Archive of SID

فصلنامه علمی پژوهشی
دانش مالی تحلیل اوراق بهادار
سال دهم، شماره سی‌وچهارم
تابستان ۱۳۹۶

# تدوین دندروگرام‌های سبد سهام بر اساس معیار فاصله اقلیدسی (مقایسه‌ای بین روش‌های گوناگون خوشه‌بندی سلسله مراتبی)

**حجت اله صادقی**[۱]
**شریفه فروغی دهنوی**[۲]

تاریخ دریافت: ۹۵/۱/۱۹ تاریخ پذیرش: ۹۵/۴/۱۵

**چکیده**

امروزه تجزیه‌وتحلیل بازارهای مالی به‌عنوان بخشی از بازار سرمایه و تأثیر آن در توسعه و طراحی پرتفوی و استراتژی سرمایه‌گذاری هر کشور، به موضوعی مهم و بسیار حیاتی تبدیل گشته است. هدف از این پژوهش، بررسی چگونگی ارتباط و توزیع سهم‌های مربوط به شاخص ۳۰ شرکت بزرگ بورس اوراق بهادار تهران و اثرات ارتباطی بین خوشه‌های حاصل از سهم‌های مرتبط با هر صنعت است. در این پژوهش با استفاده از انواع روش‌های خوشه‌بندی سلسله مراتبی، ساختار، طبقه‌بندی و سلسله‌مراتب این سهم‌ها در سال ۱۳۹۳ موردبررسی قرارگرفته است. نتایج نشان داد که با تمرکز بر هر یک از روش‌های خوشه‌بندی سلسله مراتبی و پیاده‌سازی آن‌ها بر روی سهم‌های موردنظر، خوشه‌های مختلفی از سهام با توجه به شباهت و روابط اقتصادی و حوزه صنعت مربوط به هر سهم شناسایی گردید و همچنین خوشه‌های کلیدی و سهم‌های حیاتی در مجموعه موردنظر به دست آمد. درمجموع نتایج حاکی از آن است که انتخاب بهترین روش خوشه‌بندی سلسله مراتبی برای خوشه‌بندی سهم‌ها، به هدف موردنظر از تجزیه‌وتحلیل خوشه‌ای و در نظر گرفتن مزایا و معایب هر روش بستگی دارد.

**واژه‌های کلیدی:** تجزیه و تحلیل خوشه‌بندی سلسله مراتبی، دندروگرام، سبد سهام، معیار فاصله اقلیدسی.

۱- استادیار دانشکده اقتصاد، مدیریت و حسابداری، دانشگاه یزد، ایران Sadeqi@yazd.ac.ir
۲- کارشناس ارشد مدیریت بازرگانی (مالی)، دانشکده اقتصاد، مدیریت و حسابداری، دانشگاه یزد، ایران (نویسنده مسئول)
forooghi_sh@yahoo.com



www.SID.ir

Archive of SID



## ۱- مقدمه

در کشورهای توسعه‌یافته کنونی، اکثر سرمایه‌گذاری‌ها از طریق بازارهای مالی انجام می‌گیرد و حیات بازار سرمایه وابسته به مشارکت فعال افراد جامعه در بورس است. از عمده‌ترین مشکلات کشورهای جهان سوم و مخصوصاً کشور ما، نبود مسیر و زیرساخت مناسب برای سرمایه‌گذاری‌های افراد و سازمان‌ها است. ازجمله دلایل عدم گرایش افراد به فعالیت در بازارهای مالی را می‌توان در عوامل فرهنگی، اقتصادی، قوانین و مقررات دولتی و مهم‌تر از همه عدم اطمینان حاکم بر بازارهای مالی همچون بورس اوراق بهادار عنوان کرد. بازار سهام ازجمله ارکان اقتصادی هر کشوری به شمار می‌رود. رسالت این بازار جذب سرمایه‌های کم عموم و سوق دادن آن‌ها به‌سوی استفاده در سرمایه‌گذاری هایی است که بتواند تضمین‌کننده منافع عموم مردم باشد. این امر سبب می‌شود تا سرمایه‌های عموم صرف فعالیت‌های کاذب اقتصادی نشود که از بین رفتن این سرمایه‌ها را به دنبال دارند و در اصل زمینه‌ای ایجاد می‌کند تا از این سرمایه‌ها به‌طور مناسب بهره‌برداری شود. لیکن سرمایه‌گذاران صرفاً با مقاصد انتفاعی در بازار سهام مشارکت می‌کنند. حضور مؤثر این افراد در بورس اوراق بهادار در پی کسب سود از سرمایه‌گذاری‌های خود است. ازاین‌رو کسب سود از جانب تک‌تک سرمایه‌گذاران می‌تواند تأثیر مثبتی بر جلب سرمایه‌های عموم داشته باشد.

با توجه به شرایط عدم اطمینان حاکم بر بورس اوراق بهادار، یکی از عمده‌ترین مسائل پیش روی سرمایه‌گذاران در بازارهای سرمایه، تصمیم‌گیری در جهت انتخاب اوراق بهادار مناسب برای سرمایه‌گذاری و تشکیل سبد بهینه سهام است. سرمایه‌گذاران در بازار سرمایه به دنبال افزایش بازده از یک‌سو و کاهش ریسک از سوی دیگر هستند. بدین منظور سرمایه‌گذاران سبدی از سهام تشکیل می‌دهند تا از طریق آن بتوانند با متنوع سازی، ریسک خود را حداقل سازند. سبد سهام مجموعه‌ای از سهام است که هر سهم موجود در آن، بازدهی و ریسک مشخصی دارد. برای تشکیل یک سبد سهام مناسب همبستگی بین قیمت‌های سهام نقش مهمی را ایفا می‌کند. آنچه در ایجاد یک سبد سهام دارای کمترین مقدار ریسک اهمیت بسیاری دارد، یافتن سهم‌هایی است که دارای کمترین میزان ارتباط با یکدیگر باشند. هرروزه تلاش‌های گسترده‌ای برای بهبود روش‌های انتخاب سبد سهام در بازارهای مالی دنیا صورت می‌گیرد. تلاش در جهت بهبود روش‌های تجزیه‌وتحلیل سهام به‌ویژه در بازارهایی که تنوع سهام در آن‌ها بسیار بالا است، منجر به پدید آمدن روش‌های نوینی گردیده است که در کنار روش‌های گذشته درصدد یافتن پاسخی برای حداکثر سازی سود در بازارهای مالی هستند. روش‌های تجزیه و تحلیل خوشه‌بندی سلسله مراتبی[1] از مصادیق این روش‌های نوین است.

خوشه‌بندی فرآیندی از گروه‌بندی داده‌ها به طبقات و یا خوشه‌ها است، به‌طوری‌که اشیاء در درون یک خوشه در مقایسه با یکدیگر شباهت بالایی دارند اما نسبت به اشیاء در خوشه‌های دیگر بسیار بی-شباهت هستند (هان و کمبر، ۲۰۰۱). از میان روش‌های خوشه‌بندی، روش خوشه‌بندی سلسله مراتبی دارای ویژگی‌هایی است که می‌تواند در خوشه‌بندی سهم‌های مختلف به کار گرفته شود. روش‌های خوشه‌بندی سلسله مراتبی سری‌های تودرتویی از بخش‌ها را تولید می‌کنند (جین، مارتی و فلاین، ۱۹۹۹). همبستگی بین قیمت‌های سهام و شناسایی گروه‌هایی با پویایی‌های مشترک، در متنوع سازی پرتفوی باهدف کاهش ریسک نقش بسیار مهمی دارد. روش خوشه‌بندی سلسله مراتبی می‌تواند ابزار مناسبی برای کمک به سرمایه‌گذاران در تشکیل سبد بهینه سهام باشد. ساختارهای سلسله مراتبی حاصل از این روش، ابزار مفیدی برای درک و تشخیص ساختار، طبقه‌بندی و سلسله‌مراتب در بازارهای مالی است. با استفاده از ساختارهای



www.SID.ir

Archive of SID



خوشه‌بندی حاصل‌شده در درختان سلسله مراتبی می‌توان خوشه‌هایی از سهام را با توجه به نزدیکی و روابط اقتصادی‌شان شناسایی کرد. این ساختارها به ایجاد یک پایگاه ارزشمند برای مدل‌سازی نظری و تجزیه و تحلیل بیشتر بازارهای سهام کمک می‌کنند.

بنابراین به صورت خلاصه مسئله این پژوهش عبارت است از خوشه‌بندی سهام ۳۰ شرکت بزرگ بورس اوراق بهادار تهران که بر طبق تعریف، شرکت-های حاضر در یک خوشه دارای بیشترین شباهت با یکدیگر و بیشترین فاصله (عدم تشابه) با شرکت‌های خوشه‌های دیگر هستند. هدف اصلی این تحقیق به‌صورت کلی عبارت است از خوشه‌بندی شرکت‌هایی که در قلمرو مطالعه حضور دارند. در این پژوهش سعی می‌شود تا انواع روش‌های خوشه‌بندی سلسله مراتبی به‌تفصیل شرح داده شود و با به‌کارگیری این روش‌ها بر روی مجموعه سهام موردمطالعه، بهترین روش برای ایجاد پرتفوی‌ای بهینه از سهام شناسایی گردد. بر اساس آنچه گفته شد، دو سؤال پژوهشی زیر موردبررسی قرار می‌گیرد:

۱) با روش‌های خوشه‌بندی سلسله مراتبی، خوشه‌های مختلف سهام در قلمرو موردمطالعه، چگونه صورت‌بندی می‌شوند.

۲) کدام‌یک از انواع روش‌های خوشه‌بندی سلسله مراتبی به‌کار رفته در قلمرو مورد مطالعه با واقعیت‌ها نزدیکی بیشتری دارد.

## ۲- مبانی نظری و مروری بر پیشینه پژوهش

### ۲-۱- سبد سهام

سرمایه‌گذاران همواره در پی راه‌هایی برای کسب درآمدی مناسب از سرمایه‌گذاری هایشان هستند. مسئله انتخاب سهام یکی از مسائل مهم و پیچیده درزمینهٔ مالی و تصمیمات سرمایه‌گذاری است. این شاخه از مسائل به‌صورت اساسی به مسئله عملکرد مناسب سهام در آینده و ساختن سبد بهینه‌ای از سهام مرتبط است. انتخاب یک سهم و یا سبدی از سهام (پرتفوی) که بتواند سودآوری بهینه‌ای را به دنبال داشته باشد از مهم‌ترین دغدغه‌های سرمایه‌گذاران در بازارهای مالی محسوب می‌شود. مسئله انتخاب سهام شامل ایجاد سبد سهامی است که مطلوبیت سرمایه‌گذار را حداکثر می‌سازد و تنوع‌بخشی و تشکیل سبد سهام و بهینه‌سازی آن یکی از شروط برای موفقیت در بازارهای سرمایه کارآمد است.

مدیریت سبد سهام یکی از اساسی‌ترین مشکلات درزمینهٔ سرمایه‌گذاری است. در اصل نگرانی از این است که چگونه افراد تصمیم می‌گیرند تا اوراق بهادار را در سبدهای سرمایه‌گذاری نگه‌داری کنند و چگونه باید بودجه بین انواع مختلفی از دارایی‌ها مانند سهام یا اوراق قرضه اختصاص داده شود. هدف اصلی از انتخاب یک مجموعه از دارایی‌های دارای ریسک، ایجاد یک سبد سهام به‌منظور حداکثر ساختن بازده تحت یک ریسک خاص یا حداقل ساختن ریسک برای به دست آوردن یک بازدهی خاص است. برای سرمایه‌گذاران حیاتی است تا بهترین انتخاب سبد سهام را به‌منظور حمایت از تصمیم‌گیری‌های مالی داشته باشند (ژانگ و ژوو، ۲۰۰۴).

بهینه‌سازی سبد سهام عبارت است از انتخاب بهترین ترکیب از دارایی‌های مالی به‌نحوی‌که سبب شود تا جایی که امکان دارد، بازده سبد سرمایه‌گذاری حداکثر و ریسک سبد سهام حداقل شود. ایده اصلی نظریه پرتفوی این است که اگر سرمایه‌گذاری بر روی دارایی‌هایی انجام گیرد که به‌طور کامل با یکدیگر همبستگی ندارند، در این صورت ریسک آن دارایی‌ها یکدیگر را خنثی کرده و بنابراین می‌توان یک بازده ثابت را با مقدار ریسک کمتر به دست آورد (مارکوویتز ۱۹۵۲). با انتشار مدل مارکویتز، تغییرات و بهبودهای فراوانی در شیوه نگرش مردم به سرمایه‌گذاری و سبد سهام ایجاد شد و به‌عنوان یک ابزار کارا برای بهینه‌سازی سبد سهام مورداستفاده قرار گرفت (لای، یو، وانگ و ژو، ۲۰۰۶). این مدل در مواجهه با محدودیت‌هایی همچون افزایش داده‌ها و انواع مختلف و تعداد زیادی از



www.SID.ir

Archive of SID



سرمایه‌ها در بازارهای سرمایه کنونی کارایی ندارد؛ بنابراین می‌توان از روش‌های اکتشافی چون روش‌های خوشه‌بندی سلسله مراتبی برای حل این مسئله استفاده کرد.

### ۲-۲- خوشه‌بندی سلسله مراتبی

بسیاری از سیستم‌های پیچیده مشاهده‌شده در علوم فیزیک، بیولوژی و اجتماعی در یک ساختار سلسله مراتبی تودرتو سازمان‌یافته‌اند؛ یعنی عناصر سیستم می‌تواند در خوشه‌هایی تقسیم شوند که به‌نوبه خود می‌تواند در زیر خوشه‌ها بخش‌بندی شوند و به همین ترتیب تا یک سطح معین. ساختار سلسله مراتبی از تعامل بین عناصر به‌شدت پویایی، سیستم‌های پیچیده را تحت تأثیر قرار می‌دهد؛ بنابراین توصیف کمی از سلسله‌مراتب‌های سیستم یک گام نهایی در مدل‌سازی سیستم‌های پیچیده است (سایمون،۱۹۹۱). تجزیه‌وتحلیل داده‌های چند متغیره در تحقیقات مربوط به طیف گسترده‌ای از سیستم‌ها، اطلاعات حیاتی را فراهم می‌کنند. در میان روش‌های چند متغیره، نامزد بدیهی برای تشخیص ساختار سلسله مراتبی از یک مجموعه از داده‌ها، روش‌های خوشه‌بندی سلسله مراتبی هستند (آندربرگ، ۱۹۷۳). خوشه‌بندی سلسله مراتبی به دلیل رشد نمایی داده‌های جهان واقعی، از اهمیت زیادی در تجزیه‌وتحلیل داده‌ها برخوردار است. اغلب این داده‌ها بدون برچسب هستند و دامنه دانش در دسترس قبلی برای تسهیل برچسب زدن دستی این داده‌ها در دسترس یا کافی نیست. در اصل چالش اصلی هزینه محاسباتی این کار است. در چنین شرایطی، خوشه‌بندی یک گزینه مناسب‌تر نسبت به روش‌های یادگیری نظارت‌شده مانند طبقه‌بندی و رگرسیون است. خوشه‌بندی سلسله مراتبی یکی از روش‌های خوشه‌بندی است که قابل‌اجرا بر روی انواع داده است. روش‌های خوشه‌بندی سلسله مراتبی خوشه‌هایی باکیفیت بالاتر تولید می‌کنند (بوگتیا و همکاران، ۲۰۱۵).

تجزیه‌وتحلیل خوشه‌بندی سلسله مراتبی یکی از رایج‌ترین روش‌ها در تجزیه‌وتحلیل خوشه‌ای است. این روش یک درخت دودویی از داده‌ها را می‌سازد که به‌طور پی‌درپی گروه‌های مشابهی از نقاط داده را ادغام می‌کند (آریه و آریه، ۲۰۱۴). ساختار درخت ایجادشده توسط خوشه‌بندی سلسله مراتبی، دندروگرام[۲] نامیده می‌شود که معمولاً برای نشان دادن روند خوشه‌بندی استفاده می‌شود. این نشان می‌دهد که چگونه اشیاء به‌صورت گام‌به‌گام باهم گروه‌بندی می‌شوند (هان و کمبر، ۲۰۰۱). دندروگرام در لغت به معنای یک نمودار درختی از دندروم یونانی برای درخت و گرام برای ترسیم نشأت می‌گیرد (پائولوس و کریستوفک، ۲۰۱۵).

روش سلسله مراتبی، یک تجزیه سلسله مراتبی از مجموعه داده اشیاء ارائه‌شده را ایجاد می‌کند. این روش را می‌توان بر مبنای چگونگی شکل‌گیری تجزیه‌وتحلیل سلسله مراتبی به‌صورت یکی از دو روش تجمعی[۳] و یا تقسیم‌کننده[۴] طبقه‌بندی کرد. در روش تجمعی، عملکرد الگوریتم پایین به بالا است، یعنی الگوریتم با خوشه‌هایی که هرکدام شامل یک نقطه داده است شروع می‌شود و ادغام خوشه‌ها ادامه می‌یابد. این روش به‌طور متوالی اشیاء یا گروه‌هایی که نزدیک به هم هستند را ادغام می‌کند. تا زمانی که همه گروه‌ها درون یک خوشه ادغام شوند (بالاترین سطح از سلسله‌مراتب) و یا تا زمانی که یک شرط در نظر گرفته‌شده پایان یابد. در روش تقسیم‌کننده، عملکرد الگوریتم از بالا به پایین است، یعنی الگوریتم با یک خوشه بزرگ شامل تمام نقاط داده در مجموعه داده شروع می‌شود. در هر تکرار پی‌درپی، یک خوشه به خوشه‌های کوچک‌تر تقسیم می‌شود تا درنهایت هر شیء در یک خوشه قرار گیرد و یا تا زمانی که یک شرط در نظر گرفته‌شده پایان یابد (گان، ما و وو، ۲۰۰۷).

در اصل الگوریتم‌های خوشه‌بندی تجمعی با هر مشاهده نشان‌دهنده یک خوشه تک شروع می‌شود و در هر یک از N-1 مرحله، نزدیک‌ترین دو خوشه



www.SID.ir

Archive of SID



(حداقل غیرمشابه) به یک خوشه منفرد ادغام می‌شوند، درواقع تولید یک خوشه کمتر در سطح بالاتر بعدی است. استراتژی تجمعی در پایین شروع می‌شود و در هر سطح به‌صورت بازگشتی یک جفت انتخاب‌شده از خوشه‌ها را به یک خوشه تک ادغام می‌کند. این جفت انتخاب‌شده برای ادغام، دو گروه با کوچک‌ترین تفاوت میان گروهی را شامل می‌شوند (هیستی، تیبشیرانی و فریدمن، ۲۰۰۹). ازآنجاکه شباهت برای تعریف یک خوشه اساسی است، یک معیار شباهت بین دو الگوی رسم شده از یک فضای ویژگی یکسان، برای بیشتر روش‌های خوشه‌بندی ضروری است. فاصله و شباهت مفاهیم متقابلی هستند. اغلب معیارهای شباهت و ضرایب شباهت برای توصیف کمی چگونگی شباهت دو نقطه داده یا چگونگی شباهت دو خوشه استفاده می‌شوند. رایج‌ترین محاسبه عدم تشابه بین دو الگو، از یک معیار فاصله تعریف‌شده بر روی فضای ویژگی استفاده می‌کند. نزدیکی الگوها به‌طورمعمول به‌وسیله یک تابع فاصله تعریف‌شده بر روی جفتی از الگوها اندازه‌گیری می‌شود (آربی، بیر، کریزلی و کینس، ۲۰۰۶).

چهار معیار گسترده برای اندازه‌گیری فاصله بین خوشه‌ها وجود دارد که به شرح زیر است:

رابطه ۱) حداقل فاصله: $d_{min}(c_i, c_j) = min_{p\in c_i, p'\in c_j}|p-p'|$

رابطه ۲) حداکثر فاصله: $d_{max}(c_i, c_j) = max_{p\in c_i, p'\in c_j}|p-p'|$

رابطه۳) متوسط فاصله: $d_{mean}(c_i, c_j) = |m_i - m_j|$

رابطه ۴) میانگین فاصله: $d_{avg}(c_i, c_j) = \frac{1}{n_i n_j}\sum_{p\in c_i}\sum_{p'\in c_j}|p-p'|$

که در آن $|p-p'|$ فاصله بین دو شیء یا نقطه است و $m_i$ میانگین خوشه $c_i$ و $n_i$ تعداد اشیاء درون $c_i$ است (هان و کمبر، ۲۰۰۱). با توجه به معیارهای فاصله مختلف بین گروه‌ها، روش‌های سلسله مراتبی تجمعی را می‌توان به روش‌های تک پیوندی[۵]، پیوند کامل[۶]، پیوند میانگین گروهی[۷]، روش مرکزی[۸] و روش وارد[۹] طبقه‌بندی کرد (گان، ما و وو، ۲۰۰۷).

### ۲-۳- پیشینه پژوهش

بازارهای مالی به‌عنوان سیستم‌های پیچیده‌ایی که به خوبی تعریف شده‌اند، توسط اقتصاددان‌ها، ریاضی‌دان‌ها و همچنین به‌تازگی توسط فیزیک‌دان‌ها موردمطالعه قرارگرفته‌اند. با توجه به کاربردهای موفق الگوریتم‌های خوشه‌بندی سلسله مراتبی، این روش‌ها به صورت گسترده‌ای در پژوهش‌های مالی مورد استفاده قرار گرفته‌اند. یکی از پیشرویان این زمینه منتگنا[۱۰] (۱۹۹۹) است. وی با استفاده از روش سلسله مراتبی کشف کرد که یک سازمان سلسله مراتبی از سهم‌ها در بازارهای مالی وجود دارد. در پژوهشی دیگری بننو[۱۱]، لیلو[۱۲] و منتگنا (۲۰۰۱) با تجسمی از همبستگی متقابل توسط یک درخت حداقل پوشا و همچنین با انجام خوشه‌بندی سلسله مراتبی فوق متریک، نشان دادند که همبستگی متقابل قوی بین سهم‌های مشغول به فعالیت در یک بخش همسان وجود دارد. مطالعه همبستگی بین سهم‌های مختلف بر اساس تجزیه‌وتحلیل خواص ماتریس همبستگی آن‌ها با توجه به نوسانات قیمتی‌شان، برای فهم رفتارهای بازارهای مالی از اهمیت قابل‌توجهی برخوردار است. مطالعات متعددی حاکی از آن است که استفاده از روش‌های خوشه‌بندی سلسله مراتبی در تشخیص ساختار و چگونگی ارتباط سهم‌های مختلف در بازارهای مالی، عملکرد مطلوبی را به دنبال داشته‌اند که ازجمله آن‌ها می‌توان به پژوهش‌های گرس[۱۳] و آرگیورکیس[۱۴] (۲۰۰۷)، بریدا[۱۵] و ریزو[۱۶] (۲۰۰۷)، تاباک[۱۷]، سرا[۱۸] و کاجوئیرو[۱۹] (۲۰۱۰) اشاره کرد.



www.SID.ir

Archive of SID



در پژوهشی یک روش ترکیبی تجزیه‌وتحلیل نمادین سری‌های زمانی با الگوریتم خوشه‌بندی پیوند منفرد نزدیک‌ترین مجاور برای ساخت پرتفوی و برای شناسایی خوشه‌ها در شبکه‌های صنعتی شرکت‌های بزرگ امریکای شمالی به‌کاررفته است. پس از ساخت درختان سلسله مراتبی، برخی خوشه‌ها در شبکه‌هایی همبستگی یافته‌اند که از یک دیدگاه اقتصادی مفهومی را ایجاد می‌کنند (بریدا و ریزو، ۲۰۱۰). تجزیه‌وتحلیل توپولوژی ۴۰ شرکت از بزرگ‌ترین شرکت‌های انگلیسی و یا دارای سرمایه بسیار برای تأیید پیش‌بینی بازار در طی دوره ۲۰۰۶ تا ۲۰۱۰ مطالعه دیگری است که با استفاده از مفهوم تجزیه‌وتحلیل درخت حداقل پوشا و درخت سلسله مراتبی انجام‌گرفته است. نتایج این پژوهش نشان می‌دهد که درختان ابزار مفیدی برای درک و تشخیص ساختار جهانی، رده‌بندی شناخته‌شده و سلسله‌مراتب در داده‌های مالی هستند (اولوزوی و همکاران، ۲۰۱۲). در تحقیقی باهدف بررسی خواص توپولوژیکی شبکه‌های همبستگی در بین شرکت‌های خودروسازی جهانی بر اساس داده‌های قیمت و با استفاده از مفهوم درخت سلسله مراتبی دریافتند که از توپولوژی ساختاری این درختان، خوشه‌های مختلفی از شرکت‌ها با توجه به نزدیکی جغرافیایی و روابط اقتصادی‌شان شناسایی‌شده‌اند و نتایج نشان می‌دهد که برخی از شرکت‌ها به دلیل اتصال نزدیکشان با دیگر شرکت‌ها، در این شبکه مهم‌تر هستند (کوکیکاپلین، دوگان، دیوایرن و کسکین، ۲۰۱۳). در مطالعه‌ای برای بررسی شبکه‌های تشکیل‌شده توسط شاخص صنعت CITIC (شرکت اعتماد بین‌المللی و سرمایه‌گذاری چین) در طی سه دوره ۲۰۰۶ تا ۲۰۱۳ از دو تکنیک درخت حداقل پوشا و درخت سلسله مراتبی استفاده‌شده است. در این مطالعه ویژگی‌های ساختاری شبکه از طریق مدل‌سازی شبکه پیچیده‌ای از بازار سهام چین قبل و بعد از بحران مالی مورد تجزیه‌وتحلیل قرارگرفته است. نمودار درخت سلسله مراتبی ایجادشده، اطلاعات طبقه‌بندی‌شده‌ای از شاخص‌های صنعت CITIC را ارائه کرده است و توزیع و اثرات ارتباطی بین خوشه‌های صنعت را آشکار ساخته است (یانگ، لی و زانگ، ۲۰۱۴).

تن[۲۰] و همکارانش (۲۰۱۵) در پژوهشی برای نشان دادن ثبات همبستگی متقابل و سلسله‌مراتب بازارهای مالی در طول زمان، خوشه‌بندی سری‌های زمانی را در طی زمان‌های مختلف در طول سال‌های ۲۰۰۶ و ۲۰۰۷ انجام دادند. در این پژوهش با انجام خوشه‌بندی سلسله مراتبی جزئی پیوند کامل، دریافته شد که خوشه‌هایی از سهم‌ها در بازارهای مختلف وجود دارد که این خوشه‌ها به دو ابر خوشه سازمان‌یافته‌اند که این فعل‌وانفعالات ممکن است مسئول علی برای سقوط در بازارهای بورس اوراق بهادار باشند. روش‌های خوشه‌بندی سلسله مراتبی بر روی بازارهای مالی مختلف دیگری ازجمله نرخ‌های تبادلات خارجی اعمال‌شده‌اند که از آن جمله می‌توان به کارهای مزیونو[۲۱]، تاکایوزیو[۲۲] و تاکایوزیو (۲۰۰۶)، بریدا، گومز[۲۳] و ریزو (۲۰۰۹)، کوکیکاپلین[۲۴]، دیوایرن[۲۵] و کسکین[۲۶] (۲۰۱۲) اشاره کرد که با استفاده از روش‌های مختلف خوشه‌بندی سلسله مراتبی توانستند تا ساختارهای خوشه‌ای از ارزها و ارز کلیدی در هر خوشه در بازار موردمطالعه خود را شناسایی کنند.

## ۳- روش‌شناسی پژوهش

در این پژوهش به‌منظور اندازه‌گیری همبستگی بین قیمت‌های سهام، شناسایی روابط بین آن‌ها و شناسایی گروه‌هایی از قیمت‌های سهام مشابه از روش‌های خوشه‌بندی سلسله مراتبی تجمعی استفاده‌شده است. نوع پژوهش بر اساس هدف، کاربردی است و بر اساس چگونگی به دست آوردن داده‌های مورد نیاز از نوع توصیفی- تحلیلی محسوب می شود.



www.SID.ir

Archive of SID



**۳-۱- الگوریتم خوشه بندی سلسله مراتبی**

برای انجام خوشه‌بندی سلسله مراتبی روش‌های مختلفی وجود دارد که به‌طورکلی مراحل زیر را دنبال می‌کنند:

۱) شروع با تعریف ضرایب همبستگی متقابل از تفاوت لگاریتمی محاسبه‌شده از قیمت سهام در یک افق زمانی انتخاب‌شده، در اصل ابتدا باید تابع همبستگی بین یک جفت از قیمت‌های سهام تعریف شود. اجازه داده می‌شود تا $p_i(t)$ قیمت بسته‌شده سهام شرکت i در زمان t و $p_i(t-1)$ قیمت سهام در دوره t-1 باشد. سپس یک تغییر در یک فاصله زمانی $\Delta t$ به‌صورت زیر تعریف می‌شود:

$$Y_i(t) = lnp_i(t) - lnp_i(t-1) \qquad \text{رابطه ۵)}$$

برای هر مقدار از $\Delta t$، ضریب همبستگی برای تفاوت‌های لگاریتم قیمت (که تقریباً همزمان با بازده‌های سهام است) بین تمام جفت‌های ممکن از سهام موجود در پرتفوی موردنظر محاسبه می‌شود. با استفاده از $Y_i(t)$، ضریب همبستگی پیرسون بین یک جفت از قیمت‌های سهام یا قیمت‌های منفرد سهام می‌تواند توسط تابع همبستگی متقابل به‌صورت زیر محاسبه شود:

$$\rho_{ij} = \frac{\langle Y_i Y_j\rangle - \langle Y_i\rangle\langle Y_j\rangle}{\sqrt{(\langle {Y_i}^2\rangle - \langle Y_i\rangle^2)(\langle {Y_j}^2\rangle - \langle Y_j\rangle^2)}} \qquad \text{رابطه۶)}$$

که در آن $i$ و $j$ نمادهای عددی از سهم‌ها هستند و $\langle \dots \rangle$ نشان‌دهنده متوسط آماری در طول دوره موردمطالعه است. ضریب همبستگی $\rho_{ij}$ می‌تواند محدوده‌ای از ۱- تا ۱+ داشته باشد، که در آن به ترتیب ۱- بدین معنا است که دو قیمت سهام همبستگی کاملاً مخالف دارند و ۱+ یعنی دو قیمت سهام کاملاً همبسته هستند. زمانی که $\rho_{ij} = 0$ باشد، قیمت‌های دو سهم غیر همبسته هستند.

ماتریس ضریب همبستگی یک ماتریس متقارن با $\rho_{ii} = 1$ در قطر اصلی است. در اصل یک ماتریس متقارن N×N با عناصر مورب مساوی است.

۲) مرحله بعدی محاسبه یک فاصله متریک بین تمام جفت‌های سهام یا در اصل محاسبه ماتریس مجاورت است. ضریب همبستگی از یک جفت از سهم‌ها نمی‌تواند به‌عنوان یک فاصله بین دو سهم استفاده شود، چراکه آن سه اصل متریک را برآورده نمی‌سازد. بااین‌حال یک متریک می‌تواند با استفاده از یک تابع فاصله از ضریب همبستگی تعریف شود (منتگنا، ۱۹۹۹).

یکی از فواصلی که به‌طورمعمول مورداستفاده قرار می‌گیرد، فاصله اقلیدسی[۲۷] است. چند هزار سال پیش اقلیدس اظهار داشت که کوتاه‌ترین فاصله بین دو نقطه، یک رابطه خطی است. فاصله اقلیدسی رایج‌ترین فاصله‌ای است که تاکنون برای داده‌های عددی مورداستفاده قرارگرفته و به‌صورت زیر تعریف می‌شود (چا، ۲۰۰۸).

$$d_{EUC} = \sqrt{\sum_{i=1}^{d} |P_i - Q_i|^2} \qquad \text{رابطه ۷)}$$

یک تابع فاصله متریک مناسب بین جفت‌هایی از سهام به‌صورت زیر است:

$$d(i,j) = \sqrt{2(1-\rho_{ij})} \qquad \text{رابطه ۸)}$$

با این انتخاب $d(i,j)$، سه اصل از یک فاصله متریک به‌صورت زیر برآورده می‌شود:

$$(1)\ d(i,j) = 0 \leftrightarrow i = j \qquad \text{رابطه ۹)}$$

$$(2)\ d(i,j) = d(j,i) \qquad \text{رابطه ۱۰)}$$

$$(3)\ d(i,j) \leq d(i.k) + d(k,j) \qquad \text{رابطه ۱۱)}$$



www.SID.ir

Archive of SID



اولین اصل معتبر است، چراکه اگر و تنها اگر همبستگی کامل باشد ($\rho = 1$)، یعنی اگر دو سهم یک فرآیند تصادفی همانند را انجام دهند، آنگاه $d(i,j) = 0$ است.

اصل دوم معتبر است، چراکه ماتریس ضریب همبستگی و ازاین‌رو ماتریس فاصله D با توجه به تعریفشان متقارن هستند.

اصل سوم معتبر است، چراکه معادله ارائه‌شده معادل با فاصله اقلیدسی بین دو بردار $\overrightarrow{Y_i}$ و$\overrightarrow{Y_j}$ است که از سری های زمانی $Y_i$ و$Y_j$ و با درنظر گرفتن هر ثبتی از سری های زمانی یک جزء از بردار به دست آمده است. بردار به دست آمده باید یک هنجار واحد داشته باشد. یعنی به وسیله کم کردن مقدار میانگین هر ثبت و به وسیله نرمال کردن آن به انحراف استانداردش به دست آید. تابع فاصله معرفی شده $d(i,j)$ می تواند در محدوده $0 < d(i,j) \leq 2$ قرار گیرد. در حالی که همبستگی ها در محدوده ای از ۱- تا ۱+ قرار دارند. همبستگی بالا با مقادیر کم $d(i,j)$ تطابق دارند. یعنی هر چه همبستگی بالاتر باشد، فاصله بین دو سهم کمتر است و بالعکس (منتگنا، ۱۹۹۹). برای ساخت درخت سلسله مراتبی از قیمت‌های سهام با استفاده از ماتریس فاصله D، ابتدا باید یک ماتریس فوق‌متریک تبعی $D^<$ با عناصر $d_{ij}^<$ را تعیین‌کرد. فاصله $d_{ij}^<$ سه اصل فاصله اقلیدسی را برآورده می‌سازد و از نابرابری فوق متریک زیر با یک شرط قوی‌تر نسبت به اصل سوم پیروی می‌کند (وانگ و زیه، ۲۰۱۵).

$$d_{ij}^< \leq max\{d_{ik}, d_{kj}\} \qquad \text{رابطه ۱۲)}$$

۳) مرحله بعدی جستجو برای حداقل فاصله در ماتریس به‌دست آمده است.

۴) سپس ترکیب دو خوشه با حداقل فاصله.

۵) در ادامه به روز رسانی ماتریس مجاورت با محاسبه فاصله بین خوشه های جدید با دیگر خوشه ها.

۶) تکرار سه مرحله قبلی در صورتی که بیش از یک خوشه باقی مانده باشد (بوگتیا و همکاران، ۲۰۱۵).

### ۳-۲- انواع روش‌های خوشه‌بندی سلسله مراتبی تجمعی

در این پژوهش روش‌های خوشه‌بندی سلسله مراتبی تجمعی جهت ساخت درختان سلسله مراتبی به کار گرفته می‌شوند. همه این روش‌ها مراحل بیان شده در بخش قبل را دنبال می‌کنند. تنها تفاوت این الگوریتم‌ها در روشی است که آن‌ها شباهت بین یک جفت از خوشه‌ها را مشخص می‌کنند. روش تک پیوندی یکی از ساده‌ترین روش‌های خوشه‌بندی سلسله مراتبی است که با عنوان روش نزدیک‌ترین مجاور نیز شناخته می‌شود (بریدا و ریزو، ۲۰۱۰).

روش تک پیوندی فاصله نزدیک‌ترین مجاور یعنی حداقل تفاوت را برای اندازه‌گیری عدم تشابه بین دو گروه بکار می‌برد (هیستی، تیبشیرانی و فریدمن، ۲۰۰۹). اگر X و Y دو خوشه در نظر گرفته شوند، فاصله بین نزدیک‌ترین نقطه در خوشه‌های X و Y به‌صورت زیر تعریف می‌شود:

$$D_{single}(X,Y) = \min d(x,y) \qquad \text{رابطه۱۳)}$$
$$x \in X, y \in Y$$

برخلاف روش تک پیوندی، روش پیوند کامل دورترین فاصله مجاور را برای اندازه‌گیری عدم تشابه بین دو گروه استفاده می‌کند. این روش با عنوان روش دورترین مجاور نیز شناخته می‌شود. اگر X و Y دو خوشه در نظر گرفته شوند، فاصله بین دورترین نقطه در خوشه‌های X و Y به‌صورت زیر تعریف می‌شود:



www.SID.ir

Archive of SID



$$D_{CL}(X,Y) = max\, d(x,y) \quad x \in X, y \in Y$$ رابطه ۱۴)

روش میانگین گروهی با عنوان UPGMA نیز شناخته می‌شود که برای گروه‌بندی جفت‌های بدون وزن، از میانگین حسابی استفاده می‌کند. در روش میانگین گروهی، فاصله بین دو گروه به‌صورت میانگین فاصله بین تمام جفت‌های ممکن از مشاهدات مربوط به هر خوشه تعریف می‌شود.

اگر x و y دو خوشه در نظر گرفته شوند، میانگین تفاوت بین گروها درروش میانگین گروهی صورت زیر تعریف می‌شود (گان، ما و وو، ۲۰۰۷):

رابطه۱۵)

$$d_{GA}(X,Y) = \frac{1}{N_X N_Y} \sum_{x \in X} \sum_{y \in Y} d(x,y)$$

روش وارد یکی از الگوریتم‌های خوشه‌بندی تجمعی است که مبتنی بر ارتباط متریک نیست (کوگن، نیکولاس و تیبول، ۲۰۰۶).

در پیوند وارد، کل افزایش مجموع مربع درون خوشه‌ای به‌عنوان فاصله بین دو خوشه برای پیوند در نظر گرفته می‌شود، به‌طوری‌که به پیوند دو خوشه منجر می‌شود. این فاصله به‌صورت زیر تعریف می‌شود (گان، ما و وو، ۲۰۰۷):

رابطه ۱۶)

$$\mathrm{D_{Ward}(X,Y)} = (2[\mathrm{SS(X \cup Y)} - (\mathrm{SS(X)} + \mathrm{SS(Y)}]^{\frac{1}{2}})$$

روش مرکزی با عنوان روش گروهی جفت بدون وزن با استفاده از مرکز (UPGMC) نیز شناخته می‌شود. دو خوشه بر اساس فاصله بین مراکزشان ادغام می‌شوند.

$$D_{X,Y} = D_{\bar{x},\bar{y}}$$ رابطه ۱۷)

که در آن $\bar{x}$ و $\bar{y}$ به ترتیب مراکزی از خوشه X و Y هستند که به‌وسیله میانگین حسابی از اشیاء مربوطه در خوشه محاسبه می‌شوند (بوگتیا و همکاران، ۲۰۱۵).

## ۴- یافته‌های پژوهش

### ۴-۱- داده‌های پژوهش

داده‌های مورداستفاده در این تحقیق دربرگیرنده قیمت‌های روزانه مربوط به سهم‌های شاخص ۳۰ شرکت بزرگ بورس اوراق بهادار تهران در سال مالی ۱۳۹۳ است. این اطلاعات بر اساس گزارش‌های بورس اوراق بهادار تهران گردآوری‌شده است. ازآنجاکه درباره برخی از این شرکت‌ها داده آماری یافت نشد، سهم‌های مربوط به این شرکت‌ها از نمونه حذف شد و تنها ۲۷ سهم را که باوجود اطلاعات کامل و تبادلات تجاری کافی می‌توانند همبستگی را برآورد کنند، برای انجام تحقیق در نظر گرفته شدند. با توجه به اینکه در تاریخ‌های محدودی از روزهای معاملاتی، داده‌های قیمت در مورد مجموعه سهام موردمطالعه موجود نبود، برای رفع این مشکل و دستیابی به این داده‌ها، روش درون‌یابی به کار گرفته‌شده است. برای این منظور از نرم‌افزار Rapid Miner استفاده‌شده است.

### ۴-۲- پیاده‌سازی انواع روش‌های خوشه‌بندی سلسله مراتبی

به‌منظور ایجاد یک درخت سلسله مراتبی، ابتدا باید بازده‌های لگاریتمی قیمت‌های روزانه در طی دوره زمانی موردنظر محاسبه شوند. به کمک رابطه ۵، بازده‌های لگاریتمی قیمت‌های مربوط به ۳۰ شرکت بزرگ بورس اوراق بهادار تهران در سال ۱۳۹۳ محاسبه‌شده است.

در این پژوهش، ضریب همبستگی بین بازده سهم‌ها با توجه به رابطه ۶ محاسبه شد؛ به‌طوری‌که ماتریسی ۲۷×۲۷ با قطر اصلی یک به دست می‌آید. برای اندازه‌گیری فاصله بین سهم‌ها از معیار فاصله اقلیدسی استفاده‌شده است. در یک ماتریس فاصله، هر سهم بر اساس فاصله اقلیدسی با دیگر سهم‌ها مرتبط است. در این مورد فاصله نشان‌دهنده تشابه‌هایی از مجموعه موردنظر است. برای مثال اگر

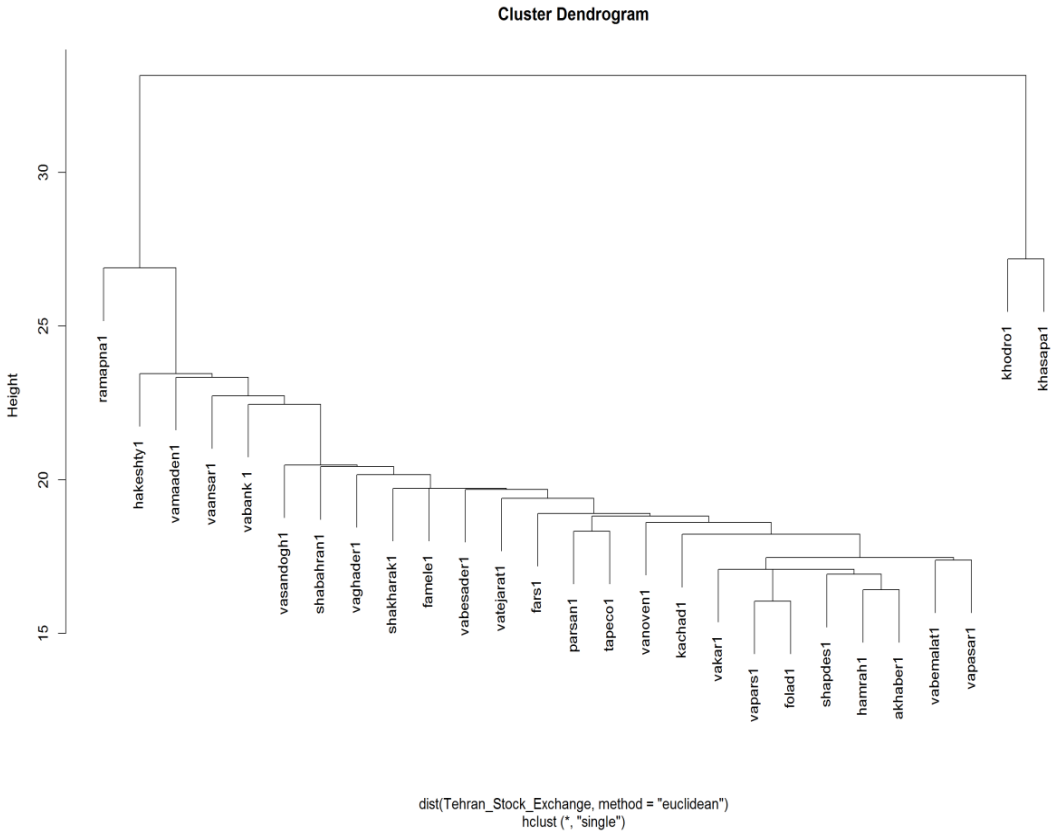


www.SID.ir

Archive of SID



بدانیم که میزان شباهت دو سهم بالا است، در اصل یعنی میزان فاصله کم است. با در نظر گرفتن ماتریس فاصله کامل به‌دست‌آمده با استفاده از ضرایب همبستگی، از طریق کد نویسی در نرم‌افزار R الگوریتم‌های مختلف سلسله مراتبی اجرا و درختان سلسله مراتبی از ماتریس فاصله همسان حاصل‌شده است. درختان سلسله مراتبی به‌دست‌آمده توسط هر یک از الگوریتم‌ها، به ترتیب در شکل‌های ۱ الی ۷ ارائه‌شده‌اند.




**شکل ۱- دندروگرام سهم‌های شاخص ۳۰ شرکت بزرگ بورس اوراق بهادار تهران در سال ۱۳۹۳ با استفاده از روش تک پیوندی و معیار فاصله اقلیدسی**

روش تک پیوندی حداقل فاصله بین دو جفت از سهم‌ها را جهت ادغام دو خوشه در نظر می‌گیرد. برای شروع با دندروگرام حاصل از روش تک پیوندی در شکل ۱، پنج خوشه قابل‌شناسایی است و می‌توان دید که دو سهم وپارس ۱ و فولاد ۱ دارای کمترین مقدار فاصله در نمونه هستند که نشان‌دهنده یک ارتباط قوی بین این دو سهم از خوشه اول است. خوشه دوم عمدتاً از سهم‌های مربوط به صنعت رادیویی ساخته‌شده است و سهم شپدیس ۱ از صنعت شیمیایی نیز به آن متصل است. همان‌طور که مشاهده می‌شود، این سه سهم ازنظر بازده روزانه دارای همبستگی بالا و فاصله کمی با یکدیگر هستند. سومین خوشه از سهم‌های مربوط به صنعت مالی یا بانک‌ها تشکیل‌شده‌اند. دو سهم از صنعت شیمیایی با توجه به فاصله بین آن‌ها، همبستگی بالایی دارند و چهارمین خوشه را تشکیل می‌دهند. آخرین خوشه بیشترین فاصله را نسبت به سایر خوشه‌ها دارد و از پیوند دو سهم از صنعت خودرو تشکیل‌شده است.

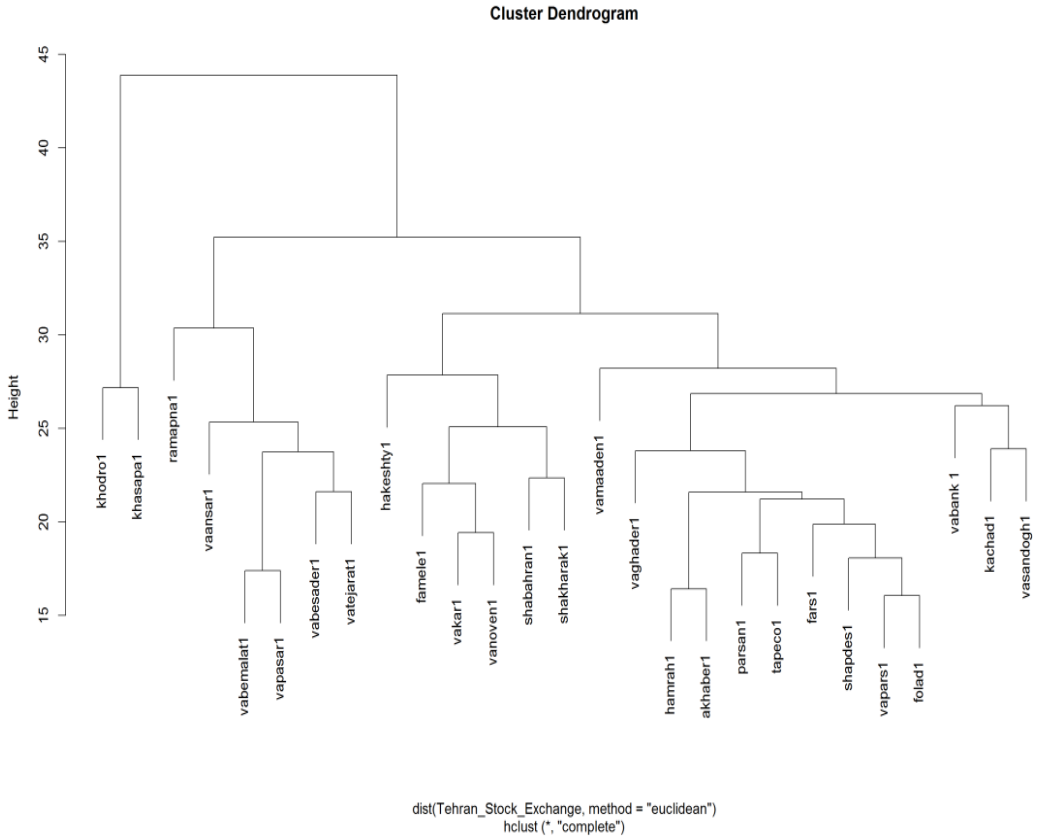


www.SID.ir

Archive of SID



خوشه‌های ایجادشده توسط روش تک پیوندی طویل و زنجیره مانند است که به‌عنوان یکی از خصوصیات این روش محسوب می‌شود. از این اثر زنجیری می‌توان برای تشخیص داده‌های پرت و خوشه‌های منفرد و دورافتاده استفاده‌کرد.




**شکل ۲- دندروگرام سهم‌های شاخص ۳۰ شرکت بزرگ بورس اوراق بهادار تهران در سال ۱۳۹۳ با استفاده از روش پیوند کامل و معیار فاصله اقلیدسی**

روش پیوند کامل دورترین فاصله در بین دو جفت سهام را جهت اندازه‌گیری تشابه دو خوشه به کار می‌برد. در دندروگرام حاصل از روش پیوند کامل که در شکل ۲ نمایش داده‌شده است، هفت خوشه‌ی مختلف از سهم‌ها با توجه به نزدیکی‌شان مشاهده می‌شوند. خوشه اول شامل دو سهم از صنعت رادیویی با توجه به دارا بودن کوتاه‌ترین فاصله، دارای بیشترین همبستگی در بین سایر سهم‌ها می‌باشند. دومین خوشه، خوشه‌ای غیر همگن و شامل سهم‌هایی از صنعت شیمیایی به همراه سهم فولاد ۱ از صنعت فلزات اساسی و سهم وپارس ۱ از صنعت مالی است. خوشه سوم از دو سهم از حوزه مالی تشکیل‌شده است و سهم فملی ۱ از صنعت فلزات اساسی با پیوندی به این خوشه متصل است. چهارمین خوشه از پیوند سهم شبهرن ۱ از صنعت فرآورده‌های نفتی و سهم شخارک ۱ از صنعت شیمیایی ایجادشده است. در این دندروگرام خوشه همگنی از سهم‌های مربوط به صنعت مالی قابل‌شناسایی است که خوشه پنجم را تشکیل می‌دهد و نشان از وجود ارتباط قوی بین این سهم‌ها است. ششمین خوشه شامل دو سهم از صنایع چند رشته‌ای و وکچاد ۱ از صنعت کانه فلزی است.

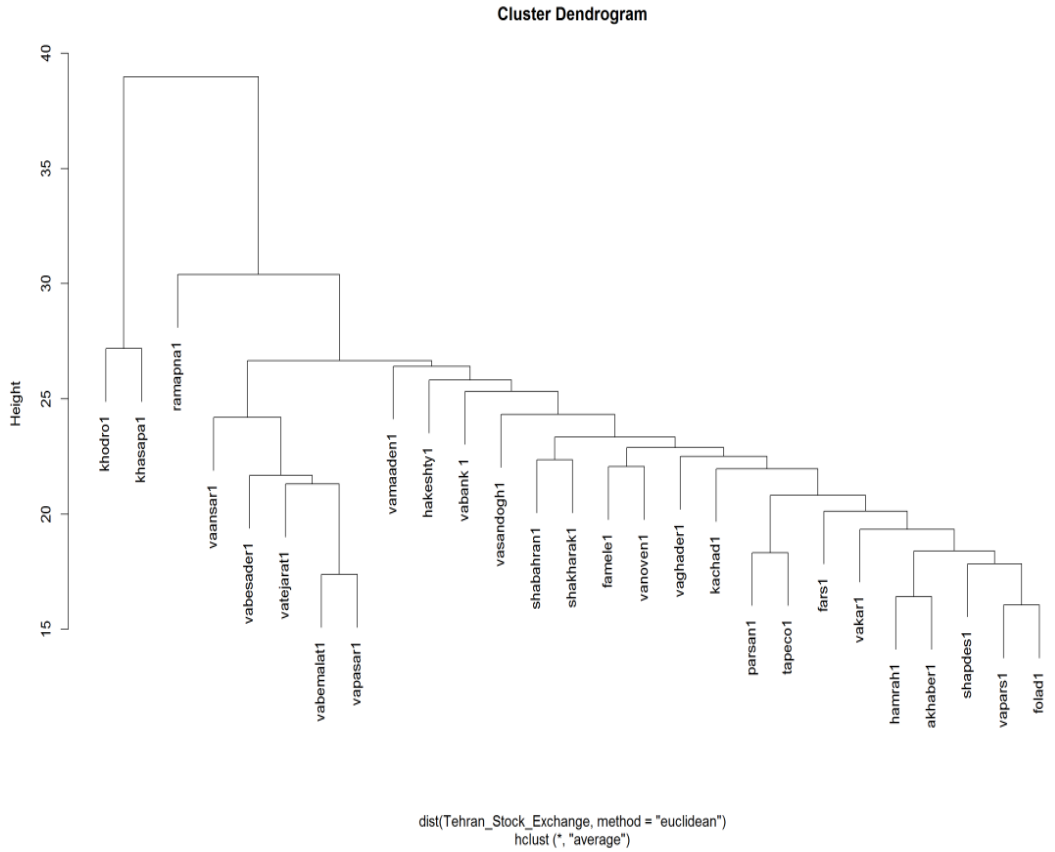


www.SID.ir

Archive of SID



آخرین خوشه که به‌صورت خوشه‌ای مجزا در سمت چپ دندروگرام دیده می‌شود، از دو سهم صنعت خودرو تشکیل‌شده است. با توجه به شکل ۲ و در مقایسه با شکل ۲ می‌توان دید که با استفاده از روش پیوند کامل، سلسله‌مراتب تودرتویی از نمودار همراه با خوشه‌های متراکم و با قطرکمی تولیدشده‌اند که ازنظر بصری، خوشه‌ها به‌راحتی قابل‌تشخیص هستند.




**شکل ۳- دندروگرام سهم‌های شاخص ۳۰ شرکت بزرگ بورس اوراق بهادار تهران در سال ۱۳۹۳ با استفاده از روش میانگین گروهی و معیار فاصله اقلیدسی**

روش میانگین گروهی، میانگین پیوندها را به‌عنوان فاصله بین دو خوشه در نظر می‌گیرد. با به‌کارگیری روش میانگین گروهی، یک ساختار تولیدشده از خوشه‌های نسبتاً متراکم و همچنین خوشه‌های نسبتاً دور از هم یافت می‌شود. دندروگرام شکل ۳ یک حالت سازشی بین روش‌های تک پیوندی و پیوند کامل را نمایان می‌کند. دندروگرام حاصل شامل هفت خوشه مجزا است. اولین خوشه حاوی سهم‌هایی از صنعت رادیویی است که فاصله بین این دو سه کمترین مقدار فاصله در کل مجموعه سهم‌ها است و نشان از همبستگی قوی این دو سهم دارد. خوشه دوم در کل خوشه ناهمگنی از سه سهم شپدیس ۱، وپارس ۱ و فولاد ۱ فعال در صنایع مختلفی را نشان می‌دهد. خوشه همگن سوم مربوط به دو سهم تاپیکو ۱ و پارسان ۱ از صنعت شیمیایی است. خوشه چهارم از پیوند سهم فملی ۱ از صنعت فلزات اساسی و سهم ونوین ۱ از صنعت مالی تشکیل‌شده است. سهم شبهرن ۱ از صنعت فرآورده

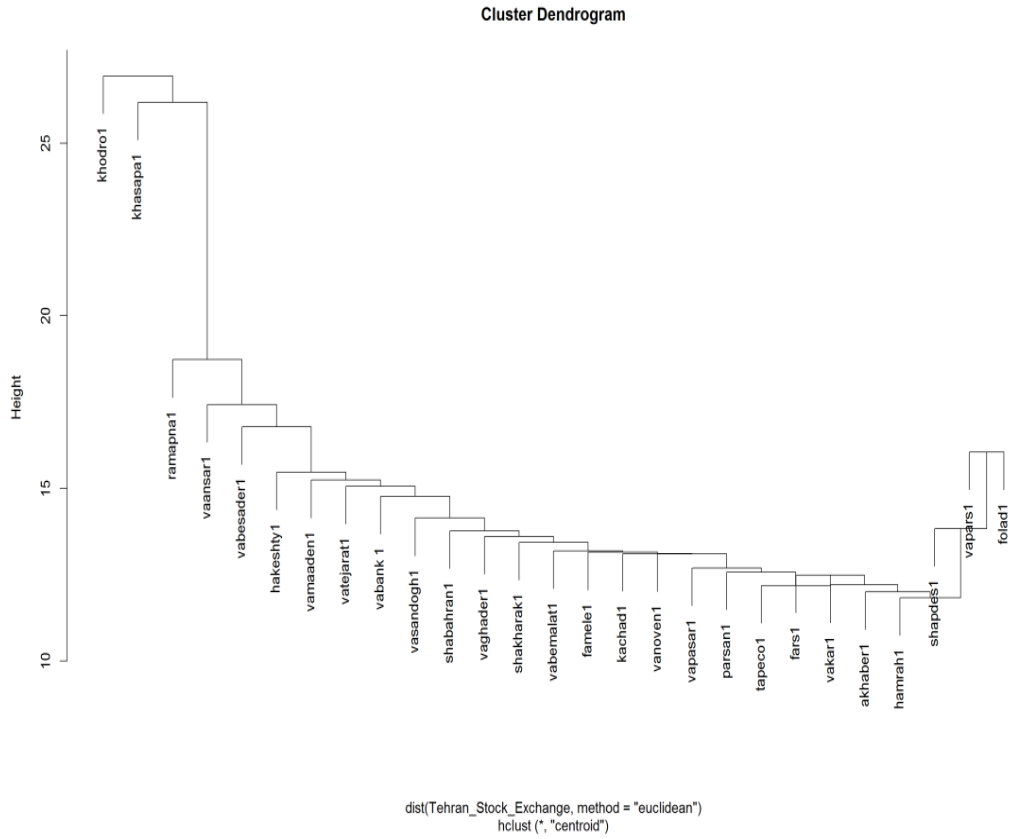


www.SID.ir

Archive of SID



نفتی و سهم شخارک ۱ از صنعت شیمیایی نیز خوشه ناهمگن پنجم را می‌سازند. خوشه ششم یک خوشه همگن از صنعت مالی است. سهم‌های صنعت خودرو در چپ دندروگرام با فاصله بیشتری نسبت به سایر سهم‌ها خوشه هفتم را نشان می‌دهد.




**شکل ۴- دندروگرام سهم‌های شاخص ۳۰ شرکت بزرگ بورس اوراق بهادار تهران در سال ۱۳۹۳ با استفاده از روش مرکزی و معیار فاصله اقلیدسی**

روش مرکزی با در نظر گرفتن فاصله بین مراکز ثقل دو خوشه آن‌ها را با یکدیگر ادغام می‌کند. با توجه به دندروگرام حاصل از روش مرکزی، سهم‌های موردمطالعه به‌طور متوالی در یک خوشه زنجیره مانند به یکدیگر پیوند خورده‌اند. همچنین دو سهم خودرو ۱ و خساپا ۱ با یک فاصله بسیار زیاد نسبت به دیگر سهم‌ها خوشه‌بندی‌شده‌اند. مشاهده می‌شود که دندروگرام موردنظر عملاً اطلاعات مفیدی در مورد چگونگی قرار گرفتن سهم‌ها با در نظر گرفتن حوزه فعالیت صنعتی‌شان و همچنین میزان همبستگی آن‌ها را ارائه نمی‌دهد.

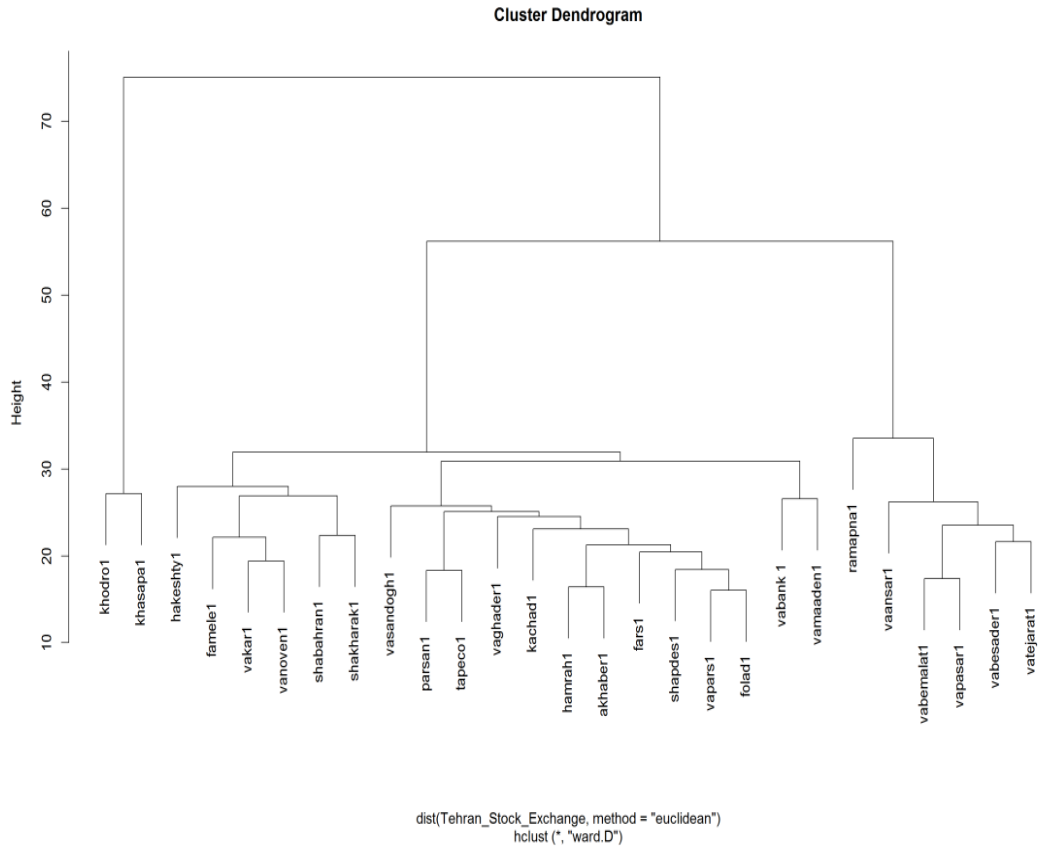
در روش وارد، فاصله بین دو خوشه با در نظر گرفتن کل مجموع مربع تفاضل درون خوشه‌ای محاسبه می‌شود. در شکل ۵ حاصل از به‌کارگیری روش وارد، می‌توان هفت خوشه را به‌وضوح مشاهده کرد. در این دندروگرام فاصله بین سهم‌های همراه ۱ و اخابر ۱ کوتاه‌ترین در بین سایر سهم‌ها است. این امر نشان‌دهنده همبستگی قوی بین این دو سهم همگن از صنعت رادیویی است که اولین خوشه را تشکیل می‌دهند.



www.SID.ir

Archive of SID






**شکل ۵- دندروگرام سهم‌های شاخص ۳۰ شرکت بزرگ بورس اوراق بهادار تهران در سال ۱۳۹۳ با استفاده از روش وارد و معیار فاصله اقلیدسی**

این خوشه به همراه خوشه دوم با سهم‌های تاپیکو ۱ و پارسان ۱ از صنعت شیمیایی ، خوشه چهارم شامل پنج سهم وانصار ۱، وبملت ۱، وپاسار ۱، و بصادر ۱ و سهم وتجارت ۱ از صنعت مالی یا بانک‌ها و خوشه هفتم حاصل از پیوند دو سهم از صنعت خودرو که یک خوشه قوی را نشان می‌دهد، همگی خوشه‌های همگن و دارای همبستگی بالایی را تشکیل می‌دهند. خوشه سوم با پیوند دو سهم از صنعت شیمیایی و دو سهم از صنایع مالی و فلزات اساسی، خوشه پنجم شامل سهم‌های فعال در صنایع مختلف فلزات اساسی، مالی، شیمیایی و فرآورده نفتی و ششمین خوشه با دو سهم وبانک ۱ از صنایع چند رشته‌ای و سهم ومعادن ۱ از صنعت کانه فلزی، هر سه خوشه‌های ناهمگنی را نمایش می‌دهند که شاید پیوند سهم‌هایشان در یک خوشه به دلیل فعالیت‌های اقتصادی مشترک و تبادلات مالی و همچنین نقش تأمین‌کنندگی برخی از این شرکت‌ها برای مواد اولیه یا اعتبارات مالی موردنیاز دیگر شرکت‌های موجود در این خوشه‌ها باشد. در کل روش وارد توانسته است به‌وضوح سلسله مراتبی را ایجاد کند که در آن خوشه‌ها توانسته‌اند تعداد زیادی از سهم‌های موردمطالعه را دربر گیرند.

## ۵- نتیجه‌گیری و بحث

روش‌های خوشه‌بندی سلسله مراتبی به‌عنوان راهکاری نوین درزمینه‌ی بهبود روش‌های انتخاب سبد سهام در بازارهای مالی، می‌تواند سبب انتخاب



www.SID.ir

Archive of SID



اوراق بهادار مناسب برای سرمایه‌گذاری و تشکیل سبد بهینه سهام توسط سرمایه‌گذاران گردد. تجزیه‌وتحلیل این درخت‌ها می‌تواند به‌عنوان بررسی‌ای از این موضوع باشد که شرکت‌های مشغول به کار در یک صنعت یکسان، تمایل به تولید یک خوشه همسان دارند. دندروگرام حاصل از درختان سلسله مراتبی می‌توانند اطلاعات طبقه‌بندی‌شده‌ایی از شرکت‌های پذیرفته‌شده در بورس اوراق بهادار ارائه دهند و توزیع و اثرات ارتباطی بین خوشه‌های حاصل از سهم‌های مربوط به هر صنعت را آشکار سازد. هر روش خوشه‌بندی سلسله مراتبی دارای ویژگی‌های منحصربه‌فردی است که این ویژگی‌ها سبب خوشه‌بندی هوشمندانه‌ای از شرکت‌ها می‌شوند و نتایج مختلفی را به همراه دارند. در بسیاری از مطالعات، تجزیه‌وتحلیل مقایسه‌ای در خصوص انتخاب بهترین روش از میان روش‌های مختلف خوشه‌بندی سلسله مراتبی صورت نگرفته است؛ اما برخی از مطالعات روش‌های پیوند کامل، وارد، میانگین گروهی و همچنین به‌کارگیری معیار فاصله اقلیدسی به‌عنوان گزینه بسیار مناسب در هنگام ساخت درختان سلسله مراتبی را تائید کرده‌اند.

در این تحقیق و برای پاسخگویی به اولین سؤال پژوهشی مطرح‌شده، ما روش‌های خوشه‌بندی سلسله مراتبی تک پیوندی، پیوند کامل، میانگین گروهی، مرکزی و روش وارد را بر روی سهم‌های موردمطالعه اعمال کردیم و مشاهده کردیم که هریک از این روش‌ها توانسته‌اند تا ساختار خوشه‌بندی خاصی را با توجه به خصوصیات مربوط خود، نمایان سازند. در ساختارهای سلسله مراتبی به‌دست‌آمده در این پژوهش، می‌توان خوشه‌هایی از سهام را با توجه به نزدیکی و روابط اقتصادی‌شان شناسایی کرد. البته دراین‌بین برخی از ساختارهای ایجادشده استثنا هستند. همچنین در پاسخ به دومین سؤال مطرح‌شده در این پژوهش، از بین روش‌های خوشه‌بندی سلسله مراتبی به‌کاررفته در این پژوهش، دو روش وارد و پیوند کامل با تشکیل بیشترین تعداد خوشه و قرار دادن تعداد بیشتری از سهم‌ها در هر خوشه با توجه به میزان نزدیکی و صنعت مربوط به هر سهم، توانسته‌اند ساختار سلسله مراتبی متراکم و مناسبی از سهم‌های موردمطالعه را نمایش دهند. به دنبال آن روش خوشه‌بندی میانگین گروهی نیز ساختارهای سلسله مراتبی قابل تفسیری از سهم‌های موردنظر را از منظر حوزه فعالیت و همبستگی سهم‌ها ارائه می‌کند. آنچه در اینجا باید متذکر شد این است که با توجه به مشترک بودن برخی از عوامل اقتصادی تأثیرگذار بر روی بسیاری از شرکت‌ها، پیوند نزدیک بین سهم‌های مختلف در برخی از خوشه‌های ایجادشده دور از ذهن نیست.

در مورد روش تک پیوندی و مرکزی آنچه قابل‌ملاحظه است، شناسایی دو سهم خودرو ۱ و خساپا ۱ از صنعت خودرو توسط این روش‌ها به‌صورت یک خوشه مجزا و دورافتاده نسبت به سایر سهم‌ها است. آن‌طور که استنتاج می‌شود، به‌کارگیری این روش‌ها برای شناسایی نقاط دورافتاده می‌تواند در تشخیص خوشه‌های کلیدی و سهم‌های حیاتی در یک مجموعه از پرتفوی موردنظر به کار گرفته شوند. هدف از ساخت یک پرتفوی، کاهش ریسک و افزایش بازده سرمایه‌گذاری یا بالعکس از طریق متنوع سازی سهام است. بر این اساس سهم‌هایی برای تشکیل یک پرتفوی انتخاب می‌شوند که دارای کمترین میزان همبستگی با یکدیگر باشند. در ساخت درختان سلسله مراتبی سهم‌ها بر اساس میزان فاصله‌شان با یکدیگر در یک خوشه قرار می‌گیرند. در اصل سهم‌های پیوند خورده در هر خوشه مشابه هستند و دارای پویایی‌های مشترک‌اند و این باعث می‌شود تا انتخاب بیشتر از یک سهم از هر خوشه ایجادشده برای تشکیل یک پرتفوی بهینه از سهام مناسب نباشد. با توجه به بااهمیت بودن سهمی که بیشترین فاصله و در اصل کمترین میزان تشابه را با سایر سهم‌ها دارد، قرار دادن یک سهم از هر خوشه در کنار سهمی که دارای بیشترین فاصله در ساختار



www.SID.ir

Archive of SID



سلسله مراتبی است، می‌تواند در ایجاد یک پرتفوی بهینه مناسب باشد.

با توجه به آنچه در این پژوهش مطرح شد، سرمایه‌گذاران و فعالان در عرصه بورس اوراق بهادار می‌توانند برای مدیریت بهینه پرتفوی‌های سرمایه‌گذاری خود از نتایج به‌دست‌آمده توسط این روش‌ها به‌جای روش‌های سنتی متداول استفاده نمایند. روش‌های خوشه‌بندی به‌عنوان راهکاری نوین درزمینه‌ی بهبود روش‌های انتخاب سبد سهام در بازارهای مالی، می‌تواند سبب انتخاب اوراق بهادار مناسب برای سرمایه‌گذاری و تشکیل سبد بهینه سهام توسط سرمایه‌گذاران گردد. نتایج ساختار سلسله مراتبی به‌طور استاندارد در قالب یک دندروگرام گزارش می‌شود. این نمودار درختی مهم‌ترین اتصالات را به‌عنوان یک خروجی از تجزیه‌وتحلیل خوشه‌ای و یک ساختار سلسله مراتبی ممکن در تجزیه‌وتحلیل مجموعه داده نشان می‌دهد. اساساً منطق روش‌های خوشه‌بندی سلسله مراتبی ارائه یک نمای بصری از مشاهدات است و این نمایش گرافیکی مفید از خوشه‌بندی سلسله مراتبی یکی از دلایل اصلی برای محبوبیت این روش‌ها است و این جنبه بصری سازی جزء نقاط قوت این روش محسوب می‌شود؛ ولی طبیعی است که آماره و یا عدد کمی‌ای که مبنای تصمیم‌گیری باشد در این روش تعریف‌نشده است و تحلیل نتایج بر عهده کاربر است. بر همین اساس پیشنهاد می‌شود که در پژوهشی دیگری که امکان تعریف آماره وجود دارد، از روش‌هایی استفاده شود که متکی بر آماره و شاخص عددی باشند و نتایج آن‌ها با نتایج این پژوهش مقایسه گردد.

این مطالعه، چارچوبی را برای ایجاد یک راهنما برای انتخاب یک روش مناسب برای ایجاد یک پرتفوی بهینه از دارایی‌ها در بازار سهام ارائه کرده است؛ بااین‌وجود پیشنهاد می‌شود تا روش خوشه‌بندی سلسله مراتبی به سایر بازارهای مالی همچون بازار ارز، خودرو، کالا و غیره بسط داده شود و مناسب‌ترین روش‌های خوشه‌بندی سلسله مراتبی برای شناسایی چگونگی ارتباط بین سایر دارایی‌های مالی نیز اعمال شود.

## فهرست منابع


* Anderberg, M. R. (1973). Cluster analysis for applications. Monographs and textbooks on probability and mathematical statistics: Academic Press, Inc., New York.
* Arabie, P., Baier, N. D., Critchley, C. F., & Keynes, M. (2006). Studies in Classification, Data Analysis, and Knowledge Organization: springer.
* Bonanno, G., Lillo, F., & Mantegna, R. N. (2001). High-frequency cross-correlation in a set of stocks,Quantitative Finance, 1(1), 96-104.
* Bouguettaya, A., Yu, Q., Liu, X., Zhou, X., & Song, A. (2015).Efficient agglomerative hierarchical clustering. Expert Systems with Applications, 42(5), 2785-2797.
* Brida, J. G., Gómez, D. M., & Risso, W. A. (2009). Symbolic hierarchical analysis in currency markets: An application to contagion in currency crises. Expert Systems with Applications, 36(4), 7721-7728.
* Brida, J. G., & Risso, W. A. (2010). Dynamics and structure of the 30 largest North American companies. Computational Economics, 35(1), 85-99.
* Brida, J. G., & Risso, W. A. (2007). Dynamics and structure of the main Italian companies. International Journal of Modern Physics C, 18(11), 1783-1793.
* Cha, S.-H. (2008). Taxonomy of nominal type histogram distance measures. City, 1(2), 1.
* Gan, G., Ma, C., & Wu, J. (2007). Data clustering: theory, algorithms, and applications (Vol. 20): Siam.
* Garas, A., & Argyrakis, P. (2007). Correlation study of the Athens stock exchange. Physica A: Statistical Mechanics and its Applications, 380, 399-410.
* Han, J., & Kamber, M. (2001). Data mining: concepts and techniques: Morgan Kaufmann San Francisco, Calif, USA.
* Hastie, T. J., Tibshirani, R. J., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction: Springer.




www.SID.ir

Archive of SID



Correction of February 2007: How financial hierarchies change in a market crash. Physica A: Statistical Mechanics and its Applications, 424, 225-241.

* Ulusoy, T., Keskin, M., Shirvani, A., Deviren, B., Kantar, E., & Dönmez, C. Ç. (2012). Complexity of major UK companies between 2006 and 2010: Hierarchical structure method approach. Physica A: Statistical Mechanics and its Applications, 391(21), 5121-5131.
* Wang, G.-J., & Xie, C. (2015). Correlation structure and dynamics of international real estate securities markets: A network perspective. Physica A: Statistical Mechanics and its Applications, 424, 176-193.
* Yang, R., Li, X., & Zhang, T. (2014). Analysis of linkage effects among industry sectors in China's stock market before and after the financial crisis. Physica A: Statistical Mechanics and its Applications, 411, 12-20.
* Zhang, D., & Zhou, L. (2004). Discovering golden nuggets: data mining in financial application. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 34(4), 513-522

**یادداشت‌ها**

[1]. Hierarchical Clustering
[2]. Dendrogram
[3]. Agglomerative
[4]. Divisive
[5]. Single-Linkage Algorithm
[6]. Complete-Linkage Algorithm
[7]. Group Average Method
[8]. Centroid Method
[9]. Wards Method
[10]. Mantegna
[11]. Bonanno
[12]. Lillo
[13]. Garas
[14]. Argyrakis
[15]. Brida
[16]. Risso
[17]. Tabak
[18]. Serra
[19]. Cajueiro
[20]. Ten
[21]. Mizuno
[22]. Takayasu
[23]. Gomes
[24]. Kocakaplan
[25]. Deviren
[26]. Keskin
[27]. Euclidean Distance

* Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review.ACM computing surveys (CSUR), 31(3), 264-323.
* Kocakaplan, Y., Deviren, B., & Keskin, M. (2012). Hierarchical structures of correlations networks among Turkey's exports and imports by currencies. Physica A: Statistical Mechanics and its Applications, 391(24), 6509-6518.
* Kocakaplan, Y., Doğan, Ş., Deviren, B., & Keskin, M. (2013). Correlations, hierarchies and networks of the world's automotive companies. Physica A: Statistical Mechanics and its Applications, 392(12), 2736-2774.
* Kogan, J., Nicholas, C., & Teboulle, M. (2006). Grouping multidimensional data: Springer.
* Lai, K. K., Yu, L., Wang, S., & Zhou, C. (2006). A double-stage genetic optimization algorithm for portfolio selection. Paper presented at the Neural Information Processing.
* Mantegna, R. N. (1999). Hierarchical structure in financial markets. The European Physical Journal B-Condensed Matter and Complex Systems, 11(1), 193-197.
* Markowitz, H. (1952). Portfolio selection. The journal of finance, 7(1), 77-91.
* Mizuno, T., Takayasu, H., & Takayasu, M. (2006). Correlation networks among currencies. Physica A: Statistical Mechanics and its Applications, 364, 336-342.
* Paulus, M., & Kristoufek, L. (2015). Worldwide clustering of the corruption perception. Physica A: Statistical Mechanics and its Applications, 428, 351-358.
* Rea, A., & Rea, W. (2014). Visualization of a stock market correlation matrix. Physica A: Statistical Mechanics and its Applications, 400, 109-123.
* Simon, H. A. (1991). The architecture of complexity: Springer.
* Tabak, B. M., Serra, T. R., & Cajueiro, D. O. (2010). Topological properties of stock market networks: The case of Brazil. Physica A: Statistical Mechanics and its Applications, 389(16), 3240-3249.
* Teh, B. K., Goo, Y. W., Lian, T. W., Ong, W. G., Choi, W. T., Damodaran, M., & Cheong, S. A. (2015). The Chinese



www.SID.ir

Archive of SID

# Codification of Dendrograms Portfolio Based on Euclidean Distance Measure
## (A Comparison Between Different Methods of Hierarchical Clustering)

**Hojatollah Sadeghi**[1]
**Sharifeh Forooghi Dehnavi**[2]

Receipt: 2016 / 4 / 8 Acceptance: 2016 / 7 / 6

**Abstract**

Today analysis of financial markets as a part of the capital market and its impact on development and portfolio design and investment strategy of each country has become an important and most critical issue. The aim of this study was to investigate how the connection and distribution of stocks related to 30 large companies index of Tehran Stock Exchange and the effects of relationship between clusters of related stocks to every industry. In this study, using a variety of methods of hierarchical clustering, structure, classification and hierarchy of the stocks in the year 1393 reviewed. The results showed that With a focus on each of the hierarchical clustering methods and their implementation on the target stocks, were identified different clusters of stocks due to the similarity and economic relationships and also the key clusters and the vital stocks in the desired set were obtained. The results indicate that the choice best hierarchical clustering algorithm for clustering stocks depends on the desired purpose of cluster analysis and consideration of the advantages and disadvantages of each method.

**Keywords:** Dendrogram, Euclidean Distance Measure, Hierarchical Clustering Analysis, Portfolio

1- Prof Department of Accounting& Finance Faculty of Economic, Management and Accounting, Yazd University, Yazd, Iran
sadeqi@yazd.ac.ir
2- Masters Business Management (finance), Faculty of Economic, Management and Accounting, Yazd University, Yazd, Iran
forooghi_sh@yahoo.com Corresponding Author



www.SID.ir