فصلنامــه علمی پژوهشی دانش سرمایه‌گــذاری
سال نهم/ شماره سی‌وچهارم/ تابستان ۱۳۹۹

تشکیل سبد سرمایه گذاری بهینه در بورس اوراق بهادار تهران با استفاده از روش‌های خوشه‌بندی سلسله مراتبی و تفکیکی

سیدمجتبی میرلوحی
استادیار دانشکده صنایع و مدیریت، دانشگاه صنعتی شاهرود، شاهرود، ایران (نویسنده مسئول)
mirlohism@shahroodut.ac.ir

نیما محمدی تودشکی
دانشجوی دکتری مدیریت مالی، دانشگاه تهران پردیس بین المللی کیش، کیش، ایران

تاریخ دریافت: ۹۷/۰۳/۰۷ تاریخ پذیرش: ۹۷/۰۳/۲۷

چکیده

نیاز برای دست‌یابی به معیارهای درست بهینه سازی پرتفو که علاوه بر دقت به سرعت در تصمیم گیری‌های سرمایه‌گذاری بیانجامد، امری فراتر از نیازهای آکادمیک است چرا که نیاز سرمایه‌گذاران، شرکت‌های سرمایه‌گذاری و مدیران سرمایه‌گذاری، به کاهش زیان‌های ناشی از سرمایه گذاری، افزایش بازده متناسب با ریسک همواره مورد بحث بوده است. با این وجود دستیابی به روشی جهت بهینه سازی پرتفو به گونه‌ای که شکاف بین نیازهای کاربردی و مدل‌های تئوریک را پر کند کار بسیار دشواری است. با توجه به اینکه یکی از مهمترین عوامل موثر در کسب بازدهی مطلوب، متنوع‌سازی است. این تحقیق با موضوع «تشکیل سبد سرمایه‌گذاری بهینه در بورس اوراق بهادار تهران با استفاده از روش‌های خوشه‌بندی سلسله مراتبی و تفکیکی»، سعی دارد با استفاده از داده‌های بازار و خوشه‌بندی آن، روشی مناسب جهت بهینه‌سازی پرتفو ارائه کند. نتیجه این مقایسه میزان موفقیت بهینه‌سازی بر اساس خوشه‌بندی را نسبت به پرتفوی شاخصی روشن خواهد کرد.

واژه‌ها کلیدی: خوشه‌بندی، بهینه‌سازی، سلسله‌مراتبی، تفکیکی، ارزیابی پرتفو،شارپ.

۱- مقدمه

دستیابی به رشد بلندمدت و مداوم در اقتصاد نیازمند تجهیز و تخصیص بهینه منابع در سطح اقتصـاد ملـی است و این مهم بدون کمک بازارهای مالی به ویژه بازار سرمایه گسترده و کارامـد امکـان پـذیر نیسـت. در یـک اقتصاد سالم، وجود سیستم مالی کارامد در توزیع مناسـب سـرمایه و منـابع مـالی نقـش اساسـی دارد. معمـولا بازارهای مالی را به عنوان سیستمی‌مرکب از افراد و موسسات، ابزارها و رویه‌هایی که پس اندازکنندگان و قـرض گیرندگان را در یک جا جمع می‌کند، تعریف می‌نمایند. یکی از مهمترین بازارهای مـالی، بـازار سـهام اسـت. در دنیای مالی امروز برای سرمایه گذاری در سهام، کسـب دانش،کـاهش هزینـه،انتخاب سـهام برتـر و سـودآورتر و استفاده بهینه از سرمایه،جزء لاینفک اقدامات و فعالیت‌های سرمایه گذاران است. همه تلاش محققـان در طـول تاریخ، کاهش ریسک سرمایه گذاری در بازار سهام و افزایش بازدهی است. یکی از راه‌هـای کـاهش ریسـک،تنوع بخشی در انجام سرمایه گذاری‌هاست. شرکت‌های سرمایه گـذاری بـا سـرمایه گـذاری در دارایـی‌هـای متنـوع و تشکیل پرتفوی اقدام به کاهش ریسک سرمایه گذاری‌های خود می‌کنند.بحث انتخاب سبد سرمایه گذاری بهینه یکی از مباحث مهم در بازار سرمایه است که باید مورد توجه اشخاص حقیقی و حقوقی قرار گیرد. .انتشار نظریـه پرتفوی سهام هری مارکوویتز ،اصلی ترین و مهمترین موفقیت در این راستا بـود (فـابوزی و دیگـران ۲۰۰۷)از زمانی که مارکویتز مدل خود را منتشر کرد این مدل تغییرات و بهبودهای فراوانی را در شـیوه نگـرش مـردم بـه سرمایه گذاری و سبد سهام ایجاد کرد و به عنوان ابزاری کارا برای بهینه سازی سبد سهام بـه کـار گرفتـه شـده است (لایکینگ و دیگران ۲۰۰۶) مارکوویتز پیشنهاد کرد که سرمایه گذاران ریسک و بازده را به صورت توامان در نظر میگیرندو میزان تخصیص سرمایه بین فرصتهای سرمایه گذاری گوناگون را بر اساس تعامـل بـین ایـن دو انتخاب نمایند(فابوزی و دیگران ۲۰۰۷) .

مطالعات نشان می‌دهد سرمایه‌گذاران در جهت انتخاب بهترین سبد سرمایه گذاری به میزان ریسک و بـازده آن توجه می‌کنند. معمولا فرض بر این است کـه سـرمایه گـذاران ریسـک را دوسـت ندارنـد و از آن گریزاننـد و همواره در پی آن هستند تا در اقلامی‌از دارایی‌ها سرمایه گذاری کنند که بیشـترین بـازده و کمتـرین ریسـک را داشته باشند. به عبارت دیگر سرمایه گذاران به بازده سرمایه گذاری به عنوان یک عامل مطلوب مـی‌نگرنـد و بـه واریانس بازده‌ها(ریسک) به عنوان یک عنصر نامطلوب نظر دارند. در بهینه سازی پرتفوی مسـئله اصـلی انتخـاب بهینه دارایی‌ها و اوراق بهاداری است که با مقدار مشخصی سرمایه میتوان تهیه کرد اگر چه کمینه کردن ریسک و بیشینه نمودن بازده سرمایه گذاری به نظر ساده می‌رسد اما در عمل روش‌های متعددی برای تشکیل پرتفـوی بهینه به کار رفته است.

مسئله بهینه سازی مارکوویتز و تعیین مرز کارای سرمایه گذاری،زمانی که تعداد دارایی‌هـای قابـل سـرمایه گذاری و محدودیت‌های موجود در بازار کم باشد ،توسط مدل‌های ریاضیات حل شـدنی اسـت. امـا هنگـامی‌کـه شرایط و محدودیت‌های دنیای واقعی در نظر گرفته شود ،مسئله پیچیده و مشکل خواهد بود، سالهاسـت کـه در حل چنین مسائل پیچیده ای ریاضیات پیشرفته و کامپیوترها به کمک انسان شتافته تا هر چـه بیشـتر وی را در بیرون آوردن از شرایط عدم اطمینان محیطی و ابهام یاری رساند. مدل‌هایی مثل برنامه ریزی خطی، برنامه ریزی

عدد صحیح، برنامه ریزی عدد صحیح مختلط و برنامه ریزی (صفر –یک)، در برنامه ریزی‌های ریاضی وجود دارند که می توانند با در نظر گرفتن هدف و شرایط حاکم بر مسأله، ترکیبی بهینه با مقدار بهینـه مشـخص از عناصـر تشکیل دهنده سبد را ارائه دهند . در نتیجه ، می توان برای رسیدن به چنین هدفی، اطلاعات مالی را با در نظـر گرفتن تمام شرایط حاکم بر سرمایه گذاری در دنیای واقعی وارد برنا مه ریزی ریاضی کرد. روشهای ابتکاری کـه با هدف رفع کاستیهای روشهای کلاسیک بهینه سازی معرفی شدند بـا جسـتجویی جـامع و تصـادفی، احتمـال دستیابی به نتایج بهتر را تا حد زیادی تضمین میکنند. ازجمله ی این الگوریتم‌ها میتـوان بـه الگـوریتم ژنتیـک، کوچ پرندگان، تبرید فلزات، جستجوی ممنوعه، کولونی مورچگان، رقص زنبورهـا، رقابـت اسـتعماری و ... اشـاره کرد.(آرش طالبی ۱۳۸۹).

مسئله انتخاب سهامی‌که به خوبی متنوع شده باشند می‌تواند به وسیله خوشه بندی داده‌ها حل شود. خوشه بندی اولین بار توسط میرکین(۱۹۹۶) به عنوان یک روش ریاضی برای آشکار سازی ساختارهای طبقه بندی در داده‌های جمع آوری شده از پدیده‌ها در دنیای واقعی طراحی شد.روش‌های خوشه بندی مجموعه ای از داده‌ها را به گونه ای سازماندهی می‌کند که اطلاعات با نقاط مشترک در یک خوشه قرار بگیرند.

در این تحقیق، با استفاده از روش‌های خوشه بندی مناسب معرفی شـده توسـط میـرکین(۱۹۹۶) از جملـه خوشه بندی سلسله مراتبی و تفکیکی برای داده‌های سهام بورس اوراق بهادار تهران می‌پردازیم. داده‌های سهام با استفاده از روش‌های خوشه بندی سلسله مراتبی و تفکیکی خوشه بندی خواهند شد. داده‌های سهام شامل بازده در دوره‌های مختلف همچنین نسبت ارزش بازار سهام به فاکتورهـای مختلـف از جملـه سـود هـر سـهم، ارزش دفتری، سود هر سهم، سود تقسیمیو ... که توسط نظر خبرگان انتخاب خواهند شد، خوشه بندی می‌شوند. ایـن داده‌های معیارهای مستقیم سرمایه گذاری هستن که در انتخاب سهام مـورد اسـتفاده قـرار مـی‌گیرنـد. پـس از بررسی روایی خوشه‌های به دست آمده با استفاده از مدل مارکویتز نسبت به تشکیل سبد بهینه برای هـر یـک از الگوریتم‌های مورد استفاده،اقدام خواهد شد.در نهایت با استفاده از شاخص‌های ارزیابی عملکرد پرتفو نسـبت بـه انتخاب بهترین روش خوشه بندی اقدامی‌می‌شود.

هدف از تحقیق حاضر استفاده از روش خوشه بندی برای طبقه بندی سهام در بازار بورس تهران است تـا از این طریق بتوان سهام مشابه را در طبقات یکسان قرارداد و سپس به بهینه سازی پرتفو با استفاده از خوشه‌هـای به دست آمده پرداخت به عبارت دیگر هدف اصلی این تحقیق ارائه یک روش کاربردی برای بهینـه سـازی سـبد سهام است که بتواند سریعتر تغییرات صورت گرفته در بازار را در پرتفو اعمال نماید و نسـبت بـه مـدل پرتفـوی شاخصی بهتر عمل نماید. در نهایت نیز تلاش خواهد شد با استفاده از نتایج به مقایسـه عملکـرد پرتفـوهـای بـه دست آمده از روش‌های مختلف بهینه سازی سبد سهام پرداخته شود. برای دستیابی بـه اهـداف ذکـر شـده، دو فرضیه زیر مورد آزمون قرار گرفته است:

۱) بازده پرتفوی بهینه به دست آمده از الگوریتم‌های خوشه بندی تفکیکی بیشتر از بازده پرتفوی شاخصـی است.

۲) بازده پرتفوی بهینه به دست آمده از الگوریتم‌های خوشه بندی سلسله مراتبی بیشتر از بازده پرتفوی شاخصی است.

۲- مبانی نظری و مروری بر پیشینه پژوهش

- **مدیریت پرتفو**

لغت پرتفوی، در عبارات ساده، به ترکیبی از دارایی‌ها گفته می‌شود که توسط یک سرمایه گذار برای سرمایه گذاری تشکیل می‌شود. این سرمایه گذار می‌تواند یک فرد یا یک موسسه باشد. از نظر تکنیکی، یک پرتفوی در برگیرنده مجموعه‌ای از داراییهای واقعی و مالی سرمایه گذاری شده یک سرمایه گذار است.

در سال ۱۹۵۲ هری مارکویتز مدل اساسی پرتفوی را ارائه کرد که مبنایی برای تئوری مدرن پرتفوی قرار گرفت. قبل از مارکویتز سرمایه گذاران با مفاهیم ریسک و بازدهی پرتفوی آشنا بودند. اگر چه آنها با مفهوم ریسک آشنا بودند ولی معمولاً نمی‌توانستند آن را اندازه گیری کنند. سرمایه گذاران از قبل می‌دانستند که ایجاد تنوع مناسب است و نباید" همه تخم مرغ‌هایشان را در یک سبد بگذارند". با این حال، مارکویتز، اولین کسی بود که مفهوم پرتفوی و ایجاد تنوع را به روش علمی‌بیان کرد. او به صورت کمی‌نشان داد که چرا و چگونه تنوع سازی پرتفوی می‌تواند باعث کاهش ریسک پرتفوی (مجموعه سرمایه گذاری) یک سرمایه گذار شود.

مدل مارکوویتز بر اساس مفروضات زیر بیان شده است:

۱) سرمایه گذاران ریسک گریزند و مطلوبیت مورد انتظار افزایشی است.
۲) سرمایه گذاران پرتفو را بر اساس مبنای میانگین و واریانس مورد انتظار عایدی انتخاب می‌کنند
۳) هر سرمایه گذاری تا بی نهایت قابل تقسیم است.
۴) سرمایه گذاران افق زمانی یک دوره‌ای داشته و برای همه مشابه است.

چرا ایجاد تنوع در سرمایه گذاری برای سرمایه گذاران مهم است؟ می‌توان گفت که قانون شماره یک مدیریت پرتفوی، ایجاد تنوع است. از آنجا که سرمایه گذاران نسبت به آینده مطمئن نیستند باید برای کاهش ریسک دست به ایجاد تنوع در سرمایه گذاری خود بزنند. به عبارت دیگر تشکیل یک پرتفوی متنوع، میزان ریسک را تا حد زیادی کاهش می‌دهد. به عنوان مثال در بحران اقتصادی سال ۱۹۸۷ آمریکا، فقط کمتر از ۹ درصد صندوق‌های مشترک سرمایه گذاری (که اقدام به تشکیل پرتفلیو می‌کردند) با ضرر و زیان مواجه شدند. مجموعه کارا پرتفویی است که با یکسان بودن عایدی بین تمام پرتفوهای مورد انتظار واریانس آن از همه کمتر باشد و یا بیشترین عایدی مورد انتظار در بین تمامی‌آنهایی که واریانس مشابه دارند.

مارکویتز درصدد بر آمد تا روشها و ایده‌های موجود را در قالب یک چارچوب رسمی‌سازماندهی کرده و به این سوال اساسی پاسخ دهد: آیا ریسک پرتفوی با مجموع ریسک اوراق بهادار منفرد، که در مجموع پرتفوی را تشکل می‌دهند برابر است؟ مارکویتز با ارائه روش اندازه گیری ریسک پرتفوی به محاسبه ریسک و بازدهی

پرتفوی مورد انتظار پرتفوی پرداخت. مدل او بر مبنای بازدهی پرتفوی مورد انتظار و ویژگی‌های ریسک اوراق بهادار که چارچوب تئوریکی برای تجزیه و تحلیل گزینه‌های ریسک و بازدهی پرتفوی است، استوار شده است.
تنوع بخشی به قضیه حد مرکزی برمی‌گردد .براساس این قضیه، اگر x1,x2,….xn متغیر تصادفی مستقل باشند و هر xi دارای یک توزیع احتمال با میانگین μ و واریانس σ^2، آنگاه:

$$\lim_{N\to\infty} P\left(\frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}(X_i-\mu)\le y\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{1}{2}s^2}ds$$

با در نظر گرفتن مطالب فوق، برای پرتفویی با دارایی با بازدهی پرتفوی‌های R1,R2,….RN در صورتی که وزن سرمایه‌گذاری هر یک مساوی باشد، بصورت زیر می‌باشد:

$$R_P=\frac{1}{N}\sum_{i=1}^{N}R_i$$

براساس قضیه حد مرکزی، واریانس پرتفوی بصورت زیر خواهد بود:

$$var(R_P)=\frac{1}{N^2}\sum_{i=1}^{N}var(R_i)$$
$$=\frac{1}{N^2}.N.\sigma^2$$
$$=\frac{\sigma^2}{N}_{N\to\infty}\to 0$$

در چارچوب کلاسیک بهینه سازی میانگین واریانس ، سرمایه گذاری را در نظر می‌گیریم که می‌خواهد پرتفویی با N دارایی ریسک پرتفویی تشکیل دهد. .w=(w1,w2,….wN) نشاندهنده بردار وزن سرمایه گذاری در هر یک از دارایی‌هاست. فرض کنیم بردار بازدهی پرتفوی دارایی‌ها بصورت بردار R=(R1,R2,….RN) باشد، دراین صورت بردار بازدهی پرتفوی موردانتظار داراییها ب بصورت μ=(μ1, μ2,…. μN) می‌باشد. Σ ماتریس واریانس کوواریانس بازدهی پرتفوی داراییهاست که بصورت زیر تعریف می‌گردد:

$$\begin{bmatrix}\sigma_{11} & \cdots & \sigma_{1N}\\ \vdots & \ddots & \vdots\\ \sigma_{N1} & \cdots & \sigma_{NN}\end{bmatrix}$$

که در آن σ ij نشاهنده کوواریانس بین بازدهی پرتفوی دارایی i و j است که براساس رابطه $\sigma_{ij}=\sigma_i.\sigma_j.\rho_{ij}$ بدست می‌آید که در آن ρ_{ij} ضریب همبستگی بین بازدهی پرتفوی دو دارایی i و j

است. شایان ذکر است که $\sigma_{ii} = \sigma_i^2$. با توجه به مفروضات بالا، مساله بهینه سازی میانگین واریانس کلاسـیک- بصورت زیر توسط مارکویتز ارائه شد:

$$\mu_p = w'\mu$$
$$\sigma_p^2 = w' \sum w$$

برای مثال اگر سرمایه گذار در دو دارایی با وزنهای $w = (w_1, w_2)$ سرمایه گذاری کند، نرخ بازدهی پرتفـوی موردانتظار و واریانس پرتفوی بصورت زیر محاسبه می‌شوند:

$$\mu_p = w_1\mu_1 + w_2\mu_2$$
$$\sigma_p^2 = [w_1 \quad w_2]\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{bmatrix}\begin{bmatrix}w_1\\ w_2\end{bmatrix}$$
$$= [w_1\sigma_{11} + w_2\sigma_{21} \quad w_1\sigma_{12} + w_2\sigma_{22}]$$
$$= w_1^2\sigma_{11} + w_2^2\sigma_{22} + 2w_1w_2\sigma_{12}$$

بنابراین بهینه سازی پرتفوی در مدل کلاسیک مارکویتز بصورت زیر محاسبه می‌شود:

$$\min_w \; w'\sum w$$
$$\text{s.t:} \quad \mu_0 = w'\mu$$

با این فرض که

$$w'\tau = 1 \;\; , \;\; \tau = [1,1,\dots,1]$$

این مساله، یک مساله بهینه سازی درجه دو است.

$$w = g + h\mu_0$$

μ_0 نرخ بازدهی پرتفوی موردانتظار هدف است.

$$g = \frac{1}{ac - b^2}\sum^{-1}[c\tau - b\mu]$$
$$h = \frac{1}{ac - b^2}\sum^{-1}[a\mu - b\tau]$$
$$a = \tau'\sum^{-1}\tau$$
$$b = \tau'\sum^{-1}\mu$$
$$c = \mu'\sum^{-1}\mu$$

همانگونه که ملاحظه می‌گردد، مارکویتز در مسئله انتخاب پرتفوی استاندارد خـود فـرض میکنـد کـه همـه سرمایه گذاران، انتخاب‌های خود را بر اساس دو معیار بازدهی و ریسک انجام میدهند. در مدل مارکویتز، بازدهی پرتفو معادل میانگین وزنی بازدهی اجزای پرتفو و ریسک آن برابر با انحراف معیار بازدهی‌های دارایی‌های موجود در پرتفو در نظر گرفته شده است. وزن کارای پرتفوها مجموعه‌ای از پرتفوهاسـت کـه در هـر سـطح از ریسـک،

بیشترین بازدهی را ارائه می‌دهند. سرمایه گذاران، کاهش ریسک ناشی از متنوع سازی سرمایه گـذاری را انـدازه گیری می‌کنند. این در حالی است که تحقیقات زیاد، همگی نادیده گرفتن سایر ترجیحات سرمایه گـذاران را در مدل مارکویتز مورد انتقاد قرار داده‌اند.

معمولاً سرمایه گذار در مسئله انتخاب پرتفوی به طور همزمان اهـداف متعارضـی مثـل بـازدهی، ریسـک و نقدشوندگی را به کار می برد. تکنیک‌های برنامه ریزی چند هدفه مثل برنامه ریزی آرمانی و برنامه ریزی توافقی برای انتخاب پرتفویی که به بهترین شکل اهداف و ترجیحات سرمایه گذار را برآورده می سازد، به کار می رود .

فؤاد بن عبدالعزیز و همکارانش در سال ۲۰۰۷، مدل برنامه ریزی توافقی با محدودیت تصادفی را که در واقع ترکیبی از مدل برنامه ریزی و مدل برنامه ریزی با محدودیت تصادفی بود را برای انتخاب پرتفوی ارائه دادند.

ماتسیاس در سال ۲۰۰۴ به همراه همکارانش از مدل چند معیاره برای انتخاب پرتفو استفاده نمودنـد. آنهـا مدلی برای بهینه سازی پرتفوی سهام با استفاده از مدل مارکوویتز ارائه کردند. آنها با استفاده از شـاخص S&P و پنج هدف مرتبط با ریسک و بازدهی و با در نظر گرفتن اولویتهای سرمایه گذاران، از توابع تصمیم گیرنده و تـابع مطلوبیت جهانی استفاده کردند. نتایج نشان داد، الگوریتم ژنتیک، شبیه سازی و استفاده از جستجوی سفارشـی محلی می‌توانند تا حدود زیادی مشکلات موجود در انتخاب پرتفوی را حل کند.

- **الگوریتم‌های طبقه‌بندی**

طبقه‌بندی تشکیل طبقات با استفاده از واحدهای(نقاط) جامعه ی مورد بررسی است بـه طـوری کـه واحـدهای درون هر طبقه دارای حداکثر شباهت و واحدهای بین طبقات دارای حداکثر اختلاف باشند.

یک طبقه‌بندی مناسب باید منجر به تولید طبقاتی شود که حداکثر تشابه درون طبقات و حداقل تشابه بین طبقات حاصل شود. مبحث طبقه‌بندی اساس بسیاری از سیستم‌های پشتیبانی تصمیم را تشکیل می‌دهد.

قره چلو (۱۳۸۹)، در مقاله ای تحت عنوان ارزیابی و مقایسه الگوریتم‌هـای مختلـف طبقـه‌بنـدی در تهیـه نقشه شوری سطحی خاک، به مقایسه روش‌های مختلف الگوریتم طبقه‌بندی پرداخته است. در ایـن مقالـه روش حداکثر احتمال به عنوان دقیق ترین و روش حداقل فاصله تا میانگین با کسب درصد ارزیابی دقت پایین، روشی نامناسب شناخته شده است. نتایج حاکی از آن است که برای بهره گیری از روش حداکثر احتمال، نیاز است کـه طبقات از یک توزیع نرمال برخودار باشند.

جیانسی، سال ۲۰۰۷ در مقاله ای تحـت عنـوان الگـوریتم طبقـه‌بنـدی کـارا بـا اسـتفاده از ماشـین بـردار پشتیبانی (SVM)برای داده‌های وسیع، ماشین بردار پشتیبانی رایکی از روش‌های یادگیری بانظارت معرفی کرده است که از آن برای طبقه‌بندی استفاده شده اسـت. بـه عقیـده ی پژوهشـگر ابـزار مـذکور شـالوده ی تئـوریکی نیرومندی دارد و الگوریتم طبقه‌بندی بر مبنای آن ، کارایی خوبی را موجب می‌شود. این ابزار دقت طبقه‌بندی را افزایش می‌دهد.

هاچلوفی۱ در سال ۲۰۱۲ روش‌های طبقه‌بندی را به دو دسته تقسیم کرده است، روش طبقه‌بندی خود کار و روش تخصیص. وی این روش‌ها را اینگونه تعریف می‌کند: روشهای طبقه‌بندی خودکار روش‌هـایی هسـتند بـر

مبنای فراگیری (یادگیری) نظارت نشده . هدف این روش تشکیل طبقه‌های همگنی است که با توجه بـه معیـار هدف ایجاد شده‌اند. روش تخصیص روشی است بر مبنای یادگیری نظارت شده که از مثال‌هـا و شـاخص‌هـای از قبل شناخته شده بهره می‌گیرد.

ناندا و همکاران در سال ۲۰۱۰ در مقاله‌ای با عنوان«خوشه‌بندی داده‌های بازار سهام هنـد جهـت مـدیریت پرتفو» از رویکرد داده کاوی برای طبقه‌بندی سهام در خوشه‌های مختلف استفاده کردند.پس از طبقه‌بندی، سهام مورد نظر برای پرتفوها از بین این خوشه‌ها انتخاب شدند. این روش منجر بـه کـاهش ریسـک حاصـل از متنـوع سازی شد. رویکرد خوشه‌بندی، سهام را بر اساس معیارهای سرمایه گذاری تقسیم بندی می‌کند. نتـایج تحقیـق ایشان نشان داد، روش تحلیلی K-means نسبت به روش‌های SOM و فازی C-means خوشه‌های متـراکم تـری ایجاد می‌کند.

رشیدی و انالوئی در سل ۱۳۸۵ تحقیقـی بـا عنـوان « اسـتفاده از الگـوریتم K-means تعـدیل یافتـه بـرای خوشه‌بندی شرکتهای فعال در بازار سهام» از این روش برای خوشه‌بندی شـرکتهای فهرسـت شـده در شـاخص داوجونز استفاده کردند. نتایج تحقیق نشان داد، بیش از هشتاد درصد از شرکتهای مـذکور در یـک خوشـه قـرار می‌گیرند.

انواری رستمی‌و ختن لو در مقاله‌ای در سال ۱۳۸۵ به بررسی مقایسه‌ای رتبه‌بندی شرکت‌های برتر بر اساس نسبت‌های سودآوری و شاخص‌های بورس و اوراق بهادار تهران پرداختند. این تحقیق بر اسـاس داده‌هـای پنجـاه شرکت برتر طی سه سال به اجرا در آمد و همبستگی میان نتایج این دو نوع رتبه‌بندی را مورد بررسی مقایسه‌ای قرار می‌دهد. نتایج آزمون تجربی نشان گر آن است که همبستگی ضعیفی میان این دو گروه رتبه‌بندی به عمـل آمده وجود دارد. این بدان معنی است که شرکتهای برتر منتخب بـورس لزومـا دارای رتبـه‌هـای بـالاتر از حیـث نسبت‌های سودآوری نمی‌باشند. بنابراین آنچه را که نسبت‌های حسابداری برتر تلقی می‌کنند از حیث بورس برتر محسوب نمی‌شوند. به عبارتی دیگر، نسبت‌های سودآوری شرکت‌های بورس برای پیش بینی و تعیین شرکت‌های برتر از نظر بورس مناسب نمی‌باشند.

فروغی و همکاران در سال ۱۳۸۹ نیز در پژوهشی با عنوان «مقایسه ریسک سهام رشدی و سهام قیمتـی در بورس اوراق بهادار تهران» برای بررسی ارتباط بین ریسک سیستماتیک بازار و ریسک سیستماتیک سهام قیمتی و رشدی، بازه زمانی پژوهش را به چهار گروه رکود، میانی، توسعه و اوج تقسیم کردند. نتایج پژوهش آنهـا نشـان می‌دهد که در گروه رکود، ارتباط بین ریسک سیستماتیک سهام رشدی با ریسک سیسـتماتیک بـازار، بیشـتر از سهام قیمتی است. در گروه‌های میانی و توسعه، این ارتباط برای ریسک سیسـتماتیک سـهام قیمتـی بیشـتر از سهام رشدی است و در گروه اوج نیز هیچگونه ارتبـاط معنـی داری بـین ریسـک سیسـتماتیک بـازار و ریسـک سیستماتیک سهام رشدی و قیمتی وجود ندارد.

با توجه به پیشینه تحقیقات متوجه می‌شویم که بر خلاف تعداد زیاد مطالعات صورت گرفته پیرامون بهینه سازی پرتفو، موضوع خوشه‌بندی کمتر مورد بررسی قرار گرفته است.

مدل پژوهش و متغیرهای آن

داده کاوی فرایند محاسباتی کشف الگوها در مجموعه‌های بزرگ دادگان است که شامل شیوه‌های مرتبط با مباحث یادگیری ماشین، آمار و سیستم‌های پایگاه است[۲]. هدف کلی فرایند داده کاوی استخراج اطلاعات از یک مجموعه داده‌ای و تبدیل آن به یک ساختار قابل فهم برای استفاده بیشتر است[۳].

خوشه‌بندی یکی از مهم‌ترین مسائل در حوزه یادگیری بدون ناظر است که هدف آن کشف ساختار درونی داده‌های موجود بدون وجود برچسب است.

آنالیز خوشه عمل دسته‌بندی اعضای یک جمعیت به منظور کشف ساختار دادگان است. آنچه مورد نظر است این است که اعضای یک خوشه به یکدیگر شبیه بوده و نزدیک به یکدیگر قرار داشته باشند در حالیکه متفاوت از اعضای دیگر گروه‌ها باشند. در واقع خوشه بیانگر مجموعه‌ای از داده‌هاست که بیشترین شباهت را از میان کل دادگان موجود به یکدیگر دارند (A. R. Webb, Statistical Pattern Recognition, John Wiley & Sons,2003)

مدل‌های خوشه‌بندی عموما شامل مدل‌های زیر می‌باشند:

- مدل‌های مبتنی بر اتصال

  برای مثال خوشه‌بندی سلسله مراتبی مدلهایی بر اساس اتصال فاصله‌ای می‌سازد.
- مدل‌های مبتنی بر مرکز

  برای مثال، الگوریتم k-means هر خوشه را توسط یک بردار میانگین نمایش می‌دهد.
- مدل‌های مبتنی بر توزیع

  خوشه‌ها با استفاده از توزیع‌های آماری، همچون توزیع‌های نرمال چند متغیره با استفاده از الگوریتم بیشینه کردن امید ریاضی مدل می‌شوند.
- مدل‌های مبتنی بر چگالی

  برای مثالDBSCAN[۴] وOPTICS[۵] خوشه‌ها را به عنوان نواحی چگال متصل در فضای داده‌ای تعریف می‌کنند.
- مدل‌های مبتنی بر گراف

  یک clique[۶] که زیر مجموعه‌ای از گره‌ها در یک گراف است به گونه‌ای که دو گره در زیرمجموعه توسط یک یال می‌توانند به عنوان یک نمونه اولیه از خوشه در نظر گرفته شوند.

در این تحقیق از روش خوشه‌بندی سلسله مراتبی و افرازبندی استفاده شده است. روش افراز بندی به دو روش سخت و فازی (نرم) تقسیم می‌شود. در روش پارتیشن بندی سخت هر شی از مجموعه داده‌ها باید به دقت به یک خوشه اختصاص داده شوند مانندK-Mean . همچنین در یکی از انواع تقسیم‌بندی‌ها روش‌های خوشه‌بندی به دو دسته سلسله‌مراتبی و افرازبندی تقسیم می‌شوند.

الگوریتم خوشه‌بندی K-means

الگوریتم K-means اولین بار توسط مک کوئین در سال ۱۹۷۶ ارائه شد. این الگوریتم برای خوشه‌بندی مجموعه داده‌های بزرگ بسیار کارآمد است و سریع ترین راه خوشه‌بندی برای حجم وسیعی از داده‌ها محسوب مـی‌شـود. در این الگوریتم قانون کلی برای تعداد بهینه خوشه‌ها وجود ندارد و تعداد خوشه‌ها بستکی بـه مشـکل و مسـئله مورد نظر دارد. الگوریتم K-means نیاز به سه پارامتر مشخص شده دارد:

۱) تعداد خوشه K
۲) مقدار اولیه خوشه
۳) فاصله متریک (داده‌های عددی)

الگوریتم خوشه‌بندی فازی

خوشه‌بندی فازی (که تحت عنوان خوشه‌بندی نرم نیز شناخته می‌شود) به دسته‌ای از خوشه‌بندی‌هـا تعلـق دارد که در آن‌ها هر یک از نقاط دادگان می‌تواند به بیش از یک خوشه تعلق داشته باشد. در خوشه‌بنـدی غیـر فـازی (که با عنوان خوشه‌بندی سخت نیز شناخته می‌شوند) داده‌ها به خوشه‌های مجزا تقسیم شده که در آن هر نقطـه دادگان می‌تواند تنها به دقیقا یک خوشه متعلق باشد. این در حالی ست که نقاط دادگان می‌توانند به طور بالقوه به چندین خوشه تعلق داشته باشند. در اینگونه الگوریتم‌هـا درجـه‌هـای عضـویت بـه هـر یـک از نقـاط دادگـان تخصیص داده می‌شود. بنابراین نقاط موجود در مرزهای یک خوشه احتمالا با درجه کمتری از نقـاط مرکـزی در خوشه قرار دارند.

ارزیابی مدل خوشه‌بندی

ارزیابی مدلهای خوشه‌بندی در سه موضوع زیر مورد بررسی قرار می‌گیرد[۷].

۱) ارزیابی جهت‌دار بودن خوشه‌ها:

ارزیابی جهت‌دار بودن خوشه‌ها تعیین می‌کند که داده‌ها ساختار غیرتصادفی دارند. در غیر این‌صورت حتـی اگـر الگوریتم چند خوشه را به عنوان خروجی به ما بدهد این خوشه‌ها معتبر و معنـی‌دار نخواهنـد بود.آمـاره‌هـاپکینز شاخصی در آمار است که تصادفی بودن مقدار متغیرها را می‌آزماید. این سنجش بر اساس یک آزمون فرض است که به صورت زیر تعریف شده است:

- فرض صفر: داده‌ها به صورت یکنواخت توزیع شده‌اند و هیچ خوشه معنی‌داری در آنها وجود ندارد.
- فرض یک: داده‌ها به صورت یکنواخت توزیع نشده‌اند و در آنها خوشه معنی‌دار مشاهده می‌شود.

این آزمون را می‌توان با تکرارهای مختلف انجام داد. در صورتی که آماره آزمون بزرگتـر از ۰٫۵ باشـد مـی‌تـوانیم فرض یک را رد کنیم و بپذیریم که داده‌ها به صورت یکنواخت توزیع شده‌اند.

۲) ارزیابی تعداد خوشه‌ها

تعیین درست تعداد خوشه‌ها نقش مهمی‌در الگوریتم‌های خوشه‌بندی ایفا می‌کند به این علت که دانه‌بندی تحلیل خوشه‌ای را مشخص می‌کند. تعیین تعداد خوشه‌ها را می‌توان به صورت موازنه‌ای بین قابلیت فشرده‌سازی و دقت تعریف کرد. برای توصیف این دو مفهوم می‌توان دو حالت را در نظر گرفت. در حالت اول کل داده‌ها در یک خوشه جای می‌گیرند. در این حالت قابلیت فشرده‌سازی در وضعیت بیشینه است ولی خوشه مفهومی‌ندارد. در حالت دوم هر شیء در داده‌ها در یک خوشه قرار می‌گیرد. در این حالت دقت در وضعیت بیشینه است زیرا بین هر شیء و مرکز خوشه فاصله صفر وجود دارد. در این حالت هم خوشه‌بندی مفهومی‌ندارد. بنابراین در تعیین تعداد خوشه‌ها باید موازنه بین این دو شاخص در نظر گرفته شود.

۳) ارزیابی کیفیت خوشه‌ها

ارزیابی یا اعتبار سنجی خوشه‌بندی به سختی خود خوشه‌بندی است[۸]. شیوه‌های معمول شامل ارزیابی "داخلی" که در آن خوشه‌بندی در یک امتیاز کیفیتی واحد خلاصه می‌شود، ارزیابی "خارجی" که در آن خوشه‌بندی با یک دسته‌بندی "زمینه حقیقی " مقایسه می‌شود، ارزیابی "دستی" توسط فرد خبره، و "ارزیابی غیر مستقیم" توسط ارزیابی سودمندی خوشه‌بندی در کاربرد مورد نظر هستند. معیارهای ارزیابی داخلی این مشکل را دارند که نماینده توابعی هستند که خود می‌توانند به عنوان تابع هدف خوشه‌بندی در نظر گرفته شوند. به عنوان مثال می‌توان مجموعه دادگان را توسط ضریب سیلوئت خوشه‌بندی کرد هر چند که هیچ الگوریتم کارامد شناخته شده‌ای برای این کار وجود ندارد. با استفاده از این چنین معیارهای داخلی برای ارزیابی ما ترجیحا شباهت مسائل بهینه سازی و نه صرفا چگونگی سودمندی خوشه‌بندی را مقایسه خواهیم کرد[۹].

ارزیابی خارجی نیز مشکلات مشابهی دارد: در صورتی که ما چنین برچسب‌های "زمینه حقیقی" ای داشته باشیم پس به خوشه‌بندی نیازی نیست و در کاربردهای عملی ما اغلب چنین برچسب‌هایی نداریم. از سوی دیگر، برچسب‌ها تنها بازتاب افراز ممکن از مجموعه دادگان هستند که دال بر این موضوع نیستند که یک خوشه‌بندی متفاوت و حتی بهتر وجود ندارد.

بنابراین در نهایت هیچ یک از این شیوه‌ها نمی‌توانند کیفیت حقیقی یک خوشه‌بندی را ارزیابی کنند بلکه به ارزیابی انسانی نیاز است که بسیار وابسته به طرز تفکر شخص است. با این وجود چنین ارقامی‌می‌توانند کاملا حاوی اطلاعات مفید در شناسایی خوشه‌بندی‌های بد باشند، با این حال نباید ارزیابی انسانی را رد کرد[۱۰].

کیفیت خوشه‌ها در روشهای یادگیری بدون نظارت از طریق روشهای ارزیابی درونی انجام می‌شود. این روشها ارزیابی می‌کنند که خوشه‌ها تا چه حد از هم جدا هستند و تا چه حد به هم فشرده‌اند.

۳- روش شناسی پژوهش

در پژوهش حاضر، شاخص‌های انتخابی برای خوشه‌بندی سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران از طریق مطالعه مقالات بین‌المللی و بررسی شاخص‌های مشابه، همچون S&P و Russell گردآوری شده و پس از گفتگو با ۱۲ نفر از خبرگان در این زمینه (مدیران و کارشناسان سازمان بورس، اساتید رشته

مدیریت مالی و حسابداری)، ۵ شاخص نهایی برای خوشه‌بندی انتخاب گردید. با توجه به مبانی نظری پـژوهش و اجماع گروهی خبرگان، متغیرهای تحقیق برای خوشه‌بندی سهام به شرح ذیل می‌باشند:

۱) نسبت قیمت به درآمد هر سهم :(P/E)عبارت است از نسبت قیمت به درآمد هر سهم کـه نشـان‌دهنـده مدت زمانی است که برای بازگشت اصل سرمایه‌گذاری از محل عایدات آتی سهم مورد نیاز است.
۲) شاخص ریسک سیستماتیک :(β)ریسک سیستماتیک نشان‌دهنده آن بخش از کل ریسک مجموعه سهام است که به علت وجود عواملی کع کل سهام موجود در بازار را تحت تاثیر قرار می‌دهند، به وجود آمده و قابل کاهش نیست. از عوامل مهم ریسک سیستماتیک تحولات سیاسی و اقتصادی، چرخه تجاری، تورم و بیکاری است.(قائمی، محمدحسین و طوسی، سعید، ۱۳۸۴)
۳) سود هر سهم :(EPS)عبارت است از سود خالص پس از کسر مالیات، تقسیم بر تعداد سهام شرکت.
۴) بازدهی :(Return)به عنوان مهمترین پارامتر در مدیریت پرتفو که با اسـتفاده از قیمـت سـهام محاسـبه می‌گردد.
۵) نسبت ارزش بازار به ارزش دفتری:(M/B) نسبت ارزش بازار به ارزش دفتری فاصله ارزش بازار هر سهم را نسبت به ارزشی که در فاتر حسابداری آن شرکت ثبت شده است، نشان می‌دهد.

به منظور قابل انجام بودن عمل خوشه‌بندی و محاسبه فاصله درست میان سـهام، داده‌هـا بـر اسـاس معیارهـای اصلی (P/E و M/B) فیلتر شده و تنها سهام حاوی این ویژگی‌ها برای امر خوشه‌بندی در نظر گرفته شدند. لازم به ذکر است؛ مقادیر این ویژگی‌ها در تمامی‌دوره‌های سه ماهه مورد فیلتر قرار گرفتند.

خوشه‌بندی به روش K-means

در هر سال خوشه‌بندی به ازای ۲ خوشه، ۳ خوشه، ... و ۱۳ خوشه انجام می‌گیرد. یعنی تعداد خوشـه‌هـایی کـه داده‌های مربوط به هر سال به آن‌ها تقسیم می‌شوند ۲ تا ۱۳ خوشه می‌باشد. پیش از انجـام عمـل خوشـه‌بنـدی مقادیر هر یک از شاخص‌ها نرمال شده و برای محاسبه فاصله بهینه می‌شوند. این نرمال سازی بدین صورت است که کمترین مقدار هر یک از ویژگی‌ها از مقادیر آن ویژگی کاسـته شـده و سـپس بـه یـک مقـدار ثابـت تقسـیم می‌شود. این مقدار ثابت به ازای هر ویژگی اختلاف مطلق حداکثر و حداقل مقدار آن ویژگی است. فرمول نرمـال سازی به شرح ذیل می‌باشد.

$$x_{ij} = \frac{max_j - x_{ij}}{max_j - min_j}$$

پس از اجرای کد خوشه‌بندی افرازبندی، دادگان را به صورت زیر خوشه‌بندی می‌نماید.

```
Browse[2]> clusters
K-means clustering with 13 clusters of sizes 4, 3, 219, 7, 4, 1, 23, 15, 26, 26, 1, 1, 39

Cluster means:
        P/E          M/B        EPS       RETURN          BETA      P/E.1       M/B.1     EPS.1    RETURN.1     BETA.1       P/E.2       M/B.2     EPS.2
1  5.721990     714.9408   527.2500  0.66189142  0.070253948 12.565497    514.3102  285.7500   9.5550198 -0.1242056   12.5472823     0.00000 9318.0000
2 14.296673     636.2726  1375.0000 -0.29733611  1.008001822 12.440693    136.2741  254.0000   3.6460138  1.3090771 1225.9185604 699375.00000  163.0000
3 22.182592     589.7394   454.6301  0.30068586  0.402611343 21.039238    632.8258  384.4475   3.8707314  0.7823693   62.2618036   332.19791  150.2009
4  6.740333     483.6553   173.4286  2.13865958  0.657962624 11.192173    647.1862  301.2857   4.4764475  1.4405057    7.4998404 18991.56052  186.4286
5 10.376951   42169.5670   995.7500  3.42541775  1.205747534  9.861718  33108.9533  746.2500   5.1171209  0.3529441    7.1038486  8323.38522  536.5000
6 11.570175 -454827.5862   608.0000 -0.07786344 -0.707857616 11.517544   -1724.2285   46.0000 -11.0322649  1.7921523   27.6346154    51.19882 -203.0000
7  9.384407    4015.2017   487.9130  2.50371494  0.580385327 18.504186    6537.5312  665.6522   6.8886943  1.4999685   10.0875445  3087.24261  176.1304
8 14.756024   18786.7527   456.0667  2.36953154  0.938447103  7.966461  17131.2602 1052.9333   8.9333621 -0.6186988    5.8795335  8080.82484 1291.8000
9 14.350161     505.6422  3206.6154  1.48761239  0.375667509 12.723711    428.5056  501.1923   3.7208155  0.6422514    9.9178810   307.93781  312.5385
10 75.622806   -136.0779 -1114.0000  1.12344130  0.461383106  9.820585   -115.8758 -707.1154   7.9642956  0.5125368   20.5049157    26.19879  169.0000
11  9.120588   86189.6054    61.0000  1.88216068 -0.006956663 10.256637 130249.6038  441.0000   0.6030205  1.5905823    0.6700201     0.00000   82.0000
12 30.235294   -8691.0594   486.0000 -0.07786344  1.585188932 38.425926 -71393.4959  337.0000  15.2379169  2.2351673   11.4231434   108.40846  185.0000
13  7.478637     427.6973   411.9744  1.22725175  0.468132872  7.639791    390.6708 2344.2308   2.2334206  1.0772547    8.0143857   404.95904  398.0513
     RETURN.2      BETA.2      P/E.3        M/B.3     EPS.3     RETURN.3      BETA.3
1   5.2335950   2.2937885   6.293671    767.71481  285.7500  1.96040296  0.05147008
2   5.0622410  13.9134377   7.176998   3918.84579  254.0000  0.06148412  0.51629198
3   0.1170205   1.6298750  20.022528    692.57552  428.3425  4.46479680  0.33839424
4   1.7112070   8.1716357 102.497194    638.08603  280.8571  9.88816475 -0.39248560
5  -0.8021704   6.0043205   7.032867  28006.97399  746.2500  2.75630330  0.64790100
6  -5.1984193  -0.8278542  11.263158   -363.31739 -343.0000 -0.68920178  2.23502192
7   3.6165669   6.5526608  13.263109   9006.53627  671.2174 26.22623728  0.51897474
8   2.6084129   5.7404641   6.781750  15576.51204 1305.7333  1.39700084  0.13774505
9   0.7406372   2.7791069  10.951897    315.25954  575.2308 -0.44617439 -0.05601581
10  9.7412958  -9.1173839  37.419176    -14.98398 -686.3077  5.88835164 -1.10891272
11  4.0921490   8.3641060   9.049587 121403.40684  441.0000  1.61772015 -0.97856599
12 14.7415449 -18.7454612  24.496063    368.48040  337.0000  7.96935734 -2.28802914
13  0.7795452   1.5743308   6.525232    379.66007 2282.0513 14.85720302 -0.31359856

Clustering vector:
  [1]  3  9  9  3  9 13 10  3 10  3  3  8  3  3 13  3  3  3  9  3  9 13  8  3  9 10  3 13  3 13  8  7  7  7  8  3  3  7  7  8  8  8  5  7  7  2 13  3  3
 [50]  9  3  3  3  3  3  3  3 10  3  3 10  3  3  3  3  3  3  3  3  3  3  3  3  7  3  3  9  7  3  3  3 13 13  3  3  9 13 13  9  9 13  3  3  3 10  3  3  3
 [99]  7  3  3  3  3  3  9  3 13 13  3  8  4 10  3 13  3  3  3  3  3  8  4  3  3  3  3  1  3  7  7  3  3  7  3  3  3  3  3  3  8  4  3  8  3  3  3  3  3
[148]  3  3  3  3  3  3  3  3  3  3 10  3 13  9 13  3  9 13  3  9  3  3  1 13  3  3  3  3  3  3  9 13  3  3 13  3 13  9 13  3  9  3  3 13  3  3 13
[197]  3  3 10 10  3  3  3  3 13 10 13 13  3  3  3  3  8  3  3  3  1  3  3  3  6  3  9  3  3 10  4  3  3  3  3  3  7  5  4  7  3  4 13 10 10 10  3 10 13
[246]  9 13  3  3  3  3  3  3  3  3  3  9  3  3  3  3  3  3  3  3  3 11 10  3  3 13  2 13  9 13  3 12  8  3  3  3 13  3  3 13 10  3 13  3  3  3  3
[295]  3  3  3  3 13  3  1  3  3  9 13  3  3  3  7  3  9  3  3  7  7  7 10  3  7  3  3  3  7  2  8  3  3  3  3  3  3 10  3 10  5 10 10 10  3  9  3  3
[344]  3  3  3  3  3  3  3  3  9  3  3  7 10  3  3  3 10  3  3  8  3  3  7  5 13  4

Within cluster sum of squares by cluster:
 [1]    5964265   51464577  913720114  914159402 2400591280          0 1532305455 2570362422  132327658  221500280          0          0  159105098
 (between_SS / total_SS =  99.5 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"    "size"         "iter"         "ifault"
```

شکل ۱خروجی کد خوشه‌بندی به روش k-means

خوشه‌بندی به روش سلسله مراتبی

روش‌های تعیین فاصله که در تابع خوشه‌بندی سلسله مراتبی با استفاده از زبان برنامه نویسی R وجود داشته و مورد استفاده قرار گرفتند شامل توابع زیر بوده‌اند:

۱) Euclidean؛

۲) Maximum؛

۳) Manhattan؛

۴) Canberra؛

۵) Binary؛

۶) Minkowski؛

خوشه‌بندی با استفاده از تمام روش‌های تعیین فاصله فوق‌الذکر در کد R صورت گرفت و نتایج به منظور تشکیل پرتفو ذخیره شدند.

پس از اجرای کد خوشه‌بندی سلسله مراتبی، تابع خوشه‌بندیR ، متغیری حاوی اطلاعات یک سلسـله مراتـب را ایجاد می‌کند. مولفه‌های این متغیر شامل ارتفاع هر سهم در این سلسله مراتب و ترتیب مناسـب جهـت نمـایش این سلسله مراتب هستند.

تشکیل پرتفوی بهینه

پس از آنکه نتایج خوشه‌بندی از هر یک از روش‌های افرازبندی و سلسله مراتبی حاصل شد برای تشکیل پرتفوی بهینه از خوشه‌ها نیاز به محاسبه بازدهی خوشه داریم. برای بدست آوردن بازدهی خوشه i از فرمول زیر اسـتفاده می‌شود.

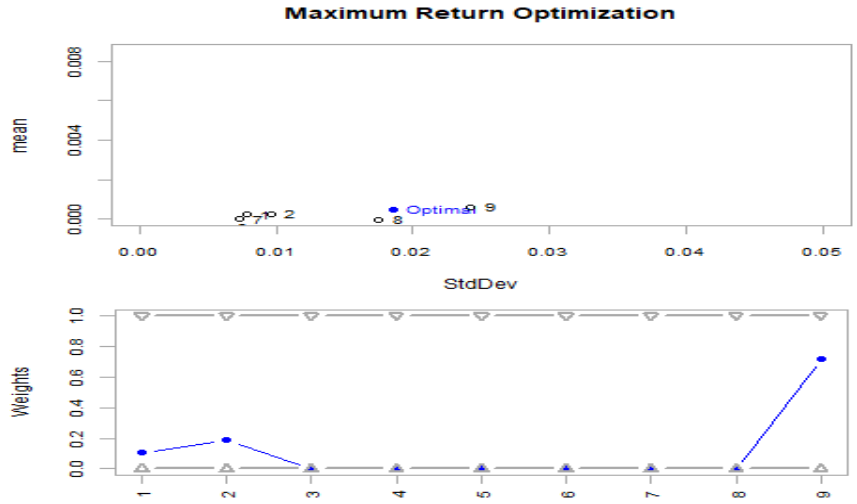
$$\mu_i = \frac{w_1\mu_1 + w_2\mu_2 + \cdots + w_n\mu_n}{w_1 + w_2 + \cdots + w_3}$$

در این فرمول w_j وزن سهم j ام در خوشه i بوده و به روش زیر بدست می‌آید:

$$w_j = \frac{M_j}{M_1 + M_2 + \cdots + M_n}$$

که در اینجا M_k ارزش بازار سهم k ام است.

به بیان دیگر بازدهی هر یک از خوشه‌ها میانگین موزون بازدهی سهام موجود در آن خوشه است که وزن‌های هر یک از آن سهام ارزش بازار آن سهم در آن خوشه است.



شکل ۲ خروجی حاصل از تشکیل پرتفوی بهینه به ازای ۹ خوشه

در شکل فوق خروجی حاصل از تشکیل پورتفوی بهینه بـه ازای ۹ خوشـه آورده شـده اسـت. در نمـودار بـالایی موقعیت هر خوشه با توجه به انحراف معیار و میانگین بازدهی آن خوشه مشخص شده است کـه در آن پورتفـوی بهینه با رنگ آبی از سایر خوشه‌ها متمایز شده است.

نحوه ارزیابی

برای ارزیابی نتیجه خوشه‌بندی در حوزه خوشه‌بندی سهام بازار بازدهی، ریسک و نسبت بازدهی بـه ریسـک بـه ازای هر یک از سال‌ها و به ازای هر یک از تعداد خوشه‌ها محاسبه می‌شود.

مقایسه پرتفوهای بهینه به دست آمده با پرتفوی شاخصی

ارزیابی عملکرد پرتفوی، آخرین مرحله در فرآیند مدیریت سرمایه‌گذاری اسـت. آن بـه مثابـه یـک سـازوکار بازخوردی و کنترلی به منظور اثربخش‌تر کردن فرآیند مدیریت سـرمایه‌گـذاری مـی‌باشـد. در ارزیـابی عملکـرد پرتفوی باید توجه داشت که عملکرد به صورت نسبی و نه به صورت مطلق، مورد ارزیابی قرار می‌گیرد لـذا بـرای امر مقایسه نیازمند یک شاخص یا مبنا هستیم. به علاوه یکی از مشکلات اصلی در ارزیابی عملکرد، تمایل انسانی به تمرکز بر بازده پرتفوی و عدم توجه کافی به ریسک متحمل شده برای کسب بازده مورد نظر است. لذا ارزیابی عملکرد پرتفوی بایستی شامل شناسایی همزمان بازده و ریسک سرمایه‌گذاری باشد(راعی و پویان‌فر، ۱۳۸۹).

تاکنون معیارهای بسیاری با توجه به این امر برای اندازه‌گیری عملکرد پرتفوی معرفی و آزمون شده‌اند. ایـن معیارهای عملکرد تعدیل شده بر اساس ریسک امکان مقایسه عملکرد پرتفوی را نسبت به بازار(شاخص بـازار) و سایر پرتفوها فراهم می‌کنند.

شاخص شارپ یکی از شاخص‌های سنتی مهم برای ارزیابی عملکرد صندوق‌های مشترک است که نسبت بـه ریسک تعدیل شده است. این شاخص بیان‌کننده مقدار مازاد بازدهی است که صندوق توانسته اسـت نسـبت بـه بازده حاصل از یک دارایی بدون ریسک کسب کند. در این شاخص بازده دارایی بدون ریسک به این سبب مـورد توجه قرار گرفت است که بتوان ملاحظه نمود آیا برای سرمایه‌گذار ریسک مطلوب‌تر است؛ زیرا بازده تعدیل شده بیشتری در ازای هر واحد ریسک به دست می‌آید.

شارپ برای منظور نمودن عنصر ریسک و خنثی کردن اثر ریسک بر میزان بازده، مـازاد بـازده کسـب شـده نسبت به بازده بدون ریسک را بر ریسک بازده تقسیم کرد. منظور از ریسک در نسبت شارپ ریسک کل است که از دو جزء ریسک سیستماتیک و غیرسیستماتیک تشکیل شده است. ریسک مورد نظـر شـارپ در ایـن شـاخص مبتنی بر تئوری مدرن پرتفوی و با فرض نرمال بودن توزیع بازده پرتفوهای به دست آمده از خوشه‌هـا مـی‌باشـد که با به دست آوردن انحراف معیار بازده هر پرتفو در دوره مورد بررسی مشخص می‌شود. این شاخص از طریـق رابطه زیر محاسبه می‌گردد:

$$l_{sharp;j} = \frac{\overline{R} - R_f}{\sigma_j}$$

$l_{sharp;j}$ نسبت شارپ پرتفوی j؛ $\overline{R}$ بازده متوسط پرتفوی j را در طول دوره؛ R_f بازده متوسط بدون ریسک در طول دوره و σ_j انحراف معیار بازده پرتفوی j را نشان می‌دهد.

بازده بازار به عنوان معیاری جهت ارزیابی عملکرد پرتفوهای به دست آمده پس از خوشه‌بندی، استفاده می‌شود. در جدول زیر میانگین بازده روزانه بازار که از طریق مقادیر شاخص کل روزانه به دست آمده؛ نشان داده شده است.

جدول ۱ متوسط بازدهی بازار

سال	متوسط بازده روزانه بازار
1390	0.0005
1391	0.0016
1392	0.0031
1393	-0.0009
1394	0.001
1395	-0.0001

لذا بر اساس بازدهی‌های فوق و شاخص شارپ، ارزیابی پرتفوهای به دست آمده از خوشه‌بندی به روش افرازبندی به شرح ذیل می‌باشد.

جدول ۲ نسبت شارپ برای پرتفوی به دست آمده از سهام خوشه‌بندی شده به روش K-means

تعداد خوشه‌ها	1390	1391	1392	1393	1394	1395	میانگین نسبت شارپ طی سالها
	نسبت شارپ	نسبت شارپ	نسبت شارپ	نسبت شارپ	نسبت شارپ	نسبت شارپ	
3	0.33692	(۶,۱۹۲۵۲)	(۱,۸۰۹۸۶)	3.69478	2.94551	1.15529	0.02169
4	1.03377	25.81412	(۰,۸۰۳۳۷)	22.60982	12.48387	1.36961	10.418
5	1.81201	(۱,۶۶۸۹۵)	0.75532	3.52823	9.14623	1.46736	2.50670
6	0.83042	6.45283	0.16532	0.92308	5.60405	4.14484	3.02009
7	0.10949	0.24228	0.48361	0.83539	4.01033	3.61872	1.54997
8	0.19226	1.40179	0.50000	0.83478	2.54462	3.50251	1.49599
9	1.17640	3.00877	0.42669	0.78437	1.89736	3.07647	1.72834
10	0.57119	2.69231	0.33871	1.05463	1.72622	2.06176	1.40747
11	0.77225	1.85606	0.29392	0.70644	1.48018	1.92890	1.17296
12	0.49815	1.75000	0.47246	0.99796	1.46788	2.36889	1.25922
13	2.61921	0.71742	0.52827	1.37393	1.43321	1.81270	1.41412

با توجه به نتایج حاصل از محاسبه نسبت شارپ طی سال‌های گذشته، زمانی که با استفاده از روش خوشه‌بندی K-means، تعداد خوشه‌ها بین چهار تا شش عدد باشد، نسبت شارپ مثبت و بالاتر از زمان خواهد بود که تعـداد خوشه‌ها کمتر از چهار و یا بیشتر از شش باشد. لذا عملکرد پرتفوهای به دست آمده از طریق خوشه‌بندی بازار به چهار تا شش خوشه، به صورت معناداری بهتر از عملکرد بازار بوده است.

لذا بر اساس بازدهی‌های فوق و شاخص شارپ، ارزیابی پرتفوهای به دست آمده از خوشه‌بندی بـه روش سلسـله مراتبی به شرح ذیل می‌باشد.

جدول۳ نسبت شارپ برای پرتفوی به دست آمده از سهام خوشه‌بندی شده به روش سلسله‌مراتبی

تعداد خوشه‌ها	1390	1391	1392	1393	1394	1395	میانگین نسبت شارپ طی سالها
	نسبت شارپ	نسبت شارپ	نسبت شارپ	نسبت شارپ	نسبت شارپ	نسبت شارپ	
3	2.7087	-12.0972	5.917105287	1.95419	34.5822	6.624204	6.61486
4	3.05892	-9.79513	6.128392084	1.52619	10.9157	3.931947	2.62767
5	1.8252	-11.072	4.692671395	3.12778	3.81818	2.214286	0.76769
6	1.37023	39.1187	4.748803828	2.14721	2.07879	2.45829	8.65367
7	1.33215	24.1087	3.259441708	1.92821	1.98457	2.663311	5.87939
8	1.20386	15.5408	2.615283267	1.80576	1.86244	2.464803	4.24883
9	1.13482	7.768	2.777932961	1.74703	1.66775	1.736689	2.80537
10	1.09001	4.69082	2.275743707	1.68274	1.34422	1.596915	2.11341
11	1.03189	2.83188	2.120469083	1.3193	1.18811	1.061525	1.5922
12	0.96504	2.59783	1.862359551	1.17663	1.04944	0.751816	1.40052
13	0.77621	2.35468	1.969306931	0.94657	1.04359	0.567716	1.27635

با توجه به نتایج حاصل از محاسبه نسبت شارپ طی سال‌های گذشته، زمانی که با استفاده از روش خوشه‌بنـدی سلسله مراتبی، سهام در بازار طبقه‌بندی شده سپس در سبد سرمایه‌گـذاری بهینـه‌سـازی مـی‌گردنـد، عملکـرد اکثریت پرتفوهای به دست آمده از پرتفوی شاخصی به صورت معناداری بیشتر بوده است. لذا به نظـر مـی‌رسـد؛ خوشه‌بندی به نوبه خود به بهینه‌سازی و عملکرد بهتر نسبت به چرتفوی شاخصی کمـک نمـوده و در ایـن بـین خوشه‌بندی به روش سلسله‌مراتبی نسبت به روش خوشه‌بندی به روش افرازبندی عملکرد بهتری داشته است.

۴- فرضیات پژوهش

در تحقیق حاضر فرضیات ذیل مورد بررسی قرار گرفته است:

۱) بازده پرتفوی بهینه به دست آمده از الگوریتم‌های خوشه بندی تفکیکی بیشتر از بازده پرتفوی شاخصـی است.

۲) بازده پرتفوی بهینه به دست آمده از الگوریتم‌های خوشه بندی سلسله مراتبی بیشــتر از بـازده پرتفـوی شاخصی است.

۵- نتیجه‌گیری و بحث

در این تحقیق که با هدف تشکیل سبد سرمایه‌گذاری بهینه در بورس اوراق بهادار تهران با استفاده از خوشه بندی و یا طبقه بندی سهام به دو روش تفکیکی و سلسله مراتبی انجام شد، شرکت‌های همگن بر اساس هر یک از شاخص‌های مورد بررسی شناسایی شدند. بدین شکل هر خوشه شامل چندین شرکت است و هر شرکت دارای ضریب عضویتی است که میزان و شدت تعلق آن شرکت را به خوشه مزبور بیان می‌دارد. پس از خوشه بندی‌های صورت گرفته، با استفاده از روش مارکویتز سبد سهام بهینه از میان خوشه‌ها و بر اساس اوزان مشخص تعیین شد. سپس نسبت شارپ به منظور بررسی عملکرد پرتفوها نسبت به پرتفوی شاخصی محاسبه گردید. نتایج تحقیق نشان داد که در بازه مشخص زمانی، بهینه سازی پرتفو بر اساس سهام خوشه بندی شده به روش سلسله مراتبی و تفکیکی نسبت به پرتفوی شاخصی از عملکرد بهتری برخوردار بوده است. همچنین در میان این دو روش خوشه بندی پرتفوهای به دست آمده پس از خوشه بندی سهام به روش سلسله مراتبی نسبت به پرتفوهای به دست آمده پس از خوشه بندی سهام به روش تفکیکی، عملکرد بهتری داشته اند.

فهرست منابع

* احمدپور، ا. و دیگران، (۱۳۸۸).«استفاده از مدل‌های تصمیم‌گیری چند شاخصه‌ای در انتخاب سهام(شرکت‌های دارویی پذیرفته شده در بورس اوراق بهادار تهران)»، فصلنامه بورس و اوراق بهادار تهران، سال ۲، شماره ۵.
* اسلامی‌بیدگلی غلامرضا، هیبتی فرشاد، رهنمای رودپشتی فریدون،(۱۳۸۸)، «تجزیه و تحلیل سرمایه‌گذاری و مدیریت سبد اوراق بهادار» چاپ سوم، تهران: انتشارات پژوهشکده امور اقتصادی، ترجمه کتاب: Investment Analysis & Portfolio Management تألیف Frank k. reilly/Keith.C.Brown
* افشارکاظمی، محمدعلی، فلاح شمس، میرفیض، کارگز، مرضیه، (۱۳۹۲).«تدوین مدلی جدید برای بهینه‌سازی پرتفوی بورس با استفاده از روش مارکویتز و اصلاح آن توسط مدل کسینوس‌ها و حل آن توسط الگوریتم ژنتیک»،مجله مهندسی مالی و مدیریت اوراق بهادار، شماره هجدهم، بهار ۱۳۹۳.
* امیری،م. و دیگران،(۱۳۸۱).«کاربرد مدل فرآیند تحلیل سلسله‌مراتبی در تعیین معیارهای موثر بر انتخاب سهام در بورس اوراق بهادار تهران»، مجله برنامه و بودجه، شماره ۷۷.
* انواری رستمی، علی اصغر و محسن ختن لو(۱۳۸۵).« بررسی مقایسه ای رتبه بندی شرکت‌های برتر بر اساس نسبت‌های سودآوری و شاخص‌های بورس اوراق بهادار تهران»، بررسی‌های حسابداری و حسابرسی، ش ۴۳، صص ۲۵-۴۳.

* پهلوان، آریا، رمضانپور، اسماعیل، قلیزاده، محمدحسن(۱۳۹۱).« اولویت‌بندی عوامل موثر بر انتخاب سهام در بورس اوراق بهادار تهران با استفاده از فرآیند تحلیل شبکه‌ای فازی»، سومین کنفرانس ریاضیات مالی و کاربردها.
* تهرانی، رضا و سعید باجلان. (۱۳۸۸).« بررسی رابطه بین ویژگی‌های شرکت و موفقیت مال»، پژوهشنامه مدیریت اجرایی، ش۱، صص ۱۳-۱۸.
* ثابتی، صالح، الهام(۱۳۸۸).« ارائه مدل تصمیم گیری چندمعیاره فازی برای رتبه بندی شرکتهای متقاضی تامین مالی بانک‌ها(مطالعه موردی: پنجاه شرکت برتر بورس اوراق بهادار تهران». مجموعه مقالات دومین کنفرانس بین المللی توسعه نظام تامین مالی در ایران، صص۱-۱۸.
* خدامرادی، سعید، ترابی گودرزی، محمد و راعی عزآبادی، محمدابراهیم،(۱۳۹۲).« رویکرد دو مرحله ریاضی در بهینه‌سازی سبد سهام». مجله مهندسی مالی و مدیریت پرتفوی.
* سپهر، ریحانه، محمدحسن مرادی، غنچه مشایخی، لاله کاردر و عطیه بامدادیان.(۱۳۸۶).«بررسی و مقایسه روش‌های مختلف خوشه بندی فازی تفکیکی مبتنی بر روش استاندارد خوشه بندی فازی FCM»، مجموعه مقالات هفتمین کنفرانس سیستم‌های فازی، دانشگاه فردوسی مشهد، صص۱-۵.
* طالبی، آرش،(۱۳۸۹) پایان‌نامه کارشناسی ارشد«انتخاب و بهینه‌سازی سبد سهام با استفاده از روش‌های فرا ابتکاری و مقایسه‌ی آن با سبدهای تشکیلی خبرگان و تازه‌کارها در بازار بورس اوراق بهاهدار تهران»، دانشکده مدیریت دانشگاه صنعتی شاهرود.
* عسگریان، احسان، حسین معین‌زاده، محسن سریانی و جعفر حبیبی.(۱۳۸۶).«رویکرد جدید برای خوشه‌بندی فازی به وسیله الگوریتم ژنتیک»، مجموعه مقالات سیزدهمین کنفرانس ملی انجمن کامپیوتر ایران، صص ۱-۳.
* فرید داریوش، پورحمیدی، مسعودی،(۱۳۹۱).«بخش‌بندی سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از تحلیل خوشه‌ای فازی»، مجله پژوهش‌های حسابداری مالی، سال چهارم، شماره سوم، شماره پیاپی(۱۳)، صص ۱۲-۱۰۵.
* فروغی، داریوش، مظاهری، اسماعیل،(۱۳۸۹)،« توانایی سود و جریان‌های نقدی عملیاتی در توضیح ارزش ذاتی تحقق یافته سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران»، دو فصلنامه پژوهش‌های حسابداری مالی،شماره۱-۲.
* قائمی، محمدحسین و سعید طوسی.(۱۳۸۴).«بررسی عومل موثر بر بازده سهام عادی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران». پیام مدیریت،ش۱۷ و ۱۸، صص ۶-۹.
* کیانی، مائده، نبوی، سیدعلی،معماریان، عرفان،(۱۳۹۳).«بهینه‌سازی سبد سهام بر اساس حداقل سطح پذیرش ریسک کل و اجزای آن با استفاده از روش الگوریتم ژنتیک».فصلنامه علمی‌پژوهشی دانش سرمایه‌گذاری، سال سوم، شماره یازدهم.
* Alexander S, Coleman T F, Li Y. 2006. Minimizing cvar and var for portfolio of derivatives. Journal of Banking and Finance, 30(2):pp.538-605.

* A. R. Webb, Statistical Pattern Recognition, John Wiley & Sons,2003
* Basalto, N., Bellotti, R., De Carlo, F., Facchi, P., & Pascazio, S. (2005). Clustering stock market companies via chaotic map synchronization, Physica A, 345 (1-2), 196–206.
* Beckwith, j., Stock Selection in Six Major Non –U.S. Markets. Journal of Investing. 7 (8), 29-44, 8001
* Bertsimas Dimitris, Christopher Darnell and Robert Soucy, (1999) "Portfolio construction Through Mixed-Integer Programming at Grantham, Mayo, Van,Otterloo and Company", ProQuest Science Journals, Interfaces 29 , pp
* Chen Y, Mabu SH, HirasawaK. 2011. Genetic relation algorithm whit guided mution for the large – scale portfolio optimization" Expert System Whit Applications ,An International journal , pp.3353-3363.
* Chi, Z., Yan, H., & Pham, T. (1996). Fuzzy algorithms: with applications to image processing and pattern recognition. Singapore: World Scientific Publishing Co. Pte. Ltd.
* D'Ursoa, P., & Giordani, P. (2006). A weighted fuzzy c-means clustering model for fuzzy data. Computational Statistics & Data Analysis, 50 (6), 1496 – 1523.
* Dobbins, Richard.; Witt. Stephen, F. & Fielding, John. (1994). Portfolio Theory and Investment Management. (2 nd ed). Massachusetts: Blackwell Publishers.
* Doherty, K. A. J., Adams, R. G., Davey, N., & Pensuwon, W. (2005). Hierarchical Topological Clustering Learns Stock Market Sectors, ICSC Congress on Computational Intelligence Methods and Applications (pp. 1-6), Istanbul.
* Edirisinghe, NCP & X Zhang, Portfolio selection under DEA-based relative financial strength indicators: case of US industries, Journal of the Operational Research Society, 57, 8002.
* Elton, J.; Edwin & Grubr, J.; Martin. (1995). Simple Criteria for Optimal Portfolio Selection, Journal of Finance, 31 (5). 1341-1357.
* Fan Cai, Nhien-An Le-Khac, M-Tahar K e chadi ,(2016). Clustering Approaches for Financial Data Analysis: a Survey,School of Computer Science & Informatics, University College Dublin, Ireland
* Faugère, C., Shawky, H. A., & Smith, D. M. (2005). Characterizing value and growth investing in institutional portfolios. Retrieved from University at Albany : http://www.albany.edu/~faugere/
* Grupe, F.H. & Jooste, S. (2004) Genetic Algorithms: A Business Perspective; Information Management & Computer Security, Vol. 12,o. 3, 289-298
* Guneri A.F. et al, A fuzzy approach to shipyard location selection, Expert Systems and Applications, 26, 9778-9777, 8007.
* J. Valente de Oliveira, J., & Pedrycz, W. (2007). Advances in fuzzy clustering and its applications. Chichester: John Wiley & Sons Ltd.
* Janani and et al., Selection of Portfolio by using Multi Attributed Decision Making (Tehran Stock Exchange), American Journal of Scientific Research, Issue 44, pp. 15-87, 8018
* Jose Figueira, Salvatore Greco and Matthias Ehrgott, MULTIPLE CRITERIA DECISION ANALYSIS: STATE OF THE ART SURVEYS, Springer, 8005
* Kalyani, S., & Swarup, K. S. (2010). Supervised fuzzy c-means clustering technique for security assessment and classification in power systems. International Journal of Engineering, Science and Technology , 2 (3), 175-185
* Konno H,Y amazaki H. 1991. Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. Management Science, 37(5):pp.519-531.
* Levy H, Sarnat M. 1984. Portfolio and investment Selection, Prentice Hall.
* Lhabitant, F. S. (2003). Evaluating hedge fund investments: the role of pure style indices. Retrieved from EDHEC-Risk Institute: http://www.edhec-

risk.com/edhec_publications/RISKReview1083060929573713739/attachments/Barry%20HF%20VaR1%2009.02.04.pdf
* Li, Jun, Jiuping Xu (2009) "A novel portfolio selection model in a hybrid uncertain environment",omega the International Journal of Management Science,
* Linsmeier T J, Pearson N D. 2000. Value at risk, Financial Analysts Journal,56(2):pp47-67.
* LN,L,Cao,L And Zhang,c,(2007),Genetec A lgorithm for Optimization in Financial Applications,University Of Sydney ,Australia.
* Markowitz H, 1952. Portfolio selection. Journal of Finance. 7: 77-91.
* Markowitz H. 1991. Foundations of portfolio theory. Journal of Finance, 46(2):pp.469-477
* Mitchel, M. (1999) An Introduction to Genetic Algorithms. Pacheco ,M,Vell asco,M, and lopez,c,(2006)cash flow planning AnD Optimization Through Genetic Algorithm Pontificia university,Brazill
* Morris, V. B., & Morris, K. M. (2007). Standard & Poor's dictionary of financial terms. New York: Lightbulb Press.
* Nanda, S.R., Mahanty, B., Tiwari, M.K. (2010). Clustering Indian stock market data for portfolio management. Expert Systems with Applications, 37 (12), 8793–8798.
* Nikoomaram H, Hemmati H, 2012. Network Model based on Sharp and Trainer Ratio Analytical Power Evaluation to Benchmark the performance of Selected Portfolio. Investment Knowledge Research and Science Quarterly. 1: 125-146.
* Ostermark R, 1996. A fuzzy control model (FCM) for dynamic portfolio management. Fuzzy Sets and Systems. 78: 243-254.
* Rashidi, P., Analoui, M. (2007). Modified k-means algorithm for clustering stock market companies. 1st Iran Data Mining Conference (pp. 201-21), Tehran: Amir Kabir University.
* Reilly F. et al. Investment Analysis and Portfolio Selection, 6d, ed., 8002.
* ROBERT JOHNSON, LUC SOENEN, Indicators of Successful Companies, European Management Journal Vol. 81, No. 2, pp. 264–267, 8002.
* Shin, H. W., Sohn, S. Y. (2004). Segmentation of stock trading customers according to potential value. Expert Systems with Applications, 27 (1), 27–33.
* S quyres .J.G. A Quick Peek According to Graham and Dodd, Journal of Financial Statement Analysis, 92-72, fall, 1772.
* Standard & Poor's (S&P) Shariah style indices: index methodology. (2009). Retrieved from Standard & Poor's Financial Services LLC: www.styleindices.standardandpoors.com
* Standard & Poor's (S&P) U.S. style indices: index methodology. (2009). Retrieved from Standard & Poor's Financial Services LLC: www.styleindices.standardandpoors.com
* Szilágyi, L., Szilágyi, S. M., & Benyó, Z. (2010). Analytical and numerical evaluation of the suppressed fuzzy c-means algorithm: a study on the competition in c-means clustering models. Soft Computing , 14 (5), 495–505.
* Talebnia Gh, Fathi M, 2010. Comparative Evaluation Optimal Stock Portfolio Selection in Tehran Stock Exchange by Markowitz and Value-at-Risk Models. Financial Study Magazine. 6: 71-94.
* Topaloglou N, Vladimirou H, Zenios SA, 2008. A dynamic stochastic programming model for international portfolio management. European Journal of Operational Research. 185: 1501-1524.
* Valente de Oliveira, J., & Pedrycz, W. (2007). Advances in fuzzy clustering and its applications. Chichester: John Wiley & Sons Ltd.
* Veerabathiran R., et al, Application of the Extent Analysis Method on fuzzy AHP, International Journal of Engineering Science and technology, Vol. 4, No. 09, July 8018.
* Wang, X., Wang, Y., & Wang, L. (2004). Improving fuzzy c-means clustering based on feature-weight learning. Pattern Recognition Letters, 25 (2), 1123–1132.

* Wen-Shiung Lee, Combined MCDM techniques for exploring stock selection based on Gordon model, Expert Systems with Applications, 8002.
* Xia, Yusen.; Liu, Baoding., Wang, Shouyang. & Lai, K.K. (1999). A Model for Portfolio Selection with Order of Expected Returns, Computers & OperationsResearch, www.armaninvestment.com
* Young, M.R. (1998). A Minimax – Portfolio Selection Rule with Linear Programming Soulotion, Management Science, 44. 673-683.

یادداشت‌ها

[1] Hachloufi
[۲] Knowledge Discovery in Databases فرایند حقیقی داده کاوی تحلیل خودکار یا نیمه خودکار مقادیر کثیر دادگان برای استخراج الگوهای از پیش شناخته نشده همچون خوشه‌بندی (تحلیل خوشه ای) رکوردهای دادگان است.
[۳] Data Mining Curriculum: داده کاوی مرحله تجزیه و تحلیل فرایند "کشف دانش در پایگاه داده‌ها" یا "From Data Mining to"KDD می‌باشد.
[4] A density-based algorithm for discovering clusters in large spatial databases with noise.
[5] Density-based clustering
[6] A graph-theoretic definition of a sociometric clique
[7] Data mining : concepts and techniques
[8] Characterization and evaluation of similarity measures for pairs of clusterings
[9] The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data
[10] The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data