چشم‌انداز مدیریت مالی
شماره ۱۱ ـ پاییز ۱۳۹۴
ص‌ص ۱۵۸ ـ ۱۳۳

مدلسازی هم‌حرکتی سهام در بورس اوراق بهادار تهران با استفاده از رویکرد خوشه‌بندی سه‌مرحله‌ای

محمد اقبال‌نیا*، احمد پویان‌فر، ملیحه مالکی*****

چکیده

خوشه‌بندی به‌عنوان یکی از روش‌های داده‌کاوی توصیفی، روشی برای گروه‌بندی مشاهده‌ها به k خوشه (گروه) مختلف است. خوشه‌بندی بازار سهام، اطلاعات سودمندی را جهت پیش‌بینی تغییرات قیمت سهام شرکت‌های مختلف در اختیار سرمایه‌گذاران شخصی و متخصصان مالی قرار می‌دهد. در سال‌های اخیر، خوشه‌بندی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار براساس شباهت در روند حرکتی یا به‌عبارتی شکل بازار آن‌ها، مورد توجه قرار گرفته است. رویکردهای موجود خوشه‌بندی، نتایج خوبی برای این روند ارائه نمی‌دهند؛ زیرا داده‌های مربوط به قیمت سهام شرکت‌ها داده‌های سری‌زمانی مالی هستند که دارای ابعاد زیاد است و تغییرات در این داده‌ها، معمولا با شیفت‌های زمانی همراه است. این ویژگی‌ها، خوشه‌بندی شرکت‌ها براساس روند حرکتی قیمت سهام آن‌ها را بسیار پیچیده کرده است. از طرفی در بین روش‌های موجود، روشی برای خوشه‌بندی بازار سهام که بتواند شرکت‌ها را به زیرخوشه‌هایی که حاوی اطلاعات مفید برای استفاده‌کننده نهایی است، تقسیم کند وجود ندارد. در تحقیق حاضر، روش خوشه‌بندی سه‌مرحله‌ای برای خوشه‌بندی شرکت‌های موجود در بورس اوراق بهادار، براساس شباهت در روند حرکتی قیمت سهام شرکت‌ها، مورد استفاده قرار گرفته است. در مرحله اول، کاهش ابعاد داده‌ها صورت گرفته است و خوشه‌بندی تقریبی شرکت‌ها انجام می‌شود. سپس در مرحله دوم، خوشه‌های ایجاد شده در مرحله قبل، به زیرخوشه‌های با کیفیت و دقت بیشتر تقسیم می‌شوند. در نهایت، زیرخوشه‌های ایجاد شده، در مرحله سوم ادغام و خوشه‌های نهایی ارائه می‌شود. مدل ارائه شده، با استفاده از روش مجموع مربعات خطا مورد ارزیابی قرار می‌گیرد. نتایج تحقیق نشان می‌دهد روش خوشه‌بندی سه مرحله‌ای در مقایسه با سایر الگوریتم‌های خوشه‌بندی مرسوم از نظر اثربخشی و کیفیت، عملکرد بهتری دارد.

کلیدواژه‌ها: بازار بورس اوراق بهادار؛ داده‌کاوی؛ هم‌حرکتی؛ خوشه‌بندی سه‌مرحله‌ای؛ سری-زمانی.

تاریخ دریافت مقاله: ۱۳۹۴/۰۹/۲۵، تاریخ پذیرش مقاله: ۱۳۹۵/۰۲/۰۶

* استادیار گروه مدیریت مالی و مهندسی مالی دانشکده علوم مالی، عضو هیات علمی دانشگاه خوارزمی.
** دکترای مدیریت مالی، دانشگاه تهران.
*** کارشناسی ارشد مهندسی مالی، موسسه آموزش عالی غیرانتفاعی خاتم (نویسنده مسئول).
Email: S_maleki @yahoo.com

۱. مقدمه

بازار سرمایه از ارکان مهم اقتصادی است که به جمع‌آوری منابع و انباشت پس‌اندازهای جامعه می‌پردازد. در این بازار منابع مالی شرکت‌های سهامی تامین شده است و امکانات لازم برای تخصیص پس‌انداز به فرصت‌های سرمایه‌گذاری فراهم می‌شود. در نظریه‌های اقتصاد کلان، سرمایه‌گذاری یکی از شاخص‌های مهم محسوب می‌شود و نقش تعیین‌کننده‌ای را در رشد و توسعه اقتصادی ایفا می‌کند. از این‌رو، به‌وجود آوردن شرایطی که سرمایه‌گذاران بتوانند از میان گزینه‌های مختلف سرمایه‌گذاری، سهام شرکت‌ها را برگزینند از اهمیت ویژه برخوردار است. در نظریه‌های اقتصاد خرد نیز اهمیت سرمایه‌گذاری ناشی از این مسئله است که در حقیقت فرد سرمایه‌گذار مصرف امروز را به امید مصرف بیشتر به آینده موکول می‌کند؛ بنابراین سرمایه‌گذاران درصدد هستند تا مناسب‌ترین سهام را برای سرمایه‌گذاری شناسایی و منابع سرمایه‌ای خود را به آن سوی هدایت کنند. از منظر معامله‌گران بازار، به‌دست آوردن هرگونه اطلاعات در تحلیل بازار از ارزش زیادی برخوردار است؛ چرا که داشتن اطلاعات می‌تواند ریسک سرمایه‌گذاری را کاهش دهد و حتی باعث سودهای بزرگ شود. در همین راستا، دانش معامله‌گران درباره حرکت چند سهم با هم، می‌تواند نقش مهمی در انتخاب سبد سهام افراد داشته باشد. هم‌حرکتی[۱] بین قیمت سهام‌شرکت‌ها پدیده‌ای شایسته است که از زمان پیدایش اولین نظریه‌های بازار مورد توجه بوده است [۷]. به علاوه، با بررسی روند حرکتی قیمت‌ها در یک گروه، می‌توان قیمت سهام شرکت‌ها را پیش‌بینی کرد. همچنین بررسی هم‌حرکتی و خوشه‌بندی شرکت‌ها، راهکار جدیدی را برای تهیه و ارائه شاخص‌های لازم برای ارزیابی وضعیت بازار در اختیار سازمان بورس اوراق بهادار قرار می‌دهد.

حجم بزرگ داده‌ها به تنهایی به مدیران سازمان‌ها در تصمیم‌گیری کمکی نمی‌کند؛ بلکه باعث سردرگمی مدیران نیز می‌شود؛ بنابراین مدیریت داده‌های خام و تبدیل داده‌های خارجی و داخلی سازمان به اطلاعات و دانش با استفاده از روش‌های گوناگون، نقش اساسی و محوری دارد. یکی از ابتدایی‌ترین روش‌ها برای تحلیل مسائل مختلف، استفاده از الگوها و بررسی آن‌ها است، این سوال برای ما پیش خواهد آمد که چه چیزی بین داده‌ها مشترک است. روش‌های مختلفی برای بررسی هم‌حرکتی بین داده‌ها وجود دارد. از روش‌های معروف در این زمینه داده – کاوی[۲] است که می‌تواند بر بانک اطلاعاتی انجام شود و دانش مورد نیاز را به‌دست آورد. داده-کاوی، هنر و علم آنالیز هوشمندانه داده‌ها است و هدف از آن پیدا کردن بینش و دانش نسبت به داده‌های پژوهش است. این کنکاش گاهی به فرایند ساختن مدل نیز تعبیر می‌شود و مدل ساخته

---

۱. Co-movement

۲. Data mining

شده برای فهم بیشتر داده‌ها و پیش بینی وقایع استفاده می‌گردد. با توجه به افزایش روزافزون داده‌ها، داده‌کاوی در اکثر حوزه‌های پژوهش از جمله مدیریت، اقتصاد، مهندسی و سایر رشته‌ها استفاده شده است [۳]. زمانی‌که می‌خواهیم به‌بررسی تاثیر تعداد کمی از عوامل بر هدف بپردازیم معمولا روش‌های آماری مناسب است؛ ولی زمانی‌که تعداد این عوامل زیاد می‌شود دیگر این روش‌ها کارایی مناسبی ندارند. این امر سبب شد که روش‌های ابتکاری دیگری علاوه بر روش‌های آماری مانند شبکه‌های عصبی[۱] و الگوریتم ژنتیک[۲] برای این منظور ایجاد شود. از داده‌کاوی می‌توان برای انجام کارهایی مانند طبقه‌بندی[۳]، پیش‌بینی[۴]، تخمین[۵] و خوشه‌بندی[۶] داده‌ها استفاده کرد. برای انجام این کارها روش‌هایی توسعه یافته‌اند که برخی از آن‌ها عبارت‌اند از: الگوریتم‌های خوشه‌بندی[۷]، شبکه‌های عصبی، الگوریتم ژنتیک، نزدیک‌ترین همسایگی[۸] و درخت تصمیم‌گیری[۹] [۱].

همان‌طور که بیان شد، یکی از روش‌های متداول برای تحلیل داده‌ها، الگوریتم‌های خوشه‌بندی است. خوشه‌بندی ما را قادر می‌سازد تا به جای پردازش حجم انبوه اطلاعات، تنها اطلاعاتی را بررسی و تحلیل کنیم که بسیار به هم شبیه هستند و این خود یک گام بلند برای ساده کردن مسئله است [۲]. خوشه‌بندی یکی از روش‌های بسیار قدرتمند برای کشف گروه‌ها و وابستگی‌های طبیعی در یک مجموعه داده و همچنین شناخت الگوهای ساختاری موجود در آن، بدون داشتن هرگونه پیش‌زمینه در مورد ویژگی‌های داده، است. در روش‌های خوشه‌بندی داده‌ها به چند گروه تقسیم می‌شوند؛ به‌طوری‌که شباهت بین داده‌های درون هر خوشه حداکثر باشد [۴]. هدف اصلی از انجام این تحقیق، دسته‌بندی شرکت‌های موجود در بورس ایران با استفاده از روش خوشه‌بندی سه‌مرحله‌ای است. خوشه‌بندی شرکت‌ها براساس داده‌های سری زمانی مربوط به آن‌ها باعث ارزیابی هم حرکتی سهام شرکت‌ها خواهد شد. در نتیجه شرکت‌هایی که روند حرکتی یکسان و یا به عبارتی شکل بازار یکسان دارند در یک خوشه قرار می‌گیرند. این دسته‌بندی‌ها، با بیان این‌که چه شرکت‌هایی روند حرکتی مشابه دارند، اطلاعات مفیدی برای بررسی و تحلیل بازار در اختیار سرمایه‌گذاران قرار می‌دهند. در واقع، خوشه‌بندی شرکت‌ها راهکاری مناسب برای پیش‌بینی قیمت سهام شرکت‌ها با توجه به روند حرکتی گذشته و یا روند

---

۱. Neural Network
۲. Genetic Algorithm
۳. Classification
۴. Prediction
۵. Estimation
۶. Clustering
۷. Cluster Detection Algorithm
۸. Nearest Neighboring
۹. Decision Tree

حرکتی قیمت سایر شرکت‌های هم گروه آن ایجاد می‌کند. به‌عبارتی، سوال اصلی تحقیق حاضر آن است که چگونه روش خوشه‌بندی را می‌توان در بررسی هم‌حرکتی قیمت سهام شرکت‌ها استفاده کرد و مزایای این روش نسبت به سایر روش‌های مرسوم مورد استفاده در بررسی هم‌حرکتی قیمت سهام چیست.

۲. مبانی و چارچوب نظری تحقیق

امروزه با گسترش سیستم‌های پایگاهی و حجم بالای داده‌های ذخیره شده، نیاز به ابزاری است تا بتوان داده‌های ذخیره شده را پردازش کرد. داده‌کاوی یکی از مهم‌ترین روش‌هایی است که با بهره‌گیری از ابزارهای آنالیز داده‌های پیچیده، به کشف الگوهای مفید موجود میان داده‌ها اشاره دارد. خوشه‌بندی به‌عنوان یکی از روش‌های داده‌کاوی توصیفی، به‌دنبال سازماندهی مجموعه‌ای از داده‌ها با استفاده از اندازه‌گیری فاصله یا انحراف، در یک سری خوشه است به‌طوری‌که داده‌ها در هر خوشه، بالاترین درجه شباهت را دارند و داده‌های متعلق به خوشه‌های مختلف دارای حداکثر درجه تفاوت هستند [۱۷].

مطالعات زیادی در رابطه با آنالیز بازار سهام و پیش‌بینی روند بازار انجام گرفته است. ناندا، ماهانتی و تیوری (۲۰۱۰)، [۲۶] با استفاده از محاسبه بازدهی سهام شرکت‌ها در زمان‌های مختلف، خوشه‌بندی شرکت‌ها را در بورس انجام دادند. مسیح و مسیح [۲۴] و آنتونیو [۱۱] گزارش کردند که روند حرکتی بازار سهام در یک کشور متاثر از روند حرکتی سهام در کشورهای دیگر است. مطالعات دیگری نیز درباره هم‌حرکتی بازارهای کشورهای مختلف انجام گرفته است. کولینز و بیکپ [۱۴] در تحقیق خود نشان دادند رویدادهای بازارهای مالی در حال ظهور در طی دهه گذشته و وجود بحران‌های مالی در کشورهای مختلف منجر به بحث درباره یکپارچگی جهانی بازار سرمایه شده است. آن‌ها با بررسی وابستگی بازارهای سهام آفریقا به نتیجه مشابه دست یافتند. لیاو و چو [۲۰] با به‌کارگیری الگوریتم‌های خوشه‌بندی به‌بررسی هم‌حرکتی بازار سهام تایوان و هنگ‌کنگ پرداختند، آن‌ها برای نشان دادن هم‌حرکتی بازار سهام شرکت‌ها شاخص بازار را در ۳۰ گروه طبقه‌بندی کردند. دیویت و میهوو [۱۶] نیز با استفاده از شاخص بازار ۵۸ کشور و محاسبه همبستگی دو به دوی آن‌ها به این نتیجه رسیدند که کشورهایی با صنایع مشابه، روند حرکتی بازار آن‌ها مشابه هم خواهد بود. این تحقیقات بیشتر برای بررسی هم‌حرکتی در بین جوامع و کشورهای مختلف انجام گرفته و به ندرت به‌بررسی هم‌حرکتی خوشه‌بندی بازار سهام در یک کشور پرداخته‌است. کوو و کستی [۱۸] بیان کردند که همبستگی زیاد بین عوامل تاثیرگذار بر داده‌های سری زمانی مورد بررسی، وجود اختلالات[1] تاثیرگذار بر داده‌ها و وجود

1. Noise

داده‌هایی با ابعاد زیاد، خوشه‌بندی داده‌ها را با مشکل مواجه می‌کند. پس مطالعات بیشتری برای یافتن راه‌کارهای مناسب برای خوشه‌بندی سری‌های زمانی با این ویژگی‌ها انجام گرفت.
راتانوماها [۲۷] و باگنال و جاناکک [۱۲] بیان کردند اندازه‌گیری فاصله، با مقیاس فاصله اقلیدسی (ED)[1] ممکن است همیشه درست نباشد؛ چرا که ED فقط برای اندازه‌گیری شباهت در فواصل زمانی یکسان درست عمل می‌کند. لین، کوو، لوناردی و چیو [۲۱] نیز گزارش کرد الگوریتم‌های داده‌کاوی کلاسیک برای داده‌های سری زمانی به‌خوبی عمل نمی‌کند. لیو و شاوو [۲۳] روش SAX[2] را برای اندازه‌گیری فاصله شباهت بین داده‌های سری زمانی مالی توسعه دادند. چو و کوو و هارت و پازانی [۱۵] بیان کردند روش DTW[3] به‌عنوان مقیاس اندازه‌گیری فاصله شباهت بین داده‌های سری زمانی مالی، شباهت در شکل بازار بین داده‌ها را نیز نشان می‌دهد. آقابزرگی و وحده [۱۰] با توجه به مطالعات انجام‌شده به‌بررسی هم‌حرکتی قیمت بازار سهام مالزی با استفاده از روش خوشه‌بندی سه‌مرحله‌ای پرداختند. آن‌ها در این تحقیق ۸۷۰ شرکت از صنایع مختلف را در طول سال ۲۰۱۰ مورد مطالعه قرار دادند و به این نتیجه رسیدند که روش خوشه‌بندی سه‌مرحله‌ای نسبت به سایر روش‌های متداول کاراتر است و نتیجه بهتری ارائه می‌دهد.

در داخل کشور نیز تحقیقاتی پیرامون هم‌حرکتی بازار سهام و گروه‌بندی شرکت‌های بورسی انجام شده‌است. عبدی [۹] به‌بررسی هم‌حرکتی بازارهای سهام تهران و دبی پرداخته است. در این تحقیق بیان می‌شود که یکی از نشانه‌های ارتباط میان بازارهای مالی، وجود سرایت میان این بازارها است. نتایج نشان‌دهنده عدم وجود سرایت در بازدهی میان بازارهای سهام تهران و دبی است. قالیباف اصل، محمدی و مظاهری [۸] با بررسی بازدهی ۴۸ شرکت فعال در بورس در طی مدت ۵ سال، هم‌حرکتی بازده سهام شرکت‌ها را براساس مدل APT و با روش عاملی مورد مطالعه قرار دادند. نتایج تحقیق نشان داد که دلیل هم‌حرکتی بازده سهام، تاثیر متفاوت عوامل کلان اقتصادی بر بازده سهام شرکت‌ها است. فرید و پورحمیدی [۵] به بخش‌بندی سهام شرکت‌های پذیرفته‌شده در بورس اوراق بهادار با استفاده از روش خوشه‌بندی فازی پرداختند. آن‌ها بیان کردند که انتخاب پرتفوی مناسبی از سهام، مستلزم استفاده از روش‌های متفاوت سرمایه‌گذاری است و هر دو نوع سهام ارزشی و رشدی باید در پرتفوی قرار گیرد تا علاوه بر کاهش ریسک، نتایج بهتری نصیب سرمایه‌گذار شود. در واقع آن‌ها طبقه‌بندی مناسبی از سهام رشدی، ارزشی و ترکیبی ارائه کرده است و بیان کردند بخش اعظم شرکت‌ها در سبد سهام ترکیبی قرار می‌گیرند؛ لیکن گرایش آن‌ها به سبد سهام رشدی است. فرید [۶] با استفاده از

1. Euclidean distance
2. Symbolic aggregate approximation
3. Dynamic time warping

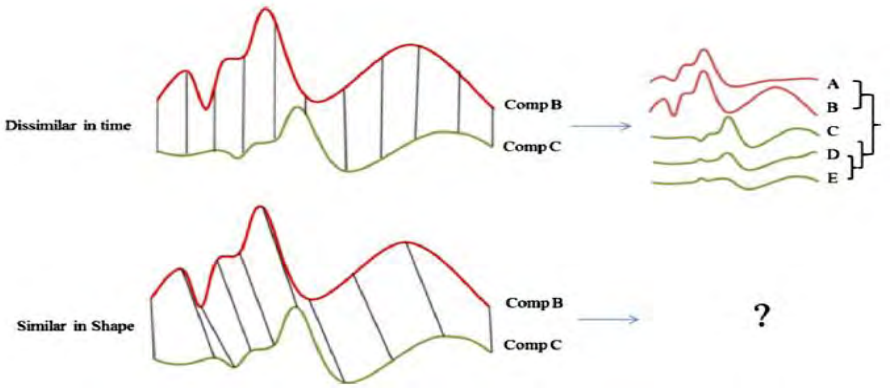
شبکه‌های عصبی خودسازنده به‌بررسی گروه‌بندی شرکت‌های سهامی فعال در بورس تهران پرداخت. وی چنین نتیجه‌گیری کرد که از میان ۱۷ متغیر مالی ۴ متغیر نسبت آنی، بازده کل دارایی‌ها، حجم معاملات و نسبت بدهی به حقوق صاحبان سهام مناسب‌ترین متغیرها در تفکیک سهام است که ۱۴۲ سهم مورد بررسی را به ۵ گروه متمایز تفکیک کرده‌اند. قاضی کلهرودی [۷] هم‌حرکتی بین قیمت‌های سهام را با استفاده از نظریه پیچیدگی در بازار سهام تهران و با توجه به شاخص هم‌حرکتی مورد ارزیابی قرار داد. مهمترین نتیجه‌ای که از این بررسی به‌دست آمد این است که یادگیری افراد که نقش تعیین‌کننده‌ای در کارایی بازار دارد، خود عامل وجود هم‌حرکتی در بین قیمت‌های سهام است.

۳. روش‌شناسی تحقیق

سری‌های زمانی مالی نسبت به سایر سری‌های زمانی، مشخصه‌های خاصی دارند. از جمله این ویژگی‌ها این است که این دسته از سری‌های زمانی با نقاط بحرانی کمی نشان داده می‌شوند، به عبارتی داده‌های مالی در هر زمان، داده‌هایی بحرانی و مهم هستند و به دقت بسیار زیادی در بررسی‌های کوتاه‌مدت و بلندمدت نیاز دارند. دومین ویژگی خاص سری زمانی مالی، ادامه‌دار بودن، بزرگ بودن و بدون مرز بودن این داده‌ها است. همان‌طور که بیان شد، الگوریتم‌های متفاوتی برای خوشه‌بندی داده‌ها وجود دارد که خوشه‌هایی با کیفیت و دقت بنیادی قابل قبول ایجاد می‌کنند و در مقطع زمانی برای داده‌های ایستا کارا هستند. تحقیقات نشان داده است که الگوریتم‌های داده‌کاوی کلاسیک، برای خوشه‌بندی سری‌های زمانی به‌خوبی عمل نمی‌کند؛ چرا که ابعاد داده‌ها و همبستگی بین داده‌ها زیاد است، از طرفی تعداد زیادی از اختلالات (نویزها) که در داده‌های سری زمانی وجود دارد، خوشه‌بندی این داده‌ها را با مشکل مواجه می‌کند.

یکی از مشکلات اساسی در داده‌کاوی سری‌های زمانی مشکل جست‌وجوی شباهت یا اندازه‌گیری فاصله است. زمانی‌که تعداد داده‌ها کم است از طریق مقایسه دو به دو آن‌ها می‌توان شباهت بین داده‌ها را بررسی کرد؛ اما زمانی‌که تعداد داده‌ها بسیار زیاد است، یا داده‌ها ابعاد بزرگی دارند این روش موثر نخواهد بود. روش‌های مختلفی برای اندازه‌گیری فاصله بین داده‌های سری زمانی وجود دارد که در این میان، روش ED و DTW روش‌های متداول در خوشه‌بندی سری‌های زمانی هستند. اگر چه اندازه‌گیری فاصله شباهت بین داده‌ها با استفاده از فاصله اقلیدسی (ED) یک روش اندازه‌گیری مناسب در بسیاری از روش‌های خوشه‌بندی موجود است؛ ولیکن دقت کافی را در بررسی هم‌حرکتی بین قیمت سهام شرکت‌ها ندارد؛ زیرا این روش فقط تشابه زمانی بین داده‌ها را بیان می‌کند. این در حالی است که قیمت پایانی سهام شرکت‌ها

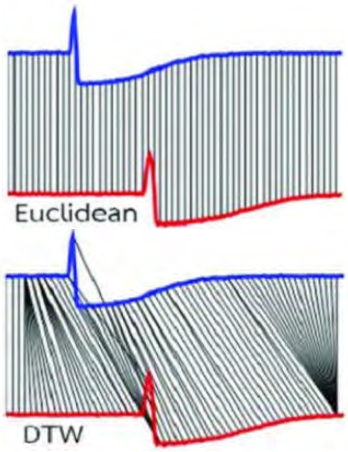
دارای شیفت‌های زمانی است و تشابه در روند حرکتی در طی زمان نیز باید مورد توجه قرار گیرد. شکل زیر روند حرکتی دو شرکت C وB را نشان می‌دهد.



شکل ۱. بررسی روند حرکتی دو سری زمانی

بر اساس شکل ۱، دو شرکت C و B در روش‌های معمول خوشه‌بندی با در نظر گرفتن داده‌ها در هر لحظه از زمان، ممکن است در دو خوشه متفاوت قرار گیرند و یا هیچ ارتباطی بین روند حرکتی آن‌ها مشاهده نشود؛ در حالی‌که نگاه دقیق‌تر در بررسی روند حرکتی نشان می‌دهد دو شرکت C و B با مقداری تاخیر با شیفت زمانی روند حرکتی مشابه دارند، یا به عبارتی بین دو سری زمانی هم‌حرکتی وجود دارد؛ بنابراین خوشه‌بندی این داده‌ها براساس تشابه در یک نقطه از زمان ممکن است نتایج درستی ارائه ندهد.

مزیت روش DTW نسبت به روش ED این است که DTW قابلیت در نظر گرفتن شیفت‌های زمانی و بررسی روند حرکتی داده‌ها را در طی زمان دارا است. شکل ۲، تفاوت بین دو روش اندازه‌گیری ED و DTW را به‌خوبی نشان می‌دهد [۲۸]. همان‌طور که مشخص است، روش ED در هر لحظه از زمان، برای بررسی هم‌حرکتی بین داده‌های دو سری زمانی فقط نقطه همتراز زمانی آن را در نظر می‌گیرد؛ اما در روش DTW هر داده از سری زمانی با چندین داده سری زمانی دیگر مورد مقایسه قرار گرفته است و در نهایت داده‌ای استفاده می‌شود که دارای کمترین فاصله با داده مورد بررسی باشد. به عبارتی روش DTW مشکل شیفت‌های محلی سری‌های زمانی را نیز در نظر گرفته و به دنبال یافتن سری‌هایی است که اگر چه در مقطع زمانی هم حرکتی را نشان نمی‌دهند؛ اما روند حرکتی آن‌ها در طی زمان با وجود شیفت‌هایی مشابه هم است [۱۹]. .



شکل ۲. بررسی دو روش اندازه گیری فاصله ED DTW

روش خوشه‌بندی سه‌مرحله‌ای شیفت‌های زمانی موجود را نیز در نظر گرفته و می‌تواند خوشه‌هایی را بر مبنای تشابه در روند حرکتی سری‌های زمانی ایجاد کند. به عبارتی روش خوشه‌بندی سه‌مرحله‌ای علاوه بر در نظر گرفتن داده‌ها در واحد زمانی، داده‌ها را در طول زمان با هم مقایسه می‌کند و نتیجه بهتری را ارائه می‌دهد [۱۰]. خوشه‌بندی سه‌مرحله‌ای(PTC) همان‌طور که از نام آن پیدا است در سه مرحله انجام می‌شود:

مرحله اول: کاهش ابعاد داده‌ها و خوشه‌بندی تقریبی. لین و کوو (۲۰۰۳)، روش جدیدی بر پایه روش PAA[۲] به نام SAX[۳] ارائه کردند. SAX اولین روش نمادین ارائه‌شده برای کاهش ابعاد سری زمانی است که با تابع فاصله تقریبی اقلیدسی تعریف می‌شود. SAX بر پایه این فرض که داده‌ها توزیع نرمال گوسین دارند استوار است؛ بنابراین داده‌ها باید ابتدا نرمال شوند. بر مبنای روش PAA، بعد از محاسبه مقدار میانگین برای هر بخش، SAX با تعریف نقاط شکست (β_i) به داده‌های هر بخش با توجه به جداول، نمادی نسبت داده و فاصله را بر مبنای این نمادهای تعریف‌شده محاسبه می‌کند. به عبارتی، داده‌های گزارش‌شده از روش PAA که در واقع میانگین هر بخش است، گسسته می‌شود. این گسست با تعریف تعداد و محل نقاط شکست $(\beta = \beta_1, \beta_2, \dots, \beta_{a-1})$ ایجاد می‌شود. $\widehat{F_i} = \left(\widehat{f_1}, \widehat{f_2}, \dots \widehat{f_W}\right)$ داده‌های نمادگذاری شده به روش SAX هستند.

---

۱. ۳-Phase Time series Clustering

۲. Piecewise aggregate approximation

۳. Symbolic Aggregate Approximation

لین و کوو و وی و لوناردی [۲۲] بیان می‌کنند که روش اندازه‌گیری حداقل فاصله[۱] (MINDIST)، یک روش مناسب برای محاسبه فاصله برای روش SAX است. MINDIST فاصله بین داده‌ها را براساس کمترین فاصله بین آن‌ها محاسبه می‌کند. مشکل روش MINDIST این است که فاصله بین شاخص محل‌های مجاور (میانگین داده‌های هر بخش) را صفر در نظر گرفته و همچنین از در نظر گرفتن نقاط ماکزیمم و مینیموم داده‌ها صرف‌نظر می‌کند. روش APXDIST، اصلاح شده روش MINDIST است که این مشکل را برطرف می‌کند. به این صورت که در این روش به ازای هر ناحیه، یک شاخص بر مبنای نقاط شکست به‌دست آمده، تعریف می‌شود. به عبارتی هر شاخص میانگین حسابی نقاط شکست ابتدا و انتهای ناحیه مذکور است.

$$0 < i < a \qquad Ind_i = \frac{\beta_{i-1}+\beta_i}{2} \qquad \text{رابطه (۱)}$$

که در آن β_i و β_{i-1} نقاط ماکزیموم و مینیموم داده‌ها است. β_0 مینیموم کلی و β_a ماکزیموم کلی به ازای داده‌های سری زمانی است. سپس فاصله بین شاخص‌ها را براساس معادله (۲) محاسبه و خوشه‌بندی سری زمانی انجام می‌شود [۲۵].

$$dis_{APXDIST} = \sqrt{\frac{n}{W}}\sqrt{\sum_{i=1}^{W}(dis(lnd_i, lnd_j))^2} \qquad \text{رابطه (۲)}$$

که در آن w تعداد بخش‌ها، n تعداد کل داده‌ها و $dis(lnd_i, lnd_j)$ فاصله بین دو شاخص lnd_i و lnd_j است. بعد از به‌دست آوردن ماتریس فاصله بین شرکت‌ها، برای خوشه‌بندی داده‌ها در انتهای مرحله اول، از روش K-Medoids که از گروه الگوریتم‌های K-Means است، استفاده خواهد شد.

مرحله دوم: تعیین زیر خوشه‌ها

هدف این مرحله اصلاح خوشه‌های ایجاد شده در مرحله قبل و بهبود کیفیت آن است. در این مرحله، تعداد مجموعه داده‌ها با تعریف نمونه در هر گروه کاهش می‌یابد. مقیاس فاصله در این مرحله فاصله اقلیدسی (ED) است.

۱. MINDIST

$$dis_{ED}(F_i, F_j) = \sqrt{\sum_{i=1}^{n}(f_i - f_j)^2}$$ رابطه (۳)

که F_j و F_i دو داده از خوشه‌بندی مقدماتی مرحله قبل هستند. به عبارتی در این مرحله خوشه‌های ایجاد شده در مرحله اول بازبینی و زیرخوشه‌ها تشکیل می‌شود. زیرخوشه‌ها (SC_{ij}) در واقع داده‌هایی هستند که روند حرکتی آن‌ها در یک بازه زمانی مشابه هم است که از گسست خوشه‌های مرحله قبل ایجاد می‌شوند.

خوشه اولیه (PC_i) را در نظر بگیرید، ماتریس فاصله بین داده‌های موجود در این خوشه محاسبه شده و به‌صورت یک ماتریس M_{n*n} نمایش داده می‌شود، به‌طوری‌که M_{ij}، فاصله بین داده i و j را نشان می‌دهد. الگوریتم توسعه یافته PCS[1] با استفاده از ماتریس فاصله تعریف شده خوشه‌های اولیه را از هم جدا می‌کند. هر خوشه در مرحله دوم بر مبنای مقدار حد آستانه هر سری زمانی ساخته می‌شود، مفهوم حد آستانه خوشه از روش CAST[2] (بندور، شمیر و یاخینی، ۱۹۹۹) به‌دست می‌آید، به این صورت که برای هر داده موجود در خوشه اولیه (PC_i)یک مقدار آستانه شباهت تقریبی بر مبنای رابطه زیر تعریف می‌شود:

$$a(F_x) = \frac{\sum_{y \in c} M_{xy}}{|SC|}$$ رابطه (۴)

که در آن M_{xy} ماتریس فاصله بین داده F_x و F_y و $|SC|$ تعداد داده‌های موجود در خوشه جدید است. PCS خوشه‌ها را به‌صورت پیوسته و بر مبنای تعریف پویای حد آستانه ایجاد می‌کند. خروجی روش PCS خوشه‌هایی هستند که از جداسازی خوشه‌های مرحله قبل ایجاد شده‌اند. در این فرآیند، هر خوشه با اضافه شدن سری‌های زمانی‌که دارای بیشترین شباهت با اعضای داخل خوشه هستند ساخته می‌شود. اضافه و کم شدن داده‌ها از هر خوشه تا زمانی ادامه می‌یابد که تغییرات بیشتر در خوشه ایجاد شده ممکن نباشد. در این مرحله برای تعیین تعداد زیرخوشه‌های ایجاد شده، یک حد آستانه( )[3]به ازای هر خوشه برای پذیرش ورود داده‌ها در آن تعریف می‌شود:

$$\alpha = \frac{\sum_{x,y \in PC_U, M_{xy} \geq \mu}(M_{xy} - \mu)}{|PC_U|} + \mu$$ رابطه (۵)

$$\mu = \frac{\sum M_{xy}}{|PC|}$$ رابطه (۶)

۱. Pure Cluster Search

۲. Cluster Affinity Search Technique

3. Affinity threshold

غ مقدار اولیه آستانه در ابتدای الگوریتم و $|PC_u|$ تعداد باقی‌مانده از داده‌ها در خوشه‌بندی اولیه در هر مرحله است. مقدار آستانه تعریف شده در این مرحله به‌صورت پویا محاسبه می‌شود، به این معنی که در هر مرحله، تعداد داده‌های باقی‌مانده در خوشه اولیه مورد بررسی قرار گرفته و در صورت برقراری شرایط در خوشه جدید وارد می‌شود. پارامتر تعریف‌شده در معادله بالا، تعداد و سایز خوشه‌های ایجاد شده را کنترل می‌کند و این یکی از مزایای روش استفاده شده در این تحقیق محسوب می‌شود [۱۰]. سپس به ازای هر خوشه ایجاد شده نمونه‌ای (CS_j) تعریف می‌شود، به این صورت که:

(CS_j) داده‌های موجود در $F_i = (F_{i1}, F_{i2} \dots, F_{iT})$

$$R_i = (r_{i1}, r_{i2}, \dots, r_{ix}, \dots, r_{iT}) \qquad \text{رابطه (۷)}$$

R_i نمونه تعریف شده برای خوشه i ام است، که مقادیر r_{ix} از رابطه زیر محاسبه می‌شود:

$$r_{ix} = \frac{\sum_{i=1}^{n} f_{ix}}{n} \qquad \text{رابطه (۸)}$$

به عبارتی تعریف نمونه برای داده‌های موجود در هر خوشه، پیچیدگی محاسبه‌ها را کاهش و باعث افزایش دقت خوشه‌بندی می‌شود، با استفاده از این تعریف نقاط دورافتاده سری زمانی نیز در نظر گرفته می‌شود. در نهایت، تمامی داده‌ها بررسی و با در نظر گرفتن حد آستانه برای هر خوشه، خوشه‌بندی دقیق‌تر و با کیفیت بهتر، نتیجه می‌شود.

مرحله سوم- ادغام[۱]. بررسی نقاط مهم یا بحرانی[۲] (PIPs) از مسائل مهم در خوشه‌بندی داده‌های سری زمانی است. مفهوم PIPs در آنالیز داده‌های مالی به معنای نگه‌داشتن نقاط مهم در داده‌ها و حذف نقاط دورافتاده است. در تحقیق دیگری [۱۳] مدلی را برای استفاده از مفهوم نقاط تاب[۳] در سری زمانی مالی، مطرح می‌کند. نقاط تاب، نقاط مهم و بحرانی در داده‌های مالی هستند که باید مورد توجه قرار گیرند.

خروجی مرحله دوم، شاخص‌هایی برای هر خوشه (R_i) ارائه می‌دهد. در مرحله سوم فاصله بین این شاخص‌ها برای ایجاد خوشه‌های نهایی محاسبه می‌شود. در این مرحله، بررسی

. Merging
۲. Perceptually important point(PIPs)
۳. Turning point

شیفت‌های محلی انجام می‌شود. همان‌طور که اشاره شد، بررسی روند حرکتی قیمت سهام شرکت‌ها با در نظر داشتن شیفت‌های زمانی نتیجه واقعی‌تری را ارائه خواهد داد. با توجه به مطالب بیان شده، از آنجاکه یکسان بودن شکل بازار شرکت‌ها مورد بررسی قرار می‌گیرد استفاده از روش DTW[1] برای اندازه‌گیری فاصله مناسب‌تر به نظر می‌رسد. این روش اندازه‌گیری فاصله، با استفاده از محور تاب[2]به بهترین هم‌ترازی بین داده‌ها در سری دست می‌یابد.

برای درک بهتر روند محاسباتی مرحله سوم، فرض بر این است از بین خوشه‌های ایجاد شده در مرحله قبل، دو خوشه X وY انتخاب شده است. فرض می‌کنیم، خوشه X شامل n شرکت و خوشه Y شامل m شرکت باشد. برای یافتن فاصله بین خوشه‌های (SC_x)و(SC_y) ، شاخص‌های محاسبه شده برای این دو خوشه را در نظر می‌گیریم:

$$R_x = \{r_{x1}, r_{x2}, \dots, r_{xi}, \dots, r_{xn}\}$$
$$R_y = \{r_{y1}, r_{y2}, \dots, r_{yi}, \dots, r_{ym}\}$$

ماتریس فاصله بین داده‌های موجود در هر شاخص، با در نظر گرفتن فاصله اقلیدسی محاسبه می‌شود؛ به‌طوری‌که نتیجه یک ماتریس n*m خواهد بود، $\mathrm{M_{ij}} = d(r_{xi}, r_{yj})$ مقدار فاصله به‌دست آمده برای شرکت iام از خوشه X و شرکت jام از خوشه Y را نشان می‌دهد. حال می‌توان مسیر بهینه بین دو خوشه X و Y را با استفاده از معادله ۹ که روش اندازه‌گیری فاصله DTW است، محاسبه کرد.

$$dis_{\mathrm{DTW}}(\mathrm{R_x}, \mathrm{R_y}) = \min(\textstyle\sum_{k=1}^{K} \mathrm{W_K}/\mathrm{K}) \qquad \text{رابطه (۹)}$$

$$W = (w_1, w_2, \dots, w_u) \quad w_u = \{(r_{x1}, r_{y1})(r_{xi}, r_{yj}), \dots, (r_{xn}, r_{ym})\} \quad \text{نقاط تاب}$$

در نهایت، با داشتن ماتریس فاصله بین شاخص خوشه‌ها که با استفاده از روابط بالا محاسبه شد، خوشه‌بندی مجدد با استفاده از روش k-medoides انجام می‌شود، نتیجه این خوشه‌بندی، ادغام خوشه‌هایی است که دارای سری‌های زمانی مشابهی بودند که در مراحل قبل در یک خوشه قرار نگرفته بودند.

. Dynamic time warping
. Axis warping

ارزیابی مدل. اگر چه نتایج خوشه‌بندی شرکت‌ها (نمودارها و جداول ارائه شده) برای متخصصان مالی قابل تفسیر است؛ ولی استفاده از معیارهای اعتبارسنجی مدل برای بیان مزیت روش سه‌مرحله‌ای نسبت به سایر روش‌های موجود ضروری به نظر می‌رسد. از طرفی، با تغییر پارامتر w که برای بخش‌بندی داده‌ها در مرحله اول خوشه‌بندی تعریف شد، نتایج متفاوت خواهد بود؛ بنابراین باید نتیجه استفاده از روش سه‌مرحله‌ای، با مقادیر مختلف w نیز مورد بررسی قرار گیرد تا مقدار مناسب این پارامتر به‌دست آید.

همان‌طور که بیان شد، خوشه‌بندی روش یادگیری بدون نظارت است؛ بنابراین کلاس‌های از پیش تعریف شده‌ای برای مقایسه نتایج خوشه‌بندی با الگوریتم‌های متفاوت و ارزیابی کیفیت مدل وجود ندارد. یکی از معیارهای متداول برای ارزیابی کیفیت خوشه‌ها، استفاده از روش مجموع مربعات خطا (SSE[۱]) است. برای هر سری زمانی، خطا به‌صورت فاصله نزدیک‌ترین خوشه از مرکز داه‌های خوشه، تعریف می‌شود. معادله ۱۰ برای محاسبه مقدار SSE مورد استفاده قرار می‌گیرد:

$$SSE = \sum_{j=1}^{K} \sum_{F_i \in SC_j} (dis(F_i, R_j))^2 \qquad \text{رابطه (۱۰)}$$

در معادله بالا، F_i سری زمانی i ام از خوشه j ام و R_j نماینده خوشه j ام است که با استفاده از معادلات ۷ و ۸ محاسبه می‌شود. برای اندازه‌گیری فاصله، معیار اندازه‌گیری ED است.

۴. تحلیل یافته‌ها

در تحقیق حاضر روش به‌کار رفته برای بررسی هم‌حرکتی بین داده‌های مورد آزمایش، رویکرد خوشه‌بندی است. با توجه به حجم زیاد داده‌ها در این بررسی و با توجه به این‌که فراوانی وقوع متغیرها یکسان نیست، برای مدلسازی این دسته از متغیرها نمی‌توان از روش‌های رایج خوشه‌بندی استفاده کرد. از طرفی، عامل دیگری که باید به آن توجه داشت، تغییرپذیری متغیرهای مورد بررسی در طی زمان است. به عبارتی متغیرها دارای شیفت‌های زمانی هستند که این عامل، بر بررسی روند حرکتی داده‌ها تاثیرگذار است.

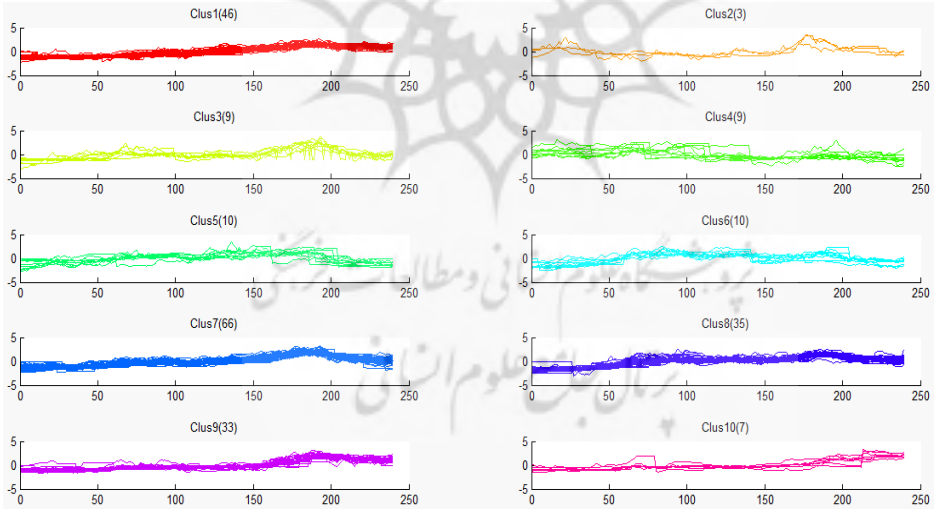
جامعه آماری پژوهش، بورس اوراق بهادار تهران و نمونه آماری تحقیق، قیمت سهام کلیه شرکت‌های پذیرفته‌شده در بورس تهران در دوره زمانی ابتدای سال ۱۳۹۲ تا پایان این سال را شامل می‌شود. در پژوهش حاضر، داده‌های مورد نیاز که همان قیمت پایانی تعدیل شده سهام شرکت‌ها است به‌صورت روزانه در فاصله زمانی ۹۲/۱/۶ تا ۹۲/۱۲/۲۸ از طریق شرکت مشاور

۱. Sum of Squared Error

سرمایه‌گذاری ارزش‌پرداز آریان دریافت شده است. شایان ذکر است که داده‌های مورد استفاده، تعدیل شده هستند، یعنی سود نقدی و افزایش و کاهش سرمایه در داده‌های مورد نظر لحاظ شده است.

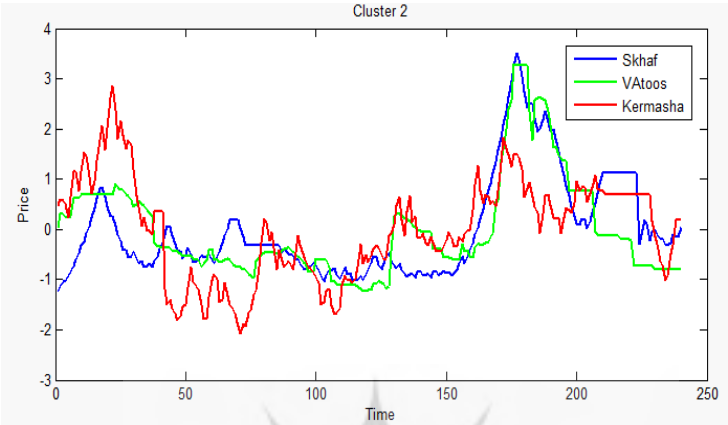
تعداد روزهای کاری در طی دوره مورد نظر، ۲۴۳ روز است. از آنجاکه برای انجام تحقیق حاضر با توجه به روش‌شناسی انجام پروژه، تعداد داده‌ها برای تمامی شرکت‌های مورد بررسی باید یکسان باشد، تعداد داده‌های هر شرکت ۲۴۰ داده در نظر گرفته شد. از طرفی با توجه به این مسئله که تعدادی از شرکت‌ها در طی زمان بررسی، داده‌های کافی ارائه نداده‌اند، شرکت‌هایی که تعداد داده‌های در دسترس آن‌ها کمتر از ۲۰۰ داده در طی سال بود از مقایسه حذف شد و در نهایت، ۲۲۸ شرکت برای اجرای مدل مورد استفاده قرار گرفت. داده‌های جمع‌آوری شده، ابتدا به‌صورت سری‌های زمانی منظم در نرم‌افزار Excel دسته‌بندی شد و سپس با وارد کردن داده‌ها در محیط نرم‌افزار Matlab و با استفاده از امکانات این نرم‌افزار زمینه بررسی و تحلیل داده‌ها و طراحی مدل تا حصول نتایج فراهم شد.

در مرحله اول مدل، مقدار تعریف شده برای تعداد خوشه‌ها (K) در روش K-Medoieds، $\sqrt{n/2}$ است. n تعداد کل داده‌ها در هر سری زمانی یا به عبارتی تعداد داده‌های هر شرکت در طی دوره مورد بررسی است. مقادیر در نظر گرفته شده برای پارامتر w، ۴ ،۶ و ۸ است. نتایج با تغییر این پارامتر متفاوت خواهد بود. نمودار ۱، خروجی مرحله اول را با در نظر گرفتن w= نشان می‌دهد.



نمودار ۱. نتیجه مرحله اول، W=۸

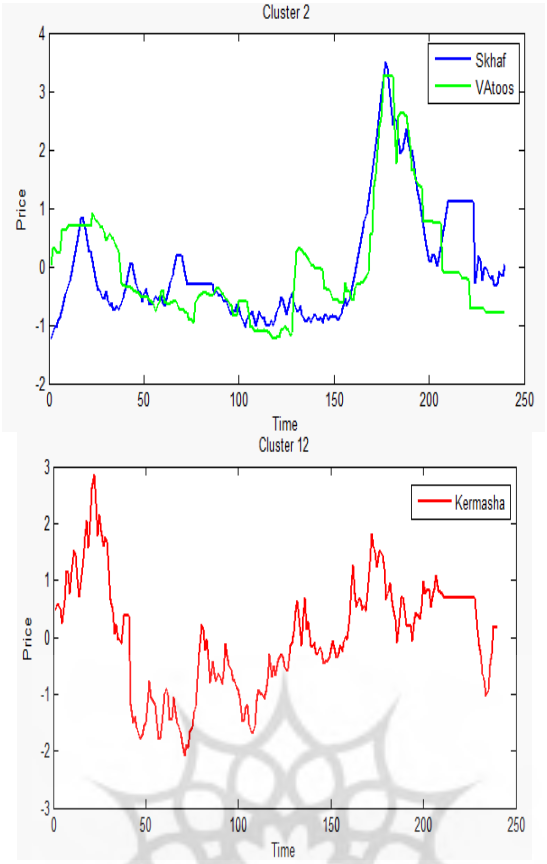
نتایج مرحله اول نشان می‌دهد که روش خوشه‌بندی سه‌مرحله‌ای، می‌تواند نقاط دورافتاده یا داده‌های پرت در سری‌زمانی را تشخیص دهد. همان‌طور که در نمودار ۱ مشاهده می‌شود، از آنجاکه به‌طور معمول شیفت‌های زمانی در روند حرکتی قیمت سهام شرکت‌ها وجود دارد، خوشه‌های رسم شده در نمودار دارای وضوح کافی نیستند. نمودار ۲، روند شرکت‌های موجود در خوشه دوم را که شامل سه شرکت است، از نمای نزدیک نشان می‌دهد.



نمودار ۲. خوشه دوم در پایان مرحله اول

در مرحله دوم، هر کدام از خوشه‌های مرحله قبل به دو خوشه مجزا تبدیل شده است. این خوشه‌های ایجاد شده نسبت به خوشه‌های مرحله قبل کیفیت قابل قبول‌تری خواهند داشت. نمودارهای زیر، نتایج خوشه‌بندی مرحله دوم برای خوشه ۲ ایجاد شده در نمودار ۲ را نشان می‌دهد. همان‌طور که در نمودارهای شکل زیر مشخص است، دو شرکت «سخافو و ەتوس» که هم‌حرکتی بیشتری را در طی دوره زمانی مورد بررسی نشان می‌دهند در خوشه باقی‌مانده و شرکت «کرماشا» که روند حرکتی متفاوت دارد در خوشه مکمل، خوشه ۱۲، وارد شده است.

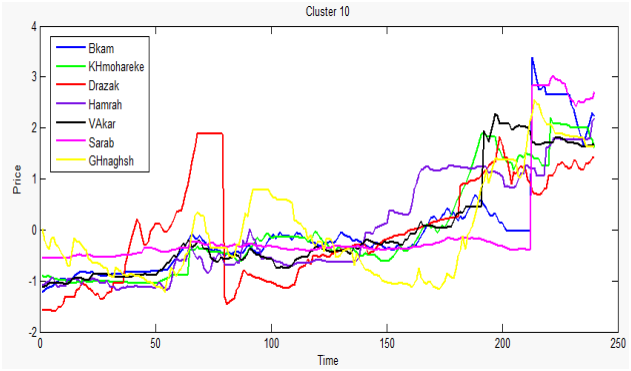
پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



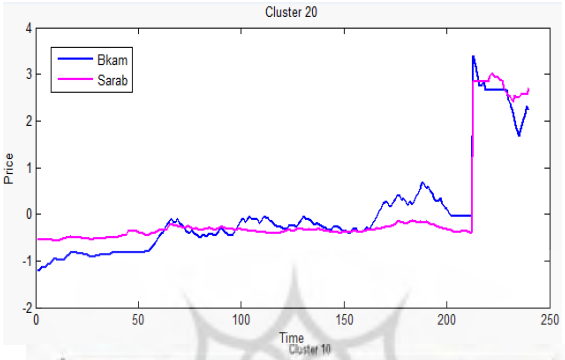
نمودار ۳. نتایج مرحله دوم برای خوشه دو، w=۸

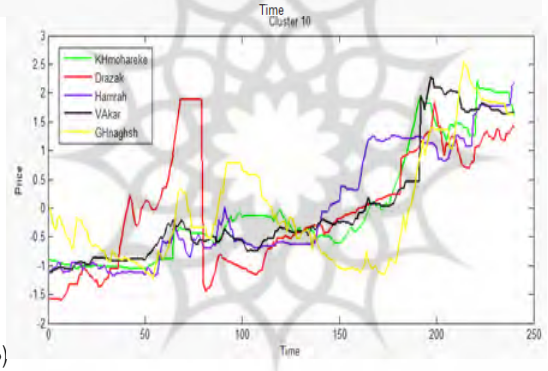
روش خوشه‌بندی سه‌مرحله‌ای توانایی ایجاد نمودار سلسله مراتبی حاصل از خوشه‌بندی در مرحله دوم را دارا است. نمودار «۴- الف»، خوشه شماره ۱۰ را که در مرحله اول خوشه‌بندی ایجاد شده و نمودار «۴- ب» نتایج مرحله دوم را برای خوشه بازبینی‌شده شماره ۱۰ و خوشه شماره ۲۰ که مکمل آن است؛ نشان می‌دهد. در قسمت «ج» از نمودار ۴، نمودار سلسله مراتبی حاصل از خوشه شماره ۱۰ در مرحله دوم نشان داده شده است.

پرتال جامع علوم انسانی

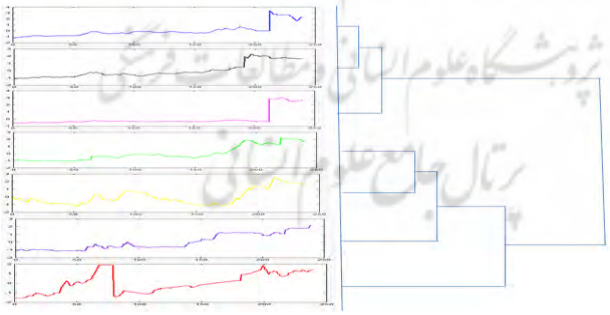


(الف)





(ب)

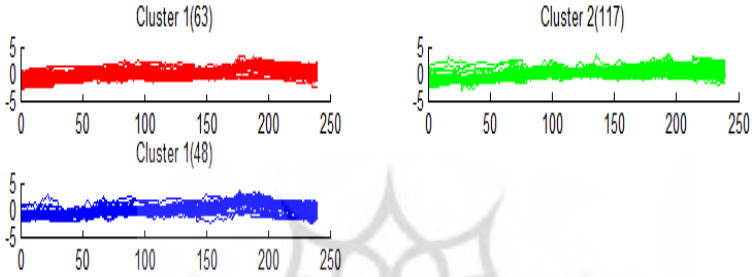


(ج)

نمودار ۴. نتایج مرحله اول و دوم و نمودار سلسله مراتبی خوشه ۱۰

نمودار دندوگرام یا سلسله مراتبی مرحله دوم می‌تواند برای پیش‌بینی هم‌حرکتی شرکت‌ها مورد استفاده قرار گیرد.

خروجی مرحله دوم، شاخص‌هایی برای هر خوشه (R_i) ارائه می‌دهد. در مرحله سوم، برای خوشه‌بندی نهایی، ابتدا شاخص‌های مربوط به هر خوشه (R_i) با استفاده از معادله‌های ۷ و ۸ محاسبه می‌شود. سپس مسیر بهینه بین خوشه‌ها با استفاده از معادله ۹ که روش اندازه‌گیری فاصله DTW است، محاسبه می‌شود. با داشتن ماتریس فاصله بین شاخص خوشه‌ها، خوشه‌بندی خوشه‌های ایجاد شده در مرحله قبل، با استفاده از روش k-Medoids انجام می‌شود، نتیجه این خوشه‌بندی، ادغام خوشه‌هایی است که دارای سری‌های زمانی مشابهی بودند که در مراحل قبل در یک خوشه قرار نگرفته بودند.



نمودار ۵. خوشه‌های ایجاد شده در پایان مرحله سوم با w=۸

همان‌طور که بیان شد در مرحله اول برای تقسیم داده‌ها به بخش‌های مختلف، پارامتر W تعریف می‌شود. در این تحقیق، مقادیر در نظر گرفته‌شده برای این پارامتر ۴ و ۶ و ۸ بوده است. نتایج اجرای مدل با در نظر گرفتن مقادیر مختلف پارامتر W متفاوت خواهد بود. حال برای تشخیص این که کدام مقدار برای این پارامتر مناسب‌تر است و خوشه‌بندی بهتری را نشان می‌دهد، مقدار مجموع مربعات خطا (SSE) برای مقادیر مختلف w با استفاده از معادله ۱۰ محاسبه می‌شود. واضح است که هر چه مقدار به‌دست‌آمده برای SSE کمتر باشد، مدل بهتر خواهد بود. نتایج حاصل برای مقادیر SSE در زیر آمده است:

از طرفی، مدل ارائه‌شده در این تحقیق، با الگوریتم متداول K-Medoied مورد مقایسه قرار می‌گیرد. برای این کار نیز داده‌های خام به روش K-Medoied خوشه‌بندی شده است و مقدار SSE برای این خوشه‌بندی نیز محاسبه می‌شود:

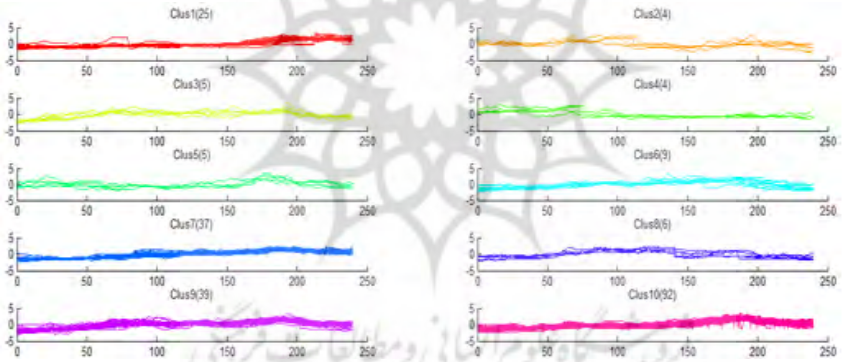
جدول ۱. مقادیر به‌دست آمده برای SSE

	W=۸	W=۶	W=۴	K-Medoids
SSE	۰.۱۲۵۰	۰.۰۱۸۵	۰.۱۳۵۵	۰.۱۹۵۳

با توجه به مقادیر به‌دست‌آمده برای SEE، انتخاب مقدار W=۶ نتیجه بهتری را به‌دست خواهد داد. از طرفی نتایج جدول ۱ نشان می‌دهد که روش خوشه‌بندی سه‌مرحله‌ای در هر صورت بهتر از روش K-Medoids عمل کرده است.

۵. بحث و نتیجه‌گیری

هدف از پژوهش حاضر، بررسی هم‌حرکتی شرکت‌های موجود در بورس اوراق بهادار تهران با استفاده از روش‌های خوشه‌بندی است. با در نظر داشتن نتایج تحقیقات انجام شده در زمینه بررسی هم‌حرکتی بازار سهام، در تحقیق حاضر، مطابق پژوهش آقابزرگی و همکاران (۲۰۱۴) از روش خوشه‌بندی سه‌مرحله‌ای برای دسته‌بندی شرکت‌های بورس تهران استفاده شد. براساس یافته‌های مدل ملاحظه می‌شود که مقدار SSE محاسبه‌شده برای آزمون کارآیی مدل، زمانی‌که مقدار پارامتر W=۶ باشد، کمتر است؛ بنابراین نتایج برای خوشه‌بندی شرکت‌ها، زمانی‌که W=۶ است، قابل قبول خواهد بود. نمودارهای زیر نتایج اجرای مراحل مختلف مدل را با مقدار W=۶ را نشان می‌دهد.



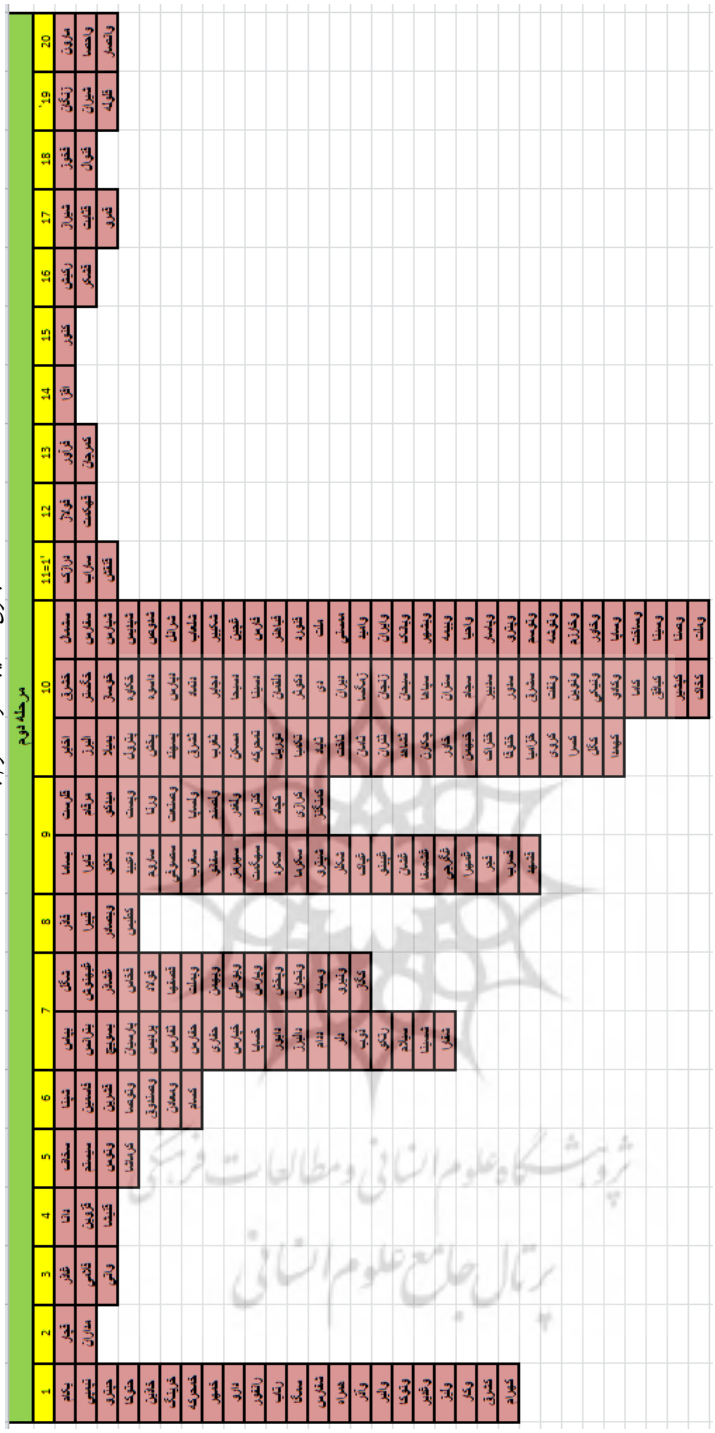
نمودار ۶. نتیجه مرحله اول برای W=۶

پرتال جامع علوم انسانی

جدول ۲. شرکت‌های عضو هر خوشه در پایان مرحله اول با $W=۶$

مرحله اول																			
۱		۲	۳	۴	۵	۶	۷		۸	۹		۱۰							
بکام	ولیز	فولاژ	غاذر	افرا	سخاف	شپنا	پیاس	فخاس	فاذر	بساما	غمهرا	سدیر	خزامیا	وایران	اخابر				
تیمپی	وکار	قجار	فرآور	دانا	سیستم	فاسمین	بترانس	فولاد	فخوز	تایرا	فجر	سدور	خشرق	وبانک	بمیلا				
حیترو	کشرق	قهکمت	فلامی	قزوین	وتوس	قشکر	بسویچ	فثابت	فنوال	تکنو	فسرب	سشرق	خکستر	وبشهر	پترول				
حتوکا	کهرام	مداران	واتی	قنیشا	کرماشا	وتوصا	پارسیان	قصفها	قهیرا	دعبید	فلوله	سشمال	خوساز	وبیمه	پخش				
خاذین			کمرجان		کنور	وصندوق	پردیس	قمرو	وبصادر	زنگان	قشهد	سفارس	خکاوه	واحیا	پسهند				
خرینگ						ومعادن	ثفارس	وبملت	کطبس	سپهان	قلارست	شپارس	داسوه	وپاسار	ثشرق				
خمحرکه						کسام	حفارس	وبهمن		سصوفی	مرقام	شپدیس	دبارس	وپترو	ثغرب				
خمهر						رکیش	حفاری	وبوعلی		سغرب	میدکو	شدوص	دنماد	وتوسم	مسکن				
دارو						قشرین	خبارس	وبارس		سفانو	وپست	شرانل	دجابر	وتوشه	تمحرکه				
درازک							خسایا	وبخش		سهرمز	ورنا	شلعاب	دسیحا	وخارزم	توریل				
رانفور							دابور	ونجارت		سهگمت	وصنعت	شکبیر	دسینا	وخاور	تکمیا				
رتاب							دالبرز	وسپه		سکرد	ولساپا	غجین	دلقمان	وساپا	تاباد				
سازاب							ددام	ونیرو		سکرما	ولصنم	فارس	دکوثر	وساخت	تاخت				
سمگا							دلر	ککاز		شپترو	ولغدر	فباهنر	دی	وسینا	تامان				
شفارس							ذوب			شکلر	کترام	فنورد	دیران	وصنا	تتران				
قنفش							رتکو			شیران	کچاد	مارون	زمگسا	وملت	تشاهد				
همراه							سیلام			غپاک	کرازی	ملت	زنجان	ونفت	چکارن				
وآذر							شسینا			غپنو	کمنگنز	ممسنی	سبحان	ونوین	خاور				
والبر							شفارا			غشان		واحصا	سپاها	ونیکی	خبهمن				
وتوکا							شکل			غشمفا		وامید	ستران	وکادو	خنراک				
وغدیر							شیراز			غگرجی		وانصار	سجام	کاما	خنوقا				
							غبهنوش					کسرا	کهمدا	کبافق	کخاک				
							غشاذر					کنگل	ثنوسا	کپشیر	کروی				

جدول ۳. نتیجه مرحله دوم با $k=17$



جدول ۴. نتیجه مرحله سوم با W=۶

مرحله سوم		
1	2	3
cluster7	cluster16	cluster1
cluster11	cluster19	cluster2
cluster17		cluster3
cluster20		cluster4
		cluster5
		cluster6
		cluster8
		cluster9
		cluster10
		cluster12
		cluster13
w=6		cluste14
		cluster15
		cluster18

از طرفی، با توجه به نتایج جدول ۱ که نتایج را بر مبنای مقادیر مختلف W نشان می‌دهد، مشاهده می‌شود که مقدار خطا زمانی که داده‌ها توسط الگوریتم K-Medoids خوشه‌بندی می‌شوند، بیشتر است. این نشان می‌دهد روش خوشه‌بندی سه‌مرحله‌ای بهتر از الگوریتم‌های معمول عمل خواهد کرد. به علاوه، روش خوشه‌بندی به‌کار گرفته شده در این تحقیق، این امکان را فراهم می‌کند تا سری‌های زمانی با ابعاد زیاد نیز به‌راحتی با تعریف نمونه برای هر خوشه ایجاد شده در مراحل مختلف مدل، دسته‌بندی شوند.

از طرفی بررسی نتایج از خوشه‌بندی نهایی با W=۶، احتمال هم‌حرکتی بین شرکت‌های واقع در یک صنعت را تقویت می‌کند. عمده شرکت‌های فعال در زمینه‌های شیمیائی، حمل و نقل، دارو، برق و بانکداری در خوشه یک قرار گرفته‌اند. از سوی دیگر، خوشه سه به‌طور عمده به صنایع مرتبط با سیمان، ساختمان، معدنی- فلزی، قطعه‌سازی خودرو، انفورماتیک و واسطه‌گری‌های مالی اختصاص دارد. از آنجاکه متغیرهای اقتصادی تاثیرگذار بر شرکت‌های واقع در یک صنعت تا اندازه زیادی با یکدیگر شباهت دارند؛ مطلب اخیر تا اندازه زیادی منطقی به نظر می‌رسد. همچنین خوشه‌بندی به‌دست‌آمده نشان‌دهنده هم‌حرکتی برخی صنایع با یکدیگر است. برخی صنایع علی‌رغم طبقه‌بندی‌های متفاوت بازهم از متغیرهای نسبتا مشابهی تاثیر می‌پذیرند و به‌عنوان صنایع پسین و پیشین با یکدیگر در ارتباط هستند. برای نمونه قرار گرفتن کلیه شرکت‌های ساختمانی، سیمان، معدنی و فلزی در خوشه سه، بیانگر مطلب اخیر است.

بررسی نتیجه تحقیق نشان می‌دهد که استفاده از روش خوشه‌بندی سه‌مرحله‌ای از آنجاکه شیفت‌های زمانی داده‌ها را هم در نظر می‌گیرد، علاوه بر دسته‌بندی شرکت‌ها باعث بررسی روند حرکتی قیمت سهام آن‌ها نیز شده است. یکی از دستاوردهای مهم ازریابی هم‌حرکتی در بازار بورس اوراق بهادار، تخصیص بهینه دارایی‌ها و مدیریت ریسک است. از دیدگاه اقتصادی، روش

به‌کار رفته قابلیت نشان دادن همبستگی بین شرکت‌ها را دارا است. این فرآیند اطلاعات مفیدی را در اختیار مدیران صندوق‌ها و سرمایه‌گذاران جهت افزایش قدرت تصمیم‌گیری برای سرمایه‌گذاری با بازده بالا براساس هم حرکتی شرکت‌ها قرار می‌دهد. از دیدگاه تکنیکال، نتایج تحقیق نشان می‌دهد که برخلاف سایر رویکردها، استفاده از روش خوشه‌بندی سه‌مرحله‌ای برای خوشه‌بندی سری‌های زمانی، نیاز به محاسبه فاصله بین تمام داده‌ها ندارد. به این ترتیب، میزان محاسبه‌ها و زمان اجرای مدل کاهش می‌یابد؛ در حالی‌که خوشه‌هایی با دقت بالاتر با تعریف نمونه در هر خوشه به‌دست می‌آید.

محدودیت‌های تحقیق. همان‌طور که بیان شد، یکی از محدودیت‌های اساسی در تحقیق حاضر، عدم وجود استانداردهای از پیش تعریف‌شده در بازار سهام داخلی است که مقایسه نتایج تحقیق را با مشکل مواجه کرد.

توقف طولانی‌مدت نماد شرکت‌ها، حجم مبنا و محدود بودن دامنه نوسان که به تشکیل صف‌های خرید و فروش منجر شده و موجب می‌شود تا سهام شرکت‌ها با تاخیر به قیمت‌های تعادلی خود دست یابند، از دیگر محدودیت‌های موجود در تحقیق است که بر داده‌های ورودی و در نهایت بر نتایج تاثیرگذار است.

پیشنهادها

پیشنهادهای کاربردی

ـ استفاده از نتایج تحقیق برای تشکیل پرتفوی بهینه به‌منظور کسب بازده بیشتر و تحمل ریسک کمتر.

ـ استفاده از نتایج در تعریف و ایجاد شاخص‌ها جهت طراحی صندوق‌های شاخصی. با توجه به این‌که، شرکت‌هایی که در یک خوشه قرار دارند، روند حرکتی یکسان دارند، می‌توان سهام شرکت‌هایی را انتخاب کرد که در خوشه‌های متفاوت قرار گرفته‌اند. به این ترتیب می‌توان صندوقی طراحی کرد که با حداقل تعداد سهام ممکن موقعیت بازار را به‌خوبی نشان دهد.

ـ تهیه و ارائه شاخص‌های لازم برای ارزیابی وضعیت بازار توسط سازمان بورس و اوراق بهادار.

پیشنهادها جهت تحقیقات آتی

ـ در تحقیق حاضر برای بررسی هم‌حرکتی قیمت سهام شرکت‌ها در بازه‌های زمانی کوتاه‌مدت از داده‌های یک ساله شرکت‌ها استفاده شده است. پیشنهاد می‌شود جهت پیش‌بینی بلندمدت

هم‌حرکتی سهام شرکت‌ها، از داده‌های دوره‌های زمانی طولانی‌تر استفاده شود و به داده‌های مورد بررسی، براساس زمان وزن داده شود.

ـ با توجه به نیازهای تحقیقات آتی می‌توان مقادیر متفاوتی برای تعداد خوشه‌های نهایی (K) در نظر گرفت.

در پژوهش حاضر در مرحله سوم مدل برای اندازه‌گیری فاصله بین داده‌ها از روش اندازه‌گیری DTW استفاده شد. در تحقیقات آتی می‌توان از سایر روش‌های اندازه‌گیری فاصله استفاده کرد.

ـ جایگزین SAX برای کاهش ابعاد داده‌ها

ـ استفاده از سایر مقادیر برای پارامتر W در تقسیم‌بندی داده‌ها و مقایسه نتایج

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

منابع

۱. ایزدپرست، محمد. (۱۳۹۰). دسته‌بندی مشتریان بیمه با استفاده از داده‌کاوی، دانشگاه پیام نور تهران. *ماهنامه تازه‌های جهان بیمه. شماره ۱۶۱.*

۲. ایمانی، محسن. (۱۳۹۱). *خوشه‌بندی متون فارسی، پایان نامه کارشناسی ارشد، دانشکده مهندسی کامپیوتر. دانشگاه صنعتی شریف.*

۳. رادمهر، فرزاد، و علم الهدائی، سید حسن (۱۳۹۳). خوشه‌بندی ابزاری برای آنالیز داده‌ها در مطالعات کمی و آمیخته. روش‌ها و *مدل‌های روانشناختی، سال چهارم، شماره پانزدهم*، ص ۳۶-۱۳.

۴. شیری. (۱۳۸۵). مقدماتی بر خوشه‌بندی. جزوه درسی یادگیری ماشین. دانشگاه صنعتی امیرکبیر، از http://www.aut.ac.ir/shiry/clustering.html.

۵. فرید، داریوش، و پورحمیدی، مسعود (۱۳۹۱). بخش‌بندی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از تحلیل خوشه‌ای فازی. مجله حسابداری مالی، سال چهارم، شماره سوم، شماره پیاپی (۱۳)، ۱۰۵-۱۲۸.

۶. فرید، سارا. (۱۳۸۶). گروه‌بندی سهام با استفاده از شبکه‌های عصبی خود سازنده. پایان نامه کارشناسی‌ارشد. دانشکده مدیریت و اقتصاد. دانشگاه صنعتی شریف.

۷. قاضی کلهرودی، سروش. (۱۳۸۵). بررسی هم‌حرکتی بین قیمت‌های سهام با استفاده از تئوری پیچیدگی در یک بازار سهام. پایان نامه کارشناسی ارشد. دانشکده مدیریت و اقتصاد. دانشگاه صنعتی شریف.

۸. قالیباف اصل، حسن، و محمدی، شاپور، و مظاهری، پگاه (۱۳۹۱). بررسی هم‌حرکتی بازار سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از مدل APT. پژوهش‌های مدیریت در ایران، دروه ۱۶، شماره۱، ۹۳-۱۰۶.

۹. عبدی، جعفر. (۱۳۸۹). *بررسی ارتباط بین بازارهای سهام تهران و دبی. پایان‌نامه کارشناسی ارشد. دانشکده مدیریت و اقتصاد. دانشگاه صنعتی شریف.*

. Aghabozorgi, E., & Wahthe, Y. ( ). Stock market co-movement assessment using a three-phaseclustering method. *Expert systems with Applications*, , - .

. Antoniou, A. ( ). Modeling international price relationships an interdependencies-between the stock index and stock index futures markets of three EU countries: A multivariate. *Journal of Business, Finance and Accounting*, , ° .

. Bagnall, A. J., & Janacek, G. ( ). Clustering time series with clipped data. *Machine Learning*, ( ), ° .

. Bao, D. ( ). A generalized model for financial time series representation and prediction. *Applied Intelligence*, ( ), ° .

. Collins, D., & Biekpe, N. ( ). Contagion and interdependence in African stock markets. South *African Journal of Economics*, ( ), ° .

. Chu, S., Keogh, E. J., Hart, D., & Pazzani, M. J. ( ). Iterative deepening dynamic time warpingfor time series. *In Second SIAM international conference on data mining*, ° .

. Dutt, P., & Mihov, I. ( ). Stock market co-movements and industrial structure. *Journal of Money, Credit and Banking, ( ),* ° .

. Kaufman, L., & Rousseeuw, P. J. ( ). Finding groups in data: an imtroduction to cluster analysis. *Hoboken, NJ: Wiley.*

. Keogh, E., & Kasetty, S. ( ). On the need for time series data mining benchmarks: A survey and empirical demonstration. *Data Mining and Knowledge Discovery, ( ),* ° .

. Keogh, E., & Ratanamahatana, C. ( ). Exact indexing of dynamic time warping. *Knowledge and Information Systems, ( ),* ° .

. Liao, S., & Chou, S. ( ). Data mining investigation of co-movements on the Taiwan and China stock markets for future investment portfolio. *Expert Systems with Applications, ( ),* ° .

. Lin, J., Keogh, E., Lonardi, S., & Chiu, B. ( ). A symbolic representation of time series, with implications for streaming algorithms. *In Proceedings of the th ACM SIGMOD workshopon research issues in data mining and knowledge discovery – KKK D' , .*

. Lin, J., Keogh, E., Wei, L., & Lonardi, S. ( ). Experiencing SAX: A novel symbolic representation of time series. *Data Mining and Knowledge Discovery, ( ),* ° . http://dx.doi.org/ . /s - - -z.

. Liu, W., & Shao, L. ( ). Research of SAX in distance measuring for financial time series data. *In First international conference on information science and engineering. (pp. –). IEEE.*

. Masih, A., & Masih, R. ( ). Dynamic modeling of stock market interdependencies: an empirical investigation of Australia and the Asian NICs. *Review of Pacific Markets and Policies, ( ),* ° .

. Muhammad, M. M., & Marteau, P. F. ( ). *Towards a faster symbolic aggregate approximation method. http:// www.arxiv.org/abs/ - .*

. Nanda, S., Mahanty, B., & Tiwari, M. ( ). Clustering Indian stock market data for portfolio management. *Expert Systems with Applications, ( ),* ° .

. Ratanamahatana, C. ( ). Multimedia retrieval using time series representation andrelevance feedback. *In Proceedings of th international conference on asian digital libraries (ICADL) (pp. –).*

. Torras, Q. ( ). Dynamic Time Warping. *Journal Club.*

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی