979 | 0.0254 | 14.19 | 0.441 | 1.678 | 3.503 |
| پالایش نفت اصفهان | 0.979 | 0.0216 | 18.68 | 0.297 | 1.722 | 1.283 |
| شیشه رازی | 0.974 | 0.0268 | 14.59 | 0.278 | 2.214 | 2.417 |
| کارخانجات تولیدی قندی | 0.924 | 0.0343 | 14.6 | 0.774 | 0.775 | 1.877 |
| کارتن ایران | 0.907 | 0.0323 | 13.35 | 0.593 | 2.699 | 3.969 |

| | | | | | | |
|-------------------------|-------|---------|-------|-------|-------|-------|
| بیسکوئیت گرجی | 0.803 | 0.0278 | 13.6 | 0.457 | 2.108 | 3.622 |
| گروه صنعتی سپاهان | 0.966 | 0.0303 | 15.12 | 0.529 | 1.516 | 3.713 |
| صنعتی معدنی گل گوهر | 0.983 | 0.0214 | 19.12 | 0.478 | 1.893 | 1.396 |
| ایران مریнос | 0.711 | 0.0223 | 12.7 | 0.499 | 0.998 | 3.053 |
| داروسازی کوثر | 0.97 | 0.0281 | 14.75 | 0.527 | 1.095 | 2.048 |
| کشاوری دامپروری مگسال | 0.983 | 0.0269 | 14.11 | 0.657 | 1.195 | 3.352 |
| تامین ماسه | 0.949 | 0.03331 | 12.68 | 0.243 | 3.268 | 2.169 |
| نورد آلومینیوم | 0.987 | 0.0271 | 14.41 | 0.52 | 1.096 | 2.299 |
| پگاه اصفهان | 0.907 | 0.0329 | 14.42 | 0.714 | 2.231 | 3.533 |
| شیشه و گاز | 0.928 | 0.027 | 14.47 | 0.581 | 0.856 | 1.941 |
| منگنز | 0.974 | 0.0306 | 13.51 | 0.294 | 0.696 | 2.406 |
| آبسال | 0.948 | 0.0296 | 14.43 | 0.636 | 0.64 | 2.387 |
| کاشی الوند | 0.991 | 0.0293 | 14.49 | 0.629 | 1.43 | 3.956 |
| لامپ پارس شهاب | 0.937 | 0.0333 | 14.73 | 0.856 | 2.174 | 2.974 |
| چرخشگر | 0.991 | 0.0303 | 15.68 | 0.197 | 2.514 | 2.24 |
| رادیاتور ایران | 0.991 | 0.0266 | 15.68 | 0.102 | 1.403 | 1.774 |
| کمک فنر ایندامین | 0.979 | 0.0305 | 13.79 | 0.803 | 1.414 | 3.187 |
| محورسازان ایران خودرو | 0.815 | 0.0325 | 14.47 | 0.508 | 1.175 | 2.559 |
| صنایع شیمیایی سینا | 0.974 | 0.0345 | 13.2 | 0.346 | 2.652 | 3.959 |
| فراورده‌های تزریق ایران | 0.891 | 0.0183 | 14.29 | 0.172 | 2.236 | 3.559 |

جدول ۴-۵. نتیجه محاسبات شاخص‌های نقدشوندگی سال ۱۳۹۸

| | PTD | ES | SIZE | LEV | CP | CET |
|-----------------------|-------|--------|-------|-------|-------|-------|
| فولاد مبارکه اصفهان | 0.97 | 0.0156 | 20.18 | 0.361 | 1.392 | 1.205 |
| ایران خودرو | 0.966 | 0.0292 | 20.14 | 0.716 | 1.611 | 1.589 |
| معدنی و صنعتی چادرملو | 0.941 | 0.0187 | 18.68 | 0.127 | 2.532 | 1.203 |
| پتروشیمی جم | 0.97 | 0.0185 | 18.61 | 0.397 | 2.321 | 1.514 |
| داروپخش | 0.958 | 0.0312 | 15.98 | 0.511 | 4.29 | 2.254 |
| ملی صنایع مس ایران | 0.962 | 0.0151 | 19.61 | 0.242 | 2.409 | 1.204 |
| شرکت سیار ایران | 0.928 | 0.0159 | 19.46 | 0.653 | 1.897 | 1.286 |
| مبین انرژی خلیج فارس | 0.916 | 0.0206 | 18.17 | 0.38 | 2.907 | 1.343 |
| پالایش نفت تبریز | 0.949 | 0.0274 | 17.47 | 0.355 | 4.439 | 1.769 |
| سرامیک صنعتی اردکان | 0.953 | 0.0284 | 14.45 | 0.254 | 4.633 | 1.927 |
| سایپا | 0.966 | 0.0295 | 19.42 | 1.09 | 2.946 | 1.853 |
| ایران ترانسفر | 0.958 | 0.258 | 17.11 | 0.669 | 3.603 | 1.381 |
| گروه صنعتی بارز | 0.958 | 0.0345 | 17.18 | 0.701 | 2.133 | 1.366 |
| بهنوش | 0.949 | 0.0288 | 15.44 | 0.797 | 3.669 | 2.002 |
| پارس الکترونیک | 0.958 | 0.0325 | 13.76 | 0.475 | 1.971 | 1.795 |
| تراکتورسازی ایران | 0.928 | 0.0271 | 16.5 | 0.498 | 1.904 | 1.598 |
| حفاری شمال | 0.924 | 0.0288 | 17.33 | 0.631 | 3.317 | 1.624 |
| داده پردازی | 0.97 | 0.0333 | 14.8 | 0.492 | 4.495 | 1.658 |
| پالایش نفت اصفهان | 0.962 | 0.017 | 19 | 0.42 | 2.993 | 1.548 |
| شیشه رازی | 0.928 | 0.0286 | 15 | 0.341 | 1.419 | 1.486 |
| کارخانجات تولیدی قندی | 0.966 | 0.034 | 14.97 | 0.763 | 0.775 | 2.027 |

| | | | | | | |
|-------------------------|-------|--------|-------|-------|-------|-------|
| کارتن ایران | 0.953 | 0.0295 | 14.89 | 0.14 | 3.443 | 2.138 |
| بیسکویت گرجی | 0.912 | 0.0347 | 14.28 | 0.383 | 3.542 | 1.535 |
| گروه صنعتی سپاهان | 0.966 | 0.0317 | 15.98 | 0.495 | 4.294 | 1.48 |
| صنعتی و معدنی گل گوهر | 0.928 | 0.0161 | 19.54 | 0.498 | 1.959 | 1.373 |
| ایران مریئوس | 0.97 | 0.032 | 12.87 | 0.491 | 0.915 | 1.792 |
| داروسازی کوثر | 0.945 | 0.0347 | 15.17 | 0.434 | 2.789 | 1.724 |
| کشاوری دامپروری مگسال | 0.958 | 0.0344 | 14.17 | 0.64 | 3.919 | 2.702 |
| تامین ماسه | 0.924 | 0.0344 | 13.03 | 0.244 | 3.886 | 2.608 |
| نورد آلومینیوم | 0.949 | 0.0303 | 14.72 | 0.52 | 4.006 | 1.906 |
| پگاه اصفهان | 0.95 | 0.0348 | 15.98 | 0.621 | 3.34 | 2.038 |
| شیشه و گاز | 0.953 | 0.0306 | 14.76 | 0.72 | 1.638 | 1.754 |
| منگنز ایران | 0.974 | 0.0329 | 13.98 | 0.232 | 4.245 | 1.892 |
| آبسال | 0.953 | 0.0282 | 14.7 | 0.531 | 1.934 | 1.902 |
| کاشی الوند | 0.951 | 0.0305 | 14.67 | 0.663 | 3.114 | 2.154 |
| لامپ پارس شهاب | 0.952 | 0.0378 | 14.75 | 0.588 | 2.351 | 2.112 |
| چرخشگر | 0.966 | 0.0349 | 14.98 | 0.566 | 2.935 | 1.602 |
| رادیاتور ایران | 0.941 | 0.0317 | 14.49 | 0.548 | 1.692 | 2.072 |
| کمک فنر ایندամین | 0.945 | 0.0376 | 14.19 | 0.767 | 3.681 | 1.744 |
| محورسازان ایران خودرو | 0.962 | 0.0352 | 14.46 | 0.514 | 3.459 | 2.767 |
| صنایع شیمیایی سینا | 0.97 | 0.0374 | 13.36 | 0.2 | 2.223 | 2.436 |
| فراورده های تزریق ایران | 0.983 | 0.0283 | 14.59 | 0.187 | 6.315 | 2.007 |

۳-۲-۴ الگوریتم خوشه‌بندی

الگوریتم‌های خوشه‌بندی مبتنی بر ضریب تشابه نیازمند یک ماتریس وقوع می‌باشند که محاسبه ضریب تشابه براساس این ماتریس باینری است. از آنجاییکه شاخص‌های نقدشوندگی بدست آمده اعدادی غیر از صفر و یک هستند، برای رسیدن به این ماتریس، میانگین هر شاخص محاسبه و با توجه به میانگین شاخص‌ها به اعداد صفر و یک کدگذاری شدند. در جدول ۴-۶ و جدول ۴-۷ زیر ماتریس وقوع برای هر سال ارائه شده است:

جدول ۴-۶. ماتریس وقوع شاخص‌های سال ۱۳۹۷

| | PBA _{it} | TO | ADJI | LM | TV | N _{it} | PTD | ES | SIZE | LEV | CP | CET |
|-----------------------|-------------------|----|------|----|----|-----------------|-----|----|------|-----|----|-----|
| فولاد مبارکه اصفهان | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| ایران خودرو | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| معدنی و صنعتی چادرملو | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| پتروشیمی جم | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |
| داروپخش | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| ملی صنایع مس ایران | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| شرکت سیار ایران | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| مبین انرژی خلیج فارس | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| پالایش نفت تبریز | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 |
| سرامیک صنعتی اردکان | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| سایپا | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| ایران ترانسفو | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| گروه صنعتی بارز | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |
| بهنوش | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |
| پارس الکترونیک | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| تراکتورسازی ایران | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |
| حفاری شمال | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| داده پردازی | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| پالایش نفت اصفهان | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| شیشه رازی | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| کارخانجات تولیدی قندی | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| کارتن ایران | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| بیسکوئیت گرجی | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| گروه صنعتی سپاهان | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 |
| صنعتی و معدنی گل گوهر | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| ایران مرینوس | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| داروسازی کوثر | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| کشاوری دامپروری مگسال | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| تامین ماسه | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| نورد آلومینیوم | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| پگاه اصفهان | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| شیشه و گاز | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| منگنز ایران | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| آبسال | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| کاشی الوند | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| لامپ پارس شهاب | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| رخشگر | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| رادیاتور ایران | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 |
| کمک فنر ایندامین | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |

| | | | | | | | | | | | | |
|-------------------------|---|---|---|---|---|---|---|---|---|---|---|---|
| محورسازان ایران خودرو | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| صنایع شیمیایی سینا | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 |
| فراورده های تزریق ایران | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |

جدول ۴-۷. ماتریس وقوع شاخص های سال ۱۳۹۸

| | PBA | TO | ADJI | LM | TV | N _{i,t} | PTD | ES | SIZE | LEV | CP | CET |
|------------------------|-----|----|------|----|----|------------------|-----|----|------|-----|----|-----|
| فولاد مبارکه اصفهان | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| ایران خودرو | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| معدنی و صنعتی چادرملو | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 |
| پتروشیمی جم | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| داروپخش | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| ملی صنایع مس ایران | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| شرکت سیار ایران | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| مبین انرژی خلیج فارس | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |
| پالایش نفت تبریز | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| سرامیک صنعتی اردکان | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| سایپا | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| ایران ترانسفو | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| گروه صنعتی بارز | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| بهنوش | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| پارس الکترونیک | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| تراکتورسازی ایران | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| حفاری شمال | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| داده پردازی | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| پالایش نفت اصفهان | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| شیشه رازی | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| کارخانجات تولیدی قندی | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| کارتن ایران | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| بیسکویت گرجی | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| گروه صنعتی سپاهان | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| صنعتی و معدنی گل گوهر | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| ایران مریئوس | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| داروسازی کوثر | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| کشاورزی دامپروری مگسال | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| تامین ماسه | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| نورد آلومینیوم | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| پگاه اصفهان | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| شیشه و گاز | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| منگنز ایران | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| آبسال | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| کاشی الوند | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| لامپ پارس شهاب | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

| | | | | | | | | | | | | |
|-------------------------|---|---|---|---|---|---|---|---|---|---|---|---|
| چرخشگر | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| رادیاتور ایران | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| کمک فنر ایندامین | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| محورسازان ایران خودرو | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 |
| صنایع شیمیایی سینا | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| فراورده های تزریق ایران | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |

با توجه به ماتریس باینری وقوع، ضریب تشابه تطابق نسبی برای هر جفت نمونه آماری محاسبه گردید و براساس این ضریب، الگوریتم خوشه‌بندی پیوند کامل اجرا شد. خوشه‌بندی تا زمانی ادامه یافت که هر نمونه آماری در یک خوشه قرار گیرد. برای هر سال، ۵ خوشه بدست آمد که تقریباً برابر با مقدار $\sqrt{\frac{n}{2}}$ است. نتایج خوشه‌بندی به صورت جدول ۴-۸ و جدول ۴-۹ است:

جدول ۴-۸. نتیجه خوشه‌بندی نمونه آماری سال ۱۳۹۷

| | |
|----------------|---|
| C ₁ | فولاد مبارکه اصفهان، ملی صنایع مس ایران، پالایش اصفهان، صنعتی و معدنی گوهر، صنعتی و معدنی چادرملو، ایران خودرو، سایپا، مبین انرژی خلیج فارس، پالایش تبریز، شرکت ارتباطات سیار ایران، ترانسفر، گروه صنعتی بارز، تراکتورسازی ایران، بهنوش، آبسال، گروه صنعتی سپاهان، پارس الکترونیک، چرخشگر، داده‌پردازی، داروسازی کوثر، کشاورزی و دامپروری مگسال، کاشی الوند |
| C ₂ | شیشه رازی، صنایع شیمیایی سینا، رادیاتور ایران، محورسازان ایران خودرو |
| C ₃ | پتروشیمی جم، حفاری شمال، کارت ایران، شیشه و گاز، کارخانجات تولید قندی |
| C ₄ | داروپخش، ایران مرینوس، پگاه اصفهان، فراورده‌های تزریقی ایران، بیسکویت گرجی |
| C ₅ | سرامیک‌های صنعتی اردکان، نورد آلومینیوم، معادن منگنز ایران، تأمین ماسه، شهاب الکترونیک، کمک فنر ایندامین |

جدول ۴-۹.

| | |
|----------------|---|
| C ₁ | فولاد مبارکه اصفهان، ملی صنایع مس ایران، پالایش اصفهان، ایران خودرو، سایپا، ترانسفر، پتروشیمی جم، پارس الکترونیک، آبسال، کارخانجات تولید قندی، شیشه و گاز |
| C ₂ | صنعتی و معدنی گوهر، صنعتی و معدنی چادرملو، مبین انرژی خلیج فارس، پالایش تبریز، شرکت ارتباطات سیار ایران، گروه صنعتی بارز، بیسکویت گرجی، شیشه رازی |
| C ₃ | داروپخش، بهنوش، حفاری شمال |
| C ₄ | سرامیک‌های صنعتی اردکان، کارتن ایران، محورسازان ایران خودرو، تأمین ماسه، داده‌پردازی، معادن منگنز ایران، نورد آلومینیوم، گروه صنعتی سپاهان |
| C ₅ | کشاورزی و دامپروری مگسال، کاشی الوند، شهاب الکترونیک، پگاه اصفهان، فراورده‌های تزریقی ایران، رادیاتور ایران، ایران مریوس، صنایع شیمیایی سینا، چرخشگر، کمک فنر ایندامین، داروسازی کوثر |

۴-۳ ارزشیابی خوشه‌بندی

از آنجاییکه خوشه‌بندی یک فرایند بدون نظارت است، بررسی صحت نتیجه خوشه‌بندی به راحتی امکان‌پذیر نیست. بنابراین احتیاج به معیارهای مناسب برای بررسی کارایی یک روش خوشه‌بندی در ارزیابی خوشه‌ها ضروری به نظر می‌رسد. به این منظور در این پژوهش از شاخص نیم رخ استفاده گردید که مقدار این شاخص بین ۱- و ۱ تغییر می‌کند. کوچک بودن مقدار نیم‌رخ برای خوشه‌ها بیانگر ضعیف بودن نتایج خوشه‌بندی است. نتایج شاخص نیم‌رخ در جدول ۴-۱۰ بیان شده است:

جدول ۴-۱۰. نتایج ارزشیابی خوشه‌بندی براساس شاخص سیلوئت

| سال ۹۷ | سال ۹۸ |
|--------|--------|
| ۰.۶۳ | ۰.۶۱۸ |

۴-۴ نتایج تحلیل واریانس و روش کمترین تفاوت معنی‌دار

در این پژوهش برای بررسی فرضیه آماری، تحلیل واریانس و روش کمترین تفاوت معنی‌دار برای یافتن علل تفاوت خوشه‌ها انجام شد. به این منظور درصد هر شاخص در هر خوشه بدست آمد و در ادامه میانگین هر شاخص محاسبه گردید که در جدول ۴-۱۱ و جدول ۴-۱۲ بیان شده است:

جدول ۴-۱۱. درصد هر شاخص در هر خوشه سال ۱۳۹۷

| | LEV | LM | PTD | PBA _{i,t} | CP | TO | CET | ADJI | SIZE | ES | TV | N _{i,t} |
|----------------|-------|-------|-------|--------------------|-------|-------|-------|-------|-------|-----------|-----------|------------------|
| C ₁ | 0.681 | 0.954 | 0.954 | 0.954 | 0.59 | 0.545 | 0.409 | 0.636 | 0.727 | 0.5 | 0.54 5 | 0.54 |
| C ₂ | 1 | 0.75 | 0.75 | 1 | 0.5 | 0.25 | 0.5 | 0.5 | 0.25 | 0 | 0.25 | 0 |
| C ₃ | 0.2 | 0.2 | 0.2 | 0.8 | 0.2 | 0.2 | 0.4 | 0.8 | 0.4 | 0.4 | 0.2 | 0 |
| C ₄ | 0.6 | 0 | 0 | 1 | 0.6 | 0 | 1 | 0.2 | 0.2 | 0.6 | 0 | 0 |
| C ₅ | 0.666 | 1 | 0.333 | 0.5 | 0.5 | 0.333 | 0.333 | 0.666 | 0.166 | 0.16 6 | 0.33 3 | |
| AVE | 0.629 | 0.580 | 0.547 | 0.850 | 0.478 | 0.365 | 0.528 | 0.560 | 0.348 | 0.33 3 | 0.26 5 | 0.175 |

جدول ۴-۱۲. درصد هر شاخص در هر خوشه سال ۱۳۹۸

| | LEV | LM | PTD | PBA _{i,t} | CP | TO | CET | ADJI | SIZE | ES | TV | N _{i,t} |
|----------------|-------|-------|-------|--------------------|-------|-------|-------|-------|-------|-------|-----------|------------------|
| C ₁ | 0.545 | 0.818 | 1 | 1 | 0.545 | 0.454 | 0.272 | 1 | 0.636 | 1 | 0.72 7 | 0.818 |
| C ₂ | 0.555 | 0 | 0.111 | 0.777 | 0.222 | 0.111 | 0.111 | 0.888 | 0.777 | 0.777 | 0.22 2 | 0.222 |
| C ₃ | 0.333 | 1 | 0.333 | 0.666 | 1 | 1 | 0.666 | 0.333 | 0.666 | 1 | 0 | 0 |
| C ₄ | 0.5 | 0.75 | 0.75 | 1 | 1 | 0.875 | 0.75 | 0.875 | 0.125 | 0.75 | 0.25 | 0.375 |
| C ₅ | 0.363 | 0.363 | 0.454 | 0.727 | 0.454 | 0.545 | 0.545 | 0.9 | 0.181 | 0.454 | 0 | 0 |
| AVE | 0.459 | 0.586 | 0.529 | 0.834 | 0.644 | 0.597 | 0.469 | 0.801 | 0.477 | 0.769 | 0.23 9 | 0.283 |

نتایج تحلیل واریانس در جدول ۴-۱۳ بیان شده است:

جدول ۴-۱۳. نتایج تحلیل واریانس برای شاخص‌های نقدشوندگی

| سال ۹۷ | | | | | |
|--------------|--------------------|------------|----------------------|-------|----------|
| منبع تغییرات | مجموع توان‌های دوم | درجه آزادی | میانگین توان‌های دوم | F | F بحرانی |
| تیمارها | 1.865 | 11 | 0.169 | 2.066 | 1.994 |
| خطاها | 3.939 | 48 | 0.082 | | |
| سال ۹۸ | | | | | |
| تیمارها | 2.018 | 11 | 0.183 | 2.096 | 1.994 |
| خطاها | 4.201 | 48 | 0.087 | | |

با توجه به آنکه مقدار F در هر دو سال از مقدار F بحرانی بزرگتر است، بنابراین فرض صفر در سطح معنی دار ۵ درصد رد شده و اختلاف در میانگین‌ها شانسی و تصادفی نمی‌باشد.

از مقایسه‌های تبعی یا پس از تجربه زمانی استفاده می‌شود که در تحلیل واریانس فرض صفر رد شده و فرضیه مخالف پذیرفته شود، یعنی دست کم میانگین دو گروه با هم اختلاف داشته باشند. در واقع از این مقایسه در مورد اینکه اختلاف میانگین کدام گروه‌ها معنی دار است استفاده می‌شود. با توجه به معنی داری تحلیل واریانس، آزمون کمترین تفاوت معنی دار به منظور یافتن آنکه اختلاف میانگین کدام شاخص‌ها معنی دار است انجام شد که نتایج به صورت جدول ۴-۱۴ و جدول ۴-۱۵ است:

| جدول ۴-۱۴. نتایج اختلاف میانگین شاخص‌های نقدشوندگی سال ۱۳۹۷ | | | | | |
|---|------------------------------------|------------------------------------|---------------------------------------|--|---------------------------------------|
| $\bar{x}_1 - \bar{x}_2 = 0.048$ | $\bar{x}_2 - \bar{x}_1 = 0.048$ | $\bar{x}_3 - \bar{x}_1 = 0.082$ | $\bar{x}_4 - \bar{x}_1 = -0.221$ | $\bar{x}_5 - \bar{x}_1 = 0.151$ | $\bar{x}_6 - \bar{x}_1 = 0.263$ |
| $\bar{x}_1 - \bar{x}_3 = 0.082$ | $\bar{x}_2 - \bar{x}_3 = 0.033$ | $\bar{x}_3 - \bar{x}_2 = 0.033$ | $\bar{x}_4 - \bar{x}_2 = -0.27$ | $\bar{x}_5 - \bar{x}_2 = 0.102$ | $\bar{x}_6 - \bar{x}_2 = 0.215$ |
| $\bar{x}_1 - \bar{x}_4 = -0.221$ | $\bar{x}_2 - \bar{x}_4 = -0.27$ | $\bar{x}_3 - \bar{x}_4 = -0.303$ | $\bar{x}_4 - \bar{x}_3 = -0.303$ | $\bar{x}_5 - \bar{x}_3 = 0.069$ | $\bar{x}_6 - \bar{x}_3 = 0.181$ |
| $\bar{x}_1 - \bar{x}_5 = 0.151$ | $\bar{x}_2 - \bar{x}_5 = 0.102$ | $\bar{x}_3 - \bar{x}_5 = 0.069$ | $\bar{x}_4 - \bar{x}_5 = 0.372$ | $\bar{x}_5 - \bar{x}_4 = 0.372$ | $\bar{x}_6 - \bar{x}_4 = 0.485$ |
| $\bar{x}_1 - \bar{x}_6 = 0.263$ | $\bar{x}_2 - \bar{x}_6 = 0.215$ | $\bar{x}_3 - \bar{x}_6 = 0.181$ | $\bar{x}_4 - \bar{x}_6 = 0.485$ | $\bar{x}_5 - \bar{x}_6 = 0.112$ | $\bar{x}_4 - \bar{x}_5 = 0.112$ |
| $\bar{x}_1 - \bar{x}_7 = 0.101$ | $\bar{x}_2 - \bar{x}_7 = 0.52$ | $\bar{x}_3 - \bar{x}_7 = 0.019$ | $\bar{x}_4 - \bar{x}_7 = 0.322$ | $\bar{x}_5 - \bar{x}_7 = -0.05$ | $\bar{x}_6 - \bar{x}_7 = -0.162$ |
| $\bar{x}_1 - \bar{x}_8 = 0.069$ | $\bar{x}_2 - \bar{x}_8 = -0.0203$ | $\bar{x}_3 - \bar{x}_8 = -0.013$ | $\bar{x}_4 - \bar{x}_8 = 0.290$ | $\bar{x}_5 - \bar{x}_8 = -0.082$ | $\bar{x}_6 - \bar{x}_8 = -0.194$ |
| $\bar{x}_1 - \bar{x}_9 = 0.348$ | $\bar{x}_2 - \bar{x}_9 = 0.232$ | $\bar{x}_3 - \bar{x}_9 = 0.198$ | $\bar{x}_4 - \bar{x}_9 = 0.502$ | $\bar{x}_5 - \bar{x}_9 = 0.129$ | $\bar{x}_6 - \bar{x}_9 = 0.016$ |
| $\bar{x}_1 - \bar{x}_{10} = 0.269$ | $\bar{x}_2 - \bar{x}_{10} = 0.247$ | $\bar{x}_3 - \bar{x}_{10} = 0.214$ | $\bar{x}_4 - \bar{x}_{10} = 0.517$ | $\bar{x}_5 - \bar{x}_{10} = 0.144$ | $\bar{x}_6 - \bar{x}_{10} = 0.032$ |
| $\bar{x}_1 - \bar{x}_{11} = 0.363$ | $\bar{x}_2 - \bar{x}_{11} = 0.315$ | $\bar{x}_3 - \bar{x}_{11} = 0.281$ | $\bar{x}_4 - \bar{x}_{11} = 0.585$ | $\bar{x}_5 - \bar{x}_{11} = 0.212$ | $\bar{x}_6 - \bar{x}_{11} = 0.1$ |
| $\bar{x}_1 - \bar{x}_{12} = 0.453$ | $\bar{x}_2 - \bar{x}_{12} = 0.405$ | $\bar{x}_3 - \bar{x}_{12} = 0.371$ | $\bar{x}_4 - \bar{x}_{12} = 0.675$ | $\bar{x}_5 - \bar{x}_{12} = 0.302$ | $\bar{x}_6 - \bar{x}_{12} = 0.19$ |
| $\bar{x}_7 - \bar{x}_1 = -0.101$ | $\bar{x}_8 - \bar{x}_1 = 0.069$ | $\bar{x}_9 - \bar{x}_1 = 0.348$ | $\bar{x}_{10} - \bar{x}_1 = 0.269$ | $\bar{x}_{11} - \bar{x}_1 = 0.363$ | $\bar{x}_{12} - \bar{x}_1 = 0.453$ |
| $\bar{x}_7 - \bar{x}_2 = -0.052$ | $\bar{x}_8 - \bar{x}_2 = 0.020$ | $\bar{x}_9 - \bar{x}_2 = 0.232$ | $\bar{x}_{10} - \bar{x}_2 = 0.247$ | $\bar{x}_{11} - \bar{x}_2 = 0.315$ | $\bar{x}_{12} - \bar{x}_2 = 0.405$ |
| $\bar{x}_7 - \bar{x}_3 = -0.019$ | $\bar{x}_8 - \bar{x}_3 = -0.013$ | $\bar{x}_9 - \bar{x}_3 = 0.198$ | $\bar{x}_{10} - \bar{x}_3 = 0.214$ | $\bar{x}_{11} - \bar{x}_3 = 0.281$ | $\bar{x}_{12} - \bar{x}_3 = 0.371$ |
| $\bar{x}_7 - \bar{x}_4 = -0.332$ | $\bar{x}_8 - \bar{x}_4 = 0.29$ | $\bar{x}_9 - \bar{x}_4 = 0.502$ | $\bar{x}_{10} - \bar{x}_4 = 0.517$ | $\bar{x}_{11} - \bar{x}_4 = 0.585$ | $\bar{x}_{12} - \bar{x}_4 = 0.675$ |
| $\bar{x}_7 - \bar{x}_5 = -0.050$ | $\bar{x}_8 - \bar{x}_5 = -0.082$ | $\bar{x}_9 - \bar{x}_5 = 0.129$ | $\bar{x}_{10} - \bar{x}_5 = 0.144$ | $\bar{x}_{11} - \bar{x}_5 = 0.212$ | $\bar{x}_{12} - \bar{x}_5 = 0.302$ |
| $\bar{x}_7 - \bar{x}_6 = -0.162$ | $\bar{x}_8 - \bar{x}_6 = -0.194$ | $\bar{x}_9 - \bar{x}_6 = 0.016$ | $\bar{x}_{10} - \bar{x}_6 = 0.032$ | $\bar{x}_{11} - \bar{x}_6 = 0.1$ | $\bar{x}_{12} - \bar{x}_6 = 0.19$ |
| $\bar{x}_7 - \bar{x}_8 = -0.032$ | $\bar{x}_8 - \bar{x}_7 = -0.032$ | $\bar{x}_9 - \bar{x}_7 = 0.179$ | $\bar{x}_{10} - \bar{x}_7 = 0.195$ | $\bar{x}_{11} - \bar{x}_7 = 0.262$ | $\bar{x}_{12} - \bar{x}_7 = 0.352$ |
| $\bar{x}_7 - \bar{x}_9 = 0.179$ | $\bar{x}_8 - \bar{x}_9 = 0.211$ | $\bar{x}_9 - \bar{x}_8 = 0.211$ | $\bar{x}_{10} - \bar{x}_8 = 0.227$ | $\bar{x}_{11} - \bar{x}_8 = 0.294$ | $\bar{x}_{12} - \bar{x}_8 = 0.384$ |
| $\bar{x}_7 - \bar{x}_{10} = 0.195$ | $\bar{x}_8 - \bar{x}_{10} = 0.227$ | $\bar{x}_9 - \bar{x}_{10} = 0.015$ | $\bar{x}_{10} - \bar{x}_9 = -0.015$ | $\bar{x}_{11} - \bar{x}_9 = 0.083$ | $\bar{x}_{12} - \bar{x}_9 = 0.173$ |
| $\bar{x}_7 - \bar{x}_{11} = 0.262$ | $\bar{x}_8 - \bar{x}_{11} = 0.294$ | $\bar{x}_9 - \bar{x}_{11} = 0.083$ | $\bar{x}_{10} - \bar{x}_{11} = 0.067$ | $\bar{x}_{11} - \bar{x}_{10} = -0.067$ | $\bar{x}_{12} - \bar{x}_{10} = 0.157$ |
| $\bar{x}_7 - \bar{x}_{12} = 0.352$ | $\bar{x}_8 - \bar{x}_{12} = 0.384$ | $\bar{x}_9 - \bar{x}_{12} = 0.173$ | $\bar{x}_{10} - \bar{x}_{12} = 0.157$ | $\bar{x}_{11} - \bar{x}_{12} = 0.09$ | $\bar{x}_{12} - \bar{x}_{11} = 0.09$ |

جدول ۴-۱۵. نتایج اختلاف میانگین شاخص‌های نقدشوندگی سال ۱۳۹۸

| | | | | | |
|-------------------------------------|------------------------------------|-------------------------------------|---------------------------------------|---------------------------------------|---------------------------------------|
| $\bar{x}_1 - \bar{x}_2 = -0.126$ | $\bar{x}_2 - \bar{x}_1 = 0.126$ | $\bar{x}_3 - \bar{x}_1 = 0.070$ | $\bar{x}_4 - \bar{x}_1 = -0.374$ | $\bar{x}_5 - \bar{x}_1 = 0.184$ | $\bar{x}_6 - \bar{x}_1 = -0.137$ |
| $\bar{x}_1 - \bar{x}_3 = -0.070$ | $\bar{x}_2 - \bar{x}_3 = 0.056$ | $\bar{x}_3 - \bar{x}_2 = -0.056$ | $\bar{x}_4 - \bar{x}_2 = -0.247$ | $\bar{x}_5 - \bar{x}_2 = 0.058$ | $\bar{x}_6 - \bar{x}_2 = -0.010$ |
| $\bar{x}_1 - \bar{x}_4 = -0.374$ | $\bar{x}_2 - \bar{x}_4 = -0.247$ | $\bar{x}_4 - \bar{x}_3 = -0.304$ | $\bar{x}_4 - \bar{x}_3 = 0.304$ | $\bar{x}_5 - \bar{x}_3 = 0.114$ | $\bar{x}_6 - \bar{x}_3 = -0.067$ |
| $\bar{x}_1 - \bar{x}_5 = -0.184$ | $\bar{x}_2 - \bar{x}_5 = -0.058$ | $\bar{x}_3 - \bar{x}_5 = -0.114$ | $\bar{x}_4 - \bar{x}_5 = 0.189$ | $\bar{x}_5 - \bar{x}_4 = -0.189$ | $\bar{x}_6 - \bar{x}_4 = 0.237$ |
| $\bar{x}_1 - \bar{x}_6 = -0.137$ | $\bar{x}_2 - \bar{x}_6 = -0.010$ | $\bar{x}_3 - \bar{x}_6 = -0.067$ | $\bar{x}_4 - \bar{x}_6 = 0.237$ | $\bar{x}_5 - \bar{x}_6 = 0.047$ | $\bar{x}_6 - \bar{x}_5 = 0.047$ |
| $\bar{x}_1 - \bar{x}_7 = -0.009$ | $\bar{x}_2 - \bar{x}_7 = 0.117$ | $\bar{x}_3 - \bar{x}_7 = 0.06$ | $\bar{x}_4 - \bar{x}_7 = 0.365$ | $\bar{x}_5 - \bar{x}_7 = 0.175$ | $\bar{x}_6 - \bar{x}_7 = 0.128$ |
| $\bar{x}_1 - \bar{x}_8 = -0.341$ | $\bar{x}_2 - \bar{x}_8 = -0.214$ | $\bar{x}_3 - \bar{x}_8 = -0.271$ | $\bar{x}_4 - \bar{x}_8 = 0.033$ | $\bar{x}_5 - \bar{x}_8 = -0.156$ | $\bar{x}_6 - \bar{x}_8 = -0.208$ |
| $\bar{x}_1 - \bar{x}_9 = -0.017$ | $\bar{x}_2 - \bar{x}_9 = 0.108$ | $\bar{x}_3 - \bar{x}_9 = 0.052$ | $\bar{x}_4 - \bar{x}_9 = 0.356$ | $\bar{x}_5 - \bar{x}_9 = 0.166$ | $\bar{x}_6 - \bar{x}_9 = 0.119$ |
| $\bar{x}_1 - \bar{x}_{10} = -0.336$ | $\bar{x}_2 - \bar{x}_{10} = -0.21$ | $\bar{x}_3 - \bar{x}_{10} = -0.266$ | $\bar{x}_4 - \bar{x}_{10} = 0.037$ | $\bar{x}_5 - \bar{x}_{10} = 0.152$ | $\bar{x}_6 - \bar{x}_{10} = -0.199$ |
| $\bar{x}_1 - \bar{x}_{11} = 0.219$ | $\bar{x}_2 - \bar{x}_{11} = 0.346$ | $\bar{x}_3 - \bar{x}_{11} = 0.289$ | $\bar{x}_4 - \bar{x}_{11} = 0.594$ | $\bar{x}_5 - \bar{x}_{11} = 0.404$ | $\bar{x}_6 - \bar{x}_{11} = 0.357$ |
| $\bar{x}_1 - \bar{x}_{12} = 0.176$ | $\bar{x}_2 - \bar{x}_{12} = 0.303$ | $\bar{x}_3 - \bar{x}_{12} = 0.246$ | $\bar{x}_4 - \bar{x}_{12} = 0.551$ | $\bar{x}_5 - \bar{x}_{12} = 0.361$ | $\bar{x}_6 - \bar{x}_{12} = 0.314$ |
| $\bar{x}_7 - \bar{x}_1 = -0.009$ | $\bar{x}_8 - \bar{x}_1 = -0.341$ | $\bar{x}_9 - \bar{x}_1 = 0.017$ | $\bar{x}_{10} - \bar{x}_1 = 0.336$ | $\bar{x}_{11} - \bar{x}_1 = 0.219$ | $\bar{x}_{12} - \bar{x}_1 = 0.176$ |
| $\bar{x}_7 - \bar{x}_2 = 0.117$ | $\bar{x}_8 - \bar{x}_2 = -0.214$ | $\bar{x}_9 - \bar{x}_2 = -0.108$ | $\bar{x}_{10} - \bar{x}_2 = 0.21$ | $\bar{x}_{11} - \bar{x}_2 = 0.346$ | $\bar{x}_{12} - \bar{x}_2 = 0.303$ |
| $\bar{x}_7 - \bar{x}_3 = 0.060$ | $\bar{x}_8 - \bar{x}_3 = -0.271$ | $\bar{x}_9 - \bar{x}_3 = -0.052$ | $\bar{x}_{10} - \bar{x}_3 = 0.266$ | $\bar{x}_{11} - \bar{x}_3 = 0.289$ | $\bar{x}_{12} - \bar{x}_3 = 0.246$ |
| $\bar{x}_7 - \bar{x}_4 = 0.365$ | $\bar{x}_8 - \bar{x}_4 = 0.033$ | $\bar{x}_9 - \bar{x}_4 = -0.356$ | $\bar{x}_{10} - \bar{x}_4 = 0.037$ | $\bar{x}_{11} - \bar{x}_4 = 0.594$ | $\bar{x}_{12} - \bar{x}_4 = 0.551$ |
| $\bar{x}_7 - \bar{x}_5 = 0.175$ | $\bar{x}_8 - \bar{x}_5 = -0.156$ | $\bar{x}_9 - \bar{x}_5 = -0.166$ | $\bar{x}_{10} - \bar{x}_5 = 0.152$ | $\bar{x}_{11} - \bar{x}_5 = 0.404$ | $\bar{x}_{12} - \bar{x}_5 = 0.361$ |
| $\bar{x}_7 - \bar{x}_6 = -0.128$ | $\bar{x}_8 - \bar{x}_6 = -0.204$ | $\bar{x}_9 - \bar{x}_6 = 0.119$ | $\bar{x}_{10} - \bar{x}_6 = 0.199$ | $\bar{x}_{11} - \bar{x}_6 = 0.357$ | $\bar{x}_{12} - \bar{x}_6 = 0.314$ |
| $\bar{x}_7 - \bar{x}_8 = -0.332$ | $\bar{x}_8 - \bar{x}_7 = 0.332$ | $\bar{x}_9 - \bar{x}_7 = 0.008$ | $\bar{x}_{10} - \bar{x}_7 = 0.327$ | $\bar{x}_{11} - \bar{x}_7 = 0.229$ | $\bar{x}_{12} - \bar{x}_7 = 0.186$ |
| $\bar{x}_7 - \bar{x}_9 = -0.008$ | $\bar{x}_8 - \bar{x}_9 = 0.323$ | $\bar{x}_9 - \bar{x}_8 = -0.323$ | $\bar{x}_{10} - \bar{x}_8 = 0.004$ | $\bar{x}_{11} - \bar{x}_8 = 0.561$ | $\bar{x}_{12} - \bar{x}_8 = 0.518$ |
| $\bar{x}_7 - \bar{x}_{10} = -0.327$ | $\bar{x}_8 - \bar{x}_{10} = 0.004$ | $\bar{x}_9 - \bar{x}_{10} = 0.318$ | $\bar{x}_{10} - \bar{x}_9 = 0.318$ | $\bar{x}_{11} - \bar{x}_9 = 0.237$ | $\bar{x}_{12} - \bar{x}_9 = 0.194$ |
| $\bar{x}_7 - \bar{x}_{11} = 0.229$ | $\bar{x}_8 - \bar{x}_{11} = 0.561$ | $\bar{x}_9 - \bar{x}_{11} = 0.237$ | $\bar{x}_{10} - \bar{x}_{11} = 0.556$ | $\bar{x}_{11} - \bar{x}_{10} = 0.556$ | $\bar{x}_{12} - \bar{x}_{10} = 0.513$ |
| $\bar{x}_7 - \bar{x}_{12} = 0.186$ | $\bar{x}_8 - \bar{x}_{12} = 0.518$ | $\bar{x}_9 - \bar{x}_{12} = 0.194$ | $\bar{x}_{10} - \bar{x}_{12} = 0.513$ | $\bar{x}_{11} - \bar{x}_{12} = 0.043$ | $\bar{x}_{12} - \bar{x}_{11} = 0.043$ |

مقدار LSD برای سال ۹۷ مقدار ۰.۲۵۷ بدست آمد. بنابراین با توجه به رابطه $|\bar{x}_i - \bar{x}_j| \geq \text{lsd}$ فرض صفر

برای شاخص‌های جدول ۴-۱۶ رد می‌شود:

جدول ۴-۱۶. شاخص‌هایی که فرض صفر برای آنها رد شده است.

| | | |
|--|--|---|
| اهرم مالی-گردش سهام | اهرم مالی-سایز | اهرم مالی-اختلاف قیمتی موثر |
| اهرم مالی-تعداد معاملات | اهرم مالی-حجم معاملات | تعداد روزهای بدون معامله براساس گردش-اختلاف قیمت پیشنهادی خرید و فروش |
| تعداد روزهای بدون معامله براساس گردش-تعداد معاملات | تعداد روزهای بدون معامله براساس گردش-حجم معاملات | اختلاف پیشنهادی قیمت خریدوفروش-عملکرد شرکت |
| اختلاف پیشنهادی قیمت خریدوفروش-گردش سهام | اختلاف قیمتی خرید و فروش-ضریب انعطاف‌پذیری تجارت | اختلاف قیمتی خرید و فروش-معیارژانگ |
| اختلاف قیمتی خرید و فروش-سایز | اختلاف قیمتی خرید و فروش-اختلاف قیمتی موثر | اختلاف قیمتی خرید و فروش-تعداد معاملات |
| اختلاف قیمتی خرید و فروش-حجم معاملات | ضریب انعطاف‌پذیری تجارت-حجم معاملات | ضریب انعطاف‌پذیری تجارت-تعداد معاملات |
| معیارژانگ-تعداد معاملات | | معیارژانگ-حجم معاملات |

مقدار LSD برای سال ۹۸ مقدار ۰.۲۶۶ بدست آمد. بنابراین با توجه به رابطه $|\bar{x}_i - \bar{x}_j| \geq lsd$ فرض صفر برای شاخص‌های جدول ۴-۱۷ رد می‌شود:

جدول ۴-۱۷. شاخص‌هایی که فرض صفر برای آنها رد شده است.

| | | |
|--|--|---|
| اهرم مالی-اختلاف قیمت پیشنهادی خرید و فروش | اهرم مالی-معیار ژانگ | اهرم مالی-اختلاف قیمتی موثر |
| تعداد روزهای بدون معامله براساس گردش-حجم معاملات | تعداد روزهای بدون معامله براساس گردش-تعداد معاملات | درصد روزهای معاملاتی-اختلاف قیمت پیشنهادی خرید و فروش |
| درصد روزهای معاملاتی-معیارژانگ | درصد روزهای معاملاتی-اختلاف قیمتی موثر | درصد روزهای معاملاتی-حجم معاملات |
| اختلاف قیمت پیشنهادی خرید و فروش-ضریب انعطاف‌پذیری تجارت | اختلاف قیمت پیشنهادی خرید و فروش-سایز | اختلاف قیمت پیشنهادی خرید و فروش-حجم معاملات |
| اختلاف قیمت پیشنهادی خرید و فروش-تعداد معاملات | عملکرد شرکت-حجم معاملات | عملکرد شرکت-تعداد معاملات |
| گردش سهام-حجم معاملات | گردش سهام-تعداد معاملات | ضریب انعطاف‌پذیری تجارت-معیارژانگ |
| ضریب انعطاف‌پذیری تجارت-اختلاف قیمتی موثر | معیار ژانگ-سایز | سایز-اختلاف قیمتی موثر |
| اختلاف قیمتی موثر-حجم معاملات | اختلاف قیمتی موثر-تعداد معاملات | معیارژانگ-تعداد معاملات |
| معیارژانگ-حجم معاملات | | |

با توجه به نتایج بدست آمده از جداول بالا، مشاهده می‌شود که شاخص‌های حجم معاملات، تعداد معاملات و اختلاف قیمت پیشنهادی خرید و فروش با اکثر شاخص‌ها دارای اختلاف میانگین معنی‌دار

هستند که این نتیجه ناشی از آن است که شاخص اختلاف قیمت پیشنهادی خرید و فروش دارای بیشترین مقدار میانگین و شاخص‌های تعداد معاملات و حجم معاملات دارای کمترین مقدار میانگین هستند.

۴-۵ خلاصه و جمع‌بندی فصل

در این پژوهش پس از معرفی الگوریتم خوشه‌بندی پیوند کامل، ۴۲ شرکت مورد مطالعه در بورس اوراق بهادار باتوجه به شاخص‌های نقدشوندگی منتخب و براساس ضریب تشابه تطابق نسبی خوشه‌بندی شدند. نتایج خوشه‌بندی شرکت‌ها در هر سال در این فصل بیان گردید. سپس با استفاده از شاخص سیلوئت، خوشه‌بندی مورد ارزیابی قرار گرفت. به منظور بررسی علل تفاوت خوشه‌ها، تحلیل واریانس و آزمون کمترین تفاوت معنی‌دار انجام شد. نتایج آزمون کمترین تفاوت معنی‌دار حاکی از آن بود که شاخص‌های اختلاف قیمت پیشنهادی خرید و فروش، حجم معاملات و تعداد معاملات با اکثر شاخص‌های منتخب دارای تفاوت معنی‌دار در میانگین می‌باشند.

فصل ۵: نتیجه گیری و پیشنهادات آتی

۵-۱ مقدمه

در فصل اول، کلیات پژوهش که شامل بیان موضوع، اهمیت و ضرورت انجام پژوهش، سوالی که محقق در جستجوی پاسخ به آن است، هدف پژوهش و قلمرو زمانی و مکانی پژوهش مورد بررسی قرار گرفته است. فصل دوم مفاهیم بنیادی و پیشینه پژوهش می باشد که در این فصل به بیان مفاهیمی همچون انواع بازار و سهام، نقدشوندگی، داده کاوی و روش های مختلف خوشه بندی پرداخته شده است و در انتهای فصل پیشینه داخلی و خارجی مطرح گردید. در فصل سوم به فرایند خوشه بندی، شاخص های نقدشوندگی مورد مطالعه، تحلیل واریانس و آزمون کمترین تفاوت معنادار اشاره شد. در نهایت در فصل چهارم جامعه و نمونه آماری مشخص شد و نتایج محاسبات و تجزیه و تحلیل آنها صورت گرفت.

مطالب این فصل به بررسی پژوهش براساس نتایج بدست آمده، محدودیت های پژوهش و پیشنهادهایی برای پژوهش های آینده اختصاص یافته است.

۵-۲ خلاصه پژوهش

در این پژوهش هدف خوشه بندی شرکت های پذیرفته شده در بورس اوراق بهادار تهران بر مبنای شاخص های نقدشوندگی و براساس ضریب تشابه تطابق نسبی است. نمونه آماری این پژوهش را ۴۲ شرکت مورد مطالعه در بورس اوراق بهادار تهران تشکیل می دهند. شاخص های نقدشوندگی منتخب شامل گردش سهام، اختلاف قیمت پیشنهادی خرید و فروش، معیار ژانگ، اهرم مالی، عملکرد شرکت، ضریب انعطاف پذیری تجارت، اختلاف قیمتی موثر، تعدیل تعداد روزهای بدون معامله براساس گردش، تعداد معاملات، حجم معاملات، درصد روزهای معاملاتی و اندازه شرکت هستند.

در ابتدا داده های مربوط به متغیرهای پژوهش از طریق شرکت مدیریت فناوری بورس تهران، گزارش های ادواری سازمان بورس اوراق بهادار تهران و سامانه اطلاع رسانی ناشران کدال جمع آوری شد،

سپس هر یک از شاخص‌های نقدشوندگی برای هر نمونه آماری محاسبه گردید. نتایج این محاسبات به منظور استفاده در فرمول ضریب تشابه تطابق نسبی به صفر و یک کدگذاری شد آن‌گاه ضریب تشابه برای هر جفت نمونه آماری محاسبه شد. سپس الگوریتم خوشه‌بندی پیوند کامل براساس ضریب تشابه انجام گرفت که در ابتدای این فرایند، هر عنصر در خوشه‌ای از خود قرار دارد. در هر مرحله دو خوشه‌ای که دارای بیشترین ضریب تشابه هستند باهم ترکیب می‌شوند و تا زمانی این روند ادامه دارد که تمام عناصر در خوشه قرار بگیرند. در این پژوهش برای هر سال ۵ خوشه بدست آمد. سپس با استفاده از شاخص سیلوئت، خوشه‌بندی ارزیابی شد. به منظور بررسی فرضیه تحقیق تحلیل واریانس و آزمون کمترین تفاوت معنی‌دار انجام گرفت.

۳-۵ نتایج پژوهش

در بازارهای مالی و بورس حجم وسیعی از داده‌ها به صورت مداوم و روزانه تولید می‌شود که هم برای فعالان این حوزه و هم محققین دانشگاهی و آکادمیک جذابیت خاصی دارد. فعالان حوزه‌های مالی روزانه با حجم وسیعی از داده‌ها سر و کار دارند و باید براساس این داده‌ها تصمیمات مهمی در زمان مناسب بگیرند. تعداد پارامترها و متغیرهای تأثیرگذار کار تحلیل اطلاعات و تصمیم‌گیری را با استفاده از روش‌های سنتی بسیار سخت می‌کند. درست همین جاست که علم داده کاوی در بازارهای مالی موثر عمل کرده و با در اختیار قرار دادن مجموعه‌های از تکنیک‌های متنوع و کاربردی امکان شناسایی الگوها، روندها و تصمیم‌گیری با اثربخشی بالاتر را فراهم می‌آورد.

در صنعت مالی با استفاده از مفاهیم ریاضیات و تحلیل‌های آماری، استفاده از الگوریتم‌ها و تکنیک‌های داده کاوی می‌توان برای درک رفتارهای بازارهای مالی اقدام کرد. براساس تجزیه و تحلیل فصل چهارم و نتایج بدست آمده از خوشه‌بندی شرکت‌های بورس با توجه به در نظر گرفتن شاخص‌های نقدشوندگی تعیین شده برای ۴۲ شرکت پذیرفته شده در بورس اوراق بهادار تهران طی دو سال ۹۷-۹۸ مشخص شد

که بیشترین تفاوت خوشه‌ها ناشی از شاخص‌های تعداد معاملات و حجم معاملات است که می‌توان از این دو شاخص برای پیش‌بینی نقدشوندگی استفاده کرد.

۴-۵ محدودیت‌های پژوهش

برخی از محدودیت‌های این پژوهش می‌توان به موارد زیر اشاره کرد:

(۱) پژوهش‌هایی که در حوزه حسابداری و مالی صورت می‌گیرد، متکی به داده‌های مالی موجود در بازار سرمایه است. کیفیت این داده‌ها نیز بر کیفیت نتیجه پژوهش اثرگذار است.

(۲) در این پژوهش به دلیل نیاز به کمی کردن شاخص‌های نقدشوندگی، زمان زیادی صرف محاسبه شاخص‌ها شد.

(۳) به دلیل آنکه محاسبه ضریب تشابه با استفاده از ماتریس باینری وقوع انجام می‌شود، ناگزیر شاخص‌ها به دو رقم یک و صفر کدگذاری کرد. در عالم واقع این طبقه‌بندی به صورت چندگانه است.

(۴) با توجه به نوآوری مدل، نرم‌افزاری برای تجزیه و تحلیل مدل بکار گرفته شده وجود نداشته است.

۵-۵ پیشنهادهایی برای پژوهش‌های آتی

برخی از موارد و موضوعاتی که پیشنهاد می‌شود در پژوهش‌های آتی مدنظر قرار گیرند، به شرح زیر است:

(۱) در این پژوهش با توجه به تفاوت ماهیت، اهداف و ساختار سازمانی شرکت‌های سرمایه‌گذاری و سایر شرکت‌ها و همچنین دشواری جمع‌آوری اطلاعات در پی مصوب هیئت مدیره سازمان بورس اوراق بهادار تهران پیشنهاد می‌گردد در پژوهش‌های آتی، نتایج با حضور و عدم حضور شرکت‌های

مذکور از جامعه آماری بررسی و مقایسه گردد.

۲) در این پژوهش از ۱۲ شاخص نقدشوندگی تعریف شده در فصل سوم استفاده شده است پیشنهاد

می شود در پژوهش های آتی، شاخص های دیگر نقدشوندگی نیز بررسی و مقایسه گردد.

۳) این پژوهش براساس خوشه بندی پیوند کامل انجام شده است پیشنهاد می شود در پژوهش های آتی

از روش های دیگر خوشه بندی استفاده و نتایج مقایسه گردد.

۴) پیشنهاد می گردد با افزایش دامنه زمانی پژوهش، تعمیم پذیری روش ارائه شده در مورد

خوشه بندی و پیش بینی نقدشوندگی براساس تعداد معاملات و حجم معاملات در سطح وسیع

تر و به صورت استنباطی مورد بررسی قرار گیرد.

۵) این پژوهش با توجه به شاخص های نقدشوندگی انجام شده است. می توان در پژوهش های آتی

شاخص های دیگری چون بازده را اضافه و در صورت استفاده از شاخص های کیفی و کلامی از

رویکرد فازی خوشه بندی بهره برد.

۵-۶ خلاصه و جمع بندی فصل

این پژوهش با هدف خوشه بندی شرکت های فعال در بورس اوراق بهادار با توجه به شاخص های

نقدشوندگی و با استفاده از ضریب تشابه تطابق نسبی انجام گرفت. شرکت های مشابه شناسایی شده و در

خوشه ها قرار گرفتند. نتیجه این خوشه بندی و تجزیه و تحلیل فرضیه پژوهش در فصل چهارم بیان

شد. در ادامه در این فصل به بررسی سوال پژوهش پرداخته و نیز محدودیت های تحقیق و پیشنهادات آتی

مورد بررسی قرار گرفت.

مراجع

- [۱] ع. سعیدی و ا.ا. دادار (۱۳۸۶) "بررسی رابطه بین نقدشوندگی سهام با شاخص نقدشوندگی دوره‌های پیشین در بورس اوراق بهادار تهران"، فصلنامه علمی-پژوهشی مطالعات مدیریت، شماره ۱۶، ص ۹۲-۷۵.
- [۲] م. نیکبخت و ح. کلهرنیا (۱۳۹۶) "تعیین رابطه بین شفافیت اطلاعاتی، نقد شونده‌گی و ارزشیابی شرکت"، پژوهش‌های حسابداری مالی و حسابرسی، شماره ۳۳، ص ۱۶-۱.
- [۳] ش. زمانی و پ. فغانی کندی (۱۳۹۵) "محاسبه میزان نقدشوندگی بورس تهران و بررسی عوامل اثرگذار بر آن"، چشم‌انداز مدیریت مالی، شماره ۱۵، ص ۸۰-۶۱.
- [۴] ا. دهقان و ح. صادقی (۱۳۹۵) "کاربرد خوشه‌بندی در بازارهای مالی"، چهارمین کنفرانس ملی مدیریت، اقتصاد و حسابداری، دانشگاه تبریز.
- [۵] س. منصوری، م. تقوی فرد، م. فرجامی و م. بینشیان (۱۳۹۵) "تحلیل شاخصهای نقدشوندگی سهام در بازار بورس با تکنیک DEMATEL-ANP" فصلنامه مهندسی تصمیم، شماره ۳۳، ص ۱-۱۶.
- [۶] ا. شهرآبادی و ن. بشیری (۱۳۹۴) "مدیریت سرمایه‌گذاری در بورس اوراق بهادار"، چاپ پنجم، انتشارات بورس، ۳۸۳ ص.
- [۷] ج. مدورا (۲۰۰۶) "نهادهای پولی و مالی"، ا. شهرآبادی و ن. رحمانی، چاپ دوم، انتشارات بورس.
- [۸] جعفری ورعی (۱۳۹۳) "بررسی نقش بازار سرمایه در امنیت خانواده"، پژوهش‌های راهبردی امنیت و نظم اجتماعی، شماره ۸، ص ۱۱۵-۱۳۱.
- [۹] م. نادعلی (۱۳۹۲) "محاسبه شاخص تنش در بازار پول اقتصاد ایران"، پژوهش‌ها و سیاست‌های اقتصادی، شماره ۶۶، ص ۱۴۲-۱۱۵.
- [۱۰] آ. خراسانی (۱۳۹۴) "انواع بازارهای مالی"، کنفرانس بین‌المللی پژوهش‌های نوین در مدیریت، اقتصاد و حسابداری، استانبول.
- [۱۱] ع. ت. روستا (۱۳۸۷) "بررسی نقش بازار اوراق بهادار خارج از بورس بر توسعه بازار سرمایه در ایران"، فصلنامه دانش و توسعه، شماره ۱۵، ص ۹۶-۷۷.
- [۲۷] م. اسماعیلی (۱۳۹۱) "مفاهیم و تکنیک‌های داده کاوی"، ص ۳۱۵.
- [۳۳] ف. داریوش و م. پورحمیدی (۱۳۹۱) "بخش بندی سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از تحلیل خوشه‌ای فازی"، فصلنامه پژوهش‌های حسابداری مالی، شماره ۴، ص ۱۲۸-۱۰۵.
- [۳۴] ا. میرزایی و م. یغمایی (۱۳۹۰) "سیستم خبره فازی برای ارزیابی کیفیت سهام و خوشه‌بندی فازی برای رتبه‌بندی سهام"، دومین همایش فناوری اطلاعات، حال، آینده، مشهد.
- [۳۵] م. امیری، م. حسکویی، م. کامی (۱۳۹۲) "رتبه بندی شرکت‌های تولیدی در بورس اوراق بهادار تهران با استفاده از مدل‌های تصمیم‌گیری با معیارهای چندگانه و شبکه عصبی مصنوعی"، فصلنامه علمی پژوهشی دانش سرمایه‌گذاری، شماره ۷، ص ۸۴-۷۳.
- [۳۶] م. طاهریو ا. خاوری نژاد (۱۳۹۲) "ارزیابی و رتبه بندی صنایع تولیدی عضو بورس اوراق بهادار تهران"، همایش سالانه سیستم‌های پولی و ارزی، ص ۵۴۵-۵۱۵، تهران.
- [۳۷] ا. پورزندی و م. کیخا (۱۳۹۳) "بهینه‌سازی سبد سهام با استفاده از روش k-means و الگوریتم ژنتیک"، مهندسی مالی و مدیریت اوراق بهادار، شماره ۱۹، ص ۱۵۱-۱۳۱.

- [۳۸] ز. فرسنادیا و م. شعبانپور (۱۳۹۴) "خوشه‌بندی سهام شرکتهای پذیرفته‌شده در بورس اوراق بهادار تهران با الگوریتم دو مرحله‌ای نگاشت ویژگی خود سامان"، پژوهش‌های نوین در مدیریت، اقتصاد و حسابداری، دوره ۲، ص ۲۰-۱.
- [۳۹] ا. سلگی، ح. باغبانان و پ. مهران (۱۳۹۴) "خوشه‌بندی سهام‌های رشدی و ارزشی در بازار بورس تهران با استفاده از الگوریتم FCM، انتشارات دانشگاه خوارزمی، ص ۱۴-۱.
- [۴۰] م. اقبالنیا، ا. پویانفرو م. مالکی (۱۳۹۴) "مدلسازی هم‌حرکتی سهام در بورس اوراق بهادار تهران با استفاده از رویکرد خوشه‌بندی سه‌مرحله‌ای" چشم‌انداز مدیریت مالی، شماره ۱۱، ص ۱۳۳-۱۰۸.
- [۴۱] ز. جلیلود، ف. سلیمانی و ج. بهاری (۱۳۹۶) "بهره‌گیری از تکنیک خوشه‌بندی در ارتقای کارایی مدیریت ریسک در شرکتهای بیمه"، فصلنامه مطالعات مدیریت و حسابداری، ص ۲۷۰-۲۵۸.
- [۴۲] ع. پاک‌مرام، ج. بحری‌ثالثو م. ولی‌زاده (۱۳۹۶) "انتخاب و بهینه‌سازی سبد سهام با استفاده از الگوریتم ژنتیک، با بهره‌گیری از مدل میانگین-نیمه‌واریانس مارکویتز"، مجله مهندسی مالی و مدیریت اوراق بهادار، شماره ۱۶.
- [۴۳] ر. مهربانپور و م. حبیبزاده (۱۳۹۷) "خوشه‌بندی و پیش‌بینی سودآوری شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران با رویکرد درخت تصمیم C"، فصلنامه علمی - پژوهشی مطالعات تجربی حسابداری مالی، شماره ۵۹، ص ۱۵۷-۱۳۵.
- [۴۴] م. مشرفی و ب. جواد (۱۳۹۸) "بهینه‌سازی چندهدفه مسئله سبد سهام با استفاده از تحلیل سلسله‌مراتبی و الگوریتم ژنتیک"، فصلنامه مدیریت مهندسی و رایانش نرم، شماره ۱، ص ۱۰۱-۷۳.
- [۴۵] ح. انصاری، ع. بهزادی و ت. فرید، "ردیابی شاخص بهبود یافته دومرحله‌ای با استفاده از الگوریتم‌های فراابتکاری" راهبرد مدیریت مالی، شماره ۲۷، ص ۲۲-۱.
- [۴۶] م. صیادی و م. امید (۱۳۹۸) "بهینه‌سازی پرتفوی مبتنی بر پیش‌بینی برای سهام گروه‌های وابسته به نفت در ایران با استفاده از روش‌های داده‌کاوی"، انتشارات دانشگاه خوارزمی، شماره ۲، دوره ۸، ص ۲۵۲-۲۲۵.
- [۴۷] ع. آذر، ا. یزدانیان و م. قندهاری (۱۳۹۹) "بهینه‌سازی سبد سهام با استفاده از الگوریتم ژنتیک و روش k-means بهبود یافته بر پایه الگوریتم ژنتیک"، چهارمین همایش ریاضیات و علوم انسانی، دوره ۴، دانشگاه علامه طباطبائی.
- [۴۸] پ. آ. مقدم، ا. تبریزو، ا. نجفی (۱۳۹۹) "طراحی یک مدل هوشمند جهت تعیین سیگنال‌های معاملات سهام با رویکرد داده‌کاوی"، پژوهش‌های نوین در ریاضی، شماره ۲۴، دوره ۶، ص ۱۷۲-۱۵۹.
- [۴۸] غ. خاکی (۱۳۹۹) "روش تحقیق با رویکرد پایان‌نامه نویسی"، چاپ نهم، فوژان، ص ۳۶۰.
- [۷۱] ر. احمدپور (۱۳۸۵) "رابطه بین معیارهای ریسک و اختلاف قیمت پیشنهادی خرید و فروش سهام در بورس اوراق بهادار تهران"، بررسی‌های حسابداری و حسابرسی، شماره ۴۶، ص ۶۰-۳۷.
- [۷۲] ن. ایزدینیا، ا. رسائیان (۱۳۸۹) "پراکنندگی مالکیت و نقدشوندگی سهام"، بررسی‌های حسابداری و حسابرسی، شماره ۷۰، ص ۲۲-۳.
- [۸۳] ن. طبری، م. میرمحمدیو م. عنایتی (۱۳۹۴) "بررسی اثر اندازه و سرمایه شرکت بر میزان ریسک شرکتهای "اولین همایش بین‌المللی نوآوری و تحقیق در هنر و علوم انسانی، ص ۴۰-۲۵.
- [۸۷] فروغی، فرهمند، ابراهیم (۱۳۹۰) "رابطه بین نقدشوندگی و عملکرد شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران"، فصلنامه بورس اوراق بهادار، شماره ۱۵، ص ۱۴۳-۱۲۵.
- [۹۰] ا. رستمی و م. لشکری (۱۳۹۳) "بررسی شاخص‌های اعتبارسنجی خوشه‌بندی"، دستاوردهای نوین در علوم مهندسی و پایه، ص ۷-۱.
- [۹۲] ع. آذر و م. مومنی (۱۳۹۹) "آمار و کاربرد آن در مدیریت"، جلد دوم، چاپ بیست و سوم، سازمان سمت.

- [12] R. August, C. Michael, K. Ramaswamy, and G. William(1986) “Asset pricing and the bid-ask spread” vol. 17, pp223-249
- [13] P. M. Panayides, N. Lambertides, and K. Cullinane(2013)“Liquidity risk premium and asset pricing in US water transportation,” *Transp. Res. Part E Logist. Transp. Rev.*, vol. 52, pp. 3–15.
- [14] H. J. . S. R(1988) “Liquidity and Execution Costs in Equity Markets” *J. Portf. Manag.*, vol. 14, pp. 10–16.
- [15] W. Liu, “A liquidity-augmented capital asset pricing model(2006)” *J. financ. econ.*, vol. 82, no. 3, pp. 631–671.
- [16] F. Black(1971) “toward a fully automated stock exchange,” *Financ. Anal. J.*, vol. 27, pp. 28–44.
- [17] N. Gârleanu(2009) “Portfolio choice and pricing in illiquid markets,” *J. Econ. Theory*, vol. 144, no. 2, pp. 532–564.
- [18] A. Rubia and L. Sanchis-Marco(2013) “On downside risk predictability through liquidity and trading activity: A dynamic quantile approach,” *Int. J. Forecast.*, vol. 29, no. 1, pp. 202–219.
- [19] D. Chung and K. Hrazdil(2010) “Liquidity and market efficiency: A large sample study,” *J. Bank. Financ.*, vol. 34, no. 10, pp. 2346–2357, 2010.
- [20] R. Hodrea(2015) “An Intraday Analysis of the Market Efficiency-liquidity Relationship: The Case of BVB Stock Exchange,” *Procedia Econ. Financ.*, vol. 32, no. 15, pp. 1432–1441.
- [21] Kyle A(1985) “Continuous Auctions and Insider Trading,” *Econometrica*, vol. 53, no. 6. pp. 1315–1336.
- [22] T. L. Abdourahmane Sarr(2002) "Measuring Liquidity in Financial Markets" ,*International Monetary Fund*,pp 60-63.
- [23] A. S. Ashour, N. Dey, and D. N. Le(2017) “Biological data mining: Techniques and applications,” *Min. Multimed. Doc.*, vol. 1, no. 4, pp. 161–172,
- [24] T. Silwattanusarn(2012) “Data Mining and Its Applications for Knowledge Management : A Literature Review from 2007 to 2012,” *Int. J. Data Min. Knowl. Manag. Process*, vol. 2, no. 5, pp. 13–24.
- [25] D. Huchon, N. Crozet, N. Cantenot, and R. Ozon(1981) “Germinal vesicle breakdown in the *Xenopus laevis* oocyte: Description of a transient microtubular structure,” *Reprod. Nutr. Dev.*, vol. 21, no. 1, pp. 135–148.
- [26] T. Hendrickx, B. Cule, P. Meysman, S. Naulaerts, K. Laukens, and B. Goethals(2015) “Mining association rules in graphs based on frequent cohesive itemsets,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9078, no. 3, pp. 637–648.
- [28] C. Giraud-Carrier and O. Povel(2003) “Characterising data mining software,” *Intell. Data Anal.*, vol. 7, no. 3, pp. 181–192.
- [29] S. R. . SNEATH P.H.A(1973)" Numerical Taxonomy.The Principles And Praticce Of Numerical Classification",*Tylor&Francis,Ltd* ,vol 24,no 2,pp 263-268.
- [30] J. O. Corliss, P. H. A. Sneath, and R. R. Sokal(1974) “Numerical Taxonomy: The Principles and Practice of Numerical Classification,” *Transactions of the American Microscopical Society*, vol. 93, no. 2. p. 303.
- [31] King.B(1976) “Step-wise Clustering Procedures,” *J. Am. Stat. Assoc.*, vol. 62, no. 86–101.
- [32] M. P. Vayssières, R. E. Plant, H. Hall, and O. S. Ave(1998) “Identification of Vegetation State- and-transition Domains in California ’ s Hardwood Table of

- content,” Rep. Fire Resour. Assess. Progr., no. May, pp. 1–101.
- H. W. Shin and S. Y. Sohn(2004)“Segmentation of stock trading customers] [49] according to potential value,” Expert Syst. Appl., vol. 27, no. 1, pp. 27–33.
- [50] N. Basalto, R. Bellotti, F. De Carlo, P. Facchi, and S. Pascazio(2005) “Clustering stock market companies via chaotic map synchronization,” Phys. A Stat. Mech. its Appl., vol. 345, no. 1–2, pp. 196–206, 2005.
- [51] K. A. J. Doherty, R. G. Adams, and N. Davey(2005) “Hierarchical topological clustering learns stock market sectors,” Proc. 2005 UK Work. Comput. Intell. UKCI 2005, pp. 190–197, 2005.
- [52] S. H. Liao, H. hui Ho, and H. wen Lin(2008) “Mining stock category association and cluster on Taiwan stock market,” Expert Syst. Appl., vol. 35, no. 1–2, pp. 19–29.
- [53] S. R. Nanda, B. Mahanty, and M. K. Tiwari(2010) “Clustering indian stock market data for portfolio management,” Expert Syst. Appl., vol. 37, no. 12, pp. 8793–8798.
- [54] D. C. Li, W. L. Dai, and W. T. Tseng(2011) “A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business,” Expert Syst. Appl., vol. 38, no. 6, pp. 7186–7191
- [55] D. Enke, M. Grauer, and N. Mehdiyev(2011) “Stock market prediction with Multiple Regression, Fuzzy type-2 clustering and neural networks,” Procedia Comput. Sci., vol. 6, pp. 201–206.
- [56] J. Straßburg, C. González-Martel, and V. Alexandrov(2012) “Parallel genetic algorithms for stock market trading rules,” Procedia Comput. Sci., vol. 9, pp. 1306–1313.
- [57] V.-G. Marica(2014) “Hyper-ellipsoid Clustering of Time Series. A Case Study for Daily Stock Returns,” Procedia Econ. Financ., vol. 15, no. 14, pp. 777–783.
- [58] Y. Wang, X. Ma, Y. Lao, and Y. Wang(2014) “A fuzzy-based customer clustering approach with hierarchical structure for logistics network optimization,” Expert Syst. Appl., vol. 41, no. 2, pp. 521–534.
- [59] I. Bose and X. Chen(2015)“Detecting the migration of mobile service customers using fuzzy clustering,” Inf. Manag., vol. 52, no. 2, pp. 227–238
- [60] S. Aghabozorgi and Y. W. Teh(2014) “Stock market co-movement assessment using a three-phase clustering method,” Expert Syst. Appl., vol. 41, no. 4 PART 1, pp. 1301–1314.
- [61] B. S. Bini and T. Mathew(2016) “Clustering and Regression Techniques for Stock Prediction,” Procedia Technol., vol. 24, pp. 1248–1255.
- [62] X. Zhong and D. Enke(2017) “A comprehensive cluster and classification mining procedure for daily stock market return forecasting,” Neurocomputing, vol. 267, pp. 152–168.
- [63] H. Li(2019) “Multivariate time series clustering based on common principal component analysis,” Neurocomputing, vol. 349, pp. 239–247.
- [64] S. Majumdar and A. K. Laha(2020) “Clustering and classification of time series using topological data analysis with applications to finance,” Expert Syst. Appl., vol. 162, p. 113868.
- [65] K. Deepak, K. S. Pradeepta, and V. Rajit(2021)“A systematic review of stock market prediction using machine learning and statistical techniques,” Mater. Today.
- [66] A. Moghar and M. Hamiche(2020) “Stock Market Prediction Using LSTM Recurrent Neural Network,” Procedia Comput. Sci., vol. 170, pp. 1168–1173.

- [67] D. Zhang and S. Lou(2021) “The application research of neural network and BP algorithm in stock price pattern classification and prediction,” *Futur. Gener. Comput. Syst.*, vol. 115, pp. 872–879, 2021.
- [69] M. Halkidi, Y. Batistakis, and M. Vazirgiannis(2001) “On clustering validation techniques,” *J. Intell. Inf. Syst.*, vol. 17, no. 2–3, pp. 107–145, 2001.
- [70] D. Chai, R. Faff, and P. Gharghori(2010) “New evidence on the relation between stock liquidity and measures of trading activity,” *Int. Rev. Financ. Anal.*, vol. 19, no. 3, pp. 181–192, 2010.
- [73] K. Kim, S & Lee(2014) “Pricing Of Liquidity In Emerging Markets”,*EMPIRICAL FINANCE*,VOL 25,PP112-133.
- [74] W. Kang and H. Zhang(2014) “Measuring Liquidity in Emerging Markets,” *Pacific-Basin Financ. J.*, vol. 27, pp. 49–71, 2014
- [75] Fong,K,Holden,C.W,&Trzcinka,C.A(201) "What Are The Best Liquidity Proxies For Global Research", *European Finance Association* ,vol 21,pp1355-1401.
- [76] Y. Amihud(2002)“Illiquidity and stock returns: Cross-section and time-series effects,” *J. Financ. Mark.*, vol. 5, no. 1, pp. 31–56.
- [77] R. Y. Goyenko, C. W. Holden, and C. A. Trzcinka, “Do liquidity measures measure liquidity?,” *J. financ. econ.*, vol. 92, no. 2, pp. 153–181, 2009, doi: 10.1016/j.jfineco.2008.06.002.
- [78] E. J. Sullivan, “A New Estimate of Transaction Costs,” *CFA Dig.*, vol. 30, no. 2, pp. 83–84, 2000, doi: 10.2469/dig.v30.n2.682.
- [79] B. Geert, “Liquidity and Expected Returns(2006)” *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699.
- [80] M. J. Fleming, “Measuring Treasury Market Liquidity(2005)” *SSRN Electron. J.*, no. June 2000.
- [81] A. Hameed, W. Kang, and S. Viswanathan(2010) “Stock market declines and liquidity,” *J. Finance*, vol. 65, no. 1, pp. 257–293.
- [82] J. R. Kale and Y. C. Loon(2011) “Product market power and stock market liquidity,” *J. Financ. Mark.*, vol. 14, no. 2, pp. 376–410.
- [84] J. Peress(2010)“Product market competition, insider trading, and stock market efficiency,” *J. Finance*, vol. 65, no. 1, pp. 1–43.
- [85] I.-L. Wang, Y.-C. Wang, and C.-W. Chen(2013) “Scheduling unrelated parallel machines in semiconductor manufacturing by problem reduction and local search heuristics,” *Flex. Serv. Manuf. J.*, vol. 25, no. 3, pp. 343–366.
- [87] R. W. Wanzala(2018) “Estimation of market immediacy by Coefficient of Elasticity of Trading three approach,” *J. Financ. Data Sci.*, vol. 4, no. 3, pp. 139–156.
- [88] K. M. S. Islam and B. R. Sarker(2014) “A similarity coefficient measure and machine-parts grouping in cellular manufacturing systems,” *Int. J. Prod. Res.*, vol. 38, no. 3, pp. 37–41.
- [89] H. Jiawei, K. Micheline, and Jian Pei(2016) "Data mining: Data mining concepts and techniques",Third edition, Morgan Kaufmann Publishers.

Abstract

Stock market forecasting using data mining methods is one of the most important topics in investment and financial market research. Many efforts have been made to predict this market using traditional methods. However, due to the increasing volume of information, these methods are not suitable for analyzing this amount of information. Data mining is able to discover hidden patterns and predict future trends in the stock market. The stock market can be considered a data mining problem and clustering is one of the data mining methods that is a good strategy to advance foresight and guidance of investors decisions. Stock liquidity means the ability to buy and sell stocks in the shortest time and at the lowest cost. Therefore, investors choose companies that have high liquidity. In this study, clustering of companies active in the stock market according to liquidity indices based on the relative matching coefficient has been done. For this purpose, 42 companies listed on the stock exchange during the years 1397-1398 based on the index selected liquidity options include turnover, proportional bid-ask spread, modified version of the amihud illiquidity measure, financial leverage, company performance, coefficient of elasticity of trading, effective spread, turnover-adjusted number of zero daily volumes, number of trades, transaction volume percentage of transaction days, the size of the company was clustered based on the Complete-Linkage Clustering algorithm and the similarity coefficient of relative matching. The results of this study show that the indicators of number of trades and volume of trades can be used to predict liquidity.

Keywords (5 to 7 keywords): Clustering, Similarity Coefficient, Liquidity, Tehran Stock Exchange



Shahrood University of Technology
Faculty of Industrial Engineering and Management
M.Sc. Thesis in Industrial Management

Stock clustering according to liquidity indices based on relative matching similarity coefficient

By: neginal sadat ahmadi

Supervisor:
Dr. reza sheikh

Advisor:
Dr. Majid Ameri

January 2021